



ESCUELA SUPERIOR DE INGENIEROS
UNIVERSIDAD DE SEVILLA



Ingeniería Superior de Telecomunicación

Proyecto Fin de Carrera

**FUNDAMENTOS Y REVISIÓN DEL ESTADO
DE LA TÉCNICA EN LA REALIZACIÓN
DE SENSORES DE IMAGEN EN
TECNOLOGÍAS CMOS CONVENCIONALES**

Autor del Proyecto:

Sergio Masferrer Oncala

Director del Proyecto:

Servando Espejo Meana

(Dpto. de Electrónica y Electromagnetismo Univ. de Sevilla
Instituto de Microelectrónica de Sevilla, CNM, CSIC)

Sevilla, Noviembre 2002

*A las tres personas para las que este Proyecto
significa tanto como para mí*

1.- SENSORES DE IMAGEN: EVOLUCIÓN HISTÓRICA Y CONCEPTOS BÁSICOS	1
1.1.- EVOLUCIÓN HISTÓRICA	1
1.2.- CONCEPTOS BÁSICOS	3
1.3.- ESTRUCTURA Y MOTIVACIÓN DEL PROYECTO	5
1.3.1.- Estructura	5
1.3.2.- Objetivos y Motivación	6
2.- ABSORCIÓN ÓPTICA EN UN SEMICONDUCTOR	9
2.1.- ESTADOS ENERGÉTICOS DEL ELECTRÓN. DEFINICIÓN DE FOTÓN	9
2.2.- BANDAS DE ENERGÍA	9
2.3.- SEMICONDUCTORES INTRÍNSECOS Y EXTRÍNSECOS	12
2.4.- DENSIDAD DE PORTADORES DE CARGA EN EQUILIBRIO	13
2.5.- CORRIENTES DE ARRASTRE Y DE DIFUSIÓN	16
2.6.- PROCESOS DE GENERACIÓN Y RECOMBINACIÓN DE PARES ELECTRÓN-HUECO	18
2.7.- RESPUESTA TRANSITORIA Y FOTOCONDUCTIVIDAD	20
2.8.- COEFICIENTE DE ABSORCIÓN Y EFICIENCIA CUÁNTICA	22
2.9.- ECUACIÓN DE CONTINUIDAD EN UN SEMICONDUCTOR	24
2.10.- UNIÓN pn	25
2.11.- UNIÓN p+n POLARIZADA	27
2.12.- CARACTERÍSTICA TENSIÓN-INTENSIDAD EN DC DE LA UNIÓN pn INVERSAMENTE POLARIZADA	29
3.- DISPOSITIVOS CMOS FOTOSENSIBLES	33
3.1.- FOTODIODOS	33
3.1.1.- Modos de operación del fotodiodo	34
3.1.1.1.- <i>Modo de operación fotocorriente</i>	34
3.1.1.2.- <i>Modo fotovoltaico de operación</i>	35
3.1.1.3.- <i>Modo de operación por integración de carga</i>	36
3.1.2.- Fotodiodos implementables en tecnología CMOS estándar	37
3.2.- FOTOTRANSISTORES	39
3.3.- FOTOPUERTAS	40
3.4.- SENSIBILIDAD ESPECTRAL Y EFICIENCIA CUÁNTICA DE LOS ELEMENTOS FOTODETECTORES	42
3.5.- OTROS ASPECTOS IMPORTANTES EN LOS SENSORES DE IMAGEN	46
3.5.1.- Dispositivos fotosensores implementados en tecnologías CMOS no estándares	46
3.5.1.1.- <i>Proceso de fabricación BiCMOS</i>	47
3.5.1.2.- <i>Fotodiodos "pinned"</i>	47
3.5.1.3.- <i>Fotodiodos fabricados en tecnología TFA</i>	49
3.5.2.- Medición de la calidad de un sistema óptico. Función de transferencia de la modulación de la señal óptica	50
3.5.3.- Absorción de fotones en las capas depositadas sobre la oblea	54
3.5.4.- Fenómeno de múltiples reflexiones	57
3.5.5.- Conclusiones de esta sección	59

4.- ADAPTACIÓN DE LA SEÑAL FOTOGENERADA	61
4.1.- FOTSENSORES EN TIEMPO DISCRETO Y EN TIEMPO CONTINUO	61
4.1.1.- Fotosensores en tiempo discreto	61
4.1.2.- Fotosensores en tiempo continuo	63
4.1.3.- Fotosensores basados en el modo de operación fotovoltaico	66
4.2.- PÍXELES PASIVOS Y ACTIVOS	67
4.2.1.- Sensor basado en píxeles pasivos (PPS)	67
4.2.2.- Sensor basado en píxeles activos (APS)	70
4.2.2.1.- APS con salida en tensión en tiempo discreto	71
4.2.2.2.- APS con salida en tensión en tiempo continuo	72
4.3.- CIRCUITOS DE ADAPTACIÓN DE LA CARGA FOTOGENERADA	73
4.3.1.- APS por integración de carga con salida diferencial directa	73
4.3.2.- Eliminación programable del offset	75
4.3.3.- Píxel con ganancia programable	77
4.3.4.- Circuito receptor adaptativo	78
4.3.5.- Píxel activo en modo corriente con resolución variable	80
4.4.- OTRAS MODULACIONES DE LA SEÑAL LUMINOSA	81
4.4.1.- Modulación por ancho de pulso (PWM)	81
4.4.2.- Modulación frecuencial del pulso (PFM)	82
5.- RUIDO EN LOS SENSORES DE IMAGEN CMOS	85
5.1.- INTRODUCCIÓN	85
5.1.1.- Fuentes de ruido temporal	85
5.1.2.- Fuentes de ruido espacial	86
5.2.- FUENTES DE RUIDO TEMPORAL	88
5.2.1.- Valor cuadrático medio y densidad espectral de potencia	88
5.2.2.- Ruido térmico	89
5.2.3.- Ruido de disparo	89
5.2.4.- Ruido flicker o ruido a baja frecuencia	90
5.3.- MODELO DE RUIDO TEMPORAL DEL DIODO Y DEL TRANSISTOR MOS	92
5.3.1.- Modelo de ruido temporal del diodo	92
5.3.2.- Modelo de ruido temporal del transistor	93
5.3.2.1.- Ruido térmico	93
5.3.2.2.- Ruido flicker	95
5.4.- ANÁLISIS DE RUIDO EN LOS SENSORES DE IMAGEN	96
5.4.1.- Análisis del ruido en los sensores PPS	96
5.4.1.1.- Ruido generado durante la fase de integración	97
5.4.1.2.- Ruido generado durante la fase de reset	98
5.4.1.2.1.- <u>Potencia de ruido a la salida debido al transistor de reset</u>	99
5.4.1.2.2.- <u>Potencia de ruido a la salida debido al amplificador y al ruido del substrato</u>	101
5.4.1.3.- Ruido generado durante la fase de lectura	101
5.4.1.3.1.- <u>Potencia de ruido a la salida debido al transistor de acceso</u>	102
5.4.1.3.2.- <u>Potencia de ruido a la salida debido al amplificador operacional y al ruido de substrato</u>	103
5.4.1.3.3.- <u>Ruido realimentado en el fotodiodo 103</u>	
5.4.2.- Análisis del ruido en los sensores APS por integración de carga	104
5.4.2.1.- Ruido generado durante la fase de reset	106
5.4.2.2.- Imagen residual generada por un reset incompleto	110
5.4.2.3.- Ruido durante la fase de integración	110
5.4.2.4.- Ruido generado durante la fase de lectura	111
5.5.- TÉCNICAS DE REDUCCIÓN DEL RUIDO TEMPORAL	112
5.5.1.- Reducción del ruido intrínseco	112
5.5.2.- Reducción del ruido de reset en los sensores APS CMOS por integración de carga	114
5.5.2.1.- Técnica "pseudo-flash"	114

5.5.2.2.- <i>Reset a través de una llave pMOS y una carga logarítmica</i>	114
5.5.2.3.- <i>Técnica de reset activo</i>	115
5.6.- RUIDO ESPACIAL EN LOS SENSORES APS CMOS EN TIEMPO DISCRETO	117
5.7.- TÉCNICAS DE REDUCCIÓN DEL RUIDO ESPACIAL	120
5.7.1.- ALGORITMOS PARA LA REDUCCIÓN DEL FPN	120
5.7.1.1.- <i>Muestreo diferencial a nivel de píxel</i>	120
5.7.1.2.- <i>Reducción del FPN en los APS en modo corriente</i>	121
5.7.1.3.- <i>Corrección del FPN en los sensores de imagen logarítmicos</i>	122
5.7.1.3.1.- <u><i>Arquitectura para la compensación del FPN en sensores logarítmicos</i></u>	123
5.8.- RUIDO MULTIPLICATIVO EN LOS SENSORES DE IMAGEN	124
6.- CONVERSIÓN ANALÓGICO-DIGITAL EN LOS SENSORES DE IMAGEN	129
6.1.- INTRODUCCIÓN	129
6.2.- ESQUEMAS BÁSICOS DE CONVERSIÓN ANALÓGICO-DIGITAL	131
6.2.1.- Convertidores flash	131
6.2.2.- CA/D de aproximaciones sucesivas	133
6.2.3.- CA/D de pendiente simple	133
6.2.4.- CA/D de doble pendiente	134
6.2.5.- Convertidores sigma-delta	135
6.2.6.- Técnica Multi-Channel Bit Serial (MCBS)	136
6.2.7.- Otras técnicas de conversión analógico-digital	138
6.2.7.1.- <i>Arquitecturas de dos etapas</i>	138
6.2.7.1.1.- <u><i>Arquitectura de reciclaje de dos etapas</i></u>	139
6.2.7.1.2.- <u><i>Arquitectura subranging de dos etapas</i></u>	140
6.2.7.2.- <i>Arquitecturas interpolativas y de plegado</i>	140
6.2.7.2.1.- <u><i>Interpolación</i></u>	140
6.2.7.2.2.- <u><i>Arquitecturas de plegado o "folding"</i></u>	142
6.2.7.3.- <i>Arquitecturas en "pipeline"</i>	142
6.3.- TÉCNICAS DE CONVERSIÓN ANALÓGICO-DIGITAL EN LOS SENSORES DE IMAGEN	144
6.3.1.- Implementación de la CA/D S-D a nivel de píxel	145
6.3.2.- Uso de un oscilador muestreado como modulador S-D de primer orden	146
6.3.3.- Implementación de la técnica MCBS a nivel de píxel	149
6.4.- OTROS ASPECTOS INTERESANTES SOBRE LA CA/D EN LOS SENSORES DE IMAGEN	150
6.4.1.- Técnica de muestreo múltiple en un sensor con CA/D MCBS a nivel de píxel	150
6.4.2.- Algoritmos de mejora del rango dinámico en sensores de imagen	152
6.4.2.1.- <i>Rango dinámico y SNR en sensores en tiempo discreto basados en integración de carga</i>	153
6.4.2.2.- <i>Mejora del rango dinámico mediante el ajuste de la capacidad del pozo</i>	155
6.4.2.3.- <i>Mejora del rango dinámico mediante la técnica de muestreo múltiple</i>	156
7.- RESUMEN Y LÍNEAS FUTURAS DE DESARROLLO	161
A.- SENSORES DE IMAGEN EN TECNOLOGÍAS CCD	163
A.1.- PÍXEL CCD	164
A.2.- PROCESO DE TRANSFERENCIA DE CARGA	165
B.- MODELO MATEMÁTICO DEL RUIDO	167
C.- ANÁLISIS DEL RUIDO EN CIRCUITOS LINEALES	171
D.- ANÁLISIS DEL RUIDO FLICKER EN TRANSISTORES DE CANAL CORTO	175

SENSORES DE IMAGEN: EVOLUCIÓN HISTÓRICA Y CONCEPTOS BÁSICOS

1.1.- EVOLUCIÓN HISTÓRICA

El desarrollo histórico de los sensores de imagen está intrínsecamente ligado al desarrollo de las tecnologías sobre las que este tipo de dispositivos se implementan. Éstas son, principalmente, la tecnología CCD (Charge-Coupled Devices) y las tecnologías MOS y CMOS.

En los años 60's comenzaron los intentos por lograr la implementación de sensores de imagen en tecnologías MOS, obteniendo diferentes grados de éxito en implementaciones con transistores nmos, pmos y bipolares. Por ejemplo, en 1963, Morrison[1] logró implementar una estructura que permitía determinar la posición de un punto de luz haciendo uso del efecto fotoconductor de los semiconductores. El "scanistor" [2], implementado por IBM en 1964, utilizaba una matriz de uniones npn direccionadas a través de una red resistiva para generar un pulso proporcional a la intensidad de luz incidente. En 1966, Westinghouse presentó una matriz de fototransistores de dimensiones 50x50 [3]. En el aspecto tecnológico, todos estos sensores generaban una salida proporcional a la intensidad instantánea de luz incidente, sin llevar a cabo ninguna adaptación de la señal óptica. Como consecuencia, la sensibilidad de estos dispositivos era baja, y se hacía necesaria la incorporación de una ganancia a nivel de píxel para mejorar las características del mismo.

En 1967, Weckler y Fairchild sugirieron el uso de uniones pn en modo por integración de carga [4]. Con esta técnica, la carga fotogenerada en la unión pn se integra en la capacidad parásita del mismo dispositivo que forma la unión. La carga fotogenerada podría convertirse en tensión mediante el uso de elementos resistivos.

Al mismo tiempo que era propuesta la integración de la carga fotogenerada, RCA publicó un sensor basado en un transistor de capa fina (TFT o Thin-Film Transistor) [5]. La matriz de píxeles implementada tenía unas dimensiones de 180x180 elementos, e incluía una circuitería lógica para llevar a cabo el direccionamiento secuencial de los píxeles.

En 1968, Noble describió varias configuraciones de matrices de sensado de imagen [6]. En este artículo se describen tanto los fotodiodos superficiales como la posibilidad de la implementación de los mismos enterrados para la reducción de la corriente oscura en estos dispositivos. Propuso llevar a cabo la lectura del píxel a través de un amplificador de carga en lugar de hacerlo simplemente a través de un elemento resistivo. También propuso por primera vez el uso a nivel de píxel de un transistor MOS en configuración de seguidor por fuente para realizar la lectura del píxel, estableciendo la teoría de lo que posteriormente se conocerá como píxeles activos (APS: Active Pixel Sensors), en contraste con las anteriores configuraciones, consideradas pasivas (PPS: Pasive Pixel Sensors).

Con la introducción del amplificador por carga se plantearon los problemas de desapareamiento entre los dispositivos que integran el sensor. Así, la descripción del ruido de patrón fijo se llevó a cabo en 1970 [7].

En respuesta a los problemas que presentaban los sensores de imagen en tecnologías MOS respecto al ruido (principalmente al ruido de patrón fijo), y a la incapacidad de esta tecnología de solventar esta barrera, en 1970 se lleva a cabo la descripción de los sensores implementados en tecnología CCD [8]. Estos sensores se caracterizaban por la relativa ausencia de ruido de patrón fijo, lo que supuso la principal adopción de esta técnica por muchas otras arquitecturas de sentido de imagen. Además de este factor, en comparación con las escalas de integración de las tecnologías MOS y CMOS, el tamaño del píxel en tecnología CCD era menor, lo que permitía alcanzar una mayor resolución en el sensor.

Desde la implantación de los sensores CCD, éstos han liderado indiscutiblemente el campo de las cámaras electrónicas en cualquier tipo de aplicación. Durante muchos años, las demandas de mercado han conducido a mejoras impresionantes en el desarrollo de las tecnologías CCD, que han respondido ofreciendo una escala de integración y una calidad de imagen que otras tecnologías no podían ofrecer mejor. Los sensores de imagen en tecnología CCD, desde el primer momento hasta hace relativamente poco, han tenido mejores características (o en el peor de los casos equiparables) que los sensores implementados en tecnologías MOS en factores tan determinantes como la eficiencia cuántica, el factor de relleno, la corriente oscura, la eficiencia en la transferencia de carga, la frecuencia de lectura, la carga residual, el ruido de lectura y el rango dinámico.

Mientras se aplicaba un gran esfuerzo en el desarrollo de los CCDs a lo largo de los 70's y de los 80's, los sensores de imagen en tecnología MOS se investigaban sólo de forma esporádica, y eran comparados de forma desfavorable con los CCDs respecto a los factores citados anteriormente. A finales de los 70's y principio de los 80's Hitachi y Matsushita continuaron con el desarrollo de los sensores de imagen en tecnología MOS [9], [10]. Sobre 1985, Hitachi combinó el sensor MOS con un registro de desplazamiento horizontal CCD [11]. En 1987, Hitachi introdujo una técnica simple on-chip para conseguir tiempos de exposición variables y eliminación del ruido flicker [12]. Sin embargo, quizás debido al ruido temporal residual, especialmente importante en condiciones de baja iluminación, Hitachi abandonó las aproximaciones que había realizado sobre los sensores de imagen en tecnología MOS.

Según Eric E. Fossum [13], a principios de los 90's, dos esfuerzos independientes han conducido al resurgimiento de los sensores de imagen en tecnología CMOS. El primero de los esfuerzos fue la creación de sistemas de visión de alta funcionalidad en los que el bajo coste era el factor determinante. El segundo esfuerzo, independiente del primero, fue consecuencia de las necesidades de la NASA por la implementación de sistemas de imagen miniaturizados y de bajo consumo como instrumental en sus equipos espaciales. Tales sistemas de imagen están limitados por la implementación de los dispositivos, y no por el coste de producción. Este esfuerzo fue llevado a cabo por la U. S. Jet Propulsion Laboratory (JPL), con la consecuente transferencia de la tecnología a AT&T Bell Labs, Kodak, National Semiconductors y varias compañías grandes norteamericanas, tales como Photobit. La convergencia de estos esfuerzos ha conducido a avances importantes en los sensores de imagen y al desarrollo de los sensores CMOS basados en píxeles activos (APS CMOS). Se han logrado implementaciones que pueden competir con los CCDs en aspectos tan claves como el ruido de lectura, el rango dinámico y la responsividad, pero con una vasta ventaja en cuanto a funcionalidad, menor consumo y el potencial bajo coste de producción de las tecnologías CMOS convencionales frente a las tecnologías CCD.

El giro de una tecnología a otra, al igual que ocurre con los procesos de selección natural, ha sido resultado de la especialización de las tecnologías CCD. Si bien estas tecnologías supusieron en un principio un avance importante en la implementación de sensores de imagen al ser concebidas para llevar a cabo esta función, a medida que se ha ido exigiendo versatilidad a los sensores de imagen, éstas no han podido responder, pues la funcionalidad que en estas tecnologías se puede añadir está muy limitada. Por otro lado, el desarrollo de las tecnologías CMOS en las últimas dos décadas ha sido exponencial en cuanto al grado de escalabilidad y de funcionalidad, lo que ha permitido que el diseño de sensores en tecnología CMOS resulte muy versátil: permite la incorporación de una circuitería relativamente compleja a nivel de píxel que corrige los defectos que esta tecnología tiene frente a las tecnologías CCD. Esto está provocando el giro en la comparación de las características de unos y otros sensores. Ya no puede afirmarse que los sensores CCDs aventajen a los sensores CMOS en los factores antes citados (eficiencia cuántica, factor de relleno, etc); de hecho, la tendencia actual posiblemente desemboque en la monopolización del mercado de cámaras digitales por las tecnologías CMOS.

1.2.- CONCEPTOS BÁSICOS

El principal reto en el diseño de un sensor de imagen es la incorporación de la mayor funcionalidad posible a nivel de píxel. Es decir, de forma óptima, el diseñador persigue llevar a cabo la corrección de errores, la amplificación de la carga fotogenerada y la conversión analógico-digital a nivel de píxel, es decir, de forma independiente para cada píxel. De esta manera se consiguen mejorar las características del sensor, tales como la minimización del ruido de patrón fijo, la mejora de la conversión de señal luminosa en eléctrica, la eliminación de la carga residual, y el aumento del rango dinámico.

Sin embargo, para la obtención de una buena calidad de imagen es imprescindible cumplir determinados requisitos de resolución de la imagen. Esto obliga a que la superficie de integrado dedicada a cada píxel sea bastante reducida. He aquí el principal problema del diseñador: demasiada circuitería para tan poco espacio.

El compromiso de diseño entre funcionalidad y superficie ocupada establece diferentes niveles en la implementación de la circuitería en los sensores de imagen: nivel de sistema; nivel de columna; y nivel de píxel (Fig. 1.1).

El sensor estará constituido por una matriz de píxeles de dimensiones $n \times m$, de forma que la **circuitería a nivel de sistema** serán todos aquellos dispositivos que estén compartidos por todos los píxeles. Esta circuitería deberá ser rápida, pues deberá responder a todos los píxeles de forma secuencial para cada imagen. En su diseño la superficie ocupada no será el principal límite de diseño.

En general, el esquema de lectura que siguen los sensores de imagen consiste en realizar el volcado sucesivo de los valores de los píxeles de la matriz por las filas. Este esquema de lectura nos lleva a definir una **circuitería a nivel de columna**, consistente en todos los dispositivos que sean compartidos por una misma columna de la matriz. La rapidez de los dispositivos deberá ser menor que la de la circuitería a nivel de sistema, pero el área disponible para su implementación se ve reducida.

Finalmente, existirán una serie de elementos que se implementen en cada píxel, definiendo así la **circuitería a nivel de píxel**. Los tiempos de operación de esta circuitería serán relativamente altos, pero la superficie disponible es muy reducida.

Los tres niveles de jerarquización quedan conectados como sigue: la circuitería a nivel de píxel hará uso del *bus de columna* para volcar la salida del píxel en la circuitería de columna; ésta a su vez hace uso del *bus de sistema* para volcar su salida sobre la circuitería a nivel de sistema.

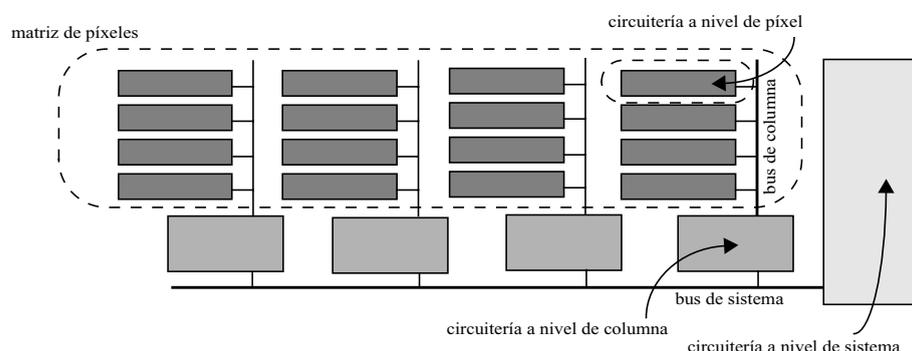


Fig. 1.1: Jerarquización de la circuitería en un sensor de imagen

Definida esta jerarquización en la circuitería de los sensores de imagen, pueden definirse diferentes estrategias que dan lugar a las diferentes arquitecturas en el diseño de los sensores de imagen. Así, se define como *píxel pasivo* a aquella estructura que a nivel de píxel incluye únicamente el elemento fotodetector y una llave para el direccionamiento del píxel. La estrategia de diseño en este tipo de sensores es aprovechar toda la superficie del píxel para la fotodetección, con lo que se pretende conseguir maximizar la cantidad de fotones capturados. La circuitería que lleva a cabo la detección electrónica de la carga fotogenerada por estos fotones, así como todas las demás funciones que realiza el sensor, se implementa, o bien a nivel de columna, o bien a nivel de sistema. Los sensores que responden a esta arquitectura se denominan *sensores basados en píxeles pasivos* o *PPS* (Passive Pixel Sensors). La incorporación de un amplificador a nivel de píxel da lugar a los *píxeles activos*. La estrategia de diseño consiste en reducir el área de fotosensado para incorporar circuitería electrónica que permita mejorar las características del sensor. Dichas arquitecturas se denominan *sensores basados en píxeles activos* o *APS* (Active Pixel Sensors). La alta escala de integración ha permitido definir el *sensor basado en píxeles digitales* (*DPS*) [14]. La complejidad de la circuitería a nivel de píxel en estos esquemas es muy elevada: cada píxel incluye una fotopuerta nmos, puertas de transferencia y de reset, una capacidad de almacenamiento, un comparador de tres etapas y 8 celdas de memoria dinámica de un bit constituidas cada una por tres transistores.

El desarrollo de sensores está dirigido tanto por una estrategia de diseño como por una limitación tecnológica. Actualmente, las escalas de integración de las tecnologías CMOS permiten implementar el amplificador a nivel de píxel y muchas más funciones, por lo que las arquitecturas PPS¹ han quedado prácticamente aparcadas.

Un concepto que cabe introducir, pues a lo largo del proyecto se hace extensa referencia a él es el *ruido de patrón fijo* o *FPN* (Fixed Pattern Noise). El ruido de patrón fijo es un ruido de naturaleza espacial, independiente del tiempo, que surge como resultado de las diferencias entre los parámetros tecnológicos de los dispositivos que se implementan en el sensor. Cuando se implementa el sensor, la celda básica (el píxel) se repite a lo largo de toda la matriz de sensado. Todos los píxeles están diseñados de forma idéntica, pero cada uno presentará una desviación de su valor nominal debido a las variaciones en el proceso tecnológico generadas por el desapa-

1. La principal ventaja de los sensores CMOS frente a los CCD es la compleja circuitería que permite añadir. Los PPS no aprovechan esta ventaja.

reamiento entre dispositivos. Igual ocurrirá entre las circuiterías a nivel de columna y a nivel de sistema. Este ruido es de especial importancia en los sensores de imagen, pues distorsiona la imagen final capturada, y será necesario el desarrollo de algoritmos para su corrección.

1.3.- ESTRUCTURA Y MOTIVACIÓN DEL PROYECTO

1.3.1.- Estructura

Expuesta la evolución histórica de las tendencias tecnológicas en la realización de sensores de imagen, el principal de este proyecto consiste en la realización de un análisis fundamentalista de las estructuras básicas utilizadas en el sensado de imagen, así como llevar a cabo una exposición de las arquitecturas actuales más generalizadas, centrandolo en torno a las tecnologías CMOS convencionales.

Para cumplir tales objetivos, el trabajo se organiza como sigue.

En el capítulo 2 se lleva a cabo un análisis resumido de las *características de los semiconductores*, y se justifica el porqué de su utilidad para el sensado de imagen. Se describe cómo se lleva a cabo la *absorción de los fotones incidentes*, se describe la unión pn y cómo aprovechar el comportamiento de la misma ante la incidencia de un flujo luminoso. Se definen conceptos tales como la *profundidad de penetración* y la *eficiencia cuántica*.

En el capítulo 3 se lleva a cabo una descripción y un análisis comparativo de los diferentes *dispositivos fotosensores*. Las propiedades de estos dispositivos dependerán básicamente de cómo implementen la unión pn, entrando en juego factores tales como la profundidad a la que se localice la unión, si ésta consiste en una unión pn simple (fotodiodo) o se forma a partir de una unión npn o pnp (es decir, a partir de un bipolar vertical), etc. El capítulo se centra principalmente en las tecnologías CMOS convencionales, pues es el principal objetivo de este proyecto, pero también se tantean otras tecnologías que, por sus características, resultan interesantes para aplicaciones de sensado de imagen. Se lleva a cabo también la definición de los parámetros que sirven para caracterizar a los sensores de imagen, tales como el *rango dinámico*, la *relación señal a ruido*, y la función de transferencia de la modulación (*MTF*).

En el capítulo 4 se analizan las diversas formas de *adaptación de la carga fotogenerada* en el dispositivo fotosensor. Entran en juego conceptos como adaptación de la carga fotogenerada en *tiempo discreto* y en *tiempo continuo*; se introducen en detalle los conceptos de *píxeles activos* y *pasivos* antes mencionados; se describe el principio de operación de los *sensores por integración de carga* y *logarítmicos*; y se proponen arquitecturas implementadas a nivel de píxel que tienen como objetivo añadir funcionalidad para la mejora de las características.

En el capítulo 5 se realiza un análisis de las distintas *componentes de ruido*. Para ello se distingue entre *ruido temporal* y *ruido espacial*. Para el ruido temporal se analizan las diferentes fuentes de ruido temporal (térmico, de disparo, flicker,...) de forma general, y se lleva a cabo una particularización de este análisis para los sensores de imagen en tiempo discreto. Respecto al ruido espacial, éste provocará una variación espacial en la imagen de salida, en condiciones de iluminación uniforme (constante en el espacio). Las fuentes principales de este ruido espacial serán las diferencias en las ganancias y los errores de desapareamiento existentes entre los dispositivos implementados. En esta categoría se incluye el *ruido de patrón fijo (FPN)*, del que se lleva a cabo una definición, se propone un modelo estadístico y se presentan algoritmos para su eliminación.

En el capítulo 6 se presentan las arquitecturas básicas para llevar a cabo la *conversión analógico-digital*, y se describen algunas arquitecturas de *moduladores* Σ - Δ y de esquemas *MCBS* (Multi Channel Bit Serial) que consiguen implementar dicha conversión a nivel de píxel.

1.3.2.- Objetivos y Motivación

El objetivo de este Proyecto de Fin de Carrera es realizar una revisión del estado de la técnica en lo referente al diseño de sensores de imagen en tecnologías CMOS. En este sentido, el proyecto es diferente a lo habitual, pues no persigue una realización o diseño concreto, sino que su resultado consiste en un “informe”, necesariamente resumido, de las características y prestaciones esperables, las ventajas e inconvenientes frente a otras alternativas, las dificultades y limitaciones más relevantes, y las técnicas de diseño reportadas para superarlas o atenuarlas, en el campo específico de los sensores de imagen CMOS.

Aparte de satisfacer el requisito académico asociado, la motivación esencial para la realización de este Proyecto ha sido la de adquirir un sustrato de conocimiento profundo en esta materia, en forma de revisión bibliográfica y comprensión de los aspectos esenciales del problema, como paso inicial de unos estudios más profundos en este campo. Este enfoque responde a un deseo claro de realización de estudios de doctorado, bien definido desde el comienzo del trabajo de este Proyecto.

El particular enfoque, con sus ventajas en lo referente al objetivo final, supone también algunos inconvenientes. Por una parte, no es fácil comprender y asimilar en detalle las limitaciones y dificultades técnicas de circuitos poco convencionales, sin realizar previamente un trabajo detallado de análisis y diseño, al menos en algunos casos concretos. En particular, la comprensión de los circuitos empleados en las configuraciones más comunes de cámaras CMOS actuales requiere manejar señales fotogeneradas (y por tanto procesos de absorción de fotones y generación en semiconductores), emplean transistores operando en inversión fuerte y débil, requieren análisis de ruido que van bastante más allá del simple análisis lineal de pequeña señal, incluyendo casos no lineales, en estado no estacionario, y también en tiempo discreto, y precisan del análisis de otros efectos de segundo orden relativamente complejos para el nivel de un recién titulado, tales como el despareamiento de diodos y de transistores MOS en todas sus zonas de inversión. Por otra, la relevancia actual del tema y su rápida evolución en los últimos años, hace que las fuentes bibliográficas existentes sean muy abundantes pero a la vez muy dispersas y en general organizadas desde puntos de vista muy dispares. Esto, unido a un objetivo definido de manera “difusa”, hacen que no esté muy claro hasta dónde se debe llegar en el desarrollo del Proyecto.

Sin embargo, no debe perderse de vista que lo contenido en este documento responde a un Proyecto Fin de Carrera, aunque sus objetivos y su enfoque apunten a la realización de un trabajo de mayor envergadura.

Referencias



- [1] S. Morrison "A new type of photosensitive junction device" *Solid-State Electron.*, vol. 5, pp. 485-494, 1963
- [2] J. Horton, R. Razza, H. Dym "The scanistor- A solid-state image scanner" in *Proc. IEEE*, vol. 52, pp. 1513-1528, 1964
- [3] M. A. Schuster, G. Strull "A monolithic mosaic of photon sensors for solid state imaging applications" *IEEE Trans. Electron Devices*, vol. ED-13, pp. 907-912, 1966
- [4] G. P. Weckler "Operation of p-n junction photodetectors in a photon flux integration mode" *IEEE J. Solid-State Circuits*, vol. SC-2, pp.65-73, 1967
- [5] P. K. Weimer, G. Sadavis, J. Meyer, L. Meray-Horvath, W. Pike "A self-scanned silicon image sensor" in *Proc. IEEE*, vol. 55, pp.1591-1602, 1967
- [6] P. Noble "Self-scanned silicon image detector arrays" *IEEE Trans. Electron Devices*, vol. ED-15, pp. 250-209, Apr. 1968
- [7] P. Fry, P. Noble, R. Rycroft "Fixed Pattern Noise in photomatrices", *IEEE J. Solid-State Circuits*, vol. SC-5, pp250-254, May 1970.
- [8] W. S. Boyle, G. E. Smith "Charge-coupled semiconductors devices" *Bell Syst. Tech. J.*, vol. 49, pp. 587-593, 1970
- [9] S. Ohba et al. "MOS area sensor: Part II- Low Noise MOS area sensor with anti-blooming photodiodes" *IEEE Trans. Electron Devices*, vol. ED-27, pp. 1628-1687, Aug. 1980
- [10] K. Senda, S. Terakawa, Y. Hiroshima, T. Kunii "Analysis of charge-priming transfer efficiency in CPD image sensors" *IEEE Trans. Electron Devices*, vol. ED-31, pp. 1324-1328, Sept. 1984
- [11] H. Ando et al. "Design consideration and performance of a new MOS imaging device" *IEEE Trans. Electron Devices*, vol. ED-32, p. 1484-1489, May 1985
- [12] T. Kinugasa et al. "An electronic variable shutter system in video camera use", *IEEE Trans. Consumer Electron.*, vol. CE-33, pp. 249-255, 1987
- [13] E. R. Fossum "CMOS Image Sensors: Electronic Camera-On-A-Chip" *IEEE Trans. Electron Devices*, vol. 44, no. 10, Oct. 1997
- [14] S. Kleinfelder, S. Lim, X. Liu, A. El Gamal "A 10,000 frames/s 0.18 um CMOS Digital Pixel Sensor with Pixel-Level Memory", *Solid State Circuits Conference*, Feb. 5, 2001

ABSORCIÓN ÓPTICA EN UN SEMICONDUCTOR

2.1.- ESTADOS ENERGÉTICOS DEL ELECTRÓN. DEFINICIÓN DE FOTÓN

Según el modelo atómico de Bohr¹, para el caso del hidrógeno en el que existe un solo electrón por átomo, el momento de éste sólo puede tomar una serie de valores discretizados, lo que da lugar a que la energía del electrón tome valores dentro de un conjunto de estados energéticos discretizado. Estos estados energéticos vienen definidos por la relación:

$$E_n = -\frac{q^4 m}{8h^2 \epsilon_0^2 n^2} \quad (2.1)$$

donde E_n es la energía del electrón, q la carga del electrón, ϵ_0 es el valor de la permitividad eléctrica del vacío, m es la masa del electrón en el espacio libre, h la constante de Planck, y n un número entero positivo.

De acuerdo con este modelo atómico el electrón, siempre que se mantenga confinado en una órbita con un determinado nivel energético asociado, no podrá radiar ni absorber energía. La radiación energética se producirá cuando el electrón salte desde una órbita de energía E_2 hasta otra de menor energía E_1 , y consistirá en la liberación de un cuanto de energía, o **fotón**, cuya energía vendrá dada por:

$$E_2 - E_1 = h \cdot \nu \quad (2.2)$$

donde ν es la frecuencia de la energía radiada. La cantidad $h \cdot \nu$ se denomina **energía del fotón**. De igual forma, un electrón con un determinado nivel energético puede saltar a una órbita de mayor energía si se produce la absorción de un fotón de energía $h \cdot \nu$.

2.2.- BANDAS DE ENERGÍA

El modelo de estados energéticos discretizados que hemos considerado es aplicable a elementos en estado gaseoso, en los que cada átomo está suficientemente separado de los demás como para que no interactúen. Sin embargo, en los sólidos los átomos no se mueven libremente, sino que se encuentran agrupados de forma que se produce la interacción entre ellos. Todos los átomos de forma individual tienen idénticos estados energéticos, como se muestra en la Fig. 2.1(a) para el caso del hidrógeno. A medida que dos átomos se acercan, las órbitas de los átomos comienzan a solaparse (Fig. 2.1(b)). De la extrapolación de esta interacción entre dos átomos a un sólido en el que existe una multitud de átomos cuyas bandas de energía interactúan surge la definición de **bandas de energía permitidas**. Para un sólido, una banda de energía permitida vendrá definida por un conjunto de estados energéticos que el electrón puede tomar. Las **bandas de energía prohibidas** estarán definidas en el sólido por los valores de energía que el electrón no

1. El modelo atómico de Bohr es demasiado simple, pero nos sirve para ilustrar el propósito de esta sección

puede tomar. Cada banda permitida contendrá un número limitado de estados que podrán acomodar a un número definido de electrones.

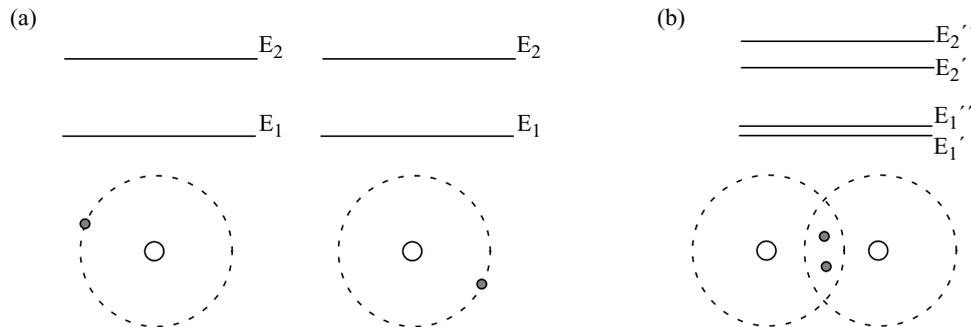


Fig. 2.1: Estados energéticos de dos átomos de hidrógeno (a) aislados; e (b) interactuando

En un semiconductor los electrones de valencia² ocupan un conjunto de estados energéticos que definen la **banda de valencia**. La banda energética permitida más próxima a la de valencia se denomina **banda de conducción**, y está separada de la banda de valencia por una banda de estados energéticos prohibidos (**banda prohibida**), de ancho E_g . Esta representación de los estados energéticos para un material se denomina **diagrama de bandas**. La Fig. 2.2 compara el diagrama de bandas de un metal, de un semiconductor, y de un material aislante.

En los metales la banda de conducción está sólo parcialmente ocupada por electrones. Las bandas de valencia y conducción se solapan, por lo que en los metales puede considerarse que no existe banda prohibida. En los materiales semiconductores la banda de valencia tiene los estados suficientes para acomodar a todos los electrones de valencia proporcionados por los átomos del material. Si no existen estados energéticos excitados la banda de valencia está completa, mientras que la banda de conducción queda vacía de electrones.

Una banda completa no contribuye al flujo de corriente en el material. Por lo tanto es necesario la absorción de una energía (proporcionada, como ejemplo, en forma de energía térmica, o mediante la incidencia de fotones sobre el material) mayor que la energía de la banda prohibida para que se generen estados energéticos excitados (electrones en la banda de valencia que pasan a ocupar estados energéticos en la banda de conducción) para que la carga contribuya al flujo de corriente en el material. La probabilidad de que ocurra la transición entre estados energéticos es pequeña cuando la energía E_g es grande, como ocurre con los materiales aislantes, caracterizados por un ancho de la banda prohibida mayor de 5 eV.

En los semiconductores el ancho de la banda prohibida no es tan grande. Por ejemplo, para el silicio $E_g=1.12$ eV, mientras que para el GaAs es de 1.43 eV. La temperatura ambiente es suficiente para que electrones en la banda de valencia reciban suficiente energía térmica para alcanzar la banda de conducción. La transición de un electrón de la banda de valencia a la banda de conducción genera

2. Una descripción profunda de la estructura electrónica de los elementos químicos puede encontrarse en [1]

un **huevo** en la banda de valencia. Un huevo se define como la existencia de un estado energético no ocupado en la banda de valencia. La aplicación de un campo eléctrico proporcionará una energía cinética a los electrones que se encuentren en la banda de conducción (generando un flujo de corriente de electrones) y a los huecos que se encuentren en la banda de valencia (generando un flujo de corriente de huecos). Por lo tanto, la conducción eléctrica en un semiconductor tiene dos componentes: el transporte de carga negativa (electrones de la banda de conducción), y el transporte de carga positiva (huecos en la banda de valencia).

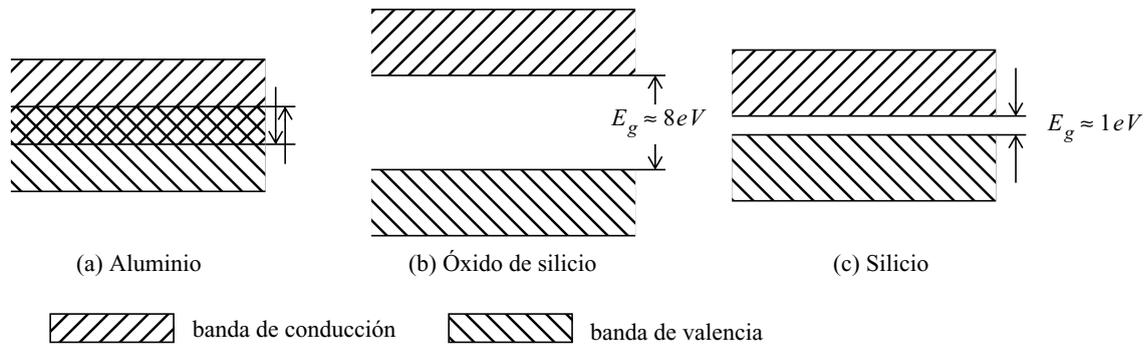


Fig. 2.2: Diagrama de bandas de: (a) un metal; (b) un material aislante; y (c) un semiconductor

Los electrones y huecos en un semiconductor son portadores de carga de valor $-q$ y $+q$ respectivamente (siendo q la carga del electrón: 1.6×10^{-19} C). Cada partícula estará caracterizada por una masa, diferente de la masa del electrón en el espacio libre m_0 . La masa del portador de carga depende de la banda de energía que ocupe, del momento de la partícula, así como de la dirección del campo eléctrico aplicado en el semiconductor. En la mayoría de los casos la masa de los portadores de carga resulta ser menor que la masa en el vacío m_0 , lo que nos obliga a tener en cuenta factores no descritos por la mecánica clásica. Una de las bases sobre las que se apoya la mecánica cuántica es en la dualidad onda-partícula, la cual permite la descripción del electrón³ como una onda de una determinada longitud de onda λ , con un determinado momento asociado p . La ecuación de De Broglie describe la relación entre la longitud de onda y el momento de la partícula:

$$p = \frac{h}{\lambda} = \frac{h \cdot k}{2\pi} \quad (2.3)$$

donde $k = 2\pi/\lambda$ es el **número de onda** de la partícula.

Las ecuaciones que describen el comportamiento del electrón⁴ de forma clásica pueden incluir la naturaleza cuántica del mismo asumiendo que el electrón tiene una masa efectiva m_e , distinta de la masa del electrón en el espacio libre. El electrón puede considerarse una partícula clásica en el seno de un semiconductor, para el que la relación entre energía y momento viene dada por:

$$E = \frac{p^2}{2m_e} = \frac{h^2 k^2}{8\pi^2 m_e} \quad (2.4)$$

donde E es la energía cinética del electrón. Puesto que energía y momento están relacionados de forma compleja, no podemos hacer uso de la expresión 2.4 para calcular m_e . Bajo consideraciones de mecánica cuántica, m_e se obtiene de la expresión⁵:

3. La dualidad onda-partícula es aplicable no sólo al electrón, sino a cualquier partícula.
4. Para ser más exactos deberíamos decir electrones libres. Por defecto, con electrón y huevo se hará referencia generalmente a los portadores de carga libres.

$$m_e = \frac{\hbar^2/4\pi^2}{\partial^2 E/\partial k^2} \quad (2.5)$$

es decir, la masa efectiva del electrón puede calcularse conocida la relación entre la energía cinética de la partícula y el número de onda de la misma.

La Fig. 2.2(c) es una representación simplificada del diagrama de bandas de un semiconductor. La representación completa exige tres dimensiones. La ecuación 2.4 permite una representación sencilla del diagrama de bandas de un semiconductor, en la cual se muestra la relación existente entre energía cinética (E) y momento (p). En la Fig. 2.3 se muestra esta representación para dos materiales semiconductores: GaAs y Si. En el GaAs el máximo de la banda de valencia y el mínimo de la banda de conducción coinciden para un valor de p (momento igual cero, pues se toma como referencia del momento el máximo de la banda de valencia), mientras que esto no ocurre con el silicio. Los semiconductores en los que el máximo de la banda de valencia y el mínimo de la banda de conducción coinciden para un mismo valor de p se denominan **semiconductores de transición directa**, mientras que si esta propiedad no se cumple se denominan **semiconductores de transición indirecta**.

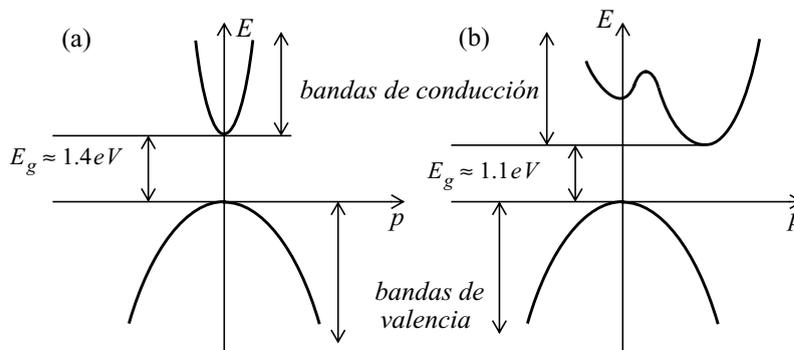


Fig. 2.3: Diagrama de bandas de: (a) GaAs (transición directa); y (b) Si (transición indirecta)

2.3.- SEMICONDUCTORES INTRÍNSECOS Y EXTRÍNSECOS

En un semiconductor perfecto las concentraciones de electrones y huecos coinciden, pues cada electrón que salta a la banda de conducción genera un hueco en la banda de valencia. Los semiconductores que cumplen que las concentraciones de electrones y huecos coinciden se denominan **semiconductores intrínsecos**, y cumplirán:

$$n = p = n_i \quad (2.6)$$

donde n y p son las concentraciones de electrones y huecos respectivamente por unidad de volumen; y n_i la **concentración de portadores intrínsecos** por unidad de volumen. La densidad de portadores intrínsecos en la mayoría de los semiconductores es baja⁶, por lo que el flujo de corriente

5. La deducción de esta ecuación pertenece al ámbito de la mecánica cuántica, por lo que no vamos a profundizar en la descripción de la misma.
6. Un valor típico para el silicio es de $n_i = 1.18 \times 10^{10} \text{ cm}^{-3}$ a una temperatura de 300 K [2], mientras que la densidad volumétrica de átomos es de $5.0 \times 10^{22} \text{ cm}^{-3}$

total que estos portadores generan no será suficiente para la implementación práctica de dispositivos electrónicos en el semiconductor. La concentración de portadores se puede controlar mediante la implantación de impurezas (sustitución de átomos de la red por átomos de la *sustancia dopante*) en el semiconductor, dando lugar a los *semiconductores extrínsecos*.

Consideremos el caso de un cristal de silicio. Si el semiconductor se dopa con átomos de elementos químicos del grupo V de la tabla periódica, cada átomo sustituido tiene un electrón más de los necesarios para completar la banda de valencia. La energía necesaria para arrancar a este electrón del átomo y que quede libre para moverse por el semiconductor es muy pequeña, de forma que la temperatura ambiente es suficiente para que el átomo se desprenda del electrón y éste se transforme en un *portador de carga libre negativa* (un electrón) en el semiconductor. La sustancia dopante en este caso se denomina *donadora*, pues dona un portador de carga negativa (un electrón) a la red cristalina. El átomo donador quedará fijado en la red cristalina como un ión positivo. Es decir, la neutralidad de carga eléctrica en el semiconductor se conserva: el número de electrones en exceso será igual al número de iones positivos. La creación del ión positivo impide que el átomo dopante correspondiente contribuya a la generación de huecos (un hueco es un portador de carga positiva, mientras que un ión será una carga positiva fijada en la red cristalina) por lo que en este tipo de semiconductores, denominados *semiconductores tipo n*, la concentración de electrones es mucho mayor que la concentración de huecos. El exceso de portadores de carga vendrá determinado por la concentración de sustancia dopante donadora.

Los *semiconductores tipo p* se generan, para el caso del silicio, mediante el dopado del semiconductor con elementos del grupo III de la tabla periódica. Los elementos del grupo III se denominan en este caso sustancias aceptoras, pues aceptan con facilidad un electrón para completar su banda de valencia. La aportación de una cantidad de energía muy pequeña a la red cristalina será suficiente para que se produzca el desplazamiento de un electrón desde la banda de valencia de un átomo de la red a la banda de valencia del átomo aceptor. Éste quedará fijado en la red como un ión negativo, mientras la sucesiva ocupación/liberación del hueco generado por la ausencia del cuarto electrón de valencia en los demás átomos se interpreta como la existencia de un portador de carga libre positiva (un hueco). En este caso el ión negativo no contribuirá a la transformación de electrones en carga libre en el semiconductor. La concentración de huecos en los semiconductores tipo p será mucho mayor que la concentración de electrones, estando el exceso de carga positiva determinado por la concentración de impurezas aceptoras.

El diagrama de energías de los semiconductores tipo n y p se muestra en la Fig. 2.4. E_c y E_v representan la energía de las bandas de conducción y de valencia respectivamente. E_a es el nivel de energía de la sustancia aceptoras, y $E_a - E_v$ se denomina la *energía de ionización* de la impureza aceptoras. De forma similar, E_d es el nivel de energía de la sustancia donadora, y $E_c - E_d$ se denomina la *energía de ionización* de la impureza donadora. Como se observa, la introducción de impurezas en el semiconductor genera estados energéticos intermedios que facilitarán los saltos energéticos entre las bandas de conducción y valencia.

2.4.- DENSIDAD DE PORTADORES DE CARGA EN EQUILIBRIO

La *densidad de electrones en la banda de conducción* representa el número de portadores de carga libre negativa capaces de contribuir a la corriente de conducción. Este parámetro puede expresarse en función de las características del semiconductor y de la temperatura como:

$$n = N_c e^{-(E_c - E_f)/kT} \quad (2.7)$$

donde:

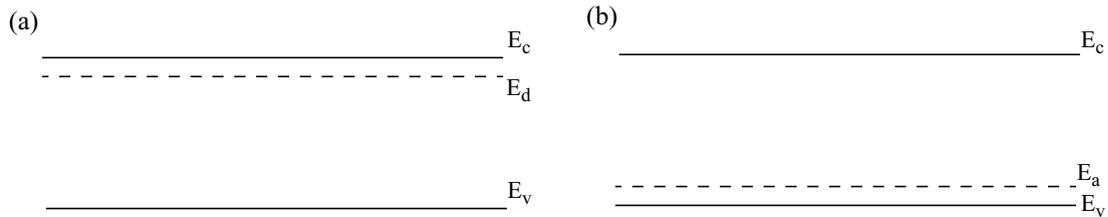


Fig. 2.4: Diagrama de bandas de un semiconductor dopado (a) tipo n; y (b) tipo p

$$N_v = 2 \left(\frac{2\pi m_e kT}{h^2} \right)^{3/2} \quad (2.8)$$

es la densidad efectiva de estados en la banda de conducción. Para el silicio este parámetro toma el valor típico de $N_c = 2.8 \times 10^{19} \text{ cm}^{-3}$ a temperatura ambiente (300 k) [1].

Igualmente, la **densidad de huecos en la banda de valencia** se calcula como:

$$p = N_v e^{-(E_f - E_v)/kT} \quad (2.9)$$

donde:

$$N_c = 2 \left(\frac{2\pi m_h kT}{h^2} \right)^{3/2} \quad (2.10)$$

es la densidad efectiva de estados en la banda de valencia. Para el silicio este parámetro toma el valor típico de $N_v = 10^{19} \text{ cm}^{-3}$ a temperatura ambiente (300 k) [1].

El producto de las ecuaciones 2.7 y 2.10 tiene la expresión:

$$np = N_c N_v e^{-E_g/kT} \quad (2.11)$$

donde $E_g = E_c - E_v$ es la energía de la banda prohibida, y resulta ser una constante característica de un semiconductor en equilibrio térmico⁷. El producto np depende de las densidades efectivas de estados en las bandas de conducción y valencia, así como de la energía de la banda prohibida, y resulta independiente de las densidades de impurezas.

En un semiconductor intrínseco, la densidad de electrones es exactamente igual a la densidad de huecos. Sin embargo, en un semiconductor extrínseco, el incremento de la densidad de uno de los tipos de portadores tiende a reducir la densidad de los otros portadores debido a los procesos de recombinación, de forma que el producto de las concentraciones de ambos tipos permanece constante para una temperatura dada. Este equilibrio viene expresado por la relación:

$$np = n_i^2 \quad (2.12)$$

que es válida tanto para semiconductores intrínsecos como extrínsecos en condición de equilibrio térmico.

7. Por equilibrio térmico se define la condición de estado estacionario a una temperatura determinada y en ausencia de excitación externa.

La densidad de portadores intrínsecos en un semiconductor, utilizando las expresiones 2.11 y 2.12, puede expresarse como:

$$n_i = \sqrt{N_c N_v} e^{-E_g/2kT} \quad (2.13)$$

Cuando el semiconductor se dopa con una densidad de impurezas donadoras N_d , la concentración de electrones en la banda de conducción aumenta en relación con la concentración intrínseca. Si toda la sustancia dopante se ioniza, es decir, N_d coincide con la densidad de átomos ionizados en el semiconductor, el número total de electrones en la banda de conducción debe ser igual a la suma de electrones originados desde el nivel energético de la sustancia donadora más los originados desde la banda de valencia. Puesto que cada electrón que salta a la banda de conducción desde la banda de valencia deja un hueco en esta última, la densidad de electrones puede escribirse como:

$$n = p + N_d \quad (2.14)$$

La ecuación 2.14 expresa la condición de neutralidad de carga para un semiconductor tipo n. Cuando en el semiconductor están presentes ambos tipos de dopado, la ecuación de neutralidad de carga puede escribirse como:

$$n + N_a = p + N_d \quad (2.15)$$

La resolución de las ecuaciones 2.12 y 2.14 nos llevan a las expresiones de las concentraciones de electrones y huecos en equilibrio térmico para un semiconductor tipo n:

$$n_n = \frac{\sqrt{N_d^2 + 4n_i^2} + N_d}{2} \quad (2.16)$$

$$p_n = \frac{\sqrt{N_d^2 + 4n_i^2} - N_d}{2}$$

El subíndice n se usa para indicar el tipo de dopado del semiconductor. Para el semiconductor tipo n el electrón será el portador dominante, por lo que se le denomina **portador mayoritario**, mientras que el hueco será el **portador minoritario**. En la mayoría de las aplicaciones prácticas, en los semiconductores tipo n, la concentración de impurezas donadoras es mucho mayor que la concentración intrínseca de portadores, es decir: $n_i/N_d \ll 1$. Bajo esta consideración las ecuaciones 2.16 pueden aproximarse por:

$$n_n \cong N_d + \frac{n_i^2}{N_d} \cong N_d \quad (2.17)$$

$$p_n \cong \frac{n_i^2}{N_d}$$

De forma análoga, para un semiconductor tipo p en equilibrio térmico, bajo la condición $n_i/N_a \ll 1$:

$$n_p \cong \frac{n_i^2}{N_a} \quad (2.18)$$

$$p_p \cong N_a$$

2.5.- CORRIENTES DE ARRASTRE Y DE DIFUSIÓN

Consideremos un bloque de semiconductor homogéneo (Fig. 2.5). Si sometemos el semiconductor a una diferencia de potencial V entre sus extremos, la presencia de este campo eléctrico generará una corriente denominada *corriente de arrastre*. La corriente de arrastre de los electrones en un volumen de semiconductor viene expresada por:

$$I_n = qAn\mu_n\xi \quad (2.19)$$

donde q es la carga del electrón, A el área de la sección transversal del volumen considerado, n la concentración de electrones libres por unidad de volumen, μ_n la movilidad de los electrones en el semiconductor⁸, y ξ el campo eléctrico aplicado.

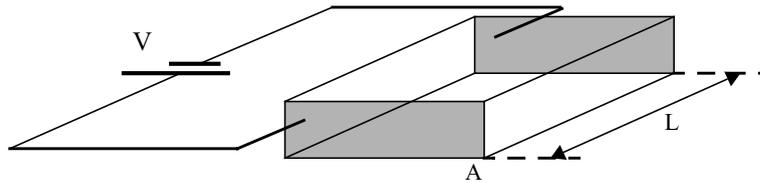


Fig. 2.5: Bloque de semiconductor homogéneo bajo un campo eléctrico aplicado

La corriente de arrastre de los huecos sigue una ley similar:

$$I_p = qAnp\mu_p\xi \quad (2.20)$$

donde p es la concentración de huecos por unidad de volumen, μ_p la movilidad de los huecos en el semiconductor, y ξ el campo eléctrico aplicado.

La movilidad de electrones y huecos es diferente, pues las masas efectivas de estos dos portadores son distintas. La movilidad es proporcional al tiempo medio entre colisiones con los átomos de la red cristalina, por lo que es de suponer que la movilidad decaerá con la concentración de impurezas en la red. La Fig. 2.6 [4] muestra la dependencia de la movilidad de los electrones y de los huecos con la concentración de dopado para el caso del silicio a una temperatura de 300 °K. Esta figura no es representativa de lo que ocurre cerca de la superficie cristalina, zona en la cual la concentración de centros de recombinación generados por las imperfecciones de la red cristalina en la interfase hace que la movilidad se reduzca

El campo eléctrico aplicado en el volumen de semiconductor homogéneo de la Fig. 2.5 puede aproximarse por V/L , donde V es la diferencia de potencial que aparece entre los extremos del semiconductor. Haciendo uso de esta relación, y considerando la relación entre las características

8. La movilidad proporciona la relación entre la velocidad media de arrastre de los portadores de carga (v_d) y el campo eléctrico aplicado (ξ): $v_d = \pm\mu\xi$, donde el signo positivo se usa para los huecos y el negativo para los electrones. Las expresiones de la movilidad para electrones y huecos, así como la dependencia de ésta con la intensidad de campo aplicado y con la temperatura se encuentra descrita en profundidad en [1].

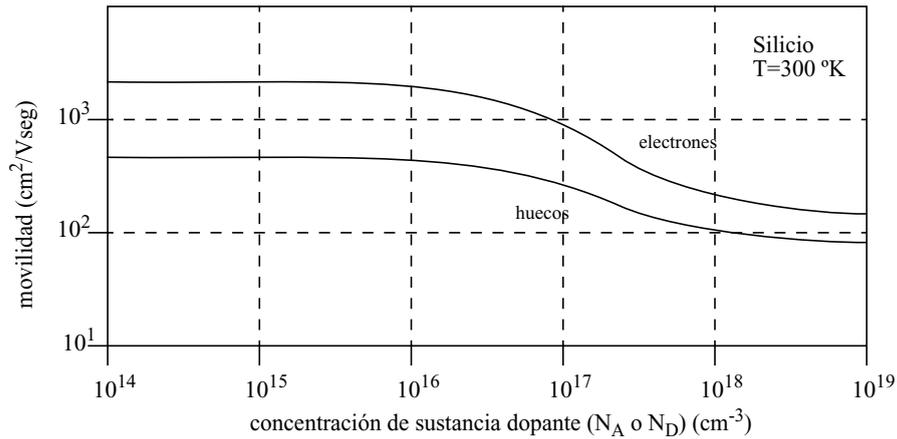


Fig. 2.6: Variación de la movilidad de electrones y huecos con la concentración de sustancia dopante

geométricas y eléctricas que proporciona la resistencia R del material ($R = \rho(L/A)$), la resistividad ρ del semiconductor puede expresarse como:

$$\frac{1}{\rho} = \sigma = q\mu_n n + q\mu_p p \quad (2.21)$$

donde σ es la conductividad del material. La resistividad de un semiconductor es un parámetro importante para el diseño de dispositivos semiconductores. La Fig. 2.7 [3] muestra la relación entre los niveles de concentración de impurezas y la resistividad, tanto para silicio dopado tipo n como para silicio dopado tipo p [1].

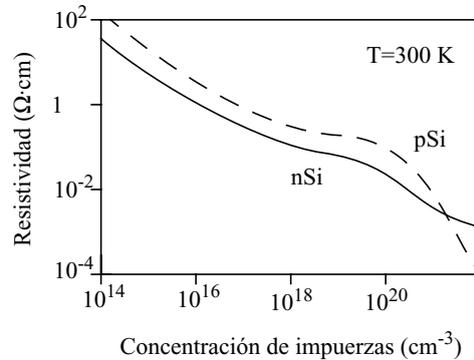


Fig. 2.7: Relación entre resistividad y concentración de impurezas en el silicio

La existencia de gradientes en la concentración de portadores de carga libre en un material genera una corriente de portadores denominada **corriente de difusión** desde las regiones en las que la concentración es alta hacia las regiones en las que la concentración es menor. Las corrientes de difusión de electrones y huecos obedecen a las siguientes ecuaciones:

$$I_n = qAD_n \frac{dn}{dx} \quad (2.22)$$

$$I_p = -qAD_p \frac{dp}{dx}$$

donde A es la sección a través de la cual se están midiendo las intensidades, D_n y D_p las constantes de difusión de electrones y huecos respectivamente, y x la dirección de difusión de los

portadores. Estas constantes guardan una relación con la movilidad de los portadores dada por la ecuación de Einstein:

$$\frac{D_p}{\mu_p} = \frac{D_n}{\mu_n} = \frac{kT}{q} \quad (2.23)$$

Para silicio poco dopado son típico los valores $D_n=38 \text{ cm}^2/\text{seg}$; $D_p=13 \text{ cm}^2/\text{seg}$.

La corriente total de electrones en un semiconductor será resultado de la suma de las componentes de corriente de arrastre y de difusión:

$$I_n = qA \left(n\mu_n \xi + D_n \frac{dn}{dx} \right) \quad (2.24)$$

Igualmente, la corriente total de huecos vendrá dada por:

$$I_p = qA \left(p\mu_p \xi - D_p \frac{dp}{dx} \right) \quad (2.25)$$

La corriente total en el semiconductor será la suma de las corrientes de electrones y huecos:

$$I = I_n + I_p \quad (2.26)$$

2.6.- PROCESOS DE GENERACIÓN Y RECOMBINACIÓN DE PARES ELECTRÓN-HUECO

El proceso de **generación de un par electrón-hueco** es el siguiente: un electrón en la banda de valencia recibe una energía suficiente (igual o mayor al ancho de la banda prohibida) para saltar la banda prohibida y pasar a ocupar un estado energético de la banda de conducción, generándose a su vez un hueco en la banda de valencia. La **recombinación de un par electrón-hueco** es el proceso por el cual un electrón salta desde la banda de conducción a la banda de valencia, ocupando uno de los huecos existentes en esta banda (se produce la aniquilación de un par electrón-hueco). En el proceso se desprende una energía igual o mayor a la energía de la banda prohibida.

Bajo la condición de **equilibrio térmico** los procesos de generación y recombinación se dan con la misma frecuencia, por lo que las concentraciones de portadores permanecen constantes. Es decir, se cumple: $pn = n_i^2$.

La condición de equilibrio se rompe con la **inyección de portadores de carga libre**. Esta inyección puede llevarse a cabo mediante medios eléctricos u ópticos. En la inyección óptica un haz de fotones de energía E_p igual o mayor a la energía de la banda prohibida E_g ($E_p = h \cdot \nu \geq E_g$) incide sobre el semiconductor. Los fotones aportan la energía necesaria para que se produzca la generación de un par electrón-hueco. La inyección de portadores incrementa la densidad de electrones y huecos, por lo que se cumplirá: $pn \geq n_i^2$. El exceso de electrones será igual al número de huecos en exceso, de forma que se conserva la neutralidad de carga en el semiconductor.

En el proceso de recombinación, la diferencia de energías entre el estado energético de la banda de conducción y el estado energético de la banda de valencia entre los que se produce la recombinación de un par electrón-hueco se emitirá en forma de fotón o de fonón, en función de la naturaleza del proceso de recombinación. Cuando se produce la emisión de un fotón el proceso se denomina

de recombinación radiativa; cuando es un fonón, se tratará de un **proceso de recombinación no radiativa**, y la energía desprendida se absorbe en el semiconductor en forma de calor.

El número de portadores inyectados en un semiconductor generalmente controla el comportamiento de los dispositivos. Consideremos el caso de un semiconductor tipo n^0 iluminado con una fuente de luz, bajo condiciones de estado estacionario. Las concentraciones de portadores en equilibrio térmico se obtienen por las expresiones 2.16, que en la práctica, considerando que la concentración de impurezas donadoras es mucho mayor que la concentración intrínseca de portadores, vienen dadas por la ecuación 2.17.

Si la luz no está encendida, el semiconductor está en equilibrio y no hay exceso de portadores. Cuando se enciende la luz, pueden darse dos situaciones: **inyección débil**, en la que la densidad de portadores inyectados es pequeña en comparación con la concentración de impurezas, por lo que la densidad de portadores mayoritarios permanecerá prácticamente constante, mientras que la densidad de portadores minoritarios será aproximadamente igual a la carga inyectada; **inyección fuerte**, en la que la densidad de portadores inyectados es comparable a la concentración de impurezas. La inyección fuerte tiene efectos de segundo orden que hacen que el comportamiento del dispositivo se desvíe del comportamiento descrito por las anteriores ecuaciones¹⁰.

Los procesos de generación/recombinación de pares electrón-hueco están gobernados por la absorción/emisión de energía por parte de los portadores de carga y dependerán de las características del material y las concentraciones de portadores en el mismo. En los semiconductores de transición directa la frecuencia U con la que un electrón realiza la transición desde la banda de conducción a la banda de valencia (**frecuencia media de recombinación**) viene dada por:

$$U = B(np - n_i^2) \quad (2.27)$$

donde B es un coeficiente de probabilidad que expresa la naturaleza estadística de los procesos de recombinación.

Si consideramos el semiconductor tipo p , bajo la condición de inyección débil, es decir, $p_o \gg \Delta n$, la frecuencia de recombinación U resulta proporcional a la densidad de **electrones en exceso** Δn (electrones ubicados en estados energéticos excitados) en la banda de conducción y al número de huecos en la banda de valencia p_o . Es decir, la ecuación 2.27 puede expresarse como:

$$U \cong B \cdot \Delta n \cdot p_o \quad (2.28)$$

La frecuencia de recombinación permite definir el **tiempo medio de recombinación** o **tiempo de vida medio** de los portadores de carga.:

$$\tau \equiv \frac{\Delta n}{U} \quad (2.29)$$

Para semiconductores de transición indirecta, las expresiones de la frecuencia media de recombinación y del tiempo medio de vida son más complejas, pues en éstos entra en juego la concentración de **trampas** (estados energéticos dentro de la banda prohibida que surgen debido a

9. Un análisis paralelo puede llevarse a cabo para semiconductores tipo p , llegando a resultados muy similares.
10. En el presente estudio se considerará en todo momento que se verifica la condición de inyección débil. Esto permite llevar a cabo un análisis más sencillo. Además la condición de inyección débil es coherente con el estudio de la unión pn aplicada al sensado de imagen, en el que las corrientes fotogeneradas tendrán dimensiones muy pequeñas.

imperfecciones en la red cristalina, y que utilizan los electrones como saltos intermedios en las transiciones entre las bandas de valencia y de conducción). La expresión de la frecuencia media de recombinación puede expresarse bajo la condición de inyección débil como:

$$U \approx cN_t(p_n - p_{n_0}) \quad (2.30)$$

donde N_t es la densidad de centros de recombinación, c es uno de los coeficientes de Einstein que expresa la probabilidad de que un electrón sea capturado por una trampa, p_n la concentración de huecos, y p_{n_0} la concentración de huecos en equilibrio térmico.

En función de la ecuación 2.30 el tiempo medio de recombinación o tiempo medio de vida del electrón para un semiconductor de transición indirecta será:

$$\tau_p = \frac{1}{cN_t} \quad (2.31)$$

Puesto que el tiempo medio de vida es inversamente proporcional a la concentración de trampas, este parámetro servirá como indicador de la pureza de un semiconductor.

En la superficie del semiconductor, debido a la existencia de una discontinuidad, así como de imperfecciones introducidas por la absorción de iones y moléculas en el proceso de fabricación de los dispositivos semiconductores, aparece un gran número de estados intermedios en la banda prohibida, denominados *estados energéticos superficiales*, que facilitan los procesos de recombinación. La frecuencia de recombinación crece considerablemente a medida que nos acercamos a la superficie del semiconductor. Las concentraciones de portadores en exceso decrecerá en las proximidades de la superficie, como ilustra la Fig. 2.8 para un semiconductor tipo n.

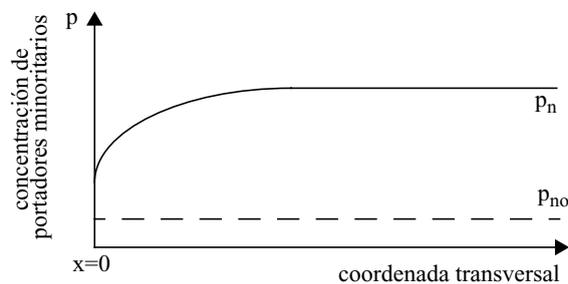


Fig. 2.8: Concentración de portadores minoritarios en las proximidades de la superficie del semiconductor

2.7.- RESPUESTA TRANSITORIA Y FOTOCONDUCTIVIDAD

Consideremos un volumen rectangular de semiconductor de sección A y longitud L con una diferencia de potencial V aplicada entre sus extremos e iluminado por una fuente de luz cuyos fotones tienen una energía mayor que la banda prohibida del semiconductor (Fig. 2.9). Suponiendo absorción uniforme en todo el volumen semiconductor, el *coeficiente de generación de pares electrón-hueco* (frecuencia de generación de portadores por unidad de volumen) vendrá dado por:

$$G_L = \frac{P_L}{h \cdot \nu} \cdot \frac{1}{V_o} \cdot \eta \quad (2.32)$$

donde P_L es la potencia óptica incidente; $h\nu$ la energía del fotón, η la eficiencia cuántica del semiconductor (número de electrones generados por fotón), y V_o el volumen total de semiconductor. La **corriente fotogenerada** en el volumen, es decir, la corriente de portadores generados mediante la absorción de fotones, vendrá dada por:

$$I_{ph} = qG_L(\mu_n + \mu_p)\frac{V}{L} \cdot A \quad (2.33)$$

donde A es la sección transversal del semiconductor, y V/L el campo eléctrico aplicado entre los contactos óhmicos.

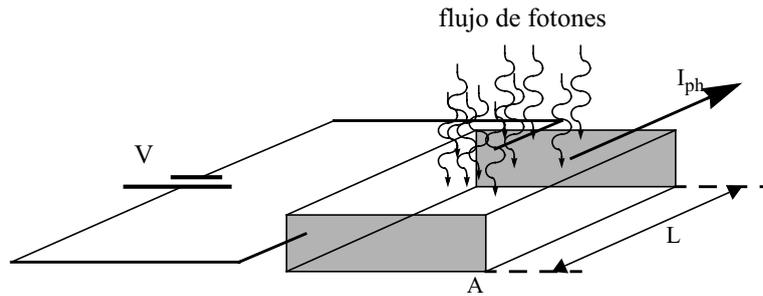


Fig. 2.9: Corriente fotogenerada en un bloque semiconductor bajo un campo eléctrico aplicado

Debido a los procesos de generación, las densidades de electrones y huecos aumentarán. Este aumento será contrarrestado por un aumento de los procesos de recombinación. Se alcanza una situación de equilibrio cuando ambos procesos se igualan. Considerando Δp la densidad de huecos en exceso, G_L la frecuencia de generación de pares electrón-hueco y U la frecuencia de recombinación, la ecuación que describe este comportamiento es:

$$\frac{d}{dt}\Delta p = G_L - U \quad (2.34)$$

Para un semiconductor de transición directa¹¹, bajo la condición de inyección débil la expresión queda:

$$\frac{d}{dt}\Delta p = G_L - \frac{\Delta p}{\tau_p} \quad (2.35)$$

Usando como condición inicial $\Delta p(t \leq 0) = G_L\tau_p$ la solución de la ecuación 2.35 es:

$$\Delta p(t) = G_L\tau_p e^{-t/\tau_p} \quad (2.36)$$

cuya dependencia con el tiempo queda reflejada en la Fig. 2.10.

La **fotococonductividad** de un semiconductor puede definirse como:

$$\sigma(t) = q[\mu_n n(t) + \mu_p p(t)] = \sigma_o + q(\mu_n + \mu_p)\Delta p(t) \quad (2.37)$$

donde $\sigma_o = q[\mu_n n_o + \mu_p p_o]$.

La expresión 2.37 pone de manifiesto la relación entre la conductividad del semiconductor y el flujo de luz incidente. Esta relación permite pensar en el semiconductor como un **dispositivo fotoconductor**: un volumen rectangular de semiconductor de sección A y longitud L con dos con-

11. Estas expresiones no son válidas para semiconductores de transición indirecta, pues no resultan tan sencillas, y no ilustran de una forma tan directa el comportamiento fotoconductor en un semiconductor.

tactos óhmicos en sus extremos (Fig. 2.5). Si sobre este volumen hacemos incidir un haz de fotones de energía superior a la banda prohibida del semiconductor, la conductividad del mismo seguirá la ley expresada en 2.37.

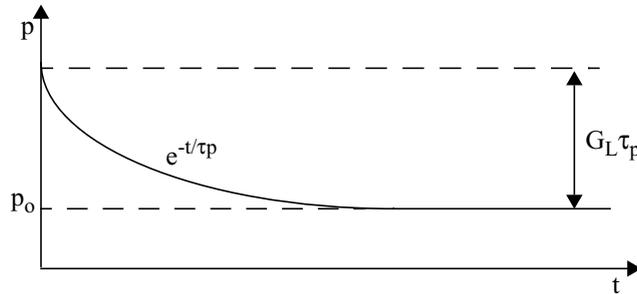


Fig. 2.10: Respuesta transitoria de la concentración de portadores en exceso

2.8.- COEFICIENTE DE ABSORCIÓN Y EFICIENCIA CUÁNTICA

Consideremos que un flujo de fotones de energía $E_{ph} > E_g$ procedente de una fuente de luz monocromática incide perpendicularmente sobre un semiconductor (Fig. 2.11). F_{ph} es el flujo de luz en fotones por unidad de área y por unidad de tiempo sobre la superficie del semiconductor ($x=0$). A medida que el flujo de fotones atraviesa el semiconductor, éste se verá reducido debido a los procesos de absorción de fotones. $F(x)$ será el flujo de fotones a una distancia x de la superficie del semiconductor.

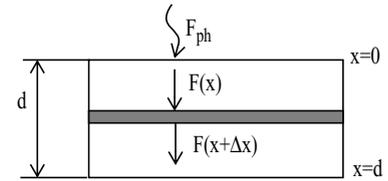


Fig. 2.11: Incidencia de un flujo de fotones sobre el SC

La cantidad de fotones absorbidos en una distancia recorrida Δx será proporcional a $F(x) \cdot \Delta x$, con un coeficiente de proporcionalidad α denominado **coeficiente de absorción**. La ecuación que modela este comportamiento es:

$$F(x + \Delta x) - F(x) = -\alpha F(x) \Delta x \rightarrow \frac{d}{dx} F(x) = -\alpha F(x) \quad (2.38)$$

Bajo la condición de contorno $F(x=0) = F_{ph}$ la solución a la ecuación 2.38 es:

$$F(x) = F_{ph} e^{-\alpha x} \quad (2.39)$$

El coeficiente de absorción es una característica del semiconductor, y es la inversa del valor de x que cumple $F(\alpha^{-1}) = F_{ph} e^{-1}$, es decir, α^{-1} , denominada **profundidad de penetración** o **longitud de absorción**, es la distancia para la cual se ha absorbido aproximadamente un 37% del flujo incidente sobre el semiconductor (flujo en $x=0$). Estadísticamente, la profundidad de penetración indica la distancia media que recorre un fotón a través del semiconductor antes de ser absorbido.

El coeficiente de absorción presenta una fuerte dependencia con la longitud de onda. La Fig. 2.12 [1] ilustra la dependencia del coeficiente de absorción con la longitud de onda para algunos semiconductores empleados en aplicaciones optoelectrónicas.

Ignorando los efectos de las pérdidas por reflexión, y asumiendo que cada fotón absorbido genera un par electrón-hueco, el **coeficiente de generación de pares electrón-hueco** vendrá dado a partir de la ecuación 2.38, en función de la profundidad de penetración, como:

$$G_L = \alpha F(x) = \alpha F_{ph} e^{-\alpha x} \quad (2.40)$$

Esta ecuación expresa la proporción de flujo luminoso que es absorbido por unidad de distancia recorrida a través del semiconductor.

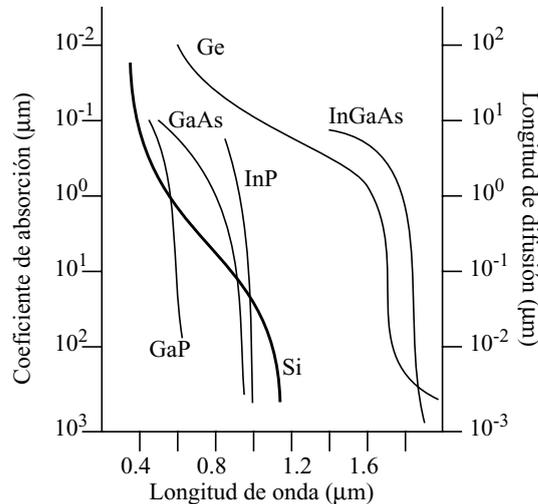


Fig. 2.12: Coeficientes de absorción para algunos semiconductores utilizados en aplicaciones optoelectrónicas

Supongamos un semiconductor de sección A y longitud L con un campo eléctrico aplicado, y expuesto a un flujo luminoso (Fig. 2.9). La relación entre el flujo luminoso incidente sobre el semiconductor (F_{ph}) y la densidad de corriente total fotogenerada en el semiconductor (J_{ph}) se conoce como **eficiencia cuántica**:

$$\eta = \frac{J_{ph}}{qF_{ph}} \quad (2.41)$$

y representa el número de pares electrón-hueco que son generados y detectados electrónicamente por cada fotón incidente. La eficiencia cuántica incluirá los efectos de todos los mecanismos que impidan que los fotones sean finalmente detectados.

Alternativamente, esta eficiencia de la conversión de señal luminosa en eléctrica puede describirse mediante la **responsividad** R , cuyas unidades son A/W. La responsividad mide la corriente que es generada en el fotosensor por unidad de potencia de luz incidente. La relación entre la responsividad (R) y la eficiencia cuántica (η) viene dada por:

$$R = \eta \frac{\lambda q}{hc} \quad (2.42)$$

donde h es la constante de Planck; c la velocidad de la luz; q la unidad de carga eléctrica; y λ la longitud de onda del fotón. Como ejemplo, considerando una eficiencia cuántica del 90%, iluminado con luz roja ($\lambda=633$ nm) de un láser de HeNe, la responsividad toma un valor de 0.46 A/W [5].

2.9.- ECUACIÓN DE CONTINUIDAD EN UN SEMICONDUCTOR

Consideremos un volumen de semiconductor de sección A y longitud Δx en el que existe un campo eléctrico aplicado (Fig. 2.5). Supongamos un flujo de portadores a través del volumen. La variación en el flujo de huecos que atraviesa el volumen puede expresarse de forma diferencial como:

$$\frac{J_p(x + \Delta x)}{q} - \frac{J_p(x)}{q} = \frac{1}{q} \cdot \frac{\partial J_p}{\partial x} \cdot \Delta x \quad (2.43)$$

Teniendo en cuenta que la continuidad de la carga depende de la variación en la concentración de portadores en el volumen debida a los procesos de recombinación (si despreciamos los procesos de generación de portadores en el volumen, pues consideramos que el semiconductor no se está excitando), y que la variación en la concentración de huecos en el volumen viene dada por $-(\partial p / \partial t) \cdot \Delta x$, la ecuación 2.43 puede reescribirse como:

$$\frac{1}{q} \cdot \frac{\partial J_p}{\partial x} + \frac{p - p_o}{\tau_p} = - \frac{\partial p}{\partial t} \quad (2.44)$$

donde $(p - p_o) / \tau$ expresa la variación en el flujo de huecos debido a los procesos de recombinación en el volumen.

Para los electrones podemos encontrar una expresión similar:

$$-\frac{1}{q} \cdot \frac{\partial J_n}{\partial x} + \frac{n - n_o}{\tau_n} = - \frac{\partial n}{\partial t} \quad (2.45)$$

Las ecuaciones 2.44 y 2.45 son las **ecuaciones de continuidad** para los huecos y para los electrones respectivamente. Estas ecuaciones, junto a las expresiones para las intensidades de corriente (2.24):

$$I_n = J_n \cdot A = qA \left(n \mu_n \xi + D_n \frac{dn}{dx} \right) \quad (2.46)$$

$$I_p = J_p \cdot A = qA \left(p \mu_p \xi - D_p \frac{dp}{dx} \right)$$

y a la distribución del campo eléctrico en el semiconductor, resultado de la resolución de la ecuación de Poisson¹²:

$$\frac{d^2 \Psi}{dx^2} = \frac{q}{\epsilon_{sc}} [(n - p) + (N_a - N_d)] \quad (2.47)$$

forman un conjunto de ecuaciones que describen las distribuciones de portadores, de corrientes y del campo eléctrico, y cuya resolución es posible conocidas las condiciones de contorno e iniciales en el semiconductor.

Para el caso en el que el semiconductor esté siendo iluminado (Fig. 2.9), a las ecuaciones 2.44 y 2.45 habrá que añadirles el término de generación de pares electrón-hueco:

12. En esta expresión $\epsilon_{sc} = K_s \epsilon_o$ es la constante dieléctrica del semiconductor y N_a y N_d las concentraciones de átomos aceptores y donadores respectivamente

$$\begin{aligned} \frac{1}{q} \cdot \frac{\partial J_p}{\partial x} + \frac{p - p_o}{\tau_p} + \alpha F_{ph} e^{-\alpha x} &= - \frac{\partial p}{\partial t} \\ - \frac{1}{q} \cdot \frac{\partial J_n}{\partial x} + \frac{n - n_o}{\tau_n} + \alpha F_{ph} e^{-\alpha x} &= - \frac{\partial n}{\partial t} \end{aligned} \quad (2.48)$$

A la corriente a través del semiconductor (ecuación 2.46) habrá que añadir una tercera componente, la corriente fotogenerada, dada por la ecuación 2.33.

2.10.- UNIÓN pn

Una unión pn se crea al poner en contacto un semiconductor de dopado tipo n con un semiconductor de dopado tipo p. Consideremos la unión pn representada en la Fig. 2.13(a) en la que se consideran las concentraciones de impurezas donadoras y aceptoras distribuidas de forma homogénea (esta condición implica la existencia de una transición abrupta de concentraciones entre las regiones n y p; en todo el posterior análisis se supondrá que se da esta condición de **unión abrupta**). A temperatura ambiente los iones donadores y aceptores serán carga eléctrica (positiva y negativa respectivamente) fija en la red cristalina, es decir, **iones positivos y negativos**. Cada ión positivo dona un portador de carga eléctrica negativa libre (**electrones libres**), mientras que cada ión negativo aporta un portador de carga eléctrica positiva libre (**huecos libres**).

En los extremos de la unión pn la neutralidad de carga se conserva (**regiones de carga neutra**), pues la carga eléctrica total libre será igual a la carga eléctrica total fija en la red. Sin embargo, en las proximidades de la unión el comportamiento es diferente debido a la existencia de un gradiente en la concentración de portadores. Los electrones libres de la región n tienden a difundirse hacia la región p de la unión (corriente de difusión de electrones), dejando en la región n un número de iones positivos igual al número de electrones que han emigrado hacia la región p; los huecos lo harán desde la región p hacia la región n (corriente de difusión de huecos), y los iones que quedarán en la región p serán iones negativos. Se genera la **región de carga espacial**¹³. A su vez los iones generarán un campo eléctrico interno que creará un corriente de arrastre en sentido opuesto a la corriente de difusión de los portadores. Se alcanzará un equilibrio cuando las corrientes de difusión (generada por el gradiente en la concentración de portadores) y de arrastre (generada por el campo eléctrico interno) se igualen.

La Fig. 2.13 muestra de forma cuantitativa: (a) la unión pn; (b) las concentraciones de portadores de iones en cada región; (c) la intensidad de campo eléctrico; (d) el potencial eléctrico; y (e) el diagrama de bandas de la unión.

En el diagrama de bandas de la unión aparece el **nivel de Fermi** (E_f). El nivel de Fermi es un parámetro importante en un semiconductor, y sirve para calcular en el semiconductor la probabilidad de que un estado energético E esté ocupado por un electrón. Esta probabilidad viene determinada por la función de distribución de Fermi-Dirac:

$$f(E) = \frac{1}{e^{(E - E_f)/kT} + 1} \quad (2.49)$$

donde k es la constante de Boltzmann, y T la temperatura absoluta. A 0 K, $f(E)$ toma el valor unidad para $E < E_f$, es decir, todos los niveles energéticos por debajo de E_f estarán ocupados, mientras que todos los niveles energéticos por encima de E_f estarán desocupados. Para $T > 0$ K,

13. Esto es una simplificación de lo que ocurre en la unión, resultado de la condición de unión abrupta.

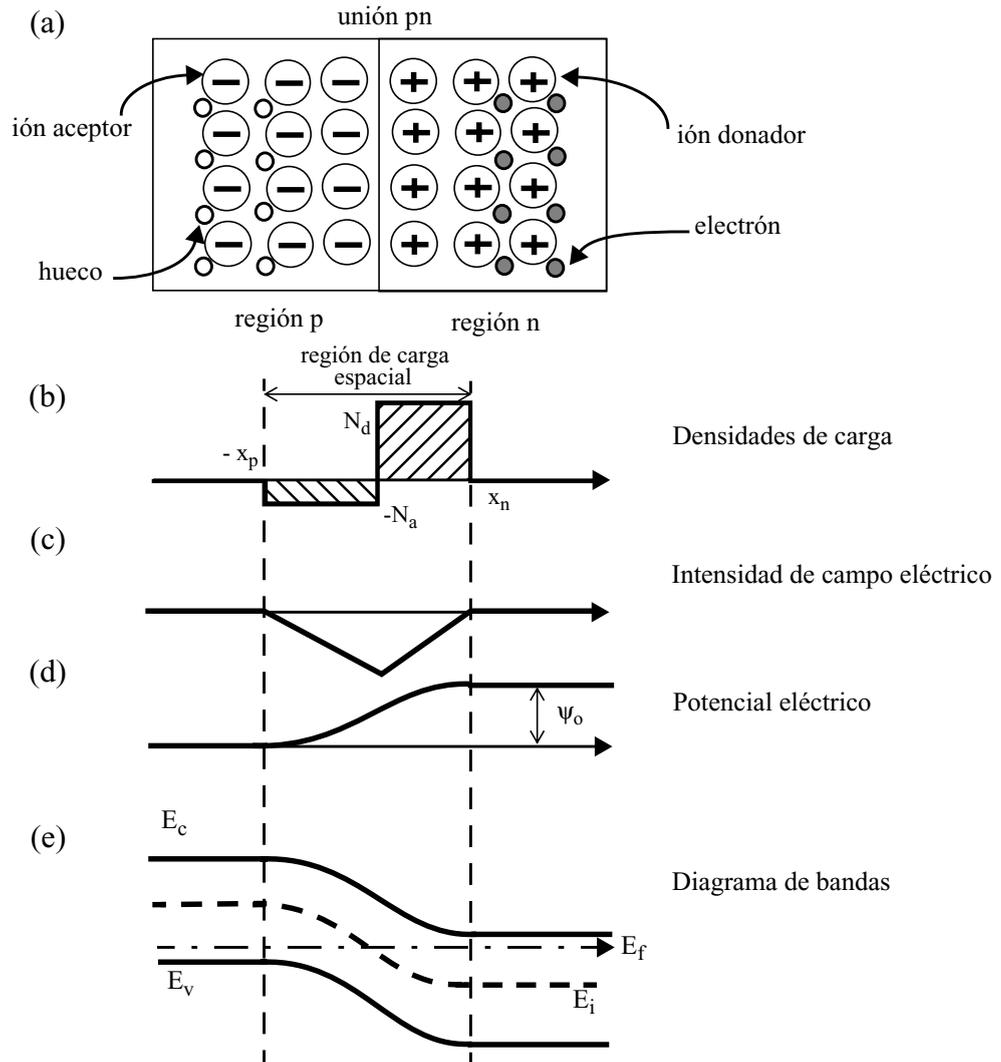


Fig. 2.13: Caracterización de la unión pn

la probabilidad de ocupación es siempre $1/2$ para $E=E_f$, independientemente de la temperatura. Además, $f(E)$ presenta simetría respecto a E_f de forma que la probabilidad de ocupación de un estado energético $E_f+\Delta E$ coincide con la probabilidad de que el estado energético $E_f-\Delta E$ esté desocupado.

La probabilidad de que un estado energético no esté ocupado vendrá dada por:

$$1 - f(E) = \frac{1}{e^{(E_f - E)/kT} + 1} \quad (2.50)$$

y servirá para medir la probabilidad de ocupación de un estado energético en la banda de valencia por un hueco.

En equilibrio térmico no existe movimiento neto de carga, y esto ocurre cuando la probabilidad de ocupación de un estado energético es la misma en cualquier parte de la unión. Si este equilibrio de

probabilidades se rompe se inicia una migración de los portadores entre estados energéticos con la intención de alcanzar una nueva situación de equilibrio. Esto se traduce en que el diagrama de bandas de la unión pn ha de ser tal que el nivel de Fermi ha de ser constante a lo largo de toda la unión.

Una característica importante de la unión pn es el ancho de la región de carga espacial, conocida como **región de deplexión**. Prescindiendo del desarrollo teórico¹⁴, el ancho de las regiones de deplexión p y n vienen dados por las expresiones:

$$x_n = \sqrt{\frac{2\epsilon_{sc}\psi_0 N_a}{qN_d(N_a + N_d)}} \quad (2.51)$$

$$x_p = \sqrt{\frac{2\epsilon_{sc}\psi_0 N_d}{qN_a(N_a + N_d)}}$$

donde ψ_0 es el potencial de contacto de la unión; q la carga del electrón; ϵ_{sc} la permitividad eléctrica del semiconductor; y N_a y N_d las concentraciones de átomos aceptores y donadores en las regiones p y n respectivamente.

El ancho total de la región de deplexión puede calcularse de las ecuaciones 2.51:

$$x_d = x_n + x_p \quad (2.52)$$

2.11.- UNIÓN p⁺n POLARIZADA

Un caso particular que resulta de especial interés por sus propiedades eléctricas es la unión p⁺n (Fig. 2.14(a)). Esta figura muestra las características de una unión p⁺n en equilibrio térmico¹⁵. ψ_0 representa la barrera de potencial que los portadores han de superar para producir corriente. Como se muestra en la figura, esencialmente la carga espacial y la barrera de potencial asociada están situadas en la región de dopado débil de la unión. Cuando se conecta una tensión externa entre los extremos de la unión el equilibrio térmico (Fig. 2.15) se rompe y, en general, fluirá corriente por el semiconductor. En general puede considerarse que toda la caída de potencial se produce entre los extremos de la región de deplexión. Si se polariza la unión en directa (se aplica una tensión V positiva entre las regiones p y n) la barrera de potencial disminuye hasta $\psi_0 - V$, como se muestra en la Fig. 2.14(b). La reducción de la barrera de potencial permite la difusión de portadores mayoritarios a través de la unión, por lo que aparece un flujo de corriente considerable. Si la unión se polariza en inversa (se aplica una tensión $V = -V_R$ entre las regiones p y n (Fig. 2.15(b))) la barrera de potencial crece hasta $\psi_0 + |V_R|$, como se muestra en la Fig. 2.14(c). El incremento de la barrera de potencial hace que el flujo de corriente a través de la unión sea prácticamente nulo: la corriente será muy baja.

Cuando la unión se encuentra polarizada en inversa el potencial de la unión ψ_0 puede sustituirse en las expresiones de los anchos de las regiones de deplexión por $\psi_0 + V$. Al ser la unión p⁺n, el ancho de la región de deplexión puede aproximarse por:

$$x_d \approx x_n = \sqrt{\frac{2\epsilon_{sc}(\psi_0 + |V_R|)}{qN_d}} \quad (2.53)$$

-
14. Este desarrollo se basa en la resolución de la ecuación de Poisson considerando homogéneas las distribuciones de sustancias dopantes como se muestra en la Fig. 2.13(b), y tomando como condición de contorno que el campo eléctrico es cero en las regiones de carga neutra.
 15. En la Fig. 2.14 los valores de x_n y x_p han de ser tales que se verifique la neutralidad de la carga, lo que equivale a que las superficies de las regiones n y p ralladas sean iguales.

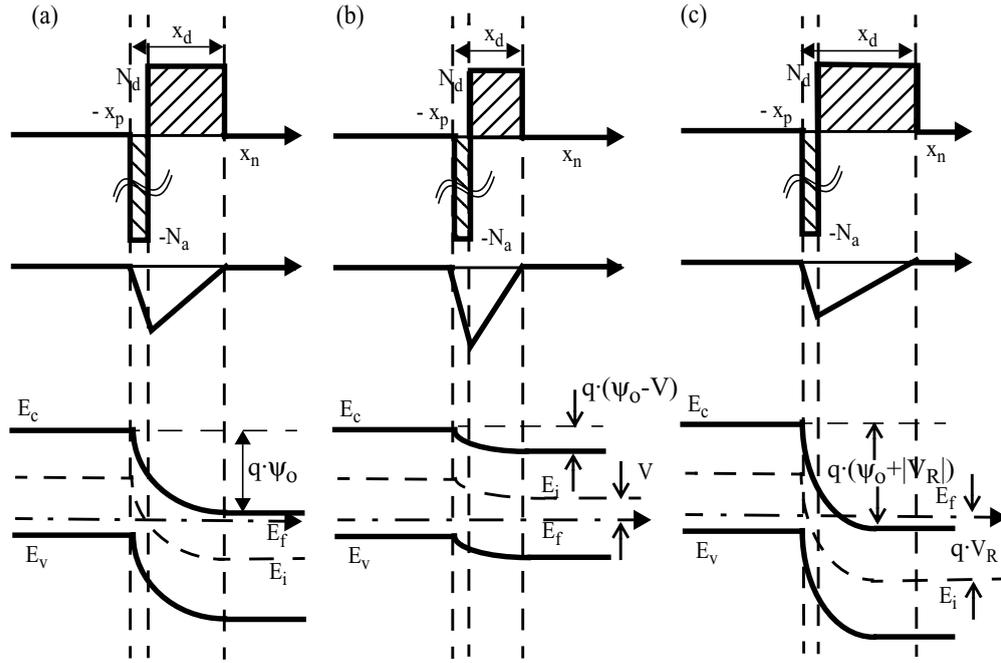


Fig. 2.14: Caracterización de la unión pn: (a) en equilibrio térmico; (b) polarizada directamente con valor de tensión V ; y (c) polarizada inversamente con un valor de tensión V_R

Según la ecuación 2.53, la tensión de polarización permite modular el ancho de la región de carga espacial.

Cuando la unión se polariza en directa los portadores se inyectan a través de la región de carga espacial. Para intensidades de corriente muy pequeñas, las densidades de corriente inyectadas no perturban la región de carga espacial de forma significativa, por lo que la aproximación de la ecuación 2.53 es válida¹⁶. Si la corriente inyectada crece, la densidad de portadores inyectados puede ser comparable a la densidad de iones en la unión, por lo que la condición de inyección débil deja de ser válida¹⁷.

La ecuación 2.53 permite modelar de forma aproximada el comportamiento capacitivo de una unión p^+n inversamente polarizada, asociando a la unión una capacidad de valor¹⁸:

$$C = A \frac{\epsilon_{sc}}{x_d} = A \sqrt{\frac{q\epsilon_{sc}}{2(\psi_0 + |V_R|)}} \left(\frac{1}{N_a} + \frac{1}{N_d} \right) \cong A \sqrt{\frac{q\epsilon_{sc}N_d}{2(\psi_0 + |V_R|)}} \quad (2.54)$$

-
16. Hay que tener en cuenta que ahora el potencial de la unión decrece, por lo que aparecerá la expresión $\psi_0 - V$ en lugar de $\psi_0 + |V_R|$, donde V es la tensión de polarización directa aplicada
 17. La condición de inyección débil se ha mantenido en el desarrollo de las expresiones desde la Secc. 2.2.6.
 18. La unión p^+n se modela como un condensador de láminas paralelas de superficie A

La expresión 2.54 es también una aproximación válida para la unión en polarización directa, siempre y cuando se verifique la condición $V < \psi_0$, donde V es la tensión de polarización directa aplicada.

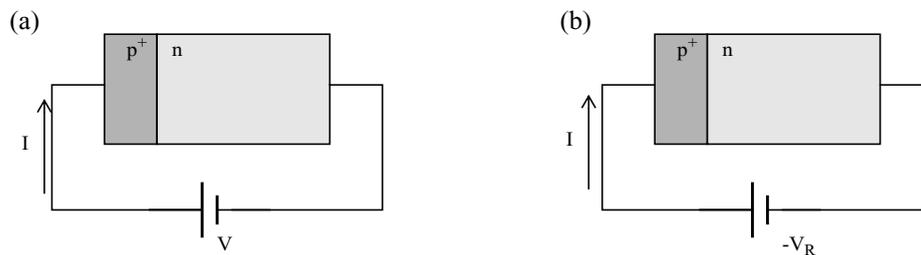


Fig. 2.15: Unión p^+n (a) directamente polarizada; e (b) inversamente polarizada

2.12.- CARACTERÍSTICA TENSIÓN-INTENSIDAD EN DC DE LA UNIÓN pn INVERSAMENTE POLARIZADA

La difusión de portadores minoritarios controla la conducción de corriente en una unión pn. Para aclarar esta afirmación, consideremos el comportamiento de los electrones y huecos en la región n de carga neutra en una unión pn inversamente polarizada. Bajo la condición de inyección débil, la densidad de huecos inyectados p_n es mucho menor que la densidad de portadores mayoritarios n_{n0} , pero mayor que p_{n0} . Los huecos en exceso se difundirán hacia la región n de carga neutra, al tener ésta una concentración menor de huecos. El transporte de huecos forma una corriente de difusión de portadores minoritarios. Al mismo tiempo, los huecos en exceso se recombinarán con los electrones, por lo que la densidad de huecos decrecerá con la distancia desde el límite de la región de deplexión.

Los huecos inyectados generarán momentáneamente un campo eléctrico. Este campo eléctrico actúa tanto sobre los portadores mayoritarios como sobre los minoritarios, generando corrientes de arrastre. Puesto que la densidad de portadores mayoritarios es mucho mayor que la de portadores minoritarios, el campo eléctrico da lugar a una corriente de arrastre de portadores mayoritarios importante, mientras que la corriente de arrastre de portadores minoritarios es despreciable. El campo eléctrico tiende a neutralizar los huecos inyectados para restablecer la neutralidad de carga espacial. La corriente de huecos es esencialmente de difusión. Los portadores mayoritarios responden a la inyección de portadores minoritarios, pero no al contrario. Por lo tanto, podemos ignorar el efecto de los portadores mayoritarios en el cálculo de la corriente de portadores minoritarios.

La densidad y la corriente de portadores minoritarios pueden calcularse asumiendo que la condición de neutralidad de carga prevalece, y que los portadores minoritarios son el único tipo de portadores presente. Estos portadores minoritarios serán transportados por difusión en las regiones de carga neutra. Esta aproximación se conoce como **aproximación de difusión**. Matemáticamente, la corriente de difusión de huecos se obtiene mediante la ecuación 2.46:

$$I_p = -qAD_p \frac{dp}{dx} \quad (2.55)$$

y la ecuación de difusión de huecos:

$$D_p \cdot \frac{d^2 \Delta p_n}{dx^2} - \frac{\Delta p_n}{\tau_p} = \frac{\partial p_n}{\partial t} \quad (2.56)$$

Consideremos el comportamiento de los portadores minoritarios en la región n de la unión. Si la región n es suficientemente ancha como para que todos los portadores inyectados en esta región se recombinen antes de alcanzar el contacto externo de la unión, es decir, puede considerarse que en el volumen de la unión se verifica $-(\partial p_n / \partial t) = 0$, la ecuación de continuidad para los huecos puede reescribirse (suponiendo distribución uniforme de las impurezas):

$$D_p \cdot \frac{d^2 \Delta p_n}{dx^2} - \frac{\Delta p_n}{\tau_p} = -\frac{\partial p_n}{\partial t} = 0 \quad (2.57)$$

donde $\Delta p_n = p_n - p_{no}$.

Para la solución a la ecuación 2.57, las condiciones de contorno a imponer:

$$p_n = \begin{cases} p_n(0) = p_{no} e^{V/\phi_t}; & \forall x = 0 \\ p_{no}; & \forall x = \infty \end{cases} \quad (2.58)$$

donde se toma como límite de la región de depleción $x=0$. La primera de las condiciones de contorno surge teniendo en cuenta la expresión del potencial de contacto¹⁹:

$$\psi_o = \phi_T \ln \frac{p_{po} n_{no}}{n_i^2} = \phi_T \ln \frac{p_{po}}{p_{no}} \quad (2.59)$$

y la condición de inyección débil: $p_p = p_{po}$.

La segunda condición de contorno es el resultado de considerar que todos los huecos se recombinan antes de alcanzar el contacto extremo de la unión.

La solución de la ecuación 2.57, bajo las condiciones de contorno 2.58, viene dada por:

$$\Delta p_n = p_{no} (e^{V/\phi_t} - 1) e^{-x/L_p} \quad (2.60)$$

donde:

$$L_p = \sqrt{D_p \tau_p} \quad (2.61)$$

es la **longitud de difusión de los huecos**. La longitud de difusión se define como la distancia media que atraviesa una carga inyectada antes de sufrir una recombinación.

La corriente de difusión de huecos se halla a partir de las ecuaciones 2.55 y 2.60:

$$I_p = I_p(0) e^{-x/L_p} \quad (2.62)$$

donde:

$$I_p(0) = qA \frac{D_p p_{no}}{L_p} (e^{V/\phi_t} - 1) \quad (2.63)$$

19. Esta expresión viene determinada por la aproximación: $\psi_o = \psi_n - \psi_p$, donde ψ_n y ψ_p son los potenciales en las regiones n y p de carga neutra

Puesto que la corriente total ha de permanecer invariante respecto al eje x para satisfacer la condición de continuidad de corriente, es decir: $I_n + I_p = I$, podemos deducir las ecuaciones equivalentes para la región p de la unión:

$$\Delta n_p = n_{po} (e^{V/\phi_t} - 1) e^{-x'/L_p} \quad (2.64)$$

$$I_n = I_n(0) e^{-x'/L_n}; \quad I_n(0) = qA \frac{D_n n_{po}}{L_n} (e^{V/\phi_t} - 1) \quad (2.65)$$

$$L_n = \sqrt{D_n \tau_n} \quad (2.66)$$

donde L_n es la longitud de difusión de los electrones, y x' es la distancia desde un punto de la región p hasta el límite de la región de deplexión (dimensión espacial en la región p de la unión).

La corriente total en la unión puede calcularse sumando las componentes de corriente de portadores minoritarios en los límites de la unión:

$$I = I_n(0) + I_p(0) = I_o (e^{V/\phi_t} - 1) \quad (2.67)$$

donde:

$$I_o = qA \left(\frac{D_n n_{po}}{L_n} + \frac{D_p p_{no}}{L_p} \right) \quad (2.68)$$

La ecuación 2.67 es aplicable para la unión polarizada tanto en directa como en inversa, siempre que se verifique la condición de inyección débil.

La expresión de la tensión en función de la intensidad resulta de interés para posteriores consideraciones:

$$V = \phi_T \cdot \ln \left(\frac{I}{I_o} + 1 \right) \quad (2.69)$$

Referencias



-
- [1] Edward S. Yang "Microelectronic Devices", McGraw-Hill
 - [2] R. F. Pierret "Modular Series on Solid State Devices, Volume I: Semiconductor Fundamentals", 2nd ed., Reading, MA: Addison-Wesley Publishing Company, 1989
 - [3] Sze. S.M. "Physics of Semiconductor Devices", 2nd Edition, Wiley, New York, 1981
 - [4] R. F. Pierret "Modular Series on Solid State Devices, Volume VI: Advanced Semiconductor Fundamentals", Reading, MA: Addison-Wesley Publishing Company, 1989
 - [5] B. Jahne, H. Haussecker, P. Geeissler "Hanbook Of Computer Vision and Applications", Academic Press, 1999

DISPOSITIVOS CMOS FOTOSENSIBLES

Una vez visto cómo se produce la absorción de fotones en un semiconductor centremos nuestro estudio en los dispositivos capaces de sensar la corriente fotogenerada por el flujo de luz incidente.

Las estructuras más comúnmente utilizadas para el sensado de la corriente fotogenerada son:

- Los *photodiodes*, basados en la absorción de fotones en una unión pn.
- Los *phototransistors*, formados por una unión npn o pnp¹.
- Las *photogates*, derivadas del principio de funcionamiento de los fototransistores.
- Otros dispositivos fabricados en tecnologías CMOS no estándares, de los que haremos un análisis muy superficial

Analicemos formalmente el comportamiento de cada una de estas estructuras.

3.1.- FOTODIODOS

La dependencia de la relación tensión-intensidad con el flujo de luz incidente en una unión pn permite el uso de ésta como dispositivo detector de la corriente que genera el flujo de luz (*corriente fotogenerada*). La unión pn utilizada con este propósito recibe el nombre de *photodiode*.

Existen principalmente tres modos de operación del fotodiode para la detección de la radiación luminosa: *modo fotocorriente*; *modo fotovoltaico*; y *modo por integración de carga*.

En los dos primeros modos de operación del fotodiode se hace una lectura continua en el tiempo del valor de la corriente fotogenerada, mientras que en el tercero, durante un cierto intervalo de tiempo, denominado *tiempo de integración*, se acumula la carga fotogenerada por el flujo de luz incidente en la capacidad de la unión pn, y es al final de este período de integración cuando se procede a la lectura de la carga acumulada.

1. Las uniones npn y pnp requieren un estudio en profundidad como extensión al que se ha realizado en el Cap. 2 para la unión pn. No vamos a entrar en detalle en este estudio, aunque sí se harán las correspondientes aclaraciones del comportamiento de estas estructuras cuando corresponda.

3.1.1.- Modos de operación del fotodiodo

3.1.1.1.- Modo de operación fotocorriente

En el modo fotocorriente el detector se polariza inversamente a una tensión determinada V_R y se mide la corriente que atraviesa el dispositivo, como se ilustra en la Fig. 3.1.

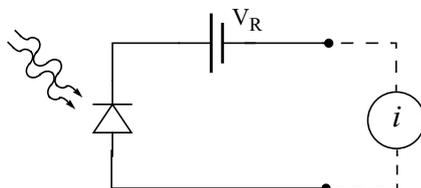


Fig. 3.1: Fotodiodo en modo de operación fotocorriente

De la ecuación 2.67 puede deducirse que la relación tensión-intensidad en el fotodiodo en modo fotocorriente, teniendo en cuenta la componente de corriente fotogenerada, viene dada por:

$$I = I_{ph} + I_o(e^{-V_R/\phi_t} - 1) \quad (3.1)$$

donde I_{ph} es la corriente fotogenerada; y V_R la tensión inversa de polarización.

La Fig. 3.2 ilustra la característica tensión-intensidad para un fotodiodo expuesto a diferentes niveles de potencia luminosa. En la gráfica aparece el concepto de **corriente oscura** de un fotodiodo. Se define como la corriente que atraviesa el fotodiodo en ausencia de iluminación.

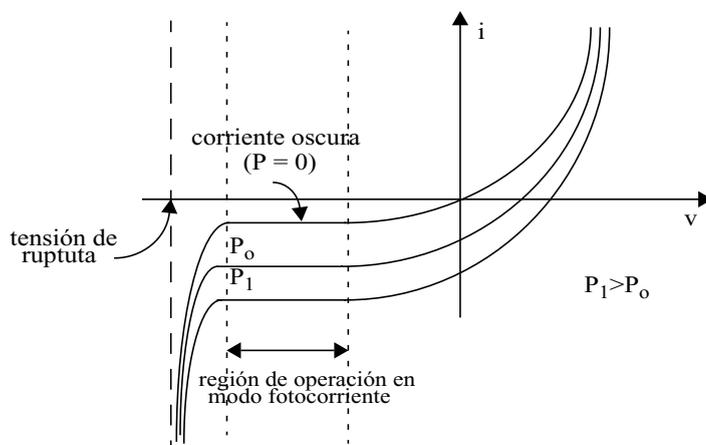


Fig. 3.2: Relación tensión-intensidad para diferentes niveles de potencia luminosa incidente (P).

La Fig. 3.3 muestra un esquema de utilización del fotodiodo en modo fotocorriente. El fotodiodo actúa como fuente de intensidad dependiente del flujo luminoso. La tensión en el nodo de salida dependerá de la intensidad que atraviesa el fotodiodo. Suponiendo que el amplificador operacional es ideal, esta dependencia viene dada por la relación:

$$V_o = V_R + R_F \cdot I = V_R + R_F \cdot (I_{ph} + I_o(e^{-V_R/\phi_t} - 1)) \quad (3.2)$$

Puede comprobarse que la tensión de salida V_o resulta proporcional a la corriente fotogenerada

$$I_{ph}: \quad V_o = k + R_F \cdot I_{ph} \quad (3.3)$$

donde k es una constante que incluye los efectos de la tensión inversa de polarización y la corriente oscura.

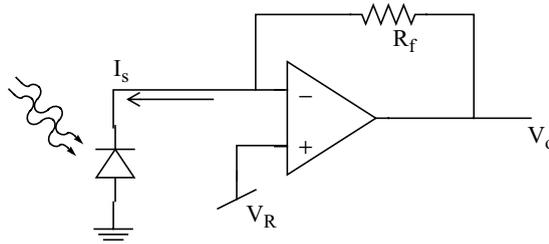


Fig. 3.3: Circuito para la medición de la corriente fotogenerada

3.1.1.2.- Modo fotovoltaico de operación

En el modo fotovoltaico no se aplica tensión de polarización, sino que se deja el fotodiodo en circuito abierto (Fig. 3.4). Para medir la tensión en circuito abierto del fotodiodo se hace uso de un amplificador con impedancia de entrada capacitiva.

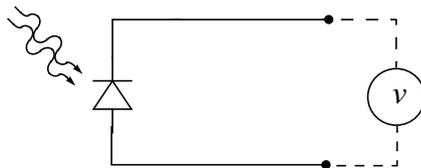
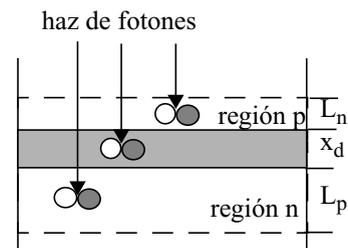


Fig. 3.4: Fotodiodo en modo de operación fotovoltaico

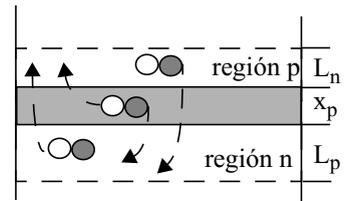
Estudiemus en detalle la conversión de la señal óptica en eléctrica en una unión pn en modo fotovoltaico.

Podemos describir el proceso en los siguientes pasos:

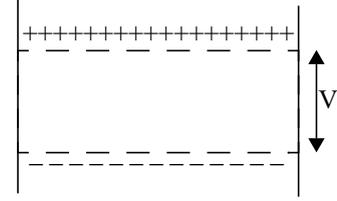
(1) Los fotones son absorbidos en las regiones p y n, generando pares de carga electrón-hueco, según los procesos descritos en la Secc. 2.6.



(2) Por difusión, en términos estadísticos, los electrones y huecos generados a una distancia de la unión menor que su longitud de difusión alcanzarán la región de carga espacial de la unión pn (zona sombreada).



(3) Los pares electrón-hueco generados dentro de la región serán arrastrados por el campo eléctrico de la unión pn.



(4) En modo fotovoltaico la unión está en circuito abierto, y los electrones y huecos a ambos lados de la unión producirán una diferencia de potencial que podrá ser sensada utilizando un amplificador de entrada capacitiva.

La corriente generada por el flujo de luz incidente tiene la misma dirección que la corriente inversa de saturación de la unión pn. Por lo tanto, la corriente total viene dada por:

$$I_R = I_{ph} + I_o(1 - e^{V/\phi_T}) \quad (3.4)$$

donde I_{ph} es la corriente fotogenerada y el segundo término es la corriente inversa del diodo.

La tensión en circuito abierto vendrá dada de imponer la condición $I_R=0$ en la ecuación 3.4:

$$V = \phi_T \cdot \ln\left(\frac{I_L}{I_o} + 1\right) \quad (3.5)$$

La expresión 3.5 proporciona la relación entre señal luminosa y eléctrica en la que basan su funcionamiento los fotodiodos en modo fotovoltaico.

3.1.1.3.- Modo de operación por integración de carga

El principio de funcionamiento del fotodiodo en modo por integración de carga se ilustra en la Fig. 3.5.

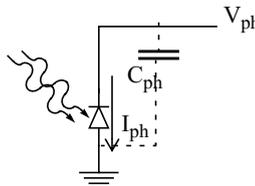


Fig. 3.5: Fotodiodo en modo de operación por integración de carga

Consideremos inicialmente la capacidad de la unión cargada a una tensión de referencia V_{ref} . Si sobre el fotodiodo se hace incidir un haz luminoso, la corriente fotogenerada por el flujo de luz incidente descargará la capacidad de la unión, de forma que la evolución en el tiempo de la tensión en circuito abierto del fotodiodo será:

$$V_{ph}(t) = V_{ref} - \int_0^t \frac{1}{C_{ph}(\tau)} \cdot I_{ph}(\tau) \cdot d\tau \quad (3.6)$$

En este modo de operación se define un periodo de tiempo, denominado **tiempo de integración**, durante el cual en la capacidad del fotodiodo se integra la carga fotogenerada. La tensión en el fotodiodo al final de la fase de integración vendrá dada por:

$$V_{ph}(t_{int}) = V_{ref} - \int_0^{t_{int}} \frac{1}{C_{ph}(\tau)} \cdot I_{ph}(\tau) \cdot d\tau \quad (3.7)$$

donde t_{int} es el tiempo de integración; $C_{ph}(\tau)$ la capacidad del fotodiodo en función del tiempo; e $I_{ph}(\tau)$ la corriente fotogenerada en función del tiempo.

Bajo la suposición de que tanto el flujo luminoso como la capacidad del fotodiodo permanecen constantes durante todo el periodo de integración, la ecuación 3.7 puede escribirse como:

$$V_{ph}(t_{int}) = V_{ref} - \frac{t_{int}}{C_{ph}} \cdot I_{ph} \quad (3.8)$$

es decir, la tensión en circuito abierto del fotodiodo decrece de forma proporcional a la corriente fotogenerada.

La lectura de la carga fotogenerada al final del periodo de integración puede llevarse a cabo siguiendo básicamente dos esquemas: inyectar, a través de un elemento resistivo, la corriente necesaria en el fotodiodo para que la capacidad de éste alcance la tensión de referencia V_{ref} de forma que la caída de tensión en el elemento resistivo (R) proporciona la información necesaria para la medición de la corriente fotogenerada (Fig. 3.6(a)); o bien llevar a cabo directamente la lectura de la tensión en el fotodiodo a través de un buffer y posteriormente fijar la tensión de la capacidad del mismo a V_{ref} haciendo uso de una llave adicional (S) (Fig. 3.6(b)).

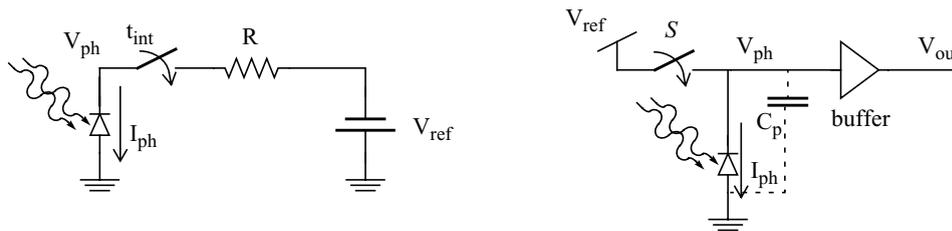


Fig. 3.6: Esquemas de lectura de la carga fotogenerada en un fotodiodo por integración de carga

3.1.2.- Fotodiodos implementables en tecnología CMOS estándar

La Fig. 3.7 muestra el corte transversal de los fotodiodos implementables en una tecnología CMOS estándar de pozo n, así como los perfiles de dopado típicos de las mismas para un proceso CMOS estándar de $0.8 \mu\text{m}$, resultados de la simulación de los procesos de AMS² [1]. Esta figura permite introducir los conceptos de **detector limitado por la longitud de difusión** (dispositivos (a) y (c) de la Fig. 3.7), en el que el espacio en el que la carga fotogenerada puede ser recogida viene determinada por la longitud de difusión de los portadores de carga; y **detector limitado en volumen** (dispositivo p⁺/n⁻ de la Fig. 3.7(b), con el pozo polarizado a V_{dd}), en el que el volumen en el que la carga fotogenerada es recogida viene limitado por una barrera de potencial creada por una unión pn. En el dispositivo (b) de la Fig. 3.7 esta barrera de potencial está generada por el potencial de contacto que aparece en la unión pn pozo/substrato.

La carga total fotogenerada en un fotodiodo puede descomponerse de la siguiente manera:

$$I_{ph} = I_{ph}|_{deplexion} + I_{ph}|_{vertical} + I_{ph}|_{superficial} + I_{ph}|_{lateral} \quad (3.9)$$

Analicemos cada componente:

- $I_{ph/deplexión}$ es la corriente fotogenerada en la región de carga activa (o región de deplexión). En principio, toda la carga fotogenerada en esta región podrá sensarse.
- $I_{ph/vertical}$ es la carga fotogenerada por los fotones que penetran profundamente en el semiconductor y que se difunde verticalmente a través del semiconductor hasta la región de carga activa.
- $I_{ph/superficial}$ es la carga fotogenerada cerca de la superficie de incidencia del flujo de fotones. Contribuirá de forma menos significativa a la corriente fotogenerada, pues la frecuencia de recombinación crece de forma considerable a medida que nos acercamos a la superficie³.
- $I_{ph/lateral}$ se debe a difusión lateral de portadores. Este fenómeno incrementa el factor de relleno efectivo, pues aumenta la zona de recogida de carga.

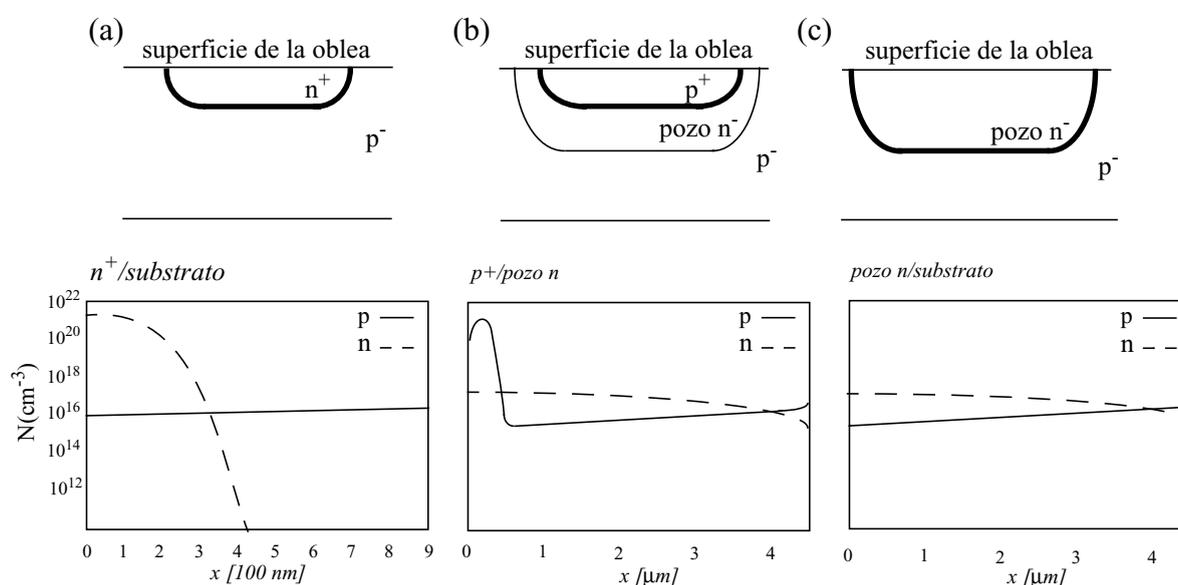


Fig. 3.7: Fotodiodos implementables en una tecnología CMOS estándar de pozo n

La absorción de fotones en las regiones profundas del semiconductor mejora la eficiencia cuántica del dispositivo, pues incrementa la componente $I_{ph/vertical}$ de la corriente total fotogenerada, llegando a ser un factor crítico para tecnologías submicra, en las que $I_{ph/deplexión}$ se reduce considerablemente al reducirse el ancho de la región de deplexión. Sin embargo, los portadores generados profundamente en el semiconductor pueden sufrir múltiples procesos, de forma que sólo una proporción de estos portadores contribuirán a $I_{ph/vertical}$ mientras que el resto o bien sufrirán procesos de recombinación (lo que implica una pérdida de la información de la incidencia del fotón en el sensor), o bien alcanzan el terminal óhmico de substrato de la oblea (normalmente fijado a la tensión de tierra por lo que se pierde la información de la generación de la carga), o bien generarán una componente de acoplamiento o “cross-talk” (la carga fotogenerada alcanza una región de deplexión distinta a la que atravesó el fotón que la generó). La pérdida de información en estas regiones y por acoplamiento conllevan una reducción de la **sensibilidad a**

3. En la Secc. 2.6 se describe la recombinación superficial en un semiconductor

las longitudes de onda largas (en el caso del espectro de luz visible se conoce como **reducción de la sensibilidad al rojo**), además de producir un deterioro del MTF⁴ para el espectro del rojo y del infrarrojo.

La reducción de $I_{ph}/superficial$ debido a los procesos de recombinación superficial produce una **reducción de la sensibilidad a las longitudes de onda cortas**, efecto conocido como **reducción de la sensibilidad al azul** para sensores que trabajen en el rango de longitudes de onda de luz visible.

Gran parte de la carga fotogenerada se recoge en la capacidad del fotodiodo por difusión de portadores a través del semiconductor, por lo que el diseño ha de optimizarse para conseguir que la carga se difunda hacia la región de deplexión correspondiente⁵ y para reducir el ruido por acoplamiento⁶. Aun optimizando el diseño para maximizar la sensibilidad, la eficiencia cuántica que se consigue en estos dispositivos es bastante baja.

Para caracterizar un dispositivo sensor de luz podemos hacer uso del *producto* $\eta * FF$, donde η es la eficiencia cuántica del dispositivo y FF es el **factor de relleno** (relación entre la superficie que ocupa el dispositivo fotosensor y la superficie del píxel⁷). Este parámetro refleja el compromiso de diseño que surge entre la eficiencia y el tamaño del sensor⁸. Es bastante difícil encontrar sensores basados en fotodiodos con un producto $\eta * FF$ mayor del 60%; lo normal es que este parámetro tome valores menores del 20% [2].

Además del bajo producto $\eta * FF$, otro inconveniente de este tipo de sensores es que la capacidad del fotodiodo (capacidad de la unión) resulta elevada para aplicaciones de sensado de luz. Como ejemplo característico de un fotodiodo, para un píxel de tamaño $14 \times 14 \mu\text{m}^2$ fabricado en un proceso CMOS estándar de $0.8 \mu\text{m}$, caracterizado por un FF del 60%, la capacidad del fotodiodo es de aproximadamente 93 fF, y la capacidad total de integración 102 fF, que equivale a una ganancia de conversión de intensidad en tensión de $1.6 \mu\text{V}/e^-$, por lo que la sensibilidad del sensor resulta bastante baja [2].

Puesto que la capacidad de la unión se ve reducida con el escalado de las tecnologías, y dado que la componente principal de la corriente fotogenerada es la correspondiente a la difusión de portadores (tanto por difusión vertical y como por difusión lateral), al escalar los dispositivos manteniendo constante la superficie del fotodiodo la eficiencia cuántica permanecerá (en primera aproximación) constante, y la capacidad de la unión disminuirá [3].

3.2.- FOTOTRANSISTORES

Se entiende por **fototransistor** al uso para la fotodetección del transistor BJT vertical que aparece en las tecnologías CMOS estándar⁹. El esquema del transistor BJT y los perfiles típicos de dopado se muestran en la Fig. 3.8 para la tecnología de AMS $0.8 \mu\text{m}$ de pozo n [1].

-
4. La definición y descripción del MTF se lleva a cabo en la Secc. 3.5.2
 5. Resulta interesante plantearse un semiconductor en el que el gradiente de concentraciones del semiconductor conduzcan la carga fotogenerada a distancias profundas verticalmente hacia las regiones de deplexión.
 6. Es necesario plantearse métodos de apantallamiento en el caso en el que las superficies de incidencia de la luz queden muy próximas, pues el ruido por acoplamiento será considerable.
 7. Superficie total ocupada por el sensor: fotodiodo, transistor de reset y circuitería a nivel de píxel
 8. Estos compromisos se estudian en detalle en la INTRO SECCIÓN, donde se discute sobre el tamaño óptimo del píxel

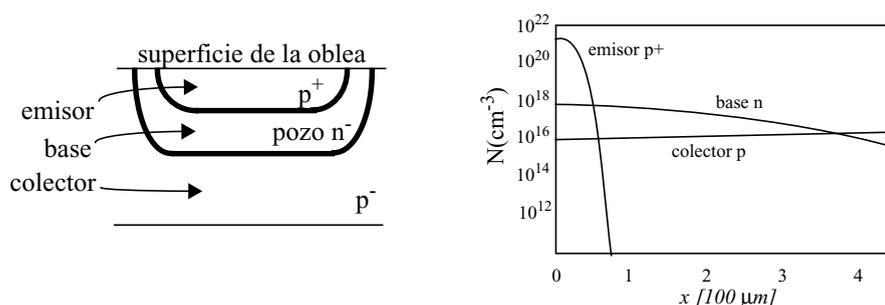


Fig. 3.8: Transistor BJT vertical: (a) corte transversal del dispositivo; y (b) perfiles de dopado típicos

Con el uso de los fototransistores se aprovecha la amplificación de la corriente de emisor que estos dispositivos presentan. El principio de operación de un transistor bipolar para el fotosensado es el siguiente: cuando la base del transistor bipolar está flotando, los portadores de carga fotogenerados en esta región que alcanzan la unión base-colector contribuyen a la corriente de colector. Además de esta componente, los portadores mayoritarios generados o inyectados en la base desde el colector se compensan con la inyección de portadores minoritarios desde el emisor a la base. El potencial resultante en la unión base-emisor hace que el transistor esté polarizado en activa. La corriente de emisor puede expresarse como:

$$I_E = I_C + I_B = (1 + h_{FE})I_B = (1 + h_{FE})I_{ph} \quad (3.10)$$

donde I_{ph} es la corriente fotogenerada en la unión pn base-colector y h_{FE} la ganancia de corriente del transistor.

La eficiencia cuántica efectiva del fototransistor resulta $(1+h_{FE})$ veces mayor que la de un fotodiodo, pudiendo ser mucho mayor que la unidad.

3.3.- FOTOPUERTAS

Una fotopuerta implementa en tecnología CMOS estándar el principio de transferencia de carga de las tecnologías CCD, y por tanto, responden a un esquema de operación por integración de carga. La Fig. 3.9 muestra el esquema de una fotopuerta. La **fotopuerta (PG)** será la encargada de generar el pozo de potencial para la recogida de la carga fotogenerada durante el periodo de integración, mientras que la **puerta de transmisión (TG)** servirá para transportar la carga desde la fotopuerta hasta el **nodo de sensado** (difusión flotante en la que se realizará la lectura de la carga transferida desde la fotopuerta).

Analicemos el funcionamiento de la fotopuerta (Fig. 3.9) sin tener en cuenta los efectos de la distancia de separación entre los terminales PG y TG. Durante la fase de integración, se polariza la fotopuerta de forma que se genera un pozo de potencial debajo del terminal. La carga fotogenerada se difundirá a través del semiconductor hacia este pozo de potencial, donde será almacenada (Fig. 3.10(a)). Al finalizar el periodo de integración se conectará TG, de forma que el pozo de potencial generado por este terminal permitirá la redistribución de la carga fotogenerada (Fig. 3.10(b)). A continuación se desconecta PG, eliminándose el pozo de potencial correspondiente, por lo que la carga fotogenerada queda distribuida entre el pozo de potencial generado por TG y la capacidad de la unión del nodo de sensado (Fig. 3.10(c)). Finalmente, se

9. Se intenta aprovechar las ventajas de amplificación de corriente que una estructura pnp o npn tiene frente a una unión pn.

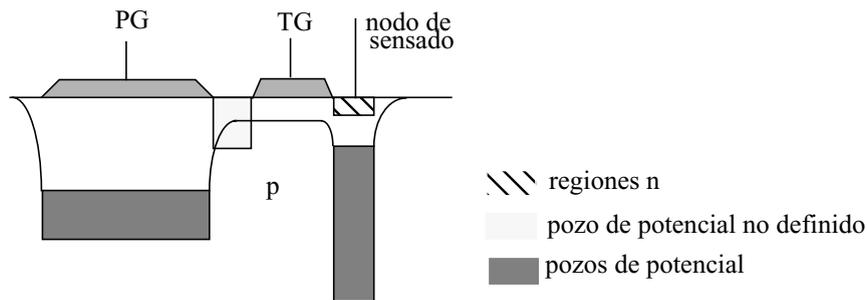


Fig. 3.9: Diagrama de una fotopuerta implementada en tecnología CMOS estándar

desconecta *TG*, y toda la carga se desplazará hacia el nodo de sensado (Fig. 3.10(d)). La transferencia desde el nodo *PG* hasta el nodo de sensado queda completada.

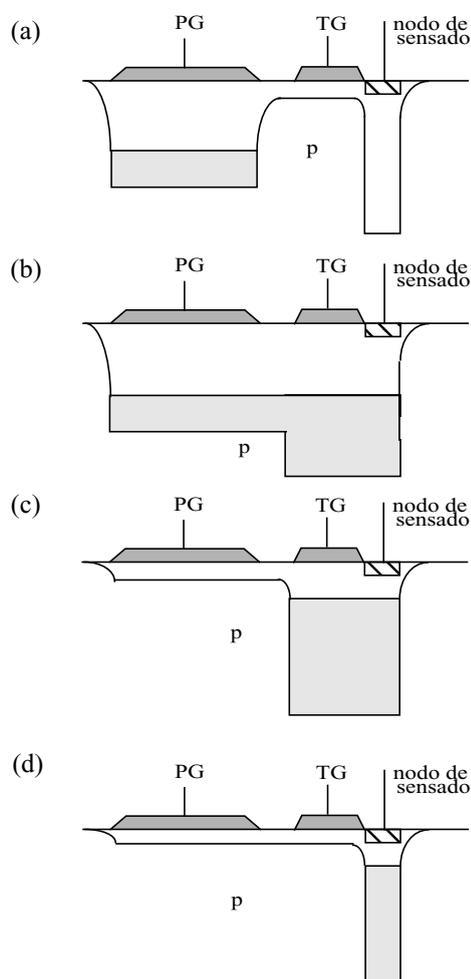


Fig. 3.10: Esquema de funcionamiento de una fotopuerta

Con este esquema de funcionamiento se consigue aislar la capacidad del dispositivo fotosensor de la capacidad empleada para sensar la corriente fotogenerada. De esta forma se logra que la mayor parte del área fotosensible (área del fotodiodo) no contribuya a la capacidad de conversión de carga en tensión. Esto permite que el coeficiente de conversión de intensidad en tensión

para una fotopuerta sea bastante mayor que el de un fotodiodo ($44 \mu\text{V}/e^-$ [4]), lográndose sensibilidades altas.

El alto factor de ganancia que consiguen estas estructuras es contrarrestado con la pérdida de flujo luminoso que provocan las capas superpuestas sobre la fotopuerta (necesarias en el proceso de fabricación CMOS estándar para la implementación de la fotopuerta), al ser éstas poco transparentes (22% de transparencia de la capa de polisilicio para una longitud de onda de 500 nm, y valores aún menores para longitudes de onda del azul [5]). Esta pérdida de sensibilidad puede contrarrestarse con un aumento de la superficie de las fotopuertas, alcanzándose valores de conversión de flujo luminoso en tensión de 17.5 V/lxs [6], pero esta medida de diseño implica un aumento del coste del sensor y una disminución de la *resolución espacial*¹⁰ del mismo. Para un píxel cuadrado de lado $16 \mu\text{m}$ implementado en tecnología CMOS estándar de $0.8 \mu\text{m}$ se consiguen valores típicos en torno a 4 V/lxs [4].

La tecnología CCD (Apéndice A) se basa en la superposición de los terminales a través de los cuales se va ir realizando el trasvase de la carga fotogenerada, de forma que en todo momento el pozo de potencial queda perfectamente definido, y la carga se va transfiriendo de un pozo de potencial a otro sin que se genere un residuo de carga en los pozos. Las tecnologías CMOS estándar no permiten esta superposición entre terminales, es decir, los terminales PG y TG no pueden superponerse; ha de existir una distancia de separación entre estos terminales definida por la tecnología. Esta separación genera una región en la que los pozos de potencial de PG y TG no se encuentran, por lo que el trasvase de carga del pozo de potencial de PG al pozo de potencial de FD no es posible. De todo esto se deriva que en la región señalada en la Fig. 3.9 como pozo de potencial no definido es necesaria la inserción de una difusión flotante para hacer posible la transferencia de carga en esta región. Esta difusión incrementa la capacidad durante la fase de transferencia, lo que deteriora el factor de conversión del dispositivo. Además, la capacidad de esta difusión flotante no se resetea, por lo que se generará una componente de carga residual, por lo que el algoritmo CDS no puede implementarse correctamente. En definitiva, las características de estos dispositivos respecto al factor de conversión de intensidad en tensión y al ruido se ven deterioradas con la inserción de la difusión flotante entre los terminales PG y TG.

3.4.- SENSIBILIDAD ESPECTRAL Y EFICIENCIA CUÁNTICA DE LOS ELEMENTOS FOTODETECTORES

Como se ha comentado en apartados anteriores, la absorción de luz en un semiconductor depende de la longitud de onda. La dependencia de la absorción de luz con la longitud de onda de la misma implica que los fotorreceptores formados a partir de uniones con difusiones a diferentes profundidades tendrán diferentes repuestas espectrales.

En este apartado estamos interesados en el estudio de la eficiencia cuántica de los diferentes elementos fotodetectores implementados en tecnología CMOS. El cálculo de este parámetro resulta complejo, pues depende de la tecnología. Es decir, para conocer cuál es el valor de la eficiencia cuántica de los elementos fotodetectores que podemos implementar en un sensor necesitaríamos previamente implementar sobre un integrado todos los componentes posibles y realizar las medidas oportunas para poder comparar resultados. Es decir, particularizar el estudio para la tecnología concreta sobre la que se vaya a realizar posteriormente el sensor. En función de estas medidas podríamos comparar el comportamiento particular de las estructuras y

10. La resolución espacial es función del tamaño del píxel, y mide la definición con la que una imagen es capturada con el sensor. Para un tamaño de sensor fijo (superficie del semiconductor donde se implementa el sensor de imagen), a menor tamaño de píxel, mayor definición; y viceversa.

decidir cuál es la que mejor se adapta a las necesidades del sensor. Esto resulta costoso, tanto en tiempo como en presupuesto, además de que la evolución tecnológica obligaría a hacer un estudio de estas características cada vez que cambiasen los parámetros de la tecnología con la que trabajásemos.

Otro planteamiento que cabe hacerse es el extrapolar los datos de un estudio previo, intentando predecir cuál será el comportamiento de cada estructura en función de los parámetros de los que dependen. Es de esta forma como vamos a proceder en este apartado: basándonos en el estudio comparativo de estructuras realizado por T. Delbrück & C. A. Mead en 1996 [7] y en los trabajos realizados por Markus Loose [8] intentaremos obtener los factores determinantes que afectan a cada estructura.

En el primero de los estudios citados se implementa en un mismo integrado los seis elementos fotodetectores ilustrados en Fig. 3.11. La tecnología usada es una tecnología BiCMOS¹¹ de pozo n de 2 μm . Los dispositivos (b), (d) y (f) se implementan aprovechando la capa enterrada (*buried layer*) disponible en tecnología BiCMOS, y por lo tanto, no son implementables en tecnología CMOS estándar. El emisor n⁺⁺ es semiconductor dopado con arsénico, con una profundidad de 0.3 μm , y una concentración de dopado de 10^{20} cm^{-3} . La base p⁺ se consigue dopando el semiconductor con boro, con una profundidad de 0.45-0.5 μm , y una concentración de dopado de $1-2 \cdot 10^{17} \text{ cm}^{-3}$. El pozo n se dopa con fósforo, y tiene una profundidad de 3 μm aproximadamente, con una concentración de dopado de $3-4 \cdot 10^{14} \text{ cm}^{-3}$. El sustrato tipo p tiene una concentración de dopado de $3-4 \cdot 10^{14} \text{ cm}^{-3}$.

Una de las principales diferencias entre los resultados obtenidos viene determinada por la distinción entre *detectores limitados por la longitud de difusión*, en los que el espacio en el que la carga fotogenerada puede ser recogida viene determinada por este parámetro; y los *detectores limitados en volumen*, en los que el volumen en el que la carga fotogenerada puede ser recogida viene determinada por una barrera de potencial creada por una unión pn.

Los transistores bipolares incorporan una ganancia de corriente. Para medir la eficiencia cuántica en ellos se contará toda la carga recogida, incluyendo la ganancia, por lo que la eficiencia cuántica podrá tomar valores mayores que la unidad.

La respuesta espectral de la eficiencia cuántica para cada una de las uniones se muestra en la Fig. 3.12. De la figura pueden obtenerse las siguientes observaciones:

- 1.- Como ya se ha comentado, hay una clara distinción entre los fotodetectores que recogen la carga fotogenerada en un volumen limitado por una barrera de potencial (estructuras (b), (c), (d) y (f) de la Fig. 3.11); y los fotorreceptores cuyo volumen de recogida de carga está limitado por la longitud de difusión de los portadores (estructuras (a) y (e) de la Fig. 3.11). Estos últimos son más sensibles a las longitudes de onda larga.
- 2.- Las respuestas espectrales son de tipo paso de banda. Todos los detectores cubren un rango que abarca al espectro de luz visible, además de tener respuestas que son aproximadamente planas dentro del espectro de luz visible.
- 3.- Las uniones poco profundas formadas dentro de los pozos (Fig. 3.11(b), (c) y (d)) son más sensibles a las longitudes de onda en torno a los 500 nm; la sensibilidad de la unión

11. Los dispositivos fotosensores se describen en la Secc. 3.5.1

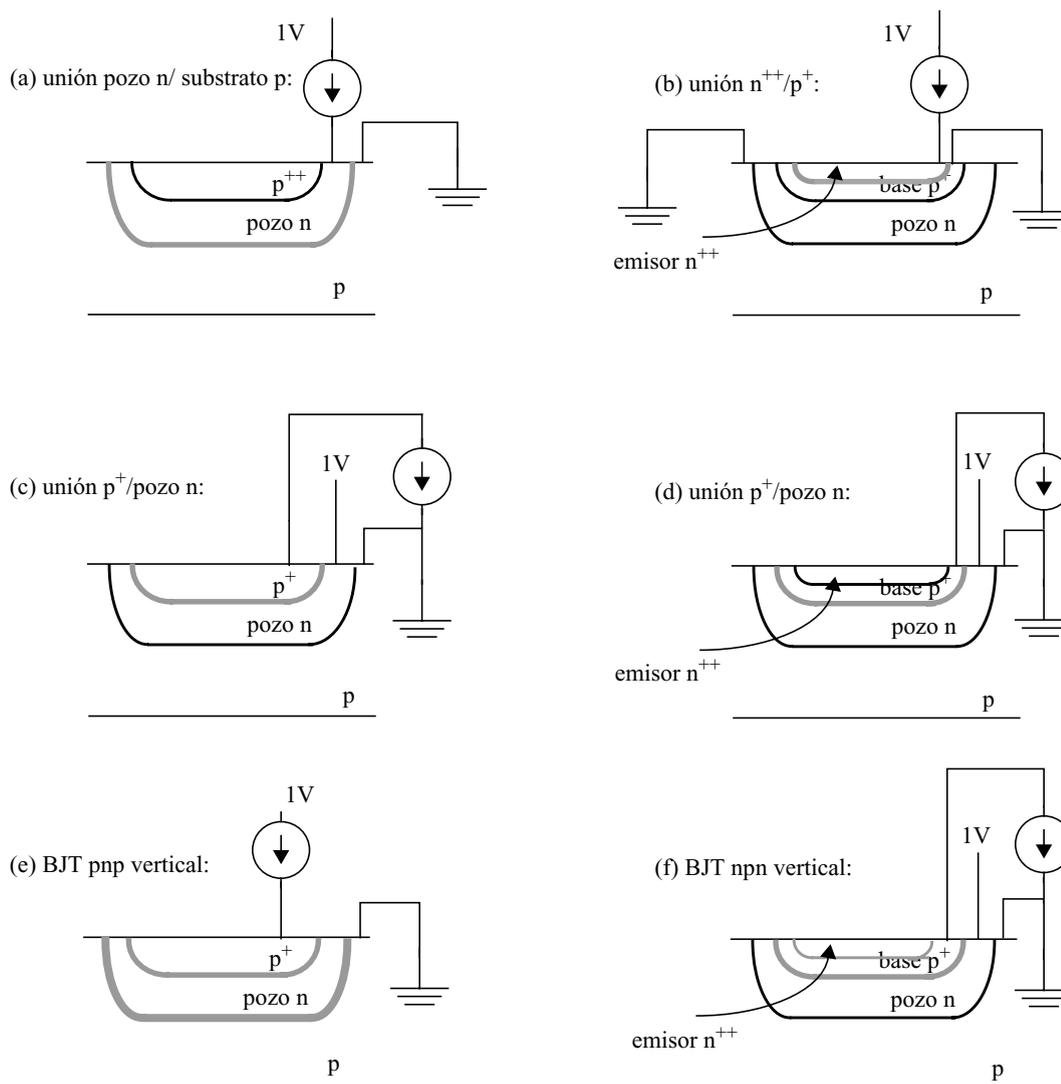


Fig. 3.11: Elementos fotodetectores implementados para el estudio de la eficiencia cuántica. Las estructuras (a), (b), (c) y (d) son fotodiodos, mientras que (e) y (f) son fototransistores. Las uniones activas (uniones pn que generan la región de carga activa) están señaladas en gris. Las fuentes de tensión en intensidad que aparecen son las utilizadas para la medición de la eficiencia cuántica.

profunda pozo n/substrato p (Fig. 3.11(a)) tiene un pico en torno a 900 nm, bien alejado del rango de luz visible.

4.- El pico absoluto de la eficiencia cuántica para los fotodiodos varía entre un máximo de 0.8 cercano al infrarrojo en la unión pozo n/substrato p (Fig. 3.11(a)), hasta un mínimo de 0.3 para las uniones poco profundas, limitadas en volumen (Fig. 3.11(b), (c) y (d)).

5.- Las respuestas de todos los dispositivos tienen un valor aproximado de 0.3 para el límite azul del espectro, tanto para los fotodiodos como para los fototransistores.

6.- Las respuestas espectrales de los fototransistores están muy cerca de múltiplos de las respuestas de la unión pn profunda. Por ejemplo, la respuesta del BJT pnp vertical es muy similar en la forma a la respuesta espectral de la unión pozo n/substrato p. Una observación similar puede hacerse para el transistor vertical npn. Esta observación significa simplemente que la mayor parte de la corriente de los fototransistores procede de la unión profunda¹².

7.- La ganancia de corriente es aproximadamente de 100 para el transistor bipolar vertical pnp, y de 30 aproximadamente para el transistor bipolar vertical npn. Esta ganancia tiene una fuerte dependencia con el nivel de corriente, y decrece tanto para intensidades altas como para bajas.

8.- Todas las respuestas espectrales muestran un corte en torno al límite de la banda en la cual la energía del fotón coincide con el ancho de la banda prohibida. Este corte no está perfectamente definido, y se extiende pasado el límite real de la banda. La eficiencia cuántica decrece una cantidad e por cada 25 meV en torno al límite teórico de la banda.

9.- Hay un aspecto interesante en la respuesta espectral de la unión p^+ /pozo n justo en torno al límite de la banda (señalado como regiones sombreadas en la figura) que puede observarse también en el transistor bipolar vertical npn. Este efecto podría ser debido a la existencia de centros de recombinación/generación superficiales generados en la implementación de la base p, y que provocan un estiramiento de la respuesta espectral en una fracción adicional de 1 eV.

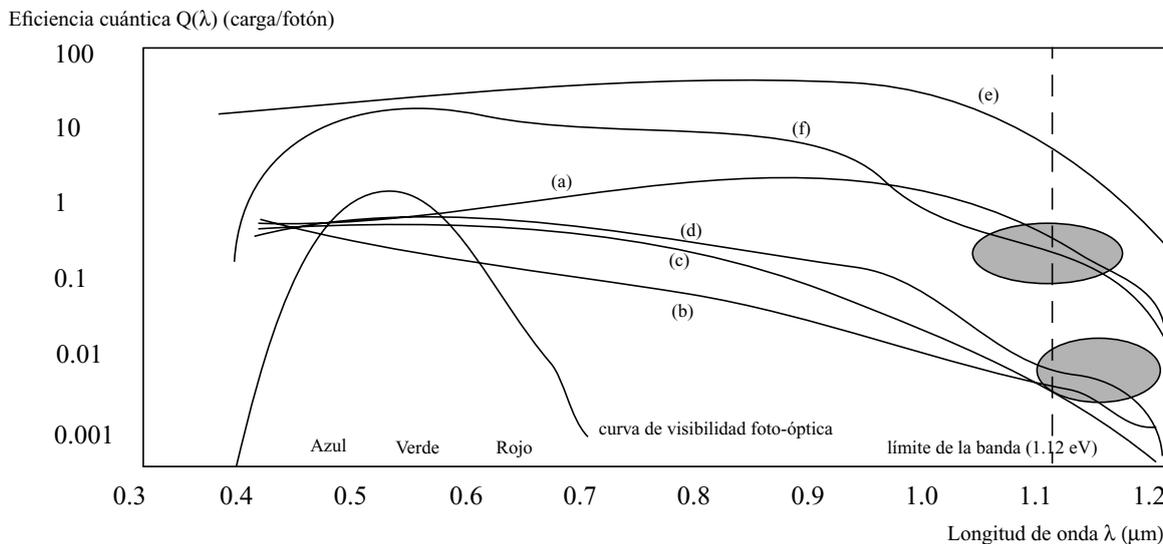


Fig. 3.12: Respuestas espectrales de la eficiencia cuántica de los dispositivos de la Fig. 3.11

El segundo de los estudios que presentamos es el realizado por Markus Loose [8]. En este estudio se realiza la simulación de las eficiencias cuánticas para los fotodiodos de la Fig. 3.7 y para el fototransistor de la Fig. 3.8. La Fig. 3.13 muestra la simulación¹³ de la eficiencia cuántica de los dispositivos implementados CMOS en tecnología de pozo n de $0.8 \mu\text{m}$ sobre sustrato de silicio en función de la longitud de onda, con una tensión inversa de polarización de -2V [8].

12. Además de la contribución de la unión pn profunda existirá una contribución de la unión pn superficial, aunque de valor mucho menor, al ser el volumen de recogida de carga en esta región mucho más limitado.
13. En realidad las curvas predichas en la simulación difieren de las medidas experimentales. La principal diferencia viene determinada por el fenómeno de múltiples reflexiones que los fotones sufren en las capas de metal, no predicho en la simulación presentada. En el Secc. 3.5.4 se lleva a cabo un estudio detallado de este fenómeno de múltiples reflexiones.

De las curvas de simulación de la eficiencia cuántica en función de la longitud de onda de la Fig. 3.13 pueden obtenerse las siguientes conclusiones:

- 1.- El diodo p^+ /pozo n es relativamente ineficiente debido a la estrecha región de depleción y a que el volumen de recogida de carga es pequeño (detector limitado en volumen). Los portadores de carga que son absorbidos fuera del pozo se encuentran con la barrera de potencial de la unión pozo n/p y no tienen oportunidad de ser detectados por el diodo. Sin embargo, esto permite que el tiempo de respuesta de los dispositivos sea muy bueno, y que la componente de ruido por acoplamiento sea pequeña. La sensibilidad máxima está localizada en el rango de longitudes de onda menores, en torno a $\lambda=530$ nm. La razón de esto es que la unión está localizada cerca de la superficie, donde la mayor parte de las longitudes de onda cortas son absorbidas.
- 2.- El diodo n^+ /substrato p es más sensible debido a que el volumen de recogida de carga es mayor (detector limitado por la longitud de difusión de los portadores). Por esta razón las longitudes de onda largas pueden detectarse de forma eficiente y el máximo de la eficiencia cuántica se localiza en torno a $\lambda=620$ nm. Sin embargo el tiempo de respuesta aumenta en comparación con el diodo p^+ /pozo n debido a que el proceso de difusión es lento.
- 3.- Finalmente, el diodo pozo n/substrato p muestra un comportamiento similar al de la estructura n^+ /substrato p. Es ligeramente más sensible debido a que la región de depleción es mayor y localizada a mayor profundidad en el semiconductor.

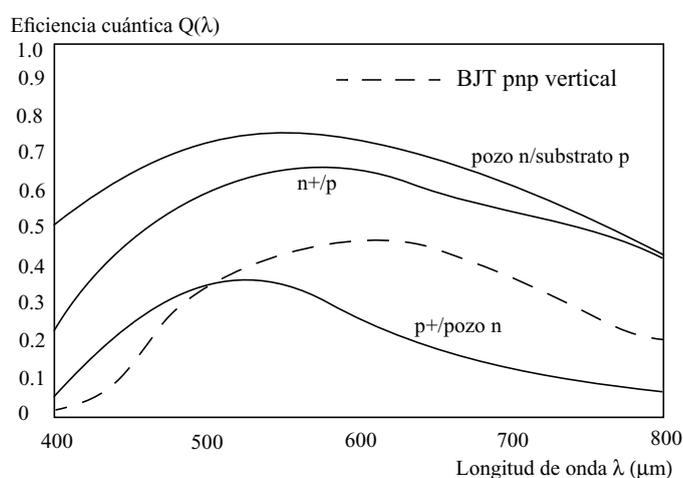


Fig. 3.13: Simulación de los espectros frecuenciales de las eficiencias cuánticas de los fotodiodos CMOS [1]. La curva del fototransistor se encuentra escalada $\times 0.1$

3.5.- OTROS ASPECTOS IMPORTANTES EN LOS SENSORES DE IMAGEN

3.5.1.- Dispositivos fotosensores implementados en tecnologías CMOS no estándares

Presentamos a continuación algunos elementos fotodetectores que, por sus propiedades, resultan interesantes, aun cuando están implementados en tecnologías CMOS no estándares. El inconveniente de los sensores implementados en tecnologías no estándares es que el coste de la implementación de los prototipos es muy alto, por lo que la investigación y desarrollo de los mismos requiere de un presupuesto mucho más elevado que el necesario para tecnologías estándares. La diferencia entre los costes de la fabricación de los sensores en una u otra tecnología

puede verse compensada por el volumen de producción de circuitos integrados, pero el coste del desarrollo de dispositivos no estándares queda sólo al alcance de fuertes inversiones económicas.

3.5.1.1.- *Proceso de fabricación BiCMOS*

En el proceso de fabricación BiCMOS se añade una capa de difusión intermedia (*buried layer*) entre la oblea de semiconductor y la posterior capa semiconductor que se añade mediante crecimiento epitaxial, como se muestra en la Fig. 3.14. Este implante añadido permite la implementación de transistores bipolares adicionales sin la necesidad de hacer uso de los pozos (en el caso de una tecnología de pozo n permite la implementación de transistores bipolares npn verticales). Por lo tanto, el uso de este proceso de fabricación permite disponer de estructuras adicionales que podrán usarse para la fotodetección. Las propiedades de estos nuevos dispositivos diferirán de los anteriormente propuestos principalmente debido a la profundidad a la que se implementan, así como a las condiciones de contorno que las rodeen, como son las posibles barreras de potencial que limiten la región de carga activa.

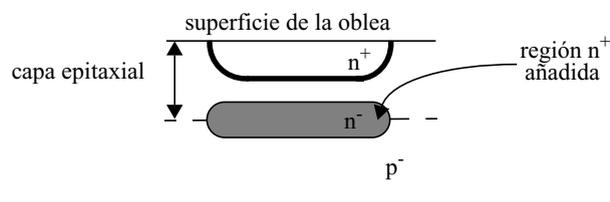


Fig. 3.14: Transistor bipolar vertical implementado en tecnología BiCMOS de pozo n

3.5.1.2.- *Fotodiodos "pinned"*

Una solución prometedora al principal problema que presentan las fotopuertas¹⁴ es el uso de fotodiodos "pinned" como elemento fotodetector (Fig. 3.15). El fotodiodo "pinned" implementado en una tecnología CMOS de pozo n consiste básicamente en una **estructura p⁺np⁻**; una **puerta de transmisión** (TG)¹⁵; una difusión flotante para el sensado de la corriente fotogenerada (**nodo de sensado**); y un **transistor de reset** (Res)

El funcionamiento del fotodiodo pinned se basa en la estructura p⁺np⁻ (para una tecnología de pozo n), en la que ambas capas p están al potencial de sustrato. Al incrementar la tensión en la capa n las regiones de deplexión de ambas uniones pn aumentan la una hacia la otra. Para una determinada tensión V_p (tensión **pinned**) las regiones de deplexión se encuentran y ningún portador mayoritario puede extraerse del dispositivo.

Para operaciones de sensado de luz el fotodiodo pinned se polariza inicialmente a la tensión V_p. Durante la fase de recogida de la carga fotogenerada, los portadores de carga mayoritarios fotogenerados se almacenan en la región de deplexión (o región de carga activa) que se genera en la unión p⁺np⁻ del fotosensor. Antes de la fase de lectura, la difusión flotante ha de fijarse a una

14. La existencia de un pozo de potencial no definida entre las puertas de fotogeneración y de transmisión obliga a introducir una difusión flotante entre los terminales de fotodetección y transmisión que deteriora notablemente las características de sensibilidad y ruido del fotodetector
15. Al igual que en las fotopuertas esta puerta sirve para transportar la carga desde la estructura p⁺np⁻ hasta el nodo en el que se sensa la carga fotogenerada, aislando el dispositivo de fotosensado del dispositivo de sensado de la carga fotogenerada

tensión de referencia¹⁶ (V_{ref}). Una vez que se ha fijado y leído esta tensión, la puerta de transferencia se abre y se lleva a cabo la transferencia completa de la carga fotogenerada. Esta secuencia de operación asegura que la señal de salida quede completamente libre de carga residual¹⁷. La transferencia completa de carga tiene lugar si la tensión en el nodo de sentido (FD) queda por encima de la tensión V_p mientras el fotodiodo pinned opera por debajo de esta tensión¹⁸.

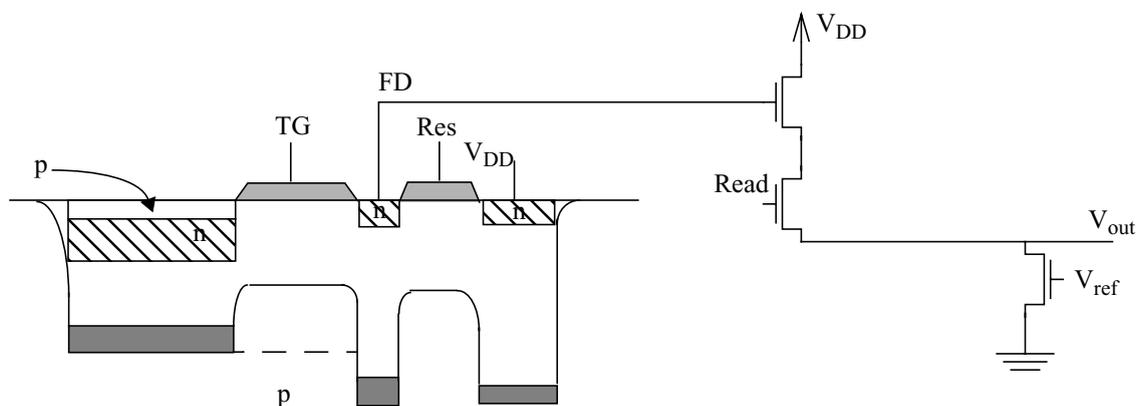


Fig. 3.15: Diagrama de un píxel basado en un fotodiodo "pinned" implementado en tecnología CMOS de pozo n

El mayor beneficio de esta estructura deriva de la completa transferencia de carga desde el dispositivo fotodetector (caracterizado por una capacidad relativamente alta) hasta el nodo de sentido (difusión flotante, de capacidad mucho menor), por lo que el principal inconveniente de las fotopuertas (transferencia incompleta de carga) es eliminado. Una segunda ventaja respecto a las fotopuertas es que la carga que es recogida por el dispositivo es fotogenerada lejos de la superficie semiconductor debido a la existencia de la capa p^+ , evitándose el aumento considerable de la corriente oscura que generan los estados energéticos superficiales. Al mismo tiempo se evita la aparición de puntos blancos sobre la imagen final, producto de las altas densidades de centros de recombinación superficiales [9].

Alternativamente a la estructura presentada en la Fig. 3.15, podría prescindirse del transistor de transferencia [10]. Bajo esta consideración, toda la carga fotogenerada se recoge en FD durante la fase de integración. Como ventajas se reduce la complejidad del píxel (reducción de la circuitería de control), y se incrementan el factor de relleno (disminuye el tamaño del píxel) y la sensibilidad. Además, V_p puede ser una tensión muy pequeña puesto que ahora no hemos de imponer la relación entre las tensiones de operación del fotodiodo y de la puerta de transmisión. Como desventaja de eliminar el transistor de transferencia el ruido kTC no puede eliminarse completamente, pues se rompe la secuencia reset-transferencia-lectura que nos permite la implementación del algoritmo CDS para completa eliminación de la carga residual de la señal de salida.

16. Esta tensión de reset debe leerse para poder implementar la técnica CDS de eliminación del ruido de patrón fijo, como se justificará en posteriores capítulos.
17. Por carga residual se entiende el residuo de carga que pueda quedar en la capacidad de integración de anteriores muestras.
18. Esta condición se expresa gráficamente en la Fig. 3.15 con línea discontinua: permite el "volcado" de la carga almacenada en la unión p^+np^- en el pozo de potencial generado por la puerta de transmisión.

Para una determinada tensión V_p como objetivo de diseño se puede elegir entre varias combinaciones de profundidad y dopado. La profundidad y los dopados deben controlarse rigurosamente para asegurar un margen suficiente de V_p por debajo y por encima del rango de operación del dispositivo. Dado que tres niveles de dopado deben superponerse, el nivel de dopado del sustrato de la primera capa debe ser moderado, lo que implica la necesidad de tecnologías de dos pozos. Estas medidas añaden complejidad y coste al proceso CMOS, desviándose éste de las cadenas de producción estándares.

A modo comparativo, tanto las fotopuertas como los fotodiodos pinned se basan en la transferencia de carga desde un dispositivo de recogida de carga (fotosensor, caracterizado por una capacidad relativamente alta) a una región de difusión (nodo de sensado, de capacidad relativamente baja). La transferencia de carga se hace de forma más efectiva en los fotodiodos pinned. Este factor, unido a que la superficie de sensado no tiene superpuesta ninguna capa conductora¹⁹ hace que la eficiencia cuántica de éstos es mayor que la de las fotopuertas. La circuitería de control de los fotodiodos pinned es más sencilla, pues no se requiere el control de la tensión del terminal de PG existente en las fotopuertas²⁰, lo que reduce el cableado del píxel. Todo esto hace que el concepto de fotodiodo pinned tenga ventajas sobre el concepto de fotopuerta. Sin embargo, la desventaja de la necesidad de máscaras adicionales en el proceso de fabricación reduce mucho las posibilidades de estudiar y desarrollar estos dispositivos.

3.5.1.3.- Fotodiodos fabricados en tecnología TFA

Una aproximación utilizada para incrementar la sensibilidad de los sensores de imagen es la implementación de sensores que permitan la integración de la carga fotogenerada en dispositivos que aprovechen la tres dimensiones espaciales, con lo que se consigue una sensibilidad muy alta sin que esto sea incompatible con la resolución del sensor (área de superficie de sensado del píxel). Este concepto es el que se persigue con la implementación de sensores en tecnología TFA (Thin-Film-on-ASIC). El sensor en esta tecnología se implementa a partir de un sistema multicapa de silicio amorfo que se deposita sobre un circuito integrado de aplicación específica (ASIC), como se muestra en la Fig. 3.16. El silicio amorfo mejora la conversión de luz en fotocorriente, mientras el circuito ASIC es el responsable de la integración, lectura y procesamiento de la señal.

Entre las propiedades del silicio amorfo destaca que el coeficiente de absorción para el espectro de luz visible es aproximadamente 20 veces mayor que el del silicio cristalino, y la respuesta transitoria a la luz es suficientemente buena para la mayoría de las aplicaciones de imágenes. El proceso de implementación de los dispositivos fotosensores permite alcanzar un contraste bastante bueno entre píxeles vecinos sin la necesidad de añadir procesos fotolitográficos en la implementación del integrado para aislar los píxeles. El contraste local entre píxeles toma valores en torno a los 86 dB [2].

El diagrama de un píxel TFA se muestra en la figura 3.10. Consiste en un electrodo “rear”, un diodo implementado en silicio amorfo y un electrodo implementado con óxido conductivo transparente (TCO). El factor determinante en la implementación es la deposición del fotodiodo sobre el ASIC. Con esto, todo el área del píxel está disponible para el fotodiodo, por lo que el factor de relleno del sensor es de prácticamente del 100%. Existe una pérdida de carga fotoge-

-
19. En las fotopuertas es necesario el terminal PG, por lo que habrá que superponer una capa conductora (bien una capa de polisilicio, bien una metalización)
 20. Más sencilla resulta aún si se prescinde de la puerta de transferencia TG

nerada en el electrodo “rear”, pues al atravesar esta capa se encontrarán con trampas en las que se producirá la recombinación de los pares electrón-hueco fotogenerados.

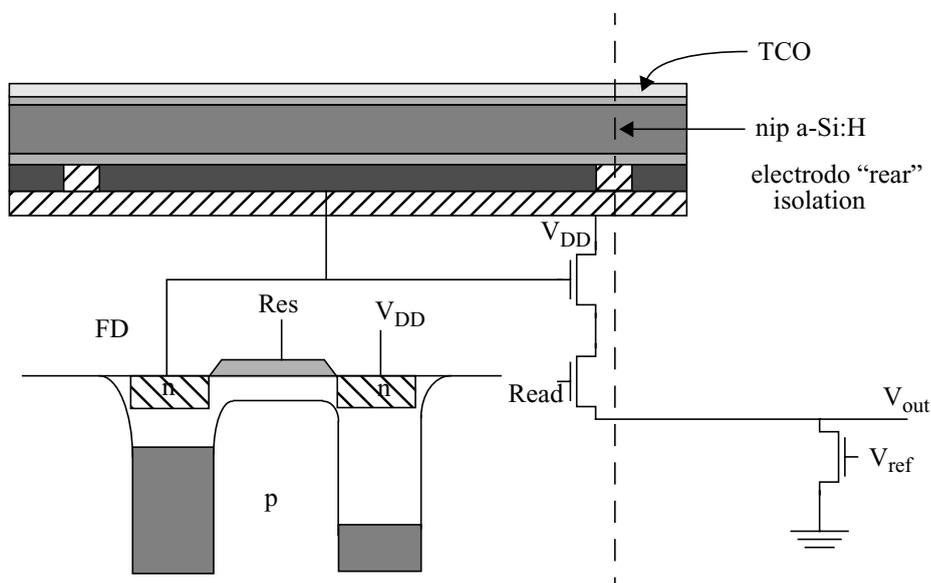


Fig. 3.16: Diagrama de un píxel fabricado en tecnología TFA

Excepto por el electrodo “rear” no existen otras capas que contribuyan a la reducción de la de carga fotogenerada por la pérdida de fotones, como ocurría en las capas de metalización y óxido grueso en el caso de los elementos fotodetectores basados en uniones pn dentro del sustrato semiconductor. Las pérdidas por reflexión se minimizan para la tecnología TFA optimizando el ancho de la capa de TCO.

Un valor típico de la eficiencia cuántica de estos dispositivos está en torno al 60% en la mayor parte del espectro de luz visible, alcanzando un máximo del 80%. Esto permite trabajar con valores de corriente fotogenerada de $1\text{ fA}/\mu\text{m}^2$ para un nivel de iluminación de 1 lux [2]. Como ejemplo de implementación en tecnología TFA de $0.8\ \mu\text{m}$, el sensor ocupa un área cuadrada de 10 mm de lado, y la capacidad de sensado toma un valor de 20 fF, equivalente a $8\ \mu\text{V}/e^{-21}$, es decir, un factor de conversión de 3.1 V/lx.

3.5.2.- Medición de la calidad de un sistema óptico. Función de transferencia de la modulación de la señal óptica

La calidad de un sistema óptico puede caracterizarse en el dominio espacial mediante la **respuesta impulsiva** del sistema, o bien en el dominio frecuencial mediante la transformada de Fourier de la respuesta impulsiva, es decir, mediante su **función de transferencia**.

Se define la respuesta impulsiva de un sistema como la respuesta de éste al ser excitado con una señal impulso. Consideremos un sensor de imagen ideal. Al excitar el sensor con un impulso, por ejemplo, una imagen en negro con un sólo punto brillante, la respuesta de éste (imagen de salida) será una nueva imagen en negro con un sólo punto brillante. Es decir, el sensor no distorsiona el impulso de entrada. Esto se traduce en que la respuesta impulsiva del sistema es un

21. frente a los $1.6\ \mu\text{V}/e^{-}$ de característico de los fotodiodos (Secc. 3.1) y los $44\ \mu\text{V}/e^{-}$ de las fotopuertas (Secc. 3.3)

impulso (Fig. 3.17(a)). Sin embargo, debido a difracciones y aberraciones en el sistema, un sistema real tiene una respuesta impulsiva con un ancho distinto de cero, de forma que la respuesta del sistema a un impulso pasa a ser un impulso distorsionado, es decir, de ancho distinto de cero (Fig. 3.17(b)).

La imagen a la salida del sistema (*respuesta*) puede calcularse como la convolución de la imagen de entrada (*excitación*) y de la respuesta impulsiva del sistema:

$$g(x, y) = f(x, y) ** h(x, y) \quad (3.11)$$

donde $f(x,y)$ es la señal de entrada (excitación); $h(x,y)$ la respuesta impulsiva del sistema; y $g(x,y)$ la respuesta del sistema a la excitación (Fig. 3.17(c)). El doble asterisco (**) indica que la convolución se lleva a cabo en las dos dimensiones del espacio. Una respuesta impulsiva estrecha implicará una mejor calidad de imagen que una respuesta impulsiva ancha.

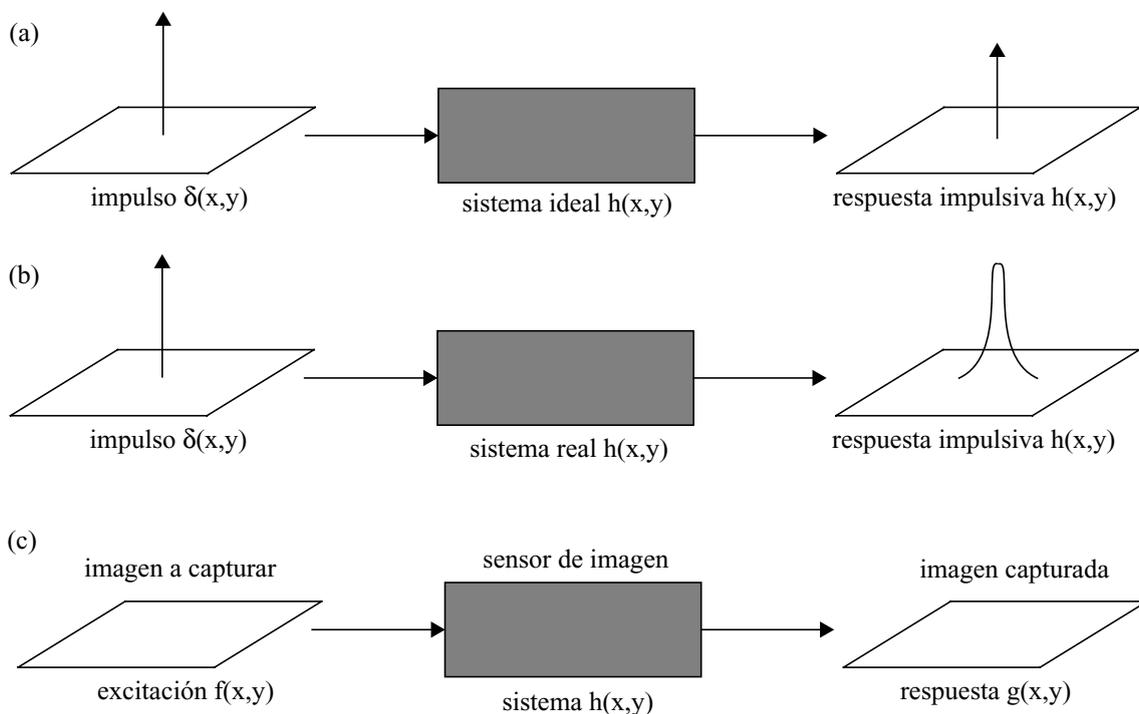


Fig. 3.17: Respuestas impulsivas de un sistema (a) ideal; (b) real; y (c) relación entre excitación y respuesta de un sensor de imagen

Alternativamente al dominio espacial, podemos considerar el comportamiento del sensor en el dominio frecuencial. En este contexto, estamos interesados en la respuesta del sensor a imágenes senoidales de diferentes frecuencias, en lugar de la respuesta a objetos puntuales (impulsos). La distribución de la irradiancia de un objeto puede considerarse como una composición de componentes de frecuencias espaciales, de la misma forma que en el dominio del tiempo una señal puede considerarse como composición de varias frecuencias temporales a través del análisis de Fourier.

Consideremos la distribución de irradiancia como una función de las coordenadas cartesianas x e y . Esta distribución de irradiancia representa a una imagen. Del perfil de una de las dimensiones de la distribución obtenemos una función de una variable que puede descomponerse en sus diferentes frecuencias espaciales mediante el análisis de Fourier.

A través de esta descripción en el dominio espacial, un objeto o imagen puede descomponerse en una serie de componentes sinusoidales caracterizadas por un periodo y una frecuencia espacial (X y $\xi=1/X$ respectivamente) (Fig. 3.18). A su vez, un objeto bidimensional puede descomponerse en un conjunto básico de senoidales de dos dimensiones. En coordenadas cartesianas, cualquier senoide dada con una orientación determinada tiene una frecuencia espacial (ξ) asociada al eje x , que se corresponderá con la inversa de la distancia entre crestas de la senoide a lo largo del eje x , así como una frecuencia espacial (η) asociada al eje y que resultará inversamente proporcional a la distancia entre crestas de la senoide a lo largo del eje y .

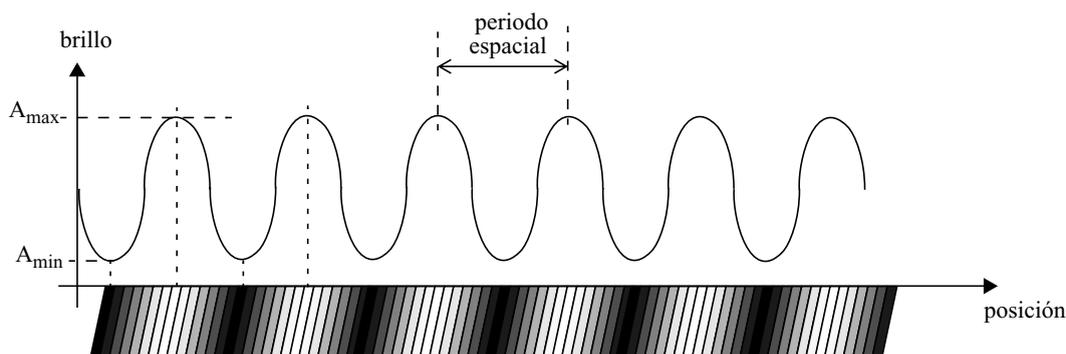


Fig. 3.18: Distribución sinusoidal de la amplitud del brillo en el espacio

La convolución de la imagen con la respuesta impulsiva para obtener la respuesta del sensor supone una pérdida de detalle de la imagen original, es decir, un filtrado. Este filtrado de imagen puede verse claramente en el dominio espacial mediante el teorema de la convolución. Aplicando la transformada de Fourier a ambos lados de la ecuación 3.11 se obtiene:

$$G(\xi, \eta) = F(\xi, \eta) \times H(\xi, \eta) \quad (3.12)$$

donde con mayúsculas se indican las transformadas de Fourier de las correspondientes funciones en minúscula. $H(\xi, \eta)$ es la transformada de Fourier de la respuesta impulsiva, es decir, la función de transferencia.

Normalmente $H(\xi, \eta)$ se normaliza, de forma que tome el valor unidad para la frecuencia espacial cero. El valor de la función de transferencia normalizada tomará el valor relativo respecto al valor de la función a baja frecuencia, con lo que se normaliza cualquier factor de atenuación que sea independiente de la frecuencia espacial. La función de transferencia $H(\xi, \eta)$, normalizada de esta manera, se conoce como **función de transferencia óptica** (OTF).

La **función de transferencia en modulación** o **MTF** (Modulation Transfer Function) puede definirse como la magnitud de la respuesta del sistema óptico a sinusoides de distinta frecuencia espacial. Se calcula en base a considerar como entradas al sistema señales sinusoidales en lugar de impulsos.

Definimos el índice de modulación como:

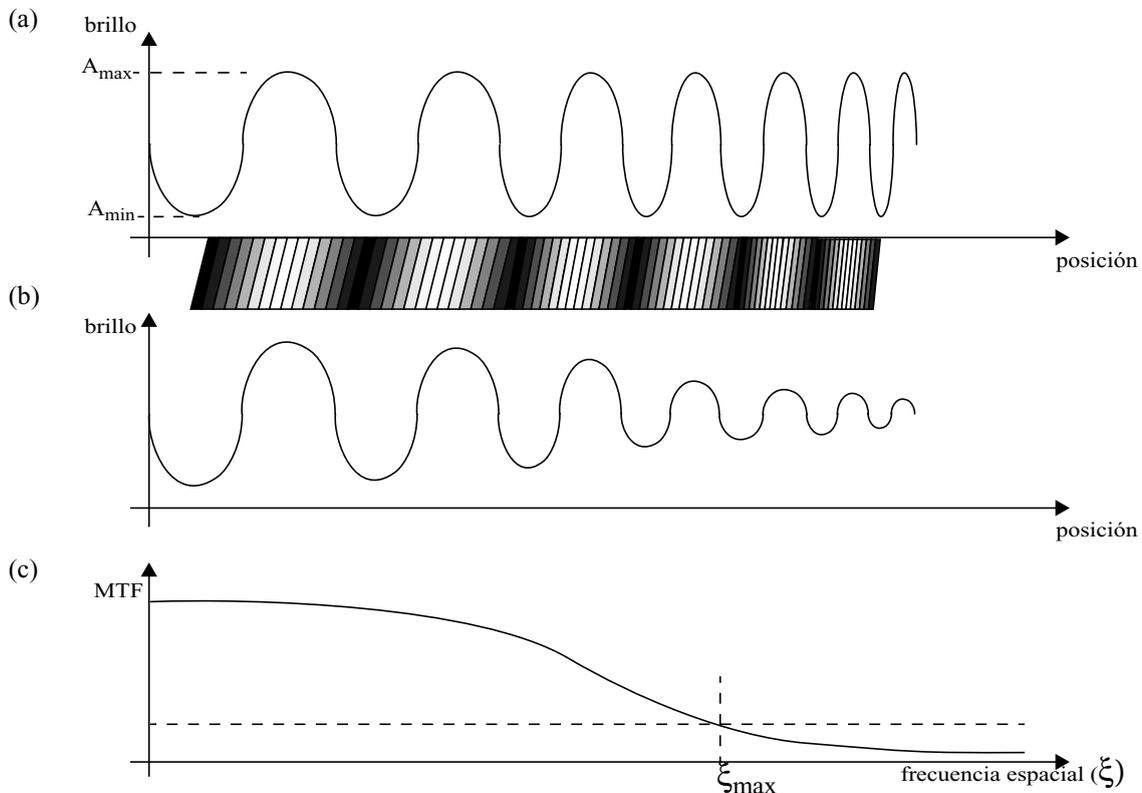
$$M = \frac{A_{max} - A_{min}}{A_{max} + A_{min}} \quad (3.13)$$

donde A_{max} y A_{min} son los valores máximo y mínimo del brillo del objeto o de la imagen (excitación) en unidades apropiadas (radiancia, excitancia, o irradiancia) (ver Fig. 3.18). El índice de

modulación es el cociente entre el valor ac de la señal y el valor dc de la misma. Puede observarse que puesto que el brillo es estrictamente positivo se verificará $0 \leq M \leq 1$.

Un sistema óptico lineal e invariante en el tiempo tendrá como respuesta a una senoide otra senoide. En general, en estos sistemas, la senoide de salida tendrá un desplazamiento en fase respecto a la señal de entrada, y el índice de modulación (en comparación con la señal de entrada) se habrá reducido. De hecho, el efecto de la resolución limitada del sistema óptico será la reducción el índice de modulación en la imagen en comparación con el índice de modulación en el objeto o imagen original.

La forma práctica de medir la MTF de un sistema óptico consistirá en capturar con el sensor una senoide de frecuencia espacial variable (Fig. 3.19(a)). La MTF podrá calcularse directamente a través del índice de modulación en la forma de onda de salida. De hecho, la MTF puede definirse como la dependencia con la frecuencia espacial del índice de modulación, asumiendo un índice de modulación constante en la excitación (Fig. 3.19(c)).



Debido a la limitación de la resolución espacial en los sistemas ópticos, los valles de las senoideas en la imagen resultante serán menos profundos, mientras que los picos serán menos altos (ver Fig. 3.19(b)), por lo que el índice de modulación decrece en comparación con el índice de modulación del objeto.

Para una frecuencia espacial dada ξ , la **transferencia de la modulación** se define como el cociente entre el índice de modulación de la imagen (respuesta) y el índice de modulación del objeto (excitación). Resulta evidente que la transferencia en modulación depende de la frecuencia espacial. La limitación de la resolución se manifiesta de forma drástica a frecuencias espaciales altas, para las cuales el nivel de detalle de la imagen es más pequeño. Con el incremento de la

frecuencia espacial, la imagen se hace más borrosa, y la transferencia de la modulación decrece. Si dibujamos la transferencia de la modulación como función de la frecuencia espacial, asumiendo una modulación constante del objeto) obtenemos la MTF:

$$MTF(\xi) = \frac{MTF_{imagen}(\xi)}{MTF_{objeto}} \quad (3.14)$$

La **resolución** de un sistema óptico, a partir de la MTF, puede definirse como la frecuencia espacial (ξ_{max}) para la cual la MTF cae por debajo de un cierto límite (índice de modulación mínimo detectable).

Las expresiones matemáticas de la MTF resultado de un análisis unidimensional de la MTF se presentan a continuación [11]. Consideremos una fila de píxeles caracterizados por un ancho de la superficie de fotosensado Δx , y una separación entre píxeles P_{pix} (Fig. 3.20). La señal sinusoidal de entrada (excitación), necesaria para el cálculo de la MTF, puede expresarse como:

$$f(x) = H_o(1 + A_o \cdot \cos(2\pi\xi x)) \quad (3.15)$$

donde H_o y A_o son las componentes dc y ac respectivamente; y ξ la frecuencia espacial.

La señal de salida $g(x)$ del píxel localizado en x_o puede calcularse como la integral de la señal de entrada entre los límites $(x_o + \Delta x/2)$ y $(x_o - \Delta x/2)$:

$$g(x) = \frac{1}{\Delta x} \int_{x_o - \Delta x/2}^{x_o + \Delta x/2} [H_o(1 + A_o \cdot \cos(2\pi\xi x))] \cdot dx = H_o[1 + M(\xi) \cdot \cos(2\pi\xi\Delta x)] \quad (3.16)$$

donde M es el índice de modulación de la señal $g(x)$:

$$M(\xi) = A_o \frac{\sin(\pi\xi\Delta x)}{\pi\xi\Delta x} \quad (3.17)$$

La MTF del sistema podemos definirla como el cociente entre los índices de modulación de las señales de excitación y respuesta:

$$MTF = \frac{MTF_{g(x)}}{MTF_{f(x)}} = \frac{\sin(\pi\xi\Delta x)}{\pi\xi\Delta x} = \text{sinc}(\pi\xi\Delta x) \quad (3.18)$$

La inversa de la distancia de separación (P_{pix}) entre dos píxeles consecutivos se define como la frecuencia de separación (ξ_s). A partir de la frecuencia de separación podemos definir la frecuencia de Nyquist del sistema $\xi_{nyquist}$ como:

$$\xi_{nyquist} = \frac{\xi_s}{2} = \frac{1}{2P_{pix}} \quad (3.19)$$

Si normalizamos la frecuencia ξ en la expresión 3.18 respecto a $\xi_{nyquist}$, la MTF queda como:

$$MTF = \frac{\sin\left(\frac{\pi\Delta x}{2P_{pix}} \frac{\xi}{\xi_{nyquist}}\right)}{\frac{\pi\Delta x}{2P_{pix}} \frac{\xi}{\xi_{nyquist}}} = \text{sinc}\left(\frac{\pi\Delta x}{2P_{pix}} \frac{\xi}{\xi_{nyquist}}\right) \quad (3.20)$$

3.5.3.- Absorción de fotones en las capas depositadas sobre la oblea

Para que el fotón incidente sea detectado electrónicamente es necesario que se produzca su absorción en la oblea semiconductor. Sin embargo, antes de alcanzar la oblea semiconductor, el

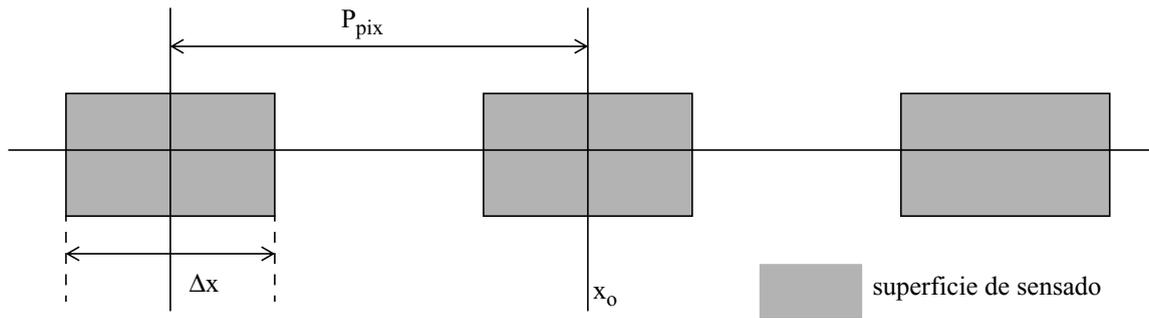


Fig. 3.20: Esquema de una fila de píxeles

fotón deberá atravesar la interfaz existente entre ésta y la superficie del integrado. Esta interfaz está compuesta por las diferentes capas depositadas sobre la oblea semiconductor para la implementación de los dispositivos. La Fig. 3.21 muestra el corte transversal de un transistor nmos y un transistor pmos implementados en tecnología CMOS estándar de un solo polisilicio. No estamos interesados en el proceso tecnológico en sí, sino en cómo cada una de estas capas superpuestas sobre la oblea afecta al fotón incidente. Desde este punto de vista analicemos cuáles son los fenómenos físicos a los que se ve sometido el fotón antes de alcanzar la oblea semiconductor, y que pueden evitar que éste sea finalmente detectado.

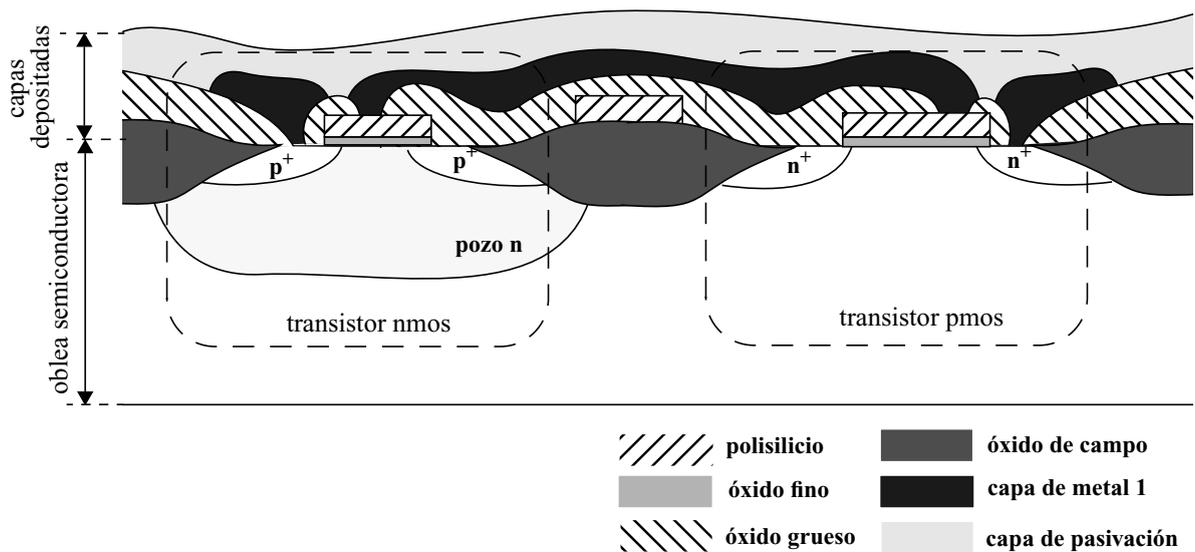


Fig. 3.21: Corte transversal de un transistor pmos y un transistor nmos en tecnología CMOS estándar de una capa de polisilicio

En primer lugar, el fotón ha de atravesar la interfaz entre el medio en el que se encuentre el sensor de imagen (generalmente aire) y la capa de pasivación. Se producirán fenómenos de **reflexión en la superficie del integrado**²², de forma que un porcentaje de los fotones incidentes serán rechazados.

Una vez atravesada la interfaz medio/capa de pasivación, el fotón viajará a través del material con el que se implementa la capa de pasivación. Éste suele ser un material aislante, por lo que la energía del fotón no será suficiente como para que sea absorbido en un proceso de generación

22. Los procesos de reflexión de los fotones se explican en la Secc. 3.5.4

de un par electrón-hueco. Sin embargo, tras atravesar esta región el fotón se encontrará con una nueva interfaz²³ por lo que podrá sufrir una nueva reflexión, de forma que, al encontrarse en su camino de vuelta con la interfaz capa de pasivación/medio, el fotón se vea sometido a un **fenómeno de múltiples reflexiones** (Fig. 3.22), quedando confinado, al menos temporalmente, en la capa de pasivación.

Aun cuando la energía de la banda prohibida del material de la capa de pasivación es suficientemente grande como para que no se produzca la absorción del fotón en el mismo, la presencia de impurezas e imperfecciones en esta región generan estados energéticos intermedios en la banda prohibida, haciendo posible la absorción del fotón. Es decir, existe una probabilidad (aunque baja) de que se produzca la **absorción del fotón en la capa de pasivación**. Si este fenómeno ocurre, el fotón no llegará a ser detectado por el sensor.

Una vez atravesada esta segunda interfaz, el fotón se encontrará bien con una capa de metal, bien con una capa de polisilicio, o bien con una capa de óxido grueso. En el caso de la capa de metal, el fotón será absorbido con una probabilidad muy alta, por lo que toda la información de la presencia del fotón en el sensor se pierde. Es decir, se produce la **absorción del fotón en la capa de metal**. Para el caso de que el fotón penetre en una capa de óxido grueso (utilizada para el aislamiento entre capas, por ejemplo, entre la capa de polisilicio y la de metal), al estar ésta compuesta de un material aislante, su comportamiento es muy similar al de la capa de pasivación. Existirán fenómenos de múltiples reflexiones en la capa de óxido grueso, así como una probabilidad de **absorción del fotón en la capa de óxido grueso**. Si el fotón penetra en una capa de polisilicio, las propiedades conductivas de este material hacen que se comporte de forma parecida a una capa de metal, por lo que la probabilidad de que se produzca la **absorción del fotón en la capa de polisilicio** es alta.

En las tecnologías CMOS actuales las difusiones n y p presentan una resistividad alta. Es por esta razón que sobre estas difusiones y sobre la capa de polisilicio se añade una capa de material siliciurado (compuesto de silicio al que se le ha añadido un compuesto metálico) para reducir la resistividad superficial de estas regiones. Ejemplos de capas saliciuradas en tecnologías de silicio son el saliciuro de tántalo (TaSi), el saliciuro de wolframio (WSi), o el saliciuro de titanio (TiSi). Existen dos tipos de procesos para añadir la capa saliciúrica. En uno de ellos sólo se añade esta capa a la puerta de los transistores. Con esto se consigue reducir la resistividad de la capa de polisilicio, pero no de las regiones de fuente y drenador. En el proceso saliciúrico autoalineado, la capa se añade sobre la puerta y las regiones de drenador y fuente. En algunos procesos esta capa saliciúrica sirve para interconectar una difusión y una capa de polisilicio adyacentes.

Debido a la naturaleza metálica de esta capa, resulta crítica su presencia para aplicaciones de sensado de luz. Las tecnologías actuales permiten añadir máscaras adicionales al proceso de fabricación para que sobre las zonas en las que se va a implantar la matriz de píxeles no se añada esta capa de saliciuro.

Si bien el esquema de la Fig. 3.21 representa una tecnología de una sola capa de polisilicio y una sola capa de metal, las tecnologías actuales suelen ser de dos capas de polisilicio y de varias capas de metal (de hasta 8 capas de metal en las tecnologías actuales), por lo que los fenómenos descritos anteriormente se dan con relativa facilidad.

23. Bien sea la interfaz capa de pasivación con la capa de metal, o bien con la capa de polisilicio, o bien con el óxido grueso.

En el comportamiento del sensor de imagen tiene especial importancia el fenómeno de múltiples reflexiones, al cual dedicamos el siguiente análisis.

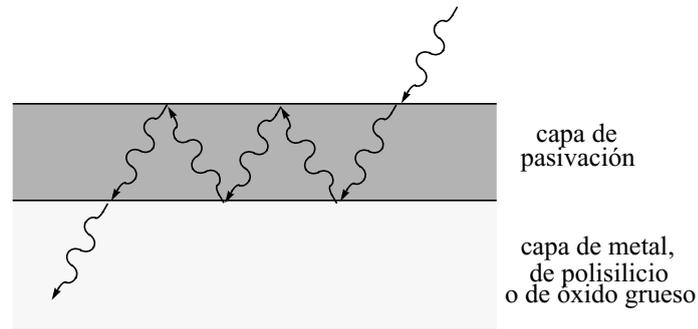


Fig. 3.22: Fenómeno de múltiples reflexiones en la capa de pasivación

3.5.4.- Fenómeno de múltiples reflexiones

Consideremos una onda de determinada frecuencia transmitiéndose a través de un medio discontinuo (Fig. 3.23). Al incidir en la interfaz entre dos medios distintos, parte de la onda se reflejará, mientras que parte de la onda se transmitirá al siguiente medio. Los porcentajes de onda reflejada y transmitida se denominan coeficientes de reflexión y transmisión respectivamente.

El medio por el que se transmite la onda presenta una impedancia característica (Z_i) a la onda que se está transmitiendo por él. Esta impedancia característica será función de la frecuencia de la onda transmitida, de las propiedades del material, y de las características geométricas del medio de transmisión. Supongamos que la onda incide sobre la interfaz entre dos materiales. Si las impedancias características de los dos medios coincide, la onda se transmitirá en su totalidad, es decir, el coeficiente de transmisión será la unidad, mientras que el coeficiente de reflexión tomará un valor cero. Si las impedancias características no coinciden, parte de la onda se transmitirá, mientras que parte de la onda se reflejará. La onda reflejada se encontrará en su viaje de vuelta con la interfaz con el anterior medio, viéndose sometida a los mismos procesos de transmisión/reflexión.

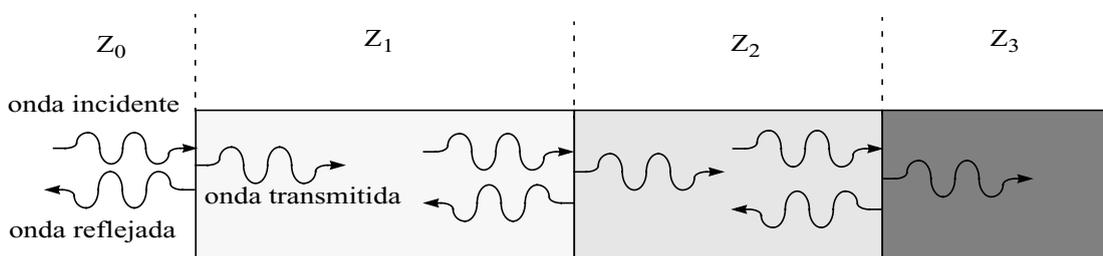


Fig. 3.23: Representación de un flujo de fotones a través de un medio discontinuo

Lo descrito anteriormente es lo que ocurre de forma transitoria con una onda que se transmite por el medio (Fig. 3.24). En resumen, la onda se ve sometida a un fenómeno de múltiples reflexiones en el medio. Si consideramos que se ha alcanzado el estado estacionario, en cada in-

terfaz podemos considerar que parte de la *onda incidente* se ha transmitido al siguiente medio (*onda transmitida*), mientras que parte se ha reflejado al anterior medio (*onda reflejada*).

Ahora bien, la magnitud de la onda transmitida será función del medio de transmisión y de la frecuencia de la onda, dado que el fenómeno de múltiples reflexiones al que se ve sometida la onda existirán fenómenos de interferencias. Fijadas las características del medio, y variando la frecuencia de la onda incidente, se observan máximos y mínimos en la amplitud de la onda transmitida, generadas a partir de interferencias constructivas y destructivas respectivamente en el seno del medio de transmisión.

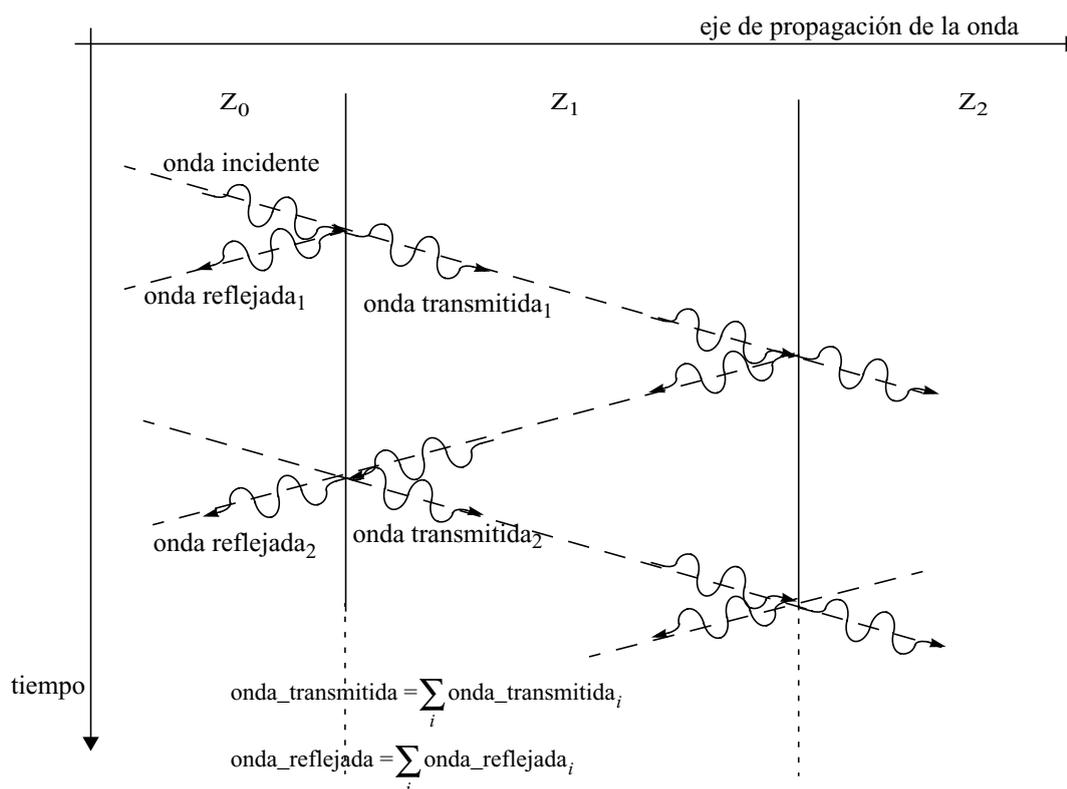


Fig. 3.24: Ondas incidente, transmitida y reflejada en la discontinuidad entre medios de transmisión

Debido a la dualidad onda-partícula del fotón, el anterior análisis es aplicable a un haz de fotones que incide sobre un medio de transmisión discontinuo, como es el caso del conjunto de capas depositadas sobre la oblea semiconductor (Fig. 3.21).

La consecuencia inmediata de todo el análisis anterior es que el espectro frecuencial de la eficiencia cuántica del sensor mostrará un comportamiento oscilatorio en torno al valor que tendría si no existiese esta discontinuidad de medios. En [12] se lleva a cabo la medición de la eficiencia cuántica para diferentes dispositivos fotosensores, y se pone de manifiesto el comportamiento antes descrito. La Fig. 3.25 compara la eficiencia cuántica para un conjunto de fotodiodos implementados en tecnología CMOS estándar de 0.8 μm . Los tres fotodiodos tienen una superficie de fotosensado de dimensiones 100x100 (μm^2). La unión pn de cada dispositivo es la siguiente. dws1: diodo de unión pozo/substrato; dwd1: diodo de unión difusión/pozo; dds1: diodo de unión difusión/substrato. La estructura de píxel utilizada se corresponde con la de un píxel activo en tiempo discreto por integración de carga²⁴.

24. Esta estructura se estudiará en el Cap. 4.

A partir de la Fig. 3.25 puede conocerse la frecuencia del fotón para la que se producen los máximos de la eficiencia cuántica a partir de la relación entre dos máximos consecutivos, resultado de fenómenos de interferencias constructivas ($\lambda_1 > \lambda_2$):

$$2l = m_1 \cdot \frac{\lambda_1}{n_{material}} \quad (3.21)$$

$$2l = m_2 \cdot \frac{\lambda_2}{n_{material}}$$

donde l es el ancho del medio de transmisión; m_1 y m_2 son enteros, λ_1 y λ_2 son las frecuencias a las que se encuentran los máximos. Si los máximos son consecutivos, entonces $m_2 = m_1 + 1$. Haciendo uso de esta relación y de las ecuaciones 3.21, m_1 puede expresarse como:

$$m_1 = \frac{\lambda_2}{\lambda_1 - \lambda_2} \quad (3.22)$$

El grosor de la capa interferente puede expresarse en función de la longitud de onda como:

$$l = \frac{m_1 \lambda_1}{2n_{material}} \quad (3.23)$$

Para hallar el grosor del medio para el que se produce un mínimo el desarrollo es exactamente el mismo, llegándose a la ecuación 3.23, donde en este caso λ_1 expresará la frecuencia para la que se da un mínimo del coeficiente de absorción.

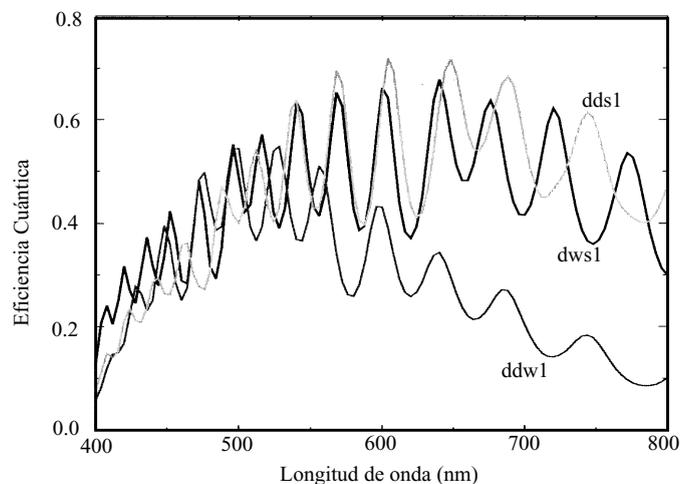


Fig. 3.25: Dependencia de la eficiencia cuántica con la longitud de onda para diferentes fotodiodos

3.5.5.- Conclusiones de esta sección

Como conclusiones de esta sección en el diseño de sensores de imagen podemos enumerar las siguientes.

En primer lugar, las capas depositadas con propiedades conductivas facilitan la absorción de fotones, por lo que impedirán que el fotón alcance la oblea semiconductor, y con ello, evitan que éste sea detectado electrónicamente. Puesto que la incidencia del haz de fotones será perpendicular a la superficie de fotosensado, habrá que evitar en la medida de lo posible la presencia de estas capas en la vertical de la superficie de fotosensado.

En segundo lugar, el fenómeno de múltiples interferencias puede reducirse añadiendo capas antirreflejantes. Sin embargo, si nos limitamos a tecnologías CMOS estándar, esta medida no es realizable. Los picos de la eficiencia cuántica serán función de los anchos de las capas depositadas, y éstas a su vez serán función de la tecnología utilizada. Tan sólo podremos afirmar que el uso de una u otra tecnología, puesto que variarán los picos de la eficiencia cuántica, podrá favorecer el sensado de ciertas frecuencias. De forma óptima, podríamos pensar en una tecnología para cual existiesen máximos de la eficiencia cuántica a las frecuencias del rojo, el verde y el azul. Desde el punto de vista práctico, el estudio genérico de sensores de imagen ha de hacerse lo más independiente posible de las características de la tecnología, por lo que pensar en la posibilidad de elegir una u otra tecnología por un factor tan variable como los picos de la eficiencia cuántica carece de sentido. Por otro lado, pensemos que el espectro de luz que va a ser sensado para cada color no se limita a una sola frecuencia, sino que se extenderá a un cierto ancho de banda en torno a las frecuencias centrales del rojo, el verde y el azul. Con esto se consigue paliar el efecto de la oscilación del coeficiente de absorción, pues se estará realizando un promedio del coeficiente de absorción en la banda considerada.

Referencias



- [1] D. Droste "Entwurf eines Wellenfrontsensor mit einem ASIC", Ph .D. Thesis. Institute of Computer Engineering, University of Mannheim, 1999.
- [2] T. Lulé, S. Benthien, H. Keller, F. Mütze, P. Rieve, K. Seibel, M. Sommer, M. Böhm "Sensitivity of CMOS Based Imagers and Scaling Perspectives", IEEE Transactions on Electronic Devices, vol. 47, no. 11, Nov. 2000
- [3] O. Yadid-Pecht, B. Mansoorian, E. R. Fossum, B. Pain "Optimization of noise in CMOS active pixel sensors for detection of ultra low light levels", Proc. SPIE, vol. 3019, pg. 125. 1997
- [4] A. J. Blanksby, M. J. Loinaz, D. A. Inglis, B. D. Ackland "Noise performance of a color CMOS photogate image sensor", en IEDM Tech. Dig., 1997, p. 205
- [5] H. S. Wong "Technology and device scaling considerations for CMOS imagers", IEEE Trans. Electron Devices, vol. 34, p. 2131, Dec. 1996
- [6] S. L. Barna, L. P. Ang, B. Mansoorian, E. R. Fossum " A low-light to sunlight, 60 frames/s, 80 kpixel CMOS APS camera-on-a-chip with 8b digital output", in Proc. IEEE Workshop on Charge-Coupled Devices and Advanced Image Sensors, 1999, p. 148
- [7] T. Delbrück & C. A. Mead "Analog VLSI phototransduction by continuous-time, adaptative, logarithmic photoreceptor circuits", California Institute of Technology Computation and Neural Systems Program, 1996
- [8] Markus Loose "A self-calibrating CMOS image sensor with logarithmic response", Ph. D. thesis, Rupertus Carola University of Heidelberg, Germany, 1999
- [9] I. Inoue et al. "Low dark current pinned photo-diode for CMOS image sensor", in Proc. SPIE Workshop on Charged-Coupled Devices and Advanced Image Sensors, 1999, p. 25.
- [10] T-H. Lee, R. M. Guidash, P. P. Lee "Partially pinned photodiode for solid state image sensors", U. S. Patent 5 903 021, 1999
- [11] Genís Chapinal "Imaging Arrays with In-Pixel Memory Implemented in CMOS Technology", University of Barcelona, Department of Electronics
- [12] Raúl Rueda Buza, "Estudio de Estructuras Fotosensoras y su Apareamiento en Tecnología CMOS", Proyecto Fin de Carrera, ETSI, Universidad de Sevilla

ADAPTACIÓN DE LA SEÑAL FOTOGENERADA

4.1.- FOTSENSORES EN TIEMPO DISCRETO Y EN TIEMPO CONTINUO

El diseño de la circuitería de adaptación de la señal fotogenerada puede realizarse siguiendo esencialmente dos procedimientos bien diferenciados: el diseño de la arquitectura en *tiempo discreto*; o bien el diseño en *tiempo continuo*.

En el diseño en tiempo discreto se establecen diferentes *fases de operación* del sensor, que se irán ejecutando secuencialmente. La secuencia de operaciones más común en el sensado de imagen es: reset del píxel (*fase de reset*), mediante el cual se lleva el píxel a un estado inicial de partida; adaptación de la carga fotogenerada (*fase de adaptación*), que generalmente consiste en la conversión de la corriente fotogenerada en tensión; y lectura del valor del píxel (*fase de lectura*) en la que el píxel vuelca el dato en la salida durante un cierto intervalo.

En contraste con los fotosensores en tiempo discreto, los fotosensores en tiempo continuo transforman la corriente fotogenerada en tensión o en intensidad mediante una relación dinámica de tiempo continuo. La señal de salida representa en todo momento al flujo de luz incidente¹.

En los sensores en tiempo discreto el tiempo de integración de la corriente fotogenerada es bastante mayor que el tiempo de lectura, por lo que la salida del píxel está definida sólo en un intervalo de tiempo corto. Esto representa un problema a la hora de plantear aplicaciones que necesiten un acceso aleatorio en el tiempo al valor del píxel, por ejemplo, aplicaciones que necesiten una continua monitorización de una imagen. Sin embargo, la definición de un estado inicial de partida permite implementar algoritmos eficaces para la reducción de la *componente de ruido de patrón fijo* o *FPN* (Fixed Pattern Noise)².

En los sensores en tiempo continuo es posible el acceso aleatorio en el tiempo al valor del píxel, pero no existe un estado inicial, por lo que la implementación de algoritmos para la reducción del FPN no es tan sencilla.

4.1.1.- Fotosensores en tiempo discreto

El proceso de adaptación de la carga fotogenerada más extendido es la integración de ésta sobre un elemento capacitivo. Estos sensores se conocen como *fotosensores por integración de carga*. El esquema básico de estos fotosensores se ilustra en la Fig. 4.1. Consiste en un *dispositivo fotosensor*; una capacidad sobre la que se va a integrar la carga fotogenerada (*capacidad del nodo de sensado* o *capacidad de integración*); y una *llave de reset*, utilizada para cargar la ca-

1. Esto es cierto bajo la consideración de que el ancho de banda del sensor es mucho mayor que la frecuencia de variación de la estimulación luminosa.
2. El ruido de patrón fijo (FPN) se considera la componente de ruido más importante en los sensores de imagen CMOS. Se estudia en el Cap. 5.

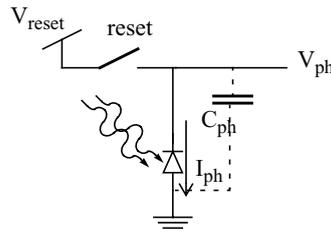


Fig. 4.1: Esquema de un fotosensor basado en la integración de la corriente fotogenerada

pacidad de integración al valor de reset. En una primera fase de operación (**fase de reset**) se conecta la llave de reset de forma que la tensión en el nodo de sentido V_{ph} queda fijada aproximadamente al valor de reset V_{reset} . Cuando la llave de reset se desconecta, la corriente fotogenerada comienza a descargar la capacidad del nodo de sentido (**fase de integración**). Esta fase durará un determinado tiempo (**tiempo de integración**), equivalente al tiempo de exposición en las cámaras fotográficas convencionales. En una tercera fase se procede a la lectura del valor de tensión almacenado en la capacidad de integración (**fase de lectura**). Esta secuencia de operaciones se conoce como **secuencia de reset-integración-lectura**.

La expresión temporal de la tensión $V_{ph}(t)$ durante la fase de integración viene dada por:

$$V_{ph}(t) = V_{reset} - \int_0^t \frac{1}{C_{ph}} I_{ph}(\tau) d\tau \quad (4.1)$$

donde C_p es la capacidad de integración. Si consideramos que tanto la intensidad fotogenerada como la capacidad de integración son constantes durante todo el periodo de integración la expresión de la tensión V_{ph} al final del periodo de integración vendrá dada por:

$$V_{ph}(t_{int}) = V_{reset} - \frac{1}{C_{ph}} \int_0^{t_{int}} I_{ph} \cdot d\tau = V_{reset} - \frac{t_{int}}{C_{ph}} \cdot I_{ph} \quad (4.2)$$

La tensión en el nodo de sentido se habrá reducido de forma proporcional a la corriente fotogenerada, es decir, proporcionalmente al flujo de fotones incidente, al tiempo de integración, e inversamente proporcional a la capacidad de integración.

Generalmente la capacidad de integración será la capacidad parásita C_p del propio dispositivo fotosensor. El valor de esta capacidad será relativamente alto, pues la superficie de sentido ha de ser grande para que la sensibilidad del sensor al flujo de fotones incidente sea alta. Para poder trabajar con tiempos de integración mayores sin que la capacidad de integración sature³, es decir, para poder ampliar el rango dinámico del sensor, la capacidad de integración puede aumentarse añadiendo dispositivos capacitivos en el nodo de sentido.

En el esquema del píxel de la Fig. 4.1 no existe una etapa de muestreo y retención del dato de salida. Se puede prescindir de esta etapa, puesto que el tiempo de lectura es muy corto en comparación con el tiempo de integración. La tensión de salida V_{out} cambiará durante la fase de lectura, pues el píxel seguirá integrando carga en el nodo de sentido. Sin embargo, la variación de tensión es suficientemente pequeña como para que sea conveniente prescindir de la etapa de muestreo y retención⁴. Hay que tener especial cuidado en el caso de que no exista etapa de

3. Si el flujo de fotones es suficientemente intenso la capacidad de integración se descargará por completo, limitando el rango dinámico de operación del sensor

muestreo y retención, pues algunos fotosensores podrían estar en la fase de integración mientras otros estén en la fase de lectura. Será necesario el uso de obturadores electrónicos⁵ que aseguren que el tiempo de integración de todos los píxeles es el mismo.

El método de integración de carga tiene varias ventajas. En primer lugar, considerando que la capacidad de integración es constante⁶, muestra una característica de transferencia lineal de conversión de la corriente fotogenerada en tensión, así como un rango de sensibilidad controlable a través del tiempo de integración. Además, el ruido de patrón fijo es bajo, puesto que la señal de salida depende de la capacidad en el nodo de sensado, y ésta tiene mejores características respecto al desapareamiento que los demás dispositivos⁷. Otra importante ventaja es el comportamiento paso de baja que muestran los dispositivos basados en algoritmos de integración. El ruido de la señal se reduce, puesto que las altas frecuencias se eliminan.

Sin embargo, aunque el rango de sensibilidad puede desplazarse fácilmente cambiando el tiempo de integración, el rango dinámico máximo dentro de una imagen es muy limitado⁸. Este inconveniente resulta del hecho de que el tiempo de integración es el mismo para todos los píxeles, y no puede ser ajustado de forma local. Para conseguir mayores rangos dinámicos existen trabajos en los que el tiempo de integración se controla para cada píxel de forma individual [2]. Otro de los inconvenientes de los sensores por integración de carga es la ya mencionada limitación en cuanto al acceso aleatorio en el tiempo a los píxeles: dos lecturas consecutivas de un píxel han de estar separadas un tiempo mayor al tiempo de integración.

4.1.2.- Fotosensores en tiempo continuo

El esquema de un fotosensor en tiempo continuo se muestra en la Fig. 4.2. Consiste básicamente en un fotosensor que actúa como fuente de intensidad controlada por el flujo de intensidad luminoso incidente sobre el píxel; y una carga resistiva para llevar a cabo la conversión de la corriente fotogenerada en tensión.

La carga que utilizemos va a definir la conversión entre corriente fotogenerada y tensión en el nodo V_{ph} . Por ejemplo, podemos hacer uso de una carga resistiva lineal, en cuyo caso la conversión de la corriente fotogenerada en tensión es lineal. Sin embargo, el diseño de sensores de imagen que ofrezcan un rango dinámico amplio, comparable al que alcanzan los sistemas de visión de los animales, motiva el desarrollo de los sensores de respuesta logarítmica. La implementación de sensores de imagen con respuesta logarítmica en tecnología CMOS es particularmente fácil de conseguir, como se muestra en el esquema de la Fig. 4.2, aprovechando la relación logarítmica entre la intensidad (I_{ds}) y la tensión puerta-fuente (V_{gs}) en MOSFET polarizado en inversión débil. A este tipo de sensores se les conoce como *sensores logarítmicos*⁹.

-
4. Eliminar esta etapa mejora el factor de relleno del sensor y elimina el ruido que esta etapa pueda introducir
 5. Estos obturadores consistirán en llaves que controlarán el tiempo durante el cual se permite el traspase de la carga fotogenerada
 6. Es decir, que no depende de la tensión entre sus terminales.
 7. Esto es consecuencia de que la capacidad vendrá determinada por la superficie de fotosensado, y ésta será elevada en comparación con los demás dispositivos, lo que mejora las características de desapareamiento entre píxeles.
 8. Un valor típico del rango dinámico de estos sensores está en torno a tres décadas [1]
 9. Mediante el uso del transistor en inversión débil se obtiene una relación logarítmica entre tensión e intensidad en tiempo continuo. Ahora bien, no es ésta la única forma de obtener una respuesta logarítmica. En posteriores capítulos veremos algoritmos para implementar una respuesta logarítmica en sensores en tiempo discreto

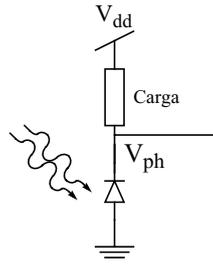


Fig. 4.2: (a) esquema básico de un píxel en tiempo continuo

La arquitectura básica del píxel logarítmico se muestra en la Fig. 4.3, tanto para una estructura nmos como para una pmos. Dicha arquitectura es particularmente sencilla: consta de un fotodiodo y de un transistor MOS como elemento de carga (M_1). En ambos píxeles la respuesta del píxel a la luz es logarítmica puesto que el transistor de carga M_1 está funcionando en la región de inversión débil. En estas condiciones la corriente de drenador del transistor M_1 viene dada por:

$$I_{ds} = \frac{W}{L} I_{do} \exp\left(\frac{V_{gs} - V_x}{n\phi_T}\right) (1 + \lambda V_{ds}) \quad (4.3)$$

donde $\phi_T = kT/q$; I_{ds} y V_{ds} son, respectivamente, la corriente y la tensión drenador-fuente; V_{gs} es la tensión puerta-fuente; λ es el coeficiente de modulación de la longitud de canal; e I_{do} y n vienen dados por las expresiones:

$$I_{do} = \mu C'_{ox} \phi_T^2 \frac{\gamma}{2\sqrt{1\phi_F + V_{sb}}} e^{-0\phi_F/\phi_T} \quad (4.4)$$

$$n = 1 + \frac{\gamma}{2\sqrt{1\phi_F + V_{sb}}} \quad (4.5)$$

$$V_x = V_{FB} + 1\phi_F + \gamma\sqrt{1\phi_F + V_{sb}} \quad (4.6)$$

En estas expresiones V_{sb} es la tensión fuente-sustrato del transistor; μ la movilidad de los portadores de carga; ϕ_F el potencial de Fermi; V_{FB} la tensión en la banda plana; y γ el coeficiente de efecto sustrato¹⁰.

La tensión V_{gs} aparece además en la tensión drenador-fuente del transistor M_1 puesto que tenemos que: $V_{gs} = V_{ds}$. Por lo tanto, la corriente I_{ds} depende sólo de V_{gs} . Esta tensión V_{gs} aparece en el nodo del fotodiodo V_{ph} puesto que $V_{ph} = V_{dd} - V_{gs}$. Substituyendo V_{gs} por $(V_{dd} - V_{ph})$ en la ecuación 4.3 obtenemos:

$$I_{ds} = \frac{W}{L} I_{do} \exp\left(\frac{V_{dd} - V_{ph}}{n\phi_T}\right) (1 + \lambda(V_{dd} - V_{ph})) \quad (4.7)$$

Dado que $\lambda \cdot V_{ds} \ll 1$, una expresión apropiada para la tensión en el nodo del fotodiodo V_{ph} puede deducirse de la ecuación 4.7:

10. Para una descripción detallada del comportamiento del transistor en inversión débil recomendamos [3]

$$V_{ph} = V_{dd} - n\phi_T \ln\left(\frac{L I_{ds}}{W I_{do}}\right) \quad (4.8)$$

De 4.8 está claro que la tensión V_{ph} en el nodo del fotodiodo es proporcional al logaritmo de la corriente del fotodiodo I_{ph} , pues coincidirá con I_{ds} .

El uso del transistor en inversión débil permite la compresión logarítmica y, por lo tanto, conseguir un rango dinámico amplio en el sensor. Sin embargo, el uso del transistor en esta región tiene serios inconvenientes derivados del desapareamiento de los transistores, especialmente de la tensión umbral V_T . La corriente I_{ds} depende exponencialmente del factor kT/q , por lo que la salida del píxel es sensible a cambios en la temperatura, pero la principal dificultad surge del poco rango de variación de tensión de salida, en torno a 0.15V para una variación de 5 órdenes de magnitud de la intensidad luminosa [4]. El desapareamiento de los transistores provoca variaciones en la tensión umbral que a su vez generarán variaciones de la tensión de salida de en torno al 5~10%, es decir, comparable a los niveles de señal medidos.

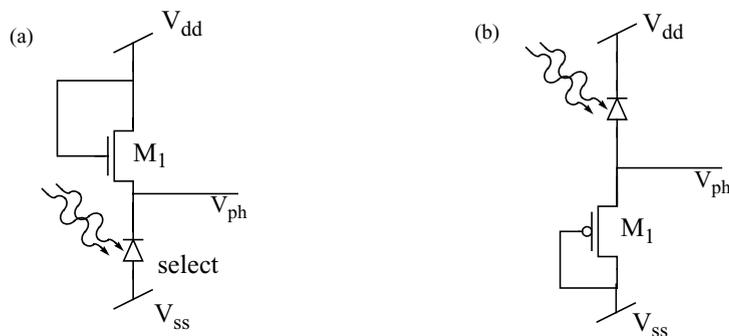


Fig. 4.3: Esquema básico de un píxel logarítmico (a) tipo nmos y (b) tipo pmos

Además del amplio rango dinámico, estos píxeles tienen una propiedad que los hace extremadamente interesantes para aplicaciones de procesamiento de imagen: un objeto con un contraste local dado que está siendo sentido con un sensor logarítmico genera una imagen con diferencias locales entre píxeles que son independientes del nivel de iluminación de la escena. Esta propiedad es fácilmente demostrable mediante la observación de que un cociente entre niveles de intensidad en un píxel I_1/I_2 genera un cambio de señal dado por $\log(I_1) - \log(I_2)$, mientras que un cambio proporcional en el nivel de intensidad $c \cdot I$ genera un cambio de señal dado por $\log(c) + \log(I)$. El mismo objeto bajo condiciones de iluminación mayor generará una salida logarítmica igual, excepto en un desplazamiento aditivo del nivel de intensidad de fondo.

En contraste con las arquitecturas de píxel en los que la carga se integra, el píxel logarítmico mide la tensión en la fuente del MOSFET en serie con el fotodiodo (Fig. 4.3). Por esta razón el comportamiento dinámico de los píxeles logarítmicos depende de la corriente fotogenerada: cuanto menor es el nivel de luz en una escena (cuanto menor es la corriente fotogenerada) más tiempo tarda el MOSFET en alcanzar el equilibrio. Por lo tanto, los píxeles logarítmicos reaccionan mucho más despacio bajo condiciones de iluminación débil.

En los píxeles en tiempo discreto se definen unas fases de operación (reset-integración-lectura), de forma que durante la fase de reset se puede hacer una lectura de la carga residual en los dispositivos de memoria del píxel para, en la fase de lectura corregir esta componente de la salida, lo que permite eliminar las componentes de ruido de patrón fijo. En los píxeles logarítmicos no

existen tales fases, sino que la salida del píxel está disponible en todo momento. Esto se presenta como una ventaja si lo comparamos con los píxeles por integración de carga, en los que la salida sólo podrá medirse en ciertos intervalos de tiempo. Pero por otra parte, el que no exista una fase en la que no se mida el valor de la carga residual no nos va a permitir implementar de forma sencilla algoritmos de corrección del ruido de patrón fijo. Será necesario el uso de una circuitería adicional que nos proporcione una referencia (en general será una referencia de intensidad) para permitir el correcto calibrado del dispositivo y así minimizar la componente de ruido de patrón fijo. De todas formas trataremos más en profundidad este aspecto, pues es fundamental en el planteamiento del diseño de una u otra arquitectura.

Otro de los inconvenientes de los sensores logarítmicos es que la sensibilidad al contraste es pequeña debido a la compresión logarítmica: aumenta el rango dinámico, pero reduce el contraste entre píxeles.

4.1.3.- Fotosensores basados en el modo de operación fotovoltaico

El esquema de un píxel basado en un sensor en modo fotovoltaico se muestra en la Fig. 4.4(a). El análisis de este tipo de estructuras puede realizarse como caso particular de los fotosensores en tiempo continuo en el que la carga se considera infinita.

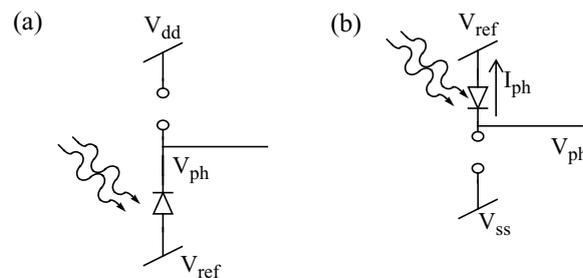


Fig. 4.4: (a) esquema básico de un píxel basado en un fotosensor en modo fotovoltaico; y (b) circuito equivalente del fotosensor

Tomando como referencia la Fig. 4.4(b), para que el fotosensor funcione en modo fotovoltaico el terminal p se coloca a una tensión de referencia V_{ref} mientras que el terminal n se deja flotante. La corriente fotogenerada I_{ph} extrae corriente del nodo V_{ph} , por lo que la tensión en el mismo disminuirá hasta que se alcance el equilibrio en el fotosensor con el diodo polarizado en directa. Este equilibrio consistirá en la igualación de las corrientes fotogenerada y directa del fotodiodo.

El análisis del fotodiodo se ha llevado a cabo en la Secc. 2.11. La ecuación que ilustra el comportamiento del fotodiodo es:

$$I = I_{ph} + I_o(1 - e^{V/\phi_T}) \quad (4.9)$$

donde I_{ph} es la corriente fotogenerada y el segundo término es la corriente inversa del diodo.

Considerando el terminal n flotante, la condición de equilibrio antes propuesta se traduce en:

$$I_{ph} = -I_o(1 - e^{V/\phi_T}) = I_o(e^{V/\phi_T} - 1) \quad (4.10)$$

La tensión en circuito abierto (resultado de colocar el nodo n flotante) puede expresarse a partir de 4.10 como:

$$V_{ph} = \phi_T \cdot \ln\left(\frac{I_{ph}}{I_o} + 1\right) \quad (4.11)$$

lo que proporciona la relación de conversión de la señal óptica (corriente fotogenerada por el flujo luminoso incidente I_{ph}) en eléctrica (V_{ph}). Nótese que la relación también es logarítmica, sin que intervenga en el esquema ningún transistor MOS como carga.

4.2.- PÍXELES PASIVOS Y ACTIVOS

Conceptualmente, la diferencia entre un *píxel activo* y un *píxel pasivo* es la implementación o no de un buffer como amplificador en la circuitería a nivel de píxel en el sensor.

El píxel pasivo no incluye amplificador a nivel de píxel. El factor de relleno, por lo tanto, estará maximizado, pues sólo será necesaria una llave para el direccionamiento de la lectura del píxel. Toda la superficie del píxel queda destinada al dispositivo fotosensor que se vaya a implementar.

El píxel activo incluye un buffer como amplificador a nivel de píxel, por lo que el factor de relleno del sensor disminuye. Sin embargo, la introducción de este elemento a nivel de píxel tiene una serie de ventajas que se discutirán en los siguientes apartados.

4.2.1.- Sensor basado en píxeles pasivos (PPS)

En una matriz de fotodiodos (PDs) el proceso de adquisición de la imagen se describe a continuación. Se asume que todos los PDs están inicialmente precargados a una cierta tensión inversa de polarización y que todas las llaves están en posición cerrada al paso de corriente. La luz incidente genera pares de carga electrón-hueco en cada píxel, lo que provoca el flujo de una corriente fotogenerada debido a la separación de los pares de carga gracias al campo eléctrico aplicado en los PDs. La capacidad del PD, cuyo valor inicial estará determinado por la tensión de polarización inicial, se descargará gracias a la corriente fotogenerada. Tras un determinado tiempo (*tiempo de exposición*), se podrá conocer el valor almacenado en la capacidad de un determinado píxel mediante el correcto direccionamiento de los interruptores de la matriz de píxeles, lo que provocará el volcado de la carga almacenada en el píxel direccionado en el correspondiente *bus de columna*. El amplificador de salida reseteará el píxel a su valor inicial al mismo tiempo que mide cuánta carga es necesaria aplicar para completar el reset. Esta carga es (sin considerar los efectos de segundo orden) la misma que la carga fotogenerada acumulada en el píxel. Esto significa que una matriz de PDs trabaja mediante el equilibrado de carga haciendo uso de líneas conductoras (generalmente metalizaciones) de gran longitud. Como se verá más adelante, este proceso de equilibrado de carga introduce un error en el proceso de detección de la señal, que será proporcional a la capacidad total del nodo del bus de columna. Esta era la principal desventaja de los PDs frente a los CCDs a principio de la década de los 80, y fue una de las principales razones por las que se descartó el uso de los PDs para el sensado de imagen a favor de los sensores CCDs.

La anterior descripción responde a la definición y funcionamiento de una matriz de píxeles pasivos. El esquema de un píxel pasivo se muestra en la Fig. 4.5.

El principal problema con los píxeles pasivos son los niveles de ruido y la escalabilidad. El ruido de lectura es típicamente del orden de 250 electrones r.m.s., comparado con los menos de 20 que se consigue en los sensores CCDs comerciales [4]. El píxel pasivo, además, no se escala bien para matrices de sensado grandes y/o para frecuencias de lectura altas. Esto se debe a que el incremento en la capacidad del bus y en la frecuencia de lectura provocan un aumento del error.

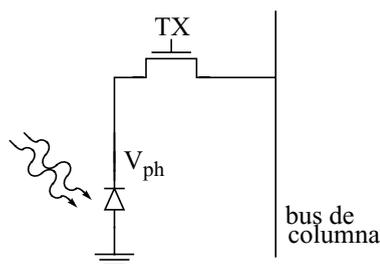


Fig. 4.5: Esquema de un píxel pasivo

La forma más sencilla de leer la carga almacenada en una capacidad consiste en medir la tensión a través de una carga resistiva, como se muestra en el esquema de la Fig. 4.6(a).

Siguiendo la secuencia de operación reset-integración-lectura, en primer lugar la llave S que se usa para direccionar el píxel se cierra y el valor del píxel se resetea a un valor de tensión inversa V_{reset} . Durante un cierto periodo de tiempo (tiempo de integración t_{int}) se abre la llave S , permitiendo que la capacidad del fotodiodo se descargue de forma proporcional al flujo de luz incidente. Transcurrido t_{int} se cierra la llave S , de forma que toda la carga necesaria para llevar la capacidad del fotodiodo a su valor de reset fluirá a través de R . La corriente que atraviesa R generará temporalmente una tensión proporcional al flujo de luz incidente en el sensor durante el periodo de integración.

En la Fig. 4.6(b) se extiende el concepto de píxel pasivo a una columna de píxeles. De esta forma se consigue reducir la circuitería a nivel de píxel, lo que mejora considerablemente el factor de relleno y se eliminan los errores que pueda introducir la fluctuación de la tensión de reset entre píxeles de una misma columna, al tener todos una misma fuente de tensión como referencia.

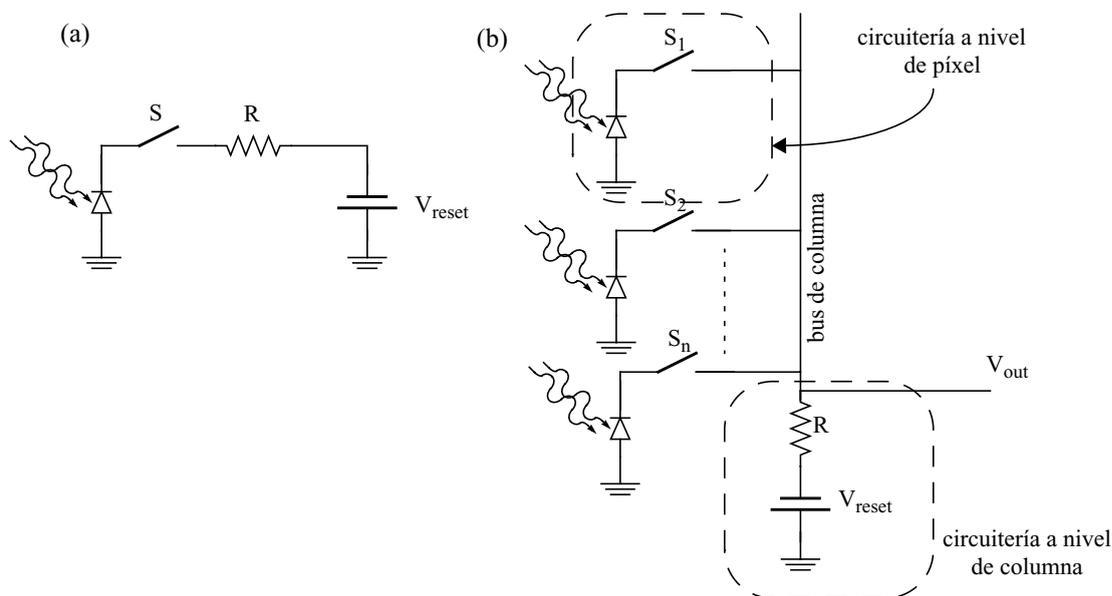


Fig. 4.6: (a) Esquema de funcionamiento de un píxel pasivo; e (b) implementación de una columna de píxeles pasivos

El principal inconveniente del esquema de la Fig. 4.6 es que el tiempo necesario para el completo reset del píxel es muy alto, por lo que el reset no se completa, lo que reduce el rango dinámico del sensor¹¹.

Como solución a este problema se propone realizar la fase de lectura a través de un amplificador de carga, como se muestra en la Fig. 4.7. Si el amplificador es ideal, la tierra virtual del mismo es la encargada de suministrar la corriente necesaria para resetear el fotosensor. El valor de tensión a la salida vendrá dado por la expresión:

$$V_{out} = -\frac{C_d}{C_f} \cdot V_{ph} \quad (4.12)$$

donde C_d es la capacidad de integración, que será aproximadamente la capacidad del dispositivo fotosensor.

El reset del fotosensor se llevará a cabo cortocircuitando la capacidad C_f

Sin embargo este esquema no es práctico, pues aparece una capacidad parásita¹² C_L en paralelo con los fotosensores que provoca que parte de C_d cargue, no sólo C_f , sino también la capacidad parásita C_L , por lo que C_f no memorizará toda la carga necesaria para el reset del fotosensor. Este efecto puede reducir significativamente el valor de la tensión de salida¹³.

Tal y como se ha comentado, la principal ventaja del píxel pasivo es que el factor de relleno se maximiza, pues sólo es necesario una llave en cada píxel. Además, el rutado de este tipo de dispositivos es muy sencillo, por lo que el número de capas depositadas sobre la superficie fotosensible es mínimo, y con ello, la eficiencia cuántica de estos dispositivos es alta en comparación con otros fotosensores basados en fotodiodos.

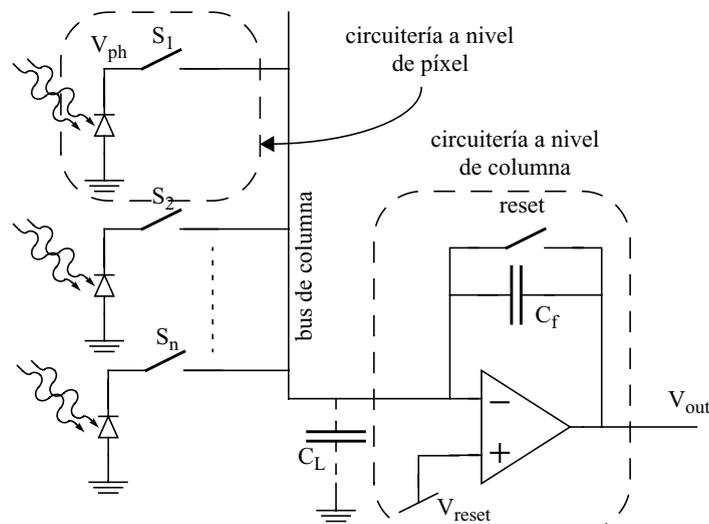


Fig. 4.7: Esquema de funcionamiento de un píxel pasivo con amplificador de carga

11. El rango dinámico del sensor está íntimamente relacionado con la variación de carga en la capacidad de integración. La existencia de una componente residual de carga en esta capacidad reducirá por lo tanto el rango dinámico del sensor
12. La componente principal de esta capacidad parásita es debida al bus de columna
13. Para un proceso de $0.5 \mu\text{m}$ caracterizado por una capacidad entre metal-1 y sustrato de $50 \text{ aF}/\mu\text{m}^2$, y considerando un bus de columna de dimensiones $3 \mu\text{m} \times 10 \text{ mm}$; un tamaño de píxel de $(30 \mu\text{m})^2$ con una capacidad equivalente de 0.15 pF ; y un valor típico de 0.2 pF para C_f , sólo el 90% de la carga necesaria para resetear el fotosensor la suministra la capacidad de integración, lo que implica una pérdida considerable de sensibilidad [4]

4.2.2.- Sensor basado en píxeles activos (APS)

Los sensores basados en píxeles activos (APS) se caracterizan por la inserción de un amplificador activo en cada píxel. El esquema más sencillo de píxel activo se muestra en la Fig. 4.8, en la que se muestran de forma conceptual tanto un APS en tiempo discreto como un APS en tiempo continuo.

En los sensores CMOS APS se reduce el factor de relleno para mejorar las características de los mismos. Estos píxeles se fabrican normalmente con un factor de relleno del 20-30%, similar a algunas implementaciones CCDs (como es el caso de los ILT (interline-trasfer) CCDs [4]).

Puesto que cada amplificador se activa sólo durante la fase de lectura, la disipación de potencia es mínima y es en general menor que en los sensores CCDs.

La tecnología APS tiene ventajas potenciales sobre los CCDs, pero presenta el inconveniente de no eliminar completamente la componente de FPN.

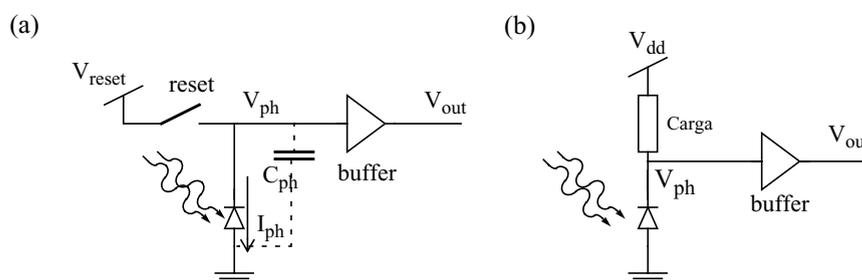


Fig. 4.8: (a) APS en tiempo discreto; y (b) APS en tiempo continuo

La salida de los APS podrá ser bien en tensión, bien en intensidad, como se muestra en la Fig. 4.9.

Con la introducción del buffer de tensión a nivel de píxel según el esquema de la Fig. 4.9(a) se pretende que la transmisión de señal a través del bus de columna sea en tensión. El esquema del **sensor APS con salida en tensión** consiste en la implementación de interruptores en cada píxel que conectan las tensiones locales a una estructura amplificadora común (amplificador de columna). En esta arquitectura la transmisión de la señal en tensión es mucho más sencilla de lograr que la transmisión de carga que se lleva a cabo en los PPS. La componente de ruido térmico en el bus de columna, filtrada a través del filtro RC paso de baja que supone la misma línea, genera un ruido en tensión que es inversamente proporcional a la raíz cuadrada de la capacidad del bus de columna, por lo que a mayor capacidad del bus, menor ruido en tensión. Por esta razón, las señales en tensión pueden transmitirse con una componente de ruido mucho menor y permiten medidas de mayor precisión que la transmisión de señales de carga. Esto implica que los sensores de imagen implementados con arquitecturas de transmisión de señal en tensión tienen ventajas inherentes sobre las arquitecturas basadas en la transmisión de señal en intensidad respecto al ruido.

Con los **sensores APS con salida en intensidad** (Fig. 4.9(b)) resulta muy sencillo implementar sensores de resolución variable, capaces de realizar el promedio de un subconjunto de píxeles vecinos. Estos esquemas permiten llevar a cabo funciones tales como el zoom real de la imagen a nivel electrónico, la reducción instantánea de los datos, la mejora de la sensibilidad a la luz en

aplicaciones robóticas, implementación de algoritmos de enfoque de objetos y de visión inspirados en modelos biológicos, etc [5].

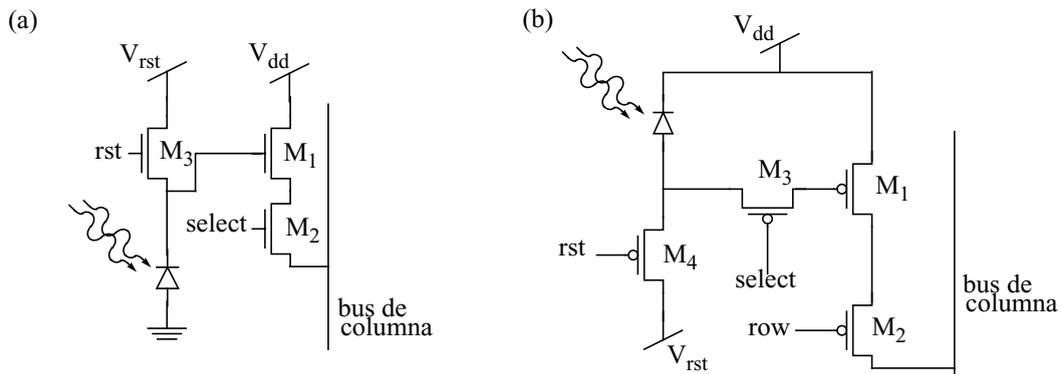


Fig. 4.9: Esquema de un píxel activo (a) con salida en tensión; y (b) con salida en intensidad

La Fig. 4.9(b) muestra el esquema de un sensor APS con salida en intensidad. Un esquema de lectura basado en corrientes en lugar de en tensiones tiene ventajas inherentes sobre los sistemas de lectura en tensión, tales como velocidad de lectura, pues en éstos la velocidad queda limitada por los rangos de tensión y la capacidad de carga/descarga.

Un inconveniente de los APS en modo corriente es que el FPN generado por el offset y por la ganancia no se consiguen reducir al mismo nivel que en los APS en modo tensión. Se consigue una compensación del offset suficiente, pero la componente de FPN debido a las diferencias en la ganancia es muy alta. Esto se debe principalmente a la fuerte dependencia de la señal de salida con los parámetros del transistor amplificador de transconductancia, al contrario que en el modo normal de funcionamiento en tensión, donde la circuitería activa en el píxel se usa como un buffer de ganancia unidad con una débil dependencia de los parámetros del mismo. Este problema se hace especialmente importante con el crecimiento del tamaño de la matriz de píxeles, lo que lleva a mayores diferencias en las corrientes de salida.

El uso de los sensores APS con salida en tensión está mucho más extendido que los sensores con salida en intensidad. Esto está motivado porque, si bien, como se ha comentado, la funcionalidad de la salida en intensidad en lo que a aplicaciones de imagen se refiere es mayor, las características en cuanto al FPN (principal problema de diseño de los sensores de imagen) son bastante peores. Nos centramos, por lo tanto, en los sensores APS con salida en tensión. Pasemos a describir las configuraciones básicas de sensores APS con salida en tensión en tiempo discreto y en tiempo continuo.

4.2.2.1.- APS con salida en tensión en tiempo discreto

El esquema más simple de implementación de un APS en tiempo discreto se muestra en la Fig. 4.10. Consta de un elemento fotodetector y de tres transistores MOS: un transistor de reset (M_1), que hace de llave de reset; un transistor en configuración de seguidor por fuente, que actúa como buffer de tensión (M_2); y un transistor de selección, utilizado para direccionar el píxel que va a ser leído (M_3).

Como figuras típicas de este tipo de sensores [4], el factor de relleno está en torno al 20 ~ 35%; el área de cada píxel está en torno a 15 veces la longitud mínima de la tecnología¹⁴; la eficiencia

14. Esta medida hace referencia a la longitud del píxel, suponiendo que ocupe una superficie cuadrada

cuántica alcanza un máximo de $\sim 40\%$ para las longitudes de onda en torno al verde; el factor de conversión de carga en tensión es aproximadamente de $3 \mu\text{V}/e^-$; y el rango dinámico está en torno a los 75 dB. Esta última característica es uno de los principales inconvenientes que motivan el desarrollo de arquitecturas que mejoren el rango dinámico.

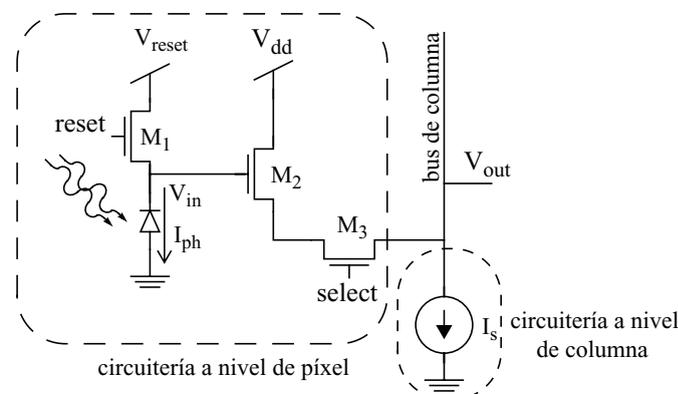


Fig. 4.10: Esquema de APS en tiempo discreto

4.2.2.2.- APS con salida en tensión en tiempo continuo

La Fig. 4.11(a) ilustra la implementación más simple de un APS en tiempo continuo, mientras que en Fig. 4.11(b) se ilustra el caso particular de un APS en tiempo continuo logarítmico implementado con transistores nmos.

Como principales características de los APS en tiempo continuo logarítmicos, el rango dinámico típicamente sobrepasa los 100 dB. El principal inconveniente es la dependencia de la ganancia del píxel con la temperatura y variación de la tensión umbral del transistor de carga M_1 .

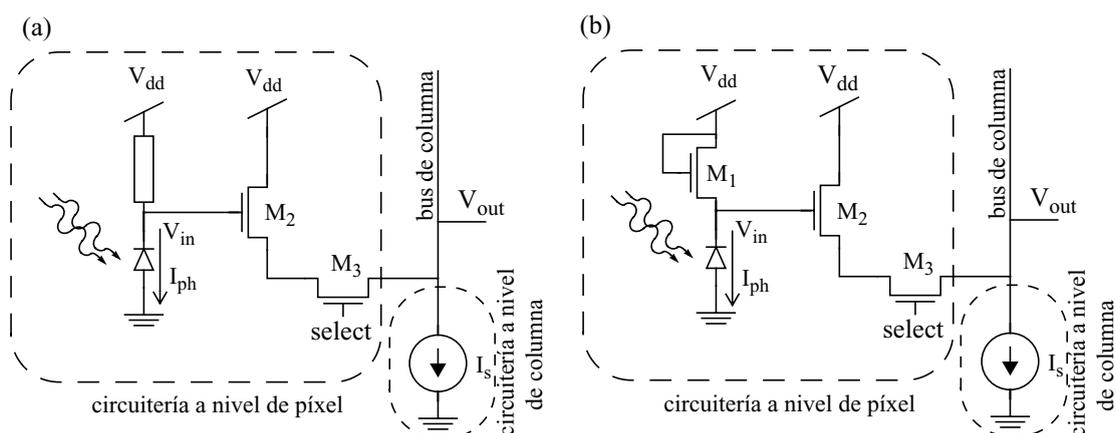


Fig. 4.11: (a) Esquema de APS en tiempo continuo; y (b) APS en tiempo continuo logarítmico

Puede observarse que la configuración de la Fig. 4.10 y de la Fig. 4.11(b) difieren tan sólo en la conexión drenador-puerta. Esto nos permite pensar en llevar a cabo la implementación de un sensor que pueda operar en ambos modos de funcionamiento. Esta propuesta resulta muy interesante, pues permite, desde el punto de vista del análisis del diseño, comparar de forma directa sobre un mismo integrado las características de uno y otro esquema; y desde el punto de vista

de la funcionalidad, poder operar en uno u otro modo de funcionamiento con sólo añadir una circuitería muy simple.

4.3.- CIRCUITOS DE ADAPTACIÓN DE LA CARGA FOTOGENERADA

En los anteriores apartados se ha llevado a cabo un análisis fundamentalista de las estructuras a nivel de píxel, tratándolas desde el punto de vista ideal. Ocurre, sin embargo, que la existencia de no idealidades desvían considerablemente el sensor de su comportamiento ideal, por lo que resulta necesario el diseño de arquitecturas que corrijan en la medida de lo posible estas no idealidades. Otro factor a tener en cuenta es la necesidad de mejorar las características del sensor, tales como el rango dinámico, la relación señal a ruido o el factor de conversión de señal luminosa en eléctrica.

Los circuitos de adaptación de la corriente fotogenerada añadirán una mayor funcionalidad a nivel de píxel que la simple lectura en tensión o en intensidad mostrada en la anterior sección. En el caso de los fotosensores en tiempo discreto las variaciones sobre la arquitectura consistirán básicamente en la incorporación de elementos que lleven a cabo fases adicionales en el proceso de reset-adaptación-lectura.

Presentamos a continuación una serie de arquitecturas a nivel de píxel que implementan algunas de las funciones antes mencionadas. La circuitería a nivel de píxel ha de ser muy sencilla, pues una arquitectura compleja a nivel de píxel, o bien degradaría la MTF pues cada píxel ocuparía una superficie relativamente grande, o bien mateniendo el área destinada a cada píxel reduciría notablemente el factor de relleno del sensor, y con ello la eficiencia del mismo. Sin embargo, por diferentes factores que se discutirán más adelante, resulta apropiado realizar la corrección del ruido de patrón fijo y la conversión analógico-digital a nivel de píxel. Este compromiso de diseño respecto a la superficie disponible para ampliar la funcionalidad a nivel de píxel supone el principal reto para el diseñador, pues verá muy limitado el número de dispositivos que podrá implementar para mejorar las características de la salida del píxel.

La implementación de todas funciones a nivel de píxel resulta en la práctica inviable, por lo que se opta por implementar parte de la funcionalidad a nivel de píxel (*circuitería a nivel de píxel*), mientras que otra parte se incorpora a nivel de columna (*circuitería a nivel de comuna*), con lo que en la matriz de píxeles finales esta circuitería será compartida por toda una columna de píxeles. Esto mejora notablemente el factor de relleno del sensor, pero no permite la corrección de determinadas no idealidades¹⁵.

4.3.1.- APS por integración de carga con salida diferencial directa

La Fig. 4.12 muestra el esquema de un APS con salida diferencial directa. Cada píxel consta de de un elemento fotosensor, de una tensión de referencia de reset, y de dos etapas de muestreo y retención. Esta configuración permite que las capacidades de muestreo C_1 y C_2 memoricen dos imágenes de forma consecutiva, permitiendo dos modos de operación del fotosensor.

El primer modo de operación consiste en llevar a cabo una fase en la cual el sensor no integra la luz incidente. Una de las capacidades memorizará la señal que el sensor genera sin considerar la carga fotogenerada, y que se considerará como ruido introducido por el sensor; y a continuación una fase de integración, en la que sobre la segunda de las capacidades se almacena la carga fotogenerada por el flujo de luz incidente.

15. Por ejemplo, se podrá realizar la corrección del FPN de la circuitería a nivel de píxel, pero no se realizará la corrección del FPN que la propia circuitería de columna introduce.

El segundo de los modos de operación consiste en hacer lecturas sucesivas de imágenes diferentes. Las capacidades C_1 y C_2 irán muestreando la carga fotogenerada correspondiente a imágenes sucesivas. Una vez almacenadas las dos imágenes podrá o no realizarse una lectura diferencial de las mismas o simplemente volcar las dos salidas sobre un mismo bus de forma secuencial. Si se realiza una lectura diferencial la salida del sensor mostrará sólo los cambios de intensidad luminosa, por lo que podría usarse esta configuración para aplicaciones de detección de movimiento. Si no se realiza una lectura diferencial, sino que simplemente se vuelcan los datos sucesivamente sobre un bus de salida (que generalmente será el bus de columna), la frecuencia de datos de salida del sensor se doblará¹⁶.

La implementación de una salida diferencial directa a nivel de píxel permite, con la máxima granularidad, sustraer de la señal luminosa sensada tanto el ruido estático que el propio sensor genera (*FPN a nivel de píxel*) como la carga residual que haya quedado de anteriores muestras en la capacidad de integración. Además, el acoplamiento (crosstalk) eléctrico y las fluctuaciones de las fuentes de alimentación serán componentes en modo común de las dos señales, por lo que la lectura diferencial los eliminará. Es por esto que resulta de especial interés la implementación de este tipo de estructuras, si bien resulta bastante costoso en área¹⁷.

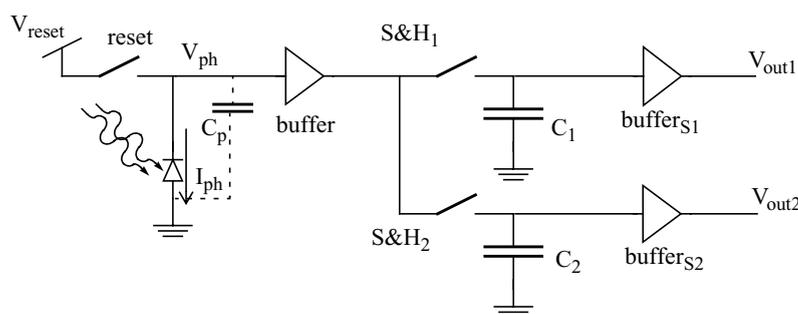


Fig. 4.12: Sensor con salida diferencial directa implementada a nivel de píxel

La implementación de la circuitería de muestreo y retención a nivel de píxel puede reducir drásticamente el factor de relleno del sensor (sobre todo en tecnologías por encima de las $0.35\ \mu\text{m}$). Una solución alternativa en estos casos es la implementación de la lectura diferencial a nivel de columna (Fig. 4.13). Con este esquema se consigue implementar las celdas de memoria¹⁸ a nivel de columna en lugar de a nivel de píxel, por lo que el factor de relleno mejora considerablemente.

La secuencia de funcionamiento es la que sigue. Tras la fase de integración la tensión de salida del píxel se muestrea y memoriza en C_S mediante la activación de *SHS*. A continuación se resetea el píxel (activación de la señal de reset) y el valor de tensión de reset en el píxel se memoriza en C_R mediante la activación de *SHR*. La diferencia de tensión entre las dos capacidades se corresponderá con el valor de la lectura diferencial. Esta secuencia de funcionamiento¹⁹ supone un problema: la medida del valor de tensión de reset es posterior a la lectura del valor de tensión.

16. El volcado sucesivo de los datos no tiene especial interés, pues las frecuencias de operación de los sensores suelen resultar bastante bajas, no suponiendo un inconveniente en el diseño.
17. En particular, por implementación de las capacidades C_1 y C_2 .
18. Las celdas de memoria serán las capacidades MOS que ocuparán bastante área.
19. El realizar el reset posteriormente a la lectura viene forzado porque la capacidad de reset es compartida por toda la columna.

Es decir, se está corrigiendo la carga residual que deja la muestra actual cuando debería corregirse la carga residual producida por la anterior imagen. Es decir, existe una componente de error de carga residual, aun cuando se está llevando a cabo una lectura diferencial del píxel.

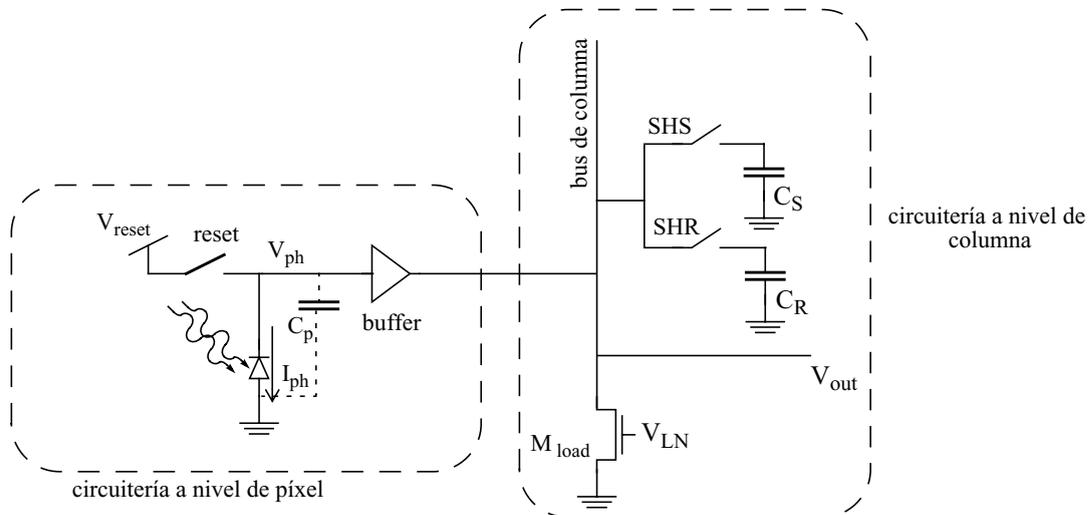


Fig. 4.13: Esquema de APS en tiempo discreto con lectura diferencial a nivel de columna

El principal problema de esta arquitectura es que, al igual que el píxel hace el volcado sobre el bus de columna, la circuitería a nivel de columna deberá hacer el volcado del dato sobre el bus de datos del sistema. Nos encontramos con la misma discusión llevada a cabo entre píxeles pasivos y activos: será conveniente realizar el volcado del dato en tensión en lugar de en intensidad, es decir, introducir buffers de tensión a la salida de la circuitería de columna. El esquema que responde a esta arquitectura se presenta en la Fig. 4.14 [6]. El principio de operación es el que sigue: tras el direccionamiento del píxel que va a ser leído se conectará la llave R_S durante un intervalo de tiempo para memorizar en C_R la señal generada por el ruido en ausencia de señal luminosa; terminada esta fase se procede a exponer el sensor a la imagen a capturar; tras el tiempo de exposición (tiempo durante el cual se habrá realizado la integración de la carga fotogenerada) se cierra S_S para memorizar el valor de salida del píxel en C_S ; tras esta fase se cierran R_X y S_X para proceder a la lectura diferencial.

La implementación de los buffers de tensión a nivel de columna no supone una reducción crítica del factor de relleno. Sin embargo, si bien con la lectura diferencial se logra contrarrestar tanto el ruido esp del píxel como la carga residual, el desapareamiento entre los buffers a nivel de columna introduce una componente de ruido de patrón fijo (**FPN a nivel de columna**) que no existía en la implementación de la lectura diferencial a nivel de píxel.

En resumen, la implementación de la lectura diferencial a nivel de columna tiene como clara ventaja que el factor de relleno del sensor mejora considerablemente, pero como inconveniente introduce una componente de FPN (FPN a nivel de columna) que no se corregirá.

4.3.2.- Eliminación programable del offset

Para determinadas aplicaciones de sensado de imagen resulta interesante que el sensor tenga un contraste espacial alto²⁰. Para conseguir esta característica en el dispositivo existe la posibilidad

20. Por contraste espacial se entiende que píxeles en una misma imagen sean capaces de distinguir niveles de intensidad luminosa bien distintos.

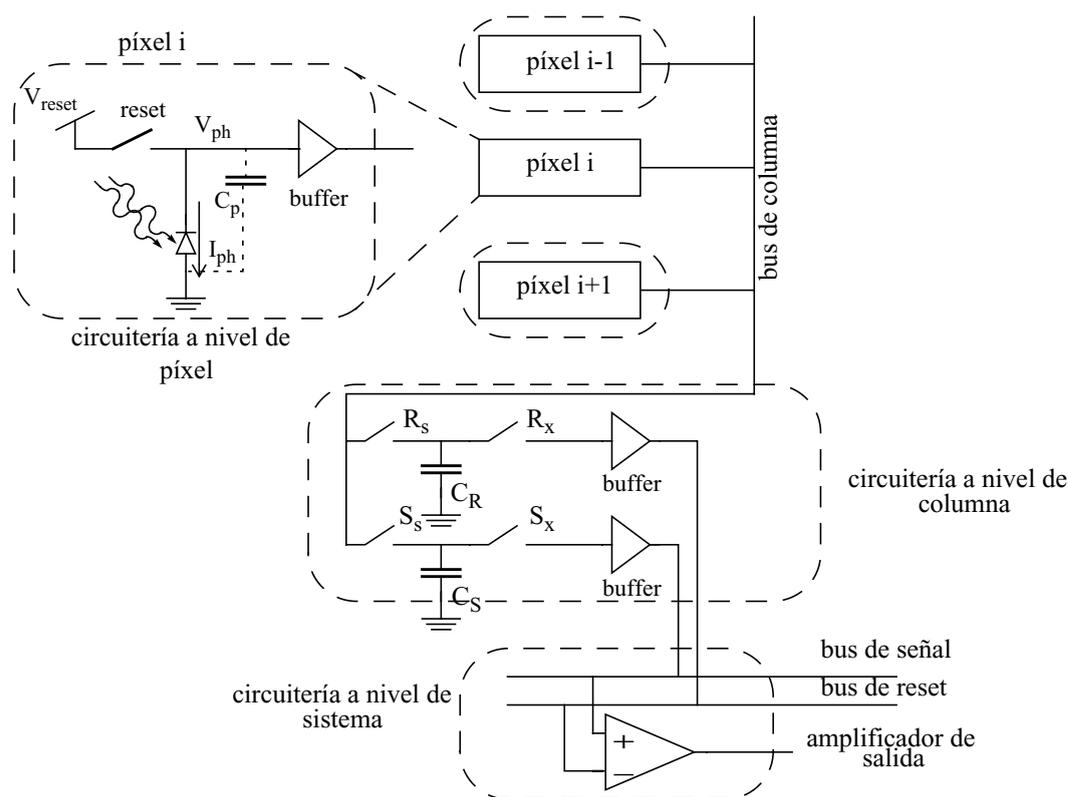


Fig. 4.14: Esquema de pixel con circuito de muestreo y retención a nivel de columna

de memorizar la corriente fotogenerada por una imagen (**corriente de fondo**) en cada píxel, de forma que el sensor responda a cambios del flujo de luz relativos a esta imagen. Esto es lo que se consigue con la **eliminación programable del offset**.

La Fig. 4.15 ilustra el esquema de eliminación programable del offset a nivel de píxel propuesto por Vietze y Seitz [7]. Cada píxel contiene un fotosensor en serie con una fuente de intensidad programable. Esta fuente de intensidad es fácilmente implementable con un MOSFET ($M_{current}$), cuya tensión de puerta puede fijarse a una determinada tensión con una capacidad (C_M) que memorice esta tensión. Además de la fuente de intensidad se implementa el transistor M_{prog} para memorizar la tensión de puerta de $M_{current}$. Los otros tres transistores responden al esquema típico de píxel por integración de la carga fotogenerada: M_{sense} es el amplificador en configuración de seguidor por fuente; M_{reset} es el transistor de reset; y M_{select} la llave que sirve para direccionar el píxel.

La fuente de intensidad opera en la región de inversión débil, de forma que la corriente de drenador depende exponencialmente de la tensión de puerta, por lo que la corriente típicamente se dobla con cada incremento de tensión de puerta de 30 mV. De esta forma la corriente de offset puede variarse fácilmente en un rango de entre 1 fA y varias decenas de μ A. El mismo mecanismo de integración de carga presentado en la Secc. 4.1.1 se emplea para la recogida de la carga fotogenerada. La carga medida representará la diferencia entre la carga total fotogenerada y la carga de offset.

El uso de este método permite alcanzar un rango dinámico de más de 150 dB, lo que permite la implementación sencilla de aplicaciones muy interesantes. Un ejemplo simple es un detector de cambio en la imagen, implementado en un proceso de dos etapas: en la primera de las etapas la

corriente de offset en cada píxel se programa de forma que el resultado neto total es cero (las corrientes de offset cancelan de forma efectiva las corrientes fotogeneradas locales); en un segundo paso la imagen objetivo es observada, con lo que el sensor extraerá una imagen diferencial resultado de las diferencias de carga fotogenerada entre la escena de referencia y la imagen objetivo. Cualquier cambio en la intensidad de luz de cualquier píxel tendrá como resultado una salida no diferencial, con lo que el cambio queda registrado por el sensor.

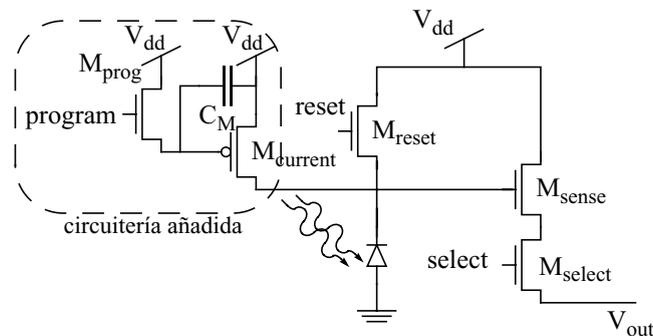


Fig. 4.15: Esquemas de píxel con eliminación programable del offset

4.3.3.- Píxel con ganancia programable

La multiplicación de la corriente fotogenerada por un factor programable a nivel de píxel resulta interesante para mejorar las características del factor de conversión de intensidad en tensión del píxel. Esta operación puede llevarse a cabo con una modificación sencilla de un espejo de corriente. En la configuración estándar del espejo de corriente se conectan los dos terminales de puerta de dos transistores; en la modificación propuesta por Vietze [8] se aplica una diferencia de potencial entre los dos terminales de puerta, tal y como se muestra en la Fig. 4.16. Esta diferencia de tensión se fija a un determinado valor (mediante el control sobre parámetros de fabricación), o bien puede implementarse como una diferencia de potencial individual programable a través de una capacidad de almacenamiento. La corriente fotogenerada en el fotodiodo en la primera rama del espejo de corriente se verá en la segunda rama multiplicada por una ganancia dada por el factor del espejo. Mediante el uso de un mecanismo muy similar al usado en el píxel de eliminación programable del offset, el píxel de ganancia programable dobla el valor de la variación de corriente por cada incremento de tensión de 30 mV. De esta forma puede conseguirse fácilmente el escalado de la corriente en un rango de varios órdenes de magnitud. Como antes, la corriente fotogenerada y multiplicada es posteriormente integrada sobre una capacidad de almacenamiento y leída por la circuitería de lectura.

Una aplicación de este esquema de operación es el sensor de alta sensibilidad descrito en [8] en el cual cada píxel tiene una ganancia aproximada de 78 dB. Se obtiene con esta ganancia una sensibilidad de 43 mV/e⁻, y un ruido referido a la entrada expresado en carga de 0.1 electrones a temperatura ambiente.

El precio a pagar es la reducción del ancho de banda del píxel, reflejado en las características paso de baja para corrientes fotogeneradas débiles, con tiempos de respuesta de varios milisegundos. Estos tiempos de operación no suponen un inconveniente en aplicaciones de vídeo.

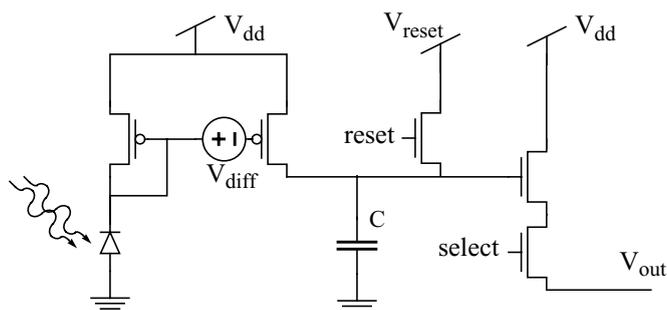


Fig. 4.16: Esquema de píxel con ganancia programable

4.3.4.- Circuito receptor adaptativo

El esquema de circuito receptor adaptativo, propuesto por T. Delbrück & C. A. Mead [9], se muestra en la Fig. 4.17. El esquema consta de cuatro bloques funcionales: **bloque de entrada**, encargado de generar la señal de entrada, que en este caso será la corriente fotogenerada en el elemento fotodetector; **bloque de referencia**, constituido por el transistor M_{fb} en configuración de seguidor por fuente, y cuya puerta será la que complete el lazo de realimentación; **bloque de comparación**, que se encargará de comparar el valor de la señal de entrada y de la señal de referencia; y **bloque de aprendizaje**, encargado de calcular una predicción de la señal de entrada para que la señal de referencia sea más precisa. El circuito se basa en la circuitería de realimentación de la puerta del transistor de realimentación (M_{fb}). Conceptualmente el circuito utiliza un modelo interno para hacer la predicción de la señal de entrada²¹. La salida del píxel será el resultado de la comparación entre la entrada y la predicción. El lazo implementa un algoritmo de aprendizaje muy sencillo para hacer que las predicciones sean más precisas.

El transistor M_{fb} (Fig. 4.17) se configura como seguidor por fuente. Para niveles normales de intensidad este transistor opera en inversión débil, por lo que la relación entre tensión de fuente del mismo (V_p) y la corriente fotogenerada (que coincidirá con la corriente de M_{fb}) será de tipo logarítmica. La tensión V_p se establece a un valor por debajo de V_f a la tensión necesaria para tener encendido M_{fb} de forma que suministre una corriente equivalente a la corriente fotogenerada. Conceptualmente, la tensión almacenada en la capacidad C_I actúa como modelo de la intensidad de entrada

La comparación entre el valor de tensión de entrada V_p y el predicho (señal de referencia) la lleva a cabo el amplificador inversor formado por M_n y M_p . Se añade un transistor cascode adicional (M_{casc}) del que se hablará más adelante. La tensión de entrada V_p controla la corriente suministrada por M_n . La corriente controlada por M_p está fijada por la tensión de polarización V_b . La ganancia en tensión (A_{amp}) de este amplificador viene determinada por el cociente entre la transconductancia de M_n y la conductancia de salida del amplificador, que a fin de cuentas viene determinada por las longitudes de los transistores que constituyen el amplificador. La tensión de polarización V_b determina la frecuencia de corte del sensor, estableciendo la corriente de polarización en el amplificador inversor.

El lazo de realimentación se completa cuando la salida V_{out} se realimenta a V_f a través del elemento adaptativo y a través del divisor capacitivo formado por C_1 y C_2 , que se muestra en la

21. Utiliza un esquema de filtrado paso de alta temporal

figura con uno de sus terminales fijado a V_{dd} (da igual que sea V_{dd} o tierra; lo importante es que el terminal tenga fijada la tensión).

Describamos cómo funciona el proceso adaptativo. Al incidir luz sobre el fotodetector se producirá un incremento de la intensidad que atraviesa el fotorreceptor. Este incremento en la intensidad hará que la tensión en el nodo V_p disminuya. Como respuesta a esta variación, la tensión de salida, al verse afectada por el amplificador inversor, mostrará este efecto multiplicado por la ganancia en tensión del mismo, es decir, por $-A_{amp}$. La tensión de salida será precisamente la variable que controle el lazo de realimentación mediante el acoplamiento de la tensión de salida a la tensión de del transistor de realimentación a través del elemento adaptativo, con una ganancia bastante baja. Resumiendo, una disminución de la tensión de fuente del transistor de realimentación hace que se aumente la tensión de puerta del mismo. Este aumento en la tensión de puerta del transistor de realimentación hará que aumente la tensión en el nodo de entrada V_p , completándose el lazo de realimentación. Existirá un conflicto entre quienes controlan la tensión V_p . Sin embargo, este conflicto queda rápidamente resuelto, pues la ganancia en tensión de la salida es varias décadas mayor que la ganancia del lazo de realimentación, por lo que el estado estacionario se alcanza con facilidad.

Con este proceso adaptativo conseguimos que las variaciones de la intensidad de entrada tomen como punto de referencia para su rango de variación el valor de referencia generado en el lazo de realimentación. La principal ventaja que se consigue con este proceso de adaptación es que el sensor se autopolariza en cada una de las adquisiciones de datos que se realizan. Ante un cambio de intensidad de entrada el sensor necesitará un cierto intervalo de tiempo en el que seguirá, según una respuesta logarítmica, un proceso de aprendizaje para que el lazo de realimentación permita que se alcance el valor central de la tensión de entrada en torno al cual se producirán las variaciones de tensión. A partir de este momento, el receptor responderá a los cambios de intensidad relativos a esta referencia. Con este modo de operación se aprovechan mucho mejor todas las características dinámicas del sensor, con lo que se podrá mejorar el rango dinámico y la sensibilidad. El inconveniente con el que habrá que trabajar es el retraso que introduce el lazo de realimentación.

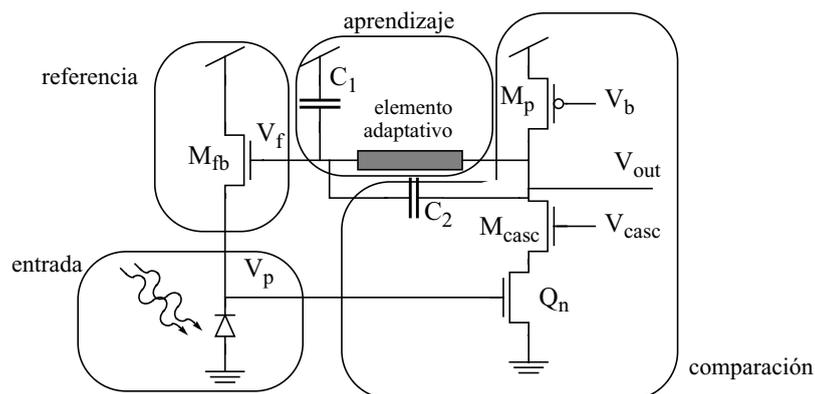


Fig. 4.17: Esquema del receptor adaptativo de T. Delbrück & C. A. Mead

Hay una gran desventaja en el uso de fototransistores frente al uso de fotodiodos en los fotorreceptores adaptativos: los fotorreceptores adaptativos basan su funcionamiento en el uso de un lazo de realimentación activo en el que existe un elemento capacitivo encargado de memorizar el valor anterior del píxel, de forma que la respuesta del valor actual se da adaptando la salida del valor nuevo del píxel partiendo del valor almacenado en el lazo de realimentación. La

realimentación activa acelera la respuesta fijando el nodo de entrada (nodo de contacto del elemento fotodetector con el resto del sensor). Cuando usamos un fototransistor en lugar de un fotodiodo, la realimentación fija el emisor, pero la base se deja flotando. De hecho, la base no puede fijarse si el transistor se va a usar para amplificar la corriente base-colector. La corriente disponible para cargar y descargar la base del transistor es aproximadamente la misma que para el fotodiodo. Como resultado de esta condición, el uso del fototransistor frena la aceleración que se consigue con el uso de la realimentación activa. El rango dinámico disminuye en torno a una o dos décadas en comparación con el uso del fotodiodo [9].

4.3.5.- Píxel activo en modo corriente con resolución variable

Las ventajas de la lectura en intensidad frente a la lectura en tensión se han expuesto en la Secc. 4.2.2. Presentamos a continuación un esquema que aprovecha la funcionalidad que este tipo de esquemas ofrece.

La Fig. 4.18(a) [5] muestra el esquema de un píxel implementado en tecnología CMOS estándar de $0.35\ \mu\text{m}$. Cada píxel está formado por un fotodiodo, interruptores de control y circuitería de amplificación de corriente. El fotosensor se implementa como una unión difusión p/ pozo n. El esquema del píxel con salida en corriente se muestra en la Fig. 4.18(a). Tras el reset, cuando M_4 está encendido, la carga generada por la absorción de luz en la unión pn del fotodiodo durante la fase de integración se acumula en el nodo V_{ph} . Al final de este periodo, M_3 se cierra y la carga se almacena en la capacidad parásita de M_1 , manteniendo la tensión de puerta de M_1 constante. La magnitud de la corriente de salida I_{pxl} , que se podrá leer en el bus de columna mediante la activación de M_2 , es una función inversamente proporcional a la intensidad de luz generada en el fotodiodo durante todo el proceso de integración. Debe hacerse notar, sin embargo, que esta función depende fuertemente de las características del transistor amplificador M_1 , que será quien introduzca la principal componente de FPN de ganancia.

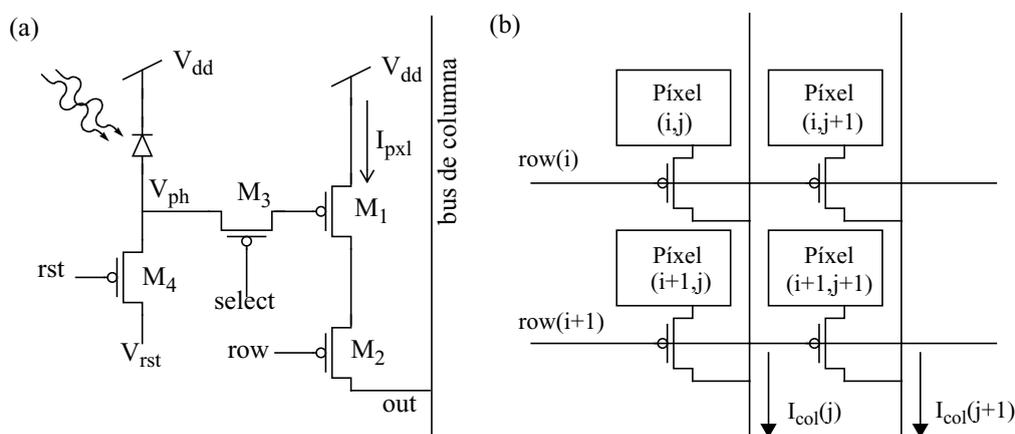


Fig. 4.18: Esquema de píxel en corriente de resolución variable

Dado que la salida de cada píxel es en corriente, nos permitirá inyectar la salida de varios nodos directamente en uno. Esto nos permite promediar las salidas de grupos de píxeles para disminuir la resolución de la matriz. Cuando la resolución de la matriz es máxima, tras el periodo de integración, las líneas se activan de una en una para su lectura.

Cuando los píxeles se agrupan por submatrices, por ejemplo, de 2×2 píxeles, se activarán las columnas correspondientes al mismo tiempo y las corrientes de salida serán la suma de todos los píxeles activados.

Cada columna tiene un espejo de corriente de cociente de ganancia I/R variable, donde R representa la resolución de la imagen (el tamaño del núcleo de píxeles (submatriz) será de $R \times R$). Una corriente de polarización se añade a cada espejo de corriente para asegurar que éstos permanecerán en saturación aún cuando la corriente de salida del espejo sea muy baja. Una señal de control permite que esta corriente de polarización se encienda o apague. La corriente de salida puede expresarse como:

$$I_{avg} = \frac{1}{R} \sum_{k=0}^{R-1} I_{col}(j+k) + I_{offset} \quad (4.13)$$

que es precisamente el promedio de las salidas de las columnas, con una componente de offset añadida, que puede ser cero o la suma de las corrientes de polarización en función de la señal de control de las corrientes de polarización.

4.4.- OTRAS MODULACIONES DE LA SEÑAL LUMINOSA

Los esquemas de lectura de la carga fotogenerada propuestos son los más extendidos en la implementación de sensores de imagen. Sin embargo, existen otras modulaciones de la carga fotogenerada que no responden a los anteriores esquemas, como son, por ejemplo, la modulación por ancho de pulso (PWM) y la modulación frecuencial del pulso (PFM). Describimos a continuación estos dos esquemas de modulación.

4.4.1.- Modulación por ancho de pulso (PWM)

Una señal modulada por ancho del pulso expresa el valor analógico del ancho de un pulso binario de tensión o de amplitud de corriente, como se muestra en la Fig. 4.19. El procesado de la señal en el dominio del tiempo elimina el procesamiento analógico de la tensión/intensidad, con lo que el sistema resulta inmune a la reducción del rango dinámico que el procesado analógico supone cuando las tensiones de alimentación del sistema son pequeñas (tecnologías de 3.3 V o menores)[10].

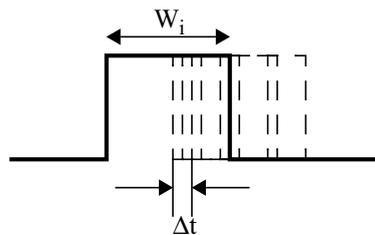


Fig. 4.19: Señal modulada por ancho del pulso PWM

La Fig. 4.20 muestra el diagrama de bloques de un chip de visión que implementa la técnica PWM [5]. La circuitería está compuesta por un fotodiodo (PD), un transistor de reset M_1 y un comparador (COMP). La tensión V_{PD} de la difusión flotante del fotodiodo estará inicialmente cargada a V_{dd} , e irá decreciendo a medida que la intensidad de luz incidente la vaya descargando. En $t = t_o$ cuando $V_{PD} > V_{ref}$ el comparador activará su salida. V_{ref} es la tensión de referencia del comparador. Si la salida permanece a nivel alto durante el intervalo, el ancho T del pulso se corresponderá con la intensidad de luz incidente. Si se desprecia el efecto sobre C_{PD} de la tensión de polarización, T viene dado por:

$$T = C_{PD} \frac{(V_{reset} - V_{ref})}{I_{ph}} \quad (4.14)$$

donde I_{ph} es la fotocorriente que atraviesa PD.

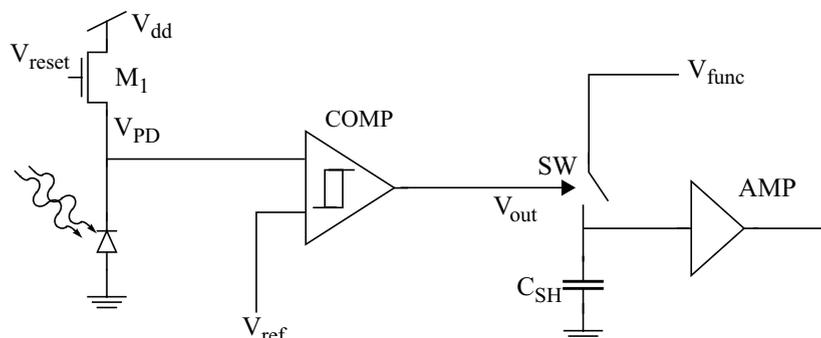


Fig. 4.20: Esquema de pixel con modulación de ancho de pulso PWM

La Fig. 4.21 muestra la implementación de la técnica de integración por corrientes en conmutación que realiza la aritmética necesaria para la implementación del PWM [10]. Las fuentes de intensidad en conmutación (SCS) convierten los pulsos PWM de tensión en pulsos de corriente. La integración de la corriente en C_{int} actúa de forma asíncrona y paralela al ancho de los pulsos. Con la reducción de las transiciones se obtienen ventajas sobre el consumo y se reduce el ruido que introduce la conmutación de los interruptores en comparación con los métodos binarios convencionales.

Estas propiedades hacen que el procesamiento de señal PWM se adapte a los algoritmos aritméticos que trabajan en paralelo para conseguir un procesamiento inteligente.

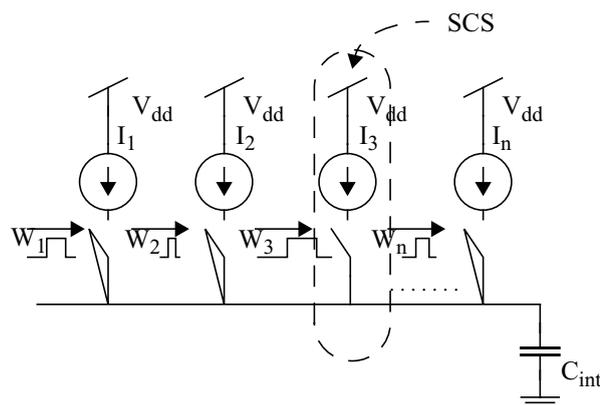


Fig. 4.21: Técnica de corrientes en conmutación para el procesamiento PWM

4.4.2.- Modulación frecuencial del pulso (PFM)

La modulación frecuencial del pulso es una representación del valor de salida del píxel en la cual se la salida analógica se convierte en un pulso frecuencial. Para el empleo de este tipo de

arquitectura en los sensores de imagen se emplean los circuitos osciladores representados en la Fig. 4.22. El fotodiodo actúa como fuente de intensidad variable controlada por la intensidad de luz de entrada, y se carga a través de M_1 . La puerta de M_1 se alimenta a través del lazo de realimentación, es decir, la salida de la cascada del Schmidt-Trigger y del inversor. En tal configuración, a mayor intensidad de luz, mayor frecuencia de pulso. El valor analógico de la intensidad de luz se convertirá en un tren de pulsos digital.

En [11] se presenta una implementación de este tipo de arquitectura. El sensor que se implementa en este estudio está diseñado para una tecnología CMOS de $1.2 \mu\text{m}$. El tamaño del píxel es de $50 \mu\text{m} \times 50 \mu\text{m}$. El fotodiodo se implementa como una unión difusión n-substrato p. La capacidad del fotodiodo es de 19 fF.

El lazo de realimentación de la puerta de M_1 consiste en un Schmidt-Trigger ST que binariza la salida del PD, y de un inversor INV para introducir el retraso necesario en la realimentación. En esta configuración, la frecuencia de oscilación f viene dada por:

$$f = \frac{I_{ph}}{C_{PD}(V_o - V_{th})} \quad (4.15)$$

donde V_o es la tensión de PD, V_{th} la tensión umbral de ST. Como puede observarse en la ecuación, la frecuencia f aumenta si la intensidad de luz se hace mayor.

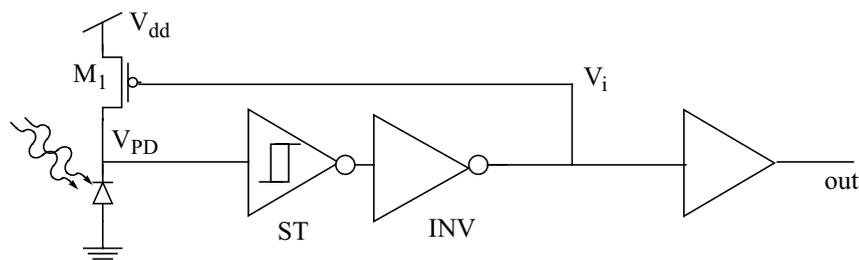


Fig. 4.22: Esquema de píxel con modulación frecuencial del pulso PFM

Referencias



- [1] Markus Loose "A self-calibrating CMOS image sensor with logarithmic response", Ph. D. thesis, Rupertus Carola University of Heidelberg, Germany, 1999
- [2] T. Lulé, H. Keller, M. Wagner, M. Böhm "LARS II- A high dynamic range image sensor with a-Si:H photo conversion layer", IEEE Workshop on CCDs and Advanced Image Sensors, Karuizawa, 1999, R34.
- [3] Y. P. Tsvividis "Operation and Modeling of the Mos Transistor" McGraw-Hill Series in Electrical Engineering
- [4] Richard Hornsey "Design and fabrication of integrated image sensors", Electrical and Computer Engineering, University of Waterloo, Waterloo, Ontario, Canada
- [5] Jonathan Coulombe, Mohamad Sawan, Chunyan Wang "Variable resolution CMOS current mode active pixel sensor" ISCAS 2000- IEEE International Symposium on Circuits and Systems, May 28-31, 2000, Geneva, Switzerland

-
- [6] Guy Meynants¹, Bart Dyerickx¹, Dirk Uwaerts¹, Jan Bogaerts² “Fixed pattern noise suppression by a differential readout chain for a radiation tolerant image sensor”, ¹Fillfactory, ²IMEC. 2001 IEEE Workshop on CCDs and AISs
 - [7] Vietze, O. and Seitz, P. “Image sensing with programmable offset pixels for increased dynamic range of more than 150 dB” in Conference on Solid State Sensor Arrays and CCD Cameras, 1996, San Jose, CA. Vol. 2654, pp 93-96
 - [8] Vietze, O. “Active pixel image sensors with application specific performance based on standard silicon CMOS processes” Ph. Thesis No. ETH-2038, Federal Institute of Technology, Zurich, Switzerland, 1997
 - [9] T. Delbrück & C. A. Mead “Analog VLSI phototransduction by continuous-time, adaptive, logarithmic photoreceptor circuits”, CNS Memo #30, Computation and Neural Systems Department, California Institute of Technology, Pasadena CA 91125
 - [10] M. Nagata, M. Homma, N. Takeda, T. Morie, A. Iwata “A Smart CMOS Imager with Pixel Level PWM Signal Processing”, Symposium on VLSI Circuits Dig. of Technical Papers, #14.4, pp. 141-144, Kyoto, June 17-19, 1999
 - [11] Jun Ohta, Akihiro Uehara, Takashi Tokuda, and Masahiro Nunoshita “Pulse-Modulated Vision Chips with Versatile-Interconnected Pixels”, IPDPS Workshops 2000: 1063-1071

RUIDO EN LOS SENSORES DE IMAGEN CMOS

5.1.- INTRODUCCIÓN

Al analizar el ruido en un sistema de adquisición de imágenes hemos de hacer frente al estudio de las componentes temporal y espacial del mismo. Atendiendo a esta descomposición podemos hacer una clasificación de las fuentes de ruido en:

- Fuentes de ruido temporal, que se modelarán como variables aleatorias dependientes del tiempo e independientes de su posición espacial en el sensor;
- Fuentes de ruido espacial, generarán un ruido independiente del tiempo, pero dependiente de la posición espacial que ocupe el dispositivo que genera el ruido en el sensor.

Describamos en detalle las fuentes de ruido presentes en un sensor de imagen.

5.1.1.- Fuentes de ruido temporal

Podemos definir el ruido temporal en un sensor de imagen como la variación en el tiempo del valor de salida del sensor bajo condiciones de iluminación constante. En general, la señal de salida en un sensor de imagen crece de forma proporcional al flujo de intensidad de luminosa incidente, mientras que el ruido crece proporcional a la raíz cuadrada de este flujo luminoso. De estas relaciones de proporcionalidad se deduce que la relación señal a ruido (SNR) del sensor mejorará con el incremento de iluminación. Dado que la calidad de la imagen depende, no del ruido, sino de la SNR, el efecto del ruido sobre la imagen final será más pronunciado en condiciones de baja iluminación. Es decir, el ruido será el factor que limite el comportamiento admisible del sensor cuando el flujo luminoso es débil. Definido el rango dinámico del sensor como el cociente entre los flujos luminosos máximo y mínimo admisibles para el funcionamiento del sensor, será el ruido el que determine el límite inferior.

Podemos hacer una clasificación de las principales fuentes de ruido temporal que aparecen en un sensor de imagen:

- **Fuentes de ruido intrínsecas** a los dispositivos semiconductores implementados. Son fuentes de ruido generadas por la propia naturaleza del material sobre el que se construyen los dispositivos que constituyen el sensor. Consideraremos como fuentes de ruido intrínsecas el ruido térmico, el ruido de disparo y el ruido flicker.
- **Ruido de cuantización**, resultado de llevar a cabo la conversión analógico-digital del valor de salida de los píxeles. Existirá sólo si la salida del sensor se digitaliza

- **Interferencias** provocadas por las variaciones de las fuentes de alimentación, presencia de campos electromagnéticos, ... Estos fenómenos provocarán fluctuaciones en la salida del sensor, y por lo tanto, podrán modelarse como fuentes de ruido temporal.

- **Ruido por acoplamiento**, provocado por la dispersión de carga a través de la oblea semiconductora hacia píxeles colindantes. Este fenómeno puede provocar una pérdida de sensibilidad del sensor a determinadas frecuencias de luz.

Algunas de estas fuentes de ruido pueden reducirse hasta niveles aceptables haciendo uso de un buen diseño del sistema, como por ejemplo:

- El ruido introducido por las fuentes de alimentación puede disminuirse haciendo un buen diseño del PSRR (Power Supply Rejection Ratio), o bien mediante el uso de fuentes de alimentación más precisas;

- Las interferencias existentes en el entorno, tales como los campos electromagnéticos, pueden reducirse mediante un mejor apantallamiento de éstas;

- El ruido por acoplamiento puede reducirse mediante el aislamiento de los píxeles con anillos de guarda, es decir, teniendo en cuenta este efecto a la hora de implementar el layout del sensor. Sin embargo, esta medida es cada vez más difícil de llevar a cabo, especialmente en tecnologías digitales, en las cuales la oblea semiconductora es bastante conductiva, permitiendo que la longitud de difusión de los portadores sea elevada, por lo que la eficiencia de los anillos de guarda sólo bloquean parcialmente la carga difundida a través de la oblea semiconductora.

Ahora bien, las componentes de ruido intrínsecas están generadas internamente por los dispositivos que componen el sensor, y resultan difíciles de suprimir. El modelado y la minimización de las mismas son los objetivos del estudio del ruido en los sensores de imagen.

Además de las fuentes de ruido expuestas, en los sensores APS CMOS por integración de carga es necesario llevar a cabo un análisis detallado del ruido en las diferentes fases de operación del mismo. En estos sensores, el ruido será, en un primer instante, muestreado en cada una de las fases de operación, y posteriormente transferido a la salida durante la fase de lectura. En la Secc. 5.4 se lleva a cabo un estudio detallado y riguroso del ruido para este tipo de arquitecturas que, por sus particulares características, necesitará de la aplicación de una metodología distinta al análisis frecuencial que comúnmente se aplica en el estudio del ruido.

5.1.2.- Fuentes de ruido espacial

El ruido espacial tiene como origen las variaciones en los parámetros tecnológicos de los dispositivos que componen el sensor. Idealmente, la salida del píxel resulta ser proporcional a la intensidad de flujo luminoso incidente¹. Sin embargo, el comportamiento del sensor se separa de esta relación lineal. Para expresar las no linealidades del valor de cada píxel en la imagen final capturada, podemos modelar la relación entre este valor y nivel de intensidad luminoso haciendo uso del desarrollo en serie de Taylor²:

$$V = \alpha_0 + \alpha_1 L + \alpha_2 L^2 + \alpha_3 L^3 + \dots \quad (5.1)$$

1. Esto no es así en los sensores logarítmicos, pero un desarrollo similar podría llevarse a cabo para este tipo de sensores

donde V es el valor de salida del píxel³; L el valor de la intensidad de flujo luminoso incidente; y α_i los coeficientes del desarrollo en serie de Taylor.

Debido a las variaciones de los parámetros tecnológicos entre dispositivos que se introduce de forma inevitable en el proceso de fabricación del sensor, los coeficientes α_i no tomarán los mismos valores para todos los píxeles. Para modelar este efecto se introducen las desviaciones típicas a través de las variables aleatorias $\Delta\alpha_i$ de los parámetros α_i en la ecuación 5.1:

$$V_{i,j}(L) = (\alpha_0 + \Delta\alpha_{0,i,j}) + (\alpha_1 + \Delta\alpha_{1,i,j})L + (\alpha_2 + \Delta\alpha_{2,i,j})L^2 + (\alpha_3 + \Delta\alpha_{3,i,j})L^3 + \dots \quad (5.2)$$

donde los subíndices i,j se utilizan para denotar la posición del píxel en la matriz de sensado (fila y columna, respectivamente). Se observa cómo con la introducción de estas desviaciones típicas la salida de cada uno de los píxeles no será la misma aun cuando sobre ellos incida el mismo nivel de luminosidad.

Podemos hacer la siguiente distinción de términos de la ecuación 5.2:

- Término en ausencia de iluminación:

$$V_{i,j}(L=0) = \alpha_0 + \Delta\alpha_{0,i,j} \quad (5.3)$$

- Término lineal:

$$V_{i,j}(\text{lineal}) = \alpha_1 L \quad (5.4)$$

- Resto de términos:

$$V_{i,j}(\text{resto}) = \Delta\alpha_{1,i,j} L + (\alpha_2 + \Delta\alpha_{2,i,j})L^2 + (\alpha_3 + \Delta\alpha_{3,i,j})L^3 + \dots \quad (5.5)$$

En general, en el diseño de sensores de imagen se persigue conseguir una relación lineal entre salida y nivel de iluminación que sea igual para todos los píxeles, es decir:

$$V_{i,j} = \alpha_1 L, \quad \forall(i, j)$$

Para conseguir este objetivo han de eliminarse los demás términos de la ecuación 5.2. Éste es el principal objetivo en el diseño de un sensor de imagen en lo que a eliminación del ruido espacial se refiere.

El término en ausencia de iluminación se denomina **ruido de patrón fijo** o **FPN** (Fixed Pattern Noise). Cuando el sensor no se ilumina, la imagen de salida debería tener un nivel de salida α_0 . Como se ha comentado, las variaciones en el proceso tecnológico introducen una variación $\Delta\alpha_{0,i,j}$ en cada píxel, de forma que la imagen en lugar de ser homogénea, presenta variaciones aleatorias del nivel de salida en cada píxel. Este efecto deteriora bastante la imagen final, por lo que se han desarrollado algoritmos para la eliminación de esta componente de ruido. El principio en el que se basan estos algoritmos es bastante sencillo: basta hacer una lectura diferencial del píxel mediante la cual restar a la salida del píxel (definida por la ecuación 5.2) la salida del mismo en ausencia de iluminación (definida por la ecuación 5.3):

$$V_{diff} = V_{i,j}(L) - V_{i,j}(L=0) = (\alpha_1 + \Delta\alpha_{1,i,j})L + (\alpha_2 + \Delta\alpha_{2,i,j})L^2 + \dots \quad (5.6)$$

2. Sería más preciso llevar a cabo el desarrollo en serie de McLaurin para modelar la existencia en el dispositivo fotosensor de una intensidad de fondo. Resulta más sencillo e ilustrativo para describir el ruido espacial centrarnos en el desarrollo en serie de Taylor
3. Consideramos la salida del píxel en tensión, de ahí que la salida se denote por V. Si fuese en intensidad el análisis sería exactamente el mismo

Con esta lectura diferencial también se elimina de la salida el coeficiente α_0 . Este término no incluye información alguna sobre el flujo luminoso incidente, por lo que su eliminación de la salida tiene efectos positivos sobre el funcionamiento del sensor (elimina la componente de continua de la salida del sensor).

La ecuación 5.5 agrupa todos los demás términos que desvían el comportamiento del sensor. Estos términos se agrupan en lo que se conoce como **fotorresponsividad no uniforme** o **PRNU** (Photo Response Non Uniformity):

$$PRNU = V_{i,j}(\text{resto}) = \Delta\alpha_{1,i,j} L + (\alpha_2 + \Delta\alpha_{2,i,j}) L^2 + (\alpha_3 + \Delta\alpha_{3,i,j}) L^3 + \dots \quad (5.7)$$

De forma general, los términos de segundo orden y superiores de la ecuación 5.7 se desprecian, por lo que el PRNU en general se estudia como un efecto de la no uniformidad de la ganancia del circuito, y no como una componente de ruido:

$$PRNU \approx \Delta\alpha_{1,i,j} \cdot L \quad (5.8)$$

5.2.- FUENTES DE RUIDO TEMPORAL

Como ya hemos comentado, determinadas fuentes de ruido pueden reducirse de forma efectiva mediante un buen diseño del sensor. No ocurre así con las fuentes de ruido intrínsecas, cuya presencia es inevitable. Analicemos, pues, cómo modelar eléctricamente el comportamiento de estas fuentes de ruido.

5.2.1.- Valor cuadrático medio y densidad espectral de potencia

Considérese una resistencia ruidosa de valor R a la que se le aplica una diferencia de potencial V entre sus extremos. La corriente total que atravesará dicha resistencia puede expresarse como:

$$I_R(t) = I_r + i_n(t) \quad (5.9)$$

donde $I_r(t)$ es la corriente ideal, es decir, en ausencia de ruido; e $i_n(t)$ la componente de ruido generado por la resistencia, o corriente ruidosa. El valor de $i_n(t)$ en un instante determinado es impredecible. El comportamiento de esta componente de corriente puede caracterizarse mediante parámetros estadísticos. Los parámetros que, en general, se utilizan con este propósito son el valor cuadrático medio $\langle \bar{i}_n^2(t) \rangle$ y la raíz cuadrada de este valor cuadrático medio $\langle \bar{i}_n(t) \rangle$ (asociado al subíndice rms: root mean square). Para la medición de estos parámetros, el ruido total observado depende del ancho de banda del instrumento de medición. Estos instrumentos, en general, tienen un ancho de banda muy estrecho Δf . Cuando el ruido esté referido a un determinado ancho de banda (en general coincidirá con el ancho de banda del aparato de medición), se expresará por notación como: $\Delta \langle \bar{i}_n^2(t) \rangle$. El cociente $\Delta \langle \bar{i}_n^2(t) \rangle / \Delta f$ se denomina **densidad espectral de potencia**⁴ (o *psd*: power spectral density) de la componente de ruido, y tendrá como unidades del Sistema Internacional A^2/Hz .

4. La densidad espectral de potencia puede expresarse integrando el ruido en toda la banda frecuencial, o considerando la integral de la misma entre 0 e ∞ . Este segundo caso es el que vamos a considerar. Teniendo en cuenta que las fuentes de ruido no tendrán parte imaginaria, la psd de la fuente de ruido será simétrica, por lo que la única diferencia entre utilizar una u otra expresión es de un factor 2.

De forma similar, si consideramos la componente de ruido en tensión en lugar de en intensidad, el ruido vendrá caracterizado por $\langle \bar{v}_n^{-2}(t) \rangle$, la densidad espectral de potencia se calculará como $\Delta \langle \bar{v}_n^{-2}(t) \rangle / \Delta f$, que tendrá unidades en el Sistema Internacional de V^2/Hz .

5.2.2.- Ruido térmico

El ruido térmico o ruido Johnson es consecuencia del movimiento aleatorio de origen térmico de los portadores de carga en las regiones resistivas (movimiento browniano). Como consecuencia del origen del mismo, este ruido depende de la temperatura del dispositivo, resultando independiente del flujo de corriente que lo atraviese. Se caracteriza por tener media cero y densidad espectral de potencia plana en un rango de frecuencias muy por encima de los GHzs, por lo que se modela como un ruido blanco gaussiano (WGN. White Gaussian Noise).⁵ El ruido térmico generado en una resistencia de valor R puede modelarse como una fuente de tensión $\langle \bar{v}_t^{-2}(t) \rangle$ en serie con una resistencia (Fig. 5.1(a)), caracterizada por:

$$\Delta \langle \bar{v}_t^{-2}(t) \rangle = 4kT \cdot R \cdot \Delta f \quad (5.10)$$

donde k es la constante de Boltzmann, y T la temperatura absoluta de la resistencia.

La densidad espectral de potencia del ruido térmico vendrá dada por⁶:

$$S_{V_t}(f) = \frac{\Delta \langle \bar{v}_t^{-2}(t) \rangle}{\Delta f} = 4kT \cdot R, \quad \forall f \quad (5.11)$$

Equivalentemente, el ruido térmico puede expresarse en intensidad, en lugar de en tensión (Fig. 5.1(b)), caracterizada por:

$$\Delta \langle \bar{i}_t^{-2}(t) \rangle = \frac{4kT}{R} \cdot \Delta f \quad (5.12)$$

con la correspondiente psd:

$$S_{I_t}(f) = \frac{\Delta \langle \bar{i}_t^{-2}(t) \rangle}{\Delta f} = \frac{4kT}{R}, \quad \forall f \quad (5.13)$$

Por ejemplo, el ruido térmico generado por una resistencia de $1k\Omega$ es de: $\langle \bar{v}_t^{-2}(t) \rangle = 16 \cdot 10^{-18} (V^2/(Hz))$, o equivalentemente $\langle \bar{v}_t^{-2}(t) \rangle = 4nV_{rms}/(Hz)$. Si consideramos ahora la contribución de este ruido en la banda de frecuencias de 100Hz hasta los 10kHz obtenemos un ruido total: $\Delta \langle \bar{v}_t^{-2}(t) \rangle = 398nV_{rms}$

5.2.3.- Ruido de disparo

El ruido de disparo o ruido shot está asociado con la discretización de la carga en los portadores cuando éstos atraviesan una región de deplexión. Estará, por lo tanto, asociado con el flujo de corriente en los diodos y en los transistores bipolares. Al igual que el ruido térmico, el ruido de disparo se modela como un proceso aleatorio WGN, es decir, de media cero, gaussiano, y con una densidad espectral de potencia plana en un ancho de banda del orden de GHzs. El

5. Las descripción detallada de este tipo de ruidos se lleva a cabo en el Apéndice B.

6. El rango frecuencial se considera Δf admitiendo como modelo un ruido WGN



Fig. 5.1: Circuitos equivalentes de una fuente de ruido térmico para una resistencia

ruido de disparo, en general, se representa como una fuente de intensidad caracterizada por el valor cuadrático medio del ruido $\langle i_s^{-2}(t) \rangle$ en paralelo con la corriente que lo genera, como se ilustra en la Fig. 5.2 para el caso de un diodo. Esta fuente de intensidad estará caracterizada por:

$$\Delta \langle i_s^{-2}(t) \rangle = 2q \cdot I_D \cdot \Delta f, \quad \forall f \quad (5.14)$$

donde q es la carga del electrón en culombios.

La Fig. 5.2 ilustra cómo la fuente de ruido de disparo aparece tanto para el diodo polarizado en directa como en inversa.

La psd del ruido de disparo resulta proporcional a la corriente I_D que atraviesa la unión:

$$S_{I_s}(f) = \frac{\Delta \langle i_n^{-2}(t) \rangle}{\Delta f} = 2q \cdot I_D, \quad \forall f \quad (5.15)$$

Por ejemplo, el ruido de disparo generado por un diodo polarizado en inversa a través del cual circula una intensidad $I_D=0.5pA$ es de: $\langle i_s^{-2}(t) \rangle = 16 \cdot 10^{-3} (A^2/(Hz))$, o equivalentemente $\langle i_s(t) \rangle = 0.4fA_{rms}/(Hz)$.

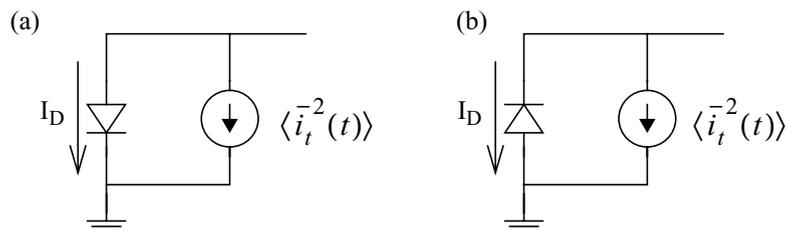


Fig. 5.2: Circuito equivalente de una fuente de ruido térmico para el diodo polarizado (a) en directa; y (b) en inversa

5.2.4.- Ruido flicker o ruido a baja frecuencia

Los orígenes del ruido flicker no se entienden del todo. Hay varios modelos propuestos que consideran todas las regiones de operación y todos los niveles de inversión de los transistores. Sin embargo, la mayoría de estos modelos son sólo válidos para transistores de canal largo y necesitan de parámetros raramente facilitados por las foundries.

En el *modelo de fluctuación de carga* [1] el ruido flicker se atribuye a la captura y posterior liberación de los portadores de carga en las trampas (estados energéticos dentro de la banda prohibida del semiconductor generados por las imperfecciones introducidas en el proceso de fabricación y por la existencia de la discontinuidad en la red cristalina que introduce la entrefase óxido-semiconductor). Estas fluctuaciones de carga generan un potencial superficial que modula la movilidad de los portadores de carga en el canal. Se asume que el canal puede intercambiar cargas con las trampas de la interfaz gracias al efecto túnel. El modelo de fluctuación de la movilidad [2] considera que el ruido flicker es el resultado de las fluctuaciones de la movilidad de la carga en el semiconductor. Ninguno de estos dos modelos explica de forma definitiva el comportamiento de la carga a baja frecuencia.

El punto importante en el que convergen las diversas teorías sobre el origen del ruido flicker y que ha sido verificado mediante medidas experimentales, es que la densidad espectral de potencia de esta componente de ruido varía de forma inversamente proporcional a la frecuencia. Debido a esta dependencia con la frecuencia el ruido flicker es un ruido rosa, a diferencia de los ruidos anteriores, que eran considerados blancos. Debido a su origen, el ruido flicker resultará independiente de la temperatura del dispositivo, resultando dependiente de la intensidad que lo atraviese.

Para un elemento resistivo plano, el ruido flicker puede modelarse como una fuente de tensión en serie con el elemento, caracterizada por un valor cuadrático medio [3]:

$$\Delta \langle \bar{v}_{flicker}^2 \rangle = KF_R \frac{R_{\square}^2}{A_R} V_R^2 \frac{\Delta f}{f} \quad (5.16)$$

donde A_R es la superficie del dispositivo; V_R la diferencia de potencial aplicada; R_{\square} la resistencia laminar; y KF_R una constante tecnológica que relaciona el ruido flicker con las características del material sobre el que se implementa el dispositivo.

De forma similar, la expresión del ruido flicker en intensidad en lugar de en tensión tiene una expresión:

$$\Delta \langle \bar{i}_{flicker}^2 \rangle = KF_R \frac{I_R^2}{A_R} \frac{\Delta f}{f} \quad (5.17)$$

La expresión general que se adapta para la densidad espectral de potencia para cualquier dispositivo es la siguiente:

$$S_{I_f}(f) = K \cdot I^c \cdot \frac{1}{|f|^n} \quad (5.18)$$

donde K es una constante de proporcionalidad, similar a KF_R , en la que se incluye la dependencia con las dimensiones del dispositivo; I la corriente que atraviesa el dispositivo; c una constante que toma valores en el rango: $1/2 \leq c \leq 2$; y n un número entero, que normalmente vale 1 para dispositivos implementados sobre semiconductor.

El modelo propuesto por la ecuación 5.18 no parece ajustarse al comportamiento real para transistores de canal corto. En el Apéndice D se incluye en modelo propuesto en [REF] y desarrollado en [4] en el cual se acepta el modelo de fluctuación de carga, y se modela cada una de las trampas genera una señal telegráfica aleatoria $N(t)$ (RTS: Random Telegraph Signal) caracterizada por tomar el valor 1 cuando la trampa captura un portador y 0 en otro caso. La densi-

dad espectral de potencia del ruido flicker será el resultado de la superposición de los efectos del conjunto de señales telegráficas generadas por el total de las trampas existentes en la interfaz.

5.3.- MODELO DE RUIDO TEMPORAL DEL DIODO Y DEL TRANSISTOR MOS

Una vez descritas las fuentes de ruido intrínsecas veamos los modelos que describen el comportamiento ruidoso de los principales dispositivos que aparecen en las arquitecturas de sensado de imagen: el diodo y el transistor MOS.

5.3.1.- Modelo de ruido temporal del diodo

El diodo aparece en los sensores de imagen como elemento fotodetector. Es por esta razón que en el tratamiento del ruido de este dispositivo consideraremos la presencia de una corriente fotogenerada, y haremos referencia al dispositivo como fotodiodo.

Las componentes de ruido dominantes en un fotodiodo son el ruido de disparo y el ruido flicker generados por las corrientes oscura (I_{dark}) y fotogenerada (I_{ph}). En la Fig. 5.7 se muestra el circuito equivalente de un diodo en el que se incluyen las fuentes de corriente I_{ph} e I_{dark} , así como las fuentes de ruido de disparo y flicker generadas por estas corrientes y caracterizadas por sus valores cuadráticos medios $\langle i_s^2 \rangle$ e $\langle i_f^2 \rangle$ respectivamente.

Las densidades espectrales de estas fuentes de ruido vienen dadas por las expresiones 5.15 y 5.18:

$$S_{I_s}(f) = 2q(I_{ph} + I_{dark}), \quad \forall f \quad (5.19)$$

$$S_{I_f}(f) = K \cdot (I_{dark})^c \cdot \frac{1}{|f|^n}, \quad \forall f \quad (5.20)$$

donde para el ruido flicker se admite la validez del modelo para cualquier frecuencia⁷.

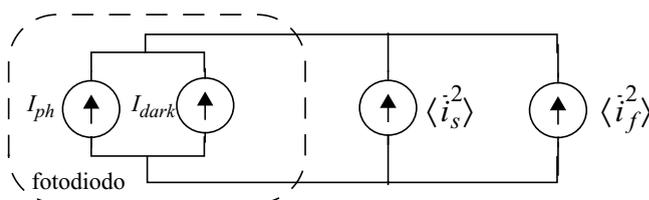


Fig. 5.3: Modelo de ruido del fotodiodo

Las fuentes de ruido de disparo y flicker son independientes. El paso de cada portador a través de una unión pn es un suceso puramente aleatorio que depende de la energía y de la velocidad que posean los portadores. El ruido de disparo es precisamente quien modela la naturaleza aleatoria de este fenómeno. El ruido flicker tiene su origen en las fluctuaciones tanto de la velocidad de recombinación superficial como de la movilidad de los portadores. Por lo tanto, resulta en general proporcional a la densidad de estados superficiales [5][6]. Como resultado, el ruido flicker no está sencillamente relacionado con la corriente total en el fotodiodo, sino que más bien depende del mecanismo mediante el cual la corriente es generada. Atendiendo a esta consideración, el ruido flicker en un fotodiodo tendrá su origen en la corriente oscura,

7. Para el rango de frecuencias de trabajo normales en los sensores de imagen esta consideración tienen bastante sentido

resultando prácticamente independiente de la corriente fotogenerada. Esto justifica que en la expresión de la psd del ruido flicker sólo aparezca la corriente oscura, y no la fotogenerada. En comparación, el ruido de disparo está generado por la suma de ambas componentes de la corriente. En el funcionamiento normal del sensor, la corriente fotogenerada es mucho mayor que la corriente oscura, por lo que el ruido flicker resulta mucho menor que el ruido de disparo. Aun cuando las corrientes sean del mismo orden, el ruido flicker seguirá siendo mucho menor que el ruido de disparo, a no ser que el fotodiodo trabaje a muy baja frecuencia (por debajo de 1Hz [4]).

5.3.2.- Modelo de ruido temporal del transistor

La Fig. 5.4 muestra la forma típica de la densidad espectral de potencia del ruido de la corriente en un transistor MOS en inversión fuerte⁸, muy parecida a la que aparece para un elemento resistivo. Pueden apreciarse dos regiones en las que el comportamiento del ruido es bien distinto. Ambas regiones se encuentran separadas por la que se denomina **frecuencia de codo** (f_c). El ruido dominante a altas frecuencias es el ruido térmico, caracterizado por su valor cuadrático medio $\langle i_t^2 \rangle$. El ruido dominante a bajas frecuencias es el ruido flicker, caracterizado por su valor cuadrático medio: $\langle i_f^2 \rangle$. Ambas componentes de ruido son incorreladas [7][8], por lo que el valor cuadrático medio de la componente de ruido de la corriente vendrá dado por la suma de los valores cuadráticos medios de ambas componentes:

$$\frac{\Delta \langle i_n^2(t) \rangle}{\Delta f} = \frac{\Delta \langle i_f^2(t) \rangle}{\Delta f} + \frac{\Delta \langle i_t^2(t) \rangle}{\Delta f} \quad (5.21)$$

Analicemos cada una de las componentes del ruido.

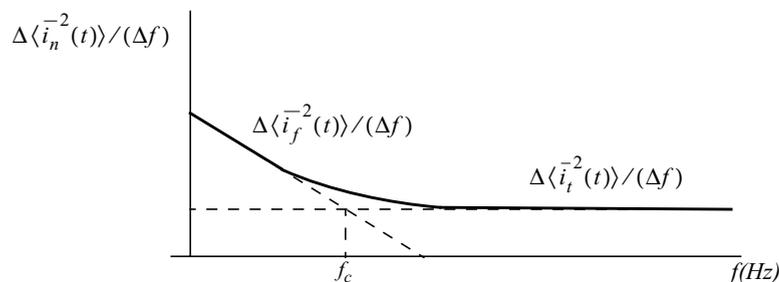


Fig. 5.4: Densidad espectral de potencia de ruido en transistor MOS, representada en escala logarítmica en ambos ejes

5.3.2.1.- Ruido térmico

El ruido térmico en un transistor MOS tiene su origen en el movimiento térmico aleatorio (movimiento browniano) de los portadores de carga en el canal.

En inversión fuerte, la corriente de drenador en ausencia de ruido puede expresarse como [9]:

$$I_D = -\mu W Q'_I(V_{CB}(x)) \frac{d}{dx} V_{CB}(x) \quad (5.22)$$

8. La curva se representa en coordenadas logarítmicas en ambos ejes

donde x es la posición a lo largo del canal; $V_{CB}(x)$ la tensión inversa de polarización efectiva del canal en inversión fuerte respecto del sustrato en función de x ; Q'_I la carga por unidad de área en la región de inversión; μ la movilidad; y W el ancho del canal.

Si consideramos un elemento de canal de longitud Δx centrado en torno a $x=x_I$, la expresión de la resistencia de este elemento puede obtenerse de la ecuación 5.22:

$$\Delta R = \frac{\Delta x}{-\mu W Q'_I(V_{CB}(x))} \quad (5.23)$$

donde el término de la derecha será positivo al cumplirse $Q'_I \leq 0$.

Asumiendo que el elemento de canal se comporta como un elemento resistivo de valor ΔR , el valor cuadrático medio del ruido térmico que se generará en este elemento, de acuerdo con la expresión 5.10, vendrá dado por:

$$\Delta \langle v_t^{-2}(t) \rangle \Big|_{elemento} = 4kT \cdot \Delta R \cdot \Delta f \quad (5.24)$$

y haciendo uso de la ecuación 5.23:

$$\Delta \langle v_t^{-2}(t) \rangle \Big|_{elemento} = \frac{4kT \cdot \Delta x}{-\mu W Q'_I(V_{CB}(x_1))} \cdot \Delta f \quad (5.25)$$

Puede demostrarse que la relación entre la variación en tensión introducida por la fuente de ruido y la correspondiente variación en intensidad para el transistor viene dada por [9]:

$$\Delta \langle i_t^{-2}(t) \rangle \Big|_{elemento} = \left[\frac{W}{L} \mu Q'_I(V_{CB}(x_1)) \right]^2 \Delta \langle v_t^{-2}(t) \rangle \Big|_{elemento} \quad (5.26)$$

De la combinación de las expresiones 5.25 y 5.26 se obtiene:

$$\Delta \langle i_t^{-2}(t) \rangle \Big|_{elemento} = -4kT \frac{\mu}{L^2} W Q'_I(V_{CB}(x_1)) \cdot \Delta x \cdot \Delta f \quad (5.27)$$

Esta ecuación expresa la contribución de un elemento de canal situado en x_I a la componente de ruido de la corriente del transistor. Las contribuciones de los elementos a lo largo del canal pueden considerarse incorreladas, por lo que el valor cuadrático medio puede hallarse como la suma de las contribuciones de cada elemento. Considerando Δx como un diferencial de longitud dx , e integrando a lo largo de todo el canal podemos obtener la expresión:

$$\Delta \langle i_t^{-2}(t) \rangle = -4kT \frac{\mu}{L^2} \int_0^L W Q'_I(V_{CB}(x_1)) dx \cdot \Delta f \quad (5.28)$$

Considerando Q_I la carga total en la región de inversión, es decir:

$$Q_I = \int_0^L W Q'_I(V_{CB}(x_1)) dx \quad (5.29)$$

la expresión 5.28 queda:

$$\Delta \langle i_t^{-2}(t) \rangle = 4kT \frac{\mu}{L^2} (-Q_I) \cdot \Delta f \quad (5.30)$$

Esta última expresión resulta válida para cualquier región de operación, con sólo modelar adecuadamente Q_I en cada caso.

Por ejemplo, en inversión fuerte⁹ [9]:

$$Q_I = -WLC'_{ox}(V_{GS} - V_T) \frac{2}{3} \frac{1 + \alpha + \alpha^2}{1 + \alpha} \quad (5.31)$$

donde:

$$\alpha = \begin{cases} 1 - \frac{V_{DS}}{V'_{DS}}, & V_{DS} \leq V'_{DS} \\ 0, & V_{DS} > 0 \end{cases} \quad (5.32)$$

expresión en la que V'_{DS} es la tensión que separa las regiones en las que el transistor opera o no en saturación.

Si particularizamos para el caso de saturación, utilizando el modelo de inversión fuerte, la expresión para el valor cuadrático medio del ruido térmico queda:

$$\Delta \langle \overline{i_t^{-2}}(t) \rangle = 4kT \frac{2}{3} \frac{W}{L} \mu C'_{ox}(V_{GS} - V_T) \cdot \Delta f \quad (5.33)$$

Puede observarse que de la ecuación 5.33 se deriva que la densidad espectral de potencia del ruido térmico en el transistor es independiente de la frecuencia.

Resulta común para caracterizar el ruido utilizar como parámetro el ruido en tensión equivalente a la entrada $\Delta \langle \overline{v_{in,t}^{-2}}(t) \rangle$. Se define como el ruido necesario en tensión entre la puerta y la fuente de un transistor no ruidoso para que el comportamiento del conjunto (fuente de ruido en tensión y transistor) equivalga al de un transistor ruidoso. A partir de la expresión de $\Delta \langle \overline{i_t^{-2}}(t) \rangle$ podemos definir el valor cuadrático medio del ruido en tensión equivalente a la entrada del transistor como:

$$\Delta \langle \overline{v_{in,t}^{-2}}(t) \rangle = \frac{\Delta \langle \overline{i_t^{-2}}(t) \rangle}{g_m^2} \quad (5.34)$$

donde g_m es la transconductancia del transistor.

Además de este valor cuadrático medio referido a la entrada, resulta bastante usual utilizar para la descripción del ruido la resistencia de ruido equivalente a la entrada R_n . Ésta es una resistencia ficticia cuyo valor es tal que la densidad espectral de potencia de la misma vale $\Delta \langle \overline{v_{in,t}^{-2}}(t) \rangle / \Delta f$.

5.3.2.2.- Ruido flicker

La expresión propuesta en [9] para la densidad espectral de potencia del ruido flicker, atendiendo a la expresión 5.18 particularizada para el caso de un transistor MOS, es la siguiente:

$$S_{I_f}(f) = \frac{M \cdot g_m^2}{WL \cdot (C'_{ox})^a} \cdot \frac{1}{f} \quad (5.35)$$

9. Para el desarrollo de las expresiones en otras regiones de inversión recomendamos la consulta de [9]

donde el exponente a toma el valor 1 ó 2; y M tiene dimensiones que dependen del valor usado para a , y que expresa la dependencia entre el ruido flicker y las características tecnológicas del dispositivo implementado. En la mayoría de dispositivos M muestra poca dependencia con el punto de polarización, aunque existen dispositivos que presentan una fuerte dependencia. Cuanto más limpio es el proceso de fabricación, menor es M .

La densidad espectral de potencia del ruido flicker referido a la entrada del transistor tiene una expresión:

$$S_{v_f}(f) = \frac{M}{WL \cdot (C'_{ox})^a} \cdot \frac{1}{f} \quad (5.36)$$

5.4.- ANÁLISIS DE RUIDO EN LOS SENSORES DE IMAGEN

Las principales fuentes de ruido presentes en los sensores de imagen APS en tiempo discreto tienen su origen en:

- El ruido generado por las corrientes oscura y fotogenerada en el dispositivo fotosensor;
- El ruido introducido por las circuiterías de lectura y amplificación;
- Las fluctuaciones de las fuentes de alimentación V_{dd} y V_{ss} .

Analicemos en detalle las fuentes de ruido tanto para los sensores PPS como para los sensores APS. Estamos interesados en el cálculo del ruido total referido a la salida al final de la fase de lectura. Para ello, se calcula la potencia de ruido media referida a la salida al final de cada fase de operación, y puesto que las contribuciones al ruido de cada fase pueden considerarse independientes, la potencia de ruido total a la salida al final de la fase de lectura vendrá dada por la suma de las contribuciones al ruido en cada una de las fases.

5.4.1.- Análisis del ruido en los sensores PPS

El esquema de sensor PPS que se va a utilizar para el análisis del ruido se muestra en la Fig. 5.5. Para el análisis se va a considerar un fotodiodo como elemento fotodetector.

Las fuentes de ruido que se van a considerar en cada fase de operación son las siguientes.

Durante la *fase de integración* tanto la llave de reset (M_1) como la de acceso (M_2) permanecen cortadas. Durante esta fase, los ruidos de disparo y flicker generados por las corrientes oscura y fotogenerada en el fotodiodo quedarán almacenados en la capacidad de integración C_{ph} .

Durante la *fase de reset*, se conecta la llave de reset, mientras que la llave de acceso permanece desconectada. Las fuentes de ruido que aparecerán son las siguientes:

- ruido térmico generado por la llave de reset;
- ruido generado por el amplificador operacional (opamp);
- ruido presente en el bus de columna debido al ruido de substrato y a otras fuentes de ruido;

Por último, durante la *fase de lectura*, la llave de acceso se conectará para realizar el volcado del dato sobre el bus de columna, mientras que la llave de reset permanece cortada. Las fuentes de ruido que habrá que tener en cuenta son:

- ruido térmico generado por la llave de acceso;
- ruido generado en el amplificador operacional;
- ruido presente en el bus de columna;

- componentes de ruido que se realimentan al fotodiodo.
- ruido introducido por los amplificadores a nivel de columna y de sistema.

Como puede observarse, los ruidos introducidos por el elemento fotodetector sólo se consideran en la etapa de integración. En el resto de las etapas el ruido introducido por el fotodiodo es despreciable, puesto que la lectura del fotodiodo se hace en tensión.

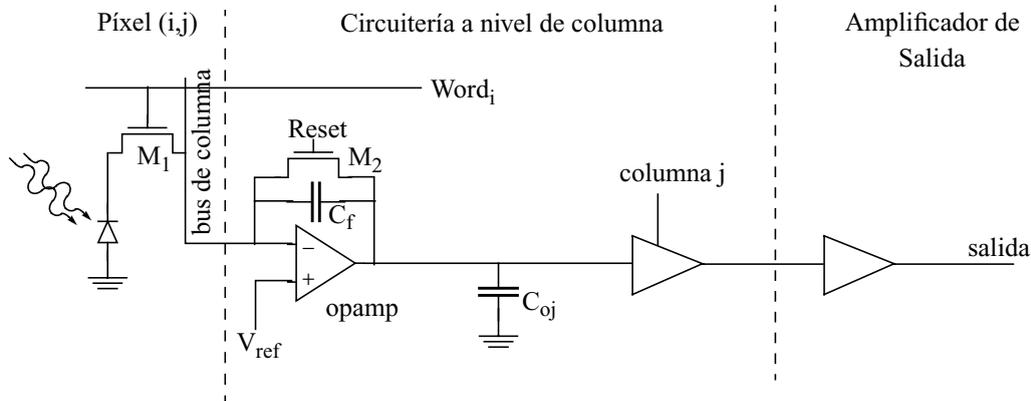


Fig. 5.5: Fotodiodo PPS CMOS

Para que el análisis de las diversas componentes de ruido sea relativamente sencillo asumimos las siguientes simplificaciones:

- se ignora el ruido flicker, por las razones expuestas en la Secc. 5.3.1.
- se modela el ruido referido a la entrada del amplificador operacional como un ruido en tensión, blanco, y limitado en banda, con un ancho de banda $B \ll A \cdot \omega_o$, donde $A \cdot \omega_o$ es el producto ganancia-ancho de banda del amplificador operacional.
- los efectos del ruido de sustrato sobre el bus de columna también se modelan como un ruido en tensión, blanco, y limitado en banda, de ancho de banda $B \ll A \cdot \omega_o$.

En el análisis del ruido no consideraremos las componentes de ruido generadas por los amplificadores de columna y de sistema, pues sus efectos podrán despreciarse frente a las demás componentes de ruido.

5.4.1.1.- Ruido generado durante la fase de integración

Durante la fase de integración las fuentes de ruido generadas por las corrientes oscura y fotogenerada se integran sobre la capacidad de integración. El circuito que modela este comportamiento se ilustra en la Fig. 5.6. Considerando que $I(t)$ es un proceso aleatorio estacionario en sentido amplio, y de media cero, caracterizada por su valor cuadrático medio $\langle i^2 \rangle$, entonces la potencia media para cualquier instante $t \geq 0$ puede calcularse como:

$$\begin{aligned}
\langle v_o^{-2}(t) \rangle &= \frac{1}{C_{ph}^2} \int_0^t \int_0^t \overline{I(t_1) \cdot I(t_2)} dt_1 dt_2 = \frac{1}{C_{ph}^2} \int_0^t \int_0^t R_I(t_1 - t_2) dt_1 dt_2 = \\
&= \frac{1}{C_{ph}^2} \int_{-t}^t (t - |\tau|) R_I(\tau) d\tau
\end{aligned} \tag{5.37}$$

donde $\tau = t_1 - t_2$, y $R_I(\tau)$ es la función de autocorrelación del proceso aleatorio.

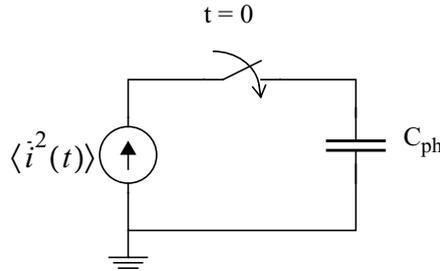


Fig. 5.6: Modelo de integración de una fuente de ruido sobre una capacidad

Si $I(t)$ es un ruido blanco, con una psd dada por:

$$S_I(f) = \frac{P}{2}, \quad \forall |f| \leq B \tag{5.38}$$

entonces la función de autocorrelación tiene una expresión:

$$R_I(\tau) = \frac{P}{2} \delta(\tau) \tag{5.39}$$

lo que permite simplificar la expresión 5.37:

$$\langle v_o^{-2}(t) \rangle = \frac{P}{2C_{ph}^2} \cdot t \tag{5.40}$$

Despreciando el ruido flicker, la única fuente de ruido presente será el ruido de disparo, caracterizado por una fuente de intensidad de valor cuadrático medio:

$$\langle \overline{i_{shot}^2} \rangle = 2q(I_{ph} + I_{dark}) \tag{5.41}$$

y psd:

$$S_{I_{shot}}(f) = 2q(I_{ph} + I_{dark}) \tag{5.42}$$

Puesto que el ruido de disparo es blanco, asumiendo que el proceso aleatorio es estacionario en sentido amplio, la potencia de ruido acumulada en la capacidad al final de la fase de integración y transferida a la salida durante la fase de lectura puede calcularse según la ecuación 5.40:

$$\langle \overline{v_{shot}^2} \rangle \Big|_{o, int} = \frac{q}{C_{ph}^2} (I_{ph} + I_{dark}) \cdot t_{int} \tag{5.43}$$

5.4.1.2.- Ruido generado durante la fase de reset

Las fuentes de ruido durante la fase de reset que analizamos a continuación son las siguientes:

- ruido térmico generado por la llave de reset, caracterizado por el valor cuadrático medio $\langle v_s^{-2} \rangle$;

- ruido generado en el amplificador operacional, caracterizado por el valor cuadrático medio $\langle \overline{v_{opamp}}^2 \rangle$, y que por aproximación se considerará un ruido blanco limitado en banda;
- ruido presente en el bus de columna, caracterizado por el valor cuadrático medio $\langle \overline{v_{bus}}^2 \rangle$, y que, igualmente, se considerará un ruido blanco limitado en banda.

La Fig. 5.7 representa el circuito equivalente del sensor en el que se incluyen las fuentes de ruido mencionadas. En tal esquema se considera el modelo de ruido de la llave de reset como el modelo de ruido descrito en la Secc. 5.3.2 para el transistor MOS operando en la región lineal. En tal región, el transistor se modela como una resistencia de valor la resistencia drenador-fuente del transistor en dicha región, en serie con una fuente de tensión de valor cuadrático medio dado por la psd del ruido térmico del dispositivo.

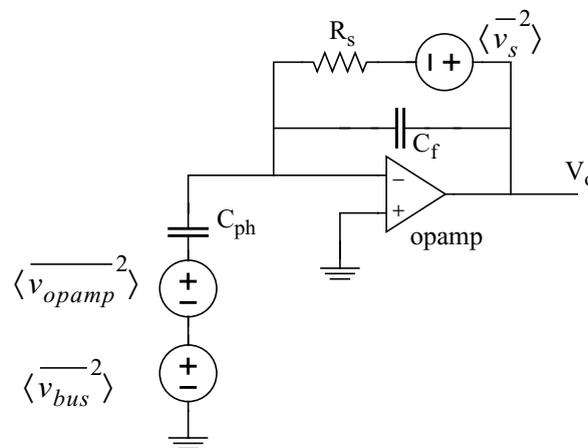


Fig. 5.7: Modelo del sensor PPS durante la fase de reset en el que se incluyen las fuentes de ruido

Las densidades espectrales de cada fuente de ruido son las siguientes:

$$\begin{aligned}
 S_{v_s}(f) &= 2kTR_s, \quad \forall f \\
 S_{v_{opamp}}(f) &= \frac{N}{2}, \quad \forall |f| \leq B \\
 S_{v_{bus}}(f) &= \frac{M}{2}, \quad \forall |f| \leq B
 \end{aligned} \tag{5.44}$$

Para calcular la potencia de ruido total referido a la salida, calculamos la potencia de ruido para cada una de las fuentes y aplicamos superposición.

5.4.1.2.1.- Potencia de ruido a la salida debido al transistor de reset

Considerando sólo el ruido generado por el transistor de reset, el circuito equivalente del sensor que incluye la fuente de ruido se muestra en la Fig. 5.8. Al final de la fase de reset, el ruido queda almacenado en las capacidades C_{ph} y C_f , y es posteriormente transferido a la salida al final de la fase de lectura.

Consideremos dos casos extremos del comportamiento del operacional. En primer lugar, consideremos que el amplificador operacional tiene un producto ganancia-ancho de banda infinito.

En tal caso, el circuito equivalente de la Fig. 5.8 se ilustra en la Fig. 5.9(a). Ocurre que en este caso no se almacena ruido alguno en C_{ph} . El ruido almacenado en C_f puesto que es un circuito RC, tendrá una potencia media igual a kT/C_f como se demuestra en el Apéndice C:

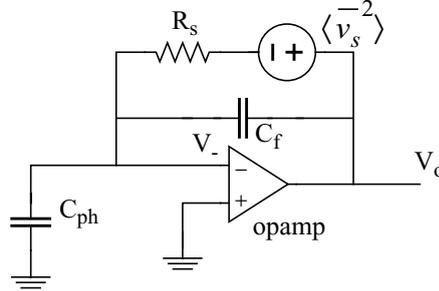


Fig. 5.8: Fuente de ruido de la llave de reset del sensor PPS durante la fase de reset

$$\langle v_s^{-2} \rangle \Big|_{o, reset} = \frac{kT}{C_f} \quad (5.45)$$

El segundo de los casos extremos consiste en considerar un producto ganancia-ancho de banda muy bajo para el amplificador operacional. En tal caso, el circuito equivalente se muestra en la Fig. 5.9(b), y al tratarse de un circuito RC, la potencia media de ruido vendrá dada por la expresión $kT/(C_{ph}+C_f)$.

Durante la fase de lectura el ruido almacenado en C_{ph} es amplificado, por lo que en el segundo de los casos la potencia media de ruido a la salida viene dada por:

$$\langle v_s^{-2} \rangle \Big|_{o, reset} = \frac{kT}{(C_{ph} + C_f)} \left(\frac{C_{ph}}{C_f} \right)^2 \quad (5.46)$$

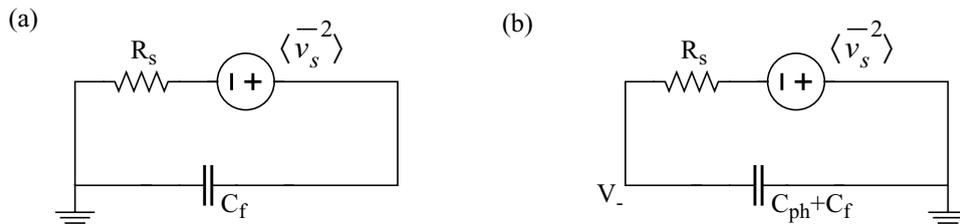


Fig. 5.9: Modelo del ruido de la llave de reset en el sensor PPS durante la fase de reset

Podemos obtener una expresión más precisa para un caso intermedio entre los propuestos. Para ello es necesario conocer la función de transferencia $V_-(f)/V_s(f)$. La calculamos utilizando para ello un modelo de un polo para el amplificador operacional:

$$\frac{V_-(f)}{V_s(f)} \approx \frac{j \left(\frac{f}{Af_o} \right)}{1 + j \frac{f}{Af_o} + j2\pi f R_s C_f - 2\pi f^2 \frac{(C_{ph} + C_f)}{Af_o} R_s} \quad (5.47)$$

A partir de la ecuación 5.47 se puede determinar la potencia media de ruido generada por la llave de reset durante la fase de reset:

$$\langle \overline{v_s^2} \rangle \Big|_{o, reset} = 2kTR_s \left(\frac{C_{ph}}{C_f} \right)^2 \int_{-\infty}^{\infty} \left| \frac{V_-(f)}{V_s(f)} \right|^2 df \quad (5.48)$$

5.4.1.2.2.- Potencia de ruido a la salida debido al amplificador y al ruido del sustrato

El circuito equivalente del sensor durante la fase de reset que incluye las fuentes de ruido debidas al operacional y al ruido del sustrato se ilustra en la Fig. 5.10. En este caso, las dos fuentes

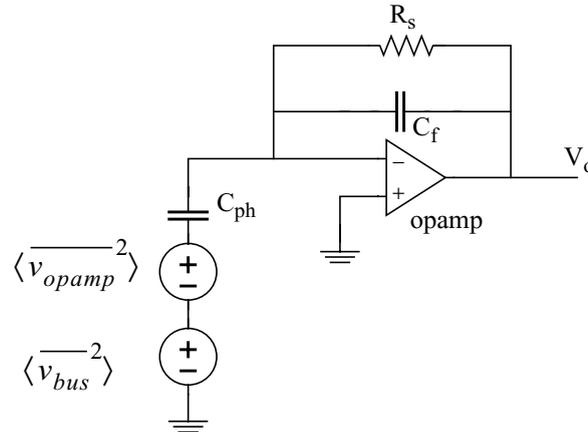


Fig. 5.10: Modelo del sensor PPS durante la fase de reset en el que se consideran las fuentes de ruido del amplificador y del sustrato

tes de ruido pueden unirse en una sola, cuya psd vendrá caracterizada por:

$$S_v(f) = \frac{N + M}{2}, \quad \forall |f| \leq B \quad (5.49)$$

Admitiendo que se cumple la condición $B \ll A \cdot \omega_o$, podemos asumir que el amplificador operacional se comporta de forma ideal, es decir, $V_- = 0$, con lo que la potencia media total almacenada en C_{ph} durante la fase de reset viene dada por $\left(\frac{N + M}{2} \right) 2B$. Esta potencia de ruido se

transfiere a la salida en la fase de lectura, por lo que la expresión de la potencia de ruido generado por el amplificador y por el sustrato durante la fase de reset, y referido a la salida, tiene la expresión:

$$\langle \overline{v_v^2} \rangle \Big|_{o, reset} = \langle \overline{v_{opamp}^2} \rangle \Big|_{o, reset} + \langle \overline{v_{bus}^2} \rangle \Big|_{o, reset} = (N + M)B \left(\frac{C_{ph}}{C_f} \right)^2 \quad (5.50)$$

5.4.1.3.- Ruido generado durante la fase de lectura

El esquema equivalente del sensor durante la fase de lectura que incluye las correspondientes fuentes de ruido se ilustra en la Fig. 5.10. La capacidad C_{bus} que aparece es la correspondiente al nodo del bus de línea. Las tres fuentes de ruido que aparecen son las siguientes:

- ruido térmico generado por la llave de acceso, caracterizado por el valor cuadrático medio $\langle \overline{v_{acc}^2} \rangle$;
- ruido generado en el amplificador operacional, caracterizado por el valor cuadrático medio $\langle \overline{v_{opamp}^2} \rangle$, y que por aproximación se considerará un ruido blanco limitado en banda;

- ruido presente en el bus de columna, caracterizado por el valor cuadrático medio $\langle \overline{v_{bus}^2} \rangle$, y que, igualmente, se considerará un ruido blanco limitado en banda.

Procedemos de forma similar a como en la Secc. 5.4.1.2: puesto que las tres fuentes de ruido son incorreladas y de media cero, calculamos la contribución de cada fuente de ruido a la potencia de ruido referida a la salida al final de la fase de lectura y usamos superposición.

5.4.1.3.1.- Potencia de ruido a la salida debido al transistor de acceso

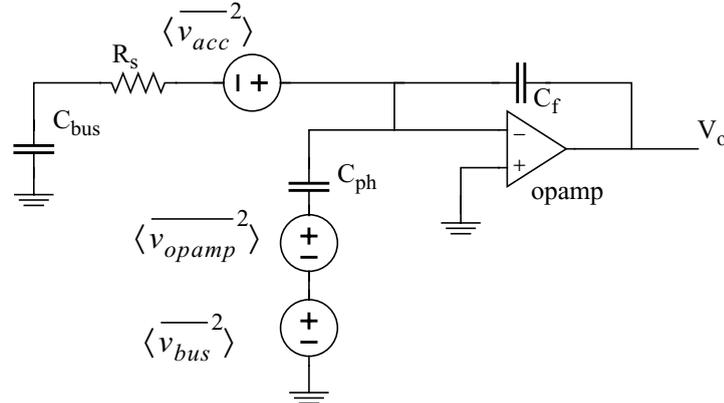


Fig. 5.11: Modelo del sensor PPS durante la fase de lectura con las correspondientes fuentes de ruido

Para simplificar el análisis, y puesto que no afecta demasiado a los resultados obtenidos, ignoramos la presencia de la resistencia R_s . El circuito equivalente para el análisis de esta sección, considerando la simplificación sobre R_s , se muestra en la Fig. 5.12. Para hallar la potencia de ruido referido a la salida calculamos en primer lugar la función de transferencia:

$$\frac{V_o(f)}{V_a(f)} = -\frac{C_{bus}}{C_f} \frac{1}{1 + j\frac{f}{\beta}} \quad (5.51)$$

donde:

$$\beta = \frac{Af_o C_f}{C_f + C_{bus} + C_{ph}}$$

Al final de la fase de lectura, la potencia media de ruido referida a la salida vendrá dada por:

$$\langle \overline{v_{acc}^2} \rangle \Big|_{o,lect} = 2kTR_s \left(\frac{C_{bus}}{C_f} \right)^2 \int_{-\infty}^{\infty} \frac{1}{1 + (f/\beta)^2} df = 2kTR_s \left(\frac{C_{bus}}{C_f} \right)^2 \beta \pi \quad (5.52)$$

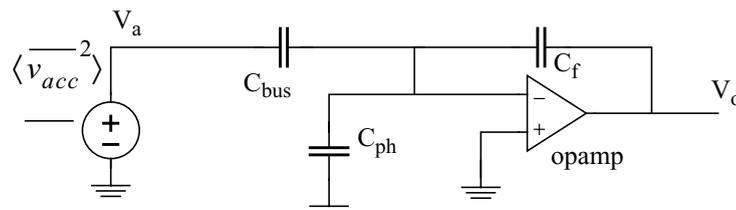


Fig. 5.12: Modelo del sensor PPS durante la fase de lectura en el que se considera el ruido introducido por la llave de acceso

5.4.1.3.2.- Potencia de ruido a la salida debido al amplificador operacional y al ruido de substrato

De nuevo, ignorando los efectos de R_s , y agrupando las dos fuentes de ruido en una sola, tal y como se hizo en la Secc. 5.4.1.2.2, el circuito equivalente que modela el efecto de estas dos fuentes de ruido durante la fase de lectura se ilustra en la Fig. 5.13.

El circuito es similar al de la sección anterior, por lo que la potencia media referida a la salida al final de la fase de lectura generada por estas dos fuentes de ruido viene dada por:

$$\langle v_v^{-2} \rangle \Big|_{o,lect} = 2kTR_s \left(\frac{C_{bus}}{C_f} \right)^2 \int_{-B}^B \frac{1}{1 + (f/\beta)^2} df = 2kTR_s \left(\frac{C_{bus}}{C_f} \right)^2 \cdot \arctan\left(\frac{B}{\beta}\right) \quad (5.53)$$

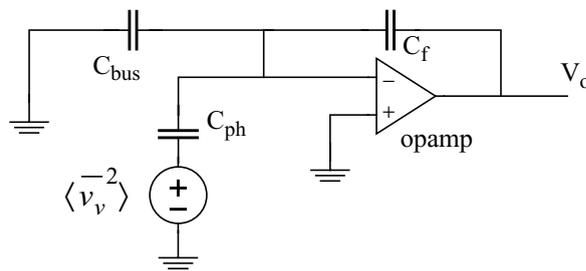


Fig. 5.13: Modelo del sensor PPS durante la fase de lectura en el que se considera el ruido introducido por el opamp y por el substrato

5.4.1.3.3.- Ruido realimentado en el fotodiodo

Al final de la fase de lectura, las tres fuentes de ruido que aparecen en esta fase de operación, almacenan una componente de ruido en el fotodiodo en la capacidad C_{bus} que será leída en la siguiente lectura. Analicemos cómo contribuye a este ruido cada una de las fuentes.

Consideremos el esquema del circuito que modela los efectos de la llave de acceso (Fig. 5.14).

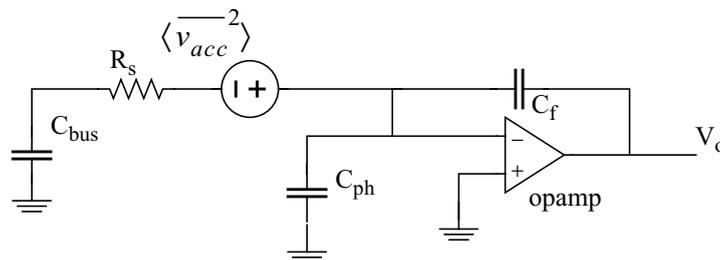


Fig. 5.14: Modelo del sensor PPS durante la fase de lectura para modelar el ruido realimentado a través de la llave de acceso

Estamos interesados en el cálculo de la potencia de ruido que quedó almacenada en la capacidad C_{bus} que se va a transferir a la salida al final de la fase de lectura. Considerando que el producto ganancia-ancho de banda del amplificador es bastante bajo, el circuito de la Fig. 5.15 representa una aproximación del circuito de la Fig. 5.14. La potencia de ruido que quedó almacenada en la anterior muestra, según el circuito RC equivalente de la Fig. 5.15, tendrá una expresión:

$$\langle \overline{v_{bus}^2}(t) \rangle \Big|_{lect} = \frac{kT}{C_{bus}}$$

Esta componente de ruido deberá añadirse a la potencia de ruido total a la salida. Esta potencia se transferirá al nodo de salida según la expresión:

$$\langle \overline{v_{bus}^2}(t) \rangle \Big|_{o, lect} = \frac{kT}{C_{bus}} \left(\frac{C_{bus}}{C_f} \right)^2$$

Respecto a las fuentes de ruido generadas por el amplificador operacional y por el ruido de

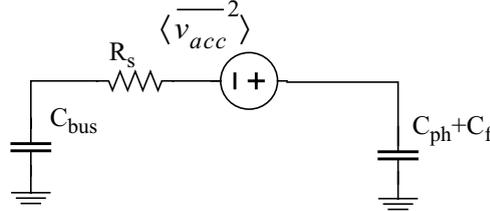


Fig. 5.15: Modelo equivalente del sensor PPS durante la fase de lectura con la correspondiente fuentes de ruido de la llave de acceso, considerando un producto ganancia-ancho de banda del opamp bajo

substrato, el circuito equivalente del sensor se ilustra en la Fig. 5.16. Para hallar la potencia de ruido almacenada en la capacidad C_{bus} debida a estas fuentes de ruido, en primer lugar calculamos la función de transferencia:

$$\frac{V_-(f)}{V(f)} = \frac{jf \frac{C_{ph}}{Af_o C_f}}{1 + jf \frac{1}{\beta}}$$

La potencia media de ruido almacenada en C_{bus} al final de la fase de lectura vendrá dada por la expresión:

$$\begin{aligned} \langle \overline{v_v^2}(t) \rangle \Big|_{lect} &= \left(\frac{N+M}{2} \right) \left(\frac{C_{ph}}{C_f} \right)^2 \frac{\beta^3}{(Af_o)^2} \int_{-B/\beta}^{B/\beta} \frac{(f/\beta)^2}{1 + (f/\beta)^2} d\left(\frac{f}{\beta} \right) = \\ &= (N+M) \left(\frac{C_{ph}}{C_f} \right)^2 \frac{\beta^3}{(Af_o)^2} \left(\frac{B}{\beta} - \arctan \frac{B}{\beta} \right) \end{aligned}$$

Este ruido se transferirá a la salida al final de la fase de lectura, con lo que la expresión de la potencia media de ruido a la salida será:

$$\langle \overline{v_v^2}(t) \rangle \Big|_{o, lect} = \left(\frac{C_{bus}}{C_f} \right)^2 (N+M) \left(\frac{C_{ph}}{C_f} \right)^2 \frac{\beta^3}{(Af_o)^2} \left(\frac{B}{\beta} - \arctan \frac{B}{\beta} \right)$$

5.4.2.- Análisis del ruido en los sensores APS por integración de carga

En el Apéndice C se presenta el estudio del ruido para sistemas lineales e invariantes en el tiempo. Sin embargo, en los sensores de imagen CMOS reales el circuito no es ni lineal ni invariante en el tiempo, por lo que para analizar el ruido en estas condiciones es necesario el desarrollo de una nueva metodología y nuevos modelos para estimar adecuadamente el ruido en estas arquitecturas [4].

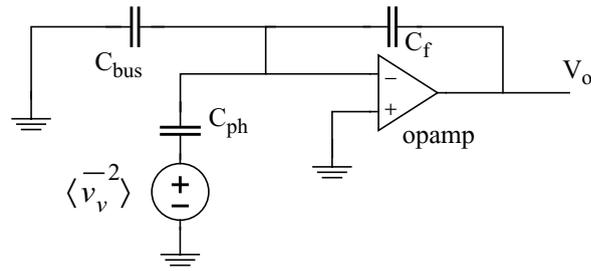


Fig. 5.16: Modelo del sensor PPS durante la fase de lectura en el que se considera el ruido realimentado en C_{bus} por el ruido del opamp y del sustrato

Analicemos el caso particular del sensor APS CMOS por integración de carga mostrado en la Fig. 5.17. Como se ha descrito en anteriores capítulos, la circuitería puede dividirse a tres niveles: circuitería a nivel de píxel, a nivel de columna y a nivel de sistema. El dispositivo fotosensor que se va a considerar para el análisis del ruido es el fotodiodo. Cada píxel contiene, además del fotodiodo, un transistor de reset M_1 , un seguidor por fuente M_2 , y un transistor de acceso M_3 . El circuito a nivel de columna incluye un transistor de polarización M_4 que actúa como fuente de intensidad para el seguidor por fuente del píxel, una capacidad C_{oj} , un amplificador, y circuitería añadida para funciones de control. A nivel de sistema la circuitería queda reducida a un amplificador (denominado amplificador de salida), y a un multiplexor que permite que el amplificador de salida sea compartido por todas las columnas.

Hay dos señales de control conectadas a cada píxel: *Word* y *Reset*. Estas señales digitales dividen el funcionamiento del píxel en tres fases u operaciones: fase de reset, de integración y de lectura. El sensor se resetea y se lee por filas. Al final de la fase de integración, los valores de los píxeles de una fila se leen y almacenan en las capacidades C_{oj} , situadas a nivel de columna. Una vez realizado el volcado de los datos de la fila sobre estas capacidades, la fila de píxeles se resetea, y los valores almacenados en las capacidades C_{oj} se transfieren secuencialmente al amplificador de salida haciendo uso del multiplexor.

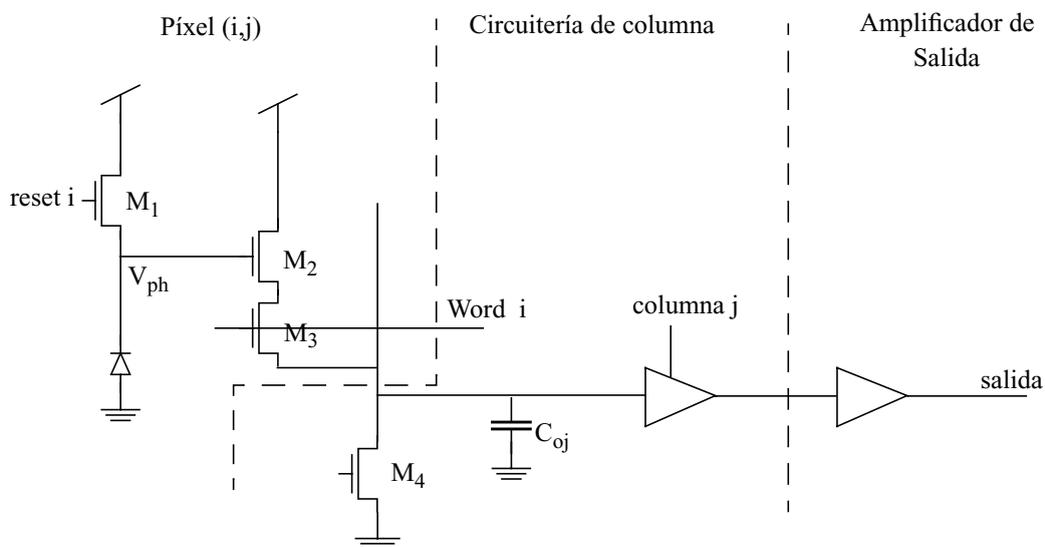


Fig. 5.17: Fotodiodo APS CMOS

Estamos interesados en el cálculo del ruido rms en tensión referido a la entrada en el nodo V_{ph} . Para calcularlo, de forma similar a como se ha procedido en la sección anterior, se calculan las potencias de ruido referidas al nodo de salida durante cada fase de operación, y una vez calculadas se transfieren al nodo de entrada haciendo uso de la función de transferencia del circuito. La potencia de ruido total será la suma de todas las contribuciones. El ruido generado durante la fase de lectura se calcula directamente en el nodo de salida, mientras que el ruido generado durante las fases de integración y de reset se muestrean sobre C_{oj} y luego se transfieren a la salida durante la fase de lectura.

En el análisis del ruido durante cada fase de operación se presta especial atención a la circuitería a nivel de píxel. Esto queda justificado por el hecho de que en los sensores CMOS típicos la circuitería a nivel de píxel es la que genera la mayor parte de la potencia del ruido.

Por las mismas razones que las expuestas en el análisis de los sensores PPS, en primera aproximación despreciamos los efectos del ruido flicker y centramos el análisis de los ruidos térmico y de disparo generados en el sensor en cada una de las fases. Resulta interesante el análisis del ruido flicker expuesto por Hui Tian en [4], por lo que recomendamos su consulta.

Respecto trabajos anteriores realizados respecto al ruido de disparo y térmico [10],[11],[12] muestran que en condiciones de baja iluminación las fuentes de ruido dominantes son las de ruido térmico y de disparo generadas por los transistores de reset y de lectura, mientras que en condiciones de iluminación alta es dominante la el ruido de disparo generado por el fotodiodo. La potencia de ruido debido al transistor de reset, el cual es muestreado al final de la fase de reset, se acota normalmente por la conocida expresión kT/C . Sin embargo, experimentos recientes [13],[14] demuestran que el ruido de reset para la estructura de la Fig. 5.17 es significativamente menor que la cota kT/C .

Otro factor a tener en cuenta en el análisis a realizar es el hecho de que en el estudio del ruido de disparo generado por el fotodiodo se asume que la relación entre carga e intensidad es lineal. A medida que las fuentes de alimentación se reducen con el escalado de las tecnologías, esta relación se hace cada vez menos lineal [4].

En esta sección se presenta un análisis detallado y riguroso del ruido térmico y de disparo para el sensor APS de la Fig. 5.17. Este análisis, llevado a cabo por Hui Tian en [4] tiene en cuenta los factores sobre el ruido de reset y la no linealidad de la capacidad del fotodiodo antes mencionados. El análisis del ruido de reset se basa en que durante la fase de reset el transistor de reset opera la mayor parte del tiempo en la región subumbral, y en que el estado estacionario no se alcanza, al cumplirse que el tiempo de establecimiento del circuito es varias veces mayor que el tiempo de reset típico en los sensores APS. Para calcular la potencia de ruido generada durante la fase de reset se considera un modelo de circuito variante en el tiempo, y se lleva a cabo un análisis del ruido en el dominio temporal, usando un modelo del transistor MOS en la región subumbral [15]. También se lleva a cabo un análisis del ruido de disparo generado en el fotodiodo que considera el efecto de las no linealidades de la relación entre carga y tensión en este dispositivo. Este análisis se lleva a cabo utilizando un modelo de circuito variante en el tiempo.

5.4.2.1.- Ruido generado durante la fase de reset

En el esquema general de reset en los sensores en tiempo discreto la capacidad de integración del dispositivo fotosensor se resetea a través de un transistor pMOS, como se muestra en la Fig.

5.18(a). La capacidad se carga a través de la resistencia de canal del MOS (R_{ch}), por lo que el circuito equivalente consiste en una capacidad (capacidad de integración C_{ph}) en serie con una resistencia (Fig. 5.18(b)).

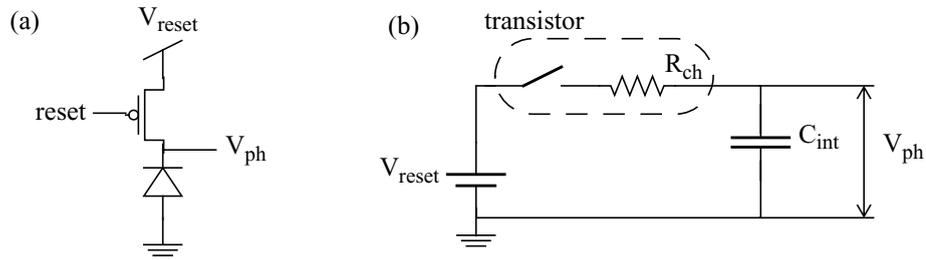


Fig. 5.18: (a) Circuito de reset del fotosensor; y (b) esquema equivalente

En el Apéndice C se lleva a cabo el análisis del circuito RC de la Fig. 5.18(b). La potencia media de salida viene dada por la expresión (C.9):

$$\langle v_n^{-2} \rangle \Big|_{reset} = \frac{kT}{C} \quad (5.54)$$

Esta expresión coincidirá con la potencia de ruido que introduce el transistor de reset.

A diferencia de este esquema de reset, en el de la Fig. 5.17 el reset de la capacidad de integración se lleva a cabo a través de un transistor nMOS. Las ventajas y desventajas son presentadas a lo largo de esta sección.

En el esquema de la Fig. 5.17, durante la fase de reset la puerta del transistor de reset M_1 se polariza a un nivel de tensión alto, típicamente a V_{dd} . Al comienzo de esta fase M_1 estará operando en inversión fuerte o en la región subumbral, dependiendo de la tensión en el fotodiodo al final de la fase de integración. Si la tensión del fotodiodo es suficientemente baja, M_1 operará inicialmente en inversión fuerte, y en un corto instante de tiempo entrará en la región subumbral en la que permanecerá durante el resto de la fase de reset.

El modelo de circuito durante la fase de reset, incluyendo las correspondientes fuentes de ruido, se ilustra en la Fig. 5.19. I_{ph} e I_{dark} son las corrientes fotogenerada y oscura respectivamente del fotodiodo; $I_d(t)$ es la corriente del transistor M_1 ; $i_d(t)$ son las variaciones de la corriente de M_1 debidas a la presencia de la fuente de ruido; e $\langle i_s^{-2} \rangle$ caracteriza el ruido de disparo generado en el fotodiodo por las corrientes oscura y fotogenerada. La psd de esta fuente de ruido vendrá dada por:

$$S_{I_s}(f) = 2q(I_{ph} + I_{dark}) \quad (5.55)$$

Si el tiempo de reset t_r es mayor que el tiempo de establecimiento t_{settle} (tiempo necesario para que la intensidad I_d iguale a la corriente $I_{ph} + I_{dark}$), entonces se alcanza el estado estacionario. En tal caso pueden aplicarse los métodos de análisis convencionales de ruido, mediante los cuales la potencia de ruido generado durante la fase de reset en el nodo V_{ph} tiene una expresión:

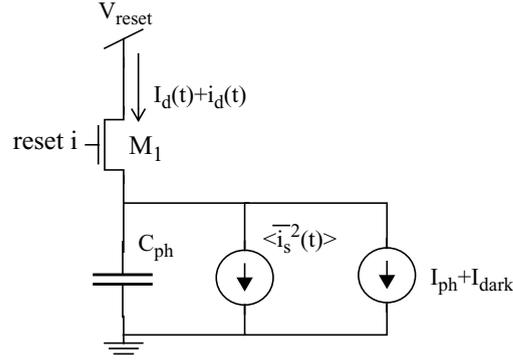


Fig. 5.19: Modelo de APS en tiempo discreto durante la fase de reset

$$\langle \bar{v}_n^2 \rangle \Big|_{reset} = \int_{-\infty}^{+\infty} \frac{2q(I_{ph} + I_{dark})}{(g_{m1} + g_{mb1})^2} \frac{1}{1 + \left(2\pi f \frac{C_{ph}}{g_{m1} + g_{mb1}}\right)^2} df \quad (5.56)$$

donde g_{m1} y g_{mb1} son las transconductancias de M_1 en la región subumbral, y el factor 2 aparece por el hecho de que en el estado estacionario $I_d = I_{ph} + I_{dark}$. Desarrollando la expresión 5.56 se obtiene:

$$\langle \bar{v}_n^2 \rangle \Big|_{reset} = \frac{q(I_{ph} + I_{dark})}{C_{ph}(g_{m1} + g_{mb1})} \quad (5.57)$$

Puesto que en la región subumbral: $I_d = \frac{kT}{q}(g_{m1} + g_{mb1})$ se llega a la expresión:

$$\langle \bar{v}_n^2 \rangle \Big|_{reset} = \frac{kT}{C_{ph}} \quad (5.58)$$

que coincide con la cota del ruido de reset típica.

Este resultado se alcanza suponiendo que se alcanza el estado estacionario durante la fase de reset. Ocurre que, por lo general, el tiempo de establecimiento es, en condiciones normales de iluminación, varias veces mayor que 1ms, mientras que los tiempos típicos de reset están en torno a varios microsegundos [4]. Por lo tanto, en general, el estado estacionario no se alcanza durante la fase de reset, lo que obliga a analizar el ruido en el dominio del tiempo.

Para llevar a cabo este análisis se parte de la aplicación de la ley de Kirchoff en intensidades en el nodo V_{ph} :

$$\frac{d}{dt}V_{PH}(t) = \frac{I_d(t) + i_n(t) - I_{ph} - I_{dark}}{C_{ph}(V_{PH}(t))} \quad (5.59)$$

donde: $i_n(t) = i_d(t) + i_s(t)$ es la componente de ruido de disparo y $V_{PH}(t)$ la tensión en el fotodiodo, que admite una descomposición en una componente no ruidosa $V_{ph}(t)$ más una componente de ruido $v_n(t)$:

$$V_{PH}(t) = V_{ph}(t) + v_n(t) \quad (5.60)$$

Es decir, la ecuación 5.59 puede escribirse como:

$$\frac{d}{dt}(V_{ph}(t) + v_n(t)) = \frac{I_d(t) + i_n(t) - I_{ph} - I_{dark}}{C_{ph}(V_{PH}(t))} \quad (5.61)$$

La capacidad C_{ph} , función de $V_{PH}(t)$, podrá aproximarse según la relación:

$$C_{ph}(V_{PH}(t)) = C_{ph}(V_{ph}(t) + v_n(t)) \approx C_{ph}(V_{ph}(t)) + \left(\frac{dC_{ph}}{dt}\right)v_n(t) \quad (5.62)$$

La intensidad $I_d(t)$ será igualmente función de $V_{PH}(t)$, por lo que linealizamos esta intensidad en torno a $V_{ph}(t)$:

$$I_d(t) \approx I_d(V_{ph}(t)) - g(t)v_n(t) \quad (5.63)$$

donde $g(t) = -dI_d/(dt)$.

Considerando, además de esta linealización, que la capacidad C_{ph} durante la fase de reset varía muy poco, y que, por lo tanto puede considerarse $C_{ph}(V_{PH}(t)) \approx C_{ph}(V_{ph}(t))$, la componente de ruido de la ecuación 5.61 queda:

$$\dot{i}_n(t) = C_{ph}(V_{ph}(t))\frac{d}{dt}v_n(t) + g(t)v_n(t) \quad (5.64)$$

Esta expresión es una ecuación diferencial lineal de primer orden. La solución de la misma en $t=t_r$, es decir, al final de la fase de reset, viene dada por:

$$v_n(t_r) = \int_0^{t_r} \frac{\dot{i}_n(\tau)}{C_{ph}(\tau)} \exp\left(-\int_\tau^{t_r} \frac{g(\tau_o)}{C_{ph}(\tau_o)} d\tau_o\right) d\tau \quad (5.65)$$

Para el caso en el que la función autocorrelación del ruido sea una función δ (que es el caso de los ruidos de disparo y térmico), la ecuación 5.65 nos lleva a:

$$\langle v_n^{-2}(t_r) \rangle \Big|_{reset} = \int_0^{t_r} \frac{R(\tau)}{2C_{ph}^2(\tau)} \exp\left(-2\int_\tau^{t_r} \frac{g(\tau_o)}{C_{ph}(\tau_o)} d\tau_o\right) d\tau \quad (5.66)$$

donde $R(\tau)$ es la psd de la fuente de ruido.

De la ecuación 5.66 puede comprobarse que la contribución del ruido una vez que el transistor opera en la región subumbral es extremadamente pequeña, y por lo tanto puede ignorarse, al igual que también puede ignorarse el ruido de disparo asociado a las corrientes $I_{ph} + I_{dark}$ puesto que estas corrientes son mucho menores que la corriente del transistor de reset I_d [4].

Bajo estas simplificaciones, $R(\tau) = 2qI_d(\tau)$, y considerando C_{ph} constante, se llega a la expresión:

$$\langle v_n^{-2}(t_r) \rangle \Big|_{reset} = \frac{1}{2} \frac{kT}{C_{ph}} \left(1 - \frac{\delta^2}{(t_r - t_1 - \delta)^2} \right) \quad (5.67)$$

donde:

$$\delta = \frac{U_T C_{ph}}{I_d(0)} \quad (5.68)$$

es el tiempo térmico, es decir, el tiempo necesario para cargar la capacidad del fotodiodo a una tensión U_T usando una corriente $I_d(0)$; y t_1 es el momento en el cual el transistor de reset para de estar por encima del umbral a estar por debajo, es decir, el tiempo que el transistor permanece en saturación.

De la ecuación 5.67 se deriva que el valor cuadrático medio del ruido generado durante la fase de reset está acotado por $\frac{1}{2} \frac{kT}{C_{ph}}$.

5.4.2.2.- Imagen residual generada por un reset incompleto

Se ha demostrado que el ruido de reset está acotado, para el esquema de reset estudiado, por la cantidad $\frac{1}{2} \frac{kT}{C_{ph}}$. Ocurre que esta cota se alcanza a expensas de que aparece en el sensor una imagen residual. Puesto que no se alcanza el estado estacionario, la tensión al final de la fase de reset será función de la tensión en el fotodiodo al comienzo de esta fase, generándose por lo tanto un residuo en la imagen.

La componente residual inducida por el reset incompleto puede eliminarse utilizando como transistor de reset un pMOS, o sobrecargando la puerta del transistor de reset. En ambos casos el transistor opera en la región lineal durante la fase de reset, con lo que la tensión final en la capacidad de integración será V_{reset} , independientemente de la tensión inicial en el fotodiodo, por lo que la imagen residual es eliminada. Sin embargo, si se recurre a alguna de estas dos técnicas, el ruido de reset se incrementa significativamente hasta la cota típica de $\frac{kT}{C_{ph}}$.

5.4.2.3.- Ruido durante la fase de integración

Durante la fase de integración resulta dominante el ruido de disparo generado por las corrientes oscura (I_{dark}) y fotogenerada (I_{ph}). La psd de este ruido vendrá dada por:

$$S_{I_s}(f) = 2q(I_{ph} + I_{dark}) \quad (5.69)$$

Partiendo de la ecuación 5.61 y considerando que el transistor de reset está apagado, y asumiendo que C_{ph} permanece constante durante toda la fase de integración, la potencia media de ruido al final de la fase de integración puede calcularse según los medios analíticos convencionales, llegándose a una expresión:

$$\langle v_n^{-2}(t_{int}) \rangle \Big|_{int} = \frac{q(I_{ph} + I_{dark})}{C_{ph}^2} t_{int} \quad (5.70)$$

Este resultado se alcanza considerando la capacidad de integración C_{ph} constante. Ocurre que C_{ph} es función de la tensión aplicada, y puede variar significativamente durante la fase de integración. Es necesario, por lo tanto, llevar a cabo un análisis del circuito en esta fase de operación en el dominio del tiempo.

Recordando la ecuación 5.64, y considerando la capacidad C_{ph} como función de la tensión aplicada, puede escribirse:

$$i_n(t) = C_{ph}(V_{ph}(t)) \frac{d}{dt} v_n(t) + g_c(t) v_n(t) \quad (5.71)$$

donde:

$$g_c(t) = -\frac{1}{C_{ph}(V_{ph}(t))} \left(\frac{d}{dt} C_{ph}(V_{ph}(t)) \right) (I_{ph} + I_{dark}) \quad (5.72)$$

es la conductancia de la capacidad variable.

La ecuación 5.71 es una ecuación diferencial lineal de primer orden, de la que puede extraerse la expresión del valor cuadrático medio del ruido al final de la fase de integración:

$$\begin{aligned} \langle \overline{v_n^2}(t_{int}) \rangle \Big|_{int} &= \\ &= q(I_{ph} + I_{dark}) \int_0^{t_{int}} \frac{1}{2C_{ph}^2(V_{ph}(\tau))} e^{-2(I_{ph} + I_{dark}) \int_{\tau}^{t_{int}} \left(\frac{d}{dV_{ph}(\tau_o)} \left(\frac{1}{C_{ph}(V_{ph}(\tau))} \right) \right) d\tau_o} d\tau \end{aligned} \quad (5.73)$$

Si consideramos ahora la simplificación de unión pn abrupta, entonces la dependencia entre C_{ph} y V_{ph} viene dada por:

$$C_{ph}(V_{ph}(t)) = C_{ph}(V_{ph}(0)) \sqrt{\frac{V_{ph}(0) + \phi}{V_{ph}(t) + \phi}} \quad (5.74)$$

donde ϕ es el potencial de contacto de la unión y $V_{ph}(0)$ la tensión en C_{ph} el comienzo de la fase de integración.

Partiendo de nuevo de la ecuación 5.61, y considerando los términos en ausencia de ruido, puede escribirse para la fase de integración:

$$\frac{d}{dt} V_{ph}(t) = -\frac{I_{ph} + I_{dark}}{C_{ph}(V_{ph}(t))} \quad (5.75)$$

La ecuación 5.75 permite la resolución de la ecuación 5.73:

$$\langle \overline{v_n^2}(t_{int}) \rangle \Big|_{int} = \left[\frac{q(I_{ph} + I_{dark})}{C_{ph}^2(V_{ph}(0))} \cdot t_{int} \right] \cdot \left[1 - \frac{1}{2(V_{ph}(0) + \phi) C_{ph}(V_{ph}(0))} t_{int} \right]^2 \quad (5.76)$$

5.4.2.4.- Ruido generado durante la fase de lectura

Durante la fase de lectura, el ruido predominante es el térmico y flicker generado por los transistores M_2 y M_3 , así como por los circuitos a nivel de columna y de sistema. Ignorando el ruido introducido por las circuiterías a nivel de columna (excepto la contribución de M_4) y de sistema, pues son muy pequeñas en comparación con las demás componentes, el ruido de lectura puede calcularse según el modelo de pequeña señal de la Fig. 5.20. En esta figura, $\langle \overline{i_{M2}^2}(t) \rangle$, $\langle \overline{v_{M3}^2}(t) \rangle$ y $\langle \overline{i_{M4}^2}(t) \rangle$ son las fuentes de ruido térmico asociadas a M_2 , M_3 y M_4 respectivamente, g_{m2} es la transconductancia de M_2 , g_{d3} es la conductancia de canal de M_3 , y C_o es la capacidad a nivel de columna, incluida la capacidad parásita del bus de columna. Asumiendo estado estacionario, lo cual está bien justificado durante la fase de lectura, puede demostrarse fácilmente que las componentes de ruido referido al bus de columna debido a M_2 , M_3 y M_4 vienen dadas respectivamente por [4]:

$$\begin{aligned}
\overline{\langle v_{M2}^2(t) \rangle} &= \frac{2kT}{3C_o} \frac{1}{1 + g_{m2}/g_{d3}} \\
\overline{\langle v_{M3}^2(t) \rangle} &= \frac{kT}{C_o} \cdot \frac{1}{g_{d3} \left(\frac{1}{g_{d3}} + \frac{1}{g_{m2}} \right)} \\
\overline{\langle v_{M4}^2(t) \rangle} &= \frac{2kT}{3C_o} g_{m4} \left(\frac{1}{g_{d3}} + \frac{1}{g_{m2}} \right)
\end{aligned} \tag{5.77}$$

Estas ecuaciones muestran que diferentes fuentes de ruido están asociadas con diferentes anchos de banda, y por lo tanto tienen diferentes efectos sobre el ruido referido al bus de columna.

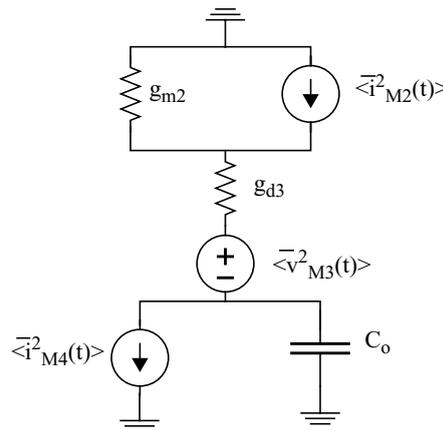


Fig. 5.20: Esquema del circuito de pequeña señal del APS en tiempo discreto durante la fase de lectura

5.5.- TÉCNICAS DE REDUCCIÓN DEL RUIDO TEMPORAL

En la Secc. 5.1.1 se describe, grosso modo, cómo reducir el ruido temporal generado por las fuentes de alimentación, las interferencias, y por acoplamiento. En la Secc. 5.4 se han propuesto modelos eléctricos para llegar a definir la densidad espectral de potencia de las fuentes de ruido intrínsecas (ruido térmico, de disparo y flicker). Conocidas estas expresiones, en esta sección analizamos qué medidas se pueden tomar para minimizar esta potencia.

Además de las fuentes de ruido intrínsecas, queda por analizar cómo llevar a cabo la reducción del ruido para la arquitectura APS CMOS por integración de carga. En esta arquitectura, se han de analizar de forma independiente las fases de reset, integración y lectura. Durante la fase de integración, la fuente de ruido dominante es el ruido de disparo que aparece en el dispositivo fotosensor, por lo que no podemos hacer prácticamente nada al respecto para su reducción. En la fase de lectura, la arquitectura del sensor está compuesta por el buffer de salida del píxel y la capacidad que aparece en el nodo del bus de columna, por lo que las medidas a tomar para la reducción del ruido serán las mismas que para cualquier circuito analógico similar, resultando un análisis sencillo. La reducción del ruido de reset es la más compleja de analizar, pues el comportamiento de esta fase dependerá fuertemente del valor de tensión inicial en la capacidad de integración. En esta sección abordamos el análisis de estructuras que consiguen la reducción del ruido de reset sin deteriorar otras características del sensor.

5.5.1.- Reducción del ruido intrínseco

Para ver cómo podemos minimizar la densidad espectral de potencia consideremos la potencia total de ruido intrínseco en un dispositivo como la suma de las componentes térmica, de dis-

paro y flicker, cuyas densidades espectrales de potencia vienen dadas por las ecuaciones 5.13, 5.15, y 5.18 respectivamente:

$$S_{I_i}(f) = \frac{4kT}{R}, \quad \forall f \quad (5.78)$$

$$S_{I_s}(f) = 2q \cdot I_s, \quad \forall f \quad (5.79)$$

$$S_{I_f}(f) = K \cdot (I_f)^c \cdot \frac{1}{|f|^n} \quad (5.80)$$

donde I_s es la corriente que determina el ruido de disparo, e I_f la que determina el ruido flicker. La expresión del ruido flicker resulta válida para el rango de frecuencias de operación normal en los sensores de imagen.

Atendiendo a las ecuaciones 5.78, 5.79, y 5.80 podemos destacar tres factores importantes que determinan la potencia de ruido total en el dispositivo:

- La corriente que atraviesa el dispositivo;
- Las constantes que aparecen en las expresiones, las cuales vienen determinadas por las características tecnológicas del material sobre el que se implementa el dispositivo, así como de las propiedades geométricas del mismo;
- El rango de frecuencias en el que trabaje el dispositivo, que determinará la potencia total de ruido en el dispositivo.

Respecto a las **corrientes**, como medida a tomar para la reducción de la potencia de ruido, parece lógico a priori minimizarlas. Si particularizamos las psds para el caso de un fotodiodo, ocurre que:

$$I_s = I_{ph} + I_{dark}$$

$$I_f = I_{dark}$$

Minimizar la corriente fotogenerada I_{ph} carece de sentido, puesto que como se ha comentado en anteriores secciones, la relación señal a ruido, que es el parámetro que realmente determina la calidad del sensor, se degrada con la reducción de esta componente de corriente. La única medida que parece lógico tomar es la reducción de la corriente oscura I_{dark} . Esta corriente depende de los parámetros tecnológicos del dispositivo, en particular de las concentraciones de dopado de las regiones p y n con las que se implementa el fotodiodo. Como medida a tomar para la minimización de la corriente oscura, se puede optar por la elección de la unión pn dentro de las disponibles en la tecnología cuyos niveles de dopado favorezcan la reducción de esta corriente.

Los coeficientes R y K , y los exponentes c y n depende también de los **parámetros tecnológicos** del dispositivo implementado, por lo que cabe una discusión similar a la llevada respecto a la corriente oscura.

Respecto al rango frecuencial, una medida de carácter general a tomar es la limitación en banda del sistema mediante el uso de filtros paso de banda. Con esta medida, la potencia de ruido total será resultado de integrar la psd total sólo en el ancho de banda considerado. Esto favorece de forma particular a la reducción del ruido flicker, cuya principal contribución se

encuentra en las frecuencias cercanas al cero. En posteriores secciones se describe el algoritmo CDS (Correlated Double Sampling) para la eliminación de la componente de ruido de patrón fijo. Este algoritmo implementa un filtrado paso de alta que favorecerá de forma indirecta a la reducción del ruido a baja frecuencia.

En resumen, las psds de los ruidos intrínsecos de un dispositivo están fuertemente ligadas a los parámetros tecnológicos del mismo, por lo que no existen medidas de carácter general que se puedan tomar para la minimización de estas potencias.

Ocurre que, en general, en los sensores de imagen existen componentes de ruido más importantes, como son el ruido generado durante la fase de reset¹⁰ y el ruido de patrón fijo, por lo que el diseño del circuito suele centrarse en la minimización de estas componentes. Del ruido generado durante la fase de reset nos ocupamos en la siguiente sección, mientras que el ruido de patrón fijo se trata un poco más adelante, al ser éste una componente de ruido espacial, y no temporal.

5.5.2.- Reducción del ruido de reset en los sensores APS CMOS por integración de carga

Presentamos a continuación diversas técnicas para la reducción de reset que pretenden minimizar la cota de este ruido sin generar una imagen residual, principal inconveniente del esquema de reset propuesto en la Secc. 5.4.2.1.

5.5.2.1.- Técnica “pseudo-flash”

Para conseguir que la cota de reset se mantenga en $kT/2C_{ph}$ sin que se genere imagen residual en [4] se propone el esquema de reset ilustrado en la Fig. 5.21, denominado *reset pseudo-flash*. En este esquema el drenador del transistor de reset se conecta a la señal *Reset_Dr*. Como se muestra en el diagrama de tiempos (Fig. 5.21(b)), la señal *Reset_Dr* cae a una tensión baja, por ejemplo a tierra, al comienzo de la fase de reset, con lo que se producirá una descarga rápida del residuo de carga que hubiese quedado en la capacidad de integración al final de la fase de integración. Al mismo tiempo, el ruido de reset no se incrementa al realizarse el reset a través de un transistor nMOS, por lo que la cota se mantiene en $kT/2C_{ph}$.

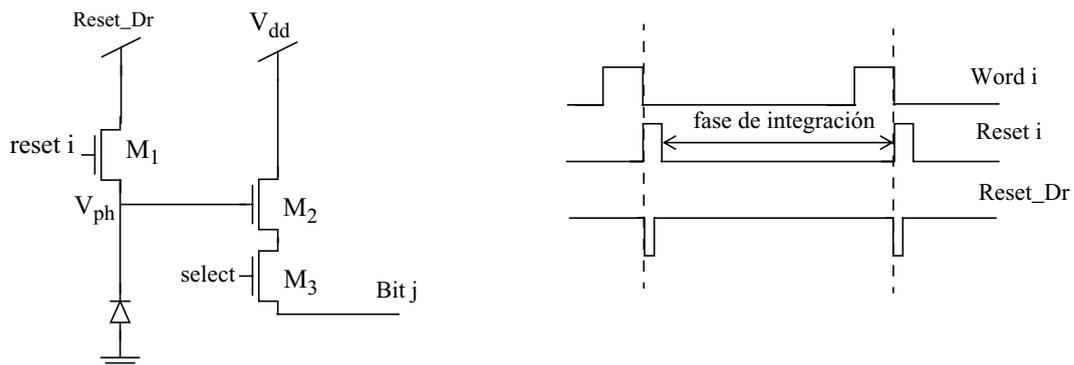


Fig. 5.21: (a) Esquema de APS en tiempo discreto y (b) diagrama de tiempos para la reducción del ruido de reset

5.5.2.2.- Reset a través de una llave pMOS y una carga logarítmica

La Fig. 5.22 [16] muestra el esquema de la circuitería a nivel de píxel y a nivel de columna propuesta para conseguir la reducción del ruido de reset sin generar una imagen residual. El

10. Este ruido aparece en los sensores APS por integración de carga

principio de operación es bastante parecido al reset pseudo-flash. Inicialmente, la señal HTS está a nivel bajo, por lo que M_{short} conecta los drenadores de los transistores del píxel M_{SF} y M_{rst} a V_{dd} . Cuando HTS pasa a nivel alto, M_{short} se corta y M_{drop} conduce corriente, permitiendo que la tensión del drenador de M_{reset} caiga una tensión V_T por debajo de V_{dd} . La referencia inicial de tensión en el nodo V_{ph} al comienzo de la fase de reset será $V_{dd}-V_T$, sea cual sea la tensión en la capacidad de integración al final de la fase de integración. Cuando HTS vuelve a nivel bajo, el píxel se resetea según el comportamiento descrito en la Secc. 5.4.2.1. Por lo tanto, puesto que el potencial inicial es siempre el mismo, sea cual sea la tensión al comienzo de la fase de reset, el nivel de reset alcanza el mismo potencial, eliminando la imagen residual que se genera con el esquema básico.

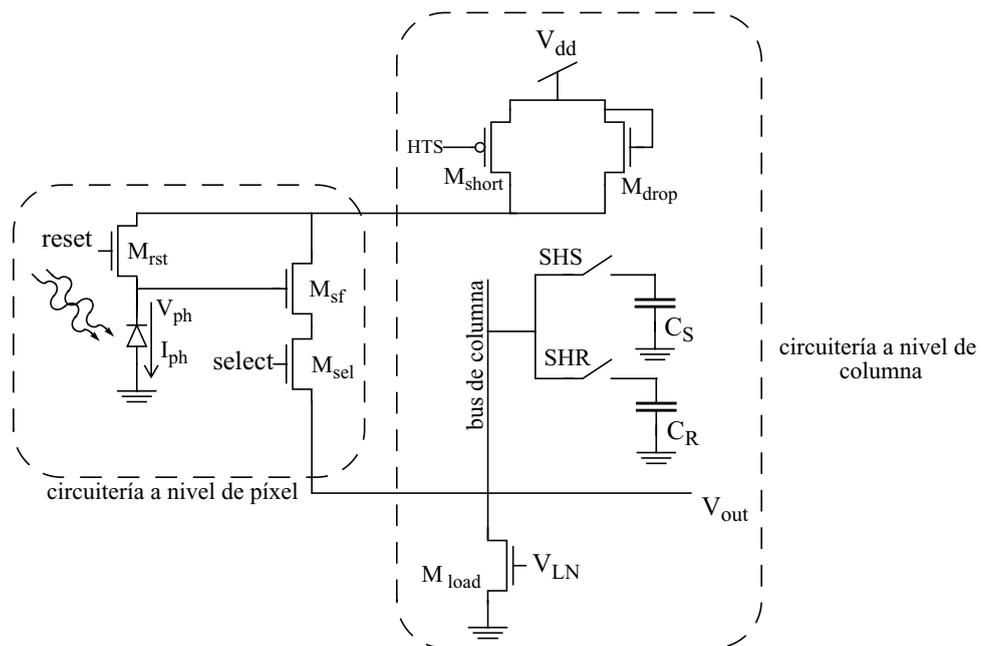


Fig. 5.22: Esquema de APS en tiempo discreto para la reducción del ruido de reset propuesta en [16]

5.5.2.3.- Técnica de reset activo

La técnica de reset activo reduce el ruido de reset sin añadir carga residual. Esta técnica usa limitación en banda y capacidades de realimentación para reducir el ruido. Veamos cómo opera el píxel con reset activo.

La forma general de un píxel APS con reset activo se muestra en la Fig. 5.23 [17]. Consiste en dos circuitos independientes, un circuito de lectura y un circuito de reset. El circuito de lectura consiste en dos transistores NMOS M_2 y M_3 , y el circuito de reset consiste en tres transistores NMOS, M_1 , M_5 y M_4 , y un amplificador. Las formas de onda del circuito de reset se muestran en la Fig. 5.24.

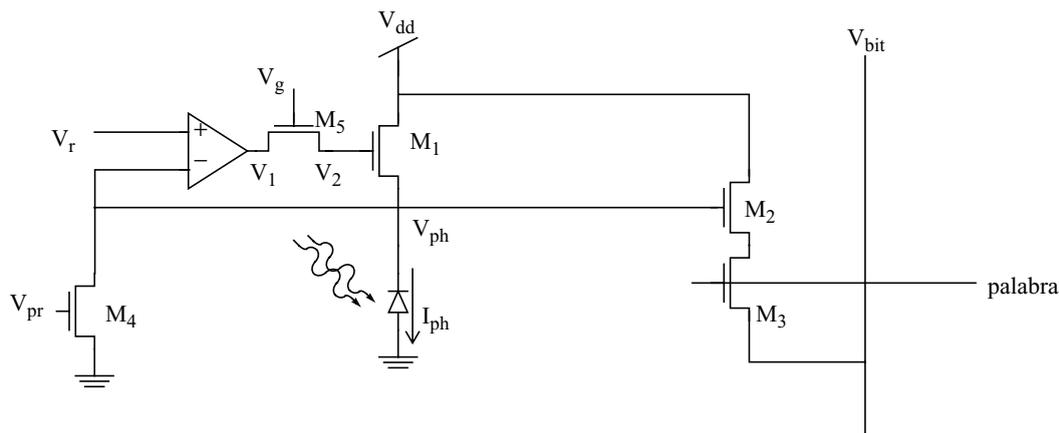


Fig. 5.23: APS CMOS con reset activo

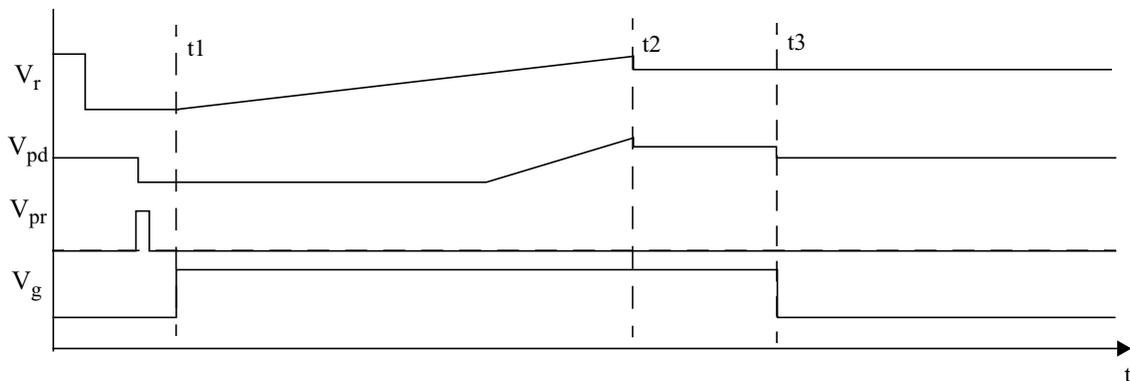


Fig. 5.24: Diagrama de tiempos del algoritmo de reset activo

La descripción del funcionamiento de la circuitería de reset es la siguiente. Justo antes del instante t_1 la señal V_{pr} se activa durante un corto intervalo de tiempo, fijando la tensión V_{ph} a tierra. En el instante t_1 la señal V_g se coloca a V_{dd} , con lo que el transistor M_5 se enciende. En el mismo instante, la señal de control V_r comienza a crecer lentamente, partiendo de la tensión cero, hasta alcanzar el valor $\max(V_r)$ en el instante t_2 . En el momento en el que la tensión V_r excede a la tensión V_{ph} , la salida del amplificador se dispara y enciende M_1 . En el momento en el que M_1 se enciende, la tensión V_{ph} seguirá a la tensión V_r hasta que ésta deje de crecer en t_2 . En t_2 la señal de control V_r disminuye bruscamente unos cuantos milivoltios. Esto provocará que la salida del amplificador baje bruscamente y corte al transistor M_1 . La tensión V_{ph} disminuirá una pequeña cantidad debido a la inyección de los portadores del canal en la capacidad de integración, pero esta disminución será menor que el salto de V_r , por lo que la salida del amplificador mantendrá M_1 cortado. En el instante t_3 la señal de control V_g pasa a nivel bajo, cortando al transistor M_5 . La tensión V_{ph} disminuirá una cierta cantidad, que dependerá únicamente de la capacidad de solapamiento entre los nodos de puerta y fuente del transistor M_1 .

En [17] se presenta un análisis completo del circuito, en el que se incluyen las expresiones para la carga residual, así los modelos de pequeña señal del circuito en cada fase de operación. En el análisis del ruido de reset llevado a cabo en este estudio, mediante la apropiada selección de la ganancia del amplificador de reset, g_m/g_o , de la capacidad de solape de M_1 , C_f , y de las capacidades que limitan de la banda, C_{l1} y C_{l2} , el ruido de píxel puede hacerse mucho menor que kT/C . Esta reducción del ruido de reset se consigue cortando M_1 , limitando en banda el ampli-

ficador de reset, y controlando el lazo del reset a través del divisor capacitivo. Eligiendo los parámetros de forma adecuada se consigue una reducción del ruido de reset equivalente a $kT/(18C)$.

5.6.- RUIDO ESPACIAL EN LOS SENSORES APS CMOS EN TIEMPO DISCRETO

Como se ha comentado en la Secc. 5.1.2, el ruido espacial tiene como origen las variaciones en los parámetros tecnológicos entre dispositivos, introducidas de forma inevitable por el proceso de fabricación. Para modelar estas variaciones se expresa la relación entre el valor del píxel de la imagen final capturada ($V_{i,j}$) y la intensidad luminosa incidente (L) como (ecuación 5.2):

$$V_{i,j}(L) = (\alpha_0 + \Delta\alpha_0|_{i,j}) + (\alpha_1 + \Delta\alpha_1|_{i,j})L + (\alpha_2 + \Delta\alpha_2|_{i,j})L^2 + \dots \quad (5.81)$$

Pretendemos ver cómo dependen los parámetros que aparecen en esta ecuación con la arquitectura del sensor, así como un método para caracterizar las desviaciones típicas de los parámetros en función de las características de los dispositivos y la distancia entre ellos. Por simplicidad, y puesto que el desarrollo es extrapolable para los demás términos, vamos a ocuparnos del ruido de patrón fijo, que coincide con el término de orden cero de la ecuación 5.81.

Para un nivel de intensidad luminoso nulo, el valor del píxel en la imagen final capturada tiene una expresión:

$$V_{i,j}(L=0) = \alpha_0 + \Delta\alpha_{0,i,j} \quad (5.82)$$

donde α_0 es el valor nominal del valor de salida para $L=0$, y $\Delta\alpha_{0,i,j}$ una variable estadística que modela la desviación típica que define al ruido de patrón fijo.

El parámetro $\Delta\alpha_{0,i,j}$ puede descomponerse de forma que exprese la dependencia con la jerarquía de la arquitectura. Haciendo uso de esta jerarquización, podemos escribir:

$$\Delta\alpha_{0,i,j} = \Delta\alpha_{0,i,j}|_{pixel} + \Delta\alpha_{0,i}|_{columna} + \Delta\alpha_0|_{sistema} \quad (5.83)$$

$\Delta\alpha_{0,i,j}|_{pixel}$ es la **componente de FPN a nivel de píxel**, es decir, caracteriza la desviación típica de la salida del píxel debido a las diferencias paramétricas entre píxeles, es decir, introducida por las diferencias paramétricas entre las circuiterías a nivel de píxel; $\Delta\alpha_{0,i}|_{columna}$ es

la **componente de FPN a nivel de columna**, y se genera por las desviaciones de los parámetros entre la circuitería a nivel de columna, por lo que, puesto que todos los píxeles de una misma columna comparten la circuitería a nivel de columna, no dependerá del subíndice j ; y $\Delta\alpha_0|_{sistema}$ será la **componente de FPN a nivel de sistema**, es decir, caracteriza la desviación típica respecto del valor nominal que introduce la circuitería a nivel de sistema, y que por tanto será igual para todos los píxeles del sensor.

La principal contribución al FPN a cada nivel está determinada por las diferencias entre las ganancias de los amplificadores. Por ejemplo, en los sensores APS, el hecho de que incorporen un amplificador a nivel de píxel implica que la componente de FPN a nivel de píxel sea considerable (problema que no presentarán los sensores PPS).

La componente de FPN a nivel de píxel tiene una distribución aleatoria espacialmente uniforme a lo largo de toda la matriz. Si consideramos una matriz de 10x10 píxeles en la que sólo

consideramos esta componente, la salida del píxel a excitación luminosa nula, se verá distorsionada de la forma que ilustra la Fig. 5.25(a). La distorsión de la imagen que introduce la componente de FPN a nivel de columna, como se ilustra en la Fig. 5.25(b), tendrá una distribución por columnas, pues todos los píxeles de una misma columna compartirán la misma desviación; mientras que la Fig. 5.25(c) ilustra la variación de la imagen ante la componente de FPN a nivel de sistema.

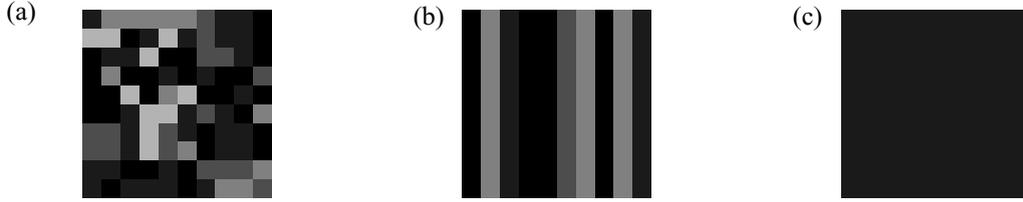


Fig. 5.25: Representación del ruido espacial introducido sobre una imagen oscura ($L=0$) por las componentes de FPN (a) a nivel de píxel; (b) a nivel de columna; y (c) a nivel de sistema

Atendiendo a la ecuación 5.83, se observa la importancia del nivel jerárquico en el cual llevar a cabo la corrección del ruido de patrón fijo. Tendremos tres posibilidades:

- Corregir el FPN a nivel de sistema, es decir, introducir la circuitería de corrección del FPN de forma que sea compartida por todos los píxeles. Con esta opción la lectura diferencial propuesta en la Secc. 5.1.2 se lleva a cabo a nivel de sistema, por lo que sólo estaremos eliminando la componente de FPN a nivel de sistema. El FPN vendrá caracterizado por una desviación típica:

$$\Delta\alpha_{0,i,j}|_{diff} = \Delta\alpha_{0,i,j} - \Delta\alpha_0|_{sistema} = \Delta\alpha_{0,i,j}|_{pixel} + \Delta\alpha_{0,i}|_{columna} \quad (5.84)$$

La distorsión de la imagen final será suma de las distorsiones ilustradas en las Fig. 5.25(a) y (b).

- Corregir el FPN a nivel de columna, es decir, introducir la circuitería de corrección del FPN a nivel de columna. En este caso el FPN estará generado por las desviaciones paramétricas entre las circuiterías a nivel de columna:

$$\Delta\alpha_{0,i,j}|_{diff} = \Delta\alpha_{0,i,j} - \left(\Delta\alpha_{0,i,j}|_{pixel} + \Delta\alpha_0|_{columna} + \Delta\alpha_0|_{sistema} \right) = \Delta\alpha_{0,i}|_{columna} \quad (5.85)$$

La distorsión de la señal tendrá la forma ilustrada en la Fig. 5.25(b).

- Corregir el FPN a nivel de píxel, en cuyo caso se elimina por completo el FPN:

$$\Delta\alpha_{0,i,j}|_{diff} = \Delta\alpha_{0,i,j} - \left(\Delta\alpha_{0,i,j}|_{pixel} + \Delta\alpha_0|_{columna} + \Delta\alpha_0|_{sistema} \right) = 0 \quad (5.86)$$

La dispersión de los parámetros que definen a un componente electrónico se conoce como desapareamiento (“*mismatch*”), y los efectos de esta dispersión se conocen como efectos de desapareamiento. Un modelo bastante extendido para el modelado de los efectos de desapareamiento es el descrito por Pelgrom *et al.* en [18]. De acuerdo con este trabajo, las propiedades de apareamiento para dos transistores MOS diseñados de idéntica forma pueden describirse en términos de tres variables aleatorias estadísticamente independientes y de media

cero. Estas variables aleatorias definen la desviación típica de tres parámetros del modelo de gran señal de primer orden del transistor MOS. En particular, dada la expresión de la corriente de un transistor MOS en inversión fuerte, operando en la región de saturación:

$$I = \beta(V_{GS} - V_T)^2(1 - \lambda V_{DS})$$

$$V_T = V_{T_o} + \gamma(\sqrt{V_{SB} + \phi_B} - \sqrt{\phi_B})$$
(5.87)

los parámetros β' , V'_{T_o} y γ' para un segundo transistor idéntico al primero se modelan como:

$$\beta' = \beta + \Delta\beta$$

$$V'_{T_o} = V_{T_o} + \Delta V_{T_o}$$
(5.88)

$$\gamma' = \gamma + \Delta\gamma$$

donde $\Delta\beta$, ΔV_{T_o} , $\Delta\gamma$ son tres variables estadísticas de media cero, y cuyas desviaciones típicas dependen del área de canal del dispositivo y de la distancia D entre los dos dispositivos de la siguiente manera:

$$\sigma^2(\Delta V_{T_o}) = \frac{A_{V_{T_o}}^2}{WL} + S_{V_{T_o}}^2 D$$

$$\sigma^2(\Delta\gamma) = \frac{A_{\gamma}^2}{WL} + S_{\gamma}^2 D$$

$$\sigma^2\left(\frac{\Delta\beta}{\beta}\right) = \frac{A_{\beta}^2}{WL} + S_{\beta}^2 D$$
(5.89)

donde β viene dada por:

$$\beta = \frac{\mu_o C'_{ox} W}{2L}$$

W y L son el ancho y largo del transistor respectivamente, y los coeficientes A_i y S_i (donde por i se entiende a cada una de las variables) son coeficientes de proporcionalidad. Como puede observarse, la varianza de cada parámetro contiene dos términos. El primero de ellos, inversamente proporcional al área de canal del dispositivo, modela el fenómeno aleatorio conocido como correlación de corta distancia (short-correlation-distance) en el proceso de fabricación (ruido espacial blanco); mientras que el segundo de los términos da muestras de variaciones no aleatorias de los parámetros en el proceso de fabricación en la oblea semiconductora, normalmente en distribución circular en torno al centro de la oblea. Puesto que la localización y orientación de la oblea es en general desconocida para el diseñador, esa variación puede modelarse como un proceso estotástico, como se hace en las ecuaciones 5.89. Excepto para dispositivos de área muy grande separados una larga distancia, el término dependiente del área es claramente dominante, y por lo tanto:

$$\sigma(i) = \sqrt{\frac{A_i^2}{WL} + S_i^2 D} \approx \frac{A_i}{\sqrt{WL}}(1 + \epsilon_i D^2 WL)$$
(5.90)

donde i es cualquiera de las seis variables: $\Delta\beta_n/\beta_n$, $\Delta\beta_p/\beta_p$, $\Delta V_{T_{on}}$, $\Delta V_{T_{op}}$, $\Delta\gamma_n$, $\Delta\gamma_p$, y:

$$\epsilon_i = \frac{1}{2} \frac{S_i^2}{A_i^2}$$

Estas expresiones de la desviación típica de los parámetros, junto con la estructura del ruido de patrón fijo en función de la arquitectura, conforman una base para el modelado y estimación del ruido de patrón fijo en los sensores de imagen CMOS.

Como puede derivarse de la ecuación 5.90, el ruido de patrón fijo será especialmente crítico en la circuitería a nivel de píxel, pues a este nivel la superficie disponible obliga a que los transistores sean de dimensiones mínimas, por lo que la varianza de los parámetros según el modelo de Pelgrom se hará máxima.

En [19] se desarrolla un modelo estadístico para la estimación del FPN en los sensores de imagen, tanto para un sensor PPS como para un APS en tiempo discreto por integración de carga.

5.7.- TÉCNICAS DE REDUCCIÓN DEL RUIDO ESPACIAL

El análisis del ruido espacial, como se comentó en la Secc. 5.1.2, se centra básicamente en el estudio del ruido de patrón fijo, modelando los efectos del PRNU dentro de las no linealidades de la ganancia del sistema. De igual manera, los algoritmos desarrollados para la reducción del ruido espacial se centran en la reducción del FPN. De estos algoritmos nos encargamos a continuación.

5.7.1.- ALGORITMOS PARA LA REDUCCIÓN DEL FPN

5.7.1.1.- Muestreo diferencial a nivel de píxel

El esquema más común para la eliminación del FPN consiste en llevar a cabo un muestreo diferencial del valor de salida del dispositivo fotosensor. La Fig. 5.26 muestra el esquema de muestreo diferencial a nivel de píxel. Consiste en llevar a cabo una doble lectura del valor de salida del dispositivo fotosensor: una primera lectura del fotosensor para almacenar el valor de tensión de C_{ph} al final de la fase de integración en la capacidad C_s ; y una segunda lectura para almacenar en la capacidad C_r la tensión en el nodo de sensado tras la fase de reset. Finalmente se lleva a cabo la lectura diferencial de los dos valores almacenados. Puesto que los dos valores son leídos por la misma circuitería, cualquier desapareamiento y/o variación en el proceso en la circuitería se cancela tras la lectura diferencial, por lo que el FPN se reduce bastante.

La técnica de llevar a cabo un muestreo diferencial en el que se restan los valores de la capacidad de integración al final de la fase de integración y al final de la fase de reset se conoce como **correlación por doble muestreo** o **CDS** (Correlated Double Sampling), y es una de las técnicas más extendidas de reducción del ruido de patrón fijo.

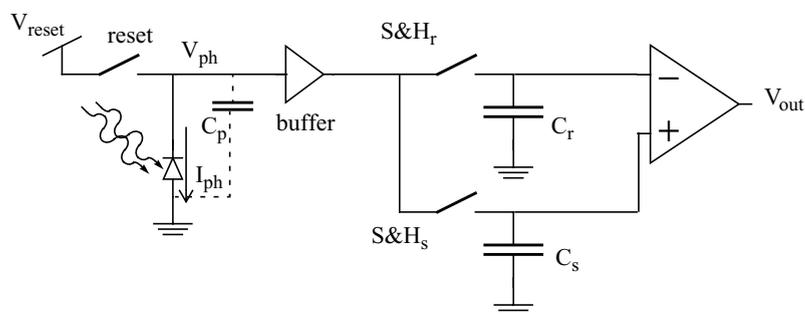


Fig. 5.26: Sensor con salida diferencial directa implementada a nivel de píxel

Un ejemplo de implementación del esquema y modo de operación del píxel con salida diferencial queda ilustrado en la Fig. 5.27 [20]. El elemento fotodetector se implementa en este ejemplo con un diodo difusión n-substrato p.

El circuito de muestreo y retención consiste en un interruptor de muestreo y un nodo de retención (nodo capacitivo). Los interruptores (M_2 y M_3) actúan como obturadores electrónicos para controlar el tiempo de exposición del píxel. Los nodos de retención serán las puertas de los transistores en configuración de seguidor por fuente (M_4 y M_6). Puesto que la corriente residual asociada a la corriente inversa de la unión pn del drenador de los obturadores puede destruir la señal si la capacidad en la puerta del seguidor de tensión es demasiado pequeña, se añaden dos capacidades MOS (M_8 y M_{10}) para incrementar la capacidad en estos nodos.

El interruptor de reset M_1 y los dos obturadores M_2 y M_3 son dispositivos pmos, mientras que los otros transistores de la celda son dispositivos nmos. Normalmente todos los transistores de una misma celda tienen la misma polaridad para lograr un buen espaciado entre los pozos en el layout del sensor. Sin embargo, si el transistor de reset fuese nmos, la tensión de reset del fotodiodo estaría a una tensión umbral por debajo de la tensión de alimentación ($V_s = V_{dd} - V_{th}$). La tensión umbral de los dispositivos nmos, incluido el efecto substrato, oscila entre 0.6 y 1 V (el rango de variación depende también en gran medida de la tecnología utilizada). Esta caída de tensión será una parte significativa del rango de variación de tensiones si la tensión de alimentación es de 3.3 V o menor. Por lo tanto, el uso de dispositivos pmos como transistor de reset y como obturadores para evitar esta caída de tensión resulta conveniente. De esta forma el rango de variación de tensiones se maximiza a cambio de un incremento en la superficie de implementación del píxel.

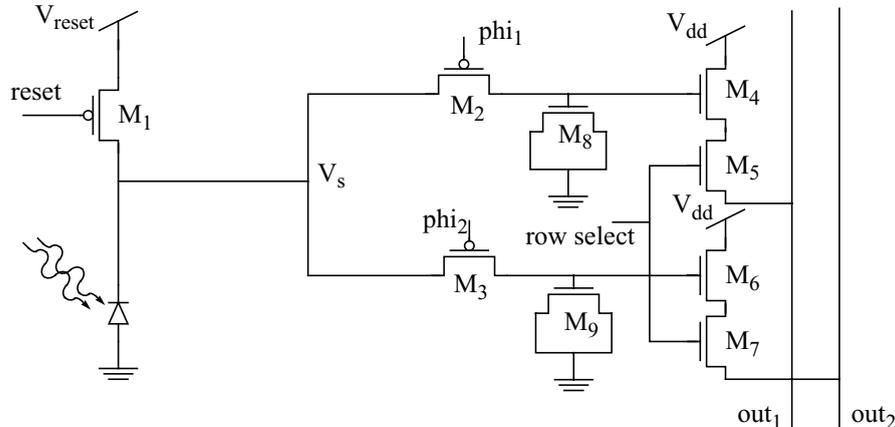


Fig. 5.27: Esquema de píxel con salida diferencial directa

5.7.1.2.- Reducción del FPN en los APS en modo corriente

En la Secc. 4.3.5 se ha descrito el sensor APS en modo corriente. El esquema de un píxel en modo corriente se ilustra en la Fig. 5.28. La forma de implementar la técnica CDS en los sensores en modo corriente consiste en llevar a cabo la operación:

$$I_{diff} = I_{rst} - I_{sig} \quad (5.91)$$

Con esto conseguimos reducir el error de offset, y se consigue reducir en gran medida el ruido espacial para imágenes oscuras. Sin embargo, los sensores en modo corriente resultan especialmente sensibles a las desviaciones de los parámetros tecnológicos de los transistores del píxel,

por lo que la técnica CDS no se implementa correctamente para niveles de iluminación altos (corrientes de salida pequeñas).

Es decir, para esta configuración la implementación de la técnica CDS funciona bien para imágenes oscuras, mientras que para señales de iluminación alta las diferencias paramétricas (β , V_T ,...) de los transistores provocan un error importante (PRNU elevado) para imágenes brillantes.

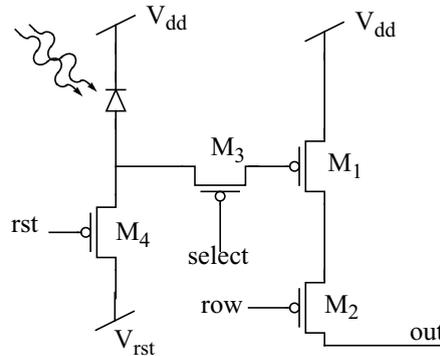


Fig. 5.28: Esquema de píxel en modo corriente

El cociente de las corrientes diferencial y de reset se presenta como una evaluación de la intensidad de luz mucho más independiente de los parámetros de los transistores de la fotocélula [20]. El cálculo de este cociente es sencillo si se tiene en cuenta que cualquier convertidor A/D implementa la conversión mediante la evaluación del cociente entre la entrada a medir y una escala de referencia. Por lo tanto, escalando la escala de referencia de acuerdo con I_{rst} para cada píxel se obtiene una salida digital que tiene corregidos los errores de desapareamiento de los transistores:

$$OUT_{digital} = (2^n - 1) \frac{I_{meas}}{I_{fsr}} = (2^n - 1) \frac{I_{rst} - I_{sig}}{I_{rst}} \quad (5.92)$$

donde n es el número de bits de la salida, I_{meas} es la entrada a medir por el convertidor, e I_{fsr} es la escala de referencia del convertidor A/D.

5.7.1.3.- Corrección del FPN en los sensores de imagen logarítmicos

Los fotorreceptores logarítmicos muestran un FPN muy alto debido a los efectos del desapareamiento, en especial de la tensión umbral, cuando los transistores trabajan en inversión débil. Las variaciones de offset son muy altas, llegando a alcanzar variaciones equivalentes a 3 décadas en la intensidad de luz medida. Desafortunadamente, cualquier estructura para la corrección de este efecto es mucho más complicada en los sensores logarítmicos que en los sensores de integración de la carga. La razón es la inexistencia de un estado de reset en los sistemas que trabajan en tiempo continuo, en los que no existe un estado inicial de reposo. Los sensores de integración aprovechan este estado de reposo para muestrear la imagen residual que queda en el sensor para proceder a su posterior eliminación en la imagen final.

La forma más común de corregir las no uniformidades en los sensores de imagen logarítmicos es el implementar un método de corrección digital. Inicialmente, los errores de los píxeles son medidos mediante la iluminación del sensor de forma uniforme, y se almacenan los resultados en una memoria digital. Durante la operación de lectura, las señales leídas por los píxeles se convierten en valores digitales y posteriormente son corregidas de acuerdo con el valor del error

almacenado. Para este proceso se podrán usar diferentes algoritmos: algoritmos de un solo punto (sólo el offset); algoritmos de dos puntos (offset y pendiente); o bien algoritmos de mayor orden. La corrección digital del FPN se implementa normalmente fuera del chip de sensado, aunque en principio no existe ningún impedimento para implementarlo en el mismo integrado. Sin embargo, una solución que lo implementase todo en un solo chip requeriría un área extra para la memoria digital y una lógica de control adicional, lo que conduciría a chips mayores. Por lo tanto, en el caso de una corrección digital del FPN es preferible la implementación de la memoria y la lógica adicional fuera del sensor, lo que directamente provoca un aumento del coste del sistema.

En [21] se propone la calibración mediante la estimulación del receptor con una corriente muy alta añadida a la fotocorriente. En este caso, la salida del píxel es prácticamente independiente de la iluminación real, pues la variación de la fotocorriente sólo sería una pequeña fracción de la corriente total. En esta situación, la salida del fotorreceptor que corresponde a esta referencia puede substraherse de la señal de salida correspondiente a la iluminación real mediante el uso de la técnica de muestreo por doble correlación (CDS). Este método de calibración permite la implementación de sensores de pequeño tamaño, pues sólo es necesario para su implementación un transistor adicional. Sin embargo, tras esta corrección el FPN restante es bastante alto. Esto se debe principalmente a la no uniformidad que producen las variaciones en la pendiente. Dado que la intensidad de calibración es mucho más alta que la fotocorriente, el punto de calibración está muy lejos del punto de operación del sensor. Por otro lado, la frecuencia de lectura está limitada, pues tras la calibración el píxel necesita un tiempo para alcanzar su nivel original.

En [22] se propone un método de reducción del FPN que se basa en este sistema de calibración, pero usa corrientes de calibración muy bajas, del mismo orden de magnitud que las fotocorrientes. Por lo tanto, las variaciones en la pendiente tienen sólo un efecto secundario sobre el FPN, y por lo tanto, las variaciones del offset son reducidas en mayor medida.

5.5.7.1.4.- Arquitectura para la compensación del FPN en sensores logarítmicos

La Fig. 5.29(a) muestra el elemento clásico de sensado logarítmico, que usa un transistor MOS conectado en modo diodo y un fotodiodo de difusión. La fotocorriente en el fotodiodo se convierte en tensión V_{ph} , que será proporcional a la respuesta logarítmica de la corriente. La componente que domina el FPN es la variación en la tensión umbral tanto del transistor de conversión M_1 como del transistor seguidor por fuente de la salida M_2 . Para poder medir esta componente de FPN, el píxel debe colocarse en condiciones de oscuridad total.

La Fig. 5.29(b) [23] muestra la nueva estructura, en la que se obtiene una referencia oscura precisa in-situ (valor de la salida del píxel en ausencia de iluminación). En este esquema el fotodiodo opera en modo fotovoltaico. En este modo de operación el fotodiodo proporciona una tensión en circuito abierto proporcional a la intensidad de luz incidente. La tensión en circuito abierto en el fotodiodo nos da una conversión logarítmica de forma natural, que puede estimarse como:

$$V_{ph} = -\phi_t \ln\left(\frac{I_s + I_{ph}}{I_s}\right) \quad (5.93)$$

De esta ecuación queda claro que la tensión en circuito abierto será cero para ausencia de iluminación, e irá hacia tensiones negativas con la iluminación. La novedad introducida es el añadir una nueva llave M_3 . Cuando V_z está a nivel alto, M_3 cortocircuitará el fotodiodo a tierra y

fijará V_{ph} a tierra. Esto da una referencia precisa de la respuesta del píxel de ausencia de iluminación.

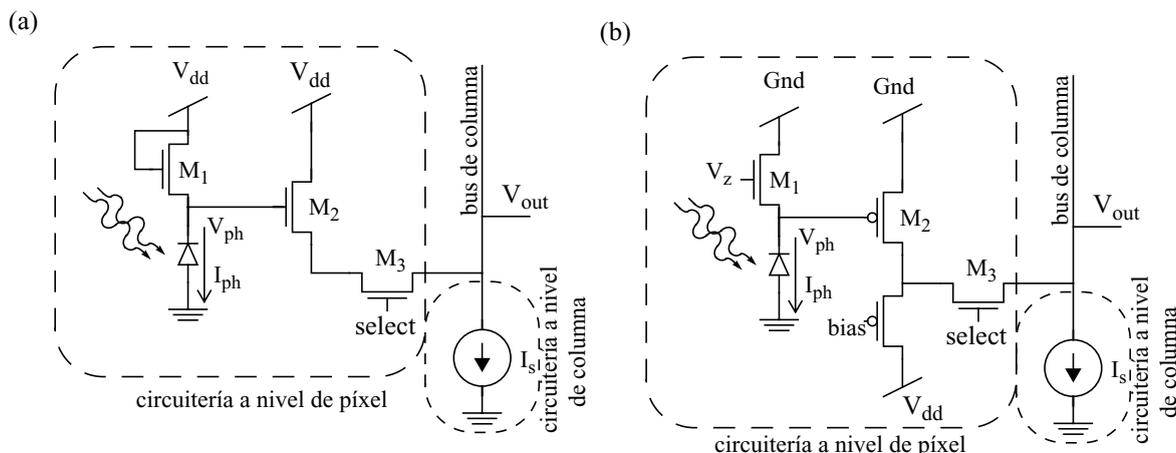


Fig. 5.29: Esquema de APS en tiempo continuo (a) típico; y (b) con esquema de reducción del FPN

Con este mecanismo de funcionamiento, el FPN introducido por la circuitería a nivel de píxel puede compensarse. Las fases de operación para llevar a cabo esta compensación son las siguientes: 1º: lectura del píxel fijando la señal de control $V_z = '0'$; 2º: lectura del píxel con $V_z = '1'$; 3º: lectura diferencial de ambas salidas para obtener una salida libre de FPN.

Con este método, el fotodiodo se resetea cada vez que es leído. La operación de reset tendrá dos consecuencias: 1º eliminación de la carga residual debida a la conversión logarítmica; y 2º respuesta mixta lineal y logarítmica a la salida del píxel. La segunda consecuencia deriva de lo siguiente: tras el reset del fotodiodo, su tensión evolucionará hacia valores negativos con la iluminación y finalmente se establecerá en el valor dado por la ecuación 5.93. Pero este proceso toma su tiempo, especialmente cuando el nivel de iluminación es bajo. Si el píxel se lee antes de que se establezca este valor, su respuesta no será logarítmica, sino lineal frente al nivel de iluminación.

La única dificultad aquí es que V_s es una tensión negativa, por lo que deben aplicarse el diseño particular de algunos circuitos y/o modificaciones en el proceso. Se han usado transistores pmos para la lectura de este valor negativo sin ninguna modificación en el proceso. Pero este método implica que el tamaño del píxel sea grande debido al espaciado entre pozos requerido por el proceso de fabricación de los integrados. Si es posible una tensión umbral negativa en los transistores nmos mediante la modificación del proceso de fabricación, podrían obtenerse sensores de imagen mucho más compactos (ceranos a los tamaños de los sensores de 3 transistores tradicionales).

5.8.- RUIDO MULTIPLICATIVO EN LOS SENSORES DE IMAGEN

Uno de los problemas de los sensores de imagen es el ruido generado por las fuentes de luz artificial que existen en el ambiente, principalmente a frecuencias de gran interés (50Hz o 60Hz). En los detectores convencionales, tales como los CCDs, la corriente fotogenerada se integra en la capacidad de entrada del fotodetector y luego se muestrean. Aunque esto no reduzca del todo el ruido, el efecto no es visible para el ojo humano debido a los efectos residuales de los sensores de imagen en la retina (*tiempo de retención de la retina*). La integración y el muestreo, sin em-

bargo, no pueden realizarse si la imagen necesita ser procesada en tiempo continuo, o si la frecuencia de salida necesaria es mayor que la frecuencia de muestreo del sensor.

El **ruido multiplicativo** aparece como un factor multiplicativo en las fotocorrientes detectadas. La razón es que la luz reflejada en la superficie de los objetos es proporcional a la intensidad de la fuente de luz y a la superficie refractante para superficies lambertianas¹¹ sencillas. Este ruido multiplicativo llega a ser muy severo en los chips de detección de movimiento, en los que este ruido se amplifica en las etapas diferenciales temporales, requeridas para la detección de movimiento. El uso de un filtro rechazo de banda para la eliminación la frecuencia prohibida, además de ser de difícil implementación en los circuitos VLSI, también elimina parte de la señal. Por lo tanto es necesario el uso de otros métodos.

Hay dos factores que sugieren otra solución a este problema. En primer lugar, este ruido es común a un conjunto de píxeles vecinos. En segundo lugar, es un ruido multiplicativo. Una solución sencilla sería dividir la señal obtenida de un detector entre la media ponderada de varios fotodetectores.

La **cancelación del ruido multiplicativo (MNC: Multiplicative Noise Cancellation)** es esencialmente imposible. En los ruidos aditivos, las características frecuenciales del ruido y de la señal se suman, y pueden ser separadas si la señal y el ruido están separados en frecuencia. Pero en el ruido multiplicativo las características frecuenciales de la señal ruidosa será una convolución de las respuestas frecuenciales de la señal y del ruido. Por lo tanto, resulta impracticable separar ruido y señal en el dominio frecuencial. Sin embargo, si el ruido tiene una sola componente frecuencial, la respuesta frecuencial de la señal ruidosa será una versión desplazada de la respuesta frecuencial de la señal sin ruido.

Encontramos la respuesta para la reducción del ruido multiplicativo en consideraciones en el dominio del tiempo de la señal y del ruido [24]. Hay dos métodos para cancelar el ruido multiplicativo (es decir, la componente de a ruido a 50 Hz). El primer método de cancelación es el muestreo sincrónico, que consiste en muestrear la señal ruidosa con la misma frecuencia que la frecuencia de la componente de ruido. La señal muestreada tendrá un factor de escala constante relacionado con el ruido, si la amplitud del ruido es constante. Este método, sin embargo, limita el ancho de banda de la señal de entrada a 25 Hz, que no resulta ser suficiente para aplicaciones en las que la imagen debe ser muestreada y procesada a frecuencias mucho mayores. El segundo método se basa en el procesamiento de señales de forma colectiva (en contra de un procesamiento de canales de señal) en una matriz de sensores. Se asume que el ruido multiplicativo es común a las células vecinas en una matriz de elementos de procesamiento analógicos. Por lo tanto, si la señal de un canal se divide entre el promedio de señales de una zona (promediado sobre el tamaño de una ventana), el ruido será reducido de forma efectiva.

Debe matizarse que el uso de este método elimina la componente en dc de la señal, y sólo los cambios son detectados. En los sistemas de detección de movimiento sólo se requiere la detección de cambios, de hecho en la detección de movimiento sólo el gradiente espacial resulta de interés, y no los niveles absolutos de intensidad.

11. Una superficie lambertiana se define como aquélla que refleja la energía incidente de forma uniforme (para cualquier ángulo de incidencia), por lo que la energía reflejada depende únicamente de la longitud de onda

Para la formulación del proceso de MNC mediante el uso del promedio espacial y división, se asume una matriz de tamaño N . Las entradas se denotan por $I_{in}(j)$. Las salidas se representan por $I_{out}(j)$. Si S es una función de filtrado resultado de promediar espacialmente las muestras, entonces el método puede describirse por:

$$I_{out}(j) = \frac{I_{in}(j)}{S(j)} \quad (5.94)$$

donde j representa la posición de la señal en la matriz. La función de promedio $S(j)$ juega un papel importante en el resultado del proceso. Para una ventana de tamaño infinito el proceso se reduce simplemente a un parámetro de escalado, mientras que para un tamaño de ventana tendiendo a cero el resultado es una salida plana de valor 1. Si el ruido está presente en un área muy pequeña y se elige un ventana relativamente grande, otras zonas se verán afectadas por la señal promediada, que portará una proporción del ruido.

Para la implementación del método en VLSI, son necesarios dos elementos: un circuito promediador; y un circuito divisor. Estos circuitos deberán tener un tamaño lo más reducido posible para poder ser implementados en los sistemas de sensado de imagen.

Referencias



- [1] McWhorter, A. L. 1957 "1/f noise and germanium surface properties" In R. H. Kingston (Ed.), *Semiconductor Surface Physics* (Philadelphia: University of Pennsylvania Press), pp. 207-228
- [2] Hooge, F. N. 1994 "1/f noise sources" *IEEE Transactions on Electron Devices*, No. 41, pp. 1926-1935
- [3] K. Laker, W. Sansen "Design of Analog Integrated Circuits and Systems" McGraw-Hill International Editions, Electrical Engineering Series
- [4] Hui Tian "Noise Analysis In CMOS Image Sensors" PhD thesis
- [5] J. S. Kim, C. H. Park, H. S. Min, Y. J. Park "Theory of 1/f Noise Current in n⁺p Diodes, n⁺p Photodiodes, and Schottky Diodes", In *Quantum 1/f Noise and Other Low Frequency Fluctuations in Electronic Devices: Seventh Symposium*, pages 123-128, 1999
- [6] E. Simoen, C. L. Clayens "On the Geometry Dependence of the 1/f Noise in CMOS Compatible Junctions Diodes" *IEEE Trans. Electron Devices*, 46(8): 1725-1732, August 1999
- [7] A. van der Ziel "Noise: Sources, Characterization, and Measurements" Prentice-Hall. Englewood Cliffs, N. J. 1970
- [8] A. Ambrozy "Electronic Noise" McGraw-Hill, New York, 1982
- [9] Y. T. Tzividis "Operation and Modeling of the MOS Transistor" McGraw-Hill
- [10] A. J. P. Theuwissen "Solid-State Imaging with Charge-Coupled Devices", Kluwer Academic Publishers, 1995
- [11] G. C. Holst "CCD Arrays Cameras and Displays: Second Edition", JCD Publishing, 1998
- [12] S. K. Mendis, E. R. Fossum, et al. "A 128x128 CMOS Active Pixel for Highly Integrated Imaging Systems", In *IEEE IEDM Technical Digest*, San Jose, CA, Dec. 1993
- [13] O. Yadid-Pecht, B. Pain, C. Staller, C. Clark, E. R. Fossum "CMOS Active Pixel Sensor Star Tracker with Regional Electronic Shutter", *IEEE J, Solid-State Circuits*, 32(2), pg. 285-288, Feb. 1997
- [14] K. Singh "Noise Analysis of a Fully Integrated CMOS Image Sensor" In *Proc. SPIE*, vol. 3650, pg. 44-51, San Jose, CA, Jan. 1999
- [15] A. van der Ziel "Noise in Solid State Devices and Circuits", Wiley, 1986

-
- [16] B. Pain, G. Yang, M. Ortiz, K. McCarty, B. Hancock, J. Heynssens, T. Cunningham, C. Wrigley, C. Ho "A Single-chip Programmable Digital CMOS Imager with Enhanced Low-Light Detection Capability" 13th International Conference on VLSI Design - January 2000, pp 342-347.
 - [17] Boyd Fowler, Michael D. Godfrey, ... "Low Noise Readout using Active Reset for CMOS APS" Proceedings of SPIE Vol. 3965, January 2000
 - [18] M. J. M. Pelgrom, A. C. J. Duinmaijer, A. P. G. Welbers, "Matching Properties of MOS Transistors", IEEE J. Solid-State Circuits, Vol. 24, pp. 1433-1440, Oct. 1989
 - [19] A. El Gamal, B. Fowler, H. Min, X. Liu "Modeling and Estimation of FPN Components in CMOS Image Sensors", In Proceedings of SPIE, vol. 3301, pp. 168-177, April 1998.
 - [20] S. Ma, L. Chen "A Single-Chip CMOS APS Camera with Direct Frame Difference Output", IEEE Journal of Solid-State Circuits", Vol. 34, No. 10, October 1999
 - [21] S. Kavadias, B. Dierickx, D. Scheffer, "On-chip offset calibrated logarithmic response image sensor", IEEE Workshop on CCDs and Advanced Image Sensors, Karizawa, 1999, P8.
 - [22] Markus Loose "A Self-Calibrating CMOS Image Sensor with Logarithmic Response" PhD thesis, 1999
 - [23] Yang Ni, Karine Matou "A CMOS Log Image Sensor with on-chip FPN Compensation", ESSCIRC'2001, 18-20 Sept. 2001 Villach, Austria, pp. 128-132.
 - [24] A. Moini, A. Blanksby, A. Bouzerdoum, K. Eshraghian, R. Beare, " Multiplicative Noise Cancellation (MNC) for Analog VLSI Vision Sensors," ETD2000, Electronics Technology Direction for the year 2000, May 1995, pp. 253-257.

CONVERSIÓN ANALÓGICO-DIGITAL EN LOS SENSORES DE IMAGEN

6.1.- INTRODUCCIÓN

A la hora de implementar la circuitería que lleva a cabo la conversión analógico-digital (CA/D) de la salida del píxel podemos optar por hacerlo bien a nivel de sistema, bien a nivel de columna, o bien a nivel de píxel (Fig. 6.1).

Las diferencias entre llevar a cabo la CA/D a uno u otro nivel han de estudiarse desde dos puntos de vista: por un lado, las prestaciones que la circuitería de CA/D ofrece en cuanto a velocidad, resolución, linealidad, robustez frente al ruido, características de desapareamiento entre dispositivos (“*missmatch*”), etc; y por otro lado, cómo las características del sistema de captura de imagen (factor de relleno, SNR, DR, etc) se ven afectadas por la incorporación de la circuitería de CA/D. Estudiemos cómo se ven afectadas cada una de estas características por separado y los compromisos de diseño que surgen entre ellas.

Para que el sensor cumpla especificaciones aceptables sobre el *factor de relleno* para un tamaño fijo de píxel, es necesario que la superficie ocupada por la circuitería añadida a nivel de píxel sea pequeña. Es decir, para poder incorporar la CA/D a nivel de píxel resulta necesario que la circuitería que desempeña tal función sea muy sencilla. Esto va en detrimento de las características deseables de esta circuitería, principalmente en lo que a velocidad de operación y resolución se refiere. Si optamos por la incorporación de la CA/D a nivel de columna o sistema, al sacarse la circuitería de CA/D del píxel el factor de relleno no se ve afectado por la superficie ocupada por ésta. En ambos casos, la superficie ocupada por la circuitería de CA/D no será la principal limitación en el diseño, por lo que el CA/D podrá ofrecer características muy aceptables de resolución y velocidad.

La realización de la CA/D a nivel de píxel permite que esta fase de operación del sensor se lleve a cabo en paralelo para todos los píxeles; si se implementa a nivel de columna, se realizará la CA/D de la matriz de píxeles de forma secuencial por filas; si la CA/D es a nivel de sistema, la conversión deberá realizarse secuencialmente píxel a píxel. Esto nos da una idea de la *velocidad* a la que deberán operar los CA/D en cada caso. Suponiendo una matriz de píxeles de tamaño $n \times m$ donde n es el número de filas y m el número de columnas; $t_{CA/D}$ el tiempo asignado para llevar a cabo la CA/D en el sensor; y N la resolución de la CA/D en número de bits, mientras que a nivel de píxel el CA/D deberá convertir el valor analógico de un sólo píxel en un tiempo $t_{CA/D}$, el CA/D a nivel de columna deberá realizar la conversión de n píxeles en el mismo tiempo, y a nivel de sistema el CA/D deberá convertir la matriz entera, es decir, $n \times m$ píxeles en un tiempo $t_{CA/D}$. Este hecho favorece la condición antes expuesta sobre el factor de relleno: las especificaciones sobre velocidad que deberá cumplir el CA/D a nivel de píxel serán relajadas en comparación con las otras dos opciones, lo que facilita un diseño sencillo.

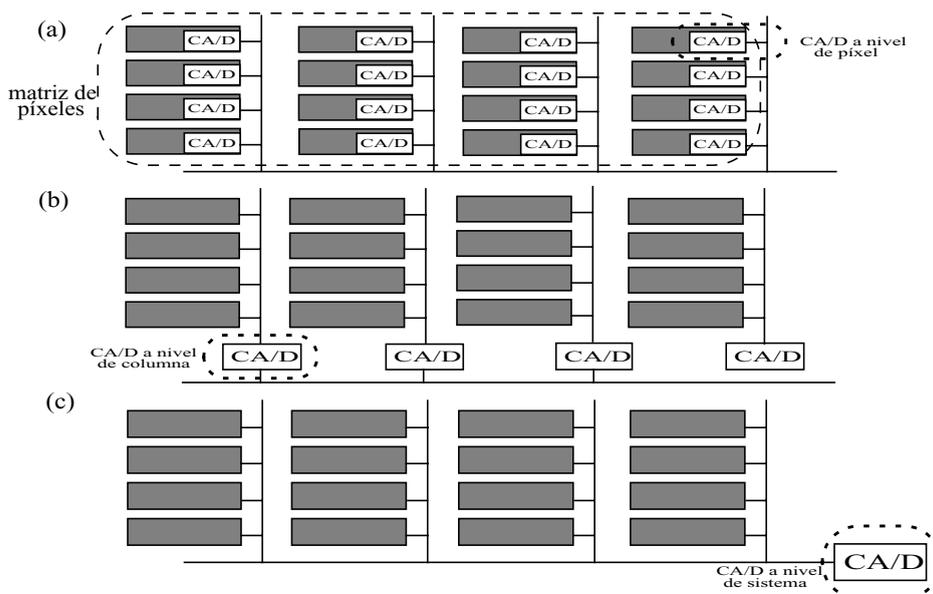


Fig. 6.1: Diferentes niveles de implementación de la conversión analógico-digital en un sensor de imagen: (a) a nivel de píxel; (b) a nivel de columna; y (c) a nivel de sistema

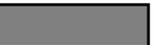
En cuanto a **precisión**, basando el razonamiento en la superficie disponible, la CA/D a nivel de píxel jugará en el límite de la resolución necesaria para ofrecer unas prestaciones aceptables. A nivel de columna, puesto que se dispone de más superficie, la precisión de la CA/D puede aumentarse; y más aún en el diseño de la CA/D a nivel de sistema, nivel al cual la precisión de la CA/D no vendrá limitada por la condición de área ocupada, sino por las características de ruido del sistema.

Respecto al **ruido de patrón fijo** (FPN) introducido por la circuitería de CA/D, la discusión es similar a la llevada a cabo respecto a la circuitería de amplificación en la Secc. 5.6, el FPN introducido es mínimo si se utiliza un sólo circuito de CA/D para llevar a cabo la conversión de todos los píxeles; y se hace máximo si se incorpora un circuito de CA/D en cada píxel.

La digitalización del dato a nivel de píxel favorece la **relación señal a ruido** (SNR), pues la circuitería analógica que procesa la señal antes de su digitalización es mínima. Una vez obtenido la salida digital del píxel será más sencillo eliminar el ruido que introduzcan las posteriores etapas. Cuando la CA/D se lleva a nivel de sistema, la digitalización se lleva a cabo una vez la señal ha atravesado todo el sistema, y todo el tratamiento de la señal se habrá llevado a cabo de forma analógica. A la hora de digitalizar el dato de salida del píxel, la componente de ruido se habrá amplificado en las diferentes etapas del sensor, por lo que la SNR se habrá degradado. La CA/D a nivel de columna es una situación intermedia a las dos expuestas.

Toda la discusión llevada a cabo hasta el momento puede resumirse en que la CA/D a nivel de píxel resulta interesante desde el punto de vista de las características del sensor, pero hace difícil el diseño del CA/D, pues al limitarse la superficie disponible para su implementación, se limitan las especificaciones que éste ofrece. Por otro lado, en el diseño de la CA/D a nivel de sistema la superficie no es la principal limitación, por lo que el diseño del CA/D permite ofrecer altas prestaciones de velocidad y resolución. Basándonos en estas directrices podemos afirmar que, en general, el **consumo de potencia** de la circuitería que lleva a cabo la CA/D será menor si se implementa a nivel de píxel, pues será más sencilla, diseñada a partir de especificaciones de resolución y velocidad más relajadas, y opera durante un periodo de tiempo corto.

El análisis comparativo expuesto se resume en la tabla [6.1] [1]. Por las ventajas que supone, la tendencia general es el diseño de la CA/D a nivel de píxel. El escalado de las tecnologías favorece este propósito, pues permite la incorporación de una circuitería más compleja a nivel de píxel, y con ello la implementación de una CA/D que ofrezca mejores características. Además, la posibilidad de incorporar la CA/D a nivel de píxel ha impulsado el desarrollo de algoritmos que mejoran el rango dinámico del sensor. La mejora del rango dinámico ofrecido por los sensores de imagen es uno de los principales retos en el diseño de este tipo de sistemas, por lo que el desarrollo de estos algoritmos supone un paso importante en la evolución de estos dispositivos.

Característica	A nivel de píxel	A nivel de columna	A nivel de sistema
Factor de relleno			
Velocidad			
Precisión			
SNR			
FPN			
Consumo de potencia			

 pobre
  regular
  bueno

Tabla 6.1: Relación entre características del sensor y nivel de implementación del CA/D

6.2.- ESQUEMAS BÁSICOS DE CONVERSIÓN ANALÓGICO-DIGITAL

Analicemos a continuación las estructuras típicas que se utilizan para llevar a cabo la conversión analógico-digital.

6.2.1.- Convertidores flash

El convertidor flash (Fig. 6.2), también llamado CA/D en paralelo, es el convertidor más rápido, debido a que realiza la conversión del dato analógico a n-bits en paralelo. Se utiliza una tensión de referencia V_{ref} a la cual se aplican sucesivos divisores de tensión (resistencias R). Cada una de las divisiones en tensión de V_{ref} se aplica como primera entrada a un comparador. Estos comparadores tomarán como segunda entrada la tensión a digitalizar, de forma que todos los comparadores cuya referencia de tensión está por debajo de la tensión de entrada tendrán como salida un '1', mientras que el resto dará un '0', con lo que la conversión analógico-digital es directa. Las señales generadas a la salida de los comparadores serán de la forma¹: 111...1000...0, por lo que es necesario el uso de un codificador para obtener la salida digital de forma habitual (por ejemplo, código binario). La tabla Tabla 6.2 ilustra un ejemplo de relación entre la salida de los comparadores y la salida digital (palabra de código).

1. Esta forma de presentación digital se conoce como código termométrico

Como se observa, la obtención de los n -bits se hace en paralelo, por lo que este convertidor requerirá tan sólo de un ciclo de reloj para llevar a cabo la conversión. El tiempo de operación de este tipo de convertidores viene fijado por los tiempos de retraso en los comparadores y en el codificador.

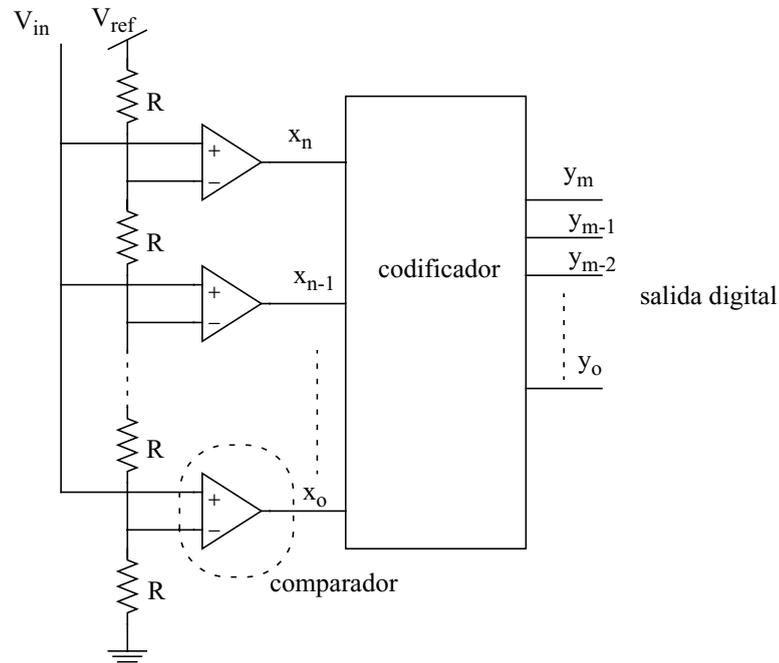


Fig. 6.2: Esquema de un CA/D flash

$x_5 \dots x_0$	$y_2 \dots y_0$
000000	000
000001	001
000011	010
000111	011
001111	100
001111	101
011111	110
111111	111

Tabla 6.2: Tabla de verdad del codificador de un CA/D flash de 3 bits

Un problema de este tipo de convertidores es que para la conversión de N bits se requieren $2^N - 1$ comparadores, por lo que el número de comparadores se dobla con cada bit. El área y la potencia requeridas son muy grandes, lo que hace imposible su implementación a nivel de píxel.

Una limitación adicional viene dada por las resistencias divisoras de tensión. En primer lugar, existe una corriente continua que atraviesa las resistencias. Para minimizar esta corriente, las resistencias han de ser grandes, con las implicaciones que esto conlleva en cuanto a superficie;

y en segundo lugar, la precisión del CA/D dependerá de la precisión con que se implementen las resistencias y los comparadores.

Respecto al codificador, la circuitería de este bloque incluirá la corrección de posibles defectos que puedan producirse en el código termométrico. Por ejemplo, si se produce el retraso en la salida de uno de los comparadores entre dos señales consecutivas cuya codificación en 8 bits sea: 11110000 y 11000000, un retraso podría generar como salida: 11010000. El codificador será el encargado de la corrección de este tipo de situaciones.

Un ventaja fundamental para cierto tipo de aplicaciones es que el convertidor flash es monótono por construcción. Las desviaciones en los valores de las resistencias provocan que la curva de conversión entre el valor analógico de entrada y las entradas a los comparadores a través de los divisores de tensión no sea lineal del todo. Aun a pesar de estas desviaciones, el convertidor flash asegura que la relación entre la entrada de los comparadores es monótona.

6.2.2.- CA/D de aproximaciones sucesivas

Estos convertidores obtienen la palabra de código digital de salida mediante comparaciones sucesivas de la entrada analógica con la salida de un CD/A cuya entrada se va generando mediante intervalos encajados. La Fig. 6.3 ilustra este concepto.

El convertidor en primer lugar evalúa el bit más significativo de la palabra, y considera que es '0' o '1' en función de si el resultado del valor de referencia del CD/A es menor o mayor que la entrada. A continuación, mediante aproximaciones sucesivas, cada uno de los bits de la palabra se van probando hasta que se alcanza la mejor referencia de la palabra digital. Es decir, el CA/D de aproximaciones sucesivas requiere $n+1$ ciclos de reloj para llevar a cabo la conversión para una resolución de n bits por palabra.

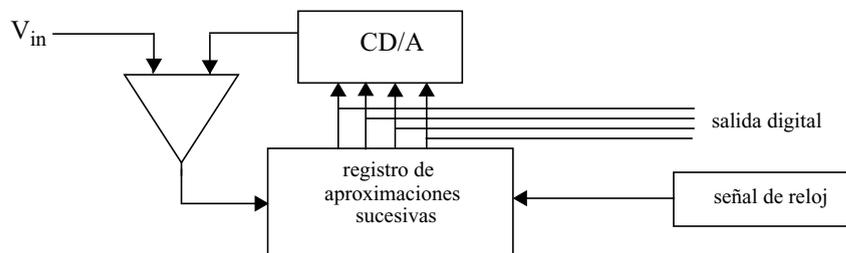


Fig. 6.3: CA/D de aproximaciones sucesivas

6.2.3.- CA/D de pendiente simple

Su funcionamiento es similar al de aproximaciones sucesivas (Fig. 6.4). En este caso un contador binario se incrementa (o decrementa) de forma continua hasta que la salida del CD/A sobrepase por exceso (o por defecto) a la entrada analógica, momento en el cual el comparador desactivará el contador. La salida binaria del contador coincidirá con la mejor aproximación digital de la tensión analógica.

Los principales inconvenientes de este convertidor son la velocidad y la precisión. Por un lado, la velocidad de la CA/D depende de la tensión a convertir, requiriendo en el peor de los casos 2^n ciclos de reloj para una precisión de n bits, mientras que por otro, la conversión depende de la precisión del convertidor A/D, por lo que la precisión dependerá fuertemente tanto del cir-

cuito generador de la señal de reloj como del circuito contador binario. La degradación con el tiempo de los componentes de estos circuitos introduce imprecisiones considerables en la conversión analógico-digital. Será necesaria la incorporación de mecanismos de autoajuste para corregir esta degradación.

Como principales ventajas, el área y la potencia consumidas son relativamente bajas.

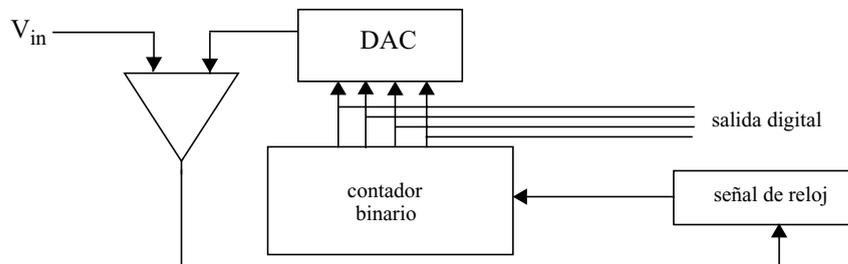


Fig. 6.4: Diagrama de bloques de un CA/D de pendiente simple

6.2.4.- CA/D de doble pendiente

El CA/D de doble pendiente (Fig. 6.5) funciona de forma parecida al de pendiente simple, mejorando principalmente en la robustez ante la degradación de su funcionamiento con el tiempo, al hacer que el valor de salida depende de la pendiente de la salida del contador binario de forma relativa, y no de forma absoluta como ocurre en los CA/D de pendiente simple.

En este CA/D se integra la entrada analógica durante un periodo fijo de tiempo (t_{int}), y tras este periodo se mide el tiempo necesario para que el sistema alcance una tensión de referencia (Fig. 6.6). La medida de este tiempo es lo que determina la conversión digital del valor analógico de entrada. Este modo de operación se consigue mediante el uso de un integrador con un interruptor analógico que conmute entre el valor de tensión de entrada a digitalizar y el valor de referencia.

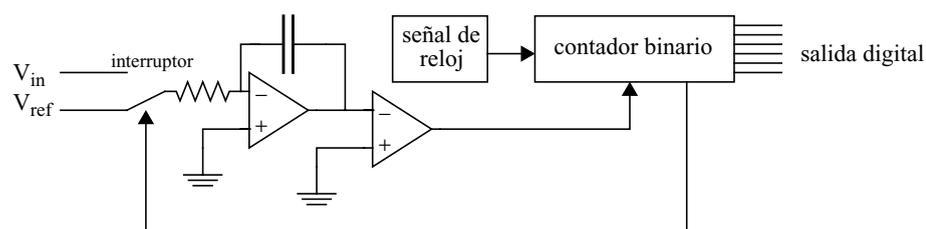


Fig. 6.5: Diagrama de bloques de un CA/D de doble pendiente

La precisión que se alcanza con este tipo de convertidores es muy alta, pues, como se ha comentado, el valor final de la conversión depende de parámetros relativos (el valor del tiempo para recuperar el valor inicial respecto del valor de tensión al final del periodo de integración). Además, gracias a esta dependencia relativa, estos esquemas resultan inmunes a la degradación temporal de los dispositivos que implementan cada bloque, así como a la constante de tiempo del integrador y a la frecuencia de reloj. El precio a pagar por esta precisión es el tiempo de operación, pues t_{int} puede resultar excesivo para determinadas aplicaciones.

Como ventajas adicionales, el consumo de potencia y área no es demasiado grande.

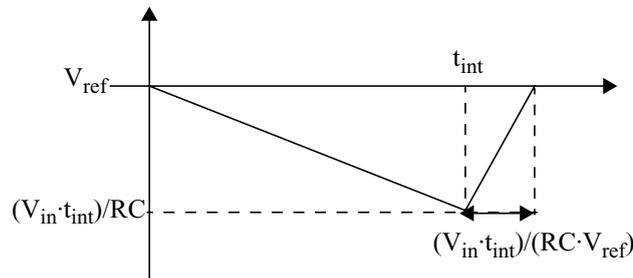


Fig. 6.6: Funcionamiento de un CA/D de doble pendiente

6.2.5.- Convertidores sigma-delta ($\Sigma-\Delta$)

Los moduladores $\Sigma-\Delta$ se basan en el sobremuestreo de la señal. Esta técnica de sobremuestreo relaja los requerimientos de la circuitería analógica. El precio a pagar básicamente consiste en un aumento de la velocidad y complejidad de la circuitería digital [2].

La Fig. 6.7 [3] muestra el esquema básico de un modulador $\Sigma-\Delta$. Puede observarse que la salida y se resta, a través del lazo de realimentación, a la entrada x , la cual ha sido sobremuestreada a una frecuencia mucho mayor que la frecuencia de Nyquist. La señal resultante, tras procesarse por un filtro en tiempo discreto $H(z)$, sirve como entrada al cuantizador, que en general trabajará con un número reducido de niveles de cuantización. Si la ganancia del filtro es alta en el rango de frecuencias de interés, y baja fuera de este rango, el error de cuantización se atenúa en la banda de interés debido al lazo de realimentación.

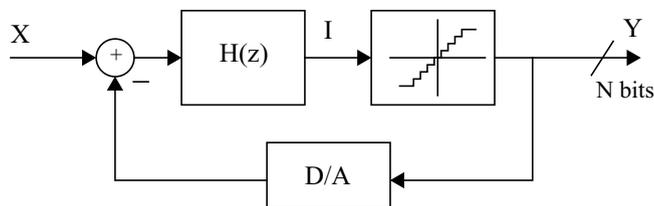


Fig. 6.7: Esquema básico de un modulador $\Sigma-\Delta$

Los convertidores $\Sigma-\Delta$ resultan inmunes en gran medida a las variaciones en el proceso de fabricación, pues mientras las anteriores estructuras requieren de una precisión alta en la relación entre componentes, los convertidores $\Sigma-\Delta$ basan su funcionamiento en promediados de señales [1]. El diagrama de bloques de un convertidor $\Sigma-\Delta$ básico se ilustra en la Fig. 6.8.

El principio de funcionamiento es el siguiente (Fig. 6.9). En cada ciclo de reloj, el integrador va acumulando múltiplos de la tensión de entrada hasta que la tensión total excede el valor de referencia V_{max} , momento en el cual la salida q del comparador se pone a '1', se activa el interruptor analógico y el lazo de realimentación resta una cantidad V_{max} a la señal de entrada. Este ciclo se repite durante 2^N ciclos de reloj (donde N es el número de bits de la palabra digital), de forma que contando el número de veces que la señal q se ha puesto a '1' obtenemos la conversión digital.

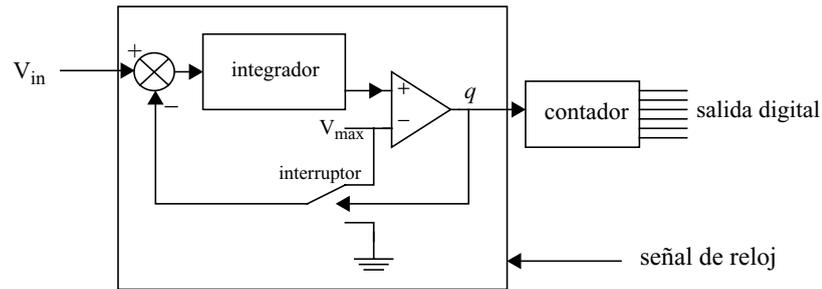


Fig. 6.8: Diagrama de bloques de un CA/D Σ - Δ

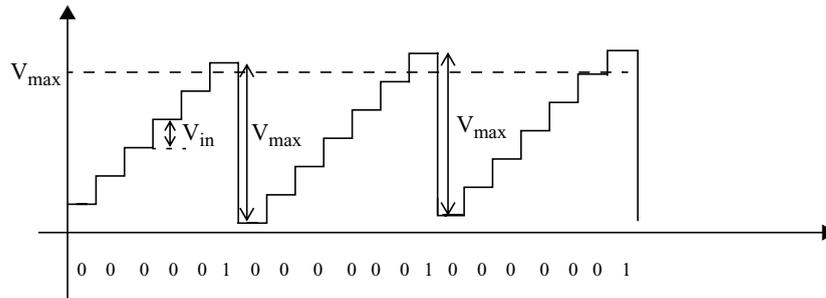


Fig. 6.9: Principio de funcionamiento de un CA/D Σ - Δ

El modulador Σ - Δ genera datos de salida a L veces la frecuencia de Nyquist. El dato sobremuestreado se diezma hasta la frecuencia de Nyquist mediante el uso de un filtro digital.

Los convertidores Σ - Δ resultan ventajosos en cuanto a resolución, consumo, área y robustez frente al ruido. Como inconveniente, la velocidad de operación no es alta. Dado que los tiempos de operación en los sensores de imagen son relativamente altos, estos convertidores resultan muy útiles para la implementación de la CA/D en aplicaciones de sensado de imagen.

La principal desventaja de esta técnica de CA/D es que el ancho de banda del dato de salida antes del diezmo es mucho mayor que el del dato final a la frecuencia de Nyquist. Si el diezmo se lleva a cabo fuera del sistema, es decir, "off-chip", los requerimientos de ancho de banda de entrada/salida del sistema serán fuertes, especialmente para velocidades altas o sensores con un formato de bit grande.

6.2.6.- Técnica Multi-Channel Bit Serial (MCBS)

Las técnicas tradicionales de CA/D, tales como la CA/D de pendiente simple, requieren una celda de memoria de m bits en cada píxel para implementar la conversión de m bits, lo cual no resulta factible con los valores típicos de m en aplicaciones de sensado de imagen. Además, las técnicas de conversión en serie, tales como las aproximaciones sucesivas o algoritmos de CA/D requieren circuitos relativamente complejos, de componentes analógicos muy precisos, por lo que no se adaptan a la implementación de la CA/D a nivel de píxel.

La *técnica MCBS (Multi-Channel Bit Serial)*, introducida por D. Yang, B. Fowler y A. El Gamal [4] consigue la implementación de la CA/D a nivel de píxel mediante un algoritmo que requiere sólo una celda de memoria de 1 bit en cada píxel. La descripción de esta técnica se incluye a continuación.

La CA/D realiza la conversión de una señal analógica en una representación digital (denominada *palabra de código*) de acuerdo con una tabla de cuantización. Por ejemplo, en la tabla

Tabla 6.3 se ilustra el caso de una conversión a código Gray², donde S es una señal analógica que toma valores en el intervalo continuo $(0,1]$. La tabla muestra la asignación de cada intervalo de entrada a una palabra digital de 3 bits. La observación que permite simplificar la CA/D es que cada bit de la palabra de código puede generarse de forma independiente: el bit menos significativo (LSB) estará a '1' si la señal de entrada pertenece al intervalo $(0.125, 0.375] \cup (0.125, 0.375]$; el segundo bit estará a '1' si la señal de entrada pertenece al intervalo $(0.250, 0.625]$; mientras que el bit más significativo (MSB) estará a '1' si la señal de entrada pertenece al intervalo $(0.500, 1.000]$. Según esta observación, el CA/D es esencialmente un detector que indica para cada rango qué bit de la palabra de código está a '1'. Por lo tanto, este detector puede implementarse con un sólo comparador de 1 bit, por lo que sólo se requiere una celda de memoria de 1 bit.

Rangos de entrada de la señal analógica	Palabra de código
$(0.000, 0.125]$	000
$(0.125, 0.250]$	001
$(0.250, 0.375]$	011
$(0.375, 0.500]$	010
$(0.500, 0.625]$	110
$(0.625, 0.750]$	111
$(0.750, 0.875]$	101
$(0.875, 1.000]$	100

Tabla 6.3: Código Gray para un CA/D de 3 bits ($m=3$)

Veamos el diagrama de bloques del detector (Fig. 6.10(a)) para la generación del bit LSB. La descripción del algoritmo es la siguiente. Al terminal positivo del comparador se conecta la señal analógica S , mientras que al negativo se conecta la señal $RAMP$ (Fig. 6.10(b)). La salida del comparador será la puerta de la celda de memoria. El dato de entrada de la celda de memoria será la señal BITX (Fig. 6.10.(b)). En el momento en el que la señal $RAMP$ excede el valor de S , el comparador cambia su salida a '0' y la celda memoriza el valor de BITX. El valor almacenado será el bit menos significativo de la palabra de código. Dado que $RAMP$ crece monótonicamente, el comparador sólo cambiará el valor de salida una vez, por lo que la celda de memoria almacenará el valor deseado. Una vez que $RAMP$ alcanza el valor máximo se procede a resetear las señales $RAMP$ y BITX para proceder a la siguiente comparación. Para la detección del resto de bits la secuencia es idéntica excepto en la definición de los intervalos (señal BITX).

La extracción de los datos se lleva a cabo por *planos de bits*: en primer lugar se calcula el bit LSB en paralelo para todos los píxeles, generando así el plano del bit LSB. El plano se descarga fuera de la matriz de sensado y se procede a calcular el siguiente bit en paralelo para todos los píxeles. Este mecanismo se repite para todos los bits, con lo que se consigue sacar de la matriz de píxeles la memoria necesaria para almacenar los valores digitalizados de los píxeles.

-
2. El código Gray se caracteriza porque entre dos palabras de código consecutivas sólo existe un bit de diferencia. Esta propiedad es muy útil a la hora de implementar algoritmos para la reducción del ruido, pues minimizar el error cuadrático medio, entendiendo por error u fallo en la conversión A/D.

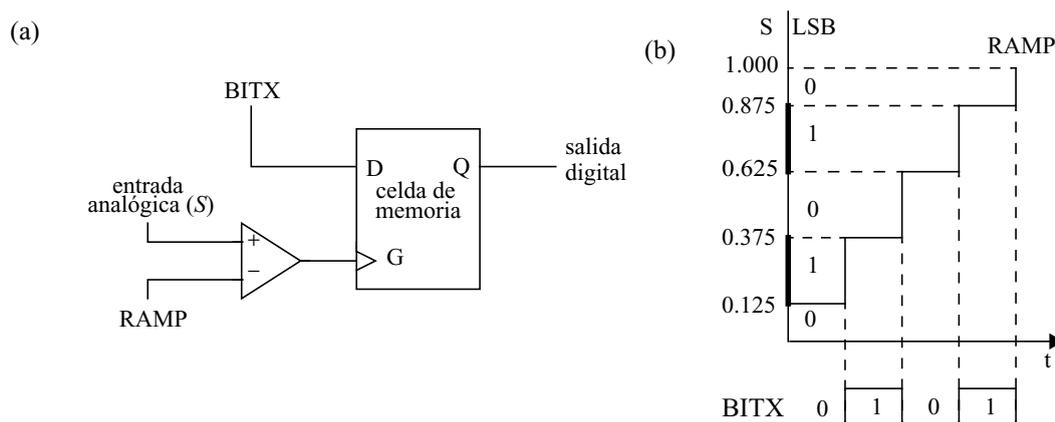


Fig. 6.10: (a) Diagrama de bloques de un canal del CA/D MCBS; y (b) Señales de control para la generación del bit LSB

Este ejemplo de 3 bits puede generalizarse para la implementación de un CA/D de m -bits. Para alcanzar una precisión de m -bits el intervalo de entrada se divide en 2^m rangos de entrada, de forma que cada rango está representado a la salida por una palabra. Para determinar la palabra, el CA/D genera los bits de la palabra en serie, mediante sucesivas comparaciones. En cada comparación la señal RAMP no varía, mientras que BITX indicará qué intervalos han de dar salida '1'.

El número de comparaciones necesarias para la generación de cada bit varía. En el ejemplo de la palabra de 3 bits, para la generación del bit LSB son necesarias cuatro comparaciones, mientras que para el bit MSB sólo es necesaria una. Como límite superior, para cualquier código, el número de comparaciones necesarias es de $2^m - 1$. La codificación Gray trabaja por debajo de este límite, minimizando el número de comparaciones necesarias, reduciendo por tanto la probabilidad de error en la conversión.

Aunque se ha asumido una cuantización uniforme, resulta muy sencillo la implementación de cualquier cuantización, por ejemplo, de una distribución logarítmica de rangos de la señal de entrada.

La Fig. 6.11 muestra el diagrama de bloques de un CA/D MCBS. Las señales RAMP y BITX están distribuidas de forma global. La circuitería compartida consta de una máquina de estados sencilla y un CDA de m -bits, necesario para la generación de la señal RAMP. Los únicos bloques necesarios a nivel de píxel son un comparador y una celda de memoria de 1 bit.

6.2.7.- Otras técnicas de conversión analógico-digital

Presentamos a continuación, de forma resumida, otras arquitecturas de conversión analógico-digital que, por sus características, no resultan interesantes para su implementación a nivel de píxel.

6.2.7.1.- Arquitecturas de dos etapas

El incremento exponencial de consumo, área, y capacidad de entrada de los convertidores flash en función de la resolución los hacen impracticables para resoluciones por encima de los 8 bits, lo que hace necesario el desarrollo de otras arquitecturas para las cuales el compromiso entre los parámetros citados y la resolución sea admisible. Las arquitecturas de dos etapas reducen la velocidad a cambio de potencia, área y capacidad de entrada.

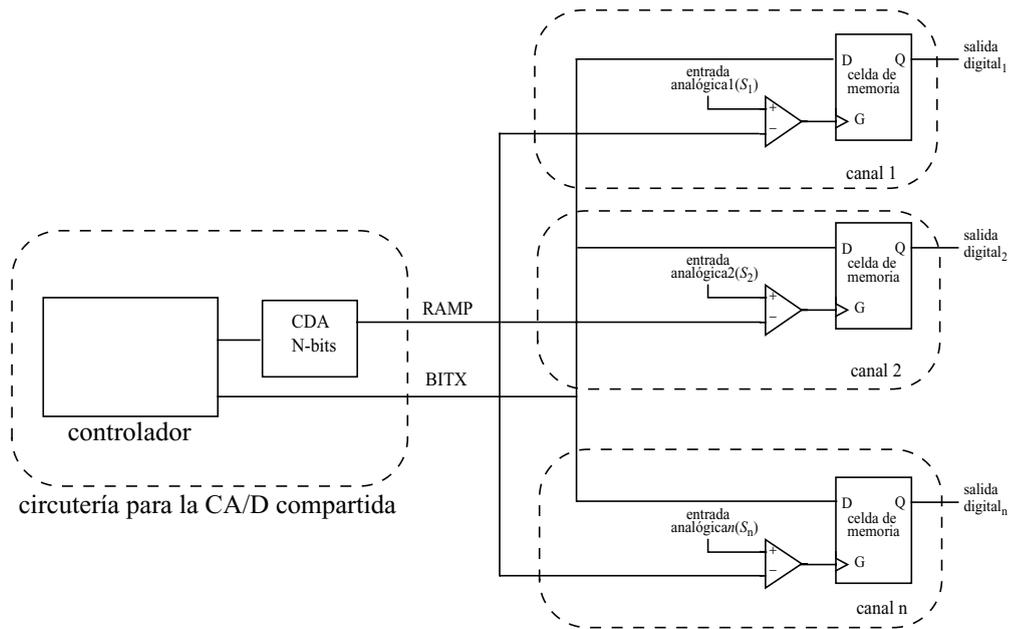


Fig. 6.11: Diagrama de bloques de un CA/D MCBS de n canales

En un convertidor de dos etapas, en primer lugar se obtiene una aproximación grosera de la entrada para poder adaptar un pequeño rango de tensiones alrededor del nivel de entrada. A continuación, el nivel de entrada se determina con una mayor precisión dentro del rango establecido. La Fig. 6.12 [5] ilustra una arquitectura de dos etapas. Consiste en una etapa de muestreo y retención front-end (SHA), un CA/D flash muy vasto (conversión grosera), un CD/A, una etapa de resta, y una etapa CA/D flash de precisión (conversión fina).

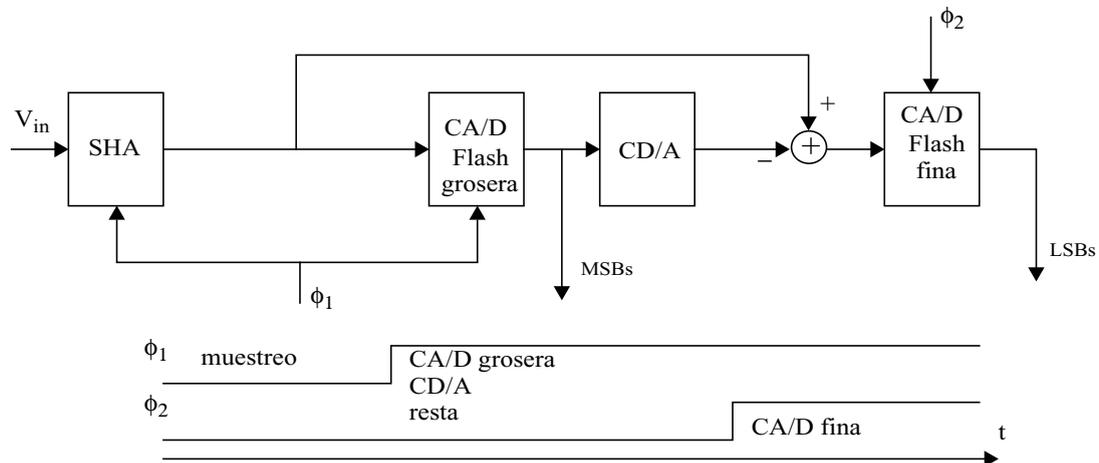


Fig. 6.12: Arquitectura de CA/D de dos etapas

Presentamos a continuación algunas arquitecturas que responden a esta estructura de dos etapas.

6.2.7.1.1.- Arquitectura de reciclaje de dos etapas

Un convertidor A/D de dos etapas no necesita utilizar dos etapas flash separadas para llevar a cabo las conversiones basta y precisa. Una sola etapa puede utilizarse para ambas conversiones.

nes. En tal caso el convertidor utiliza lo que se denomina una arquitectura de reciclaje (Fig. 6.13 [5]). En esta estructura, durante la conversión grosera, la etapa flash sensa la salida del front-end SHA, V_A , y genera la salida digital basta. Esta salida es entonces convertida en analógica por el CD/A y se resta a V_A utilizando la etapa de resta. Durante la conversión precisa, la salida de la etapa de resta es digitalizada por la etapa flash. Debe hacerse notar que en esta segunda fase, el fondo de escala del CA/D debe ser igual al rango de salida de la etapa de resta. Es decir, para la conversión precisa, o bien la referencia del CA/D debe reducirse, o bien el residuo debe ser amplificado.

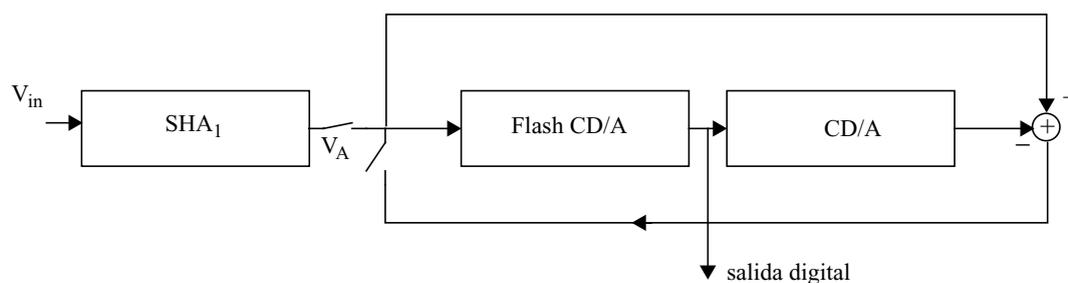


Fig. 6.13: Arquitectura de CA/D de reciclaje

6.2.7.1.2.- Arquitectura subranging de dos etapas

Una variante de las arquitecturas de dos etapas que no necesita un restador explícitamente es la topología “subranging”. En esta arquitectura, la etapa grosera identifica y subdivide un rango de tensiones de referencia en torno a la tensión de entrada. La técnica de subranging puede explicarse utilizando las características mostradas en la Fig. 6.14 [5]. La Fig. 6.14(a) es la salida intermedia del CD/A en un CA/D ideal en función de la entrada analógica. La estimación generada por el CD/A debe restarse a la entrada para producir un residuo sobre el que llevar a cabo la cuantización fina. Ahora, supongamos que el CD/A puede proporcionar dos salidas diferentes, una desplazada verticalmente el tamaño del escalón de cuantización, como se muestra en la Fig. 6.14(b). En tal caso, la señal de entrada será siempre mayor que V_{D1} y menor que V_{D2} , y si la diferencia entre V_{D1} y V_{D2} se subdivide en 2^n , siendo n el número de bits de resolución del CA/D, se proporcionará un conjunto de 2^n tensiones de referencia alrededor de la señal de entrada. La etapa fina podrá entonces comparar la entrada frente a estas tensiones de referencia y determinar los bits menos significativos de la muestra.

6.2.7.2.- Arquitecturas interpolativas y de plegado

Estas arquitecturas pretenden la mejora de las características de la CA/D manteniendo la naturaleza de una sola etapa de estas arquitecturas, es decir, sin añadir circuitos de muestreo y retención al CA/D. La interpolación y el plegado o “folding” han demostrado ser técnicas muy beneficiosas.

6.2.7.2.1.- Interpolación

Para la reducción del número de preamplificadores en la entrada de un convertidor flash, la diferencia entre la entrada analógica y cada tensión de referencia puede cuantizarse a la salida de cada preamplificador. Esto es posible gracias a la ganancia finita (y por lo tanto, rango lineal de entrada distinto de cero) de los preamplificadores típicos utilizados como front-end de los comparadores.

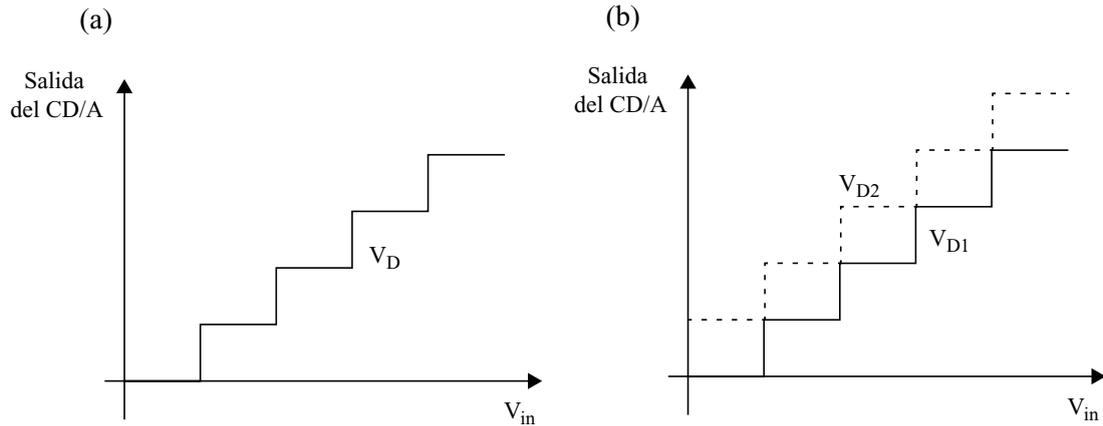


Fig. 6.14: (a) Salida de la etapa CD/A intermedia de un CA/D ideal de dos etapas; y (b) salidas CD/A para una arquitectura subranging

El concepto de interpolación puede ilustrarse con la Fig. 6.15 [5], en la cual los preamplificadores A_1 y A_2 comparan la entrada analógica con V_{r1} y V_{r2} , respectivamente. Puede demostrarse que, siendo $V_m = (V_{r1} + V_{r2})/2$, la polaridad de la diferencia entre V_{X2} y V_{Y1} es la misma que la diferencia entre V_{in} y V_m . Basándose en esta observación, la resolución equivalente de una etapa flash puede incrementarse mediante la interpolación de las salidas de los preamplificadores. La Fig. 6.16 muestra cómo una celda de memoria adicional detecta la polaridad de la diferencia entre las salidas de dos preamplificadores adyacentes. El número de preamplificadores se reduce a la mitad ($N/2$) respecto a la arquitectura flash convencional (N), pero el número de celdas de memoria necesarias será igual a N .

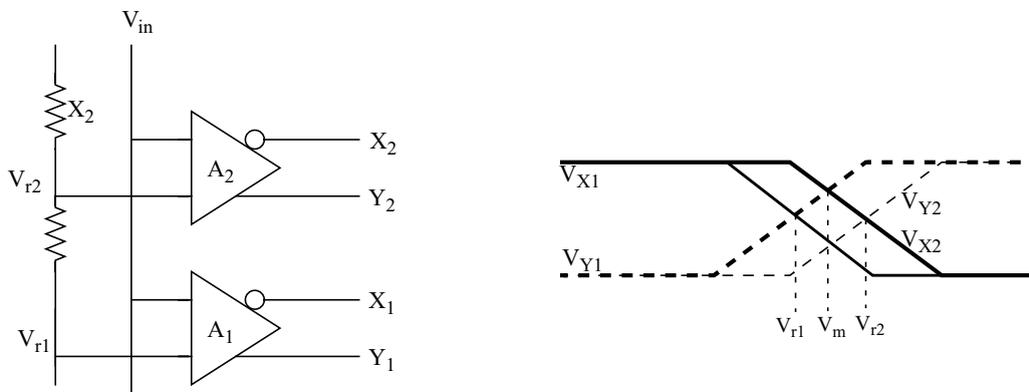


Fig. 6.15: Interpolación entre las salidas de dos amplificadores

La técnica de interpolación reduce sustancialmente la capacidad de entrada, la disipación de potencia, y el área ocupada por los convertidores flash, al mismo tiempo que preserva la naturaleza de una etapa de la arquitectura. Esto es posible porque todas las señales llegan a la entrada de las celdas de memoria de forma simultánea, y por lo tanto pueden capturarse en un solo ciclo de reloj.

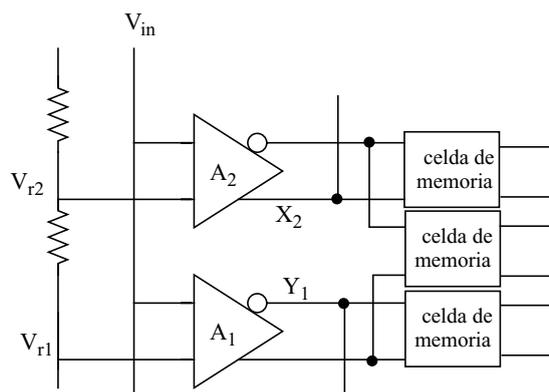


Fig. 6.16: Interpolación en un CA/D flash

6.2.7.2.2.- Arquitecturas de plegado o “folding”

Las arquitecturas de plegado o “folding” llevan a cabo un preprocesado de la señal analógica para reducir la circuitería necesaria, manteniendo la naturaleza de una sola etapa de las arquitecturas flash.

El principio de operación de las arquitecturas de plegado es la generación de una tensión residual mediante un preprocesado analógico y la consecuente digitalización de este residuo para obtener los bits menos significativos. Los bits más significativos pueden obtenerse utilizando una etapa flash grosera que opere en paralelo con el circuito de plegado y, por lo tanto, muestree la señal a aproximadamente el mismo tiempo en el que el residuo es muestreado. La Fig. 6.17 muestra la generación del residuo para una arquitectura de dos etapas y una de plegado.

6.2.7.3.- Arquitecturas en “pipeline”

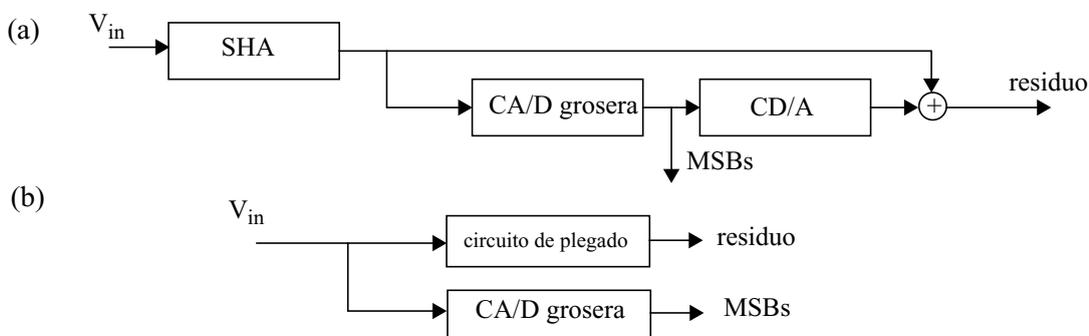


Fig. 6.17: Generación del residuo en las arquitecturas (a) de dos etapas; y (b) de plegado

El concepto de “pipeline”, normalmente utilizado en circuitos digitales, puede también aplicarse en el dominio analógico para conseguir velocidades mayores donde un conjunto de operaciones deben llevarse a cabo en serie. La Fig. 6.18 [5] muestra un sistema pipeline genérico. Aquí, cada etapa lleva a cabo una operación sobre una muestra, proveyendo una salida a la siguiente etapa, y, una vez que la siguiente etapa ha adquirido la muestra, comienza la misma operación sobre la siguiente muestra. Por lo tanto, en cualquier momento, todas las etapas están procesando diferentes muestras de forma concurrente, y por lo tanto la frecuencia de salida depende tan sólo de la velocidad de cada etapa y del tiempo de adquisición de la siguiente muestra.

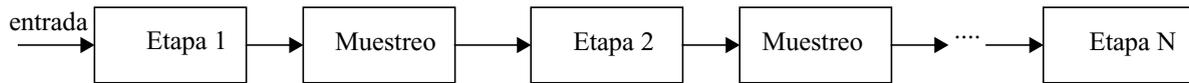


Fig. 6.18: Arquitectura en pipeline genérica

Como ejemplo sencillo de estructura pipeline, considérese un CA/D de dos etapas en el que deben realizarse cuatro fases u operaciones en serie, denominadas: conversión A/D grosera, conversión D/A intermedia, resta, y conversión A/D fina. Con tal esquema, el CA/D no puede empezar a procesar la siguiente muestra hasta que las cuatro operaciones se hayan completado. Supongamos ahora que una etapa de muestreo (SHA) y retención se añade entre la resta y la etapa de conversión fina, como se ilustra en la Fig. 6.19, de forma que el residuo se almacena antes de que la conversión fina comience. De esta forma, el SHA del front-end, el CA/D grosero, la CD/A intermedia, y la resta pueden comenzar a procesar la siguiente muestra mientras que la etapa de CA/D fina procesa la anterior muestra, y por lo tanto, potenciando una mayor rapidez de conversión.

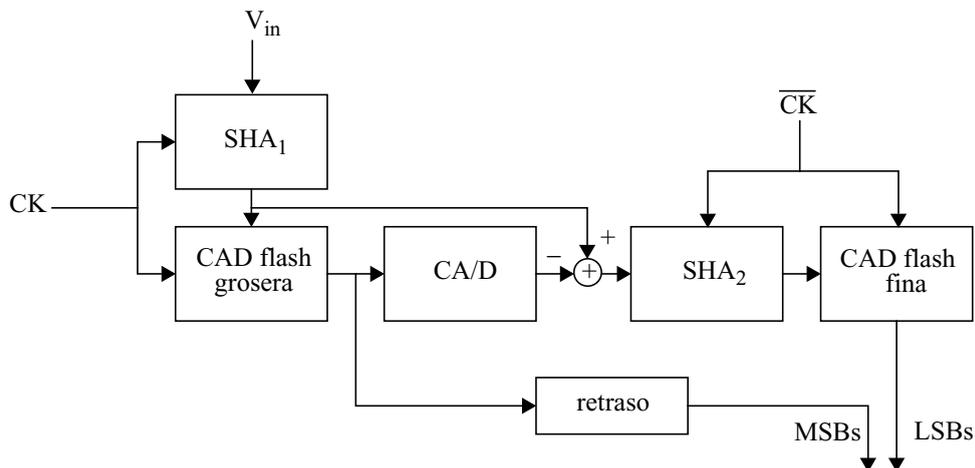


Fig. 6.19: Arquitectura de CA/D de dos etapas basada en pipeline

La Fig. 6.20 [5] muestra un esquema más general de los CADs basados en una arquitectura pipeline. La arquitectura consiste en N etapas, cada una de las cuales incluye una etapa SHA, un CA/D, un CD/A, una etapa de resta, y un amplificador opcional. El esquema de conversión será el siguiente. La primera etapa muestrea y mantiene la entrada analógica, produce una estimación digital de k -bits de la entrada mantenida, convierte esta estimación en una señal analógica, resta esta señal a la entrada mantenida, generando el residuo, y, en algunas implementaciones, amplifica este residuo por 2. A continuación, la siguiente etapa de la estructura muestrea el residuo amplificado y lleva a cabo la misma secuencia de operaciones mientras que la primera etapa comienza el procesamiento de la siguiente muestra a convertir.

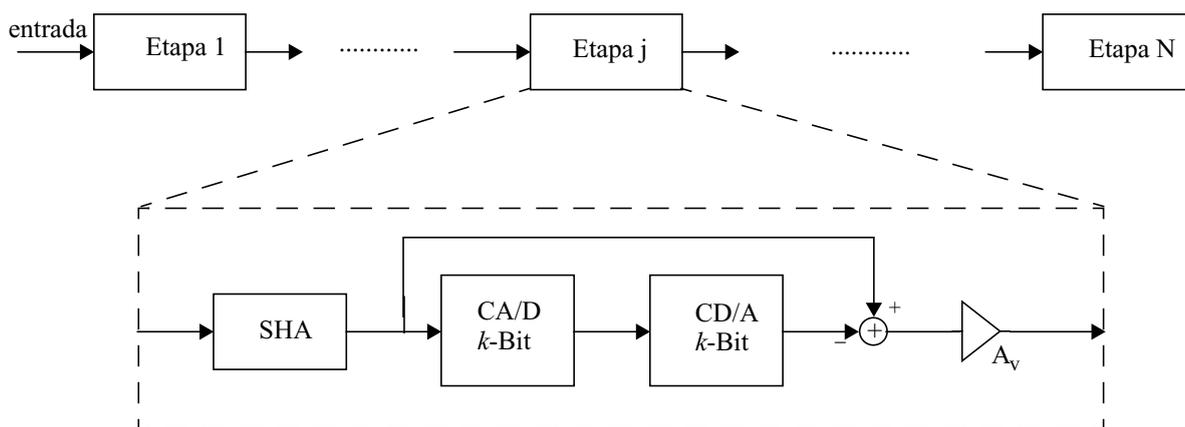


Fig. 6.20: Arquitectura de CA/D de N etapas basada en pipeline

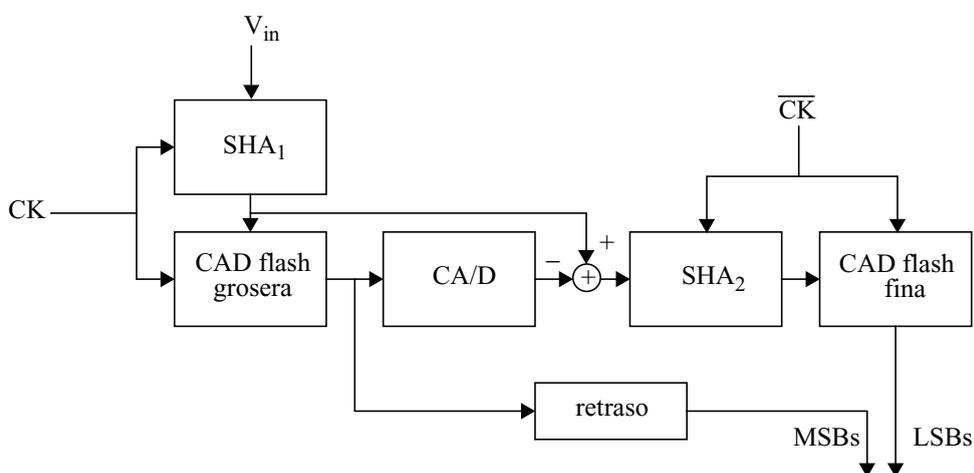


Fig. 6.21: Arquitectura de CA/D de dos etapas basada en pipeline

6.3.- TÉCNICAS DE CONVERSIÓN ANALÓGICO-DIGITAL EN LOS SENSORES DE IMAGEN

Presentamos a continuación una serie de técnicas para llevar a cabo la conversión analógico-digital en los sensores de imagen. Estas técnicas están orientadas a lograr la realización de la CA/D a nivel de píxel, por lo que se basarán en circuiterías muy sencillas que minimicen el área ocupada. A pesar de esto, en la mayoría de los casos la superficie ocupada por la circuitería que realiza la conversión no permite que la estructura se implemente realmente a nivel de píxel. En estos casos resulta típico agrupar los píxeles en submatrices, de forma que, a través de un multiplexor analógico, el convertidor se comparte entre todos los píxeles de la submatriz³.

3. Una configuración típica es el uso de un multiplexor 4:1, de forma que cada CA/D es compartido por 4 dispositivos fotosensores

6.3.1.- Implementación de la CA/D Σ - Δ a nivel de píxel

La implementación de la CA/D Σ - Δ a nivel de píxel propuesta en [6] se muestra en la Fig. 6.22. Consta de 14 transistores, y consiste en un comparador gobernado por una señal de reloj (CK), y un CD/A de 1 bit. El modo de operación de la CA/D es el que sigue. Una vez direccionado el dispositivo fotosensor para llevar a cabo la conversión A/D con la señal $SELECT$, la tensión resultado de la integración de la carga fotogenerada es cuantizada usando un comparador con realimentación regenerativa gobernado por la señal de reloj CK . El comparador se polariza en la región subumbral con la intención de reducir la potencia y el ruido, aumentar la ganancia. La corriente de polarización será la mínima necesaria para conseguir la frecuencia de muestreo deseada [7]. El dispositivo fotosensor se realimenta con el valor cuantizado tras ser convertido en carga utilizando un CD/A de 1 bit. El CD/A se implementa con un registro de desplazamiento analógico, similar a una estructura de transferencia de carga (CCD) de tres fases. Si la salida del comparador $FEEDBACK$ es débil, una secuencia de control de tres pulsos $DUMP$, $STORE$ y CK (Fig. 6.23) inyecta una cantidad de carga fija en la capacidad del dispositivo fotosensor. La carga total q_d , transferida a la capacidad del fotodiodo viene dada por la expresión:

$$q_d = (V_{dd} - V_{STORE} - V_{tp})C_{ox}W_{16}L_{16} \quad (6.1)$$

donde V_{dd} es la tensión de alimentación; V_{STORE} es el valor bajo de tensión de la señal $STORE$; C_{ox} la capacidad de óxido de M_{16} ; y W_{16} , L_{16} las dimensiones de M_{16} . Como puede observarse, la señal V_{STORE} puede utilizarse para controlar el total de la carga transferida a la capacidad del dispositivo fotosensor. Al final de cada ciclo de reloj se genera un bit y se lee a través del bus de columna.

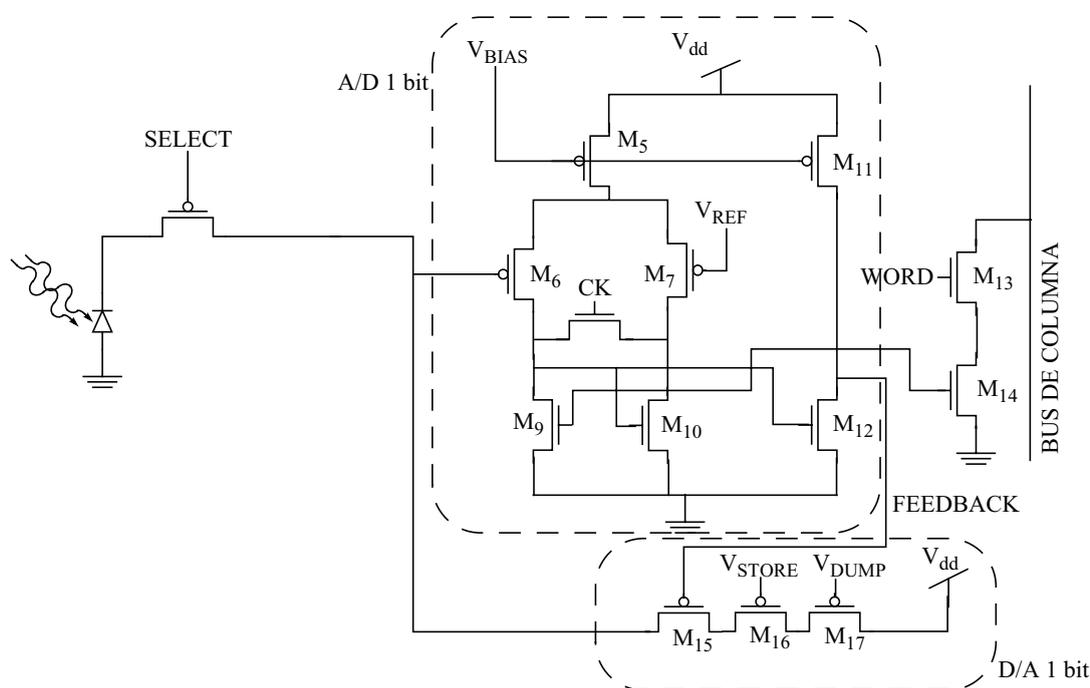


Fig. 6.22: Implementación del CA/D Σ - Δ a nivel de píxel

Las principales ventajas de este circuito [7] son su reducido tamaño, baja sensibilidad al ruido de los transistores, ausencia de FPN de offset, y una capacidad de integración total elevada. El ruido referido a la entrada generado por el CD/A de 1 bit, y las componentes de ruido flicker y térmico del comparador son muy pequeños en comparación con el ruido de cuantización, y por lo tanto pueden ser despreciados para la mayoría de las aplicaciones. Más aún, el ruido intro-

ducido por el CD/A se reduce mediante la integración de la carga inyectada en el dispositivo fotosensor, mientras que el ruido del comparador se reduce a través del lazo de realimentación. La componente de FPN debida al offset del modulador $\Sigma-\Delta$ es cero debido a la diferenciación de primer orden de todos los ruidos y errores de offset que se lleva a cabo a través del lazo [8]. La capacidad de integración total elevada se debe a la inyección de carga a través del lazo de realimentación a la capacidad de integración del dispositivo fotosensor. Por ejemplo, si la tasa de sobre muestreo es 64, la capacidad de carga del circuito es aproximadamente 64 veces mayor que la de un APS con el mismo dispositivo fotosensor.

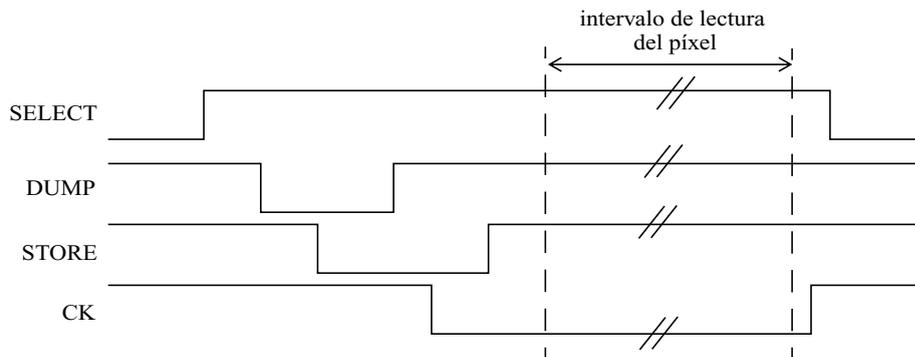


Fig. 6.23: Diagrama de tiempo de las señales de control del modulador $\Sigma-\Delta$

Los principales inconvenientes de este circuito incluyen la componente de FPN de ganancia, respuesta pobre a niveles bajos de intensidad, respuesta no lineal, y frecuencia de salida de datos elevada [6]. El FPN de ganancia está provocado por las variaciones en los parámetros de M_{16} de un modulador a otro. La respuesta pobre a niveles bajos de intensidad está asociado a que en la modulación $\Sigma-\Delta$ el intervalo asociado al rango menor de la señal de entrada tiene la menor resolución. La respuesta no lineal se debe a la característica no lineal entre carga y tensión en el dispositivo fotosensor. Este efecto puede reducirse considerablemente incrementando la tensión inversa de polarización o reduciendo el rango de variación de tensión en el dispositivo fotosensor.

6.3.2.- Uso de un oscilador muestreado como modulador $\Sigma-\Delta$ de primer orden

El diseño de un oscilador muestreado como modulador $\Sigma-\Delta$ de primer orden se basa en el esquema de la Fig. 6.24. El nodo de sensado se conecta a un comparador que cambia de estado cuando la tensión V_{ph} disminuye por debajo de una tensión de referencia V_{low} . La salida del comparador se utiliza para realimentar la puerta de la llave de reset. Con esta configuración, el píxel actúa como un oscilador que genera pulsos en el nodo de salida a una frecuencia inversamente proporcional al tiempo que tarda la tensión V_{ph} en alcanzar la referencia V_{low} .

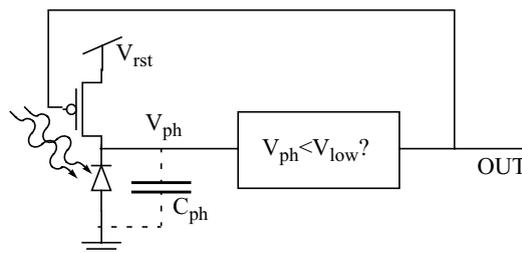


Fig. 6.24: Diagrama de bloques de un oscilador muestreado como modulador $\Sigma-\Delta$

Consideremos la forma canónica del modulador Σ - Δ de primer orden mostrado en la Fig. 6.25. Una entrada, $x[m]$, generada mediante el muestreo continuo de $x(t)$ en los instantes $m \cdot \tau$, alimenta a un acumulador. Si la salida del acumulador está por encima de un valor umbral W_{ref} la salida $y[m]$ se pone a '1' y una cantidad x_{max} se resta de la siguiente muestra (Fig. 6.25(a)). Si la salida del acumulador está por debajo de ese umbral, $y[m]$ se pone a '0', y la siguiente entrada pasa sin modificarse al acumulador.

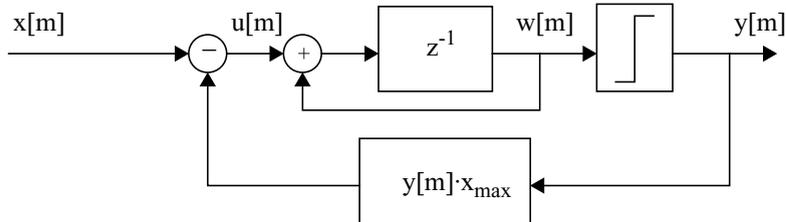


Fig. 6.25: Diagrama de bloques de un cuantizador Σ - Δ canónico de primer orden para un dato muestreado

Si consideramos ahora el funcionamiento de un circuito continuo en el tiempo que integra una entrada x hasta que el valor acumulado alcanza un nivel de referencia W_{ref} , resulta sencillo demostrar la equivalencia entre un modulador Σ - Δ de primer orden y un oscilador muestreado que opere de forma asíncrona respecto de la señal de reloj [9]. En el momento en el que el circuito integrador alcanza el nivel de referencia W_{ref} , el integrador se resetea a cero, obteniéndose como patrón de salida la señal mostrada en la Fig. 6.25(b). Tras un escalado dimensional para eliminar t , de forma que $w[m]=w(m \cdot t)$, puede comprobarse gráficamente que las formas de onda continua y digital (Fig. 6.25) se ajustan perfectamente. Puede observarse que si el modulador síncrono cruza el valor umbral en el límite de un ciclo de reloj, entonces el oscilador asíncrono equivalente debe ser reseteado durante el intervalo $((m-1) \cdot \tau, m \cdot \tau]$. Supongamos que el circuito asíncrono genera un pulso de ancho τ cada vez que se resetea, y que este pulso se muestrea en el siguiente límite del ciclo de reloj para generar la salida. Si el ancho de este pulso se configura para que sea igual al periodo de reloj, se garantiza que el pulso se muestreará sólo una vez. Claramente, la salida binaria del oscilador muestreado y del modulador Σ - Δ de primer orden será la misma.

El patrón de la señal de salida puede generarse a partir de cualquier oscilador. El muestreo proporciona información sólo de la frecuencia de oscilación. Esta frecuencia resulta proporcional a la entrada, y es la inversa del tiempo que lleva integrar la señal de entrada desde 0 hasta W_{ref} , o lo que es lo mismo, $T=W_{ref}/x$. Si $T \leq \tau$, la salida se saturará (salida a 1 en todo momento) no dando más información. La entrada máxima que puede observarse es, por lo tanto $x_{max}=W_{ref}/\tau$, y la frecuencia de oscilación puede expresarse como:

$$f = \frac{1}{T} = \frac{x}{x_{max} \cdot \tau} \quad (6.1)$$

La ecuación 6.1 describe un mapeo lineal de los valores de frecuencia dentro del rango $[0, 1]$ en pasos de $1/\tau$.

Un ejemplo de oscilador asíncrono como modulador se describe en [10]. El esquema del oscilador se muestra en la Fig. 6.27. Se basa en un oscilador simple de relajación diseñado para minimizar el consumo. El píxel contiene un comparador simple como único elemento

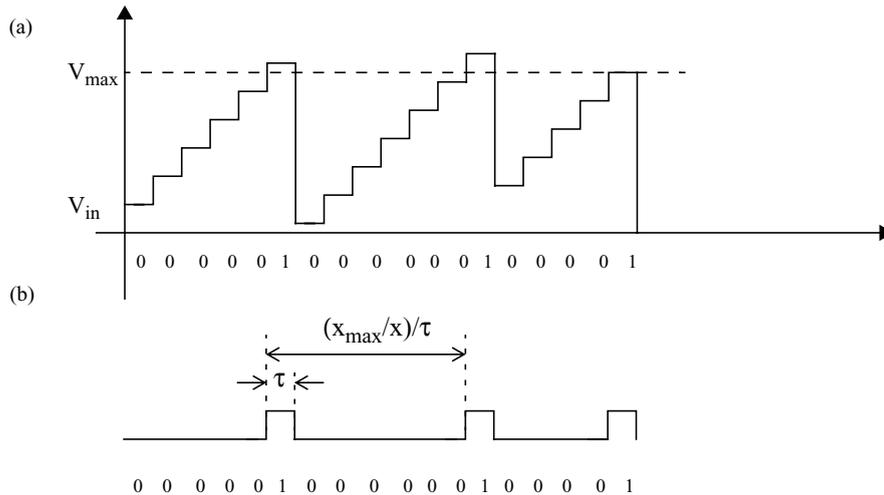


Fig. 6.26: Salidas (a) de un CA/D Σ - Δ ; y (b) la generada por un oscilador asíncrono

análogo, y un circuito de reset de pulso. En lugar de conectar la salida del comparador a un

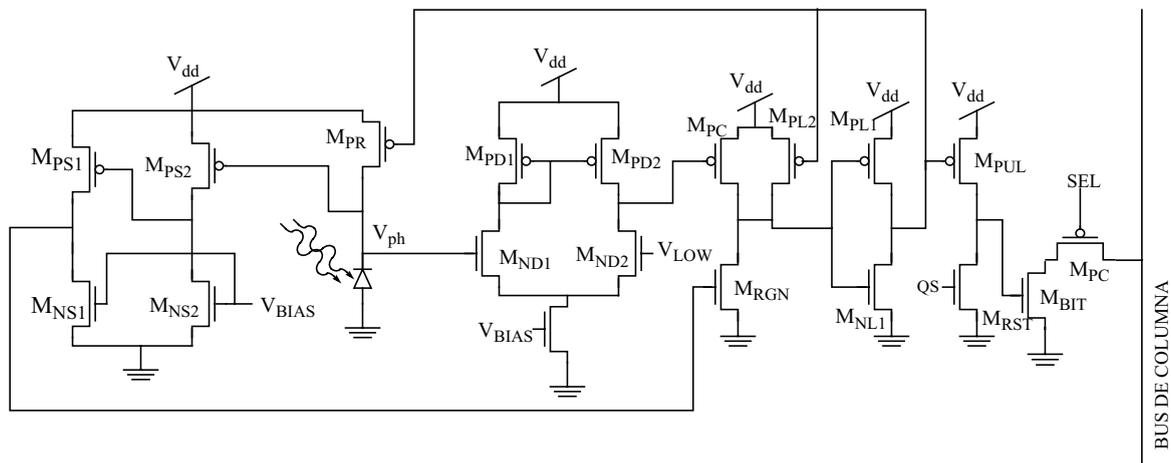


Fig. 6.27: Implementación de un CA/D Σ - Δ basado en un oscilador asíncrono [10]

bus, la información de si ha cambiado de estado la salida del comparador se almacena en la capacidad de puerta de un MOS. Este valor almacenado es entonces leído y reseteado a una frecuencia determinada por un reloj externo. Puesto que la lectura de este bit no afecta al oscilador, sólo la corriente fotogenerada en el dispositivo fotosensor determina la frecuencia del píxel.

La descripción del esquema de la Fig. 6.27 es el siguiente. La señal de entrada es la corriente fotogenerada en el dispositivo fotosensor. La tensión en el nodo de integración (V_{ph}) disminuye con el tiempo de forma proporcional al flujo luminoso incidente. Esta tensión será reseteada a V_{dd} cuando disminuya por debajo de un determinado nivel de referencia V_{low} . El circuito se compone de cuatro secciones: un amplificador diferencial, que continuamente compara la tensión del nodo V_{ph} y la tensión de referencia V_{low} ; un biestable, encargado de disparar el reset; una sección regenerativa que conmuta al biestable y recomienza la integración; y una lógica de captura del pulso, que almacena un bit justo antes del reset. Una señal de selección de fila SEL vuelca la salida del píxel sobre el BUS DE COLUMNA para proceder a la lectura del bit. Tras la lectura, la señal QS resetea el bit almacenado a GND.

6.3.3.- Implementación de la técnica MCBS a nivel de píxel

Las características de la técnica MCBS se adaptan muy bien para conseguir la implementación de la CA/D a nivel de píxel, pues requiere sólo de una celda de memoria. El diseño de la técnica MCBS a nivel de píxel propuesto en [4] se ilustra en la Fig. 6.28. Consta de 18 transistores. El comparador consiste en un amplificador de transconductancia seguido de un espejo de corriente Wilson y una carga cascode de salida. Se polariza para que opere en la región subumbral para maximizar la ganancia y minimizar el consumo. La celda de memoria opera como una celda DRAM de 2 transistores, donde M_1 es el transistor de selección de escritura y M_2 es el buffer de lectura.

Las no uniformidades del píxel se reducen al compartirse la circuitería que genera las señales globales, que constituyen a su vez los circuitos más complejos de la CA/D. La no uniformidad que introduce el detector tiene dos fuentes: la variación introducida por la realimentación a través de los elementos parásitos de las llaves analógicas; y el offset del comparador. El offset del comparador puede reducirse implementando la técnica de autocero, lo que se consigue almacenando el valor de offset en la capacidad de integración. Esta operación se lleva a cabo conectando el transistor M_4 durante un intervalo de tiempo una vez que la conversión se ha llevado a cabo.

Una propiedad interesante del algoritmo MCBS es que puede testearse electrónicamente. Para ello todos los fotodetectores se desconectan. La entrada de los detectores se conecta al valor de la señal RAMP deseado, simplemente conectando M_4 , y a continuación se lee la salida. De esta forma puede determinarse la función de transferencia del sensor variando la señal de entrada.

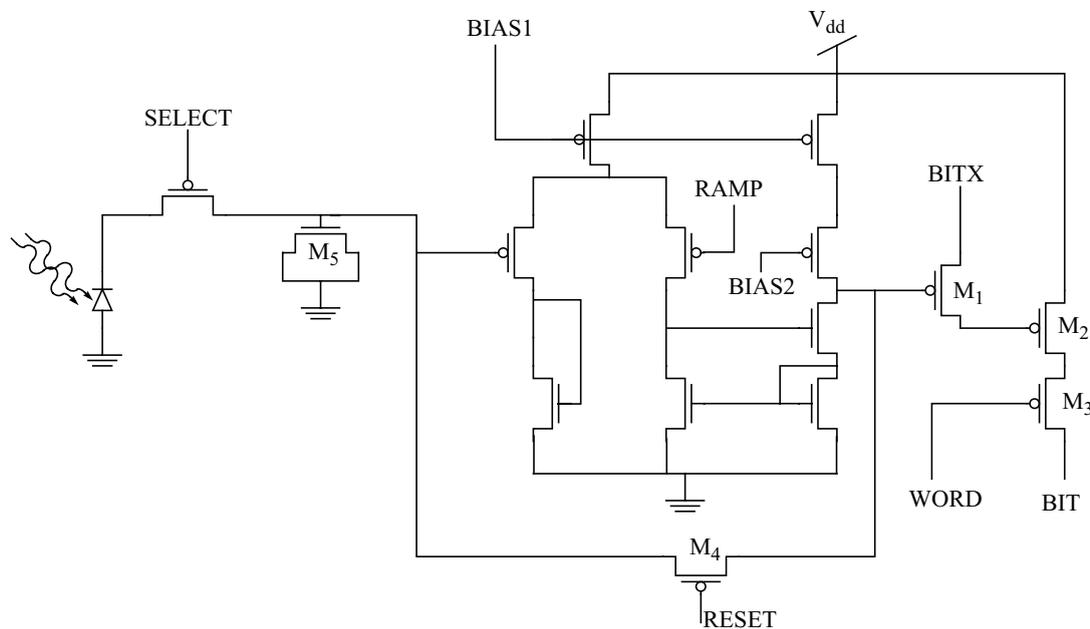


Fig. 6.28: Implementación del circuito detector para la CA/D MCBS

Las principales desventajas de este circuito son básicamente un ancho de banda del comparador moderado, y una sensibilidad del fotosensor reducida, producto de la capacidad de muestreo y retención M_5 . El ancho de banda moderado está causado por la limitación en área que

supone el diseño de la circuitería a nivel de píxel. La resolución máxima del CA/D está determinada por la ganancia del comparador, y la velocidad de conversión está fijada por la ganancia y el ancho de banda del comparador. Puede demostrarse que en el esquema propuesto, asumiendo que se trabaja en la región subumbral, y siendo C_d la capacidad de integración y C_{gs5} la capacidad puerta/fuente de M_5 , la tensión inducida en el fotosensor se atenúa por un factor $C_d/(C_d + C_{gs5})$ cuando se muestrea sobre la capacidad M_5 [6].

6.4.- OTROS ASPECTOS INTERESANTES SOBRE LA CA/D EN LOS SENSORES DE IMAGEN

En esta sección se describen dos aspectos sobre la conversión analógico-digital que resultan de interés desde el punto de vista de los sensores de imagen.

El primer aspecto que se desarrolla deriva de la técnica MCBS. Si bien en la Secc. 6.6.2.6 se ilustra esta técnica utilizando intervalos de la señal de entrada uniforme, la técnica de muestreo múltiple permite que los tiempos de exposición y el número de bits del formato de salida puedan elegirse libremente.

El segundo aspecto viene como consecuencia del muestreo múltiple. Consiste en la mejora del rango dinámico del sensor. Existen dos técnicas básicamente para la mejora del rango dinámico del sensor: ajuste de la capacidad de pozo; y uso del muestreo múltiple. En la Secc. 6.6.2.4 se describen y comparan ambos métodos.

6.4.1.- Técnica de muestreo múltiple en un sensor con CA/D MCBS a nivel de píxel

Un sensor de imagen con CA/D MCBS a nivel de píxel puede realizar un muestreo múltiple de una forma muy general. Puesto que las señales están disponibles para los CA/Ds en todo momento y todos los CA/Ds operan en paralelo, el número de muestras, los tiempos de exposición, y el número de bits del formato de salida pueden elegirse libremente. Vamos a considerar la implementación del muestreo múltiple usando tiempos de exposición crecientes de forma exponencial, es decir, $T, 2T, 4T, \dots, 2^k T$. Cada muestra del píxel se digitaliza en m bits. Las muestras digitalizadas para cada píxel se combinan en un número binario de $m+k$ bits, en un formato de punto flotante con un exponente en el rango $[0, k]$ y una mantisa de m bits. Esto incrementa el rango dinámico del sensor en un factor de 2^k , al mismo tiempo que proporciona una resolución de m bits para cada rango de exposición de la iluminación.

En el muestreo múltiple no es necesario realizar una lectura del píxel en cada muestreo. Las muestra de m bits para cada píxel puede combinarse en un número binario de $m+k$ bits simplemente escalando la última muestra de m bits antes de que el sensor sature por un factor de 2^{k+1-l} , donde $1 < l \leq k+1$ es el índice de la muestra. No es necesario sacar los m bits en cada muestra. De hecho, si la relación entre flujo luminoso y tensión es lineal, sólo es necesario leer un total de $m+k$ bits (mínimo requerido).

Para ilustrar cómo se salvan los bits durante la fase de lectura, consideremos un ejemplo con $k=2$ y $m=2$, es decir, se muestrea una imagen tres veces, en $T, 2T$ y $4T$, y se digitaliza cada muestra en 2 bits de resolución. En la Fig. 6.29 se muestra la evolución de la tensión en el dispositivo fotosensor a lo largo del periodo de integración para tres niveles de intensidad luminosa distintos (I_1, I_2 e I_3), asumiendo una relación de conversión de intensidad en tensión lineal. La tensión se muestrea en los tiempos de exposición $T, 2T$ y $4T$. Cada muestra se digitaliza con una resolución de 2 bits, usando un código binario, como se muestra en la Tabla 6.4. La pri-

mera muestra se digitaliza en x_1x_2 , es decir, 11 para I_1 . La segunda intensidad también se digitaliza en 2 bits. Puesto que el segundo periodo de muestreo es doble del primero, los dos bits en los que se codificará la muestra serán x_2x_3 si la muestra no satura el sensor, es decir, $V < V_{max}$, y 11 en el caso en el que sature. En tal caso, el CA/D sólo necesita generar el bit menos significativo x_3 . De forma similar, para la tercera muestra en $4T$, sólo el bit menos significativo necesita ser generado. La Tabla 6.5 muestra los valores de los bits para cada rango de entrada de iluminación y la correspondiente representación en punto flotante. Este esquema puede fácilmente extenderse para cualquier exponente k y mantisa m . La primera muestra se digitaliza en m bits; después, sólo el bit menos significativo es generado en cada muestreo consecutivo, y se van añadiendo a la primera muestra durante el diezmado. Por lo tanto, considerando que la relación entre intensidad luminosa y tensión es lineal, sólo es necesario leer $m+k$ bits, siendo esta cifra el mínimo requerido. Trabajando en este límite se reduce el ruido de lectura, aunque en general es necesario leer más bits para corregir no linealidades en el sensor, así como errores de offset y ruido.

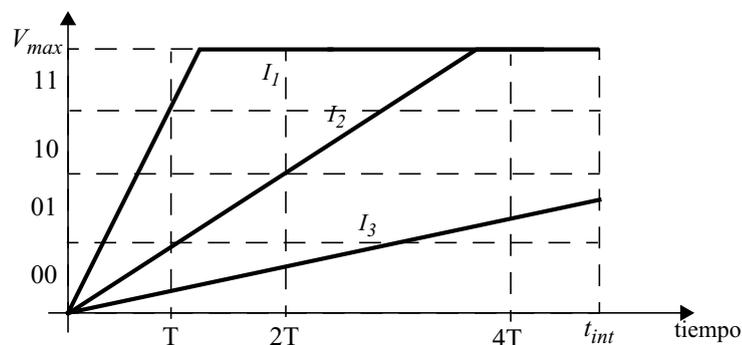


Fig. 6.29: Representación de la tensión en la capacidad de integración frente a tiempo durante la fase de integración para un esquema de muestreo múltiple

Iluminación	$x_1x_2x_3x_4$	Exponente	Mantisa
I_1	1 1 1 1	2	11
I_2	0 1 0 1	1	10
I_3	0 0 0 1	0	01

Tabla 6.4: Palabras de código para tres niveles de iluminación

La Tabla 6.5 proporciona la tabla completa de cuantización y la curva que relaciona el correspondiente nivel de iluminación con la palabra de código para el ejemplo de $m=2$ y $k=2$. Se usa código Gray en lugar de código binario para minimizar el error cometido en la codificación de las muestras. La curva de transferencia se muestra en la Fig. 6.30. El tamaño de los escalones crece exponencialmente, lo que indica que el CA/D tiene una resolución en punto flotante.

La Fig. 6.31 muestra el esquema del circuito implementado en [11]. Cada CA/D es multiplexado de forma que es compartido entre 4 píxeles. El comparador de 1 bit se ha descrito en la Secc. 6.6.2.6. Antes de la CA/D la carga recogida por el dispositivo fotosensor se muestrea en la capacidad de muestreo y retención M5. Tras la CA/D, la capacidad de muestreo y retención se resetea a través de la señal RST, que provoca que el comparador opere como un amplificador operacional con ganancia de realimentación unidad. El autocero se implementa durante la fase de reset, pues la tensión de offset del comparador se almacena en la capacidad del dispositi-

Rango de iluminación de entrada	$x_1x_2x_3x_4$	Número binario	DN	Exponente	Mantisa
(0, 1/16]	0 0 0 0	0 0 0 0	0	0	00
(1/16, 2/16]	0 0 0 1	0 0 0 1	1	0	01
(2/16, 3/16]	0 0 1 1	0 0 1 0	2	0	10
(3/16, 4/16]	0 0 1 0	0 0 1 1	3	0	11
(4/16, 6/16]	0 1 1 1	0 1 0 1	5	1	10
(6/16, 8/16]	0 1 0 0	0 1 1 1	7	1	11
(8/16, 12/16]	1 1 1 0	1 0 1 1	11	2	10
(12/16, 1]	1 0 0 0	1 1 1 1	15	2	11

Tabla 6.5: Relación entre niveles de iluminación, palabras de código y número digital (DN)

tivo fotosensor. Con esto se logra implementar el algoritmo CDS, por lo que se reduce considerablemente el FPN y el ruido flicker. El amplificador de sensado a nivel de columna mostrado en la Fig. 6.31 se diseña para que trabaje a alta velocidad, bajo ruido, y bajo consumo. Un amplificador de carga se usa para minimizar la variación de tensión en el bus de columna. Para sensar el bus de columna, en primer lugar se resetea la capacidad M_c , y la corriente del píxel y el transistor M7 se integran. El bit sensado se memoriza usando un flip-flop. El inversor triestado es parte de un multiplexor 10:1. Las señales necesarias para controlar la CA/D, incluidas las señales RAMP y BITX se generan de forma global a través de un CD/A off-chip y una circuitería de control digital.

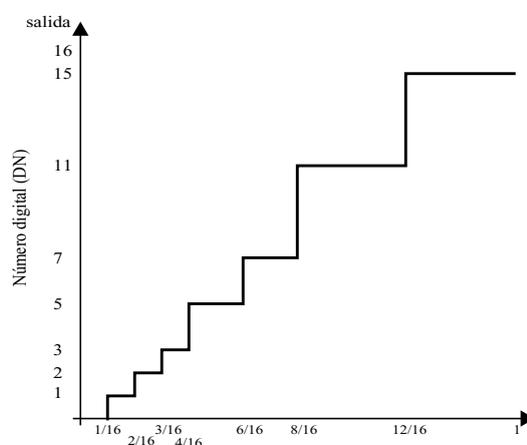


Fig. 6.30: Curva de transferencia nivel de iluminación-número digital correspondiente a la Tabla [6.4]

6.4.2.- Algoritmos de mejora del rango dinámico en sensores de imagen

Presentamos a continuación dos métodos para la mejora del rango dinámico de los sensores de imagen.

El primero se basa en el *ajuste de la capacidad de pozo*. El esquema de ajuste de la capacidad de pozo [12][13][14] comprime la curva de respuesta de la corriente del sensor frente a la carga integrada mediante el uso de una puerta lateral superpuesta, es decir, la puerta del transistor de reset en el caso del CMOS APS. La tensión aplicada a esta puerta determina la capacidad del pozo. Durante el tiempo de integración la capacidad de pozo crece de forma monótonica hasta

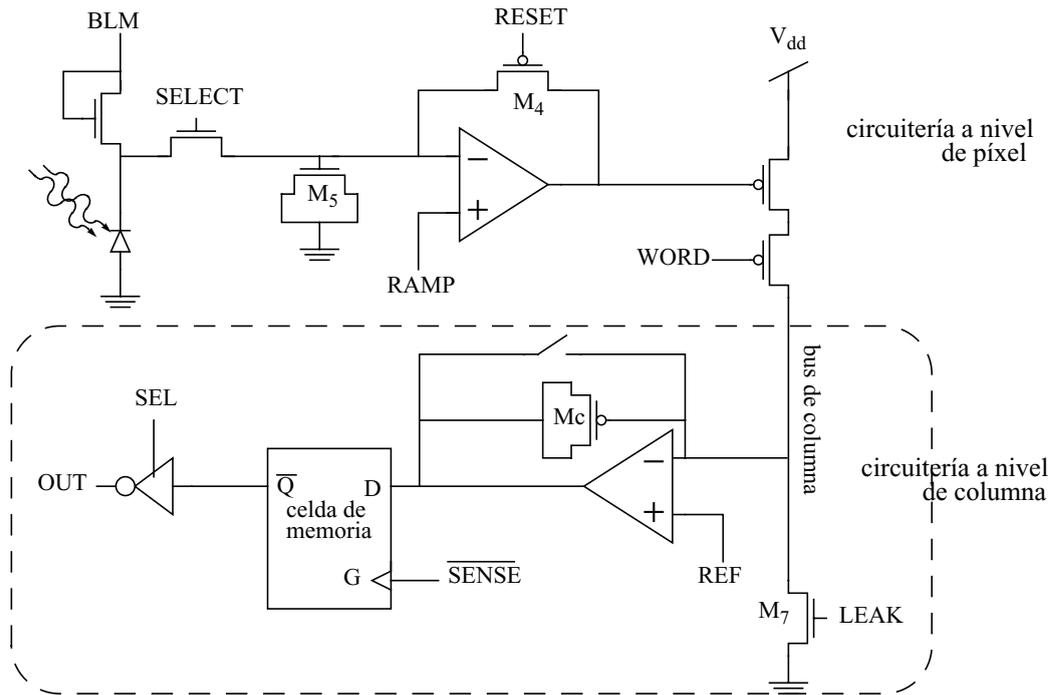


Fig. 6.31: Esquema de la circuitería a nivel de píxel y a nivel de columna para la implementación de la técnica de muestreo múltiple

su máximo valor. La carga fotogenerada que excede esta capacidad es drenada a través de la puerta superpuesta, de forma que se evita que la capacidad del dispositivo fotosensor sature.

El segundo de los métodos es el **muestreo múltiple** [15], descrito en la sección anterior. Este método se basa en medir el valor de la tensión del píxel en determinados instantes durante la fase de integración. Esto permite la implementación de una codificación binaria del dato de salida en punto flotante, lo que mejora considerablemente el rango dinámico.

Ahora bien, no sólo estamos interesados en hallar la mejora sobre el rango dinámico, sino cómo estas arquitecturas degradan la relación señal a ruido. Para este estudio se considerará un modelo simplificado del sensor y se propondrán expresiones del DR y de la SNR para comparar en qué grado afecta cada estructura a cada parámetro.

6.4.2.1.- Rango dinámico y SNR en sensores en tiempo discreto basados en integración de carga

En este apartado pretendemos hallar la expresión para el rango dinámico (**DR**) y la relación señal a ruido (**SNR**) en los sensores de imagen en tiempo discreto por integración de carga. En este modo de operación, durante la fase de integración tanto la corriente fotogenerada (i_{ph}) como la corriente oscura del fotosensor (i_{dark}) se integrarán sobre la capacidad de integración. Asumimos que se verifican las condiciones: $i_{ph} \geq 0$ e $i_{dark} \geq 0$ durante todo el periodo de integración. Se supone también que el sensor tiene una capacidad de carga limitada q_{max} . La Fig. 6.32 ilustra la carga recogida frente al tiempo en dos situaciones: una primera en la que la carga integrada durante la fase de integración no alcanza el límite de carga q_{max} , por lo que la salida del sensor es la correcta; y una segunda situación en la que la capacidad de integración se satura, sobrepasando el límite de carga q_{max} , por lo que el valor de salida no es el correcto.

La primera de las situaciones se corresponde con un nivel de iluminación del píxel bajo, mientras que la segunda se corresponde con una iluminación elevada del píxel.

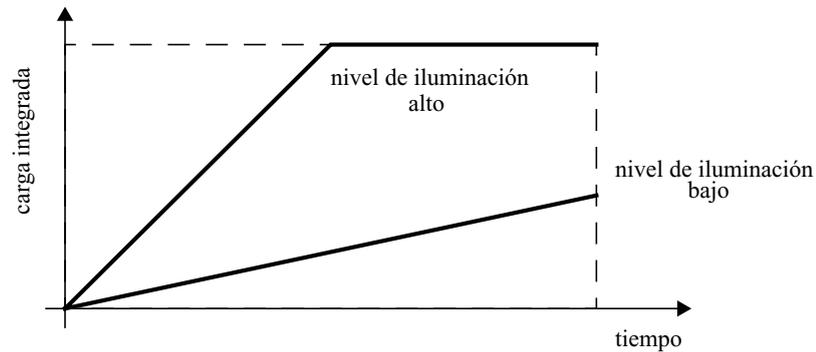


Fig. 6.32: Representación de la carga integrada frente a tiempo durante la fase de integración

En esta sección introducimos un modelo simplificado que relaciona la fotocorriente y la tensión de salida del sensor, y lo usamos para analizar el rango dinámico y el SNR para sensores en tiempo discreto que operan en modo de integración de carga sin mejora del rango dinámico. Este mismo modelo se empleará para analizar el rango dinámico y el SNR cuando se usan esquemas de mejora del rango dinámico.

El modelo se muestra en la Fig. 6.33 [6]. La fuente de corriente $I_s(t)$ representa el ruido de disparo debido a las corrientes fotogenerada y oscura, y se modela como un ruido blanco gaussiano con una densidad espectral de potencia:

$$S_{I_s}(f) = 2q(i_{ph} + i_{dark})$$

La carga acumulada Q al final del periodo de integración es una función $f[\]$ de la corriente total $I(t)$ durante la fase de integración ($0 \leq t \leq t_{int}$). Para el sensor en modo integración de carga implementado sin mejora del rango dinámico esta función $f[\]$ es simplemente:

$$f = \min \left\{ \int_0^{t_{int}} I(t) dt, q_{max} \right\} \quad (6.2)$$

Se asume que la relación entre carga y tensión es lineal, con un factor de amplificación g . La carga añadida Q_r representa el ruido debido a los circuitos de lectura, lo que incluye el ruido referido a la entrada del amplificador y el ruido de reset del APS. En el caso de llevarse a cabo la CA/D incluye también el ruido cuantización. Se asume que el ruido total tiene media cero y una potencia dada por σ_r^2 . Para simplificar el modelo se ignora el ruido de patrón fijo.

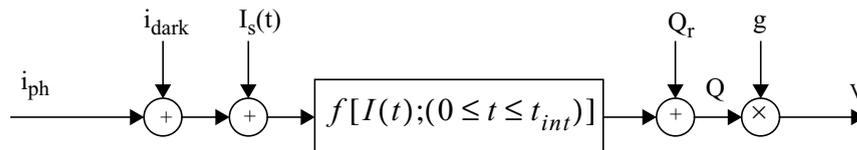


Fig. 6.33: Esquema de bloques del sensor para el cálculo del DR y del SNR

Basándose en el modelo definido en la Fig. 6.33, la expresión del rango dinámico (DR) del sensor operando en modo de integración de carga viene dada por [6]:

$$DR = \frac{i_{max}}{i_{min}} = \frac{q_{max} - i_{dark}t_{int}}{\sqrt{q(i_{dark}t_{int}) + \sigma_r^2}} \quad (6.3)$$

La relación señal a ruido, definida como el cociente entre la potencia de la señal de entrada i_{ph}^2 y la potencia media de ruido referida a la entrada, i_{ph} queda, en función de la corriente fotogenerada [15]:

$$SNR(i_{ph}) = \frac{(i_{ph}t_{int})^2}{q(i_{ph} + i_{dark})t_{int} + \sigma_r^2} \quad (6.4)$$

expresión válida para $i_{ph} \leq i_{max}$. El SNR no se define para $i_{ph} > i_{max}$, pues la distorsión introducida por la saturación del sensor complicaría el desarrollo matemático sin ofrecer más información de la necesaria para esta discusión.

6.4.2.2.- Mejora del rango dinámico mediante el ajuste de la capacidad del pozo

Como se ha comentado, este método se basa en llevar a cabo un drenaje de la corriente que carga la capacidad de integración cuando ésta es suficiente como para saturar el dispositivo. Asumiendo el caso más sencillo en el que la capacidad de pozo se ajusta sólo una vez desde el momento en el que la corriente fotogenerada alcanza el valor $q_{max} \cdot \theta$ hasta el instante t_1 . θ expresa la proporción de la carga q_{max} en la cual se realiza el ajuste de la capacidad de pozo. La Fig. 6.34 muestra la carga media recogida frente al tiempo para dos valores de fotocorriente. Puede verse que cuando la carga recogida alcanza $q_{max} \cdot \theta$ en el ejemplo de una iluminación alta, la carga de salida queda fijada hasta t_1 .

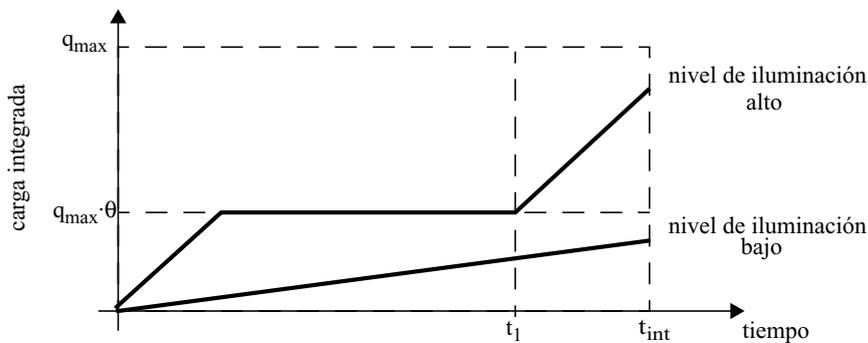


Fig. 6.34: Representación de la carga integrada frente a tiempo durante la fase de integración mediante el esquema de ajuste de la capacidad del pozo

Basándose en el modelo de fotosensor de la Fig. 6.33, las expresiones a las que se llega con este esquema de mejora del SNR son las siguientes [15]:

$$SNR(i_{ph}) = \begin{cases} \frac{(i_{ph}t_{int})^2}{q(i_{ph} + i_{dark})t_{int} + \sigma_r^2}; & 0 \leq i_{ph} < \frac{q_{max} \cdot \theta}{t_1} - i_{dark} \\ \frac{i_{ph}^2(t_{int} - t_1)^2}{q(i_{ph} + i_{dark})(t_{int} - t_1) + \sigma_r^2}; & \frac{q_{max} \cdot \theta}{t_1} - i_{dark} \leq i_{ph} < \frac{q_{max} \cdot (1 - \theta)}{t_{int} - t_1} - i_{dark} \end{cases} \quad (6.5)$$

Para valores de i_{dark} pequeños, el rango dinámico mejora en un factor [15]:

$$DRF = \frac{1 - \theta}{1 - t_1/t_{int}} \quad (6.6)$$

La Fig. 6.35 [15] muestra la relación entre SNR e i_{ph} para una simulación en la que se toman como parámetros: $q_{max}=125000e^-$; $\sigma_r=20e^-$; $t_{int}=30ms$; $i_{dark}=1fA$; $\theta=7/8$; y $t_1/t_{int}=255/256$. Para estos parámetros resulta $DRF=32$.

El análisis puede fácilmente extenderse a un número k de ajustes de la capacidad del pozo.

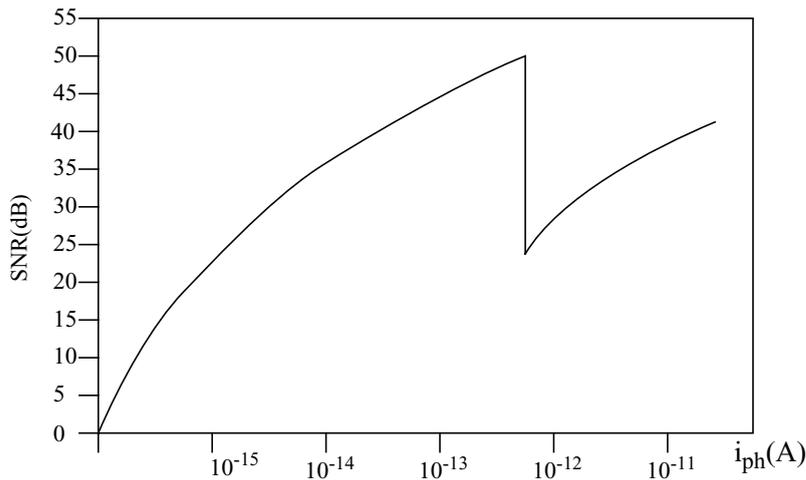


Fig. 6.35: SNR vs. i_{ph} para el esquema de ajuste de la capacidad de pozo

Puede observarse que el rango dinámico se amplía, mientras que el $SNR(i_{max})$ se degrada en un factor $1-\theta$ relativo al valor de pico del SNR de la arquitectura que no implementa la mejora del rango dinámico.

6.4.2.3.- Mejora del rango dinámico mediante la técnica de muestreo múltiple

La técnica de muestreo doble se ha aplicado para mejorar el rango dinámico en los sensores CCDs, y en los CMOS APS. En esta técnica la imagen se toma dos veces, una vez con un corto periodo de integración y otra tras un tiempo de integración mucho mayor, y las dos imágenes se combinan para mejorar el rango dinámico. Conceptualmente, el tiempo de integración corto captura las áreas de alta iluminación antes de que el pozo se sature y los tiempos de integración altos capturan las áreas de baja iluminación después de un tiempo de integración adecuado para estas condiciones. Las dos imágenes, sin embargo, no resultan suficientes para representar las zonas de la escena que son demasiado oscuras para que sean capturadas en la primera imagen, o demasiado luminosas para ser capturadas en la segunda. Yang et al. [4] muestran que el CA/D a nivel de píxel se ajusta de forma ideal para la implementación de algoritmos que usen múltiples muestras. Se considera la implementación de un muestreo múltiple para tiempos de integración crecientes de forma exponencial. En este caso, el rango dinámico se amplía en un factor 2^k y la imagen combinada tiene una resolución en punto flotante con un exponente k , donde k es el número de veces que se lleva a cabo el muestreo.

Utilizando el modelo del sensor propuesto en la Fig. 6.33, y considerando el caso sencillo de un doble muestreo en los instantes t_{int}/a y t_{int} , para $a>1$. La Fig. 6.36 muestra la carga media recogida frente al tiempo para tres niveles de iluminación. Puede observarse que mediante el

muestreo en t_{int}/a , la señal de iluminación moderada puede muestrearse antes de que el sensor sature.

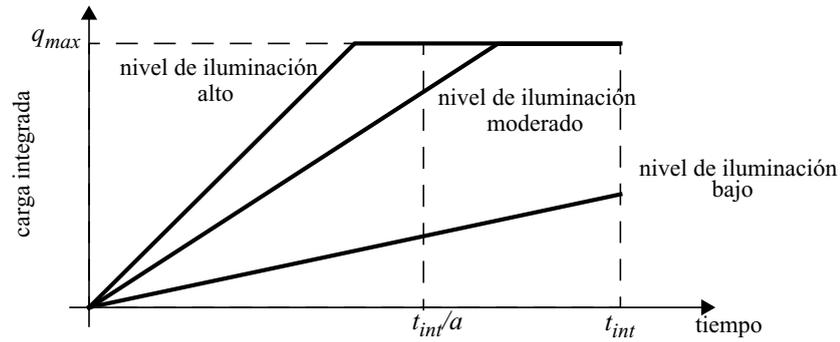


Fig. 6.36: Representación de la carga integrada frente a tiempo durante la fase de integración para un esquema de doble muestreo

Las expresiones de la SNR a la que se llegan son [15]:

$$SNR(i_{ph}) = \begin{cases} \frac{(i_{ph}t_{int})^2}{q(i_{ph} + i_{dark})t_{int} + \sigma_r^2}; & 0 \leq i_{ph} < \frac{q_{max}}{t_{int}} - i_{dark} \\ \frac{i_{ph}^2 \left(\frac{t_{int}}{a}\right)^2}{q(i_{ph} + i_{dark})\frac{t_{int}}{a} + \sigma_r^2}; & \frac{q_{max}}{t_{int}} - i_{dark} \leq i_{ph} < \frac{a \cdot q_{max}}{t_{int}} - i_{dark} \end{cases} \quad (6.7)$$

El rango dinámico mejora en un factor [15]:

$$DRF = \frac{\frac{a \cdot q_{max}}{t_{int}} - i_{dark}}{\frac{q_{max}}{t_{int}} - i_{dark}} \approx a \quad (6.8)$$

bajo la condición de i_{dark} pequeña.

La Fig. 6.37 [15] muestra la curva SNR frente a i_{ph} en las mismas condiciones que la curva de la Fig. 6.35. Mientras que el factor de mejora del rango dinámico es el mismo ($DRF=32$), la caída de la SNR en t_{int}/a es menor que la caída de la SNR en el momento del ajuste de la capacidad de pozo. Es decir, con el método de muestreo múltiple el rango dinámico del sensor se amplía, perjudicando en menor medida la SNR en comparación con el ajuste de la capacidad de pozo.

Para finalizar, la Fig. 6.38 [15] muestra la comparación entre los dos métodos de mejora del rango dinámico en el caso de múltiples ajustes de la capacidad de pozo y de muestreo múltiple. Las curvas se obtienen de nuevo en las mismas condiciones que las anteriores, además de las siguientes consideraciones.

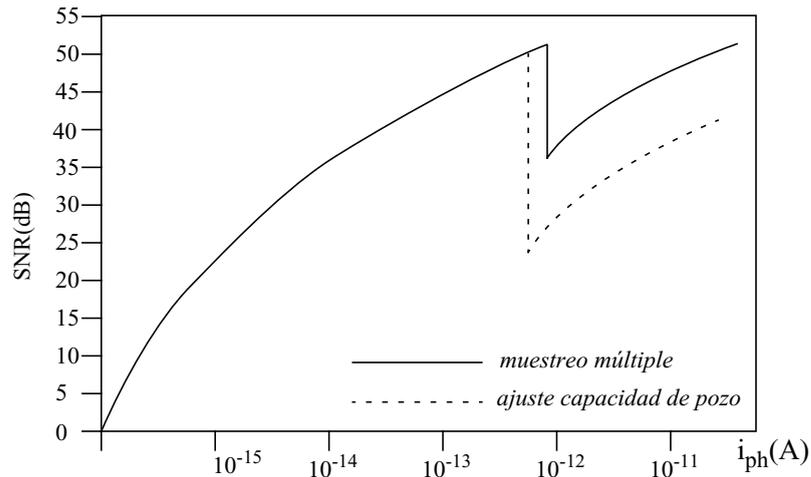


Fig. 6.37: Comparación de la curva SNR vs. i_{ph} entre el esquema de ajuste de la capacidad de pozo y el muestreo múltiple

Para el ajuste múltiple de la capacidad de pozo se toma un número de ajustes $k=8$, unos niveles de capacidad de ajuste $\theta_i=i/10$, y unos tiempos de ajuste:

$$t_i = 1 - \frac{1}{10 \cdot 2^{i-1}} \quad (6.9)$$

para $i=1, 2, \dots, 8$. Se obtiene un factor de mejora $DRF=256$, es decir, unos 48 dB. La suma de todas las caídas del SNR es de 31 dBs, y el $SNR(i_{max})$ está 7dB por debajo de la curva de la arquitectura sin mejora del rango dinámico.

El muestreo múltiple se lleva a cabo utilizando como parámetro $k=8$. Se obtiene una mejora del rango dinámico $DRF=256$. En la Fig. 6.38 se observa que el muestreo múltiple logra una mejora en torno a los 10 dBs.

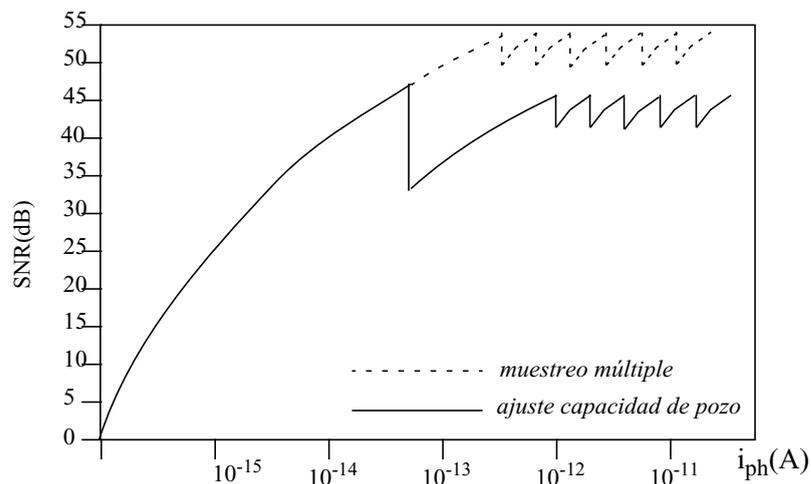


Fig. 6.38: Comparación de la curva SNR vs. i_{ph} entre el esquema de ajuste múltiple de la capacidad de pozo y el muestreo múltiple (ambos para $k=8$)

Referencias



- [1] Richard Hornsey "Design and Fabrication of Integrated Image Sensors" Electrical and Computer Engineering, University of Waterloo, Waterloo, Ontario, Canada
- [2] J. Candy, G. C. Temes "Oversampling Sigma-Delta Converters" IEEE Press, 1992
- [3] F. Medeiro, A. P3rez-Verd3, A. Rodr3guez-V3zquez "Top-Down Design of High-Performance Sigma-Delta Modulators", Kluwer Academic Publishers
- [4] Boyd Fowler, Abbas El Gamal, and David Yang "A Nyquist Rate Pixel Level ADC for CMOS Image Sensors" In Proceedings of IEEE 1998 Custom Integrated Circuits Conference, pp.237-240, Santa Clara, CA May 1998
- [5] Behzad Razavi "Principles of Data Conversion System Design", IEEE Press, 1995
- [6] Boyd Fowler, Abbas El Gamal, and David Yang "Techniques For Pixel Level Analog To Digital Conversion" In Proceedings of SPIE, Infrared Readout Electronics IV, pp. 2--12, Vol. 3360, Orlando, April 1998
- [7] B. Fowler "CMOS Area Image Sensors with Pixel Level A/D Conversion" PhD thesis, Stanford University, 1995
- [8] J. Candy, G. Temes "Oversampled Delta-Sigma Data Converters" IEEE Press, 1992
- [9] J. C. Candy, O. J. Benjamin "The Structure of Quantization Noise form Sigma-Delta Modulation" IEEE Trans. Commun., Vol. COM-29, pp. 1316-1323, Sept. 1981
- [10] Lisa G. McIlrath "A Low Power-Noise Ultrawide-Dynamic-Range CMOS Imager with Pixel-Parallel A/D Conversion", IEEE Journal of Solid-State Circuits, Vol. 36, No. 5, May 2001
- [11] D. Yang, A. El Gamal, B. Fowler, H. Tian "A 640x512 Image Sensor with Ultrawide Dynamic Range Floating-Point Pixel-Level ADC" IEEE Journal of Solid-State Circuits, Vol. 34, No. 12, Dec. 1999
- [12] T. F. Knight "Design of an Integrated Optical Sensor with On-Chip Preprocessing" PhD thesis, MIT, 1983
- [13] M. Sayag "Non-linear Photosite Response in CCD Imagers" U. S. Patent No. 5,055,667, 1991. Filed 1990
- [14] S. Decker, R. MacGrath, K. Brehmer, C. Sodini "A 256x256 CMOS imaging array with ultra-wide dynamic range pixels and column-parallel digital output" in ISSCC Digest of Technical Papers, pp.176-177 (San Francisco, CA), Feb. 1998
- [15] D. Yang, A. El Gamal "Comparative Analysis of SNR for Image Sensors with Enhanced Dynamic Range" In Proceedings of the SPIE Electronic Imaging '99 conference, Vol. 3649, San Jose, CA, January 1999

RESUMEN Y LÍNEAS FUTURAS DE DESARROLLO

En este proyecto se han abordado los aspectos fundamentales del diseño de sensores de imagen en tecnologías CMOS. Para ello, se han repasado los fenómenos de absorción óptica en semiconductores, y los fundamentos del diodo de unión como elemento sensor fundamental. Se ha seguido con un repaso a las características esenciales de los dispositivos fotosensibles disponibles en estas tecnologías, a las figuras de mérito relevantes en los sistemas de adquisición, a los procedimientos de adaptación de la señal fotogenerada más comunes, y a los problemas de ruido espacial y temporal como dificultades más significativas del diseño de este tipo de sistemas. Finalmente, se han estudiado los procedimientos de conversión analógica-digital apropiados para estos propósitos. Todo esto constituye un inmejorable punto de partida para la continuación de este trabajo hacia objetivos aún más ambiciosos.

Son generalmente conocidas las ventajas e inconvenientes de los sensores de imagen basados en tecnologías CMOS convencionales, frente a la alternativa basada en dispositivos acoplados por carga (CCD), y por tanto no se repetirán aquí. También es ampliamente conocida la evolución actual de la tecnología de sensores de imagen basados en semiconductores: la predominancia comercial de las soluciones CCD está dando paso últimamente a una incorporación progresiva de las soluciones CMOS, en ciertas aplicaciones. Previsiblemente, esta tendencia continuará, debido fundamentalmente a la continua progresión en el escalado tecnológico y a la consecuente tendencia a la integración de sistemas completos (sensor, procesador, actuador), a la necesidad de reducir el consumo de potencia, y a la conveniencia de reducir los costes de producción. Debe remarcar especialmente que, con toda probabilidad (ya hay algunos ejemplos reportados), el futuro cercano nos muestre dispositivos sensores unidos a elementos de procesamiento, e incluso actuadores, en un mismo sustrato. En este sentido, la continuación de este trabajo se dirigirá, además de a intentar contribuir de alguna manera al perfeccionamiento de los sensores de imagen CMOS en sí, a estudiar y perfeccionar el funcionamiento de los mismos en cooperación con circuitería de procesamiento, en un mismo chip, para la realización de tareas de visión artificial.

SENSORES DE IMAGEN EN TECNOLOGÍAS CCD

Los dispositivos por acoplamiento de carga o CCD basan su funcionamiento en el almacenamiento de la carga fotogenerada por el flujo de luz incidente en los elementos capacitivos contruidos a partir de la unión Metal-Óxido fino-Semiconductor, o bien Polisilicio-Óxido fino-Semiconductor. La Fig. A.1 muestra el elemento capacitivo que se utiliza normalmente para la implementación del dispositivo CCD. Este elemento se denomina capacidad de canal enterrado, y se caracteriza porque la carga fotogenerada se recoge en la región de deplexión (o pozo de potencial) formada por una unión pn alejada de la interfaz semiconductor-óxido fino. La generación de la región de recogida de carga lejos de la interfaz tiene ventajas sobre los dispositivos que generan dicha región cerca de la superficie, tales como que la corriente oscura es mucho menor.

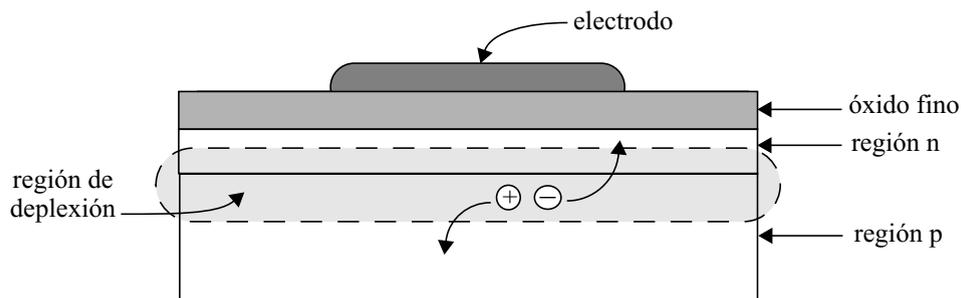


Fig. A.1: Esquema de capacidad de canal enterrado y desplazamiento de un par electrón-hueco fotogenerado en en la región de deplexión

El pozo de potencial está generado mediante la correcta polarización de uno de los terminales del dispositivo (electrodo). Al colocar a una determinada tensión este electrodo se genera por debajo de éste una región de deplexión cuya capacidad es la que se utiliza para recoger la carga fotogenerada por el flujo de luz incidente. Además de este electrodo, existen otros componentes que se utilizan para delimitar la región de recogida de carga, de forma que la carga recogida en dicha región de deplexión no escape de ella, evitando que el paquete de carga fotogenerada se degrade. La Fig. A.2 muestra el esquema de la superficie de un píxel CCD.

Los elementos que limitan la región de carga espacial son:

- en dirección horizontal, nos encontramos con los **delimitadores de canal**. La Fig. A.3 muestra el corte transversal del píxel CCD en esta dirección. Esta limitación de la estructura limita el canal enterrado que forma el dispositivo CCD. Estas regiones limitadoras normalmente tienen regiones de difusión tipo p fuertemente dopadas por debajo de la capa de óxido de campo y debido a la combinación del fuerte dopado y del ancho del óxido de campo, el campo electrostático por debajo de la puerta en esta región es relativamente insensible a la tensión o a cambios

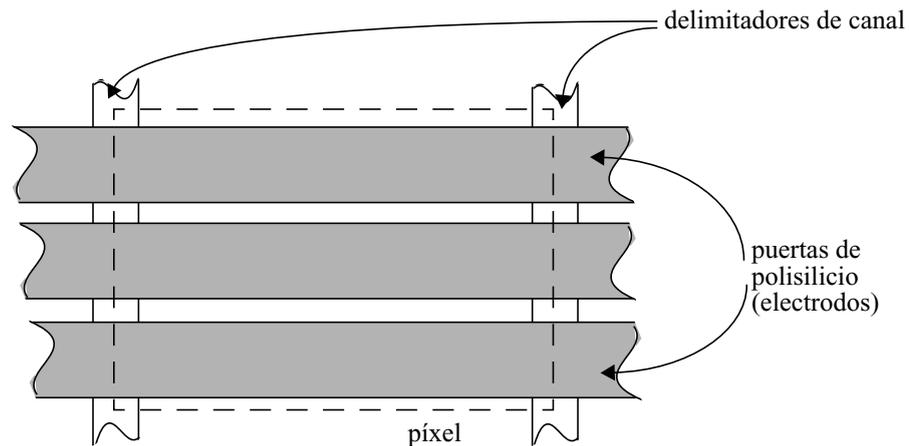


Fig. A.2: Esquema de la superficie de un píxel CCD

en la tensión de puerta. En este sentido, la carga que se recoge en el dispositivo de canal enterrado puede quedar confinada sobre la región por debajo de la capa de óxido fino.

- en dirección vertical no es necesario añadir delimitadores de canal, pues la correcta polarización de los **electrodos vecinos** anula el pozo de potencial por debajo de ellos, con lo que la región de depleción queda limitada a la zona justo por debajo del electrodo que la genera. La Fig. A.4 muestra el corte transversal del píxel en dirección vertical.

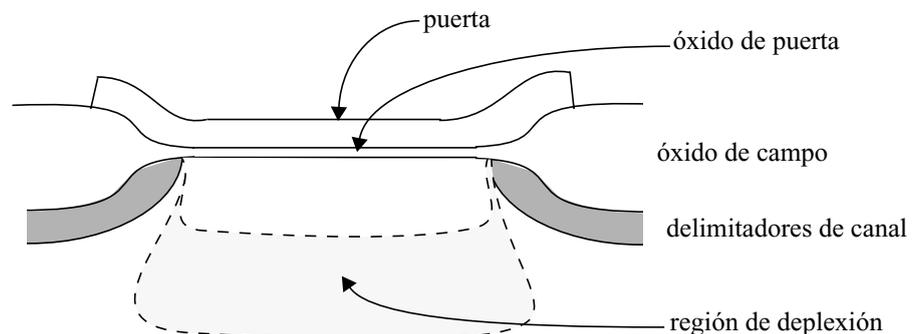


Fig. A.3: Corte transversal horizontal del píxel CCD donde se muestran las regiones delimitadoras de canal

A.1- PÍXEL CCD

La estructura de un píxel CCD simple se ilustra en la Fig. A.2. Esta figura muestra tres puertas de polisilicio orientadas perpendicularmente a dos regiones de canal delimitadoras. Entre las regiones de canal delimitadoras se encuentra el canal enterrado. Si el potencial en el electrodo intermedio es más positivo que el aplicado a las otras dos puertas, se formará un mínimo local de energía potencial bajo la puerta intermedia. Cuando los fotones chocan contra el píxel, se generarán pares electrón-hueco debido al efecto fotoeléctrico. Los electrones creados dentro de la región de potencial mínimo recogerán en esta región. Los electrones que se generan cerca de las regiones de canal delimitadoras y lejos de la región de potencial se difundirán hacia el potencial mínimo y también serán recogidos. En ambos casos, los huecos se difunden y se recogen en el sustrato tipo p. La cantidad de carga recogida en el pozo está relacionada linealmente con la intensidad del flujo de fotones y con el tiempo durante el cual la luz incide sobre el píxel (tiempo de integración).

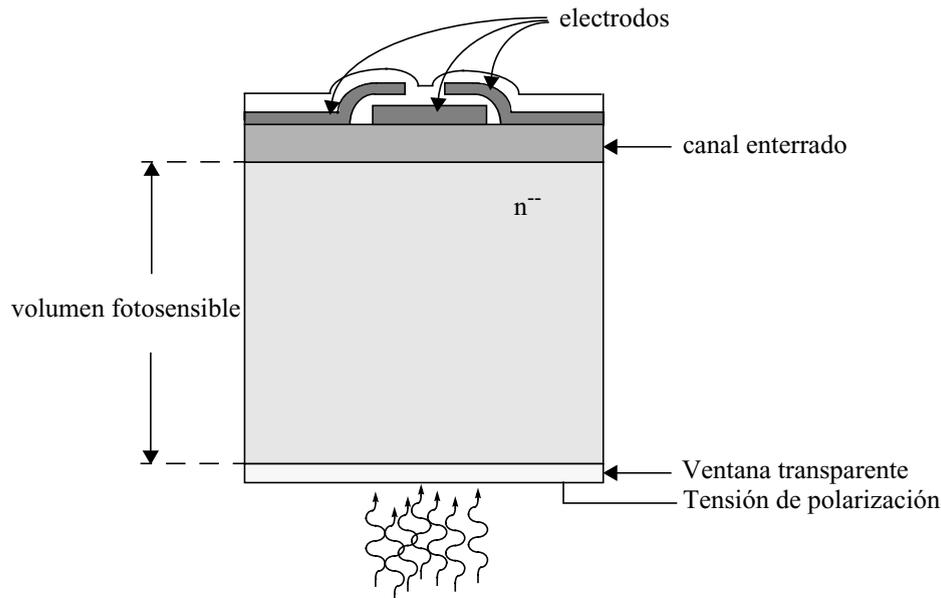


Fig. A.4: Esquema del corte transversal en sentido vertical del píxel CCD

Existen otros formatos de píxel CCD. Por ejemplo, estructuras que utilizan dos puertas de polisilicio para definir un píxel y algunas que utilizan cuatro. Sin embargo, la tecnología más utilizada en la fabricación de CCDs en una escala científica es la estructura de tres fases ilustrada en la Fig. A.4. La prevalencia de la tecnología de tres fases se debe principalmente al mayor yield y tolerancia en el proceso de esta tecnología.

A.2- PROCESO DE TRANSFERENCIA DE CARGA

Un dispositivo por carga acoplada consiste en una matriz regular de píxeles individuales. La Fig. A.5 ilustra un registro lineal de píxeles, cada uno con tres puertas separadas, referidas como Phase1, Phase2 y Phase3 (P1, P2, P3). Como se muestra en el registro lineal, todas las puertas P1 están conectadas a la misma línea de reloj. Igual ocurre con P2 y P3.

Un píxel CCD no tiene sentido definido de forma individual para la lectura de la cantidad de carga acumulada en su pozo de potencial cerca del electrodo de integración. El proceso de lectura de esta señal en carga implica mover el paquete desde el sitio en el que se ha recogido la carga hasta un amplificador de carga colocado al final de la línea. Éste es el proceso de transferencia de cara ilustrado en la Fig. A.5. Se asume que la carga es recogida debajo de los electrodos P2. En t_1 , se aplica una tensión positiva al electrodo P3 igual a la aplicada a P2, con lo que se encienden los electrodos P3. La carga localizada debajo de P2 se redistribuye hacia la región generada por la polarización de P3 mediante procesos de difusión y arrastre. En t_2 se desconecta la tensión aplicada a P2. La carga acumulada debajo de P2 se moverá rápidamente a la región generada debajo del electrodo P3. La carga se ha desplazado del electrodo P2 al electrodo P3.

Repetiendo el proceso para P3 y P1 se consigue mover la carga hasta la región generada debajo de P1. El ciclo se repite cíclicamente, de forma que todos los paquetes de carga se mueven de forma simultánea de un píxel hacia el que tienen a la derecha. El píxel situado en el extremo derecho volcará su carga en un amplificador de carga, mediante el cual se medirá el número de electrones que han sido recogidos en el correspondiente píxel. Serán necesarios N ciclos de reloj para realizar la lectura del registro lineal de N píxeles.

En todo momento en cada píxel individual existe un pozo y una barrera con los píxeles colindantes. Estos dos elementos son necesarios para almacenar la carga y que el paquete de carga no se degrade. Puede observarse que si se intercambian los papeles de dos de los electrodos la carga se desplazará hacia la izquierda.

La efectividad con la cual la transferencia de carga se lleva a cabo se mide con un parámetro denominado CTE (Charge Transfer Efficiency). Típicamente la carga se transfiere con una efectividad por encima del 99.999% por píxel.

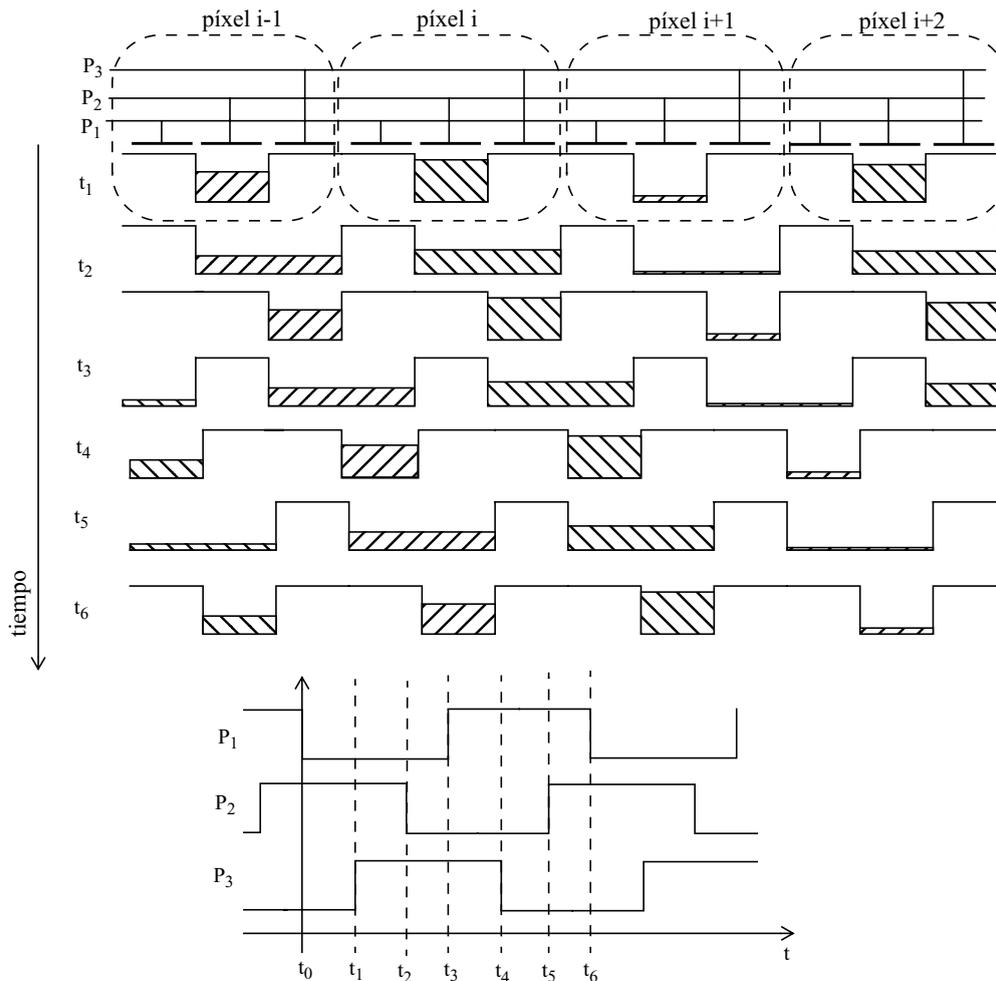


Fig. A.5: Esquema de transferencia de carga para una línea de píxeles CCD

MODELO MATEMÁTICO DEL RUIDO

El ruido se modela generalmente como una variable aleatoria continua, y las formas de onda de ruido se modelan como procesos aleatorios. Una variable aleatoria continua X puede caracterizarse por su **función de densidad de probabilidad (pdf)**:

$$f_x(x) > 0, \int_{-\infty}^{\infty} f_x(x) dx = 1 \quad (\text{b.1})$$

Conocida la pdf se pueden calcular todos los momentos de la variable aleatoria. Los dos momentos más importantes son la **media** o **valor medio**:

$$\bar{X} = \int_{-\infty}^{\infty} x f_x(x) dx \quad (\text{b.2})$$

Y el **valor cuadrático medio** de la variable:

$$\overline{X^2} = \int_{-\infty}^{\infty} x^2 f_x(x) dx \quad (\text{b.3})$$

El valor cuadrático medio representa la potencia de la señal X . La raíz cuadrada del valor cuadrático medio (denotado como RMS) representa una señal constante de potencia equivalente a la potencia media de X .

De la media y del valor cuadrático medio podemos calcular la **varianza** de X :

$$\sigma_x^2 = \overline{X^2} - (\bar{X})^2 \quad (\text{b.4})$$

que se interpreta como la distancia cuadrática de X a su valor medio. Cuando X tiene valor medio cero, su varianza coincidirá con su valor cuadrático medio. La varianza se utiliza, por lo general, para estimar la potencia de ruido.

La variable aleatoria más usada para describir el ruido es la **variable aleatoria gaussiana**. Con esta variable aleatoria se describen las fuentes de ruido objeto más importantes: el ruido térmico, el ruido de disparo, y el ruido flicker. La pdf de la variable aleatoria gaussiana tiene forma de campana, como se muestra en la Fig. B.1.

La función de densidad de la variable aleatoria gaussiana viene dada por:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma_x^2}} e^{-\frac{(x-\bar{X})^2}{2\sigma_x^2}}, \quad -\infty \leq x \leq \infty \quad (\text{b.5})$$

En el análisis de ruido, también necesitamos conocer la relación entre las muestras de ruido muestreadas en diferentes momentos. Esto se modela mediante **variables aleatorias de varias variables**, que tendrán asociadas una función de distribución de probabilidad conjunta. En el caso más simple de dos variables aleatorias X e Y , la función de probabilidad conjunta viene dada por:

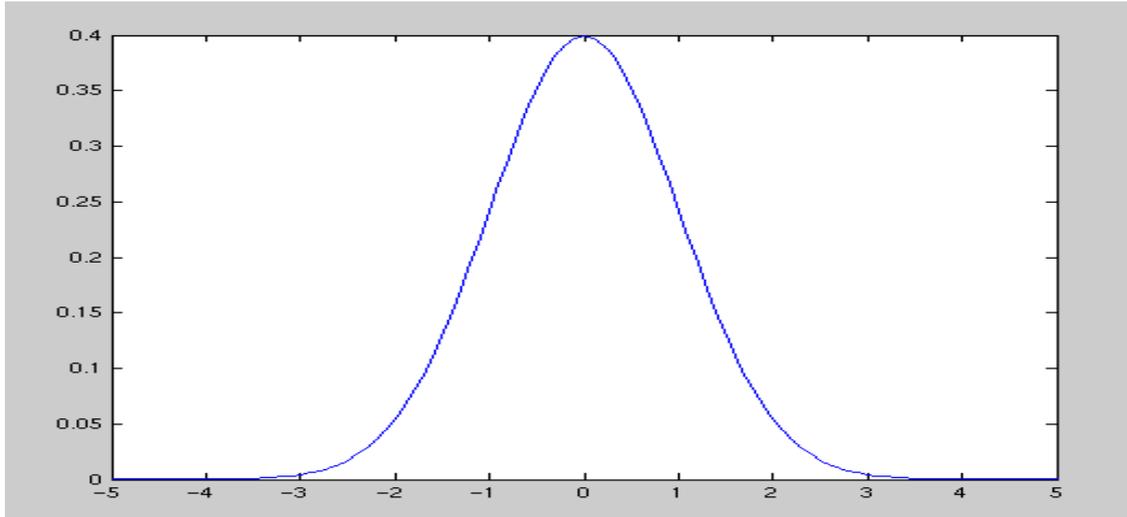


Fig. B.1: Función de densidad de probabilidad gaussiana $N(0,1)$

$$f(x, y) > 0, \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_x(x, y)(dx)dy = 1 \quad (\text{b.6})$$

La relación entre las dos variables aleatorias viene descrita por la función de autocorrelación:

$$\overline{XY} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy f_x(x, y)(dx)dy \quad (\text{b.7})$$

Si se verifica la condición: $\overline{XY} = \bar{X} \cdot \bar{Y}$ entonces X e Y están incorreladas. Si $f(x, y) = f_X(x) \cdot f_Y(y)$ entonces las variables aleatorias son independientes. Es fácil de demostrar que dos variables aleatorias independientes son incorreladas. La implicación inversa no es cierta.

Un proceso aleatorio $X(t)$ se usa para modelar la forma de onda del ruido. El ruido se considerará como un conjunto infinito de variables aleatorias muestreadas en tiempo. Para los instantes t_1, t_2, \dots, t_n , las muestras $X(t_1), X(t_2), \dots, X(t_n)$ serán variables aleatorias. En el análisis del ruido, será importante conocer el valor medio y la función de autocorrelación de las muestras:

$$- \overline{X(t)}$$

$$- R_x(t + \tau, t) = \overline{X(t + \tau)X(t)}$$

Muchos de los procesos de ruido se modelan como **procesos aleatorios estacionarios**, es decir, como procesos con estadísticos invariantes en el tiempo. En éstos, tanto el valor medio como la función de autocorrelación son invariantes en el tiempo, es decir:

$$- \overline{X(t)} = \mu$$

$$- R_x(t + \tau, t) = \overline{X(t + \tau)X(t)} = R_x(\tau)$$

Bajo estas consideraciones $X(t)$ se considera un **proceso aleatorio estacionario en sentido amplio (WSS)**, y su función de autocorrelación tiene las siguientes propiedades:

$$- R_x(0) = \overline{X^2(t)} \text{ que se interpreta como la potencia media del proceso}$$

$$- R_x(\tau) \text{ es una función impar}$$

$$- |R_x(\tau)| \leq R_x(0) \text{ para todo } \tau$$

La densidad espectral de potencia de un WSS es la transformada de Fourier de su $R_x(\tau)$:

$$S_x(f) = F[R_x(\tau)] = \int_{-\infty}^{\infty} R_x(\tau) e^{-j2\pi f\tau} d\tau, \quad -\infty \leq f \leq \infty \quad (\text{b.8})$$

Puede demostrarse que la psd debe verificar:

- $S_x(f) \geq 0$ siendo una función impar

$$- P = \overline{X^2(t)} = \int_0^{\infty} S_x(f) df$$

$$- P(f_1, f_2) = 2 \int_{f_1}^{f_2} S_x(f) df$$

donde P es la potencia media, y $P[f_1, f_2]$ es la potencia de la señal en el rango de frecuencias $[f_1, f_2]$

De especial importancia en el análisis de ruido es el **ruido blanco limitado en banda**. Es un proceso estacionario en sentido amplio, con valor medio cero, y una psd plana limitada en banda, como se muestra en la Fig. B.2:

$$S_x(f) = \begin{cases} \frac{P}{2}, & |f| < B \\ 0, & \text{en otro caso} \end{cases} \quad (\text{b.9})$$

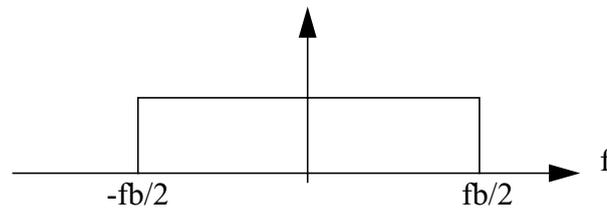


Fig. B.2: Densidad espectral de potencia del ruido blanco limitado en banda

Su función de autocorrelación es:

$$R_x(\tau) = PB \frac{\sin(2\pi B\tau)}{2\pi B\tau} = PB \text{sinc}(2B\tau) \quad (\text{b.10})$$

Considerando el ancho de banda infinito, tendremos un ruido blanco. Esto implica también las siguientes condiciones:

$$- \overline{X(t)} = 0,$$

$$- R_x(\tau) = \frac{P}{2} \delta(\tau),$$

$$- S_x(f) = \frac{P}{2}, \forall f$$

Si $X(t)$ es una variable aleatoria gaussiana, entonces el ruido blanco es un ruido blanco gaussiano (WGN). El ruido térmico y el de disparo se modelan como procesos WGN.

En el análisis de ruido normalmente se requiere la estimación del valor medio y de la función de autocorrelación del proceso estacionario y ergódico. La forma de obtener estos parámetros es con la medida simultánea de diferentes muestras en diferentes sistemas idénticos, los cuales se verán afectados únicamente por aquellos parámetros de naturaleza aleatoria. Ocurre que en el caso de circuitos integrados la medida de las características del ruido usando este sistema implicaría un alto coste, pues como paso necesario está el tener una muestra suficiente de integrados de iguales características para su posterior análisis estadístico. Ahora bien, al considerar que nuestro proceso es ergódico podemos reducir el análisis a diferentes muestras de un sistema muestreando en el tiempo, es decir, muestrear la variable aleatoria en un mismo sistema tomando como variable el tiempo. En esta media temporal, el valor medio se estima como:

$$\langle X(t) \rangle_T = \frac{1}{T} \int_0^T X(T) dt \quad (\text{b.11})$$

y la función de autocorrelación como:

$$\langle R_x(\tau) \rangle_T = \frac{1}{T} \int_0^T X(T) X(T + \tau) dt \quad (\text{b.12})$$

La densidad espectral de potencia, en este caso, se puede expresar como:

$$S_x(f) = \frac{1}{n} \sum_{i=1}^n F[\langle R_x(\tau) \rangle_T^i] \quad (\text{b.13})$$

El estimador en este caso se construirá a partir de la estimación de n muestras de $R_X(\tau)$, y tomando la media de todos los periodos de muestreo.



ANÁLISIS DEL RUIDO EN CIRCUITOS LINEALES

En un circuito real existen múltiples fuentes de ruido. Estamos interesados en hallar el ruido total en un determinado nodo, que por lo general será un nodo de entrada o de salida de circuito. Esto requiere conocer la función de transferencia desde la fuente de ruido al nodo en cuestión. Dado que el ruido es por lo general muchomás pequeño que las señales del circuito, podremos usar modelos lineales ara analizar las contribuciones de las diferentes fuentes de ruido a la potencia total de ruido en el nodo en particular.

Veamos dos ejemplos de circuitos: el integrador y el circuito RC, pues se usan bastante en la estructura de los sensores de imagen CMOS.

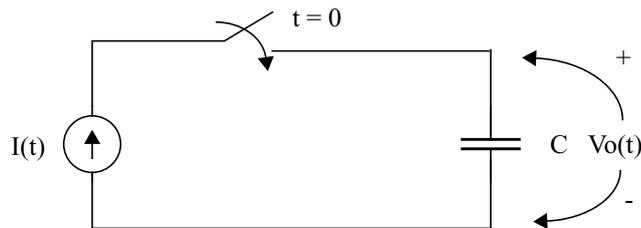


Fig. C.1: Circuito integrador

La estructura del circuito integrador se muestra en la Fig. C.1. Asumiendo $I(t)$ como un proceso estacionario en sentido amplio, de media cero, aplicado a una capacidad de integración en $t = 0$, la tensión de salida para $t > 0$ será:

$$V_o(t) = \frac{1}{C} \int_0^t I(\tau) d\tau \quad (c.1)$$

Por lo tanto, dado que su valor medio vale cero:

$$\begin{aligned} \overline{V_o^2}(t) &= \frac{1}{C^2} \int_0^t \int_0^t \bar{I}(t_1) \bar{I}(t_2) dt_1 dt_2 = \int_0^t \int_0^t R_I(t_1 - t_2) dt_1 dt_2 \\ &= \frac{1}{C^2} \int_{-t}^t (t - |\tau|) R_I(\tau) d\tau \end{aligned} \quad (c.2)$$

donde $R_I(\tau)$ es la función de autocorrelación del proceso. Si $I(t)$ es un ruido blanco con psd $P/2$, entonces:

$$\overline{V_o^2}(t) = \frac{1}{C^2} \int_{-t}^t (t - |\tau|) \frac{P}{2} \delta(\tau) d\tau \quad (c.3)$$

Si ahora consideramos nuestro circuito lineal e invariante en el tiempo con respuesta impulsiva $h(t)$ y señal de tensión de entrada $X(t)$ (Fig. C.2), de características WSS y media cero, entonces la salida $Y(t)$, en estado estacionario, también será WSS de media cero, y de densidad espectral de potencia:

$$S_y(f) = |H(f)|^2 S_x(f) \quad (\text{c.4})$$

La ecuación anterior es la base del análisis en circuitos lineales invariantes en el tiempo. El resultado típico que deriva de esta ecuación es la potencia de ruido de un circuito RC, que resulta la bien conocida expresión de kT/C .

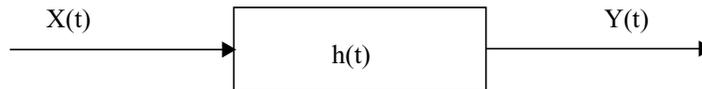


Fig. C.2: Circuito lineal invariante en el tiempo

En la Fig. C.3 se muestra el circuito RC de ejemplo, donde $V_i(t)$ es el ruido térmico asociado a R , de densidad espectral de potencia $S_{v_i}(f) = 2kTR$. La función de transferencia del circuito viene dada por:

$$H(f) = \frac{1}{1 + j2\pi fRC} \quad (\text{c.5})$$

y, por lo tanto:

$$|H(f)|^2 = \frac{1}{1 + (2\pi fRC)^2} \quad (\text{c.6})$$

Por lo que la densidad espectral de potencia de ruido a la salida vale:

$$S_{v_o}(f) = 2kTR \frac{1}{1 + (2\pi fRC)^2} \quad (\text{c.7})$$

y la potencia media de salida vendrá dada por:

$$\overline{V_o^2}(t) = \int_{-\infty}^{\infty} S_{v_o}(f) df = \frac{kT}{C} \quad (\text{c.8})$$

Se observa que la potencia media de salida resulta independiente de R . Este resultado deriva de que R está relacionada tanto con la fuente de ruido como con el ancho de banda del ruido, y ambos efectos se contrarrestan.

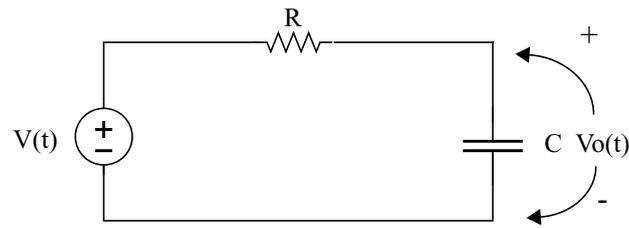


Fig. C.3: Circuito RC

Si consideramos ahora un circuito genérico con múltiples fuentes de ruido, todas ellas incorreladas, WSS y de media cero: $V_1(t)$, $V_2(t)$, ..., $V_n(t)$, en estado estacionario la tensión de ruido a la salida $V_o(t)$ será un proceso WSS, de media cero y psd:

$$S_{v_o}(f) = \sum_{i=1}^k |H_i(f)|^2 S_{v_i}(f) \quad (\text{c.9})$$

donde $H_i(f)$ es la función de transferencia del ruido desde el nodo i hasta la salida, y la potencia media a la salida vendrá dada por:

$$\overline{V_o^2}(t) = \int_{-\infty}^{\infty} S_{v_o}(f) df = \sum_{i=1}^k \overline{V_{o_i}^2}(t) \quad (\text{c.10})$$

ANÁLISIS DEL RUIDO FLICKER EN TRANSISTORES DE CANAL CORTO

Los transistores usados en las aplicaciones analógicas son normalmente dispositivos de tamaño grande. Sin embargo, en los sensores de imagen CMOS, los dispositivos implementados a nivel de píxel no pueden ser de tamaño elevado, pues se reduciría drásticamente el factor de relleno y, con ello, la eficiencia de la conversión de señal óptica en eléctrica del sensor. Las tecnologías CMOS usadas en la actualidad para la implementación de sensores son tecnologías submicrométricas en las que los transistores son de canal corto ($L < 1 \mu m$). En aplicaciones de mucha precisión, el ruido limita el rango dinámico de los sensores de imagen CMOS. En general el ruido flicker generado por los transistores CMOS usados en el bloque de lectura no presenta una limitación, pues se elimina mediante la técnica CDS. Sin embargo, uno de los mayores problemas del escalado de integración es el cambio en el comportamiento del ruido a baja frecuencia de los transistores MOS. Analicemos a continuación este factor.

Los orígenes del ruido de baja frecuencia no se entienden del todo. Hay varios modelos propuestos que consideran todas las regiones de operación y todos los niveles de inversión de los transistores. Sin embargo, la mayoría de estos modelos son válidos sólo para transistores de canal largo y necesitan de parámetros raramente facilitados por las foundries.

En el *modelo de fluctuación de carga* [1] el ruido flicker se atribuye a la captura y posterior liberación de las cargas en las trampas (estados energéticos dentro de la banda prohibida del semiconductor generados por las imperfecciones introducidas en el proceso de fabricación) que existen cerca de la interfaz semiconductor-óxido (Si-SiO₂ en el caso de una tecnología de silicio). Las fluctuaciones de carga generan un potencial superficial que modula la movilidad de los portadores en el canal. Se asume que el canal puede intercambiar cargas con las trampas de la interfaz gracias al efecto túnel. El *modelo de fluctuación en la movilidad* [2] considera que el ruido flicker es el resultado de las fluctuaciones de la movilidad de la carga en el substrato. Ninguno de estos dos modelos explica de forma definitiva el comportamiento de la carga a baja frecuencia. Sí es cierto que el comportamiento del ruido a baja frecuencia depende fuertemente del proceso de fabricación usado. Además, se han observado comportamientos muy diferentes a baja frecuencia para transistores nmos y pmos.

Recientemente se ha profundizado en el estudio de las propiedades del ruido para dispositivos de área muy pequeña. Para MOSFETs con área de canal muy pequeña es posible tener sólo una trampa en el óxido activo cerca del cuasinivel de Fermi en todo el canal. La captura y emisión del portador de carga por parte de la trampa produce una modulación discretizada de la corriente de canal, generando una señal telegráfica aleatoria (RTS). Una señal telegráfica aleatoria será una señal binaria que refleja el estado de la trampa: si ésta ha capturado a un portador tomará el valor '1', mientras que valdrá '0' en cualquier otro caso, y su densidad espectral de potencia una curva lorentziana (Fig. D.1). Para dispositivos de área mayor, el ruido flicker podrá modelarse como la superposición de las RTSs individuales. El salto entre dos niveles

discretos de la corriente de drenador se modela generalmente como superposición de los dos efectos antes mencionados: la fluctuación del número de portadores de carga, y la fluctuación de la movilidad que ocurre cuando una trampa cambia de estado.

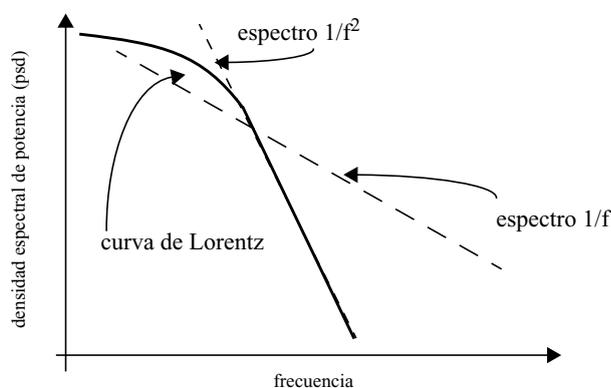


Fig. D.1: Comparación a baja frecuencia de los diferentes espectros

En [3] se realiza un estudio detallado de la potencia de ruido en la corriente de drenador tomando como variables del estudio transistores nmos y pmos de canal corto en tecnologías submicrométricas, bajo diferentes geometrías y puntos de operación. Del estudio realizado se deducen las conclusiones que a continuación presentamos.

Las características de la potencia de ruido para un nmos son bastante regulares ante variaciones geométricas y de los puntos de operación. La potencia de ruido generada por la corriente de drenador crece con la tensión de puerta en inversión moderada y débil y tiende a saturarse en inversión fuerte. En la región lineal, la potencia de ruido generada por la corriente de drenador crece con la tensión de polarización de drenador, y por lo tanto, fuera de la región óhmica crece levemente.

Para la corriente de drenador de un pmos con las mismas dimensiones se observa que la potencia de ruido depende fuertemente de la corriente de polarización. Para una polarización de puerta y de drenador dada la potencia de ruido del transistor nmos es mayor que la potencia generada por un pmos. En general, los dispositivos pmos tienen niveles de ruido menores que los nmos de iguales geometrías. Esto puede atribuirse a que el efecto túnel de los huecos en la interfaz semiconductor-óxido es mayor que el de los electrones, y la diferencia entre las densidades de trampas en el óxido cerca de los límites de las bandas de valencia y conducción, así como las diferencias en la movilidad, dando como resultado diferentes grados en la fluctuación de la movilidad superficial.

Respecto a las características del ruido de la corriente de drenador en la región de inversión fuerte y en la región subumbral para un MOSFET pmos de canal corto. En contraste con los resultados obtenidos para los dispositivos de canal largo, para estos dispositivos la potencia de ruido continúa aumentando con la tensión de puerta tanto en la región de inversión débil como en la de inversión fuerte.

Respecto a la dependencia de la potencia de ruido de la corriente de drenador para el transistor de canal n se observa que en ambas regiones (inversión débil e inversión fuerte) la potencia de ruido aumenta significativamente con la tensión de puerta tanto en la región lineal como en la

región no lineal. Puede además remarcarse que la potencia de ruido depende débilmente de la tensión de drenador en la región de saturación.

De la evolución de la potencia de ruido para el mismo dispositivo para las regiones lineal y de saturación para inversión débil e inversión fuerte puede decirse que a medida que la tensión de puerta decrece, el número de señales telegráficas aleatorias que modelan el comportamiento del ruido flicker en tecnologías submicrométricas se reduce y el espectro puede descomponerse claramente en un número finito de curvas Lorentzianas elementales. Este efecto resulta más evidente para las fluctuaciones del mismo dispositivo observado en el dominio del tiempo. Se pueden observar diferentes señales telegráficas, cada una caracterizada por diferentes tiempos entre saltos. En otras palabras, existen dos o tres trampas activas en la interfaz con diferentes características temporales y diferentes niveles energéticos.

Como conclusión, el espectro del ruido flicker en los MOSFETs de área pequeña es el resultado de una superposición de señales telegráficas aleatorias. El papel de los estados energéticos presentes en la interfaz produce el efecto de RTS. El ruido flicker se hace más importante a medida que las tecnologías se escalan y se reducen las tensiones de polarización para lograr un menor consumo.

El modelo matemático propuesto en [4] para el ruido flicker modelado como un conjunto de señales RTS es el siguiente. La captura y emisión de portadores de carga en el canal del transistor por parte de las trampas pueden modelarse por la función $N(t)$, la cual tomará un valor '1' si un portador es capturado, y '0' en otro caso, como se muestra en la Fig. D.2. La trampa se considera activa cuando su nivel de energía está próximo al nivel de Fermi [5] del semiconductor. En tal caso, las frecuencias de captura y emisión serán aproximadamente iguales. En tal caso, $N(t)$ puede modelarse como una señal telegráfica aleatoria de frecuencia λ . En equilibrio, la autocovarianza del proceso aleatorio definido por $N(t)$ puede expresarse como:

$$C_{\lambda}(\tau) = \frac{1}{4}e^{-2\lambda\tau} \quad (\text{d.1})$$

y la correspondiente densidad espectral de potencia viene dada por:

$$S_{\lambda}(f) = \frac{1}{2} \frac{\lambda}{\lambda^2 + (\pi f)^2} \quad (\text{d.2})$$

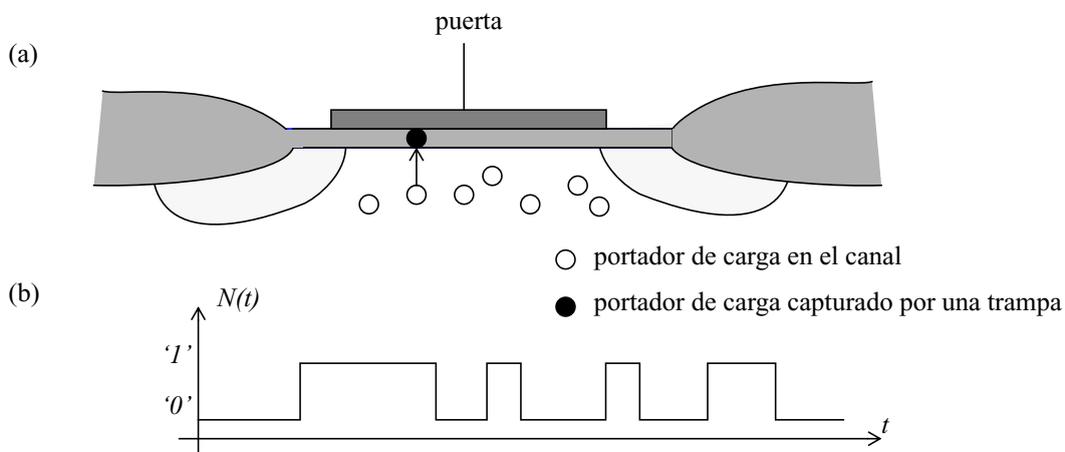


Fig. D.2: (a) Representación de un transistor MOS con una sola trampa en el óxido de puerta; y (b) la señal telegráfica aleatoria generada por la trampa

En la práctica, los transistores MOS tienen un número de trampas elevado en el óxido de puerta. Puesto que cada trampa captura y emite portadores de forma independiente, la densidad espectral de potencia del número total de electrones capturados es la suma de todas las densidades de potencia de cada electrón individual atrapado. Las trampas pueden tener diferentes frecuencias de captura y liberación, en función del lugar en el que se encuentren ubicadas. La distribución de frecuencias $g(\lambda)$ se supone que obedece a una distribución logarítmica uniforme [6]:

$$g(\lambda) = \frac{4kTAt_{ox}N_t}{\lambda \log \frac{\lambda_H}{\lambda_L}} \quad (d.3)$$

donde kT es la energía térmica, A es el área del canal, t_{ox} el ancho de óxido de puerta efectivo, N_t la densidad de trampas (en $\text{eV}^{-1}\text{cm}^{-3}$), λ_H y λ_L las frecuencias máxima y mínima de captura por parte de las trampas, respectivamente. La densidad espectral de potencia del total de electrones capturados viene dada por:

$$S(f) = \int_{\lambda_L}^{\lambda_H} S_{\lambda}(f)g(\lambda)d\lambda \cong \frac{kTAt_{ox}N_t}{2 \log \frac{\lambda_H}{\lambda_L} f}, \quad \text{para } \lambda_L < f < \lambda_H \quad (d.4)$$

Para $f < \lambda_L$, $S(f)$ es constante, mientras que para $f > \lambda_H$, $S(f)$ es proporcional a $1/f^2$. En el rango de frecuencias habitual, $S(f)$ resulta proporcional a $1/f$.

Para modelar el ruido flicker en los transistores MOS de canal corto, las fluctuaciones en el número de portadores es dominante [7], y la densidad de potencia del ruido flicker introducido por la corriente de drenador viene dada por:

$$S_{I_d}(f) = g_m^2 S_{V_g}(f) = g_m^2 \frac{1}{C_{ox}^2} S_{Q_{ch}}(f) = g_m^2 \frac{1}{C_{ox}^2} \left(\frac{q}{A}\right)^2 S(f) = \frac{g_m^2 q^2 kTAt_{ox}N_t}{2C_{ox}^2 A \log \frac{\lambda_H}{\lambda_L} f} \quad (d.5)$$

donde C_{ox} es la capacidad de óxido de puerta, $S_{V_g}(f)$ la densidad de potencia equivalente de ruido flicker en tensión en la puerta del transistor, y $S_{Q_{ch}}(f)$ es la densidad de potencia de ruido flicker introducido por la carga en el canal. Como se ve, $S_{I_d}(f)$ es inversamente proporcional al área de puerta, lo que explica que el ruido flicker tome especial importancia a medida que la tecnología se escala.

Referencias



-
- [1] McWhorter, A. L. 1957 “1/f noise an germanium surface properties” In R. H. Kingston (Ed.), *Semiconductor Surface Physics* (Philadelphi: University of Pennsylvania Press), pp. 207-228
 - [2] Hooge, F. N. 1994 “1/f noise sources” *IEEE Transactions on Electron Devices*, No. 41, pp. 1926.1935
 - [3] H. Belahrach, Y. Degerli, ... “Low-frequency noise measurements on submicrometer n-channel and p-channel MOSFETs at various operating regions” *Troisième Conférence Internationale sur les Mathématiques Appliquées et les Sciences de l’Ingénieur (CIMASI’2000)*, 23-25 Octobre 2000, Casablanca (Maroc).
 - [4] Hui Tian “Noise Analysis In CMOS Image Sensors”, PhD thesis
 - [5] M. J. Kirton, M, J, Uren “Noise in Solid-State Microstructures: A New Perspective on Individual Defects, Interface States and Low-frequency (1/f) Noise” *Advances in Physics*, 38(4): 367-468, 1989
 - [6] C. Jakobson, I. Bloom, Y. Nemirovsky “1/f Noise in CMOS Transistors for Analog Applications From Subthreshold to Saturation” *Solid-State Electronics* 42(10), pp. 1807-1817, 1998
 - [7] J. Chang, A. A. Abidi, C. R. Viswanathan “Flicker Noise in CMOS Transistors from Subthreshold to Strong Inversion at Various Temperatures” *IEEE Trans. Electron Devices*, 41 (11): 1865-1971, November 1994

