



UNIVERSITA' DEGLI STUDI DI PARMA
DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE

Dottorato di Ricerca in Tecnologie dell'Informazione
XX Ciclo

Paolo Delfanti

PROGETTO E COLLAUDO DI UN SENSORE DI
RADIAZIONE A STATO SOLIDO

Dissertazione presentata per il conseguimento
del titolo di Dottore di Ricerca

Gennaio 2008

Alla mia famiglia

CAPITOLO 1: INTRODUZIONE

Sensori di radiazione

Nell'ambito della rivelazione di radiazioni per usi commerciali, medicali o destinati agli esperimenti di fisica delle alte energie, esistono differenti tipologie di approccio. Un sensore di radiazione è fondamentalmente un tipo di sensore in grado di registrare il passaggio di un fotone o di un flusso di fotoni (e, per esteso, di particelle che si comportino come fotoni) e fornire all'utente una serie di informazioni su questo evento, come la collocazione spaziale, l'energia della radiazione, o la sua intensità. Un simile sistema si presta agli usi più svariati, quali per esempio:

- Vertex detector
- Imaging
- Spettroscopia

Solitamente, lo spazio degli stati in uscita è quadridimensionale: le quattro coordinate sono composte da due coordinate spaziali, una temporale e dall'informazione associata ad ogni coordinata e la cui natura fisica può variare da caso a caso: può essere un'informazione di intensità, di lunghezza d'onda della radiazione incidente, o anche solo un'informazione binaria di tipo luce/buio. Inoltre il segnale può essere trattato, sia da circuiteria interna all'elemento sensibile che esterna per estrapolare ulteriori informazioni, o ridurre la mole di dati in uscita ai soli elementi effettivamente interessanti.

La radiazione da rivelare si estende a gran parte dello spettro elettromagnetico, dall'infrarosso fino ai raggi gamma, oltre alla possibilità di rivelare anche vari tipi di particelle come elettroni, neutroni, protoni o particelle esotiche prodotte durante gli esperimenti di fisica delle alte energie.

Fra i differenti approcci alla rivelazione di radiazioni, quello che ha conseguito i più importanti sviluppi negli ultimi anni è l'impiego di

rivelatori a stato solido, in particolare i rivelatori a semiconduttore. Questa famiglia di componenti ha avuto un grande impulso in quanto l'industria dei semiconduttori, e la comunità scientifica a cui essa attinge stanno espandendo sempre di più gli ambiti di utilizzo delle tecniche e delle proprietà di questi materiali. Grazie ad un'ottima conoscenza dei processi fisici che avvengono all'interno dei semiconduttori e ad una matura tecnologia per la loro lavorazione, i sensori che ne fanno uso stanno gradualmente sostituendo gli altri tipi di sensore (sensori a stato gassoso, pellicole fotografiche) anche in ambiti in cui non erano stati considerati all'inizio adatti, come la rivelazione di raggi X o di MIP (Minimum Ionising Particles).

La Tecnologia CMOS nella rivelazione di radiazioni

I rivelatori di radiazione CMOS standard, pur essendone il funzionamento noto sin dai primi anni '80, hanno iniziato solo di recente ad essere competitivi verso le tecnologie dedicate. Infatti all'inizio, rispetto ai Charge-Coupled-Devices (CCD), ai Silicon Drift Detectors (SDD) e ai sensori gassosi risultavano più rumorosi, oltre ad avere un rapporto fra fotoni incidenti e segnale in uscita (Quantum Efficiency, QE) molto basso. Questo era dovuto sia alla bassa percentuale di area sensibile in ogni pixel del sensore (Fill Factor, FF), sia al poco efficiente meccanismo di raccolta di carica. Solo di recente [1], grazie al costante processo di queste tecnologie, spinto dal loro impiego nell'elettronica di consumo, è stato possibile correggere questi difetti e renderli secondari rispetto gli innumerevoli vantaggi del CMOS.

Infatti l'integrazione sempre più spinta ha permesso di aumentare l'elettronica all'interno del pixel senza penalizzare il Fill Factor. In questo modo è stato possibile inserire dell'elettronica di condizionamento del segnale all'interno del pixel, in modo da diminuire il rumore e nel contempo elaborare più efficacemente un segnale altrimenti molto più basso rispetto le tecnologie alternative. La possibilità di sfruttare la tecnologia CMOS

standard ha permesso inoltre un deciso abbattimento dei costi per la creazione e fabbricazione dei sensori di radiazione: grazie all'economia di scala, derivante dall'impiego della stessa tecnologia anche per l'elettronica di consumo, il costo per unità d'area è sensibilmente inferiore rispetto a tutte le tecnologie rivali. Inoltre la possibilità di inserire tutta la catena di processing del segnale su un unico chip, i bassi consumi e l'impiego di basse tensioni di alimentazione ne fanno un candidato ideale per i sistemi portatili.

I vantaggi del CMOS si riassumono nei seguenti punti:

- Basso costo e tecnologia facilmente disponibile
- Basso consumo di potenza
- Utilizzo di basse tensioni di alimentazione (<5V)
- Alta risoluzione (grazie al costante scaling della tecnologia CMOS secondo la legge di Moore)
- Possibilità di inserimento di elettronica (sia digitale che analogica) sullo stesso chip del sensore, per ottenere un sistema di acquisizione più compatto
- Possibilità di assottigliamento del bulk sensibile, eliminando l'effetto del multiple scattering e migliorando la risoluzione
- Possibilità di lettura non necessariamente sequenziale, con definizione di zone d'interesse

Gli svantaggi, invece, sono i seguenti:

- Alto rumore e fixed pattern noise
- Bassa efficienza nella raccolta del segnale

I sensori di tipo CMOS richiedono di conseguenza di dedicare una certa area del sensore all'elettronica ancillare del sensore, in modo da ridurre l'impatto del rumore sui valori misurati.

Funzionamento fisico della tecnologia CMOS

Il funzionamento fisico di questi dispositivi si basa[2],[3],[4] sull'impiego di un fotodiodo polarizzato in inversa: in questo modo si viene a formare lungo

la giunzione pn una zona svuotata di carica, all'interno della quale è instaurato un forte campo elettrico. Quando una radiazione incide su questa zona, se ha un'energia sufficientemente vicina al band-gap del silicio, genera delle coppie elettrone-lacuna che invece di ricombinarsi vengono separate dal campo elettrico e raggiungono gli elettrodi di raccolta. la generazione di coppie è legata all'equazione:

$$N = \frac{\Delta E}{W}$$

Dove N è il numero di coppie elettrone-lacuna generate, ΔE è l'energia rilasciata dalla singola particella (fotone, p^+ , e^- , etc) nel silicio e W è l'energia di band-gap del silicio (3.6eV). In pratica si genera una fotocorrente proporzionale alla quantità di energia rilasciata nel reticolo (a sua volta funzione dell'intensità e della lunghezza d'onda della radiazione incidente), la quale può essere o misurata direttamente o misurata come differenza di tensione ai capi del fotodiode.

Poiché la fotocorrente tende a portare il diodo verso la debole inversione, riducendo la zona svuotata e di conseguenza la sensibilità del fotorivelatore, è necessario contrastare questo fenomeno. Esistono di conseguenza vari tipi di politica per mantenere il fotodiode in inversione

- tramite polarizzazione continua, solitamente con lettura in corrente del segnale, in quanto la tensione sul nodo sensibile del fotodiode è mantenuta costante
- tramite una ricarica che periodicamente ricrei la zona svuotata e lettura del valore in tensione prima della ripolarizzazione

nel secondo caso, il fotodiode viene lasciato flottante, in modo che raccogliendo la carica generata dalla radiazione incidente abbassi la propria tensione, comportandosi come una capacità. In questo modo al momento della lettura la tensione sarà proporzionale all'intensità luminosa sul fotodiode durante il periodo in cui è rimasto isolato. Si parla di conseguenza di integrazione di carica, in quanto non viene fornita la luminosità istantanea, bensì la luminosità totale in un intervallo di tempo. Quando gli elementi fotosensibili sono strutturati in una matrice bidimensionale, come in un imager o in un vertex detector, e di conseguenza non possono esser

monitorati continuamente, ma solo letti periodicamente, questo ultimo approccio è il più impiegato.

Per raccogliere la massima quantità di carica all'interno del pixel, si tende ad impiegare come zona p il substrato (o lo strato epitassiale) del chip, mentre la zona n del diodo è solitamente costituita da una n-well. In questo modo è possibile impiegare come zona di raccolta il bulk, poco drogato, (o lo strato epitassiale quando la tecnologia ne fa uso), aumentando il volume di raccolta della carica. L'informazione sulle buche viene persa, in quanto queste vengono attratte dall'anodo, posto a massa, mentre gli elettroni sono attirati dal catodo, che è collegato all'elettronica di lettura. La possibilità di leggere l'informazione utile dagli elettroni è altresì utile perché questi hanno una mobilità molto maggiore rispetto le lacune, $1450\text{cm}^2/\text{Vs}$ contro $450\text{cm}^2/\text{Vs}$, rendendo la rivelazione del segnale più veloce ed efficiente. Lo svantaggio è nella difficoltà di impiego dei pMOS all'interno di pixel: la loro n-well agisce come centro parassita di raccolta della carica generata dalla fotoradiatione e di conseguenza diminuiscono la QE del sensore.

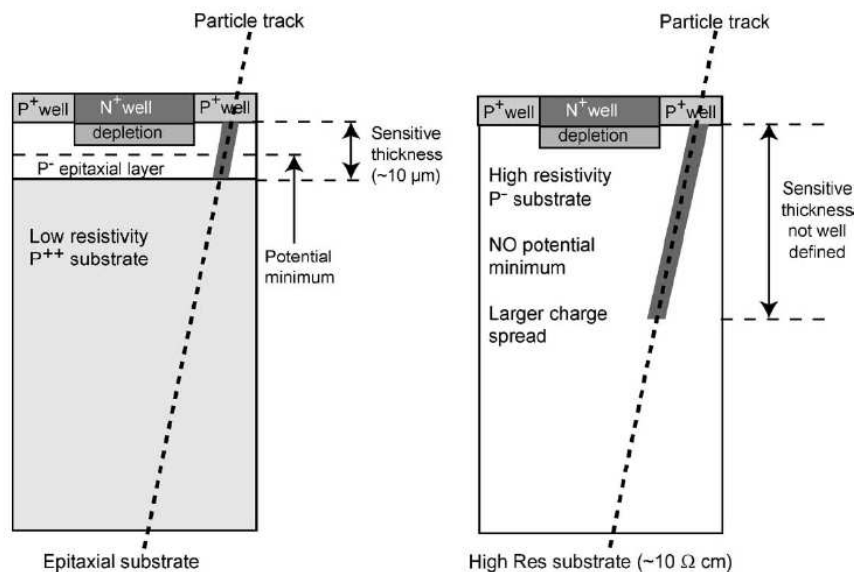


Figura 1 - Struttura di un sensore CMOS per MIP detection: con strato epitassiale (sx) e senza strato epitassiale (dx) [4]

In realtà, in un normale sensore CMOS, la zona svuotata occupa solo parte del volume sensibile: a causa dei drogaggi relativamente elevati del bulk o della zona epitassiale ($\sim 10^{15}/\text{cm}^3$) e dei limiti sulle tensioni impiegabili (1.8-3.3V), è impossibile che la zona svuotata abbracci tutto il pixel, né dal punto di vista della superficie, né della profondità. Di conseguenza la carica fotogenerata non raggiunge la n-well per drift, bensì per diffusione termica, penalizzando pertanto la velocità di raccolta, che risulta pari all'incirca alla vita media del portatore nello strato sensibile ($\sim 10\text{us}$).

Una variabile importante nella raccolta di carica è la presenza o meno di uno strato epitassiale. È infatti possibile impiegare come spessore sensibile o un bulk p- ad alta resistività o una zona epitassiale p- ad alta resistività sopra un substrato p++ a bassa resistività. Questo ultimo approccio permette un minore spreading di carica, poiché il profilo di drogaggio substrato/epi/p-well (poste dove non viene realizzato il diodo) crea un minimo di tensione locale a metà dello spessore dello strato epitassiale, dove gli elettroni vengono a raccogliersi, muovendosi quindi parallelamente alla superficie prima di esser raccolti dalla n-well e limitando così il volume della nube di carica. Di contro, il substrato p- non contribuisce alla raccolta di carica, poiché il cammino medio di una particella all'interno di esso è troppo breve rispetto alla distanza dallo strato epitassiale, a causa della maggiore facilità di ricombinazione. Pertanto tutta la raccolta della carica avviene all'interno dello strato epitassiale, meno drogato, dove gli elettroni possono raggiungere in tempo l'elettrodo di raccolta (cioè la n-well) senza esser soggetti a ricombinazione.

Per questo, la presenza di strato epitassiale pone un limite allo spessore sensibile del sensore, limitando di fatto la quantità raccolta di carica, mentre in un approccio n-well/p-bulk il sensore risulta sensibile fino ad una profondità circa pari al cammino medio di una particella nel silicio debolmente drogato.

Il pixel PPS

I dispositivi CMOS sono divisi in due famiglie: i sensori a pixel passivo (PPS) e i sensori a pixel attivo (APS). I primi sono l'approccio più semplice, in quanto prevedono di impiegare all'interno della matrice pixel dotati solo dell'elemento sensibile (fotodiodo) e di un MOS impiegato come interruttore di indirizzamento. Ogni pixel è collegato ad un bus di colonna assieme agli altri pixel della propria colonna. Durante ogni scansione, tutti i pixel di una riga vengono abilitati accendendone il transistor di indirizzamento e collegandoli al relativo bus, in modo da leggere la carica raccolta dal fotodiodo tramite un amplificatore di carica esterno alla matrice e collegato al bus, che permette inoltre di mantenere circa costante la tensione di polarizzazione del bus stesso. Al termine della lettura, la carica sul fotodiodo viene azzerata in modo da riportarlo alla massima sensibilità (reset), quindi il pass transistor viene spento e si leggono i pixel della riga successiva. In questo modo viene effettuata la scansione dell'intera matrice. I vantaggi di un simile approccio sono l'elevato Fill Factor (la zona cieca di ogni pixel è ridotta ad un unico transistor) e la possibilità di non avere strati di polisilicio sopra al pixel, come accade col CCD, che si tramuta in un'alta Quantum Efficiency (rapporto fra i fotoni incidenti e gli elettroni raccolti dal fotodiodo). Inoltre poiché l'amplificatore di carica mantiene costante la tensione di polarizzazione del bus, il rumore kTC dovuto al reset resta sotto controllo. Però il fotodiodo è collegato direttamente ad una lunga metallizzazione e di conseguenza la capacità parassita riduce sia la performance in rumore (nell'ordine di $250e_{\text{rms}}$), sia la velocità di lettura; questo rende fatto il PPS inutilizzabile per matrici di grandi dimensioni, in quanto la maggiore velocità richiesta per la lettura di un maggior numero di pixel comporta un ulteriore aumento del rumore osservato.

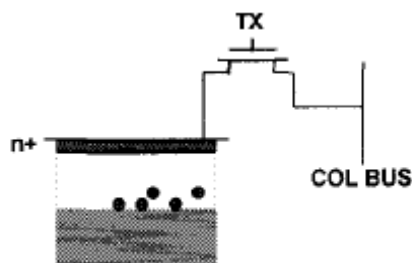


Figura 2 - Schema del pixel PPS [2]

Il pixel APS

L'APS invece rispetto al PPS prevede l'impiego di una prima elettronica di processing del segnale all'interno del pixel, direttamente collegata all'elemento sensibile. In questo modo è possibile aumentare la velocità di lettura e ridurre il rumore a causa del minore impatto dei parassiti. Il prezzo da pagare è la riduzione del FF dovuta alla presenza sul pixel di aree cieche nelle zone dove l'elettronica è presente. Lo schema standard, che è stato impiegato sin dai primi modelli, prevede l'impiego all'interno del pixel, assieme al fotodiodo, di 3 transistor:

- Transistor di reset (M1): al termine della lettura riporta in inversa il fotodiodo; il vantaggio di averlo sul pixel risiede nel fatto che rende possibile resettare il pixel senza che quest'ultimo sia selezionato, aumentando di fatto la velocità di acquisizione dei frame.
- Transistor in modalità common-drain (M2): direttamente collegato al fotodiodo, rappresenta un stadio non di amplificazione (il common-drain ha guadagno circa pari a 1), bensì di disaccoppiamento del fotodiodo dal bus di lettura. In questo modo è possibile caricare il bus di lettura più velocemente (essendo un elemento attivo) e aumentare l'escursione di segnale al fotodiodo (che in questa configurazione è collegato unicamente alla capacità di gate del transistor, di molto inferiore rispetto quella di bus). Il generatore di

corrente che polarizza questo stadio è posto alla base del bus, comune a tutti i pixel della colonna, invece che sul pixel. In questo modo viene minimizzato il fixed pattern noise fra pixel dovuto alla corrente di polarizzazione dell'amplificatore, migliora il fill-factor del dispositivo e viene ridotto il consumo di corrente, in quanto in ogni colonna c'è un solo pixel il cui amplificatore si trova in funzione in un dato istante.

- Transistor di selezione (M3): collega il source del C-D al bus, permettendo la lettura del valore.

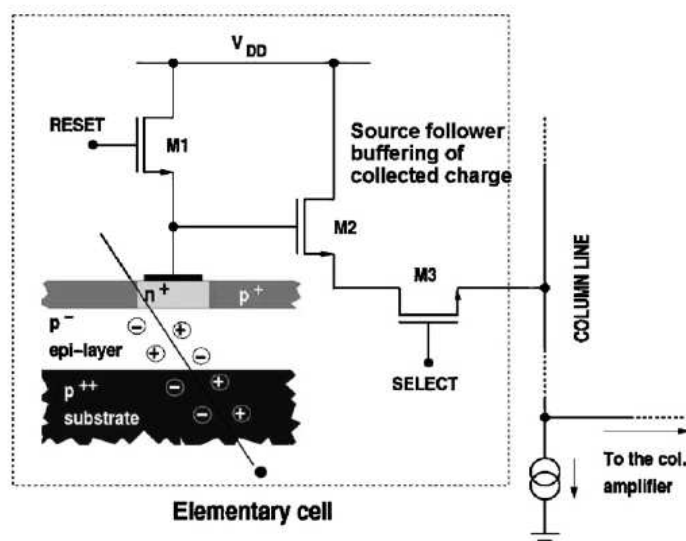


Figura 3 - Schema di un pixel APS 3T (con strato epitassiale) [2]

Come si può vedere, il pixel APS contiene un semplice follower che separa il fotodiode dall'elettronica di lettura della matrice. Permette un gran risparmio sul consumo, in quanto ogni pixel risulta acceso solo durante la lettura, cioè quando il transistor C-D, collegato al bus di colonna, è messo in condizione di condurre corrente. La scelta del source follower non è casuale, in quanto questo elemento, avendo un guadagno leggermente inferiore a 1, permette di salvaguardare il dynamic range, evitando di saturare il segnale già per basse escursioni di tensione sul fotodiode e di avere un guadagno la

cui dipendenza dalle variazioni di processo (gain mismatch) risulti meno marcata rispetto uno stadio amplificatore a guadagno maggiore di 1. La corrente di polarizzazione del transistor in C-D viene fornita da un generatore di corrente collegato al bus di colonna e comune a tutti i pixel della colonna stessa. Questo approccio permette sia di tenere sotto controllo il fixed pattern noise (la corrente di polarizzazione dello stadio è identica per tutti i pixel nella stessa colonna), sia di ridurre al minimo indispensabile i dispositivi presenti sul pixel, in quanto il MOS che fa da carico del C-D è esterno al pixel stesso.

Il funzionamento base di questo sensore si articola in tre fasi:

- Reset: M1 ON, M3 OFF: il fotodiodo viene ripolarizzato in inversa portando il catodo $V_{dd}-V_{th}$
- Integrazione: M1, M3 OFF: il fotodiodo resta flottante e inizia a scaricarsi proporzionalmente all'intensità incidente. Si noti che durante questa fase il pixel non consuma potenza.
- Lettura: M1 OFF, M3 ON: il pixel è collegato al proprio bus di colonna, con l'attivazione del SF: il valore del pixel è disponibile per la lettura

Il pixel attivo permette una lettura veloce (frequenza di lettura del singolo pixel dell'ordine dei 10MHz) e una minore dipendenza delle prestazioni rispetto le dimensioni della matrice, consentendo l'uso di matrici di grandi dimensioni (>1Mpixel).

La matrice APS e l'elettronica di lettura

Il sensore CMOS prevede, oltre all'array di pixel, che una certa area del chip venga dedicata all'elettronica di lettura e di controllo. In particolare, è necessario implementare i seguenti elementi:

- Una logica di controllo che si preoccupi di controllare la matrice e le altre strutture presenti sul chip
- Un'elettronica analogica di processing dei segnali provenienti dai pixel

- Eventualmente un'elettronica di conversione analogico-digitale per disporre in uscita del chip dei dati già convertiti

La logica di controllo è necessaria per generare i segnali di reset della matrice, per selezionare le righe e, all'interno di ogni riga, per selezionare il singolo pixel. Inoltre può essere utile per far funzionare i dispositivi di amplificazione/lettura quando questi richiedono una sequenza di operazioni successive (per esempio la cancellazione dell'offset o l'uso di ADC pipelined).

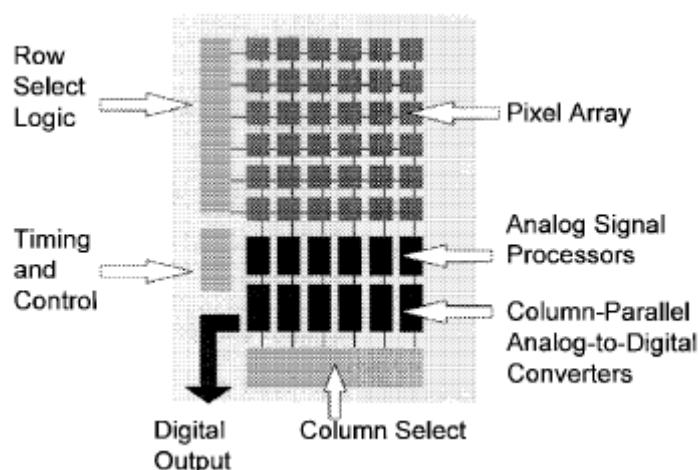


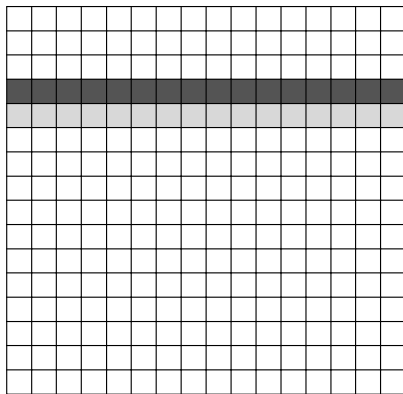
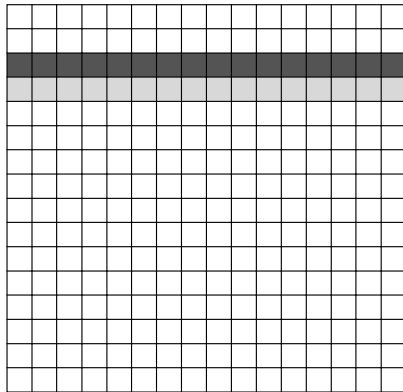
Figura 4 - Esempio di sensore CMOS completo di elettronica ancillare [[2]].

Per quel che riguarda la scelta sul funzionamento della matrice, sono possibili alcuni di gradi di libertà. La politica di generazione dei segnali di reset-lettura in particolare rappresenta l'equivalente dell'otturatore nella fotografia analogica. Per acquisire un'immagine, sia questa nel visibile o una radiografia (nella rivelazione di particelle come si vedrà gli obiettivi sono leggermente diversi) è necessario che tutti i singoli punti che la compongono siano illuminati a partire dallo stesso istante e per lo stesso periodo di tempo. Nella fotografia analogica, il funzionamento si basa sulla reazione fra la luce e una pellicola ricoperta di una sostanza fotosensibile e per controllare l'istante di inizio e la durata dell'illuminazione, si applica di fronte alla pellicola uno schermo che viene aperto e chiuso in base alla

temporizzazione voluta. Per l'imaging digitale invece la situazione è differente: la sensibilità alla luce di differenti aree di un sensore può essere decisa dall'elettronica posta sul sensore stesso e non tramite un dispositivo meccanico esterno. Però, per come il sensore è strutturato questo porta ad una serie di problemi. Infatti un pixel come quello descritto è insensibile alla luce solo durante la fase di reset, la quale è ovviamente distruttiva del valore precedentemente raggiunto. Poiché i pixel vengono letti serialmente, una riga per volta, ne deriva che le prime righe e le ultime righe lette saranno state sottoposte ad illuminazione per tempi differenti. Una possibilità a questo punto è il "global shutter" in cui tutti i pixel della matrice sono resettati allo stesso istante, ma vengono poi letti serialmente, per cui i pixel della prima riga risulteranno più "scuri" di quelli delle ultime, che sono stati integrati per più tempo. Una seconda possibilità è il "rolling shutter" in cui una riga è sempre resettata subito dopo esser letta: in questo modo il tempo di integrazione di ogni pixel resta costante, però varia l'istante di inizio: un rettangolo verticale in movimento laterale verrebbe per esempio visualizzato come un parallelogramma di luminosità uniforme in una simile politica, mentre nel primo caso risulterebbe piuttosto simile ad un trapezio di luminosità non uniforme. Per superare queste limitazioni, è necessario, come si vedrà, modificare la struttura stessa del pixel.

Per quel che concerne l'elettronica di acquisizione, l'uscita di ogni colonna è collegata ad una propria elettronica di amplificazione. Questa topologia permette di parallelizzare l'acquisizione dei dati dalla matrice con l'acquisizione contemporanea di ogni riga, rendendo la lettura più veloce e rilassando le specifiche in frequenza dell'elettronica stessa (permettendo inoltre un più efficace filtraggio del rumore ad alta frequenza), ma comporta lo svantaggio di generare un forte fixed pattern noise di colonna. Infatti poiché ogni struttura di lettura ed amplificazione ha uno scarto rispetto al valore desiderato di guadagno e punto di lavoro dovuto al processo, ne consegue che ogni colonna sia soggetta ad un offset rispetto le altre colonne, che può essere molto rilevante. Questo problema si rivela molto importante soprattutto nell'imaging, in quanto l'occhio umano risulta molto sensibile a

ROLLING SHUTTER



RESET	
LETTURA	
INTEGRAZIONE	

GLOBAL SHUTTER

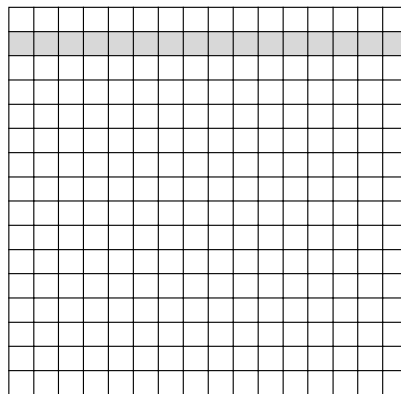
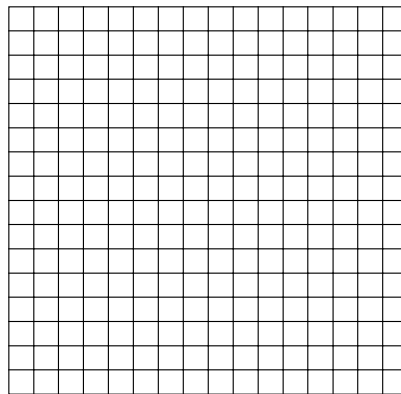
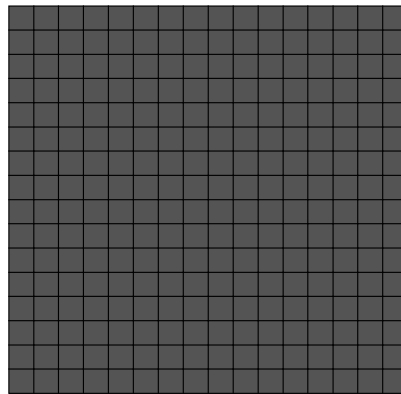


Figura 5 - Sequenza delle istruzioni per rolling shutter (sx) e per global shutter (dx): nel secondo caso, tutta la matrice è resettata e dopo l'integrazione è letta una riga alla volta

questo tipo di disturbi, ma può essere facilmente eliminato con opportune tecniche di correzione sia a livello hardware che software.

È inoltre possibile introdurre un ADC direttamente on-chip con il vantaggio di poter creare un intero sistema di acquisizione su un unico DIE, oltre alla riduzione del rumore per la vicinanza fisica fra la sorgente di segnale analogico e il convertitore. Tuttavia in tal caso una buona porzione di area del dispositivo risulta necessaria per la conversione. Si può impiegare un unico ADC per tutto il chip oppure collegare ogni colonna ad un proprio convertitore. Nel secondo caso, le specifiche dell'ADC saranno più rilassate dal punto di vista della velocità, essendocene un maggior numero che lavora in parallelo (in una matrice $N \times N$, la frequenza di lavoro è ridotta di N), ma potrebbe a sua volta portare ad un incremento nel valore di FPN (oltre che di gain mismatch) di colonna a causa della dispersione dei parametri di questi dispositivi. In alcuni approcci [5] è stato addirittura provato ad inserire gli ADC direttamente all'interno del pixel, anche se questo porta ad un evidente peggioramento del fill-factor.

Problematiche relative al reset

Il reset del chip è uno dei fattori critici del pixel. Poiché, come spiegato in precedenza, il pixel non può contenere pMOS, il fotodiode deve essere riportato in inversa collegandone il catodo ad alimentazione tramite un nMOS che si comporti da switch. Questo approccio rappresenta uno dei punti deboli del pixel APS. In particolare, l'impiego di un nMOS come pull-up fa sì che l'anodo sia caricato ad una tensione di una soglia inferiore rispetto la tensione di alimentazione. Poiché il segnale utile è rappresentato dalla diminuzione di tensione misurata sull'anodo del fotodiode rispetto la tensione di reset, è evidente come l'impiego di un nMOS come switch limiti

l'escursione utile di segnale. Questo è peggiorato dal fatto che sotto un certo valore di tensione di gate, pari $V_{th}+V_{DSat}$, essendo V_{DSat} la tensione minima di drain a cui può arrivare M4, il transistor in common-drain non lavori più in saturazione, ma tenda a spegnersi, riducendo ulteriormente lo swing di tensione rilevabile dal sensore.

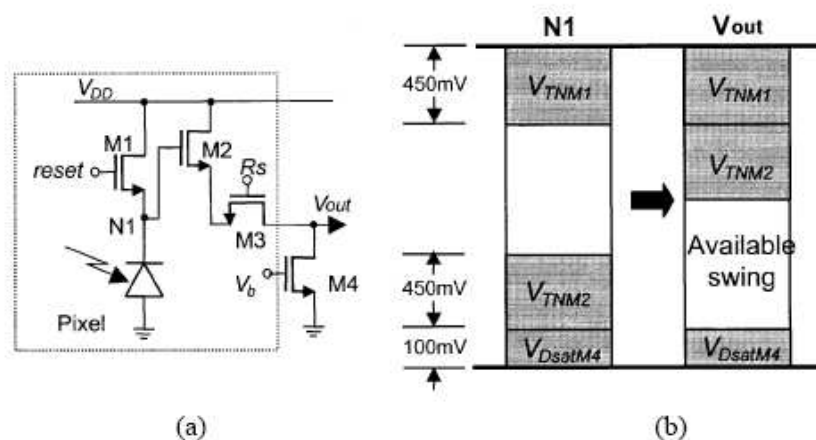


Figura 6 - Escursione di tensione disponibile per un pixel APS [6]

Un secondo aspetto negativo del basso valore di reset sta nella minore estensione di larghezza della zona svuotata, che porta ad una diminuzione della porzione maggiormente sensibile del pixel, essendo la larghezza della zona svuotata nel substrato p:

$$X_p = \frac{2\epsilon_s}{q} \frac{N_d}{N_a} \frac{1}{N_d + N_a} (\phi - V_a)$$

con N_a e N_d densità di accettori e donatori rispettivamente ($N_a \ll N_d$ perché il bulk è molto meno drogato della n-well) e V_a tensione applicata (riferita al catodo). La carica generata al di fuori della zona svuotata può sempre esser raccolta dal fotodiodo, ma in misura minore rispetto alla zona svuotata in quanto, come precedentemente spiegato, questa raggiunge il FD per diffusione termica, processo che risulta lento e poco efficiente.

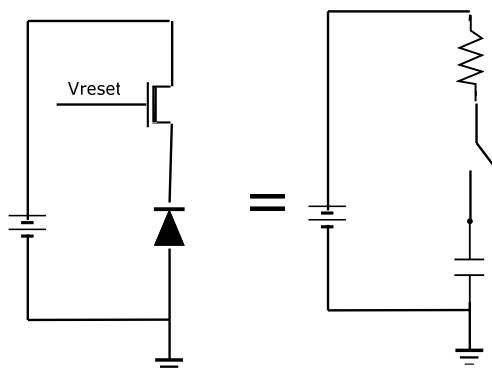


Figura 7 - Schema per lo studio del rumore kTC

Riconducibile alla fase di reset è il rumore di tipo kTC, che risulta il massimo contributo di rumore che affligge i pixel APS. Questo è un rumore termico, tipico della carica di un transistor attraverso uno switch, e può esser visto sia come un rumore in carica (potenza di rumore pari a kTC) sia come un rumore in tensione (potenza pari a kT/C). Bassi valori di capacità di conseguenza aumentano questo rumore, ma aumentano al contempo la sensibilità del sensore, e viceversa. Caricando una capacità tramite una resistenza, ad ogni istante si ha un'incertezza sulla quantità di carica presente su tale capacità, e si tratta di un rumore a potenza

$$\Delta V = 4kTR$$

che viene filtrato tramite una rete RC, la cui banda è $1/4RC$ mostrando così una potenza di rumore pari a

$$\Delta V^2 = \frac{kT}{C}$$

indipendente dal valore della resistenza, ma funzione della sola capacità.

Poiché lo switch è considerabile come una resistenza di valore trascurabile quando è acceso e come una resistenza infinita quando spento, questo rumore è visibile come rumore bianco (banda infinita) durante la carica della capacità e come offset (banda pari a 0, ovvero è stato "ibernato" il valore all'istante di chiusura dello switch) quando l'interruttore viene spento.

Ricordando inoltre che per una capacità la carica immagazzinata è proporzionale alla tensione: $Q=CV$, si ha l'espressione del rumore in carica:

$$\Delta Q^2 = kTC$$

Come si può notare, questo rumore, se considerato in tensione, diminuisce all'aumentare di C , mentre vale il contrario se è considerato dal punto di vista della carica.

Poiché questo rumore si configura come un'incertezza sul valore del diodo a fine della fase di reset, si tratta di un valore tempovariante, ma che resta costante all'interno di uno stesso periodo di funzionamento (reset-integrazione-lettura) del pixel e che varia dopo ogni fase di reset. Il segnale utile nei pixel CMOS standard è dato dall'escursione di tensione sul fotodiodo alla fine della fase d'integrazione rispetto il valore iniziale di reset, ma è di solito possibile misurare unicamente il valore finale della tensione sul fotodiodo, nell'ipotesi che il valore iniziale sia sempre pari a $V_{dd}-V_{th}$. Di conseguenza ogni incertezza sul valore iniziale d'integrazione si traduce direttamente in un'incertezza sul valore misurato. Va ricordato il rumore kTC è stato una delle cause per cui il CMOS non è stato impiegato sin dal principio negli imager o nei sensori di radiazione.

Il processo di reset, avvedendo mediante nMOS, prevede inoltre un certo grado di libertà nella scelta della zona di funzionamento del transistor di reset. Infatti se si porta la tensione di gate alla stessa tensione presente sul drain, tipicamente V_{dd} , il transistor lavora in saturazione ($V_{gs}=V_{ds}$) e sale fino a $V_{dd}-V_{th}$, quando entra in debole inversione; a questo punto però il transitorio non finisce, ma semplicemente rallenta perché il transistor conduce ancora corrente (weak inversion), di modo che sia necessario un tempo nell'ordine dei 200us per raggiungere il punto d'equilibrio, in cui la corrente di sottosoglia del MOS sia pari alla corrente di perdita del fotodiodo. Un simile tempo di reset è troppo lungo, per cui si preferisce terminare la fase di reset prima, lasciando sul fotodiodo una tensione dipendente anche dal valore precedente (image lag) e dalla durata scelta per la fase di reset stessa. Tuttavia è possibile dimostrare che proprio il fatto che il transitorio venga interrotto prima del raggiungimento del punto di equilibrio riduca la potenza del rumore kTC di un fattore 2 [7].

La seconda possibilità è forzare la conclusione del transitorio di reset facendo in modo di avere sul gate del transistor una tensione di una soglia superiore a quella di drain, per fare in modo che il transitorio avvenga col MOS in regione lineare, cioè con un comportamento simile ad un resistore a bassa impedenza. In questo modo la conclusione ($V_{ds}=0$) avviene in tempo utile eliminando ogni persistenza del valore precedente, ma si paga con un peggioramento del rumore kTC a fine transitorio, la cui potenza (in tensione) è esattamente pari a

$$\Delta V^2 = \frac{kT}{C}$$

Inoltre un simile approccio richiederebbe l'impiego di una tensione sul drain del transistor di reset pari a

$$V_{dd} - V_{th} - V_{overdrive}$$

essendo $V_{overdrive}$ (tipicamente $>100\text{mV}$) la tensione necessaria a garantire che il MOS di reset lavori in zona lineare compensando eventuali variazioni nella tensione di soglia e minimizzando la resistenza di canale. Ne consegue una minore tensione sul fotodiode alla fine del reset e il conseguente peggioramento dello swing massimo di segnale ottenibile.

Altre componenti di rumore

I contributi di rumore nella lettura dei valori provenienti dalla matrice sono di differente natura. Oltre al già citato rumore kTC, ci sono i rumori termici dei canali dei transistor, il rumore $1/f$ (pink noise) riconducibile al gate del MOS Common-Drain, il rumore di tipo shot dovuto alla corrente di perdita del fotodiode [7] e il fixed pattern noise (FPN).

Un rumore che si manifesta durante la fase d'integrazione e che è riconducibile alla corrente di buio sul fotodiode è lo shot noise. Questo è un rumore che si verifica nei rivelatori ad alta sensibilità ogni volta che i portatori di segnale (in questo caso gli elettroni) sono visibili come una grandezza granulare e non continua. In questo caso, la differente distribuzione statistica di piccole quantità può portare comunque a

variazioni rilevanti del segnale di uscita. La distribuzione statistica dei portatori di carica in movimento è data dalla distribuzione di Poisson, che modella una serie di eventi discreti che nel caso in esame hanno frequenza media pari a

$$\Lambda = \frac{I_d}{q}$$

con I_d corrente sul diodo. Ricordando che nella distribuzione di Poisson la varianza è pari al valore medio, in un periodo di integrazione vengono raccolti in media

$$\Delta n^2 = n = \frac{I_{ph} + I_{leak}}{q} T_{int}$$

elettroni. Poiché il rumore in tensione è stimabile dalla relazione $n=CV/q$ si ha che

$$\Delta V^2 = \frac{q(I_{ph} + I_{leak})}{C_{pd}^2} T_{int}$$

Ovvero, in tensione:

$$\Delta V = \sqrt{\frac{q(I_{ph} + I_{leak})}{C_{pd}^2} T_{int}}$$

In un rivelatore di tensione, la corrente di buio può di conseguenza dare origine ad un rumore tanto maggiore quanto più il tempo d'integrazione è lungo e elevata la corrente di perdita del fotodiodo.

In generale, se il tempo d'integrazione risulta inferiore a pochi ms questo contributo di rumore risulta trascurabile. Il valore della capacità del fotodiodo è inoltre considerabile in prima approssimazione costante nelle tecnologie sub-micrometriche, dove lo swing di tensione su di esso risulta notevolmente ridotto.

Il termine I_{ph} mostra inoltre l'effetto dei fotoni, il cui effetto negativo sul SNR risulta più elevato in ambiti in cui il sensore deve creare un'immagine partendo da piccole quantità di fotoni. In questo caso conviene ragionare in termini di fotoni (ricordando che il numero di elettroni generato è proporzionale al numero di fotoni incidenti). Detto N il numero medio di

fotoni, il segnale raccolto (in fotoni) è ovviamente N , mentre il rumore derivante dalla dispersione statistica di Poisson ha come deviazione \sqrt{N} . Il SNR (relativo unicamente a questo tipo di rumore) è di conseguenza pari a N/\sqrt{N} , cioè \sqrt{N} che conferma come questo contributo possa risultare fortemente penalizzante in ambiti in cui il sensore debba creare un'immagine su un numero esiguo di fotoni.

Una seconda classe di rumori sono i rumori temporali che affliggono i canali dei MOS impiegati per la lettura (MOS C-D, MOS di selezione, oltre al MOS di carico del C-D presente nell'elettronica di colonna). Si tratta di rumori a bassa frequenza ($1/f$) e rumore termico, o bianco. È possibile studiare la caratteristica di trasferimento del rumore di ognuno di questi MOS per averne il rumore termico riferito all'uscita; in particolare, per il transistor in C-D M2:

$$\Delta V^2 = \frac{2 kT}{3 C_o} \frac{1}{1 + \frac{g_{m2}}{g_{d3}}}$$

Mentre per il transistor di selezione M3:

$$\Delta V^2 = \frac{kT}{C_o} \frac{1}{g_{d3} \left(\frac{1}{g_{d3}} + \frac{1}{g_{m2}} \right)}$$

Mentre per il transistor di carico M4:

$$\Delta V^2 = \frac{2 kT}{3 C_o} g_{m4} \left(\frac{1}{g_{d3}} + \frac{1}{g_{m2}} \right)$$

Il contributo da parte del resto della struttura risulta trascurabile rispetto i transistor presi in esame, così come il contributo di rumore del transistor M3 risulta molto inferiore rispetto ai contributi di M2 e M4.

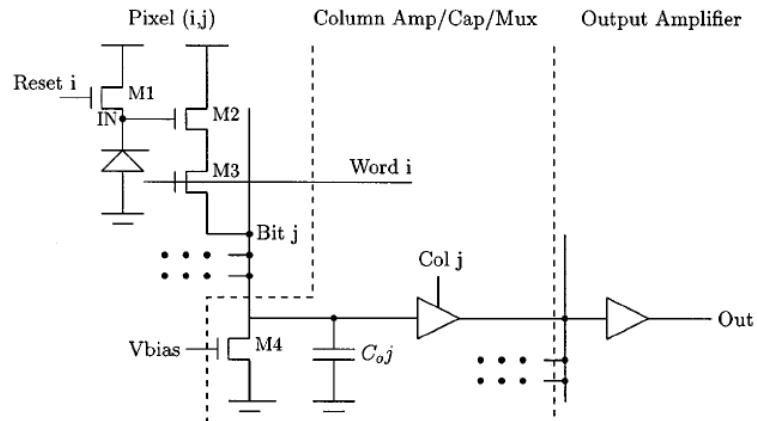


Figura 8 - Schema del pixel 3T + elettronica di colonna [7]

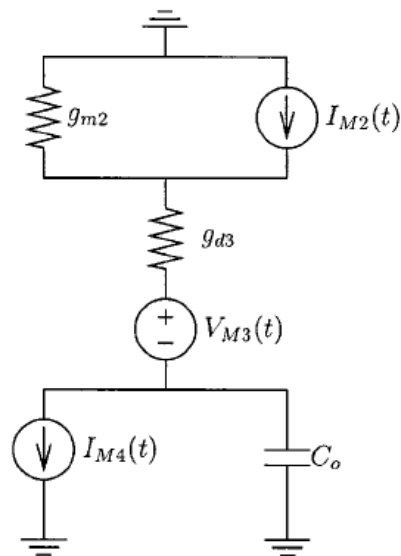


Figura 9 – Modello ai piccoli segnali per lo studio del rumore in un APS 3T [7]

Un ulteriore tipo di rumore è invece il Fixed Pattern Noise (FPN). Questo non è un rumore propriamente detto, in quanto è tempoinvariante, ed è dovuto ai mismatch fra le caratteristiche di componenti nominalmente

identici all'interno del pixel e nell'elettronica analogica di processing del segnale. In particolare ha forte dipendenza dalle variazioni delle tensioni di soglia (essendo il valore all'inizio dell'integrazione $V_{dd}-V_{th}$) e delle correnti di polarizzazione del transistor C-D. Si configura di conseguenza come un offset costante fra i valori letti dei pixel, in quanto i due parametri descritti sono determinano il punto di lavoro del sistema. In pratica a causa di questo rumore, se si sottopone il sensore ad illuminazione uniforme le immagini ottenute non risultano uniformi, ma mostrano un pattern fisso (dove il nome) e immutabile nel tempo. A sua volta all'interno del Fixed Pattern Noise si può individuare il Fixed Pattern Noise di colonna, dovuto non tanto alle variazioni fra i parametri dell'elettronica dei pixel, quanto alle variazioni dell'elettronica di processing del segnale proveniente dalle colonne. Poiché la matrice viene scandita una riga alla volta, e tutti i pixel della stessa colonna sono letti dalla medesima elettronica, si osserverà un offset fisso fra le colonne del sensore.

Un secondo contributo al FPN del pixel CMOS è costituito dalla corrente di leakage. Il fotodiode, essendo un diode flottante polarizzato in inversa, tende a spontaneamente scaricarsi per colpa della corrente di inversa, che causa una perdita di tensione. Dal punto di vista del fixed pattern noise, questa si configura come una fonte di FPN, perché la corrente di leakage varia fortemente in base ai parametri fisici del diode. In particolare, aumentando il tempo d'integrazione, è possibile notare un aumento nella dispersione dei valori di buio dei pixel di una matrice, poiché il maggior tempo d'integrazione "amplifica" le differenze di tensione derivanti dalle diverse correnti di buio fra pixel differenti, essendo (in prima istanza)

$$V_{leak} = C_{ph} I_{leak} T_{int}$$

A causa della corrente di buio dei pixel è inoltre necessario tenere sotto controllo la durata della fase d'integrazione, in quanto fasi di integrazioni troppo lunghe porterebbero ad una diminuzione dell'escursione utile del segnale, per il sommarsi di questa corrente con la fotocorrente. In base all'escursione minima di segnale accettata (cioè la differenza fra segnale saturato e valore buio) la corrente di buio setta di conseguenza un massimo periodo di integrazione (cioè una frequenza minima di funzionamento).

Metodologie di riduzione del rumore

Per ridurre il rumore esistono differenti approcci, orientati soprattutto a ridurre i rumori a bassa frequenza, come kTC e FPN. L'approccio più classico per eliminare il FPN è il double sampling (DS), che consiste nell'eliminazione degli offset costanti che affliggono i segnali sottraendo loro un valore costante afflitto dallo stesso offset. Il FPN di colonna viene eliminato inserendo circuiti di DS negli amplificatori di colonna, la cui risposta risulta così più omogenea fra dispositivi differenti. Un'alternativa è inserire alla base della colonna un pixel cieco, per cui il segnale viene dalla differenza fra il pixel cieco e il pixel illuminato, ma anche questo approccio non è perfetto per lo scopo, in quanto elimina solo il FPN di colonna.

Per eliminare il FPN dai pixel, invece esiste la possibilità di leggere due valori, tenendo lo stesso pixel abilitato per più di un ciclo di clock. In questo modo si legge prima il valore di fine integrazione, poi viene effettuato il reset del pixel, e se ne legge il valore di reset, quindi si ripete l'operazione per il pixel successivo. Il circuito di colonna, che deve acquisire il valore, salva la prima tensione misurata su un Sample/Hold analogico, quindi la sottrae dalla seconda tensione per ottenere il valore libero dall'offset causato dal FPN. Questo approccio riduce ovviamente il frame-rate, poiché ogni pixel deve restare abilitato più a lungo, aumentando la quantità di tempo necessaria per leggere la matrice. Inoltre, poiché un simile approccio sarebbe possibile solo in caso di rolling shutter, l'incremento del tempo di lettura di un pixel (read-out + reset) aumenta anche la disuniformità fra le righe dell'immagine.

Un approccio alternativo sarebbe alternare fasi di lettura e calibrazione del sensore ed effettuare la sottrazione a valle dell'ADC, via software. Mediando una serie di valori di reset (o comunque, di frame bui) acquisiti prima di far funzionare il sensore, si ottiene una matrice "buia" (pedistallo). In seguito, sottraendo dal pedistallo i nuovi valori letti, è possibile eliminare quasi del tutto la componente di FPN. Purtroppo non sempre è

possibile effettuare le preacquisizioni per calcolare il piedistallo, ed inoltre il piedistallo andrebbe aggiornato periodicamente, perché è possibile una deriva dei suoi valori dopo un lungo periodo di uso. Inoltre un simile approccio richiede la presenza di una certa quantità di logica a valle di tutto il sensore, che può essere critica se per il sensore sono richieste caratteristiche di portabilità e bassi consumi. Infine, bisogna ricordare che in certi sistemi d'acquisizione in cui bisogna processare grandi quantità di dati ad alte frequenze, può non essere sempre possibile avere tempo sufficiente per una correzione software, che è solitamente più inefficiente dal punto di vista temporale rispetto ad approcci a più basso livello.

Più complicata la possibilità di effettuare Correlated Double Sampling (CDS), che permetterebbe la riduzione non solo del FPN, ma anche del rumore kTC. Come precedentemente spiegato, il rumore kTC è modellizzabile come un'incertezza sul valore finale di reset, che di conseguenza si ripercuote sul valore finale di integrazione. Gli approcci di Double Sampling sopra descritti non sono in grado di cancellare il kTC, anzi, lo peggiorano, perché sottraggono al valore letto il valore di reset successivo, che è afflitto da un valore di rumore kTC differente (incorrelato al primo), peggiorando il kTC (in tensione) di un fattore $\sqrt{2}$. Inoltre, ogni metodo di double sampling, sia correlato che incorrelato, tende a peggiorare il rumore ad alta frequenza, sempre di un fattore $\sqrt{2}$, a causa della somma in potenza di due contributi dello stesso tipo di rumore fra loro incorrelati.

Per eliminare il kTC effettuando un vero CDS sarebbe necessario salvare prima il valore di reset e poi il valore di fine integrazione corrispondente, ma questo risulterebbe complicato, perché richiederebbe di leggere due volte non consecutive ogni pixel, alla fine della fase di reset e alla fine della fase d'integrazione. Inoltre, poiché la fase d'integrazione risulta lunga relativamente alle altre fasi del periodo di funzionamento, sarebbe necessario tenere memoria dei valori analogici di reset di tutta la matrice per poter implementare efficacemente questa tecnica; infatti nel tempo che intercorre tra il reset e la successiva lettura di un pixel, la matrice deve scandire tutti gli altri pixel. Inoltre una simile operazione richiederebbe una matrice formata da tanti S/H analogici quanti sono i pixel, con un evidente

peggioramento dei costi del sensore. Come si vedrà, è comunque possibile superare questa limitazione mediante modifiche alla struttura del pixel APS.

Effetti dello scaling sui dispositivi APS

Lo scaling del MOS verso dispositivi sempre più ridotti, oltre a migliorare le prestazioni sotto molti aspetti, richiede tuttavia una maggiore attenzione nel design, ponendo in essere nuove sfide progettuali. La diminuzione della lunghezza di canale e delle dimensioni dei componenti elettronici (transistor, diodi) permette la creazione di pixel di dimensioni sempre minori, con conseguente miglioramento della risoluzione spaziale o del fill-factor (minore area occupata dai transistor e dalle metallizzazioni). Tipicamente, data una lunghezza di canale minima λ , il minimo pitch per un pixel funzionante è di $10-20\lambda$ [1]. In realtà la relazione fra dimensioni del pixel e minima lunghezza di canale non sono così lineari, in quanto non tutte le diffusioni e metallizzazioni riescono, per vari motivi, a scalare di dimensioni seguendo lo stesso trend della lunghezza di canale, per cui un dimezzamento della lunghezza minima λ difficilmente permetterebbe di realizzare un pixel con un quarto della area. Di conseguenza oltre un certo livello di integrazione tecnologica, lo scaling dei pixel non risulta più conveniente, in quanto il miglioramento in risoluzione non è più sufficiente a compensare una serie di nuovi svantaggi. Infatti mentre la diminuzione della tensione di alimentazione tende ad essere proporzionale allo scaling, le tensioni di soglia non diminuiscono seguendo lo stesso trend, erodendo di conseguenza la tensione utile di segnale, a cui si unisce anche una maggiore conversione carica-tensione (maggiore swing in seguito alla medesima illuminazione) per la diminuzione della capacità del fotodiodo [6]. Di conseguenza, per ottenere un dynamic range accettabile, risulta necessario impiegare tecniche alternative, sfruttando la possibilità di inserire logica supplementare all'interno del pixel stesso.

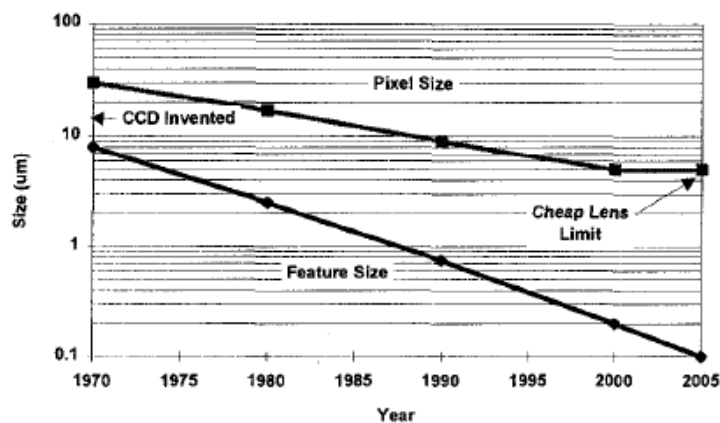


Figura 10 - Confronto fra i trend nella dimensione del pixel e minima lunghezza di canale per CMOS nel visibile [2].

Inoltre il rumore in tensione di tipo kTC a causa delle minori dimensioni del pixel e di conseguenza della minore capacità risulta peggiorare con lo scaling nella misura di $1/L$, poiché la capacità in prima approssimazione scala con $1/L^2$. La diminuzione di dimensioni del fotodiodo comporta altresì un maggiore impatto della corrente di buio sul rumore totale del pixel. Infine si ha un peggioramento del fixed pattern noise in quanto a dimensioni sub-micrometriche la variazione delle tensioni di soglia e degli altri parametri di dispositivo nominalmente identici tende ad essere molto marcata.

Approcci alternativi

Per ovviare ai problemi descritti, oltre ad implementare nuove caratteristiche utili nell'elettronica ancillare, sono state proposte una serie di modifiche allo schema classico dei pixel a 3 transistor.

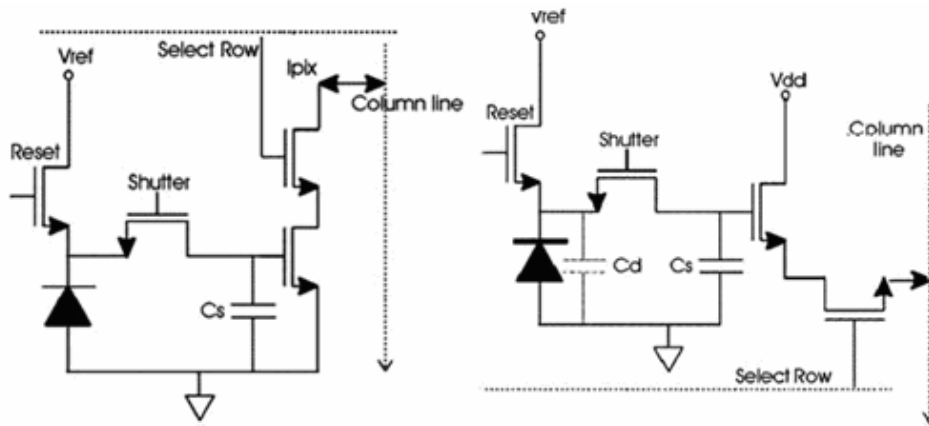


Figura 11 - Esempi di schematici di pixel 4T [28]

Uno dei più popolari è il pixel 4T [9],[28], in cui viene inserito un pass-transistor tra il diodo e il transistor C-D. In questo modo è possibile dimensionare la capacità del fotodiodo senza preoccuparsi della conversione carica-tensione. Infatti nel decidere le dimensioni del fotodiodo è necessario ricordare che una n-well di grandi dimensioni permette di raccogliere maggior carica (maggior estensione della zona svuotata). Tuttavia in tal caso il FD ha anche un valore di capacità maggiore, per cui il guadagno di tensione in funzione della carica raccolta viene penalizzato. Con un pixel 4T invece il fotodiodo può essere realizzato di dimensioni maggiori, mentre la lettura della carica viene effettuata trasferendo la carica raccolta dal fotodiodo su un nodo a bassa capacità (la floating diffusion), e consentendo così sia una maggior area di raccolta che un elevato guadagno tensione/carica. In questi approcci viene impiegato un fotodiodo 'pinned', formato da una struttura p+np- [10] con le diffusioni p tenute a massa, che svuota completamente la diffusione n. In questo modo, quando il transistor TX è acceso, la carica viene trasferita completamente sulla diffusione flottante, a patto che questa mantenga una tensione maggiore di quella delle diffusioni p.

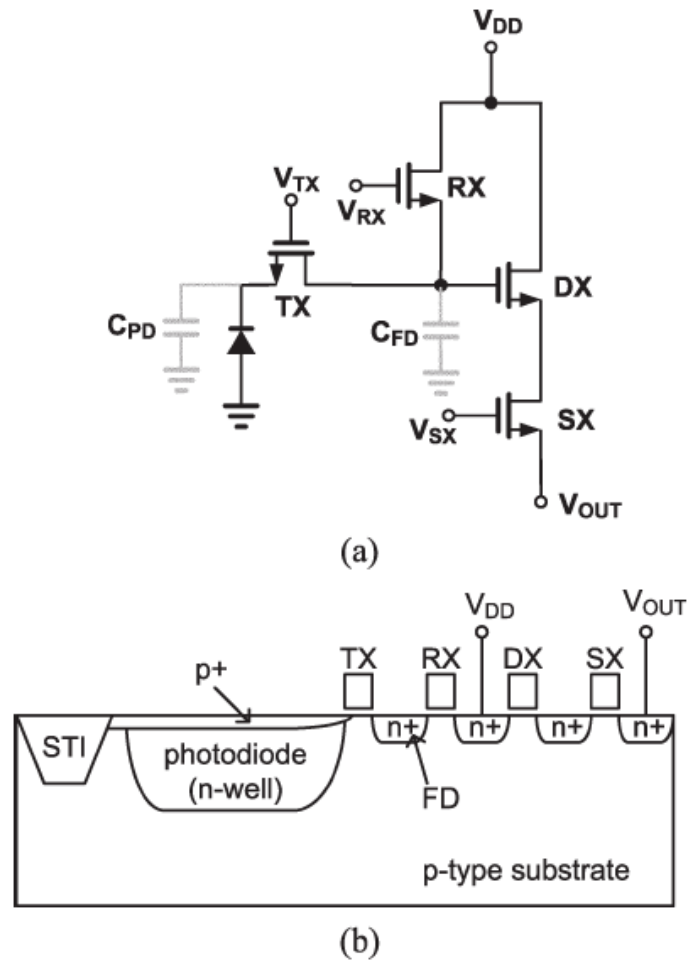


Figura 12 - Schema del pixel 4T con diodo pinned [9]

Un funzionamento alternativo è la possibilità di effettuare un vero CDS. In questo caso il S/H analogico su cui salvare il valore di reset non si trova esterno al pixel, ma dentro il pixel stesso, ed è costituito dalle capacità parassite del gate del transistor common-drain e della floating diffusion sul drain del pass-transistor. Aprendo il pass-transistor subito dopo la fine della fase di reset, sulla floating diffusion viene salvato il valore di reset. Per

effettuare la lettura, viene prima letto il valore sulla floating diffusion (valore di buio o di reset) e poi viene chiuso il pass-transistor, in modo da trasferire sulla floating diffusion la fotocarica raccolta dal diodo, e quindi tramite la differenza fra i due segnali si ottiene l'escursione utile, in cui i contributi di rumore di kTC e FPN risultano fortemente ridotti. Infine, con questi componenti è possibile effettuare un global shutter vero e proprio. In questo caso, tutti i pixel vengono resettati allo stesso momento, quindi inizia la fase d'integrazione a pass transistor acceso. Al termine di questa, il pass-transistor di ogni pixel viene spento, salvando sulla capacità del MOS C-D il valore finale del fotodiodo. Viene quindi effettuata la lettura di tutta la matrice, sapendo che il frame ottenuto è omogeneo sia come tempo di esposizione che come istante di inizio-esposizione.

Spesso il pixel 4T prevede l'impiego di un photogate [11] come elemento sensibile al posto del fotodiodo, realizzando un funzionamento simile al CCD. Si tratta di inserire un ulteriore transistor nMOS al posto del fotodiodo, che viene tenuto a tensioni maggiori di V_{th} durante la fase di integrazione. In questo modo si facilita la formazione del canale n, a cui si aggiungono le cariche negative fotogenerate dalla radiazione incidente. In lettura, il gate viene riportato a 0 mentre il pass-transistor viene attivato, in modo che il canale venga chiuso, e tutte le cariche negative vengano attratte dal nodo a tensione maggiore, cioè la diffusione n+ di lettura. In questo modo il trasferimento della carica fotogenerata al nodo di lettura a bassa capacità è più efficace, migliorando la conversione carica-tensione del pixel. Il prezzo da pagare in questo caso è la presenza di un gate di polisilicio sopra all'area sensibile del pixel, che può penalizzare l'efficienza di rivelazione di alcuni tipi di radiazione, soprattutto nel visibile.

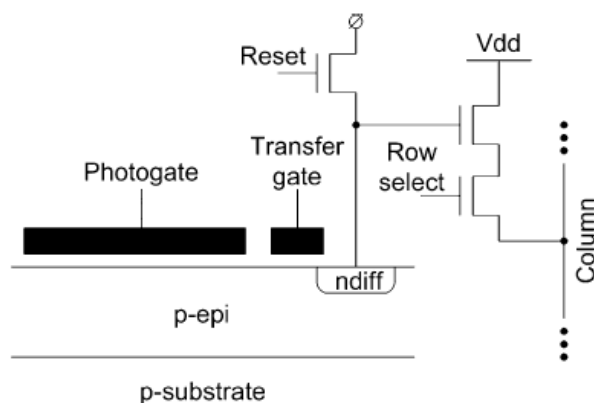


Figura 13 - Pixel 4T con photogate [11]

Come detto in precedenza, la presenza di pMOS all'interno del pixel è sconsigliabile per non penalizzare la raccolta di carica da parte del fotodiodo. In molti casi è tuttavia possibile inserire comunque dei pMOS, ipotizzando che la perdita di carica causata da questi sia comunque compensata dalla possibilità di implementare una circuiteria pienamente CMOS all'interno del pixel. Questo ultimo approccio risulta più praticabile nelle tecnologie più recenti, che permettono di creare dispositivi pMOS di dimensioni ridotte e di conseguenza in grado di sottrarre meno carica al fotodiodo. Inoltre i processi più recenti mettono a disposizione una serie di strumenti che permettono la creazione di p-well profonde in grado di isolare le n-well dei pMOS (INMAPS [14]), limitandone la raccolta parassita di carica unicamente alla propria n-well.

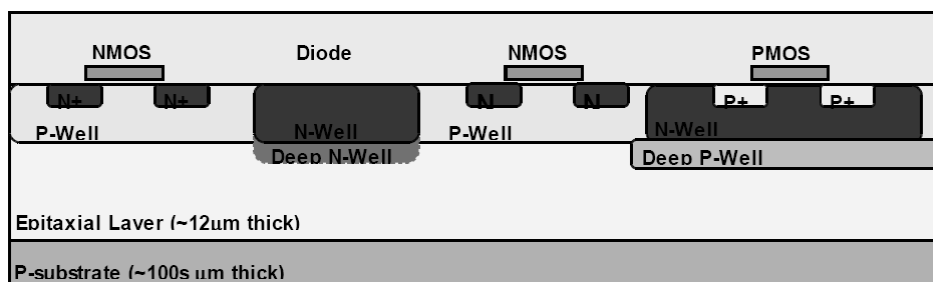


Figura 14 - Isolamento della n-well del pMOS mediante p-well profonda [14]

Un'alternativa interessante e speculare alla precedente, allo scopo di inserire pMOS senza penalizzare l'efficienza nella raccolta di carica e di impiegare circuiti più complessi, è la possibilità di impiegare un processo twin-well che permetta di creare tasche p all'interno di una n-well profonda [12]. In questo modo, gli nMOS possono essere realizzati direttamente all'interno del fotodiode, che può occupare così tutta l'area del pixel. Questo approccio permette di inserire sul pixel una gran quantità di elettronica, oltre alle n-well del pMOS, che si trovano ad occupare una superficie trascurabile rispetto al fotodiode (e a raccogliere una quantità di carica proporzionale). Il problema derivante dall'aumentata capacità del fotodiode viene superato perché, sfruttando il maggior numero di componenti così impiegabili (ora tutto il pixel può contenere nMOS), è possibile utilizzare approcci più efficienti nella conversione carica-tensione che non la semplice capacità del fotodiode, implementando per esempio un amplificatore di carica direttamente a livello di pixel. Con questi dispositivi è stata stimata un aumento di un fattore 3 nella raccolta di carica ed è stato misurato un SNR di circa 25 con una minimum ionising particle (MIP). Un simile approccio non è impiegabile però nell'ambito del visibile, in quanto la luce visibile genera carica solo in superficie, dove verrebbe bloccata dall'elettronica.

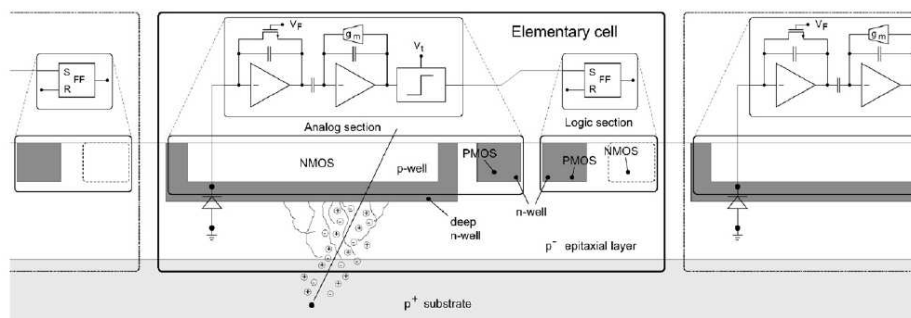


Figura 15 - Esempio di pixel con n-well profonda. Gli nMOS della sezione digitale sono creati sul bulk invece che nella p-well profonda per isolare la sezione analogica dalla sezione digitale, maggiormente rumorosa [12]

Per migliorare la velocità di lettura e ridurre il rumore, si è pensato di introdurre dispositivi dotati di convertitore analogico/digitale direttamente sul pixel (Digital Pixel Sensor, DPS) [5]. In questo caso, all'interno del pixel vengono perciò inseriti, assieme al fotodiodo, un comparatore e un registro parallelo a 8 bit, per creare un ADC a rampa singola. Durante la fase di lettura, sul secondo ingresso del comparatore viene messa una rampa di tensione, mentre il registro viene caricato da un contatore esterno (comune a tutta la matrice) tramite un bus digitale d'ingresso. Quando il valore della rampa supera il valore in uscita dal fotodiodo, il comparatore commuta, disabilitando il registro e di conseguenza salvando su di esso l'ultimo valore del contatore esterno. Alla fine della rampa, vengono letti tramite il bus digitale i valori salvati sul registro di ogni pixel e si ottiene il valore della matrice. Questo approccio permette di effettuare facilmente il CDS (con una rampa all'inizio della fase di reset si hanno i valori di buio, quindi alla fine della fase d'integrazione si hanno i valori di luce, già in formato digitale) e permette una buona velocità di lettura della matrice, in quanto si deve leggere una sequenza di valori digitali e non analogici. La durata della conversione è di circa 25us, mentre la lettura di una matrice di circa 100k pixel tipicamente richiede circa 75us, ma può essere effettuata in parallelo al periodo di integrazione del frame successivo, permettendo una frame-rate di circa 10k frames/s.

Gli svantaggi, oltre al basso fill factor, risiedono nel consumo, essendo il comparatore sempre acceso (50mW per tutto il chip, di cui 30 per i comparatori), oltre alla difficoltà di realizzare i componenti necessari sul pixel in tecnologia nMOS.

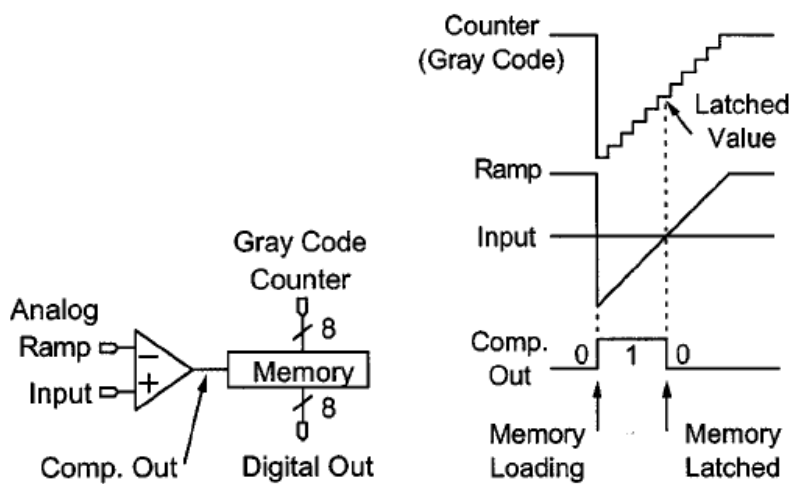


Figura 16 - Schema di funzionamento dell'ADC del DPS

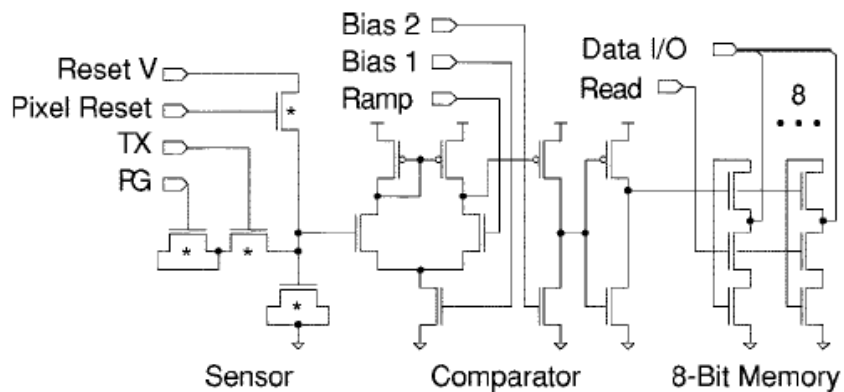


Figura 17 - Schematico del DPS con ADC a livello pixel [5]

Per ovviare ai problemi descritti nel capitolo relativo al reset del diodo, esistono in letteratura approcci differenti rispetto i reset soft e hard, come il flushed reset [13] e il reset attivo [11]. Il primo prevede di combinare hard e soft reset in modo da ottenere sia la soppressione dell'Image Lag che la riduzione del kTC. In particolare, si precarica il pixel ad un valore predefinito in hard-reset, tenendo il drain del transistor di almeno una tensione di soglia sotto a V_{dd} , in modo da eliminare l'Image Lag. Quindi si alza la tensione di drain a V_{dd} per avere una seconda fase di soft-reset, in cui il rumore kTC viene abbattuto. La parte in soft-reset non può ovviamente introdurre alcun Image Lag, in quanto il valore sul catodo del fotodiodo all'inizio di tale fase è stato portato ad una tensione sicuramente indipendente dal valore precedente.

Il reset attivo [11] invece prevede la riduzione del rumore kTC impiegando una retroazione: durante il reset l'uscita del pixel viene collegata ad un amplificatore operazionale in modo che il pixel diventi il ramo di retroazione di un op-amp in configurazione a guadagno unitario, la cui uscita viene a sua volta collegata al gate del transistor di reset. All'ingresso di questa configurazione viene quindi mandata una rampa di tensione, che sale fino al valore di reset desiderato, imponendo quel valore anche sul fotodiodo per cortocircuito virtuale. In questo modo vengono a formarsi due anelli di loop:

- un'amplificazione della capacità vista dal fotodiodo, pari a $(A+1)C$ dove C è la capacità gate-source del transistor di reset e A il guadagno ad anello aperto dell'operazionale,
- la modulazione della corrente di drain del transistor di reset

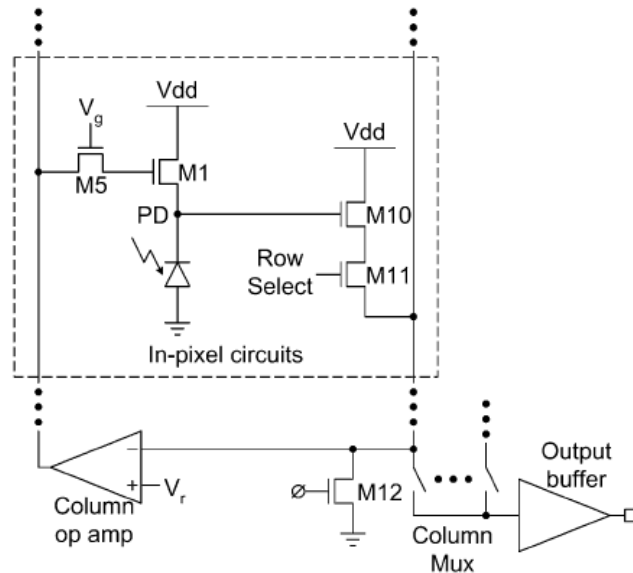


Figura 18 - Schema globale della struttura a reset attivo [11]

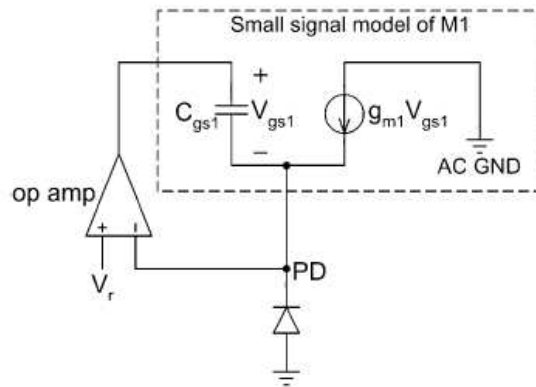


Figura 19 - Schema ai piccoli segnali della struttura a reset attivo[11]

È dimostrabile che un simile approccio riduce sensibilmente l'impatto del kTC e del FPN sul pixel, in quanto riduce il rumore di canale sul transistor di reset, principale responsabile del kTC, con una riduzione di 7.6 volte del rumore in uscita (120uV contro 910uV di un APS normale). La riduzione

del FPN è invece dovuta all'imposizione mediante retroazione di una tensione identica per tutti i pixel alla fine della fase di reset, con un miglioramento misurato di 21 volte (0.7mV di FPN di colonna contro i 15mV dell'APS normale).

È inoltre in grado di eliminare l'immagine lag, al prezzo di un fill-factor inferiore, dovuto all'introduzione di un ulteriore transistor per collegare il MOS di reset all'uscita dell'op-amp e ad un aumento del tempo necessario per terminare le operazioni di reset..

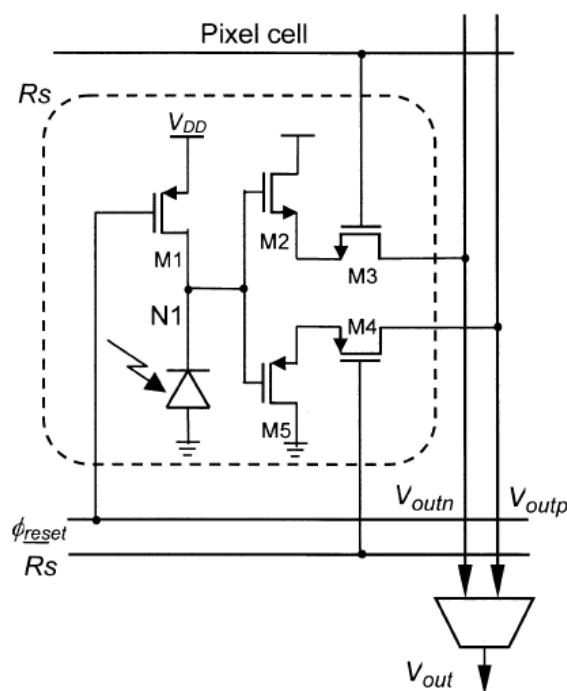


Figura 20 - Schema del pixel CAPS [6]

Per porre un rimedio ad uno dei massimi problemi posti dallo scaling, cioè la riduzione dello swing di tensione impiegabile a basse tensioni di alimentazione, è stato proposto un pixel denominato CAPS (Complementary Active Pixel Sensor) [6], in cui il fotodiode (resettato tramite un pMOS) viene letto tramite una catena nMOS normale e, in parallelo, tramite una

catena pMOS. In questo modo si hanno due segnali, uno con swing da 0 a $V_{dd}-V_{th}$ e l'altro con swing da V_{th} a V_{dd} da cui viene ricostruito il valore finale, permettendo di sfruttare tutta l'escursione di tensione possibile

Danno da radiazione

Uno dei parametri più importanti per lo sviluppo di sensori di radiazione per fisica delle alte energie e applicazioni medicali è la resistenza al danno da radiazione. Lavorando infatti in ambienti fortemente esposti a radiazioni ionizzanti, i sensori a semiconduttore sono soggetti sia a danneggiamenti a lungo termine, sia a malfunzionamenti causati da singoli eventi. Il danno da radiazione viene solitamente diviso in due categorie principali [16]:

- Danno da esposizione: dipende dalla quantità totale di particelle che hanno colpito il sensore nella sua vita, e si manifesta come un lento decadimento delle prestazioni
- Danno da singolo evento: avviene in seguito all'impatto con una sola particella ionizzante, che in certe condizioni critiche può influire in modo più o meno dannoso con il funzionamento del dispositivo. Si divide a sua volta in danno distruttivo e non distruttivo

Si tratta di due tipologie abbastanza differenti, che vengono esaminate separatamente. In generale, il danno da esposizione è più pericoloso per i circuiti analogici, mentre per i circuiti digitali tendono ad essere maggiormente critici i malfunzionamenti da singolo evento.

Danno da esposizione

Un sensore sottoposto ad un flusso continuo di particelle può subire diversi tipi di danneggiamento, in funzione della dose ricevuta e del tipo di particelle incidenti.

Per caratterizzare l'esposizione a radiazione potenzialmente dannosa, si fa riferimento ai seguenti parametri:

- Flusso: indica la quantità di particelle per unità d'area e di tempo
- Fluenza: Indica la quantità di particelle per unità d'area durante tutto il tempo di funzionamento del dispositivo. È dato dall'integrale rispetto il tempo del flusso.
- Dose: indica l'energia totale ceduta dalla radiazione per unità di massa

Un tipico sensore per raggi X con scintillatore, per esempio, riceve una dose dell'ordine dei krad durante la propria vita di funzionamento, mentre sensori pensati per applicazioni di fisica delle alte energie ricevono dosi ancora superiori (>100krad) [17].

Il tipo di danno assume inoltre forme diverse a seconda del tipo di particelle incidenti. La distinzione principale è fra

- Radiazioni non ionizzanti (neutroni): causano perlopiù danni al reticolo
- Radiazioni ionizzanti (elettroni, protoni, raggi X e γ , ioni): causano danni al reticolo e generano cariche negli ossidi. Possono inoltre modificare i profili di drogaggio

Il tipo di danno più pericoloso deriva dalla seconda tipologia di radiazioni.

Quando una di queste particelle incide sull'ossido di silicio, lo ionizza, creando delle coppie elettrone-lacuna al suo interno. Se l'elettrone, grazie alla maggiore mobilità all'interno dell'ossido ($20\text{cm}^2/\text{Vs}$) viene allontanato dalla sede dell'impatto, la lacuna, la cui mobilità è 6 ordini di grandezza inferiore ($2\cdot 10^{-5}\text{cm}^2/\text{Vs}$), resta intrappolata, causando nel lungo periodo un accumulo di cariche positive nell'ossido. A seconda del tipo di ossido in cui avviene, questo fenomeno causa due effetti differenti [18]:

- Ossido di gate: le cariche positive nell'ossido, spinte verso l'interfaccia da $V_{\text{GB}} > 0$ (nMOS) facilitano la creazione del canale
- Ossido d'isolamento: l'ossido di isolamento tende ad agire da nMOS parassita, causando un aumento della corrente di drain (anche sottosoglia).

Le lacune generate nell'ossido si spostano lentamente e tendono a restare intrappolate nei difetti dell'ossido, che sono più densi vicino all'interfaccia con il bulk, e questo effetto si ha sia soprattutto nell'nMOS, causando una

variazione nella tensione di soglia che scende perché le cariche positive facilitano la creazione di un canale n. Oltre all'accumulo di cariche nell'ossido, va considerato che la radiazione crea del danno reticolare sull'interfaccia Ossido-Silicio, creando dei centri assimilabili a trappole per le cariche mobili. Di conseguenza, questi difetti tendono ad ostacolare la creazione del canale nei MOS, portando ad un aumento della tensione di soglia sia per i transistor di canale p (in modulo) che di canale n. Questo stesso fenomeno causa una degradazione della transconduttanza (direttamente proporzionale alla mobilità nel canale), con conseguente peggioramento sia del rumore termico che del rumore 1/f. Poiché è un fenomeno preponderante ad elevate dosi, osservando la variazione di tensione di soglia dell'nMOS in funzione della dose si osserverà in particolare una diminuzione per valori bassi seguita da un aumento in seguito a irraggiamenti più elevati.

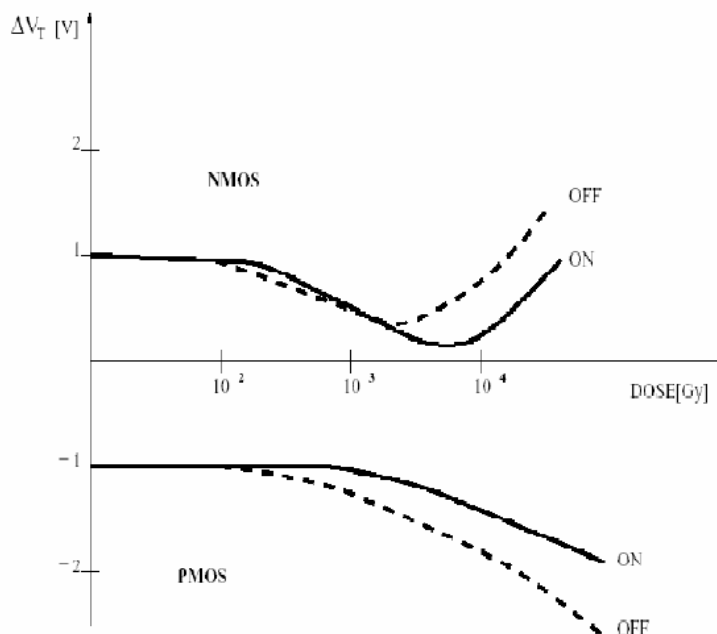


Figura 21 - Valore della tensione di soglia per MOS p e n in funzione della dose [20]

L'intrappolamento di cariche positive nel bulk, soprattutto nelle regioni di bird's beak (cioè dove l'ossido di isolamento si assottiglia in vicinanza del contatto di gate) tende, sopra a substrati di tipo p, a facilitare la creazione di canali n parassiti fra drain e source dello stesso nMOS, e fra nMOS differenti.

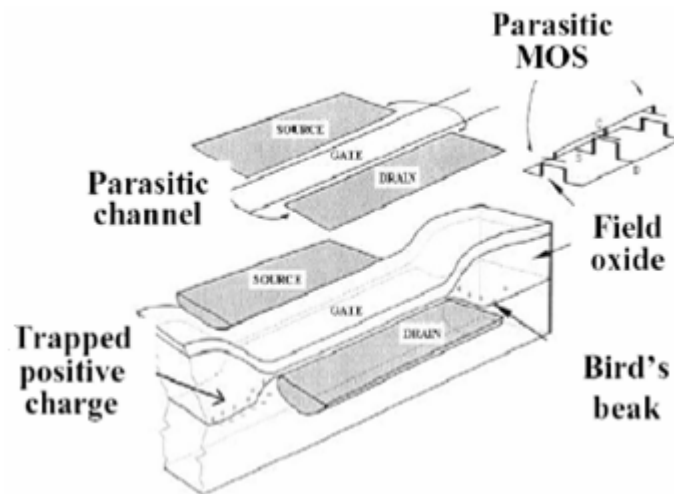


Figura 22 - Creazione di canali parassiti fra S e D per ionizzazione dell'ossido di isolamento [20]

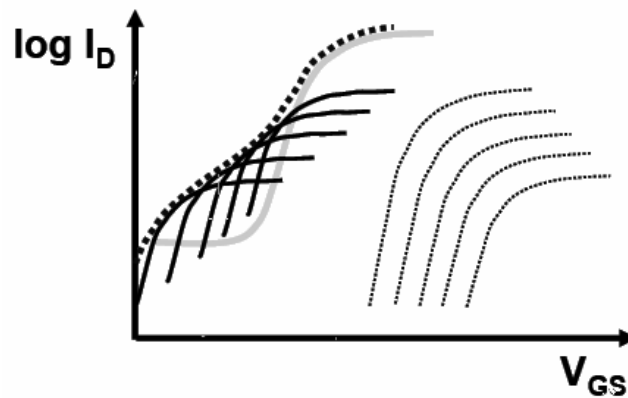


Figura 23 - Variazione della corrente di drain in funzione di V_{GS} . I transistor parassiti (tratteggiati) per l'abbassamento della tensione di soglia si attivano a V_{GS} dell'ordine

delle tensioni di segnale (linea continua) modificando pesantemente la caratteristica del MOS [20]

In condizioni normali, per evitare questo fenomeno, è sufficiente, con opportuno drogaggio del substrato, far sì che questi canali abbiano una tensione di soglia molto alta. Ma la presenza di cariche positive parassite nell'ossido indotte da dosi massicce di radiazione abbassa questa tensione di soglia, causando un aumento della corrente di sottosoglia negli nMOS, con conseguente aumento della potenza dissipata.

Inoltre le particelle incidenti, soprattutto quelle pesanti, come protoni, neutroni o nuclei atomici a bassa energia, possono danneggiare il reticolo cristallino (displacement damage), creando delle impurità locali, sia in singoli punti che in intere regioni (cluster), che agiscono come trappole per i portatori minoritari, diminuendone la vita media.

In particolare [19] a seconda della particella e dell'energia si possono trovare le seguenti tipologie di danno nel silicio

- Raggi γ : creano elettroni di Compton con energia $<1\text{Mev}$: difetti di tipo puntuale
- Elettroni: creano difetti puntuali ($E>255\text{keV}$) o cluster ($E>8\text{MeV}$)
- Neutroni: difetti puntuale ($E>185\text{eV}$) o cluster ($E>35\text{keV}$)

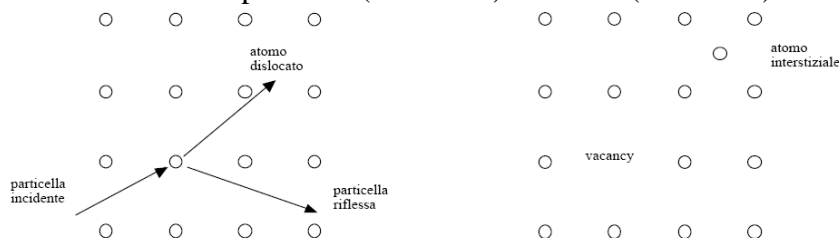


Figura 24 - Modalità di creazione del danno reticolare: una particella incidente scalza l'atomo dalla sua posizione, lasciando un interstizio vuoto [20]

Nel caso dei sensori CMOS, queste singolarità possono “catturare” gli elettroni fotogenerati nel bulk p, impedendo loro di raggiungere il fotodiode e pertanto causando una minore raccolta rispetto il dovuto. Gli elettroni catturati potrebbero quindi essere rilasciati durante un ciclo di funzionamento successivo, causando una sovrastima della carica raccolta.

Anche la corrente di perdita del fotodiode tende a peggiorare con l'irradiazione, per effetto sia del displacement damage (con forte dipendenza dal tipo di irraggiamento, soprattutto neutroni lenti) che della radiazione ionizzante.

Un ulteriore problema, nel caso di irraggiamento da ioni, è la tendenza a modificare i profili di drogaggio del substrato, spesso fino al punto di causare un'inversione (zone drogate p si comportano come drogate n e viceversa). La relazione trovata negli esperimenti di fisica che mostra la variazione del drogaggio rispetto la fluenza ϕ è [23]:

$$N_{eff} = N_{eff}(0)e^{-c\phi} - \beta\phi$$

Con β e c parametri empirici dipendenti dal tipo di irraggiamento. Questa formula mostra come dopo un certo livello di irraggiamento il substrato si comporti come se avesse un drogaggio diverso. È comunque possibile riportare il sensore ad un funzionamento accettabile mediante annealing. Infatti tenendo il sensore alle adeguate condizioni di temperatura e per un certo tempo, i profili di drogaggio e i danni reticolari tendono a ridistribuirsi seguendo la configurazione originale, permettendo un parziale recupero delle prestazioni. Allo stesso modo l'annealing termico può facilitare la ricombinazione delle cariche intrappolate nell'ossido.

Lo scaling ha altresì un effetto positivo dal punto di vista del danno da radiazione, in quanto nelle tecnologie sub-micrometriche gli ossidi risultano più sottili (~2nm), riducendo il rischio di generazione di cariche positive nell'ossido di gate. Infatti si sa che [20]

$$\Delta V_{OT} = -\frac{q}{\epsilon_{OX}} t_{OX} \Delta N_{OT}$$

Con ΔN_{OT} densità di cariche nell'ossido, a sua volta proporzionale a t_{OX} in quanto più un ossido è spesso più è facile che venga ionizzato. Di conseguenza la variazione della tensione di soglia risulta ridimensionata diminuita (dell'ordine di t_{OX}^2) con lo scaling.

Inoltre, sempre a causa del ridotto spessore dell'ossido, è più facile che le cariche imprigionate vengano disattivate tramite annealing per effetto tunnel. Infatti in questo caso le cariche intrappolate sono più vicine al canale e di

conseguenza è più facile che un elettrone superi la barriera di potenziale dell'ossido, neutralizzando una buca intrappolata. Anche l'abbassamento della tensione di soglia per il danno d'interfaccia risulta fortemente ridimensionato, in quanto nei processi tecnologici più avanzati è possibile creare interfacce ossido-bulk di migliore qualità e di conseguenza meno suscettibili al danno da radiazione.

Invece i problemi relativi all'ossido di isolamento (che anche per le tecnologie avanzate ha spessori di 100nm-1µm) sono più critici, con l'aumento di corrente sottosoglia negli nMOS che non può essere trascurato. In questo caso si cerca di correggere il problema impiegando strutture enclosed-gate, in cui cioè il gate ha forma di anello e circonda completamente il contatto di source (o di drain) in modo che non possano formarsi canali di bypass fra S e D. I canali parassiti fra due MOS differenti possono invece essere evitati mediante l'inserimento di anelli di guardia attorno ad ogni MOS. Questi anelli, formati da zone drogate p tenute a massa impediscono la formazione di un canale n, al prezzo di una maggior occupazione di area.

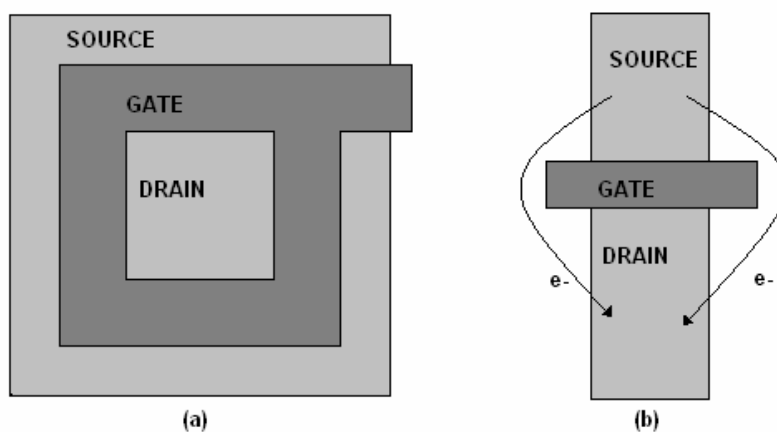


Figura 25 - Transistor enclosed-gate (a) e transistor lineare (b)

Danno da singolo evento

I danni da radiazione legati ad un singolo evento si suddividono invece in due categorie, distruttivi e non distruttivi. In ambo i casi il meccanismo dell'evento è da ricollegare al rilascio di grandi quantità di carica nel substrato da parte di una particella fortemente ionizzante. Questo fenomeno è particolarmente critico nelle applicazioni spaziali (l'atmosfera terrestre blocca la maggior parte di queste particelle, che sono invece comuni nello spazio profondo) e in quelle di fisica delle alte energie.

Come è noto, uno dei maggiori rischi della tecnologia CMOS è l'innesco del latch-up, causato dalla presenza di un circuito parassita all'interno dell'architettura base della tecnologia (nMOS in serie a un pMOS). Si tratta di due transistor bipolari in retroazione: se in un qualsiasi momento la tensione di base-emettitore di uno dei due transistor supera la tensione di soglia (cioè la tensione locale del substrato sale/scende di V_{γ} rispetto massa/alimentazione), questo si accende, causando l'accensione dell'altro, che a sua volta aumenta la corrente nel gate del primo. I due transistor, sostenendosi a vicenda, raggiungono così in fretta lo stato saturato, causando di fatto un cortocircuito fra massa ed alimentazione, che può risultare letale per il circuito. Quando una particella colpisce il reticolo, può rilasciare una notevole quantità di carica nel substrato. Se incide nella zona svuotata, in particolare, questa carica si traduce in un forte picco di corrente, che agisce da corrente di base di uno dei bipolari parassiti, innescando il latch-up e di conseguenza distruggendo il dispositivo.

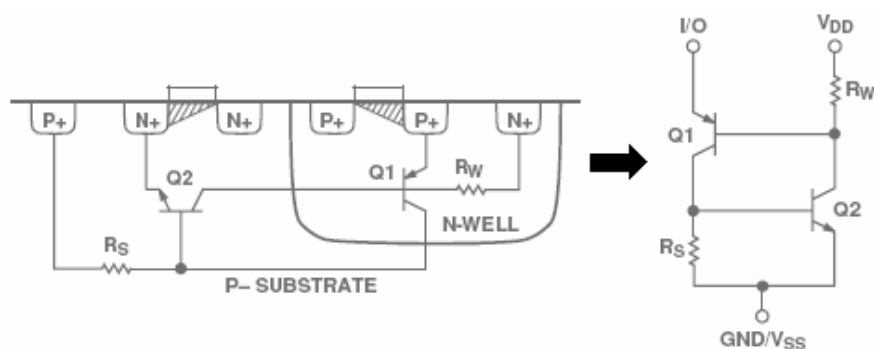


Figura 26 - Dispositivi parassiti nel CMOS e schema della struttura di latch-up [21]

Per evitare questo fenomeno, si cerca di abbassare il più possibile la resistenza parassita in parallelo alla giunzione base-emettitore del bipolare parassita, in modo che sia necessaria una corrente maggiore per raggiungere la tensione di innesco. Un metodo potrebbe essere l'impiego di un maggior drogaggio del substrato, ma questo peggiorerebbe l'isolamento del MOS rispetto al bulk, a causa di una più stretta zona svuotata. Per ovviare a ciò si può impiegare uno strato epitassiale che permette l'uso di un bulk a bassa resistività, che è schematizzabile come una resistenza di basso valore in parallelo ad una resistenza maggiore (cioè lo strato epitassiale).

In alternativa si può inserire il maggior numero possibile di contatti sustrato-massa e n-well-alimentazione per far sì che le correnti parassite abbiano un cammino breve nel substrato e di conseguenza diano origine a limitate differenze di tensione locali (nailing del substrato). Un ulteriore metodo è invece l'inserimento di trincee di ossido (LOCOS o STI) fra i pMOS e gli nMOS, per interrompere i bipolari parassiti [18],[21].

Meno gravi, ma comunque dannosi sono invece gli effetti da singolo evento non distruttivi. Una particella che rilasciasse carica in un dato punto del dispositivo, può, come si è detto, alterarne temporaneamente il potenziale anche di un valore notevole. Se in un circuito analogico normalmente questo effetto si traduce con un picco di tensione (rumore istantaneo), solitamente senza conseguenze durevoli dal punto di vista del funzionamento del dispositivo, in un circuito digitale può indurre un cambio di valore logico

- '1' al posto di '0' in una diffusione p (raccolta di lacune)
- '0' al posto di '1' in una diffusione n (raccolta di elettroni)

rendendo inaffidabili i risultati osservati. In particolare, sono possibili:

- SET (Single event transient): in una rete combinatoria, causa un picco di tensione spurio che può esser campionato dal latch a valle, causando errore. L'occorrenza di questo tipo di errore aumenta di conseguenza con la frequenza di funzionamento del circuito (più facile che il picco avvenga in corrispondenza dell'istante di campionamento)
- SEU (Single event upset): la radiazione induce la variazione del valore immagazzinato in un flip-flop o latch.

In ogni caso, si cerca di evitare questi problemi con soluzioni a livello di standard-cells per i circuiti destinati ad ambienti irradiati [18]. Al costo di una maggiore occupazione d'area, è infatti possibile rendere i circuiti digitali robusti rispetto a questi malfunzionamenti. Un primo metodo per contenere i SEU è l'aumento della capacità del nodo di memorizzazione, con l'inserimento di capacitori in parallelo, in modo che più difficilmente la perturbazione di carica causi una commutazione di valore logico. In alternativa, in un flip-flop retroazionato, il cambio dello stato di un nodo innesca una corsa critica, dove il nodo rimasto al valore corretto tende a correggere il nodo sbagliato e viceversa. Di conseguenza si può inserire una rete RC fra i due inverter, in modo che la perturbazione sull'uscita di un inverter impieghi del tempo per modificare la tensione all'ingresso dell'altro inverter (mentre il valore corretto non ha bisogno di propagarsi ed agisce immediatamente sull'ingresso dell'inverter la cui uscita è stata invertita). Ambedue questi metodi hanno il difetto di peggiorare le caratteristiche dinamiche dei flip-flop.

Sono stati pertanto sviluppati layout di flip-flop resistenti al SEU, in quanto immagazzinano l'informazione in più nodi fisicamente distanti e le cui tensioni devono esser variate contemporaneamente per commutarne lo stato. Poiché è improbabile che due o più particelle ionizzanti colpiscano contemporaneamente il substrato, e proprio nei punti critici, questa architettura risulta efficace nella soppressione dei SEU.

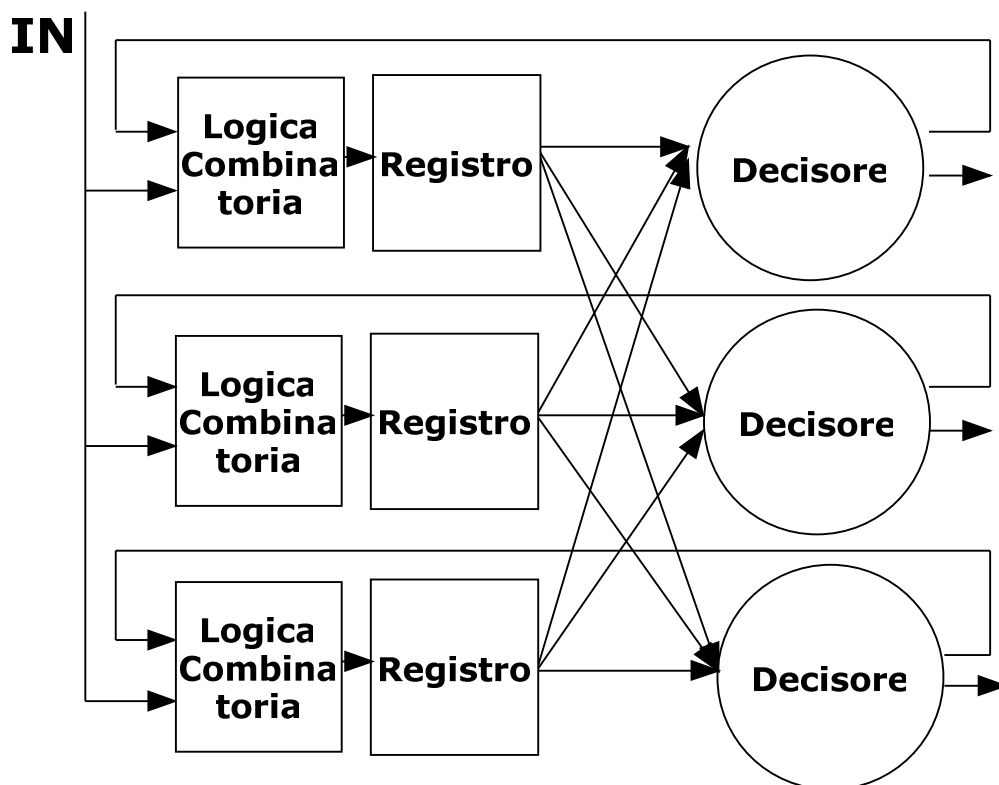


Figura 27 - Schema di circuito digitale resistente agli effetti da singolo evento. Ogni blocco è triplicato, decisore compreso, e la scelta dell'uscita giusta è fatta a maggioranza

Più complesso il metodo per la riduzione dei SET, in quanto richiede di aumentare la ridondanza dei circuiti. Ogni blocco di logica viene di conseguenza triplicato, e alla fine i 3 valori dei blocchi identici sono mandati ad un circuito “arbitro” che decide a maggioranza quale sia il valore corretto. Anche questo approccio si basa sull’ipotesi che sia altamente improbabile che un SEU colpisca contemporaneamente due o più aree correlate del dispositivo. Per rendere ulteriormente affidabile il circuito nel caso che una radiazione colpisca il circuito arbitro, anche questo può esser

triplicato in modo che la decisione venga presa in base all'interazione fra questi tre elementi. In alternativa si può campionare l'uscita dello stesso blocco in 3 istanti diversi, in modo che il picco di tensione influenzi un unico campionamento, ma questo approccio richiede l'impiego di tre clock sfasati, che può essere problematico da implementare, soprattutto se è richiesta un'elevata frequenza di funzionamento.

Rivelazione di raggi X

Perché una radiazione incidente rilasci carica nel silicio, come detto in precedenza, è necessario che l'energia di questa sia dello stesso ordine di grandezza dell'energia band-gap del silicio che è pari a 1.1eV. In realtà questa energia di band-gap non è quella che deve essere fornita dalla luce per spostare un elettrone dalla banda di valenza alla conduzione. Infatti nella banda del silicio i punti di sella delle bande di valenza e conduzione sono associati a frequenze spaziali differenti (indirect band-gap), per cui un elettrone per cambiar di banda richiederebbe, assieme all'energia di 1.1eV anche la presenza contemporanea di un fonone, che crei una deformazione del reticolo tale da sovrapporre i punti di potenziale massimo e minimo delle bande di valenza e conduzione rispettivamente.

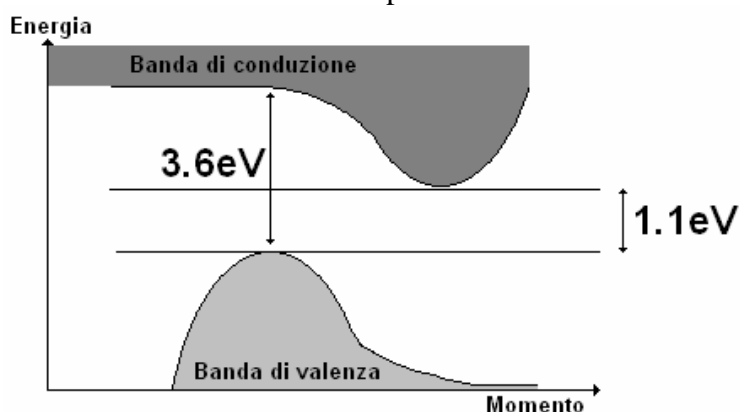


Figura 28 - Grafico qualitativo della struttura a bande del silicio

Poiché ciò difficilmente può avvenire, si assume come energia di band-gap nell'ambito dei rivelatori di radiazione in silicio la differenza di energia fra il punto di massimo potenziale della banda di valenza e la corrispondente energia della banda di conduzione, cioè 3.6eV.

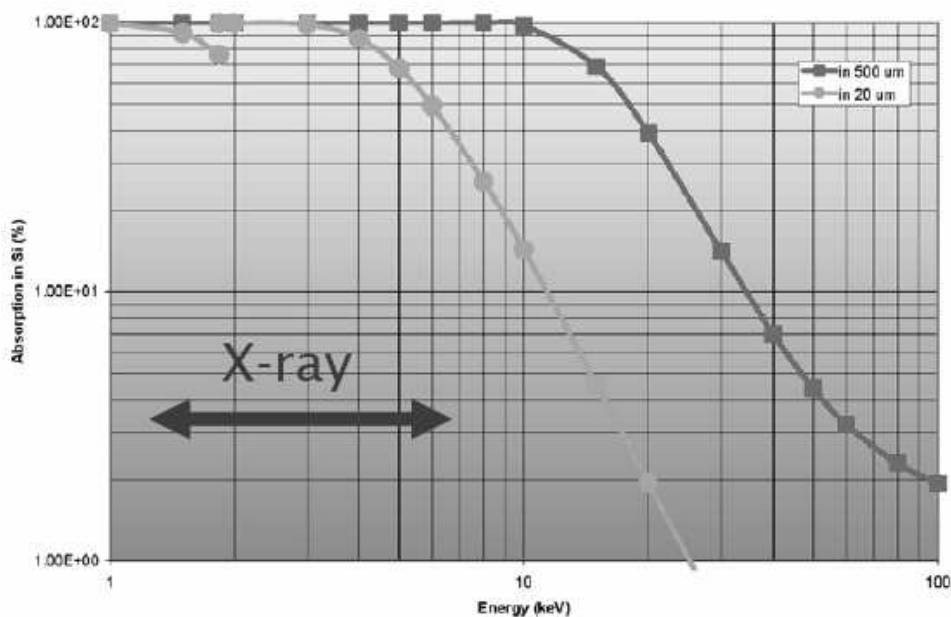


Figura 29 - Assorbimento di radiazione X in funzione dell'energia del fotone incidente per 20um e 500um di Silicio [14]

Questo riduce l'ambito di utilizzo del silicio al di fuori della luce visibile. Radiazioni nello spettro dei raggi X hard (a partire da qualche decina di keV) hanno un'interazione di tipo puntuale col reticolo: si ha una forte generazione di carica, pari al rapporto fra energia del fotone ed energia per la creazione di una coppia e-/h+ (3.6eV), e solo in alcuni punti localizzati del silicio. Tale interazione è però casuale, in quanto la probabilità che una radiazione interagisca col silicio è proporzionale allo spessore del silicio stesso e dipendente dalla lunghezza d'onda (Figura 29).

In particolare, questa limitazione rende problematico l'impiego di sensori CMOS nelle radiografie, soprattutto quando si voglia impiegare il sensore nudo. Per ottenere un'immagine con buon contrasto, serve infatti un sensore con alta quantum efficiency, in cui si abbia cioè un elevato rapporto fra l'escursione del segnale in uscita e l'intensità della radiazione incidente. Questo parametro è a maggior ragione importante perché un'elevata quantum efficiency porta a richiedere dosi di raggi X meno elevate per una radiografia, con evidenti vantaggi per il paziente che viene sottoposto ad una quantità inferiore di radiazioni. L'approccio classico medico prevede pertanto l'introduzione degli strati scintillatori. Questi sono pellicole chimiche (generalmente GADOX o CsI(Tl)), con spessore variabile da pochi micron al millimetro che vengono eccitate quando sono colpite da una radiazione X e che quindi riemettono l'energia ricevuta sotto forma di luce visibile. È così possibile convertire la lunghezza d'onda della radiazione incidente in modo che questa sia più efficientemente rilevabile dal sensore.

L'introduzione dello strato scintillatore comporta tuttavia l'insorgere di una nuova serie di problemi. Per cominciare, la radiazione luminosa visibile viene rilasciata in maniera isotropa, per cui può essere emessa con un angolo diverso rispetto quello della radiazione che l'ha originata. A causa di ciò, il segnale relativo può essere raccolto da un altro pixel, causando un "blurring dell'immagine", che ne danneggia la risoluzione spaziale. Per ovviare a questo problema si è puntato molto sulla efficace costruzione di strati scintillatori pixellati, fatti cioè di colonne parallele che incanalano tutta la luce generata in esse verso lo stesso pixel. In alternativa si può optare per uno strato scintillatore più sottile, in modo che la radiazione emessa venga generata più vicino alla superficie del sensore e perciò più difficilmente colpisca un pixel diverso. Lo svantaggio di questo ultimo approccio sta nel fatto che spesso la radiazione X ha uno spessore di penetrazione maggiore dello spessore dello strato scintillatore, per cui può accadere che colpisca il substrato del sensore rilasciando carica. Questo fenomeno, oltre ad essere pericoloso in quanto fonte di danno da radiazione, riduce anche la linearità del sensore, poiché la carica da X diretto si somma a quella da visibile generando picchi di segnale indesiderati. Un'alternativa

consiste del porre un sottile strato scintillatore di fronte alla radiazione e quindi condurre la luce generata al sensore, fisicamente distante, mediante fibra ottica, anche se questo aumenta la complessità del sistema di acquisizione. Inoltre bisogna tenere in considerazione il tempo di risposta dello scintillatore, in quanto l'emissione del fotone visibile avviene dopo un certo tempo dall'arrivo del fotone X. Tempi di risposta brevi permettono sensori con una maggiore frequenza di lettura.

È stato anche proposto, mediante l'impiego di un processo a tripla well di rendere il sensore insensibile alla carica generata nel bulk (dovuta ai raggi X) isolando n-well dal p-epi con una p-well profonda [15]. In questo modo il fotodiodo può raccogliere solo la carica generata a livello superficiale da radiazione a basso spessore di penetrazione (cioè la luce blu), aumentando il SNR del sistema scintillatore-sensore. Un secondo vantaggio di questo approccio è la possibilità di inserire nel pixel anche transistor pMOS. Anche se la loro presenza tende a ridurre il Fill Factor, essi non sono in grado di sottrarre carica al fotodiodo in quanto la loro n-well si troverebbe ad essere isolata dalla zona svuotata del diodo e a raccogliere unicamente l'eventuale carica generata nel bulk dagli X.

Ad ogni modo, l'impiego di sensori CMOS nella rivelazione di raggi X è in costante aumento, grazie alla possibilità di realizzare sensori poco costosi, di facile impiego e di dimensioni ridotte (utili per esempio nella radiografia dentale). Inoltre nell'ambito medicale risulta molto utile la possibilità di ottenere i dati in tempo reale senza attendere lo sviluppo come nell'uso di lastre chimiche.

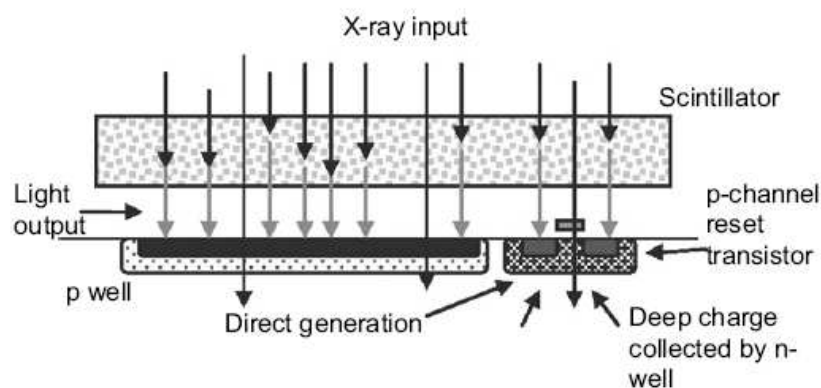


Figura 30 - Schema di pixel ottimizzato per scintillatori: solo la carica generata in superficie, nella p-well è raccolta, mentre la carica generata nel bulk viene raccolta dalla n-well dei pMOS [15]

Acceleratori di particelle

Per approfondire la conoscenza della materia nei suoi componenti basilari, scopo principale degli esperimenti di fisica delle alte energie (High Energy Physics, HEP) risultano essenziali gli acceleratori di particelle. Scopo di questi dispositivi è accelerare particelle a velocità vicine a quella della luce, quindi farle collidere contro o un bersaglio fisso o un altro fascio di particelle che si muove in direzione opposta. L'alta energia sprigionata dall'impatto permette il verificarsi di processi che raramente avvengono in natura, con la generazione di nuove particelle. Lo studio e il riconoscimento delle particelle generate consente di studiare la natura dei processi che regolano i componenti base della materia.

La misura della velocità raggiungibile dal fascio è data dall'energia del fascio (in electron-Volt) invece che dalla velocità raggiunta dalle particelle, in quanto a velocità così vicine a quella della luce variazioni di energia anche rilevanti si traducono in incrementi minimi di velocità

Le tipologie principali di acceleratore sono:

- Lineare
- Circolare

Nel lineare, le particelle vengono accelerate in linea retta, passando attraverso una serie di dischi perpendicolari al cammino che cambiano il potenziale al passaggio del fascio, in modo da attirare le particelle quando queste non l'hanno ancora passato e da respingerle dopo il loro passaggio. Il punto d'impatto è solitamente costituito da uno schermo fisso. In alternativa, e' possibile prima accelerare un fascio, tenendolo confinato in un percorso circolare al termine dell'acceleratore con dei magneti, quindi farlo collidere con un secondo fascio accelerato nello stesso condotto lineare per ottenere una maggiore energia d'interazione .

Negli acceleratori di tipo circolare, che sono in grado di generare fasci a maggiore energia, le particelle vengono invece confinate entro un anello, che percorrono più volte a velocità crescenti. Il cammino delle particelle viene tenuto circolare da campi magnetici, che crescono d'intensità ad ogni giro che viene effettuato dal fascio, in modo che il diametro rimanga costante. È inoltre possibile ottenere impatti fra fasci accelerati in direzioni opposte, di fatto raddoppiando l'energia totale del sistema, inserendo un secondo anello sotto al primo, in cui le particelle vengono accelerate in direzione opposta. Raggiunta una sufficiente energia, i due fasci vengono poi deflessi uno contro l'altro entro una camera apposta, in cui sono posti tutti gli strumenti per la rivelazione di particelle di differente natura.

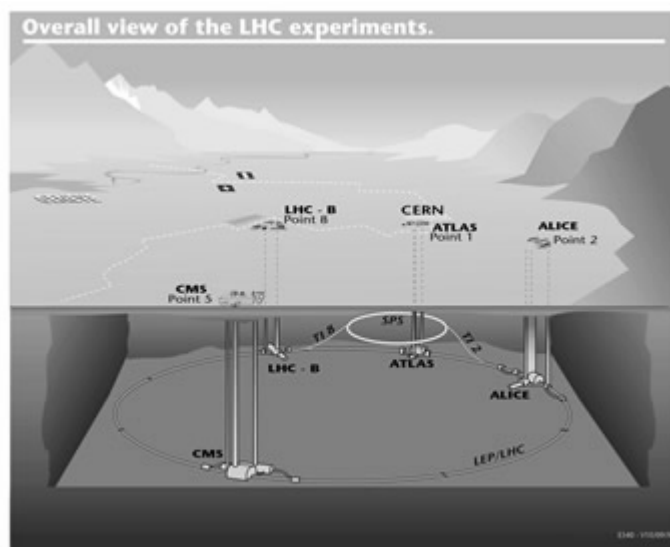


Figura 31 - L'acceleratore LHC del CERN di Ginevra

Grazie a questa struttura, sono ottenibili energie dell'ordine delle centinaia di GeV, fino ai 7TeV per i fasci di protoni (14TeV di energia dell'impatto) nei 26.6km di circonferenza del Large Hadron Collider (LHC) del CERN. Rispetto agli acceleratori lineari, i circolari permettono un più efficiente sfruttamento della lunghezza di accelerazione, in quanto il sistema di accelerazione può essere percorso più volte invece di una sola. Tuttavia non permettono di ottenere un flusso continuo di particelle, in quanto non essendo costante il campo magnetico che mantiene la curvatura del fascio, l'acceleratore non può contenere fasci a differenti velocità, per cui le particelle si muovono raggruppate in bunch che si susseguono ad una frequenza fissa. Un secondo fattore che limita le prestazioni degli acceleratori circolari è l'emissione di radiazione di sincrotrone, che avviene ogniqualvolta una particella carica si muove in una traiettoria circolare. Ad alte energie, gran parte della potenza immessa nel sistema viene pertanto persa in radiazione di sincrotrone, per cui solo un residuo può essere usato per l'accelerazione.

Di conseguenza si pensa che in futuro gli acceleratori circolari saranno sostituiti per le altissime energie da acceleratori lineari di nuova concezione, di lunghezza dell'ordine delle decine di km (attualmente il più lungo, con 3.2km di cammino, è lo Stanford Linear Accelerator Center (SLAC) in California).

La camera di collisione, in cui avvengono tutte le interazioni dell'acceleratore, può esser vista, nei casi di esperimenti a più alte energie (impatto fra due fasci) come una serie di cilindri concentrici di diametro crescente (strati o layers). Ognuno di questi layer svolge una funzione differente, rilevando un diverso tipo di particella o registrando una diversa informazione.

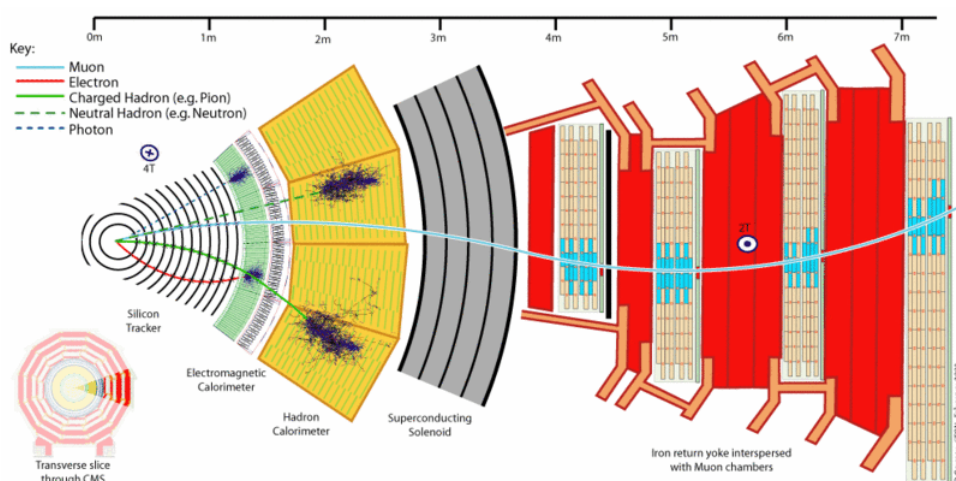


Figura 32 - Disposizione dei rivelatori nell'esperimento Compact Muon Solenoid dell'LHC

In un esperimento di fisica delle alte energie, come per esempio il CMS presso l'LHC, sono presenti una serie di strati:

- I vertex detectors, formati da più strati di sensori (pixellati o a striscia) in silicio, con il compito di registrare i punti di passaggio delle particelle. In questo modo è possibile ricostruire la traiettoria

delle particelle, da cui, conoscendo il campo magnetico presente nel mezzo, è possibile risalire a carica e energia.

- I calorimetri: tarati per misurare l'energia di differenti particelle
 - o Calorimetro elettromagnetico: misura l'energia di fotoni ed elettroni
 - o Calorimetro adronico: misura l'energia degli adroni (protoni, neutroni, pioni)
- Il solenoide superconduttivo, impiegato per generare il forte campo magnetico interno
- Lo strato per la rivelazione dei muoni
 - o Rivelazione della traiettoria
 - o Generazione dell'evento di trigger

Per ognuno di questi elementi sono possibili scelte tecnologiche differenti. Questi vengono arrangiati in anelli concentrici del diametro di 10-20m, che vengono quindi posti affiancati per tutta la lunghezza della camera di collisione.

Vertex detector

Negli esperimenti di fisica delle alte energie mediante acceleratori di particelle, uno dei componenti basilari, come si è visto, risulta essere il vertex detector. Si tratta di un componente presente nei primi strati dei sensori, il cui scopo è segnalare il passaggio di una particella e le sue coordinate. Ponendo un certo numero di questi dispositivi in successione è inoltre possibile ricostruire l'intera traiettoria di una particella, in modo da comprenderne carica, massa ed energia, il che ne permette l'identificazione. Essendo il componente maggiormente vicino alla camera di collisione, dove maggiore è il flusso di particelle per unità d'area, risulta vitale una forte resistenza al danno da radiazione, sia da esposizione che di tipo single-event. Altri requisiti essenziali sono

- Costi contenuti
- Basso consumo

- Alta frequenza di lettura
- Risoluzione spaziale almeno nell'ordine della decina μm

Infatti questi sensori devono monitorare un'area complessiva che per certi esperimenti può superare i 10m^2 , il che richiede anche bassi consumi e peso minimo, altrimenti l'intera struttura che li contiene potrebbe risultare meccanicamente critica, dovendo sopportare un elevato peso ed una complesso sistema di raffreddamento. Inoltre ricoprendo con sensori di risoluzione micrometrica e con data-rate dell'ordine delle migliaia di frames al secondo si genera una grande mole di dati, in quanto il sistema si trova facilmente a dover vagliare in tempo reale dei flussi di decine di GB/s.

Per realizzare simili dispositivi, esistono innumerevoli opzioni tecnologiche. Negli ultimi anni, in particolare, hanno raccolto crescente interesse i rivelatori a stato solido, che hanno ormai sostituito i precedenti rivelatori gassosi nell'identificazione delle traiettorie delle particelle. Il principio base di funzionamento di tutti questi dispositivi è identico a quello già descritto nel CMOS, con una giunzione p-n posta in inversa in cui quando una particella vi incide si ha un rilascio di carica. Questi dispositivi risultano vantaggiosi per l'elevata risoluzione che riescono a raggiungere, oltre alla maggiore velocità di lettura che permette una più precisa identificazione dell'istante di passaggio di una particella. Anche il valore letto di carica raccolta risulta molto più preciso.

All'interno del silicio, una particella MIP (minimum ionising particle) tende a generare una carica molto debole e costante lungo il cammino, circa 80 coppie e^-/h^+ per μm , di conseguenza la gran parte della carica risulta generata nel bulk e non in superficie, come accade nel visibile. Risulta perciò importante essere in grado di avere un elevato rapporto segnale/rumore per poter riconoscere un passaggio riducendo al minimo sia i falsi negativi (particella non rilevata) che i falsi positivi (rilevato passaggio di una particella inesistente).

Microstrip

Il primo e più impiegato sensore per gli esperimenti HEP è il sensore a microstriscia. Questo è un rivelatore 1-D in cui su un substrato n debolmente drogato (tipicamente $\sim 10^{12} \text{cm}^{-3}$) e di spessore di 150-300 μm vengono realizzate delle sottili strisce di tipo p+. Il substrato è tipicamente polarizzato tramite una tensione positiva ($\sim 100\text{V}$) posta su un contatto n+ sul fondo del sensore, che svuota completamente il volume del dispositivo, permettendo una raccolta di carica su tutto il volume. La raccolta di carica avviene tramite le strisce p+, che raccolgono le lacune generate dalla radiazione incidente, permettendo di stabilire la coordinata di impatto perpendicolare alla direzione delle strisce. Tipicamente la risoluzione effettiva, detta L la distanza fra le strisce, è $L/\sqrt{12}$, (con L tipicamente 50 μm) in quanto effettuando un confronto con le strip adiacenti è possibile avere un'informazione più accurata che non la semplice posizione della strip che ha raccolto la maggior carica. Si tende a raccogliere lacune invece di elettroni in quanto con un substrato p e strisce n, il danno da radiazione sull'ossido causerebbe la creazione di canali parassiti fra le strip, cortocircuitandole e impedendone così il corretto funzionamento.

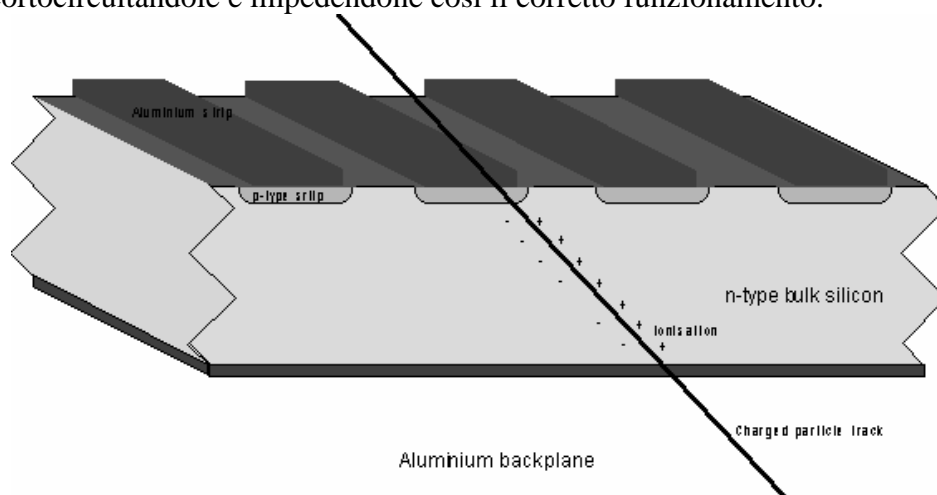


Figura 33 – Microstrip (da hepwww.rl.ac.uk)

Tipicamente questi sensori possono raccogliere una grande quantità di carica (>20000 portatori) grazie all'intero volume sensibile e alla raccolta per drift, però sono affetti da un rumore non trascurabile, per cui il loro SNR è tipicamente di 10-15.

È possibile anche ottenere un'informazione 2-D sostituendo il contatto n+ sul fondo del sensore con strisce n+ poste perpendicolari rispetto le strisce p di fronte. In questo modo mentre un lato raccoglie le lacune e dà informazione su una coordinata, l'altro lato raccoglie elettroni e dà informazioni sulla coordinata perpendicolare. In questo caso va comunque considerato che l'informazione 2D potrebbe essere affetta da errore di parallasse, se la particella è molto inclinata rispetto la perpendicolare del sensore e che le strisce n non sono rad-hard, per cui richiedono un certo processing (introduzione di una striscia p+ fra 2 strisce n, drogaggio p dell'interfaccia con l'ossido) per evitare che il danno da radiazione le cortocircuiti.

Silicon drift detectors

Una delle tecnologie maggiormente impiegate per i vertex detector a stato solido è il SDD (Silicon Drift Detector) [22]. Si tratta di sensori a semiconduttore composti da un substrato drogato p all'interno del quale sono scavate delle strisce drogate n. Ponendo lo strato p a tensioni negative e lo strato n a massa, sulla sezione trasversale del sensore si forma un avvallamento di potenziale, che tende ad attirare gli elettroni, che risultano così incanalati verso l'anodo. Grazie all'alta tensione di polarizzazione l'intera zona n è svuotata, per cui gli elettroni generati dal passaggio di particelle vengono raccolti al centro del bulk. Oltre alla componente trasversale, c'è una seconda componente del campo elettrico longitudinale, che fa sì che gli elettroni raccolti al centro tendano a muoversi perpendicolari alla strip verso un elettrodo di raccolta. In questo modo inoltre si evita l'accumulo di carica nel bulk, che potrebbe deformare il campo elettrico al suo interno. Questo sensore permette di ricostruire con

buona approssimazione (fino a 2 μ m in teoria) la distanza rispetto l'elettrodo dell'impatto, misurando l'intervallo di tempo che intercorre fra l'istante di passaggio della particella e l'istante in cui la carica viene raccolta dall'elettrodo. L'istante di passaggio della particella deve perciò esser fornito da un calorimetro esterno al sensore, e con buona precisione temporale per avere una misura affidabile. È inoltre possibile risalire alla seconda coordinata spaziale raccogliendo la carica al termine delle strip con più elettrodi drogati n [Figura 36]. In questo caso si osserva la carica raccolta dagli elettrodi adiacenti a quello che ha mostrato la maggior raccolta e se ne calcola con opportune funzioni di peso il centro di gravità.

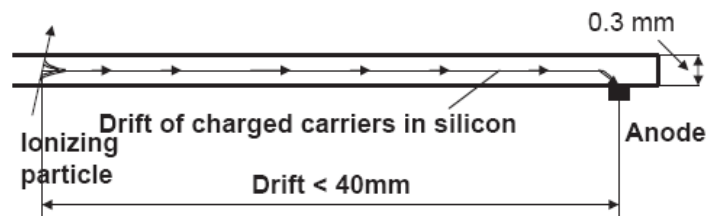


Figura 34 - Schema di funzionamento generale di un SDD [22]

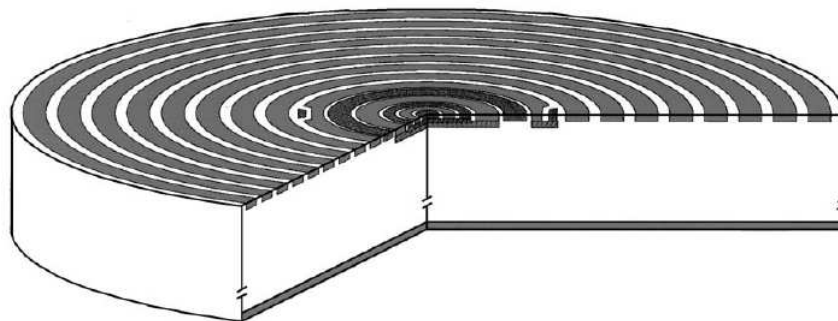


Figura 35 - SDD per radiografia: gli anelli concentrici di drogaggio p creano un campo elettrico che porta verso il centro tutta la carica generata. Usato per la spettrografia X, dove non interessano le coordinate del punto d'impatto [22]

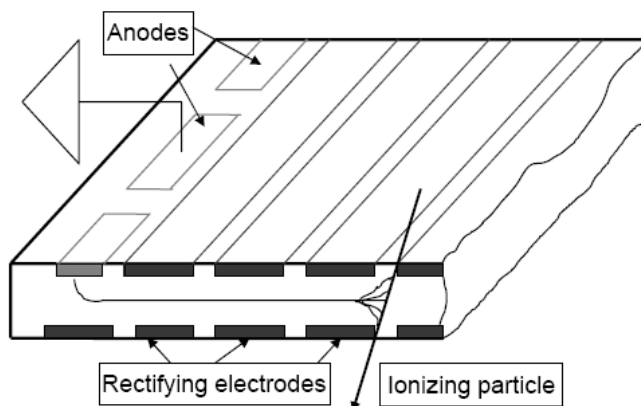


Figura 36 - Schema di un SDD multianodo, in grado di fornire le coordinate x-y del punto d'impatto [22]

Particolare cura va messa nella realizzazione di questi dispositivi, in quanto sono soggetti ad inversione di drogaggio (una elevata dose di particelle ionizzanti tende a trasformare il substrato p in substrato n), per cui ad un certo punto della loro vita funzionale risulta necessario per impiegarli invertire le tensioni di polarizzazione, con perdita di efficienza nella raccolta di carica (più difficile generare la zona svuotata) e di sensibilità.

CCD

È stata anche esplorata la possibilità di impiegare per gli esperimenti HEP la tecnologia CCD (Charge Coupled Devices). Questi sensori sono strutturati come un array di gates posti sopra ad un canale drogato n e ad un bulk di tipo p. Durante la fase d'integrazione, si impone un potenziale positivo su un gate, mentre i due gate adiacenti sono tenuti ad un potenziale negativo. Questo fa sì che sotto al gate a tensione positiva si crei una tasca di potenziale in cui gli elettroni fotogenerati sono raccolti. Per leggere l'informazione di carica, ad ogni ciclo di clock il gate accanto a quello a tensione positiva viene portato ad una tensione positiva anch'esso, quindi il

precedente viene spento: la carica immagazzinata nella tasca viene di conseguenza spostata verso la nuova tasca creata dal gate appena acceso. In questo modo la carica raccolta nei vari pixel si sposta di pixel in pixel verso il nodo di raccolta, dove viene misurata.

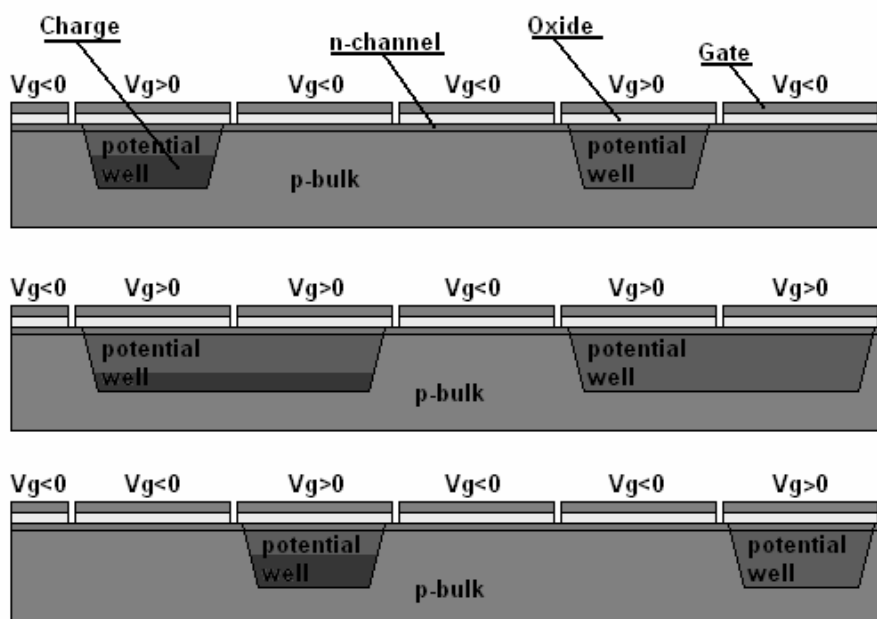


Figura 37 – Schema di funzionamento e spostamento della carica per un CCD

Questo approccio ha vari vantaggi dal punto di vista del rumore, che risulta molto più basso rispetto l'equivalente CMOS, soprattutto relativamente al rumore kTC (praticamente assente) e FPN, e della efficienza di raccolta della carica fotogenerata. Gli svantaggi sono una minore affidabilità, in quanto la rottura di un pixel causa l'impossibilità di leggere tutti i pixel a monte di esso, e soprattutto la radiation-softness, in quanto i gates che compongono il CCD hanno ossidi relativamente spessi e sono di conseguenza più suscettibili al danno da esposizione. Inoltre la creazione di trappole nel canale sepolto (dove la carica viene trasferita) degrada notevolmente la raccolta di elettroni. Infine, la tecnologia CCD non supporta la creazione di dispositivi elettronici al suo interno, per cui tutta

l'elettronica di controllo e acquisizione deve essere realizzata su chip esterni, rendendo estremamente complesso il sistema di acquisizione.

DEPFET

Il DEPFET è un ulteriore componente studiato per gli esperimenti di fisica delle alte energie. Anche esso richiede una tecnologia ad-hoc per essere creato, al momento non facilmente disponibile.

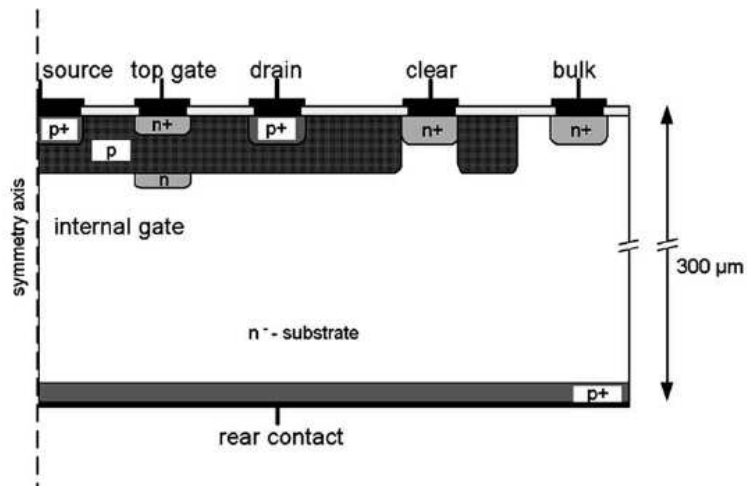


Figura 38 - Sezione di un DEPFET [24]

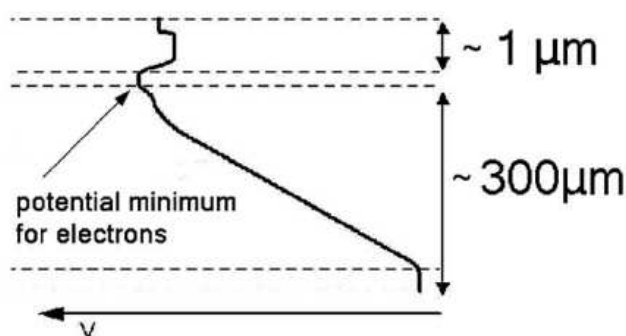


Figura 39 - Profilo del potenziale nel bulk di un DEPFET [24]

Le idee alla base di questo dispositivo sono le seguenti:

- Totale svuotamento del substrato di tipo n applicando la tensione inversa a diffusioni di tipo p poste sia sotto che sopra il sensore: in questo modo viene a crearsi una zona di minimo di potenziale all'interno dell'area di raccolta nella posizione voluta, permettendo di confinare la carica in quel punto e di avere una raccolta di carica molto veloce
- Lettura non distruttiva: la carica viene raccolta all'interno di una diffusione n+ sepolta sotto al canale fra drain e source, che si comporta da gate interno, modulando la I_{DS} e la V_G del DEPFET, che danno informazioni circa la quantità di carica raccolta e che possono essere lette in ogni momento.
- Controllo del segnale, grazie alla possibilità di modulare il canale fornendo un'appropriata tensione sul gate esterno del DEPFET

Questo dispositivo richiede inoltre che periodicamente venga dato un impulso di clear, attivando un secondo MOS che svuota la diffusione sepolta dalla carica. I vantaggi di questo approccio, oltre alle caratteristiche già citate, sono un basso rumore ($2.2e^-$ per un fotone a 6keV [24]) e la possibilità di raccogliere carica a transistor spento (con consumi di potenza inferiori). Gli svantaggi sono la richiesta di tensioni comunque alte e non

standard (6-7V per il substrato, 14V per l'impulso di clear), la necessità di un chip esterno per la lettura e la lettura di tipo sequenziale, che limita la velocità di acquisizione dei dati.

Problematiche per un vertex detector CMOS

I problemi che un simile dispositivo deve affrontare sono però molteplici. Negli acceleratori di nuova concezione, infatti, l'aumento dell'energia dei fasci che interagiscono comporta anche un aumento della dose totale a cui vengono sottoposti i sensori, per cui è necessario un approccio che sia in grado di resistere ai danni da radiazione derivanti da esposizione. Inoltre in questi esperimenti, l'area attraverso la quale si deve monitorare il passaggio risulta molto estesa, nell'ordine delle decine di m^2 : risulta pertanto necessario gestire una gran mole di dati, soprattutto considerando che il passaggio della particella deve essere individuato con la massima precisione possibile (sotto i 10 μm). Di conseguenza possono risultare utili sensori con vari gradi di signal processing interno, che possono spaziare dal comprimere i dati in uscita, fino al riconoscere autonomamente la presenza o meno di un punto colpito e di fornire al sistema unicamente i valori dei pixel colpiti e dei vicini.

Inoltre la grande area richiede che il costo per unità di superficie del sensore sia contenuto, così come il suo consumo di potenza. Un secondo fattore di interesse è la possibilità di assottigliare il sensore senza perdere sensibilità e senza minarne la resistenza meccanica, che permetterebbe una struttura molto più leggera.

Alla luce di questi fattori, si è presa in considerazione la possibilità di impiegare la tecnologia CMOS per la realizzazione di questi dispositivi. Questa infatti è resistente alle radiazioni, in quanto gli ossidi di gate sono più sottili rispetto alle tecnologie rivali (come il CCD) e di conseguenza tendono a ionizzarsi meno all'aumentare della dose di radiazione. Inoltre, trattandosi di una tecnologia commerciale, l'economia di scala implica che sia disponibile a prezzi economici e che i dispositivi siano in costante

sviluppo sia dal punto di vista dell'occupazione di area che delle prestazioni. Questo fa sì che l'impiego della tecnologia CMOS migliori la risoluzione spaziale e le prestazioni in velocità del sensore, consentendo una migliore aderenza alle specifiche. I consumi risultano fortemente ridotti, in quanto ogni pixel è acceso unicamente durante la fase di lettura, mentre la possibilità di integrare elettronica digitale ed analogica sullo stesso chip permette di avere del signal processing on-chip senza l'aggiunta di componenti esterne.

Inoltre i sensori di tipo CMOS hanno il vantaggio di poter essere assottigliati senza perdere in funzionalità. Questo permette di contenere il peso del sensore, che negli esperimenti di fisica delle alte energie può rivelarsi un fattore assolutamente non trascurabile. Gli svantaggi di tale approccio provengono, come precedentemente descritto, dalla scarsa efficienza nella raccolta di carica e dalla presenza di elevato rumore kTC.

In generale, il funzionamento da vertex detector è abbastanza differente rispetto l'imaging. Se in questo ultimo caso le particelle da rilevare (fotoni) sono talmente numerose da essere considerate una grandezza continua e di cui si vuole trovare la quantità in funzione della posizione spaziale, nel primo caso c'è un numero finito di particelle da rilevare (spesso una sola), di ognuna delle quali si desidera la coordinata di attraversamento del sensore. Di conseguenza, l'output di un vertex detector sarà un'immagine totalmente scura, con pochi punti illuminati in corrispondenza dei passaggi delle particelle.

Tuttavia l'identificazione di un punto di passaggio non è un problema banale. Infatti mentre nell'imaging visibile la carica viene generata in superficie, all'interfaccia fra ossido e silicio (a causa dello scarso spessore di penetrazione), la carica generata da una MIP si localizza nel bulk. Ciò è dovuto al fatto che le MIP non vengono fermate dal silicio, ma lo attraversano, generando una circa 80 coppie elettrone-lacuna per micron attraversato. Se da un lato questo fenomeno consente di avere un fill-factor vicino al 100%, in quanto anche il bulk sottostante la superficie occupata dall'elettronica di pixel può contribuire alla generazione di carica, dall'altro porta con sé anche un grave svantaggio. Essendo infatti le coppie elettrone-

lacuna generate nel bulk, e muovendosi perlopiù per diffusione termica, la loro traiettoria non viene indirizzata se non in minima parte verso la superficie: di conseguenza tendono a creare una “nube” di carica, che influenza anche i pixel confinanti con quello colpito. Quello che viene osservato in conseguenza al passaggio di una particella è pertanto un “cluster” di pixel colpiti, che riflette la diffusione e l’espansione della carica nel bulk. Questo fenomeno porta alla diminuzione della risoluzione efficace del sensore: più le dimensioni del pixel sono ridotte, più pixel risulteranno interessati al passaggio di una particella. Inoltre, dividendosi la carica su più pixel, ogni pixel singolarmente avrà una minore escursione, rendendo più difficile la scelta di una soglia che indichi infallibilmente se il pixel è stato colpito o meno. D’altronde, il fenomeno dello spreading di carica può anche incrementare, con apposito post-processing, la risoluzione del sensore oltre la risoluzione del pixel. Considerando e pesando opportunamente il valore di più pixel adiacenti, infatti, è possibile individuare il centro di gravità della distribuzione di carica con un errore inferiore alla dimensione del pixel, permettendo una risoluzione dell’ordine di 2 μ m per un sensore con strato epitassiale e 3-4 μ m per un sensore che ne sia privo [25].

Per aumentare l’affidabilità rispetto al danno da radiazione una soluzione può essere l’impiego di un maggior numero di diodi. Infatti, inserendo nel pixel 2 o 4 fotodiodi in parallelo invece di uno solo, posti ai lati del pixel, è possibile diminuire lo spreading di carica fra pixel adiacenti, in quanto la carica generata all’interno di un pixel più difficilmente viene raccolta in un pixel adiacente [8]. Il prezzo da pagare in questo caso è una maggiore capacità del nodo di raccolta della carica, con un peggioramento della conversione carica-tensione. In caso di danno da radiazione, peggiora la raccolta di carica da parte dei singoli diodi, in quanto i danni al reticolo diminuiscono il tempo di vita medio della carica all’interno del bulk, ma lavorando questi in parallelo, e diminuendo il cammino medio che una carica deve percorrere prima di raggiungere una n-well, riescono a raccogliere una quantità di carica superiore rispetto al caso con singolo diodo anche dopo un pesante irraggiamento.

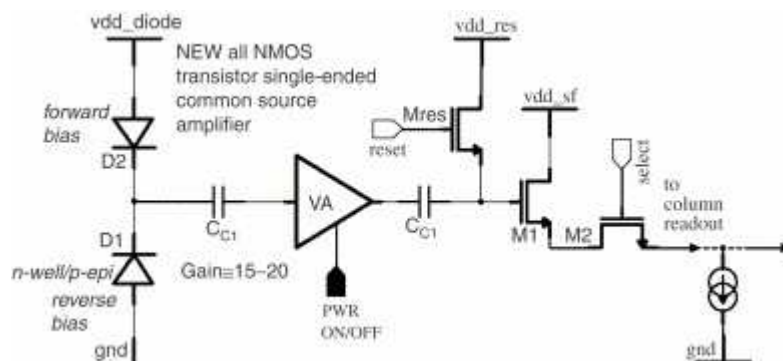


Figura 40 - Schematico del pixel di MIMOSA-8 con self-bias e accoppiamento AC all'amplificatore on-pixel [26]

Un approccio alternativo al reset normale nella rivelazione di particelle è il self-bias [26] in cui il fotodiode non viene riportato in inversa periodicamente, ma è tenuto costantemente polarizzato da un secondo diodo p-n in diretta, più piccolo, collegato al suo anodo. Questo diodo conduce una debole corrente, che in condizioni stazionarie eguaglia la corrente di perdita del fotodiode, evitando la costante diminuzione di tensione riconducibile alle correnti di buio. In questo modo è possibile amplificare il valore del diodo, cosa non possibile con l'approccio classico, in cui la diminuzione di tensione dovuta al leakage modificerebbe il punto di lavoro dell'amplificatore. Di conseguenza il segnale derivante da un impatto di particella verrebbe amplificato con guadagno differente a seconda della distanza temporale dell'istante di impatto stesso rispetto il reset. Inoltre, per avere maggiore amplificazione è possibile l'accoppiamento AC del diodo all'elettronica di amplificazione, in modo da settare il punto di lavoro dell'amplificatore indipendentemente dal valore di riposo del fotodiode.

Il reset in caso di impatto viene effettuato dal diodo di polarizzazione stesso, che riporterebbe in equilibrio il sistema in un periodo di tempo di circa una decina di us. In questo modo, equipaggiando il pixel con un'adeguata circuiteria di riconoscimento dell'impatto, in grado di salvare il valore amplificato letto prima che il diodo torni in equilibrio, è possibile creare un sistema in grado di riconoscere facilmente un evento di impatto da particella.

Ad ogni lettura è inoltre eliminato il FPN. Infatti, dopo aver letto il valore di luce, viene effettuato un reset tramite il transistor MRes che, avvenendo sul condensatore di elevato valore C_{c1} invece che sul fotodiodo, permette di ottenere il valore di buio con rumore kTC trascurabile. I due valori vengono quindi sottratti eliminando il FPN.

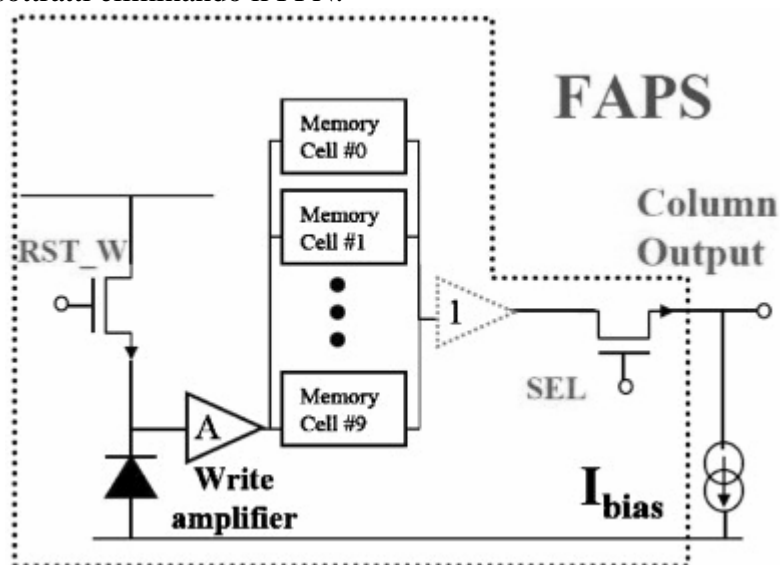


Figura 41 - Struttura del pixel FAPS

Si può altrimenti sfruttare il fatto che in numerosi esperimenti di fisica delle alte energie, la generazione di particelle non è uniforme nel tempo, ma è concentrata in bunch ad elevata frequenza intervallati da periodi relativamente lunghi privi di eventi. Questo permette di aumentare la frequenza di acquisizione del sensore. In particolare, durante un bunch di particelle, il rivelatore deve acquisire il valore di ogni pixel a framerate elevate (nell'ordine delle decine di MHz) per un periodo di tempo di centinaia di μs , quindi segue un intervallo di inattività dell'ordine delle decine di ms. Proprio per questo è possibile inserire una serie di celle di memoria all'interno del pixel (FAPS, flexible acquisition pixel sensor), formate da un condensatore MOS ed uno switch. Queste celle sono ovviamente accessibili a frequenza più elevata rispetto al bus di colonna dell'APS

standard, in quanto hanno una capacità nettamente inferiore. Durante il bunch di particelle, i valori letti vengono salvati successivamente nelle varie celle di memoria a intervalli di $\sim 100\text{ns}$, dalle quali sono successivamente letti nei periodi di inattività a frequenze molto più basse ($\sim 1\text{MHz}$). In questo modo è possibile disaccoppiare la frequenza di acquisizione del pixel dalla frequenza di lettura dell'elettronica esterna, in modo da incrementare la risoluzione temporale degli esperimenti.

CAPITOLO 2: IL PROGETTO RAPS

Introduzione

Il progetto RAPS nasce come collaborazione fra l'università di Parma, l'università di Perugia e l'INFN di Perugia. Scopo del progetto è lo studio e la realizzazione di una famiglia di sensori in tecnologia CMOS standard orientati alla rivelazione di radiazioni e particelle minimamente ionizzanti, per scopi sia scientifici che medicali. La linea guida principale è lo sviluppo di sensori con le seguenti caratteristiche, ottenibili solo con l'impiego della tecnologia CMOS:

- Basso costo
- Basso consumo
- Altamente configurabili grazie alla possibilità di implementare elettronica di controllo-processing del segnale direttamente sul chip sensibile
- In grado di eseguire le operazioni richieste senza bisogno di chip di supporto
- Resistenti al danno da radiazione

Sono stati finora realizzati tre sensori, nominati progressivamente RAPS-01, RAPS-02 e RAPS-03, con lo scopo di testare sia le soluzioni tecnologiche esistenti, che di proporre di nuove.

Scelta della tecnologia

Il RAPS-01, primo chip realizzato, ha avuto lo scopo di valutare le scelte tecnologiche compiute all'inizio del progetto in seguito a numerose simulazioni. In particolare era necessario valutare la correttezza della scelta della tecnologia, cioè la UMC 0.18 Mixed-Mode RFCMOS 1.8/3.3.

Questa tecnologia è stata scelta, sia rispetto le tecnologie alternative, sia rispetto la stessa tecnologia con lunghezza di canale a 250nm in quanto risulta essere in grado di fornire il massimo swing di tensione in risposta all'impatto di una particella. Dalle simulazioni svolte tramite Dessis, la possibilità di miniaturizzare i componenti e di conseguenza le capacità parassite ha mostrato una stima di escursione di segnale superiore per la tecnologia a 180nm.

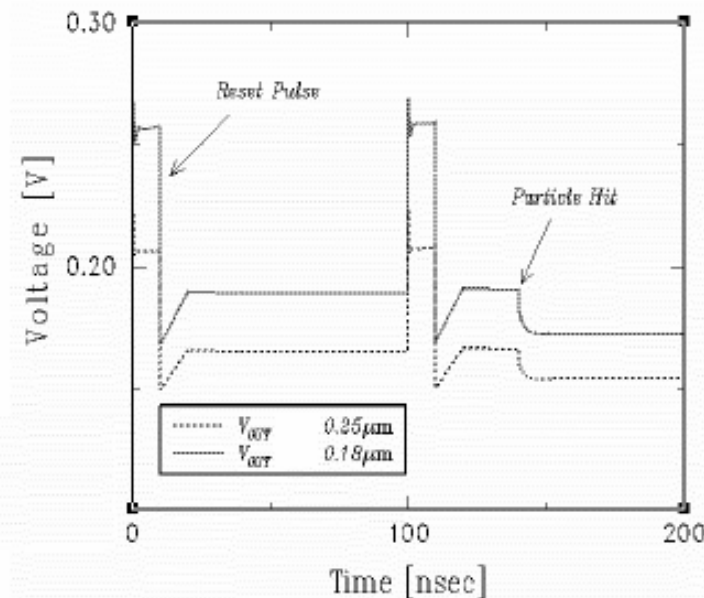


Figura 42 - Tensione di un fotodiode all'impatto di una MIP in funzione del tempo (ns) per tecnologia 0.25µm e 0.18µm

Un'altra opzione considerata è stata la possibilità di impiegare o meno uno strato epitassiale. Ricordando che l'impiego di uno strato epitassiale tende a ridurre lo spreading di carica fra i pixel (maggiore precisione nell'individuazione del punto d'impatto), ma anche a ridurre la raccolta di carica (solo la carica generata nello strato epitassiale è raccolta), sono state effettuate delle simulazioni per stimare la raccolta. In base a queste simulazioni preliminari è stato possibile verificare una maggiore escursione di segnale per una tecnologia priva di strato epitassiale (UMC) rispetto una

tecnologia che lo prevede (IBM), soprattutto in conseguenza di un impatto di centrale sul fotodiode.

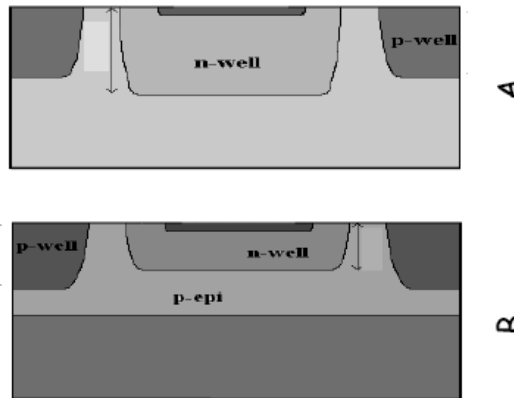


Figura 43 - Tecnologia "bulk" (A) e tecnologia a strato epitassiale (B)

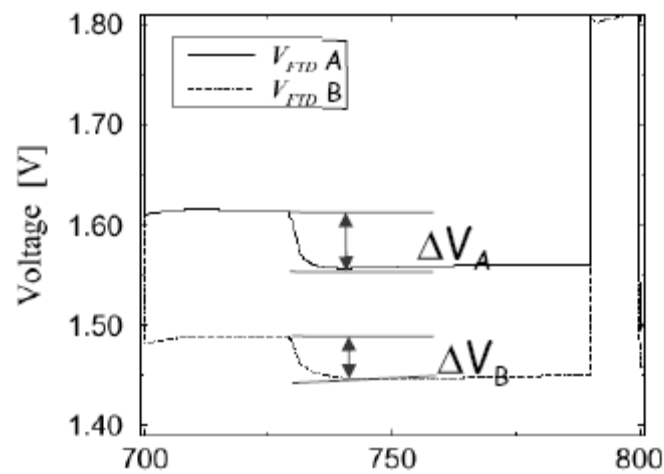


Figura 44 - Tensione di un fotodiode all'impatto di una MIP in funzione del tempo (ns) per tecnologia senza strato epitassiale (A) e con strato epitassiale (B)

Opzioni tecnologiche dei pixel APS

Il primo accorgimento di cui si è voluta valutare l'efficacia è stato il guard-ring. Per limitare lo spreading di carica, infatti, un accorgimento molto impiegato è la creazione di barriere di potenziale che impediscano agli elettroni generati sotto ad un pixel di muoversi lateralmente verso un pixel adiacente. Come detto nei capitoli precedenti, lo spreading di carica non è un effetto del tutto negativo dal punto di vista della risoluzione, tuttavia nel caso di rilevazione di MIP, con generazione di una quantità limitata di carica, si possono ottenere dei falsi negativi. Infatti se la carica risulta equamente divisa su più pixel, può essere difficile identificare un pixel il cui segnale sia molto maggiore rispetto al rumore, e di conseguenza un eventuale impatto può non essere rilevato.

Il guard-ring agisce proprio per confinare all'interno del pixel colpito la maggior quantità possibile di carica. Si tratta di strisce drogate p+ poste lungo il bordo del pixel e tenute a massa. Di conseguenza generano lungo il perimetro del dispositivo una zona a basso potenziale, che tende a formare una barriera che respinge gli elettroni (almeno quelli generati più vicino alla superficie) indirizzandoli verso il centro del pixel.

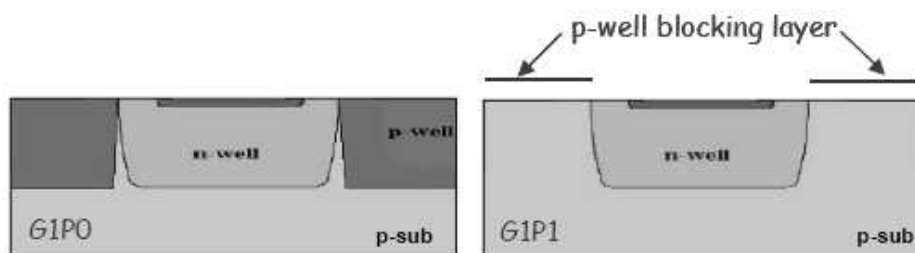


Figura 45 - Esempio di fotodiode senza (G1P0) e con il p-block (G1P1)

Il secondo accorgimento è invece la presenza o meno di uno strato p-well block. Il processo UMC, essendo un twin-tub, prevede l'uso di n e p-well sulla superficie per la costruzione dei dispositivi. Tuttavia a livello layout viene reso possibile lasciare zone della superficie prive di well, cioè con il drogaggio del bulk. Poiché il bulk della tecnologia UMC è a basso

drogaggio (alta resistività), una zona “nativa” attorno alla n-well del fotodiode permette la creazione di una zona svuotata laterale più larga, con aumento della superficie sensibile alla radiazione. Infatti, se in profondità la presenza o meno di uno strato di bulk in superficie non influisce eccessivamente sui profili di potenziale e sulla raccolta di carica, vicino alla superficie, risulterebbe difficile avere a disposizione un’elevata area di raccolta. Infatti, senza p-block, la n-well del fotodiode si troverebbe a contatto diretto con le p-well dei transistor, fortemente drogate, all’interno delle quali l’area svuotata sarebbe sottilissima, senza contare che il fotodiode si troverebbe ad avere una capacità orizzontale diversa dalla capacità laterale. Di conseguenza, per avere una maggiore area sensibile, sarebbe necessario l’impiego di una n-well molto estesa, che però avrebbe lo svantaggio di avere un’elevata capacità, che deprimerebbe il rapporto di conversione carica-tensione.

Pixel WIPS

Il WIPS è un pixel interamente sviluppato all’interno del progetto RAPS. La motivazione alla base di questo sensore deriva da uno dei maggiori problemi degli edge-detectors. Come spiegato nei capitoli precedenti, per ottenere un’informazione limitata (le coordinate x-y di un punto d’impatto), un normale sensore invia tutti i dati di tutta la matrice, gravando il sistema di acquisizione di una gran mole di informazioni inutili. Proprio per questo si è ritenuto utile un approccio alternativo, con un elemento che mandasse in output non il valore di tutti i pixel della matrice, ma solo le coordinate del pixel colpito. Ad ogni scansione, il WIPS è stato studiato per inviare lo status colpito-non colpito di ogni riga e di ogni colonna invece che di ogni pixel. Di conseguenza, se una matrice normale di dimensioni $N \times M$ fornisce $N \times M$ dati, il WIPS ne fornisce $N+M$, da cui è possibile ricavare le coordinate del pixel colpito. Ovviamente, lo svantaggio di tale approccio è che in caso di più impatti durante lo stesso periodo di integrazione, si

avranno più righe e colonne colpite, per cui la ricostruzione dei punti d'impatto non potrà essere univoca. Di conseguenza il WIPS opera nell'ipotesi che per ogni matrice sia colpito ad ogni ciclo di funzionamento al più un pixel. Questo tipo di pixel incorpora al proprio interno anche un pMOS, in quanto le simulazioni effettuate hanno dimostrato che in questo contesto la carica che la n-well di un pMOS sottrae al fotodiode non è tale da inficiare il funzionamento del pixel.

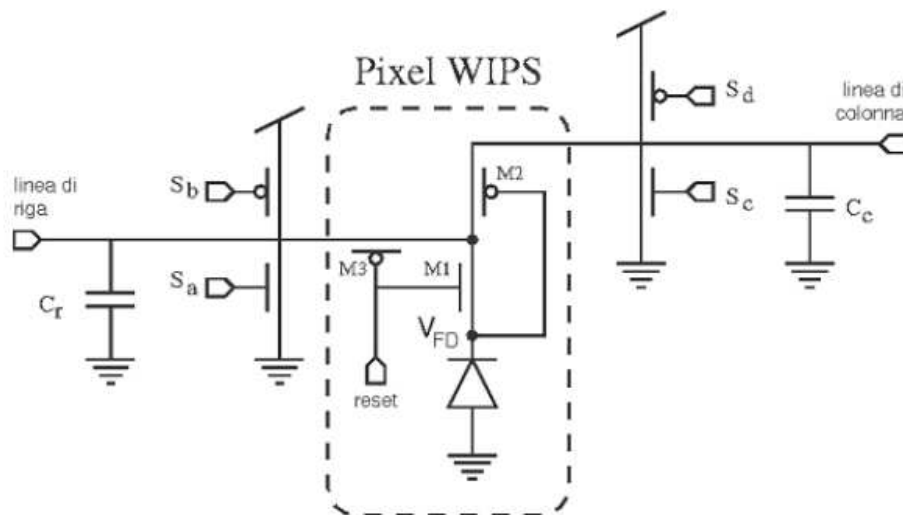


Figura 46 - Schematico del pixel WIPS con relativa elettronica di lettura

Il funzionamento è il seguente :

- Il fotodiode è ripolarizzato (reset alto, M1 ON e M3 OFF) , mentre le capacità C_c e C_r vengono portate rispettivamente a V_{dd} e a 0
- Segue una fase di attesa, in cui il fotodiode può scaricarsi in seguito all'impatto di una particella, accendendo M2
- In fase di lettura il bus di colonna viene precaricato a V_{dd} , quindi viene lasciato flottante mentre i bus di riga sono polarizzati a massa: se un pixel della colonna è stato colpito, porterà in debole inversione il proprio pMOS, che scaricherà il bus corrispondente verso massa, portandolo a 0.

- Lo stesso procedimento viene rifatto invertendo bus di colonna e di riga

Di conseguenza, l'output del sensore sarà costituito dallo stato delle 32 colonne e quindi dallo stato delle 32 righe, invece dei 1024 valori dei singoli pixel, con un evidente guadagno dal punto di vista della mole di dati da gestire e del frame-rate.

RAPS-01

I pixel sviluppati per il RAPS-01, sono i pixel di tipo G1P1, G1P0, G0P1 e G0P0, arrangiati in matrici 8x8 e 32x32. si tratta in tutti i casi di pixel 3T standard, differenti per topologia e per la presenza o meno di guard-ring e p-block. È stato inoltre inserita una matrice di tipo WIPS (weak inversion pixel sensor).

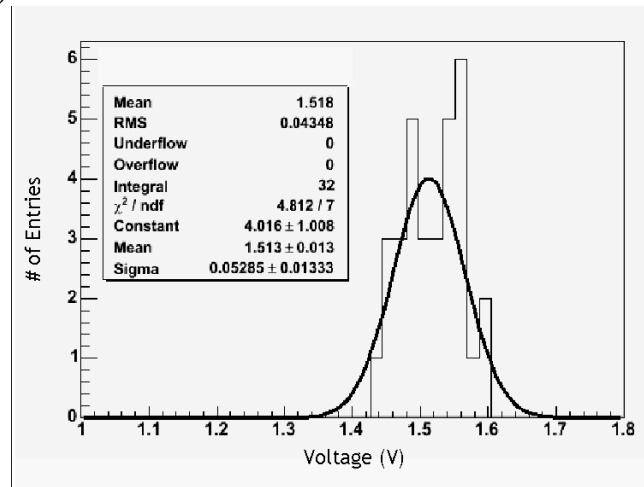


Figura 47 - Distribuzione dei valori di uscita di una matrice 8x8 APS buia. Ogni pixel è valutato 75us dopo la fine della fase di reset

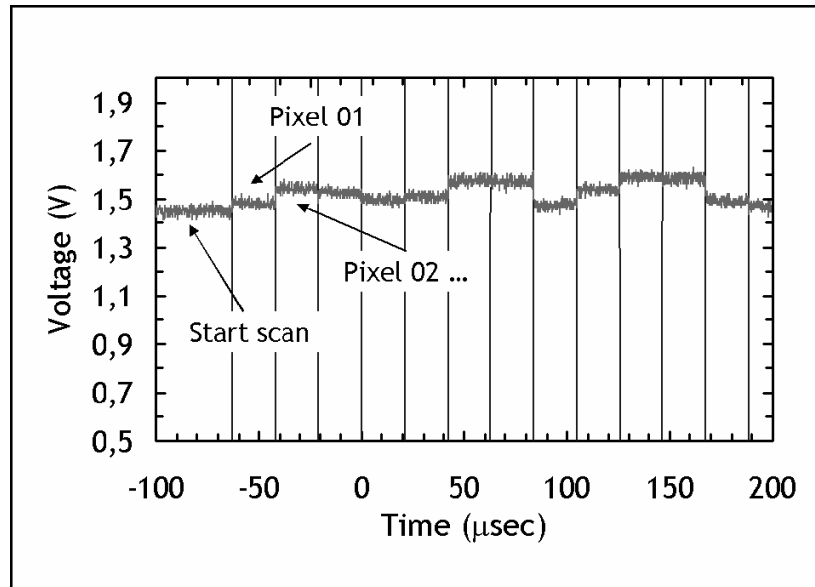


Figura 48 - Uscita di una matrice APS in condizioni di buio. Si nota la variazione dovuta al FPN

Ogni colonna possiede una propria elettronica di lettura, che prevede i seguenti componenti:

- Polarizzatore del Source-follower del pixel: è uno specchio di corrente in cui la tensione del pMOS è configurabile
- Amplificatore differenziale: uno stadio differenziale per la prima amplificazione del segnale in uscita dalla matrice
- Polarizzatore del Differenziale: identico al polarizzatore del SF, genera la corrente di polarizzazione dell'amplificatore differenziale
- MUX analogico: seleziona la colonna per la lettura
- Comparatore: formato da 2 NOT in cascata, fornisce un'uscita digitale che informa se il pixel è stato illuminato o meno.

La matrice presenta inoltre la possibilità di configurare il funzionamento dell'elettronica di lettura a valle del pixel. Sono infatti presenti nel chip alcuni registri digitali programmabili dall'esterno, collegati a dei semplici convertitori DAC. Questi dispositivi generano tre tensioni di polarizzazione

interne, che permettono di settare guadagno e punto di lavoro del circuito di lettura. È in particolare possibile selezionare:

- La tensione di polarizzazione del pMOS che guida lo specchio di corrente del bus di colonna
- La tensione di riferimento dello stadio differenziale all'uscita della matrice
- La tensione di polarizzazione del pMOS che guida lo specchio di corrente dello stadio differenziale

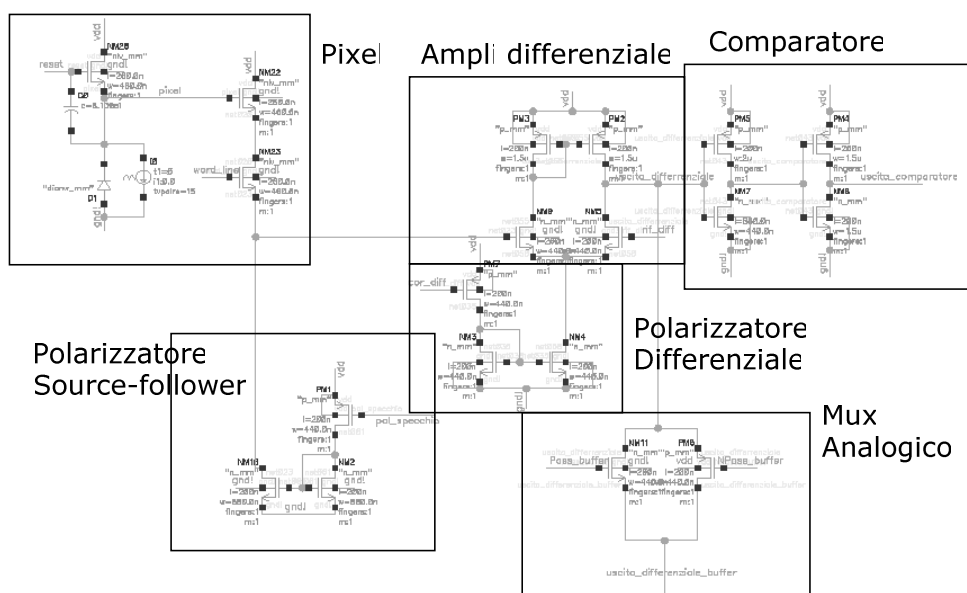


Figura 49 - Schematico dell'elettronica di lettura delle matrici del RAPS-01

Ognuna di queste tensioni è selezionabile tramite un registro a 5 bit (32 valori possibili). La caratteristica del DAC che le genera è fortemente non lineare, in modo da avere più opzioni nel range di tensioni di maggior utilità. In base alle tensioni scelte si possono ottenere caratteristiche di trasferimento lineari, oppure caratteristiche a soglia, utili nella MIP detection.

Lo svantaggio di questo approccio è la presenza misurata di un forte FPN di colonna (nell'ordine dei 15-18mV), in quanto questa elettronica non è comune a tutta la matrice, bensì è replicata per ogni colonna.

Dalle misure effettuate sui prototipi inseriti nel RAPS-01, sono risultati di maggior interesse i pixel di tipo G1P0 e G1P1. i segnali di read e di reset vengono applicati ad ogni riga tramite due registri a scorrimento, in modo che ogni riga venga letta e subito dopo resettata (rolling shutter).

RAPS-02

Il chip RAPS-02 ha incorporato le tipologie di pixel G1P0 e G1P1, che sono risultate le più performanti fra quelle inserite del sensore precedente, oltre ad una matrice WIPS (che a causa di un errore progettuale non era stato possibile testare) e ad una matrice di un nuovo pixel, SHARPS, che estende il WIPS permettendo anche una lettura event-driven.

Tre matrici 32x32 con pixel di tipo 3T standard sono presenti all'interno del RAPS-02: G1P1, G1P0, G1P0L. Come detto, le G1P1 e G1P0 differiscono per la presenza o meno dello strato p-block attorno alla n-well del fotodiode. La differenza fra G1P0 e G1P0L risiede invece nella spaziatura fra i pixel: mentre il G1P0, è formato da pixel adiacenti, i cui guard-ring si sovrappongono, con un pitch di circa 4um, il G1P0L presenta un pitch maggiore in quanto formato da pixel G1P0 distanziati di 12um l'uno dall'altro.

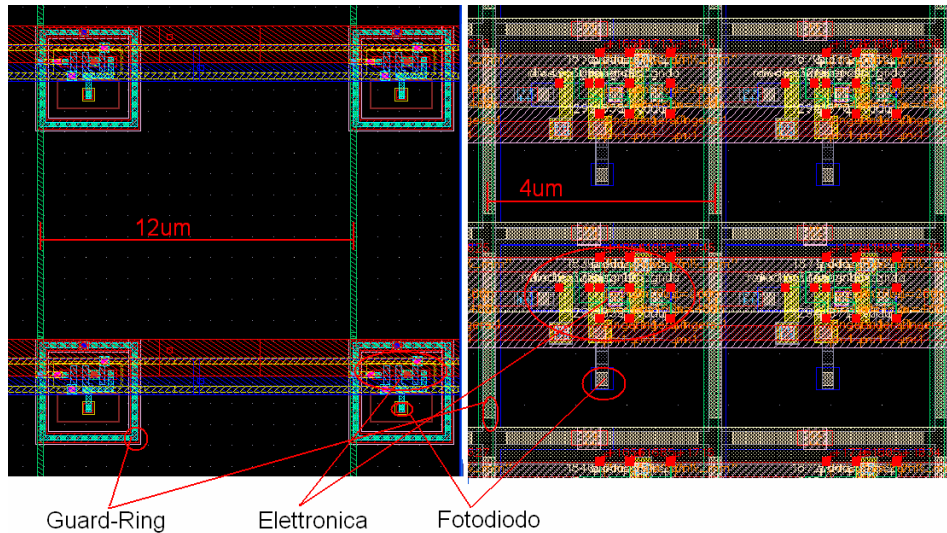


Figura 50 – 4 Pixel G1P0L (sx) e G1P0 (dx). Si noti che nel G1P0L il guard-ring non delimita i pixel, ma è posto attorno al fotodiiodo ed elettronica

L'elettronica di lettura della matrice è identica rispetto l'elettronica impiegata nel chip RAPS-01

Pixel SHARPS

Il pixel SHARPS si configura come un pixel in grado di fornire in uscita solo le coordinate del punto d'impatto, come il WIPS, ma rispetto a questo aggiunge una serie di potenzialità inedite, come la lettura event-driven e il reset automatico. Inoltre, come il WIPS, contiene al proprio interno dei pMOS, che permettono un migliore trattamento del segnale.

La struttura del pixel prevede un fotodiiodo connesso ad un source follower a sua volta connesso ad un inverter CMOS. Il source follower serve per far vedere al fotodiiodo una capacità inferiore di circa la metà rispetto ad un CMOS direttamente attaccato e per disaccoppiare il punto di lavoro del fotodiiodo e dell'inverter. Alla base dell'inverter è posto un nMOS che, collegato ad un'adeguata tensione di polarizzazione, permette di

modificarne la caratteristica di trasferimento in modo che la tensione d'ingresso, col pixel non colpito, si trovi in corrispondenza della zona ad alto guadagno. L'uscita dell'inverter è collegata al gate di un nMOS che collega i bus di colonna e di riga corrispondenti al pixel stesso; è inoltre connessa in feedback al gate del MOS di reset del fotodiodo.

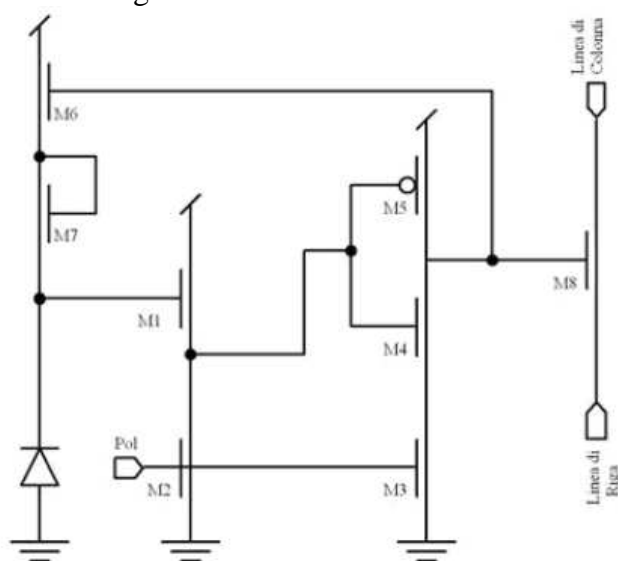


Figura 51 - Schematico del pixel SHARPS

In caso di impatto di una particella, l'abbassamento della tensione sul catodo del fotodiodo causa un forte aumento del valore d'uscita dell'inverter CMOS (essendo l'uscita del SF in corrispondenza della zona ad alto guadagno in ingresso dell'inverter), che permette l'accensione dell'nMOS che collega i bus di riga e colonna. Su questi, essendo posti uno ad alimentazione e uno a massa, si instaura una corrente che viene rilevata dalla circuiteria esterna alla matrice, generando un evento. In seguito all'evento generato, la logica invia all'esterno un burst di dati, contenente i valori di corrente letti su tutti i bus di riga e di colonna della matrice: un'informazione di N righe + M colonne, esattamente come per il WIPS. La differenza è che in questo caso il burst di dati viene generato in conseguenza di un evento, e non in seguito ad una scansione periodica come nel WIPS. Il

segnale in uscita dall'inverter, oltre a generare l'evento di lettura, viene impiegato anche per il reset del sensore, accendendo il MOS di reset. Grazie a questo, il pixel non necessita di nessun tipo di scan periodico generato dall'esterno, in quanto è in grado di compiere autonomamente tutte le operazioni necessarie al corretto funzionamento (lettura e reset).

Il segnale di impatto, generato tramite una circuiteria di rivelazione del passaggio di corrente su ogni colonna e riga, permette inoltre di discriminare tra più livelli di corrente, a seconda dell'entità della perdita di tensione sul fotodiode, effettuando un campionamento ADC a 1.5bit.

La tensione di polarizzazione del source-follower e la tensione del MOS di polarizzazione dell'inverter sono selezionabili tramite gli stessi registri dei transistor 3T.

Setup sperimentale

È stato creato presso l'Università di Parma un set-up sperimentale che consentisse il test del sensore da un punto di vista elettrico. Questo prevede la creazione di un banco flessibile, comprendente:

- Una board FPGA esterna per la generazione dei segnali di comando del RAPS-02 e per la successiva acquisizione e invio al PC dei dati
- Una scheda di sbroglio dei cavi del RAPS-02, che porti i segnali di maggior interesse ad un unico connettore
- Una scheda di supporto che colleghi la scheda di sbroglio del RAPS alla FPGA e che contenga anche l'elettronica di supporto (generazione tensioni di alimentazione, di riferimento e ADC)

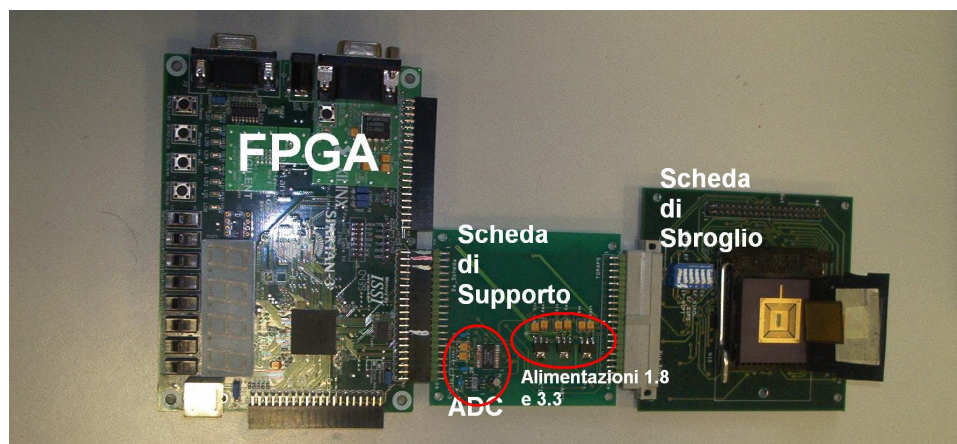


Figura 52 - Il setup sperimentale realizzato per il RAPS-02

Come si può vedere dalla figura, il sistema è molto semplice (la scheda ausiliaria impiega la stessa tensione di alimentazione della FPGA per generare le tensioni di alimentazione di ADC e RAPS) e permette inoltre una buona flessibilità nell'acquisizione dei dati grazie all'ADC a 12 canali (è possibile selezionare via VHDL quale uscita analogica fra G1P0, G1P0L e G1P1 convertire).

La scheda FPGA impiegata è la Spartan-3 Starter Kit della Digilent, che impiega una FPGA Spartan-3 (X3S200 FT256) e che può esser collegata al PC tramite porta seriale.

Caratterizzazione elettrica

Sul chip RAPS-02 è stato effettuato un intenso lavoro di caratterizzazione, sia dal punto di vista dell'elettronica che della MIP detection (presso le strutture dell'Università di Perugia).

Il rumore misurato è stato calcolato come la deviazione di ogni pixel su 100 acquisizioni uniformi di buio rispetto il proprio piedistallo, e facendo la media quadratica di questi valori per tutti i pixel della matrice. Per ogni matrice è stato calcolato il piedistallo come la media di ogni pixel su almeno

100 acquisizioni; quindi è stato stimato il FPN come la deviazione standard di tutti i pixel del piedistallo. Il FPN di colonna è stato invece calcolato calcolando prima la media di tutti i pixel di ogni colonna, e quindi stimando la deviazione standard di queste medie di colonna. Le prove sono state eseguite in prevalenza con gli amplificatori in polarizzazione lineare e ad una frequenza di funzionamento di 100kHz.

Infatti, come accennato in precedenza, il FPN di colonna è risultato la forma di rumore dominante, spaziando tra gli 11-18mV contro i 2-4mV del FPN di pixel, che è stato stimato invece come deviazione standard media dei pixel su una stessa colonna.

Pixel	G1P1	G1P0	G1P0L
Instant noise (mV)	0.97	1.3	1.25
FPN (mV)	1.95	4.7	4.5
Column FPN (mV)	9.6	17.7	16.9

Più contenuto è stato invece stimato il valore del rumore istantaneo, nell'ordine di 1-1.5mV a seconda della matrice.

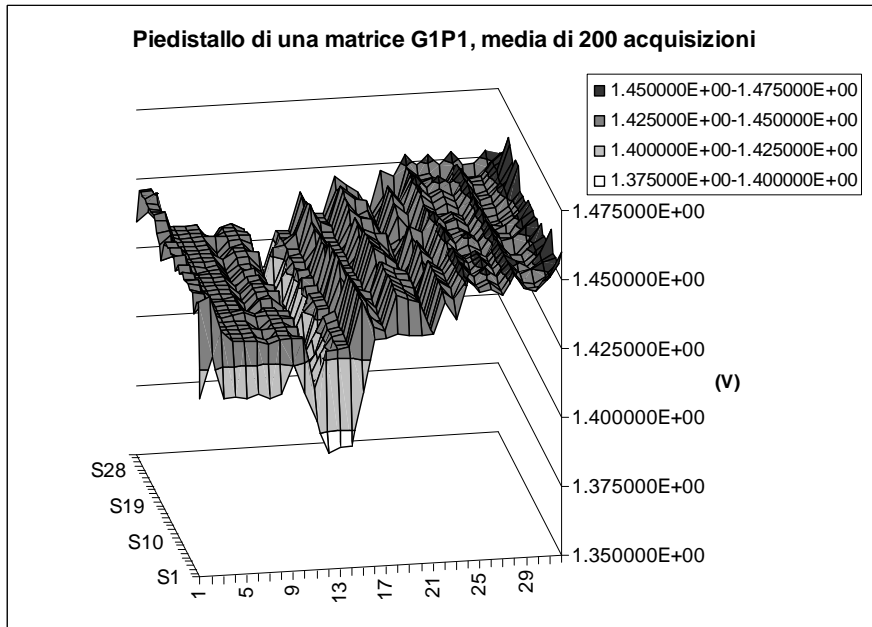


Figura 53 - Piedistallo di una matrice G1P1: è evidente il FPN di colonna

Per correggere il FPN sono state provate delle soluzioni di correzione via software, come il DS: acquisendo prima un piedistallo (cento frames bui), questo è stato poi sottratto all'immagine ottenuta, permettendo una buona riduzione del FPN (dell'ordine del 35%), soprattutto quello di colonna. Per esempio, in una matrice G1P1 illuminata (luce ambientale, pixel saturati), il FPN totale è sceso da 14.4mV a 9.2mV, mentre il FPN di colonna è stato ridotto del 50%, passando da 12.9mV a 6.2mV.

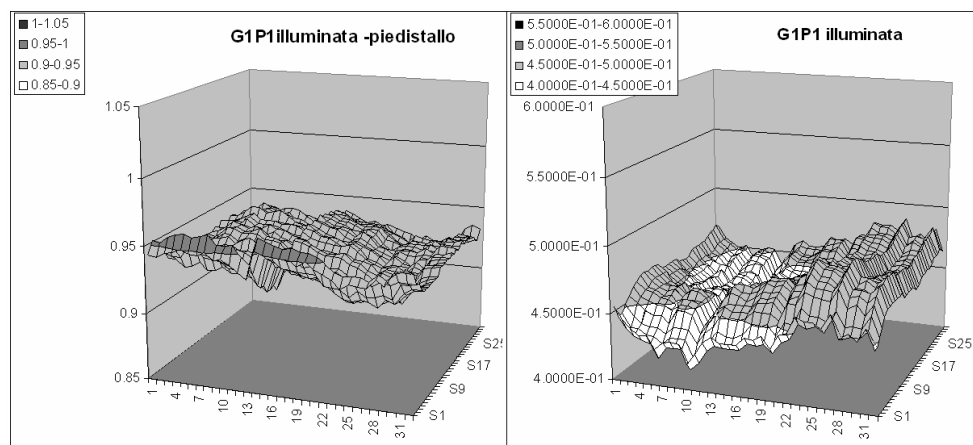


Figura 54 - Esempio di correzione con sottrazione del piedistallo dalla matrice illuminata. Si noti la riduzione del Column FPN

Caratterizzazione funzionale

Per la caratterizzazione della risposta alla radiazione incidente, il chip è stato sottoposto seguenti prove:

- Esposizione a sostanze radioattive che rilascino particelle di varia natura (raggi X, γ , particelle α , particelle β)
- Irraggiamento X
- Caratterizzazione mediante laser: il sensore è stato illuminato con fasci laser di differenti lunghezze d'onda (infrarosso, visibile, ultravioletto)

Scopo di questi esperimenti è l'individuazione di una serie di parametri di interesse:

- SNR in caso di MIP
- Segnale in funzione della frequenza della radiazione incidente
- Profondità effettiva di raccolta
- Stima dello spreading di carica
- Massima risoluzione ottenibile

Assumendo R come rumore termico della matrice, per valutare il segnale in caso di single-photon counting, la procedura adottata per caratterizzare la risposta a MIP prevede i seguenti passi:

- Acquisizione del piedistallo (cioè media per ogni pixel di una matrice su N acquisizioni al buio)
- Acquisizione delle matrici irraggiate da MIP e calcolo per ogni pixel della sua differenza rispetto al piedistallo
- Ricerca del seed pixel, cioè del pixel che mostra la maggior escursione rispetto al proprio piedistallo e superi una certa soglia S_1 , stabilita in base al rumore dei pixel di quella matrice ($S_1 = M * R$, $M > 1$). Se nessun pixel supera questa soglia, la matrice è scartata in quanto non risulta colpita
- Ricerca del cluster: procedimento ricorsivo: entro un raggio specifico dal seed pixel, ogni pixel adiacente al seed pixel o ad un pixel del cluster viene vagliato e, se mostra una deviazione rispetto al piedistallo maggiore di una seconda soglia S_2 ($R < S_2 < S_1$), viene a sua volta aggiunto al cluster.
- Valutazione del segnale: questo può essere considerato o come la variazione rispetto al piedistallo del seed pixel, o come la somma delle variazioni di tutti i pixel del cluster.
- Creazione di un grafico delle presenze: ogni valore così trovato viene inserito in un istogramma

La forma dell'istogramma, rispecchiando la perdita di energia di una particella che attraversa uno strato di materia (maggiore il valore ottenuto, maggiore la carica raccolta, che è proporzionale all'energia che la particella ha rilasciato nel silicio) tende ad avere la forma di una distribuzione Landau. Il valore considerato come most probable value (MPV) è il punto di picco della distribuzione.

Raggi X

Un primo lavoro di caratterizzazione è stato effettuato tramite irradiazione con raggi X.

Mediante l'impiego di sostanze radioattive e prove con tubi a raggi X per uso medicale e monocromatori, è stato possibile stimare l'escursione di segnale in seguito ad un impatto da fotoni a differenti energie.

Il sensore è stato sottoposto a fotoni X e sono state raccolte le escursioni di segnale dei pixel che risultavano colpiti, estrapolando così una curva del segnale in funzione dell'energia, che mostra come il segnale registrato dal RAPS risulti proporzionale alla lunghezza d'onda della radiazione X entro 1-50keV. La carica generata nei fotoni X inoltre è risultata confinata sempre entro il pixel colpito, senza dare origine a cluster di pixel.

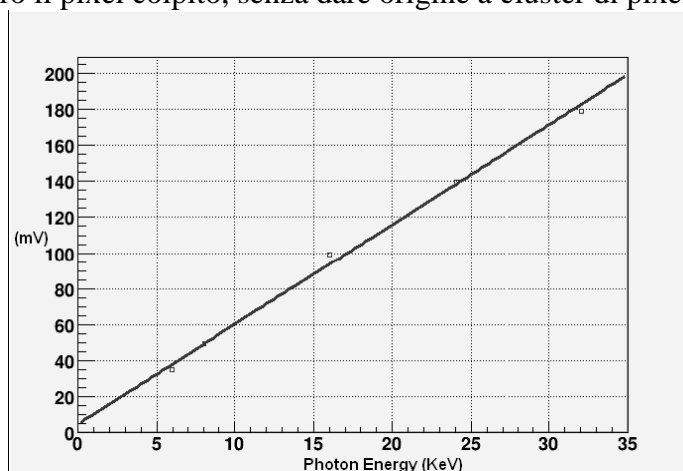


Figura 55 - Fit dei valori misurati con l'irraggiamento a raggi X di differente energia

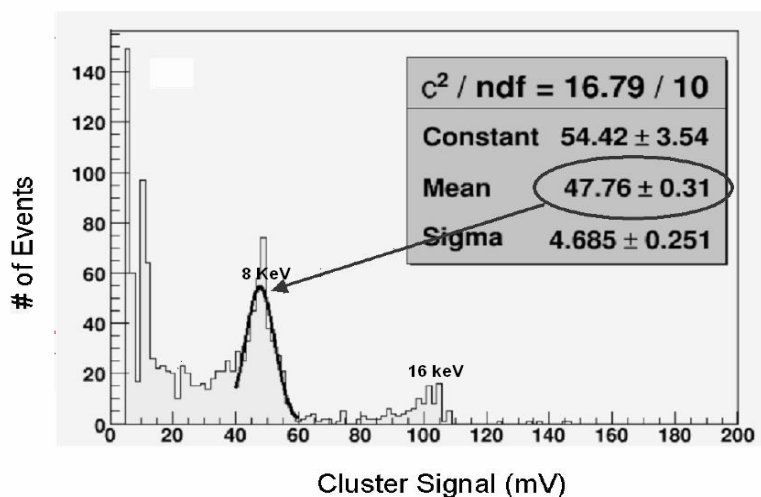


Figura 56 - istogramma delle visite per un tubo a raggi X e monocromatore a 8keV. E' visibile anche il picco dell'emissione secondaria a 16 keV

L'irraggiamento è stato effettuato mediante un isotopo ^{55}Fe , che rilascia fotoni a 5.6keV, e tubi a raggi X abbinati a un monocromatore a 8keV con 16 e 32keV di emissione secondaria. Effettuando un fit fra le escursioni osservate nei pixel colpiti in funzione dell'energia dei fotoni incidenti, si è potuto ottenere un rapporto di conversione fra l'energia del fotone e l'escursione del segnale pari a 128eV/mV. Considerando inoltre che un raggio X tende ad interagire con il silicio in maniera puntuale, rilasciando tutta la propria energia in un punto d'interazione, e che l'energia di band-gap (necessaria cioè a creare una coppia e^-/h^+) del silicio è pari a 3.6eV, si trova un rapporto di conversione che per il G1P0 è pari a 36 e^-/eV , in linea con la letteratura.

Sostanze radioattive

Le prove con sostanze radioattive si effettuano similmente a quanto precedentemente descritto, essendo la MIP detection fondamentalmente un'applicazione di photon counting. Nell'istogramma delle visite creato

(con la somma del segnale di tutti i pixel del cluster) sarà possibile riconoscere un picco in corrispondenza del buio, con una larghezza dovuta al rumore, e uno o più picchi a valori inferiori, dovuti al passaggio di particelle. La posizione di questi picchi, in particolare la distanza dal picco di buio, permette una stima della variazione di tensione indotta da una particella. Conoscendo il guadagno della catena di lettura e la capacità del fotodiode, è di conseguenza possibile stimare la quantità di carica raccolta e confrontarla con la quantità di carica nota che viene solitamente rilasciata nel substrato dalla particella in esame. In questo modo è possibile computare il SNR per le varie particelle.

Il sensore, irradiato in particolare con particelle β^- da ^{90}Sr e raggi γ da ^{141}Am , ha dimostrato di essere in grado di rivelare un gran numero di particelle dal comportamento assimilabile ad una MIP, sia pure con differenti SNR e dimensioni di cluster.

Tramite la scelta del funzionamento a soglia, in cui l'elettronica di colonna si comporta come un comparatore, e scegliendo un adeguato valore per la tensione di comparazione (selezionata caricando un registro a 5 bit), è stato possibile dimostrare la possibilità di ottenere un'escursione di segnale in uscita di circa 800mV, sufficiente a generare un evento di trigger al passaggio di una particella. Impiegando invece una polarizzazione lineare, è stato possibile effettuare misure più precise sulla raccolta e lo spreading della carica generata.

In particolare, la particella β^- genera un cluster di 4 pixel, indipendentemente dal fatto che la matrice usata sia una G1P0 o una G1P0L. Il segnale totale misurato sommando il valore di tutti i pixel nel cluster 2x2, invece, ha mostrato una distribuzione di Landau, con un MPV superiore a 25mV, che implica un SNR >20 per tutti i pixel.

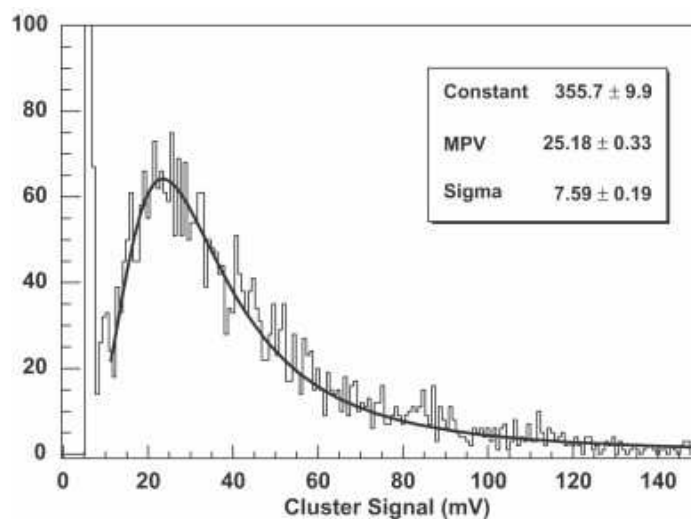


Figura 57 - Istogramma dei segnali di cluster misurati per l'irraggiamento da β^-

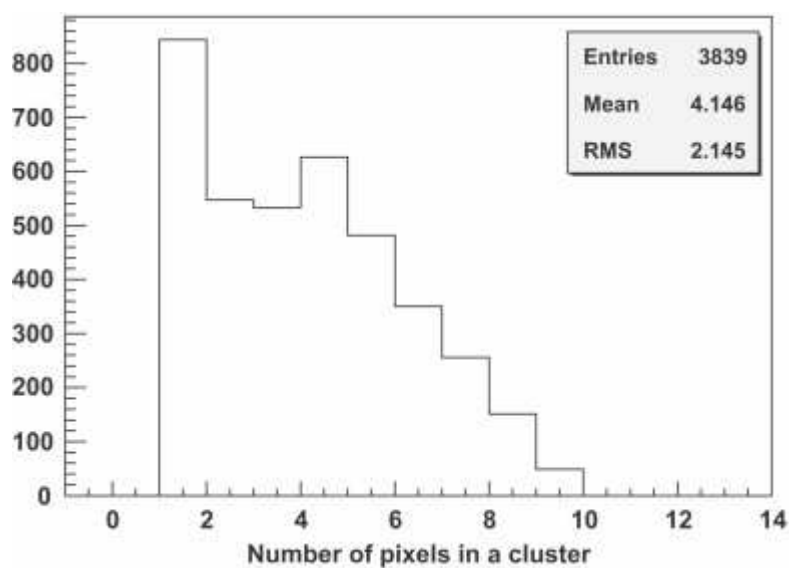


Figura 58 - Istogramma del numero di pixel facenti parte di un cluster per una β^-

Ricordando che, a differenza della gran parte della letteratura, la tecnologia impiegata nel progetto RAPS non prevede l'impiego di uno strato

epitassiale, per cui lo spessore di raccolta di carica non è noto a priori, è stata effettuata quindi una stima dello spessore effettivo di raccolta. Infatti è presumibile che la carica generata nel substrato oltre ad una certa profondità sia troppo lontana per venire raccolta, in quanto la distanza dall'elettrodo risulterebbe maggiore della lunghezza media di cammino. Di conseguenza è possibile calcolare una "profondità effettiva", per cui la carica generata da una particella è considerata come "totalmente raccolta" entro questa distanza, altrimenti è considerata "totalmente perduta". Sapendo pertanto che una MIP tende a rilasciare $80e^-/\mu\text{m}$, e conoscendo i parametri del sistema di lettura (cioè il rapporto fra la carica raccolta e il segnale osservato, nel caso in esame $36e^-/\mu\text{m}$), è possibile stimare la quantità di carica raccolta dal sensore e, dividendo tale quantità per $80e^-/\mu\text{m}$, ottenere il valore della profondità effettiva. Con 25.2mV di segnale è stata calcolata una raccolta di carica pari a circa 1000 elettroni da cui consegue una profondità di raccolta effettiva pari a circa 13 μm .

Prove laser

Infine, la caratterizzazione col laser, mediante l'impiego di diverse lunghezze d'onda, ha permesso di stimare le prestazioni nella raccolta di carica. Sono stati impiegati laser di 3 lunghezze d'onda, un infrarosso (1060nm di lunghezza d'onda), visibile (783nm) ed ultravioletto (407nm), con spot del diametro di 1-2 μm , che permettono di colpire un singolo pixel. L'impiego di tre diversi laser ha permesso di avere delle stime sulla raccolta di carica a differenti profondità, essendo la carica dell'infrarosso generata in profondità e quella dell'ultravioletto in superficie. Questo ha permesso di osservare come lo spreading di carica sia molto maggiore nel primo caso che nel secondo, dove il cluster è formato unicamente dal pixel colpito.

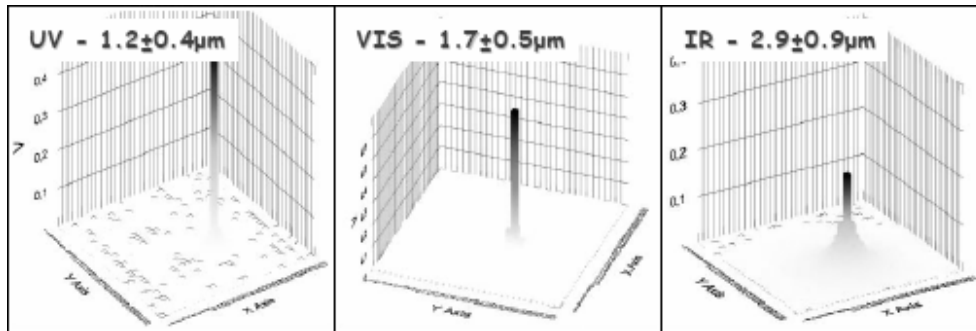


Figura 59 - risposta del G1P0 a ultravioletto, visibile e infrarosso

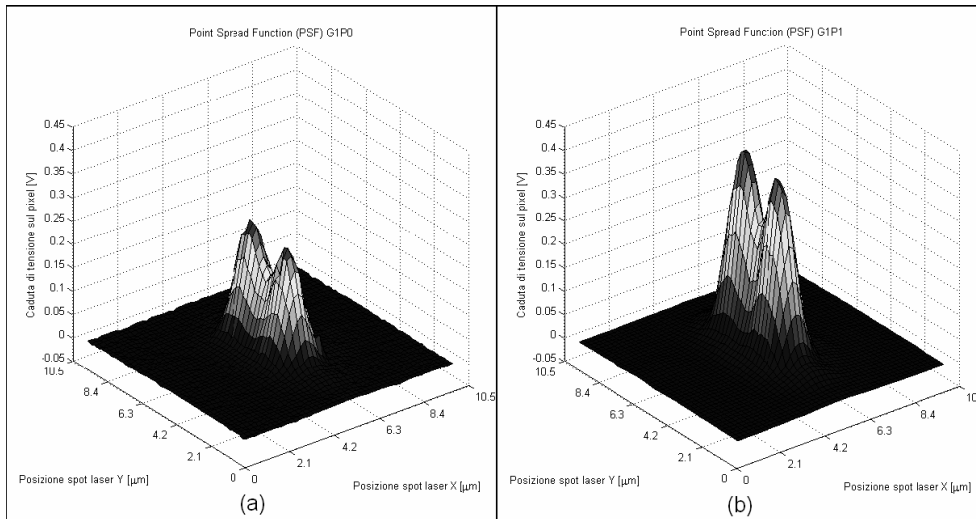


Figura 60 - Escursione di segnale per una G1P0 (a) e una G1P1 (b) con laser UV per diverse posizioni dello spot rispetto il centro del pixel. Il gap fra i due picchi è dovuto alla presenza in quel punto della metallizzazione sopra al fotodiodo

Ulteriori prove sono state effettuate sfruttando la possibilità di muovere il laser di step submicrometrici in modo da colpire punti diversi dello stesso pixel. Di conseguenza, spostando il laser e osservando l'escursione del pixel in ad ogni nuova posizione, è stato possibile mappare la sensibilità del pixel in funzione del punto d'impatto. Nel caso di laser UV, che non ha spreading di carica, sono stati trovati due picchi di segnale in corrispondenza del fotodiodo, con un gap fra i due dovuti allo schermo della metallizzazione

del contatto metal-n-well: quando il laser incide su questo punto viene schermato, causando una minore generazione di carica.

Allo stesso modo prove effettuate coi laser sulle differenti matrici hanno permesso di verificare la maggiore efficienza nella raccolta di carica da parte dei pixel G1P1 grazie al p-block, che mostrano un maggior segnale se illuminati con il laser UV rispetto i G1P0. Tuttavia questa efficienza vale solo in superficie, in quanto i test con il laser IR (carica generata in profondità, più simile ad una MIP) hanno dimostrato una differenza molto meno marcata fra i due pixel.

Infine i laser hanno permesso una stima teorica della risoluzione massima ottenibile dal sensore: illuminando il pixel in punti differenti e osservando il cluster ottenuto, si è calcolato il centro del cluster stesso semplicemente come media pesata dei pixel. Calcolando la differenza fra la posizione dello spot laser calcolata in base alle misure del RAPS e quella effettiva del laser, si è potuta stimare una risoluzione massima di 350nm per il laser visibile

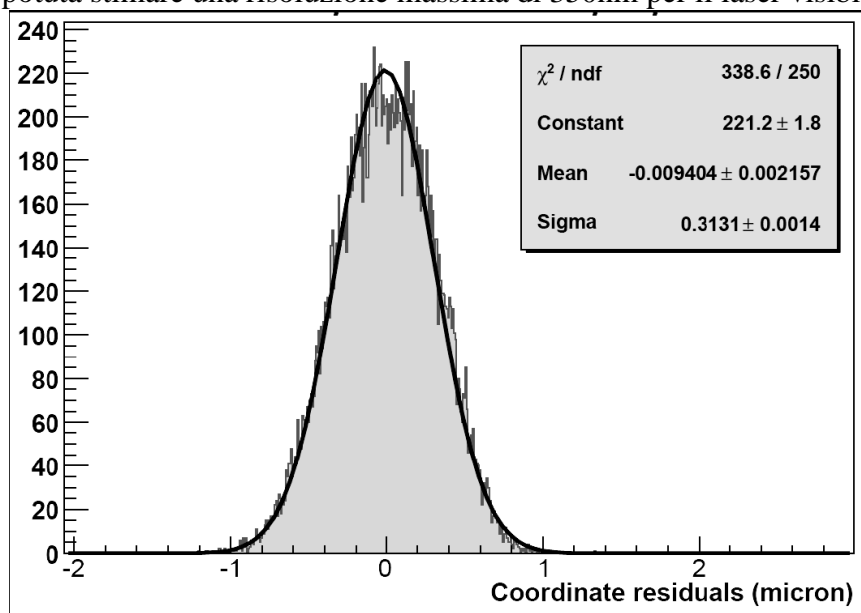


Figura 61 - istogramma degli errori fra la posizione misurata e la posizione effettiva dello spot.

CAPITOLO 3: IL RAPS-03

Limiti del chip RAPS-02

Come si è visto, i risultati delle misure e dei test del chip RAPS-02 sono risultati molto incoraggianti, sia dal punto di vista dei raggi X che dal punto di vista delle MIP. Tuttavia il chip presenta una serie di limitazioni:

- Scarsa flessibilità di impiego (il funzionamento della matrice non può essere in nessun modo influenzato dall'utente)
- Matrici di dimensioni limitate (32x32) che non permettono prove del sensore in condizioni di funzionamento realistico
- Scarsa frequenza di lavoro (~100kHz)
- Elettronica analogica di lettura
 - o poco flessibile (essendo configurata tramite registri da 5 bit, è impossibile una regolazione fine dei valori)
 - o complicata (3 gradi di libertà su cui agire: non è banale trovare un setup adatto)

Per questo è stato necessario lo sviluppo di un nuovo sensore che consenta una maggior flessibilità d'uso e un'operazione di definizione del funzionamento più semplice per l'utente, oltre alla presenza di strutture più simili a quelle di un sensore pienamente operativo.

Specifiche del chip RAPS-03

Il chip RAPS-03 è l'ultimo sensore sviluppato nell'ambito del progetto RAPS. La caratteristica principale rispetto al RAPS-02 è la presenza di una matrice di dimensioni maggiori, vicine alle dimensioni di un chip operativo, che potrebbe essere impiegata sia nella rivelazione di MIP che nella

rivelazione di raggi X per uso medico. È stata così decisa la progettazione del nuovo dispositivo, secondo le seguenti linee guida:

- Reset globale invece del rolling reset
- Semplificazione della circuiteria di lettura dal punto di vista dell'utente
- Possibilità di decidere la temporizzazione del circuito (durata delle fasi di reset ed integrazione) per meglio adattarsi a differenti impieghi
- Possibilità di forzare la lettura della matrice solo in seguito ad un evento esterno invece di scandire periodicamente la matrice
- Funzionamento ad alte frequenze (necessaria una frequenza di clock di almeno 16MHz)
- Esplorazione di architetture alternative (pixel 4T, enclosed gate) mediante apposite strutture di test

Come si può notare, è stata ricercata la maggiore flessibilità possibile, in modo da avere un sensore in grado di adattarsi ad ambiti differenti, sia medicali (raggi X) che di fisica delle alte energie (MIP).

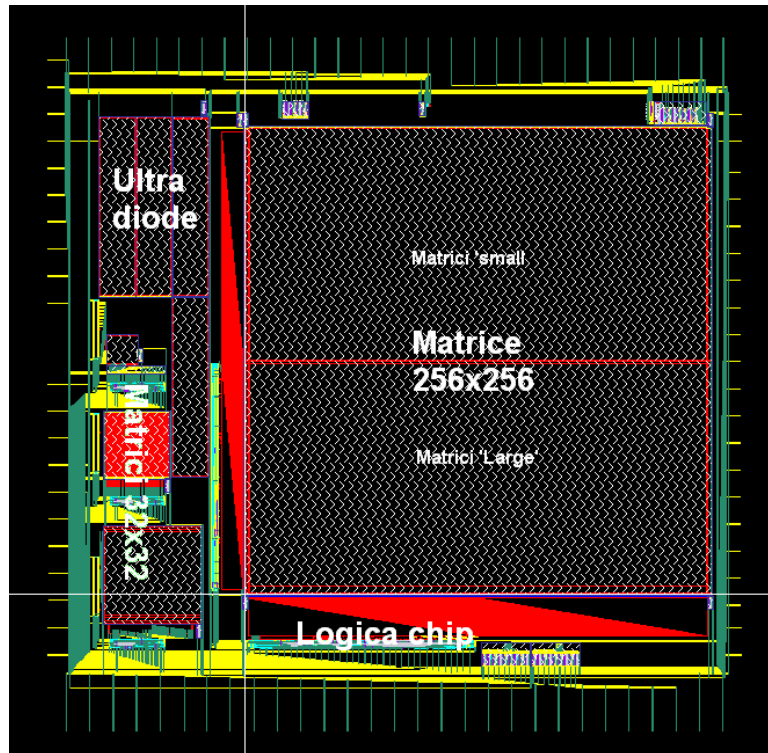


Figura 62 - Il sensore realizzato con le strutture principali

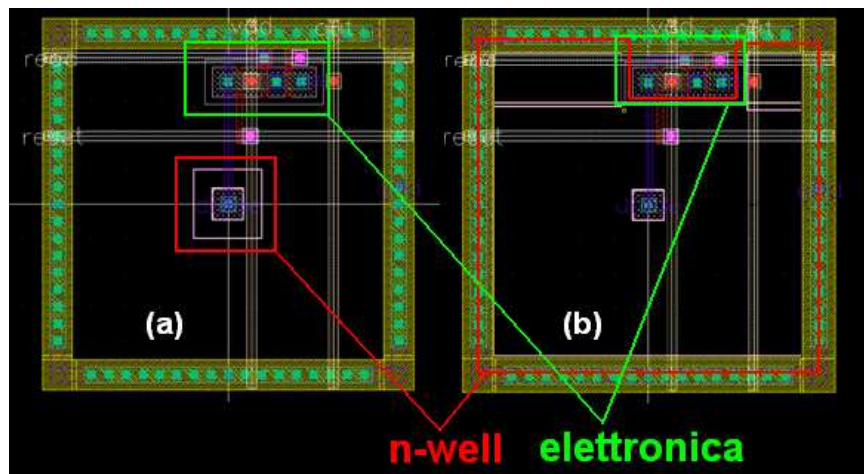


Figura 63 - I due pixel della matrice principale: 'small' (a) e 'Large' (b)

La matrice principale del RAPS-03

La matrice principale è formata da un array di 256x256 pixel, arrangiati in 4 sottomatrici di 128x128 pixel, ognuna con la propria circuiteria di lettura, che forniscono in uscita quattro valori analogici in parallelo. In realtà per evitare effetti di bordo, che nelle matrici del RAPS-02 portavano ad una raccolta di carica molto superiore per i pixel esterni, ci sono due ulteriori file di pixel attorno alla matrice che non vengono lette, e che sono state inserite per raccogliere l'eventuale carica ai bordi della matrice salvaguardando così l'uniformità della raccolta di carica degli altri pixel.

I pixel che costituiscono la matrice, mostrati in Figura 63, sono dei 3T standard di dimensione pari a $10 \times 10 \mu\text{m}^2$, differenti fra loro come dimensione e forma del fotodiodo. Il layout prevede un elevato fill-factor, superiore all'80%, e un anello di guard-ring a delimitare il perimetro del pixel, in modo da limitare lo spreading di carica fra pixel adiacenti. Il fotodiodo è formato o da un normale contatto nwell-psub di $2 \times 2 \mu\text{m}^2$ (pixel "small"), oppure da una n-well che occupa l'intera area del pixel non occupata dall'elettronica di reset-lettura (pixel "large"). Si intende in questo modo studiare la differenza, dal punto di vista della raccolta di carica, fra una zona svuotata più estesa, ma anche con capacità maggiore (miglior raccolta di carica, ma peggior conversione di questa carica in una tensione misurabile) rispetto una zona svuotata di dimensioni inferiori. Delle 4 sottomatrici, le 2 sottomatrici superiori sono composte di pixel di tipo "large", mentre le 2 sottomatrici inferiori sono composte di pixel di tipo "small".

La politica di reset prevede la possibilità di resettare contemporaneamente tutti i pixel di una matrice, mentre la lettura viene effettuata serialmente. A causa di ciò i pixel di uno stesso frame si trovano ad essere integrati per tempi differenti, pur iniziando l'integrazione al medesimo istante. Questo effetto viene tuttavia reso trascurabile da un'elevata frequenza di funzionamento, che permette di effettuare l'acquisizione di un intero frame

in meno di 1ms a fronte di tempi di integrazione previsti nell'ordine delle decine di ms, e che permette di conseguenza di mantenere una buona uniformità della matrice. Infatti in questa differenza di tempo il leakage non dovrebbe dare effetti rilevabili, e può esser semplicemente visto come una forma di FPN che penalizza leggermente il dynamic range degli ultimi pixel letti, che hanno integrato più a lungo, a parità di esposizione la fotocorrente e/o la corrente di buio.

Ricordando inoltre gli ambiti in cui è previsto il funzionamento del sensore

- MIP detection: dopo il reset, la matrice è tenuta in integrazione in attesa di un evento, e al verificarsi di questo viene effettuata la lettura per trovare il punto di passaggio di una o più particelle. Di conseguenza in un simile funzionamento non è critica la differenza di tempo d'integrazione fra il primo e l'ultimo pixel letto. Il problema potrebbe venire dalla perdita di eventi subito durante il readout, in quanto un pixel che venisse colpito nell'intervallo intercorrente fra l'istante in cui è letto e l'istante in cui verrà resettato perderebbe l'informazione di questo evento. Un simile fenomeno potrebbe esser problematico per i primi pixel letti, ma l'elevata frequenza di funzionamento prevista dovrebbe mitigare anche questo problema.
- X-ray imager: la radiografia funziona sottoponendo il tessuto da esaminare ad un "flash" di raggi X ed effettuando una lettura del sensore. Anche in questo caso, essendo generalmente il flash sincronizzato con l'istante di acquisizione di un frame, ed essendo questo l'unico evento che porti informazione (prima e dopo i pixel sono sempre uniformemente bui), la differente durata dei tempi d'integrazione si manifesterebbe solo come lieve incremento del FPN dovuta al maggior contributo della corrente di dark negli ultimi pixel letti.

La scansione può essere o periodica ("automatica") o forzata da un evento esterno ("event-triggered") selezionando l'adeguato ingresso. Questo permette di adattarsi a differenti utilizzi, in quanto in molti ambiti di richiede un sensore in grado di fornire dati solo quando richiesto. Negli

esperimenti di fisica delle alte energie, per esempio, si hanno dei sensori che registrano unicamente il passaggio di particelle (per esempio i calorimetri) e che di conseguenza generano un evento di trigger per i vertex detectors sottostanti, in modo che la loro lettura possano fornire le coordinate di passaggio della suddetta particella.

Alla base di ogni bus di colonna è presente un transistor che funge da stadio polarizzatore del common-drain del pixel letto. Durante la lettura, tutti i pixel di una riga vengono collegati al bus di colonna, quindi tramite un multiplexer analogico viene selezionato ad ogni ciclo di clock un pixel all'interno della riga viene selezionato. L'uscita del multiplexer analogico è collegata ad un op-amp in configurazione follower che guida il pad di uscita. Nella catena di lettura è possibile selezionare la tensione analogica da fornire al transistor che polarizza il bus di colonna. Bassi valori di tensione portano ad una peggiore performance dal punto di vista della velocità, mentre alti valori portano ad una riduzione dell'escursione del segnale in uscita.

Strutture di test del chip RAPS03

Il chip prevede inoltre una serie di strutture secondarie per testare e confrontare alcune opzioni tecnologiche alternative. In particolare sono presenti all'interno del chip tre matrici più piccole di 32x32 pixel, dotate di una propria logica di lettura con le stesse caratteristiche delle matrici principali (configurabilità dei tempi di reset ed integrazione, lettura event-triggered). Le matrici si differenziano per i tipi di pixel che le costituiscono

- Matrice pixel small: matrice composta di pixel simili allo small della 256x256
- Matrice pixel G1P0: matrice dotata di guard ring simile alla matrice presente sul RAPS-02, ma dal pitch leggermente maggiore (4.4um)
- Matrice pixel G1POL: pixel identici alla G1P0, ma spazati fra di loro, per poter effettuare maggiori controlli sulla dispersione della carica fra i pixel. Il pitch è di circa 12um

Le matrici già presenti nel RAPS-02 sono di dimensioni maggiori, in quanto è stato osservato che sotto una certa dimensione lo spreading di carica inter-pixel rende inutile aumentare la risoluzione spaziale. La terza matrice di pixel small è stata inserita per confrontare i dati delle matrici più grandi con i dati provenienti da una matrice di dimensioni minori, in modo da poter effettuare confronti sull'impatto del numero di pixel sulle prestazioni del sensore. Un aumento delle dimensioni della matrice infatti aumentando le capacità di colonna tende ad aumentare i tempi di risposta e di lettura del singolo pixel, proprio quando sarebbe necessario avere una maggiore velocità di funzionamento per tenere costante la frame-rate.

La logica di controllo è simile a quella della matrice principale, con la differenza che, essendo possibile effettuare la lettura di un intero frame in tempi inferiori (1024 cicli invece di 16384), anche il massimo tempo di integrazione selezionabile è molto più basso (16384 cicli di clock invece di 4 milioni), in quanto è importante unicamente evitare che la differenza relativa di tempo d'integrazione fra il primo e l'ultimo pixel letto sia paragonabile al tempo d'integrazione. Volendo integrare la matrice per tempi superiori, è possibile sia diminuire il periodo di clock, sia leggere la matrice in modalità event triggered provvedendo il segnale di start esterno a intervalli regolari.

Sul chip sono inoltre presenti una serie di strutture necessarie per caratterizzare meglio le performance dei singoli elementi della matrice, oltre ad esplorare eventuali architetture alternative. Queste strutture sono in genere collegate ai rispettivi piedini di output tramite un buffer a guadagno unitario, per pilotare la maggiore capacità delle pad di I/O del chip. Sono presenti

- Una struttura "ultradiode" composta di circa 8000 diodi del pixel small connessi in parallelo per studiare la capacità del fotodiode
- Una struttura di 2x11 pixel, indirizzabili mediante un decoder e controllabili provvedendo il segnale di reset direttamente dall'esterno.
- Due matrici 3x3 in cui ogni pixel è collegato direttamente ad un pad di uscita

- Una serie di strutture formate da singoli pixel (i cui segnali di controllo vanno forniti direttamente dall'esterno) collegati all'esterno con l'intera catena di read-out, senza elettronica di indirizzamento.
- Tre buffer con ingresso ed uscita collegati all'esterno per la loro caratterizzazione.
- Una serie di singoli transistor lineari ed enclosed-gate

La struttura "ultradiode" è composta da 8000 diodi connessi in parallelo, e collegati col catodo direttamente all'output (senza buffer) analogico, mentre l'anodo è posto direttamente a massa. Tramite questa configurazione è possibile effettuare misure dirette di conversione carica-tensione per valutare la capacità dei dispositivi.

La struttura 2x11 è invece formata da pixel "small", che sono stati opportunamente schermati. I pixel sono infatti coperti dalle metallizzazioni superiori non usate per il routing (Metal-3, Metal-4, Metal-5, Metal-6). Delle due colonne una è completamente schermata, mentre i pixel della seconda presentano un buco quadrato di $1\mu\text{m}^2$ a distanze variabili dal centro del pixel (ovvero dal fotodiodo). Sarà così possibile sia valutare i segnali di buio di ogni pixel, sia la variazione della raccolta di carica in funzione della distanza dal centro. A differenza delle matrici 32x32 e 128x128, questa matrice è priva di logica di selezione e controllo: il segnale di reset viene ricevuto dall'esterno, mentre il pixel sotto osservazione è selezionato in base a 4 segnali digitali esterni che tramite un decoder e un MUX analogico collegano all'uscita il pixel corrispondente.

Per la valutazione dell'efficacia di differenti tipologie di pixel, sono state inserite varie strutture, sia come pixel singoli che come matrici 3x3 (ma con un'uscita analogica per ogni pixel). Queste vengono controllate, come la 2x11, provvedendo gli adeguati segnali di controllo dall'esterno. Sono presenti le seguenti strutture alternative alla 3T impiegata nelle matrici principali:

- Enclosed gate
- Pixel 4T

I pixel 4T sono pixel in grado di salvare al proprio interno anche il valore alla fine della fase di reset, per facilitare il CDS (Correlated Double Sampling) o implementare un vero global shutter. Sono stati realizzati in tecnologia CMOS standard, senza l'impiego di tecniche ad hoc che facciano uso per esempio diodi 'pinned'. In questo tipo di pixel lo scopo principale è il salvataggio del valore di reset sul pixel in modo da leggere in due cicli consecutivi di clock sia il valore di reset che il valore di luce corrispondente e poter così effettuare il correlated double sampling.

In questi pixel la carica raccolta dal fotodiodo non raggiunge il pass transistor (e attraverso questo il MOS common-drain) passando attraverso una diffusione n nel bulk, bensì attraverso una metallizzazione. Sul chip RAPS-03 sono presenti singoli pixel di questo tipo fatti impiegando layout e numero di diodi differenti.

I pixel enclosed-gate sono a loro volta pixel 4T, costruiti impiegando un diverso tipo di transistor. Infatti l'enclosed-gate è un pixel in grado di sopportare, grazie alla propria forma anulare, maggiori esposizioni a radiazione senza l'incremento della corrente di sottosoglia che rende inaffidabili i normali MOS lineari.

Proprio per studiare al meglio la possibilità di impiegare questa soluzione in futuri sviluppi, una serie di singoli transistor di differenti dimensioni, sia lineari che enclosed-gate sono stati sviluppati, in modo da poter valutare le curve di I_d in funzione sia della tensione gate-source che source-drain, che ne permettono la caratterizzazione. Infatti uno dei maggiori ostacoli all'impiego di questi dispositivi è la difficoltà nel calcolare il rapporto W/L . Se nei lineari il valore di questo rapporto è intuitivo, negli enclosed, a causa della particolare conformazione del gate, risulta complicato ed anzi non si è ancora giunti ad una formula universalmente accettata [18]. In generale, si tratta di spezzare il MOS in una serie di MOS più piccoli in parallelo, per cui i 4 lati vengono a contribuire al fattore di forma come $4W/L$, mentre il calcolo del rapporto per i MOS corrispondenti agli spigoli dell'anello (che possono essere o smussati o a 90°) risulta incerto. Inoltre va considerata la presenza di una "linguetta" di polisilicio che esce dall'anello per collegare il gate con l'esterno. Ciò rende non banale la progettazione di questi

dispositivi, per cui risulta difficile passare dal layout al modello matematico, o viceversa. Per estrapolare la dipendenza del fattore di forma dalla geometria del MOS si sono di conseguenza implementati dispositivi variabili sia come dimensioni dell'anello (che tenderà ad influenzare il parametro W), sia lo spessore (corrispondente al parametro L), più una serie di MOS lineari che in base ai calcoli del modello di Giraldo[27] corrispondano agli enclosed inseriti.

Logica del chip

La logica di controllo delle matrici, come detto, è pensata in base alle seguenti linee guida:

- Funzionamento ad alta frequenza
- Flessibilità nella scelta delle temporizzazioni
- Generazione di segnali di feedback per sincronizzare il sistema di lettura esterno col flusso di dati proveniente dal chip
- Possibilità di lettura continua della matrice o in seguito ad evento esterno

Il flusso di progettazione è stato effettuato mediante il tool Synopsys di Cadence. Si è partiti da una descrizione VHDL della logica. Questa si basa sull'impiego di più macchine a stati finiti principali per controllare tutte le operazioni che la logica può eseguire. In particolare, due macchine a stati finiti distinte vengono impiegate per il controllo dei due tipi di matrici, le 128x128 e 32x32, mentre un'altra viene impiegata per gestire il flusso seriale di dati da e per i registri che controllano le temporizzazioni di reset (8bit) ed integrazione (14 bit). La logica genera inoltre una serie di segnali di feedback:

- Sync (o start frame): alto quando si legge il pixel 0,0 della matrice.
- End col: alto quando si scandisce l'ultimo pixel di una riga.
- End row: alto quando si sta scandendo l'ultima riga della matrice.

La macchina a stati finiti per le matrici prevede uno stato “idle”, in attesa di un segnale di start. A seconda del segnale di start esterno che riceve, segue una sequenza di funzionamento piuttosto che l'altra:

- Start “automatico”: ciclo continuo tra reset – integrazione – scan matrice, che termina solo in seguito ad un segnale di reset della logica o del chip. La durata di ogni ciclo dipende o dal valore caricato nei registri (reset, integrazione) o dalla dimensione della matrice (1024 per le 32x32 e 16384 cicli di clock per le 128x128).
- Start “singolo”: si esegue subito uno scan della matrice, seguito da una fase di reset, per poi tornare allo stato “idle” in attesa di un nuovo segnale di start.

Continous Scan

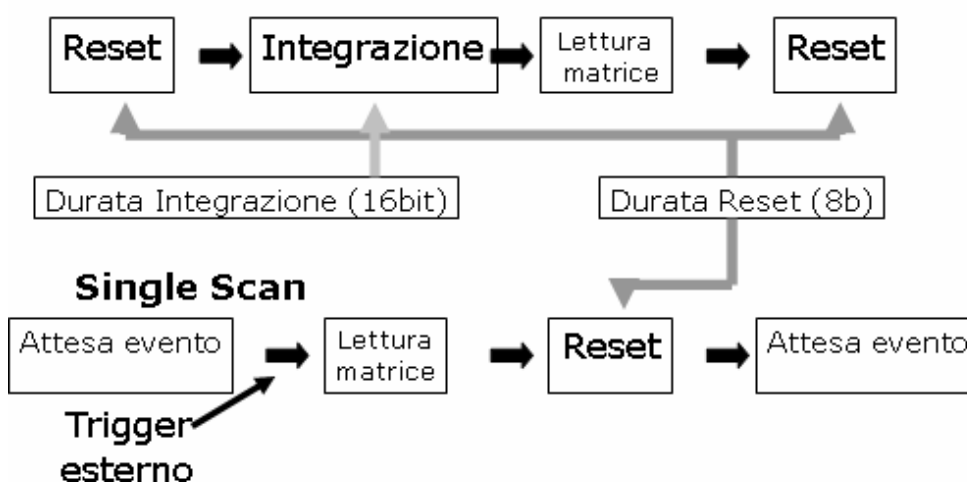


Figura 64 - Schema della successione delle operazioni nei due modi di scan continuo e event-triggered.

La logica è stata progettata per funzionare con un segnale di clock fino a 100MHz di frequenza, molto maggiore comunque rispetto la frequenza a cui si intende lavorare, che è di circa 20MHz (basata sulla necessità di leggere una matrice in meno di 1ms).

Il VHDL è stato suddiviso in 4 top-sections, due per la logica delle matrici 32x32 e due per quella delle 128x128. Questa suddivisione è stata imposta

da una riflessione per l'ottimizzazione del layout, in quanto i segnali di indirizzamento possono essere suddivisi in due classi, in base al lato della matrice da cui provengono (NON dal lato cui sono paralleli):

- Orizzontali: segnali che interessano le colonne di pixel: bus di lettura, reset globale, selezione di colonna
- Verticali: segnali che interessano le righe: segnale di select del pixel.

Proprio per questo si sono separati i blocchi di logica che generano o gestiscono questi segnali: è stata generata una sezione laterale per generare i segnali di abilitazione delle righe (fondamentalmente si tratta di un registro a scorrimento della stringa 00...001) ed una sezione inferiore per generare i segnali di abilitazione delle colonne e di reset. Il top inferiore contiene inoltre la logica della macchina a stati finiti, oltre ai registri in cui sono contenuti i valori delle temporizzazioni.

I registri per le temporizzazioni sono due registri a scorrimento da 22 bit, per la scrittura e la lettura seriale dei dati. Ponendo alto un segnale di `data_valid` e fornendo un clock per i registri (diverso dal clock del resto del circuito), il registro a scorrimento di ingresso è abilitato e riceve ad ogni clock il segnale presente all'ingresso digitale `data_in_reg`.

Per leggere i dati all'interno del registro (necessario per aver conferma dei valori caricati), bisogna prima fornire un fronte di salita ad un segnale apposito, che carica il registro di lettura coi dati del registro di scrittura, quindi fornire il clock dei registri. In questo modo il registro di lettura, che è sempre abilitato, fornisce serialmente i dati in uscita, senza comunque influenzare l'altro registro, su cui non sono eseguite operazioni. È inoltre possibile mediante i segnali "default" e "res_asin_reg" caricare immediatamente e in maniera asincrona i registri unicamente o con una stringa di '1' o di '0'.

Test

Per testare il comportamento del chip è stato progettato un setup che permettesse di effettuare le prove desiderate con la massima semplicità e

flessibilità. È stato di conseguenza scelto un approccio che riduca al minimo i componenti necessari per il test, consentendo inoltre un controllo ottimale e pienamente configurabile sul sensore e sul flusso dei dati.

Il sistema progettato si basa pertanto sui seguenti componenti

- Scheda ad hoc per alloggiamento del sensore, generazione delle tensioni di alimentazione e di riferimento, conversione A/D dei segnali
- Logica esterna programmabile via software (FPGA) per la generazione dei segnali di controllo e per l'invio dei dati digitali al PC
- PC per l'acquisizione dei dati

Come si vede il sistema progettato è facilmente trasportabile e riposizionabile, così come la possibilità di riconfigurare facilmente l'elettronica consente di modificare velocemente il funzionamento del sensore per adattarlo a condizioni differenti di funzionamento.

Setup sperimentale: hardware

Il setup sperimentale per il test del chip è stato creato in collaborazione con l'Università di Perugia. Esso consiste in una scheda contenente l'elettronica di supporto per fornire al chip le alimentazioni e i segnali di controllo e per acquisire i segnali in uscita da esso.

Per avere facilità di configurazione e costi bassi, i segnali di controllo del chip, così come il flusso di dati digitali in uscita dagli ADC (sia interni che esterni) vengono gestiti con una FPGA esterna. Attualmente è in uso una Evaluation Board della Opal Kelly. Questa scheda provvede assieme all'FPGA un'interfaccia USB per la comunicazione con PC e programmazione della scheda, ma è disponibile anche un'interfaccia JTAG per la sua programmazione tramite un VHDL creato con il tool ISE di Xilinx. Si era in precedenza impiegata una scheda di minori prestazioni, cioè l'evaluation board della Opal Kelly, che incorpora una FPGA Spartan-3,

una SDRAM da 32MB e un'interfaccia seriale col PC, ma che non poteva acquisire a frame-rate elevate a causa della scarsa bit-rate di quest'ultima.

Il PCB è stato studiato in base alle caratteristiche di funzionamento previste per il RAPS-03, in particolare è necessario:

- Fornire le alimentazioni richieste (1.8 analogica e digitale, 3.3 digitale) con il minor rumore possibile
- Fornire una frequenza di almeno 16MHz al RAPS
- Campionare ad una frequenza fino a 64MSamples/s (per avere fino 4 campioni di ogni pixel) con 12 bit di risoluzione
- Generare on-board la tensione di polarizzazione del RAPS
- Generare on-board le tensioni di riferimento degli ADC
- Interfacciarsi con facilità con l'FPGA ed eventuali ulteriori ADC esterni

Si è optato per la progettazione di un PCB a 4 strati, fabbricato presso la Sunstone Circuits.

In questo PCB i due strati interni sono usati uno come piano di massa e l'altro come piano di alimentazione, in modo da evitare il cross-talk fra le piste di segnale dello strato superiore e di quello inferiore. Inoltre è stata mantenuta la distanza minima fra due piste pari a 7 mils, evitando quanto più possibile che piste che portino i segnali a maggior frequenza (clock, segnali di sincronia) corrano parallele e vicine. Nei piani di massa ed alimentazione, il piano di massa analogica è tenuto separato da quello digitale (ovviamente più rumoroso), e lo stesso vale per le alimentazioni. I due piani sono collegati tramite un induttore, che si comporta da cortocircuito per la continua, ma che blocca il passaggio di perturbazioni ad alta frequenza fra un piano e l'altro. Sono inoltre previste capacità di bypass fra i piani di massa e di alimentazione per filtrare i disturbi AC.

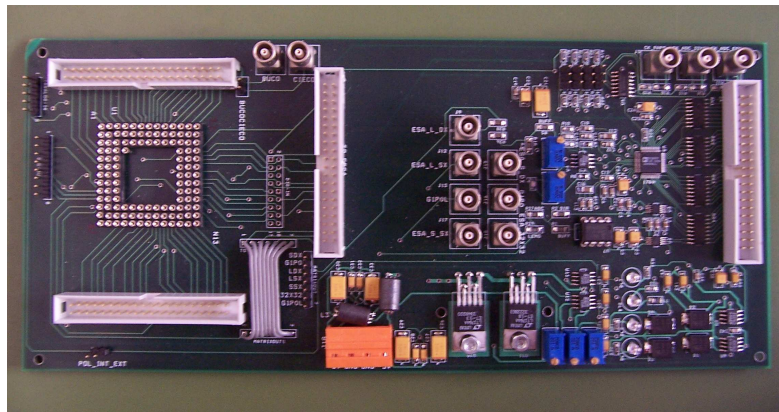


Figura 65 - La scheda sviluppata per il test del sensore RAPS-03

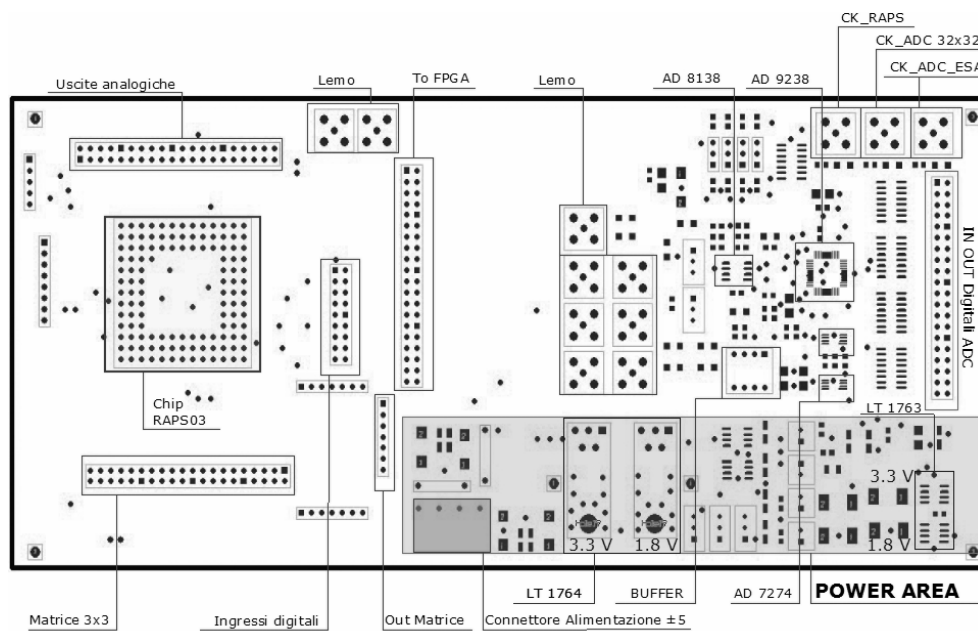


Figura 66 - Componenti della scheda

La scheda viene alimentata tramite un'alimentazione differenziale a $\pm 5V$. A bordo della scheda, oltre alle alimentazioni del chip sotto test sono generate anche le tensioni di polarizzazione del RAPS e di riferimento degli ADC. Sulla scheda sono montati due convertitori analogico-digitali:

- AD9238: un ADC flash pipelined con uscita a 12bit paralleli, 2 canali d'ingresso e di uscita, massima frequenza di campionamento di 65MHz, pensato per le matrici 128x128
- AD7274: un ADC ad approssimazioni successive da 3MS/s con uscita a 12 bit seriali per le 32x32, che sono state studiate per funzionare a frequenze più basse.

Questi ADC possono esser collegati ad una sola matrice alla volta, per cui per poter acquisire l'intera matrice 256x256 sarà necessario collegare ad un ADC esterno le uscite analogiche delle matrici rimaste scollegate.

Inoltre i due canali del convertitore flash sono collegati allo stesso ingresso analogico, ma uno riceve l'ingresso così com'è, mentre l'altro lo riceve tramite un driver (creato con un AD8138, driver ADC a bassa distorsione). In questo modo sarà possibile anche effettuare prove con soluzioni differenti per l'acquisizione, in modo da trovare quale sia la più affidabile.

Setup sperimentale: firmware

I segnali di controllo, come precedentemente scritto, sono generati da un'FPGA esterna. Nel caso dello Spartan 3 Starter Kit questa viene controllata tramite porta seriale, con la quale si possono mandare i seguenti comandi:

- Reset della logica FPGA e del RAPS
- Caricamento dei registri del RAPS
- Inizio dello scan del chip

Il controllo delle operazioni avviene anch'esso per via seriale: tramite il programma di controllo residente sul PC è infatti possibile inviare i comandi principali:

- Reset
- Carica Registri (+ i valori da caricare)
- Start automatico
- Start singolo

Ogni comando invia alla seriale una word di 8 bit, in cui 3bit contengono il codice del comando. I restanti 5 bit sono impiegati dal comando di load dei registri per selezionare un valore di periodo di integrazione da caricare (il reset invece resta fisso). In particolare, il periodo di integrazione, per abbracciare il maggior range di valori, è dato da 2 elevato il valore a 5 bit dato, il che permette di caricare i registri con qualsiasi valore da 1 (256 cicli di clock per l'integrazione) a 4096 (2M cicli di clock).

Una volta ricevuto il segnale di start, l'FPGA fornisce a sua volta il segnale di start adeguato al RAPS, quindi si mette in attesa dei segnali di feedback del chip. Una volta ricevuto il segnale di sync, l'FPGA inizia quindi ad acquisire i valori provenienti dagli ADC flash (12 bit per ogni pixel ad ogni ciclo di clock) e ad immagazzinarli nella RAM interna, fino a che legge contemporaneamente alti i segnali di end_row e end_column (cioè sta leggendo l'ultimo pixel del frame). Una volta acquisito e salvato nella RAM un frame intero, cioè 16384 valori, questa viene scandita serialmente e i dati relativi ad ogni pixel vengono suddivisi in due word da 8 bit che vengono quindi inviate al PC via seriale.

Durante questo trasferimento il RAPS continua però a funzionare, e tutte le nuove scansioni non sono acquisite fino a quando non viene terminato l'invio del frame. Di conseguenza la seriale forma un collo di bottiglia nel trasferimento dei dati, permettendo di acquisire solo una frazione dei frame generati. Infatti l'invio di 16384 word da due byte ad una frequenza di 57600 bips richiede $16384 \cdot 16 / 57600$ s, pari a 4.551s. Impiegando per esempio un tempo d'integrazione di 2M cicli di clock e una frequenza di 7.1MHz (setting a cui sono state fatte tutte le prove finora), il tempo massimo totale per ogni frame (reset + integrazione + lettura) risulta di 296ms (frame-rate di circa 3.3frames/s), il che porta alla possibilità (best-case) di osservare circa un frame ogni 15 in uscita dal RAPS.

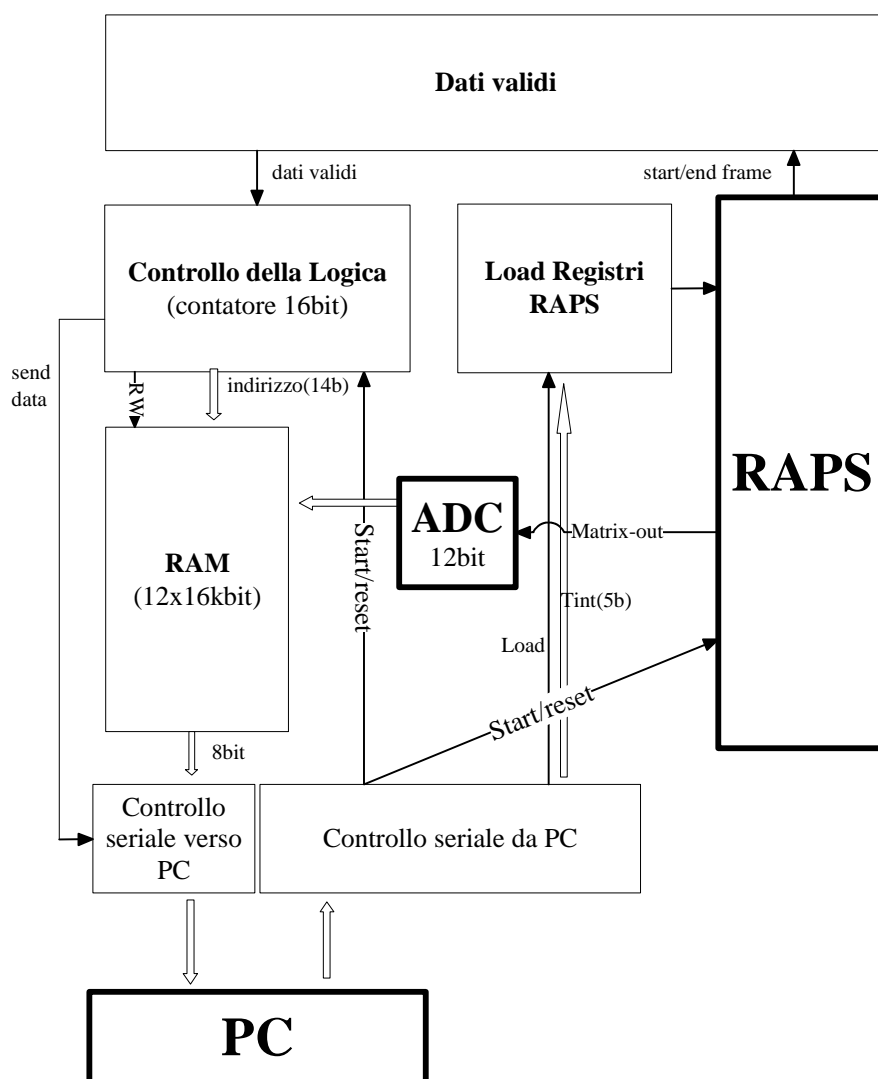


Figura 67 - Schema a blocchi del VHDL sviluppato per la Starter Kit

È stato quindi effettuato un porting del VHDL per poter utilizzare una evaluation board della Opal Kelly, che offre la possibilità di impiegare la porta USB per dialogare col PC. La ditta fornisce assieme alla scheda

l'interfaccia con la porta USB, sia a livello di primitive VHDL che a livello di librerie per il software.

Partendo dallo schema impiegato per la scheda precedentemente usata, è stato sufficiente modificare i moduli per la ricezione dei comandi via seriale e per l'invio dei dati via seriale, sostituendoli con moduli equivalenti dal lato FPGA, ma dialoganti tramite porta USB dal lato PC. Le librerie fornite dal costruttore allo scopo di trasferire dati rendono disponibili le seguenti strutture (tutte possono essere usate sia in invio che in ricezione):

- Wires: i cavi sono 16 linee ad 1 bit per il trasferimento asincrono dei dati
- Triggers: si tratta di 16 segnali a 1 bit che vengono sincronizzati con il clock del ricevente per generare un singolo impulso: vengono impiegati come segnali di avviso
- Pipes: implementate per il trasferimento sincrono di grossi blocchi di dati, permettono la lettura/invio di $2N$ words da 8bit, con N stabilito dall'utente.

L'invio dei segnali di controllo dal PC è simile a quello descritto per la scheda precedente, con la differenza che, essendo possibile trasferire 16 bit tramite wire, assieme ai 3 bit di codice del comando, rimangono disponibili 13 bit di informazioni accessorie, impiegabili per settare direttamente il valore di tempo d'integrazione da caricare nei registri (essendo il registro dei tempi di integrazione del RAPS di 14 bit, il suo LSB è in tal caso sempre a '1'). La presenza di un nuovo comando è segnalata dall'attivazione da parte del PC di un bit di trigger.

Per la lettura dei dati da RAPS, il VHDL prima scrive sulla RAM interna i dati da inviare, quindi a scrittura ultimata, invia serialmente all'interfaccia i dati nella RAM.

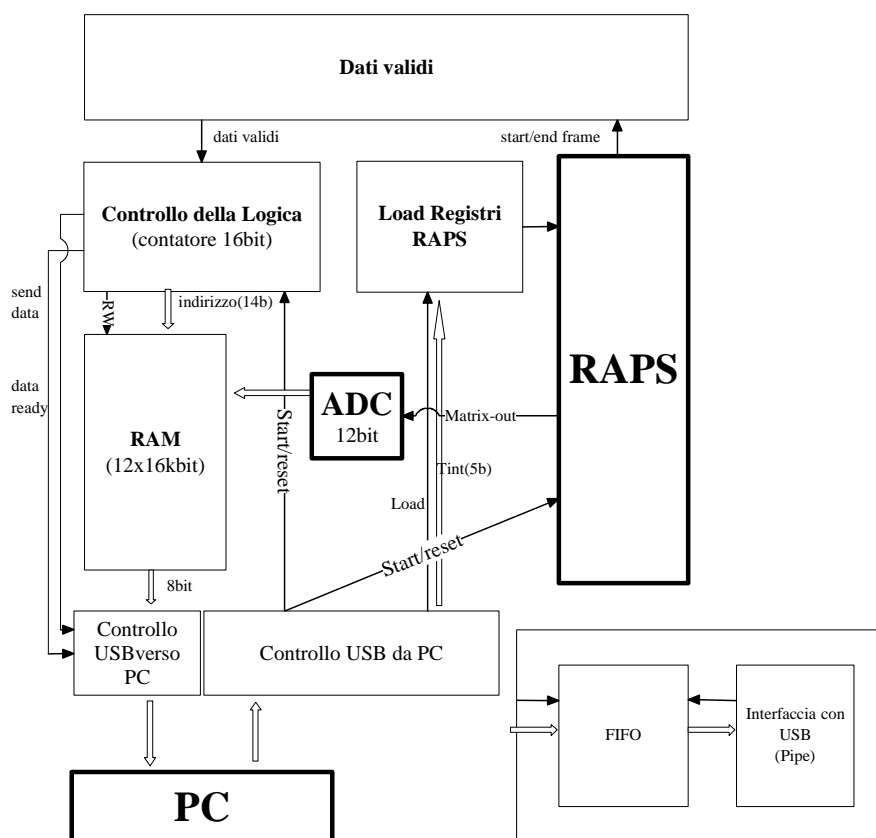


Figura 68 - Schema a blocchi del VHDL per la Opal Kelly. Nel riquadro è mostrato il diagramma del blocco “controllo USB verso PC”

L'interfaccia verso il PC è composta da una FIFO e da una pipe. I dati dalla RAM sono inseriti nella FIFO, che fa così da raccordo fra le due sezioni, che usano clock differenti (una usa il clock della scheda, l'altra usa il clock dell'USB). Si è preferito impiegare un doppio buffer (RAM e FIFO) per ridurre il rischio di acquisire frames parziali: infatti la seconda scansione della RAM per la scrittura su FIFO richiede lo stesso tempo della lettura di un frame e, come spiegato in precedenza, per ogni prestazione di interesse il tempo d'integrazione risulta comunque di molto superiore al tempo necessario per la lettura. Tuttavia può verificarsi il caso in cui la FIFO risulti

ancora parzialmente piena intanto che si sta acquisendo un nuovo frame: in questo caso potrebbe avvenire che ad un certo punto la FIFO risulti piena prima della completa acquisizione del frame e che quindi i successivi valori del RAPS vengano persi (non potendo il RAPS esser messo in pausa). Di conseguenza con l'approccio impiegato la FIFO viene scritta solo quando un intero frame è già memorizzato, per cui, nel caso pessimistico appena descritto, il riempimento della FIFO metterebbe in pausa la scansione della RAM, che riprenderebbe al liberarsi di nuovi posti nella FIFO senza la perdita di alcun valore.

Setup sperimentale: software

I dati inviati via seriale al PC vengono quindi letti tramite un programma in Visual Basic 6.0 appositamente creato che riordina i pacchetti di dati visualizzando le immagini progressivamente acquisite e salvandole su file di testo per una successiva elaborazione [Figura 69]. È inoltre possibile selezionare le dimensioni della matrice da acquisire (32x32 o 128x128).

Il programma stesso permette alcune semplici operazioni preliminari sui dati, come il calcolo del piedistallo in base ad una serie di preacquisizioni e l'eliminazione di matrici all'interno delle quali non sia stato registrato alcun evento. In pratica, il programma confronta il valore di ogni pixel col corrispondente pixel del piedistallo e se almeno uno di questi differisce di più di una soglia decisa dall'utente, il frame viene salvato, altrimenti viene scartato in quanto considerato privo di informazioni utili.

Si tratta della simulazione di un vertex detector, in quanto quando si deve monitorare un evento fisico, questo ha una frequenza di avvenimento spesso inferiore al frame-rate del sensore, per cui la maggioranza dei frames risulta priva di informazioni, e appesantisce semplicemente il sistema di acquisizione con dati inutili. Proprio per questo si tenta di impostare delle soglie che filtrino solo i dati effettivamente interessanti, che mostrino cioè un impatto (uno o più pixel più luminosi rispetto la media).

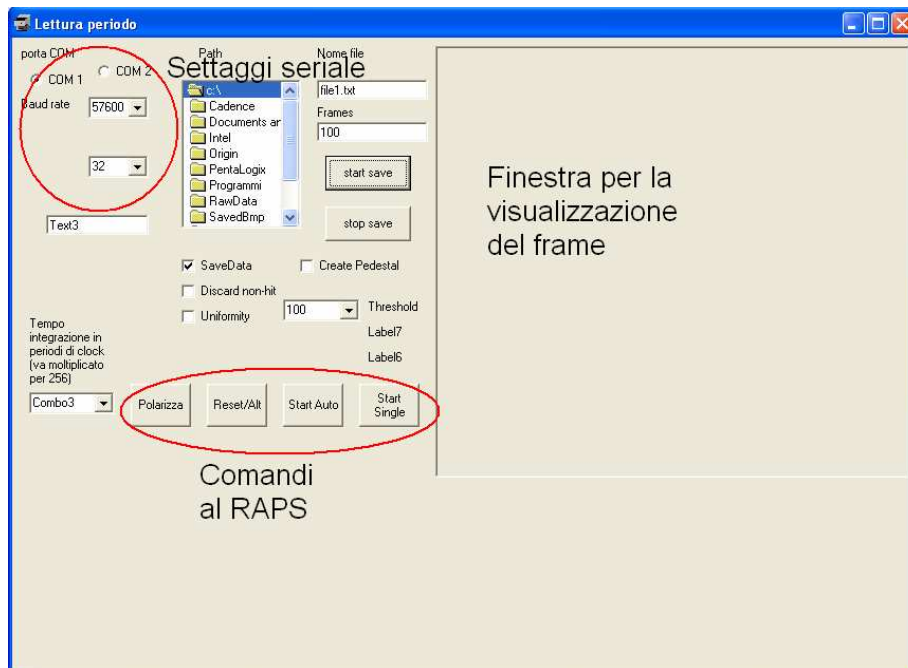


Figura 69 – Screenshot del programma in VB per l’acquisizione dei dati via seriale

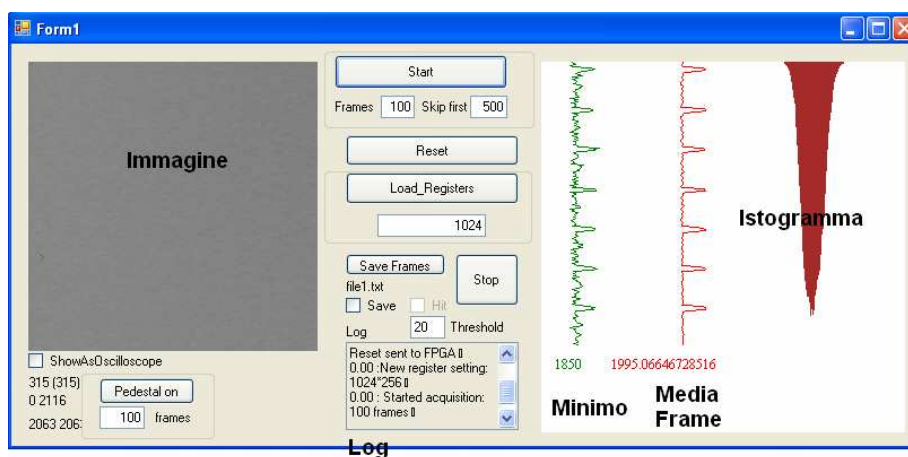


Figura 70 – Screenshot del programma in Visual C++ per l’acquisizione dei dati da USB. In questo caso è per esempio possibile notare un disturbo periodico sulla media dei valori acquisiti.

Per il controllo dell’USB si è invece preferito sviluppare un’applicazione in Visual C++ impiegando Visual Studio 2005 [Figura 70]. Anche questo programma prevede le stesse possibilità del software in Visual Basic, con la creazione di un piedistallo su un numero stabilito di acquisizioni preliminari e la possibilità di scartare i valori la cui deviazione rispetto il piedistallo risulti inferiore ad una soglia selezionabile. È inoltre possibile scegliere di non valutare un certo numero di frames preliminari, allo scopo di lasciare un certo periodo di tempo al RAPS per stabilizzare le uscite all’inizio della fase di funzionamento e visualizzare l’evoluzione nel tempo di alcuni parametri di interesse, come la forma dell’istogramma delle visite, la media e il valore minimo del frame.

Va notato che un eccessivo processing dinamico dei dati può essere problematico: poiché durante il processing il programma non può attivare la lettura dell’USB, questo può portare ad una parziale perdita di frames, in quanto l’FPGA smette di acquisire i nuovi frames quando trova la FIFO piena e in attesa di lettura.

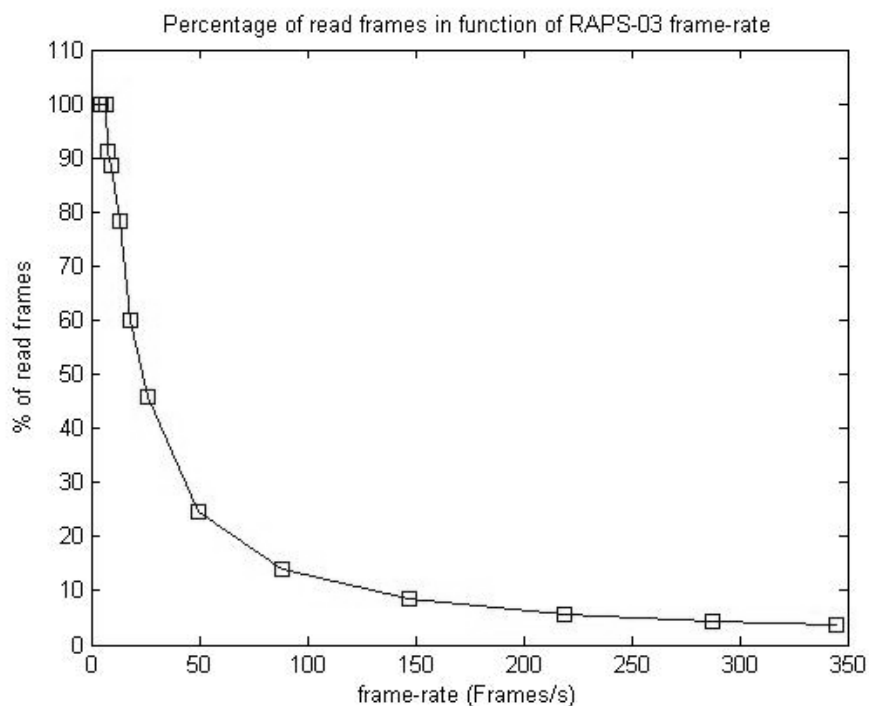


Figura 71 – Percentuale di frames acquisiti dal PC in funzione della frame-rate del RAPS

Caricando sull'Opal Kelly un firmware di prova, che genera una serie di matrici in cui il valore di tutti i pixel è il numero del frame stesso (il primo frame generato è composto da tutti 1, il secondo da tutti 2, etc...), è stato possibile contare quanti frames vengono effettivamente acquisiti dal PC per differenti frame-rates del RAPS semplicemente acquisendo 100 frames e leggendo il valore dell'ultima matrice (che dice quanti frames il RAPS ha generato nel frattempo). È stato visto che per frame-rate sufficientemente basse (<10Fps) la perdita di informazione è praticamente nulla.

Va ricordato comunque che oltre alla frame-rate, la latenza dipende fortemente anche dal programma di acquisizione, che deve finire di processare i dati di una matrice prima di acquisire la successiva. Infatti il tempo di latenza imputabile al firmware è pari al tempo necessario per

scrivere i dati dalla RAM alla FIFO dopo l'acquisizione di un frame (16384 cicli di clock), più un ciclo di clock per resettare il contatore, per cui con tempi d'integrazione maggiori di 16386 cicli (senza contare il reset), ogni eventuale latenza è da imputarsi alla connessione FPGA-PC. Questo ragionamento vale nell'ipotesi che la FIFO non venga mai riempita, ma se ciò accadesse sarebbe comunque dovuto ai componenti a valle. Il trasferimento dei dati dalla FIFO alla RAM avviene alla frequenza del cavo USB High-Speed che trasferisce 2 bytes per ogni pixel, per un totale di 32kB a una frequenza di circa 39MB/s (480Mb/s teorici, ma la frequenza effettiva è più bassa, si fa riferimento al manuale dell'utente della Opal Kelly): la lettura avviene di conseguenza in circa 800us, tempo che comunque non va sommato né alla latenza dovuta alla scrittura della RAM della FPGA, né a quella della RAM sulla FIFO, in quanto può avvenire in parallelo ad ambedue. Va invece considerata solo nel caso che la FIFO sia ancora piena (o si riempia), che potrebbe verificarsi nel caso che la fase di scrittura RAM-FIFO sia più veloce della lettura FIFO-USB e che il periodo di integrazione+reset+lettura del RAPS sia minore di 800us (difficile che accada).

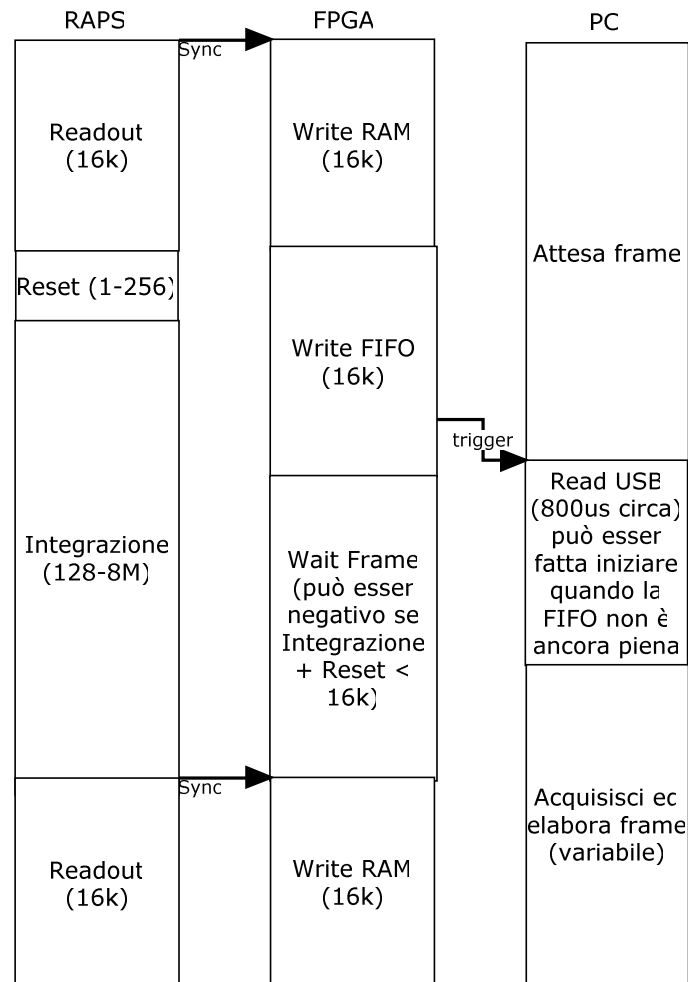


Figura 72 - Schema della sincronizzazione RAPS - FPGA - PC. I tempi sono specificati in cicli di clock

Ogni ulteriore latenza è quindi da imputarsi al programma, che prima di interrogare la porta USB per acquisire una nuova matrice deve terminare il processing sulla matrice precedente, lasciando di conseguenza il nuovo frame nella FIFO della FPGA per il tempo necessario. Nella fattispecie, la stima è stata effettuata con un programma di acquisizione che effettua una

certa dose di processing del frame (calcolo in RT di media e valore minimo, oltre alla visualizzazione dell'immagine). Impiegando una versione del programma dotata di minore calcolo ausiliario è possibile di conseguenza abbattere ulteriormente la latenza. Durante la creazione del programma di acquisizione si è comunque tenuto conto di questo aspetto e si è cercato di ridurre il tempo di latenza imputabile al programma:

- mediante l'uso di buffer grafici (gli aggiornamenti della grafica vengono prima scritti su un registro di memoria e solo alla fine disegnati su schermo)
- effettuando il salvataggio su file dei dati solo al termine di tutte le acquisizioni previste, in quanto l'accesso ai file è relativamente lento. Questo metodo potrebbe esser problematico nel caso si voglia acquisire un grandissimo numero di frames (~100000), perché potrebbe arrivare a saturare la RAM del computer.

Misure sul sensore

Una volta messo a punto il sensore, è stato svolto un lavoro preliminare di caratterizzazione elettrica. Il lavoro svolto presso l'Università di Parma, in particolare, è stato orientato verso la caratterizzazione preliminare delle strutture del sensore dal punto di vista elettrico, con la valutazione di alcuni parametri importanti per i successivi test funzionali.

In particolare sono stati svolti

- Caratterizzazione delle strutture del pixel (diodo, elettronica di lettura, logica di controllo)
- Studio della corrente di leakage
- Studio del rumore della matrice e del Dynamic Range
- Prove funzionali della matrice (su illuminazione ambientale)

Sono di seguito riportati i risultati delle misure preliminari effettuate.

Elettronica di lettura

Per l'elettronica di lettura del sensore, formata dal SF del pixel in cascata ad un buffer analogico (amplificatore operazionale in configurazione a guadagno unitario) è stato effettuato un lavoro preliminare di caratterizzazione utilizzando una serie di strutture di test apposite.

Il buffer analogico è stato caratterizzato impiegando una struttura apposta, formata da tre di questi dispositivi che condividono l'ingresso (il piedino analogico BUFFER_IN) e le cui uscite sono tre piedini analogici (OUT_0_BUFFER, OUT_1_BUFFER, OUT_2_BUFFER). Per queste strutture è stata fornita una serie di tensioni analogiche DC in ingresso ed è stata osservata la tensione in uscita, in modo da creare una caratteristica di trasferimento.

Le prove hanno mostrato una buona linearità, con un guadagno pari a 1.02 nella zona lineare della caratteristica di trasferimento.

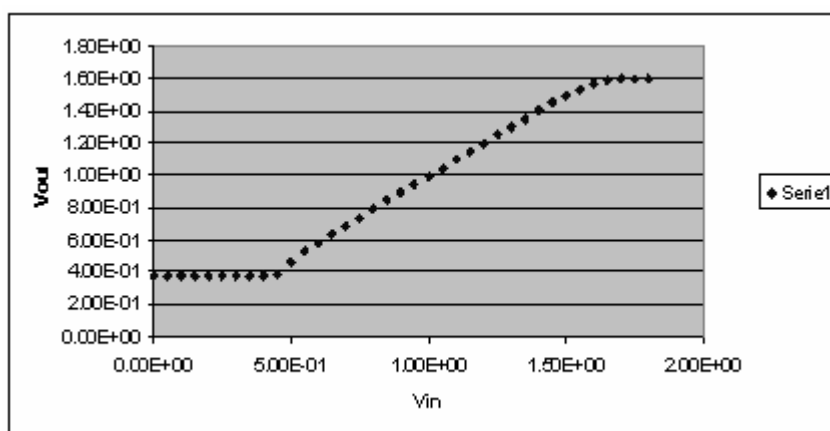


Figura 73 - Caratteristica ingresso-uscita del buffer

Più complesse le prove sul Source-Follower. Pur non essendo presente sul chip una struttura di test ad hoc per caratterizzare questo circuito, è stato comunque possibile effettuare una caratterizzazione impiegando la struttura Ultrapixel. Questa è formata dalla struttura Ultradiode a cui è collegato un pixel 3T simile a quello presente sulle matrici principali. Poiché la struttura

Ultradiodo è collegata direttamente ad un ingresso analogico, oltre che all'ingresso del source-follower, è possibile, variando questa tensione e osservando l'uscita di Ultrapixel, creare una funzione di trasferimento per la struttura, e osservare come questa tenda a variare per differenti tensioni di polarizzazione fornite in ingresso.

È stato osservato un comportamento lineare per ingressi che stiano entro il range di tensioni a cui ci si aspetta che il fotodiodo lavori ($V_{FD} < V_{dd} - V_{th}$, in questo caso pari a 1.4 circa), ma con un valore di saturazione inferiore in ingresso dipendente dalla polarizzazione. Maggiori tensioni di polarizzazione conducono ad un maggiore valore di saturazione minimo, erodendo l'escursione disponibile in uscita per la matrice. Nella zona lineare, invece, il valore del guadagno tende ad essere costante rispetto alla tensione di polarizzazione, con un valore di circa 0.83-0.85.

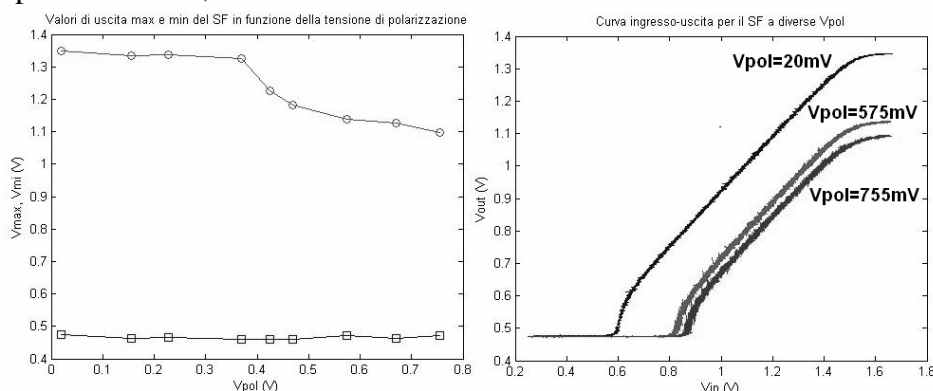


Figura 74 - Variazione di massimo e minimo valore in uscita del SF in funzione della tensione di polarizzazione (sx) e caratteristica d'ingresso-uscita del SF per tre tensioni di polarizzazione (20mV, 575mV, 755mV) (dx)

Tramite la struttura Ultradiodo è stato inoltre possibile effettuare una stima della capacità del fotodiodo 'small'. La struttura è composta da 8000 fotodiodi connessi in parallelo, in modo da rendere influente il disturbo causato dalla capacità della PAD di lettura e dalle incertezze sul sistema di acquisizione e sull'impulso di corrente. Si è ottenuto un valore medio di circa 3.5fF per le tensioni di polarizzazione inferiori a 1.4V (si ricordi che

su un pixel APS del RAPS non è possibile caricare il fotodiode a tensioni superiori a $V_{dd}-V_{th}$, oltre i quali il valore di capacità tende a scendere.

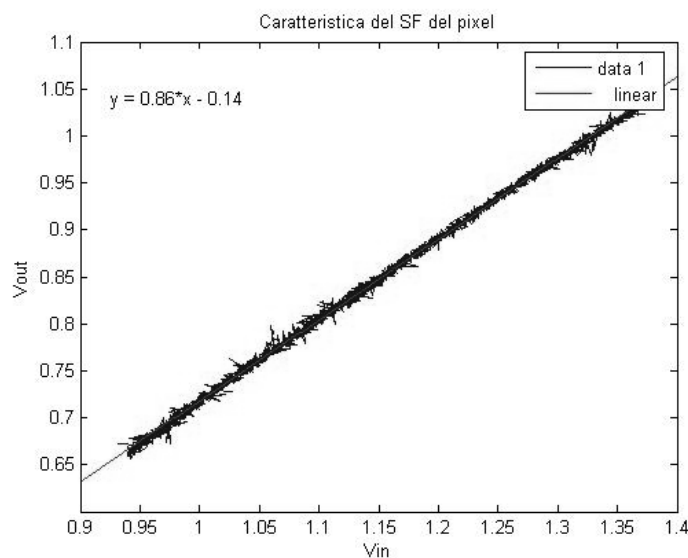


Figura 75 - Funzione di trasferimento del SF alle tensioni di interesse con la polarizzazione impiegata per le misura di leakage e di matrice

Anche la logica è stata sottoposta ad una prova di performance: poiché il sensore deve essere in grado di funzionare ad alte frequenze, è necessario che la logica supporti una frequenza di almeno 16MHz, necessaria per riuscire ad effettuare la lettura di un intero frame da 128x128 pixel in almeno 1ms. Il RAPS-03 è stato di conseguenza testato a frequenza di clock elevate, osservando via oscilloscopio se l'evoluzione dei segnali digitali logici seguisse il funzionamento previsto o se dimostrasse un comportamento errato. Le prove hanno dimostrato un funzionamento corretto della logica per frequenze fino a 120MHz, che dimostra come il RAPS-03 possa essere impiegato anche in applicazioni ad alta frequenza, con tempi di lettura di una matrice inferiori ai 150us.

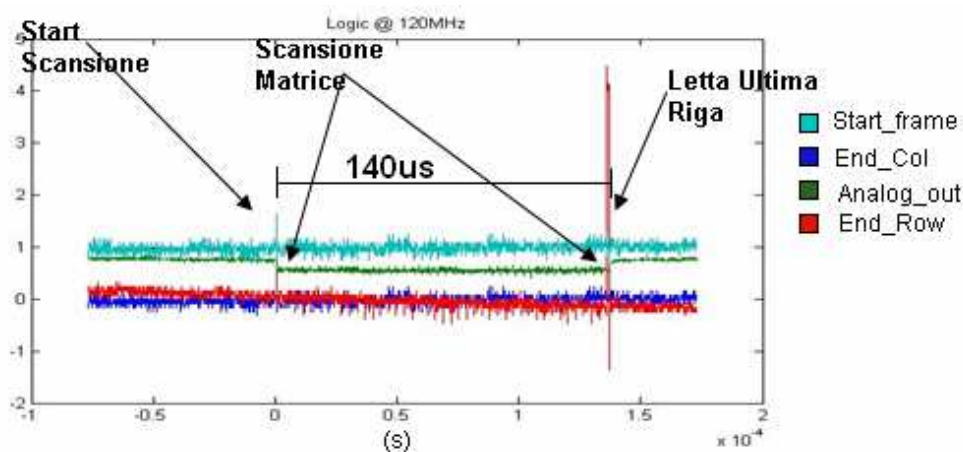


Figura 76 - Funzionamento osservato della matrice a 120MHz: la lettura di un'intera matrice 128x128 è effettuata in meno di 150us.

Misure di leakage

Negli esperimenti di fisica delle alte energie, il metodo di scansione impiegato è solitamente event-triggered, per cui la matrice viene tenuta al buio e letta solo quando un sensore esterno (tipicamente un calorimetro) rileva che c'è stato il passaggio di una o più particelle di cui si vuole conoscere il punto di passaggio. Poiché tale opzione prevede di tenere la matrice in attesa (stato d'integrazione) per un tempo indefinito, effettuando la lettura solo in seguito ad un evento esterno di trigger, è necessario effettuare una caratterizzazione del leakage a cui è sottoposta la matrice. Infatti poiché le correnti in inversa del fotodiode tendono a scaricarlo anche in assenza di luce, un tempo di integrazione eccessivamente lungo porterà ad una forte riduzione dell'escursione di tensione massima, fino al punto in cui la tensione del fotodiode scenda sotto il range di tensioni a cui il source-follower è in grado di effettuare la lettura (che nel caso del RAPS è un valore che può variare, come visto, tra 0.5 e quasi 1V). È pertanto necessario effettuare una stima della velocità di scaricamento del fotodiode,

in modo da definire una frequenza minima di funzionamento. In questo modo è possibile inviare un reset periodico alla matrice (o forzare una lettura periodica) prima che la caduta di tensione la renda insensibile ad un evento.

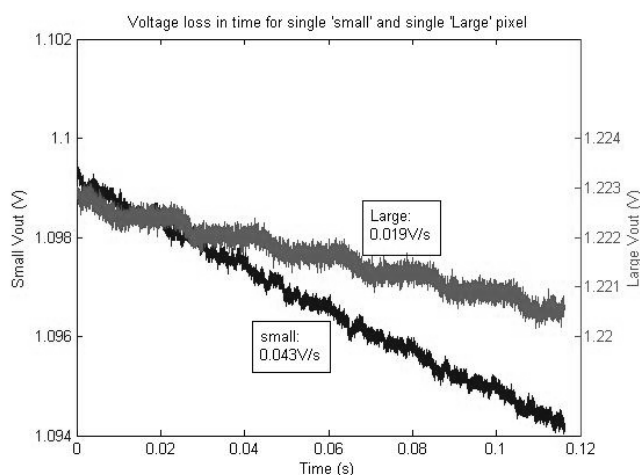


Figura 77 - Uscita delle strutture a singolo pixel small e Large: leakage

Queste misure sono state effettuate osservando l'evoluzione nel tempo dell'uscita delle strutture a singolo pixel in condizioni di buio e osservandone la pendenza. Sono state acquisite 100 forme d'onda (tensione in funzione del tempo) di cui si è fatta la media. La perdita di tensione rispetto al tempo è stata quindi calcolata come il coefficiente angolare della retta interpolante la forma d'onda ottenuta. È stato così osservata una diminuzione di tensione di circa 43mV/s per i pixel a fotodiode 'small' e di circa 19mV/s per i pixel a fotodiode 'Large'.

Noto il guadagno del source follower, 0.86, è possibile trovare la caduta di tensione al fotodiode, pari a 49mV/s, da cui viene stimata la corrente di leakage (sapendo che la capacità del diode small è pari a 3.5fF) come

$$I = C \frac{dV}{dT}$$

e che risulta uguale a 0.17fA.

Altre prove sono state inoltre effettuate su pixel arrangiati in matrici 3x3, in cui ogni pixel è dotato della propria elettronica di lettura, considerando la perdita di tensione rispetto al tempo. In questo caso è possibile notare un certo effetto di bordo, soprattutto nei pixel posti all'angolo della matrice che risultano avere perdite di maggiore entità rispetto ai pixel adiacenti.

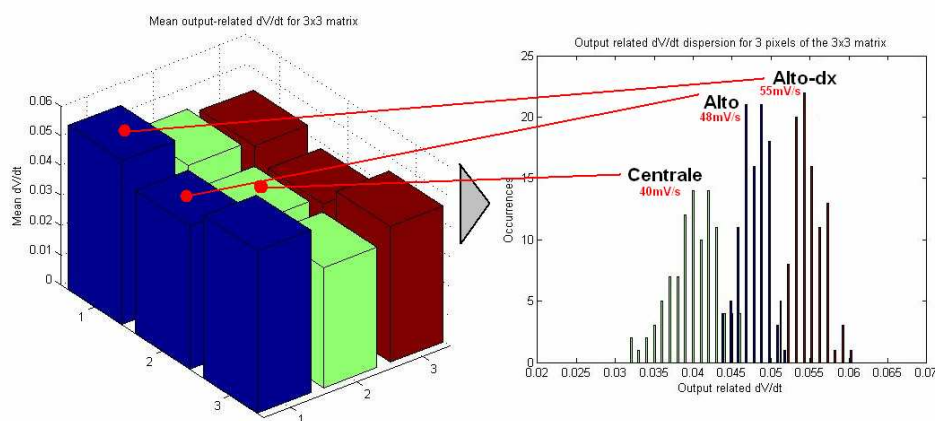


Figura 78 - Leakage misurato per la matrice 3x3 di pixel 'small' e dispersione delle pendenze misurate per 100 acquisizioni di 3 pixel campione

Misure di rumore delle matrici

La caratterizzazione delle matrici richiede la definizione dei rumori di tipo fixed pattern noise e termico. Il primo rappresenta la variazione fra pixel e pixel imputabile alle variazioni di processo (componenti nominalmente identici i cui parametri sono in realtà differenti) ed è tempo-invariante, mentre il secondo, di natura prevalentemente termica, è responsabile della variazione dei valori di un pixel sempre esposto alla stessa luminosità in letture successive. Mentre il fixed pattern noise è un rumore poco problematico (può essere eliminato facilmente mediante double sampling a livello software), il rumore termico risulta essere più pericoloso, in quanto un elevato valore può diminuire la sensibilità del sensore ad un evento che

generi piccole escursioni di segnale, come potrebbe essere il passaggio di una MIP.

Per il calcolo di questi fattori, il primo passo è la creazione di un piedistallo. Sono stati dunque acquisiti 300 campioni di una matrice non illuminata (dark frame). Per ogni pixel è stata quindi effettuata la media sui campioni letti, creando una nuova matrice delle medie. Poiché le condizioni di illuminazione sono uniformi (nessuna illuminazione) e il rumore termico è a media nulla, la matrice così creata è composta dai valori di buio dei pixel depurati dal contributo dei rumori istantanei, ed è detta piedistallo. Il FPN della matrice è di conseguenza definito come la deviazione standard dei valori dei pixel del piedistallo, in quanto questa è da attribuirsi unicamente alle differenze tempo-invarianti fra i pixel.

Per il calcolo del rumore termico, invece si ricalcola il valore di ogni pixel di ogni frame come la differenza fra il valore letto del pixel e il corrispondente pixel del piedistallo, ottenendo così 300 matrici di valori relativi. Poiché adesso il valore di ogni pixel dipende unicamente dal rumore istantaneo, e non più dal FPN, è possibile creare un istogramma dei valori comprendendo insieme tutti i pixel della matrice per avere una stima globale del rumore. La deviazione standard di tutti i valori così acquisiti (300 frames x 16384 pixel) è definita come il rumore termico o istantaneo.

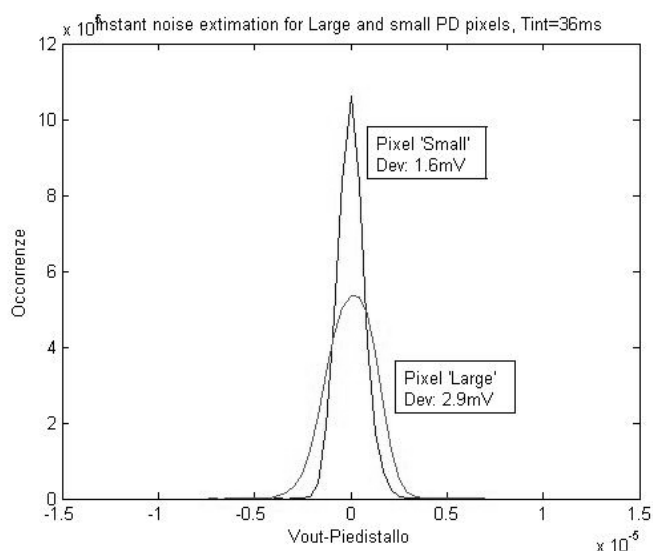


Figura 79 - Rumore istantaneo per pixel 'Large' e 'small'

Sono stati quindi fatti dei test per avere una stima dei contributi. Per le prime misure è stata impiegata una frequenza di clock di 7.1MHz e una tensione di polarizzazione analogica di 575mV, che è relativamente penalizzante dal punto di vista dell'escursione, ma che ha dimostrato una buona risposta ad alte frequenze. Il tempo d'integrazione è stato di 256k cicli di clock, ovvero 36.7ms mentre il periodo di reset è stato settato a 256 cicli di clock, pari a 36 μ s. I valori calcolati per i contributi di rumore sono i seguenti:

Matrice	Pixel 'Large'	Pixel 'Small'
Rum. Istantaneo	2.9mV	1.6mV
FPN totale	19.8mV	17.5mV
FPN Colonna	14.8mV	13.8mV
FPN Pixel	13.2mV	10.9mV

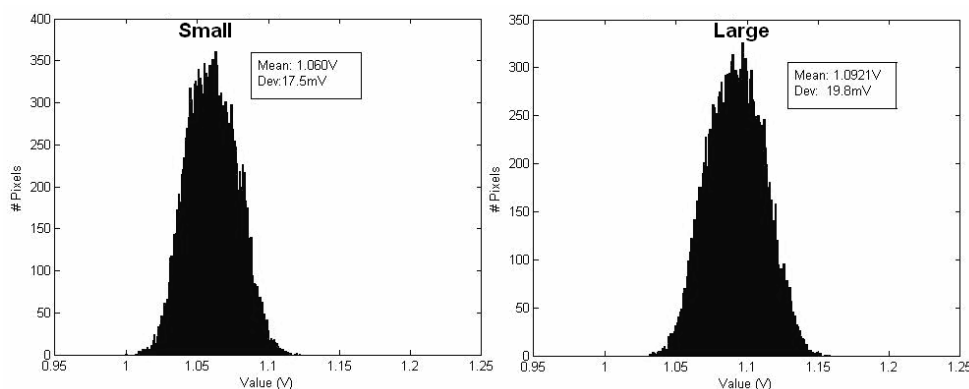


Figura 80 - Distribuzione dei valori dei pixel del piedistallo per la matrice a pixel 'small' e 'Large'

Come si può vedere, il rumore risulta più alto rispetto a quanto stimato per il RAPS-02, che aveva contributi dell'ordine di 1-1.5mV, ma bisogna tener conto di alcuni fattori:

- Il source follower, dovendo gestire frequenze più alte, ha una banda maggiore (minore filtraggio del rumore termico)
- La matrice e l'elettronica sono comunque più rumorose a causa delle maggiori dimensioni e maggiore frequenza
- Il tempo stesso d'integrazione, nonostante l'elevata frequenza è maggiore rispetto quello del RAPS-02, che è pari a 9.6ms, per cui è logico aspettarsi un aumento del contributo di shot noise
- Sono stati osservati dei piccoli contributi di cross-talk

Questo potrebbe spiegare anche il motivo per cui il pixel 'Large' mostra un valore di rumore istantaneo maggiore rispetto allo 'small', mentre sarebbe logico aspettarsi il contrario, essendo il rumore kTC in tensione inversamente proporzionale alla capacità del fotodiodo. La presenza di crosstalk porterebbe il kTC ad essere meno importante rispetto ad altri contributi.

Il maggiore FPN del Large potrebbe essere dovuto alla forma irregolare del fotodiodo, che lo porta ad avere maggior variazione dei parametri, mentre il contributo stimato del FPN di colonna è circa uguale per ambedue le matrici, come logico aspettarsi.

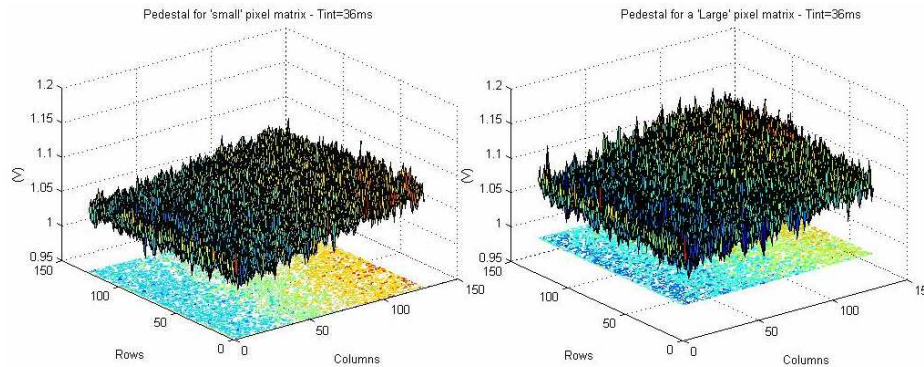


Figura 81 - Immagine dei piedistalli ($T_{int}=36ms$) per matrice a pixel 'small' (s_x) e pixel 'Large' (d_x)

Il rumore è stato anche stimato modificando le condizioni di funzionamento del sensore, in particolare cambiando il tempo d'integrazione. Se dal punto di vista del rumore istantaneo non è stata fondamentale vista una dipendenza dal tempo d'integrazione apprezzabile, il FPN ha invece dimostrato una certa dipendenza. In particolare è stato misurato un aumento del FPN misurato in funzione del tempo d'integrazione, evidentemente dovuto al contributo del FPN nella corrente di leakage, che viene amplificato da maggiori tempi d'integrazione e tende a diventare dominante rispetto al FPN indotto dalla dispersione delle tensioni di soglia e dei parametri del SF.

Calcolo del Dynamic Range

Un altro parametro importante nella valutazione del sensore è il dynamic range che questo riesce ad ottenere. Come si è visto, la scelta della tensione di polarizzazione del SF non cambia il guadagno della catena di lettura, ma ne modifica l'escursione di segnale ottenibile. Di conseguenza il Dynamic Range ottenibile, espresso come

$$DR = 20 \log_{10} \left(\frac{S_M - S_m}{N} \right)$$

Dove S_M e S_m sono il segnale massimo e minimo rispettivamente e N è il rumore. Il valore di buio è stato definito valore massimo, calcolando un piedistallo da 100 frame bui. Allo stesso modo è stato calcolato un piedistallo impiegando 100 frame fortemente illuminati in modo da ottenere la saturazione per trovare i valori minimi. I valori S_M e S_m sono stati quindi calcolati come le medie dei rispettivi piedistalli. Il rumore è invece quello calcolato al punto precedente.

Matrice	Pixel 'small'	Pixel 'Large'
S_M	1.068V	1.120V
S_m	0.654V	0.624
Rumore	1.6mV	2.9mV
SNR	258.75	171.04
DR	48.3dB	44.7dB

Le prove effettuate, per il valore di V_{pol} ottimale trovato (575mV) hanno dato un valore di DR di poco maggiore ai 45dB. Questo valore è piuttosto basso, ricordando che un APS normale ha un DR di circa 55-60dB, ottenibili con 1V di swing e 1mV di rumore, ma bisogna tener conto sia dell'impiego di una polarizzazione che favorisce la risposta dinamica all'escursione per il source-follower, sia della tecnologia impiegata. Infatti con una tecnologia in 0.18 μ m è difficile ottenere elevate escursioni di tensione e allo stesso modo il rumore relativamente alto che affligge le matrici del RAPS peggiora leggermente la situazione.

Problemi osservati

Le prove effettuate hanno portato ad osservare alcuni comportamenti inaspettati del sensore. Il malfunzionamento più eclatante è presente nelle matrici 128x128 destre, in cui è possibile notare un angolo molto più chiaro del resto della matrice in condizioni di buio. Questo è presente sia nella matrice 'Large' che 'small' di destra, e corrisponde ai due angoli adiacenti destri, il che fa supporre l'esistenza di un disturbo centrato all'incirca fra

queste due matrici, ed è stato osservato in tutti i sensori realizzati. Una spiegazione per questo fenomeno è la presenza di un dispositivo all'esterno della matrice che a causa di un cattivo isolamento inietta elettroni nel substrato (o che comunque causa una variazione di tensione locale del substrato) aumentando la carica raccolta dai pixel presenti entro un certo raggio. Alcune prove sono state effettuate per meglio comprendere questo comportamento:

- variando la frequenza di funzionamento mantenendo però costante il tempo d'integrazione: i pixel luminosi hanno mostrato che la loro escursione aumenta con la frequenza
- variando il tempo d'integrazione: i pixel luminosi hanno mostrato che la loro escursione aumenta con il tempo d'integrazione

Questo fa sospettare che si tratti di un disturbo in corrente che viene integrato nel tempo e non di un problema di reset (il disturbo in tal caso originerebbe valori costanti a meno del leakage).

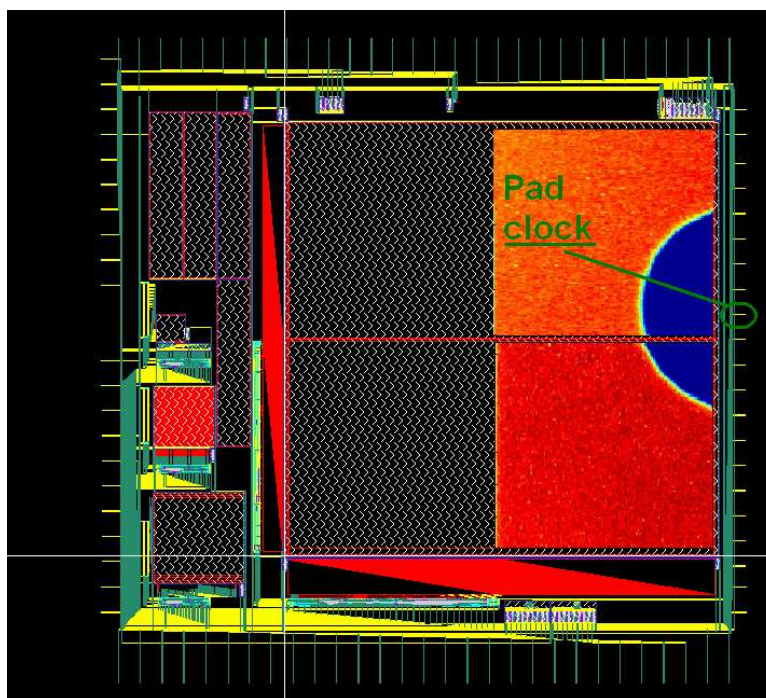


Figura 82 - Posizione della zona colpita nel layout. Si può notare come il centro di questa zona corrisponda all'incirca al punto dove è presente il PAD di clock.

L'angolo colpito si trova in una zona della matrice lontana dall'elettronica di supporto o di test, ma vicina alla sezione digitale dell'anello di I/O, i cui PAD sono dotati di buffer digitale interno. In particolare si è osservato che il PAD di clock (il segnale d'ingresso a maggior frequenza) è ubicato proprio dove dovrebbe essere il centro del disturbo, per cui è possibile che sia proprio questo a generare un disturbo di carica che influenza i pixel vicini.

Un secondo problema riscontrato è la presenza di alcuni cross-talk, che non è ancora chiaro se imputabili al chip o alla scheda sperimentale. Questi si traducono in alcuni rumori periodici, uno a bassa ed uno ad alta frequenza.

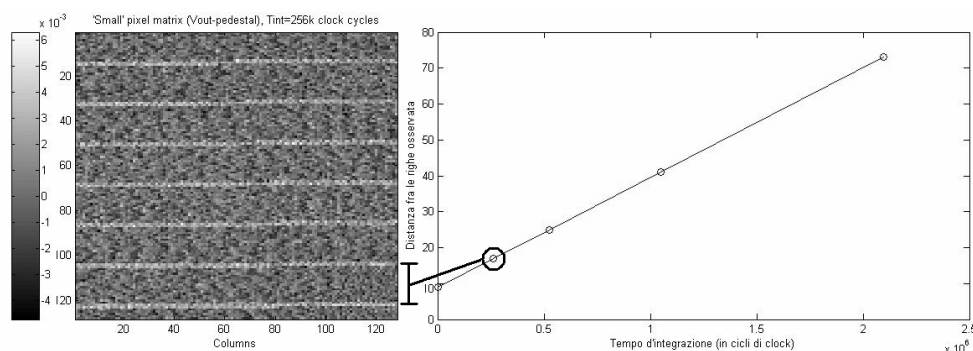


Figura 83 - Cross-talk ad alta frequenza osservato su una matrice (differenza fra immagine e piedistallo, immagine in negativo) e distanza fra le righe scure in funzione del tempo d'integrazione selezionato per la matrice 128x128 (che è direttamente proporzionale a sua volta al tempo d'integrazione delle 32x32)

Quello ad alta frequenza mostra un pattern che si ripete ogni N righe lette, dove N è un numero di cui si è osservata una dipendenza lineare con il valore di integrazione selezionato e che non è sincronizzato con la matrice. Questo disturbo si manifesta con 256 pixel consecutivi a valore più alto degli altri (circa 5mV), seguiti da una lettura a valori normali di M pixel, dove M risulta pari alla somma di tempo d'integrazione e lettura della matrice 32x32. È di conseguenza evidente, essendo tutte le misure state fatte con tempo di reset pari a 256 cicli di clock, che questo rumore risulti nascere da un cross-talk con il segnale di reset della matrice 32x32.

Il cross-talk a bassa frequenza è invece osservabile come un rumore che causa periodicamente delle perturbazioni sul valore medio dei frame, affliggendone tutti i pixel. Questo varia molto con il periodo d'integrazione, e si manifesta solitamente con la lettura, ad intervalli fissi, di un frame identico ai frames normali, ma differente per un offset fisso.

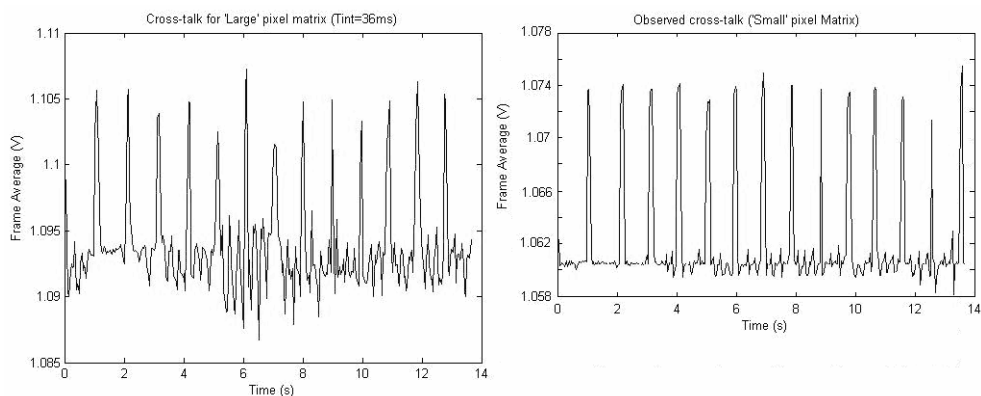


Figura 84 - Evoluzione della media del frame nel tempo per 'Large' (sx) e 'small' (dx). Si può notare la frequenza del disturbo (circa 1Hz) e l'escursione che esso genera (circa 10-15mV)

Prove di imaging

Per provare il funzionamento del sensore, sono state inoltre effettuate prove di imaging. Il sensore è stato sottoposto ad illuminazione ambientale ed è stato quindi coperto parzialmente con piccoli pezzi di materiale, come capelli o quadratini di carta. Il tempo d'integrazione è stato settato in modo da aver il miglior contrasto fra le zone buie e illuminate e contemporaneamente da rendere trascurabile l'effetto derivato dal differente tempo d'integrazione tra i primi pixel letti e gli ultimi. Una durata inferiore a 1ms è stata necessaria per evitare la saturazione delle ultime righe del sensore.

Il sensore ha mostrato piena funzionalità in questa modalità di funzionamento, con un buon contrasto fra le zone illuminate e quelle scure e una buona risoluzione. È stato anche effettuato qualche tentativo di post-processing, per valutare il miglioramento ottenibile con tecniche di double-sampling: acquisendo una matrice di piedistallo come media di cento acquisizioni della matrice illuminata a luce ambientale, e sottraendola all'immagine ottenuta con la stessa illuminazione e l'oggetto posto sopra di

essa, è possibile aumentare il contrasto e stabilizzare meglio l'immagine, anche per tempi d'integrazione di lunghezza paragonabile al tempo di lettura (Figura 86). In questo caso il mismatch fra i pixel derivante dall'aver integrato la luce per un tempo diverso viene trattato come un FPN, e dà luogo ad una migliore immagine in quanto durante i test la luminosità ambientale non è stata soggetta a variazioni.

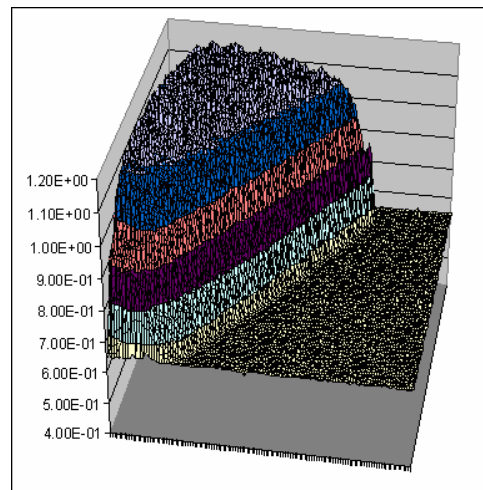


Figura 85 - Prova con pezzo rettangolare di carta ($T_{int}=1.22ms$)

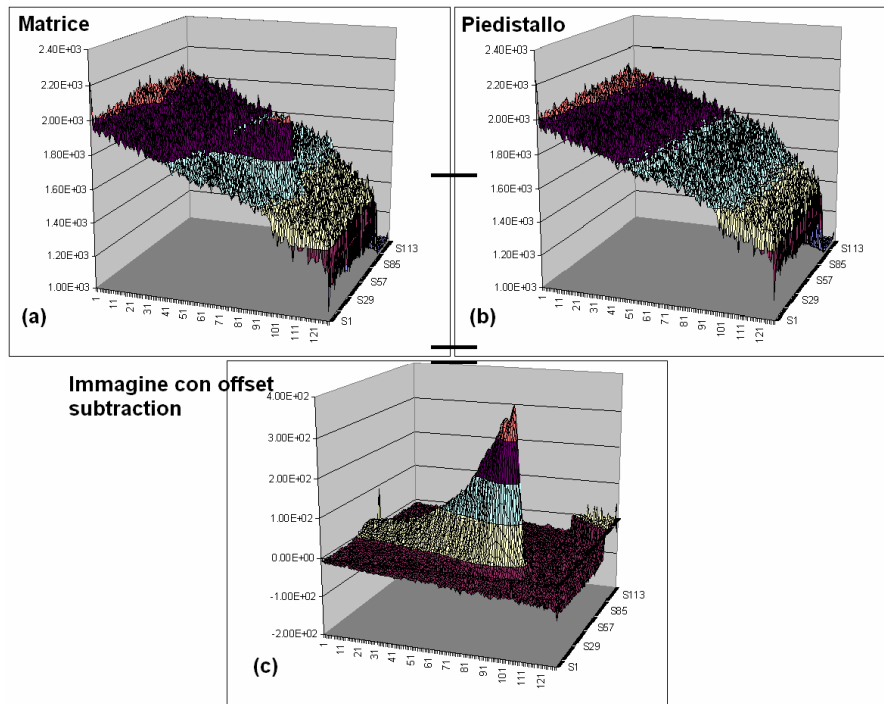


Figura 86 – Matrice con capello. Dalla matrice risultante viene sottratto l'offset, in questo caso la media di 100 matrici illuminate: il capello risulta chiaramente visibile, con una zona maggiormente buia nel punto del bulbo (Tint=573us)

CONCLUSIONI E SVILUPPI FUTURI

Conclusionone

Il sensore RAPS-03, terzo esemplare della linea RAPS è stato seguito durante l'intero iter progettuale:

- Definizione delle specifiche
- Progettazione
- Definizione e realizzazione delle strutture di test
- Test preliminari di caratterizzazione elettrica.

Scopo del lavoro è la creazione e piena caratterizzazione di un sensore per radiazioni minimamente ionizzanti (principalmente raggi X e particelle) orientato sia a impieghi in esperimenti HEP (high energy physics) che per alcune applicazioni medicali, e che sfrutti le caratteristiche della tecnologia CMOS come flessibilità d'impiego, primo processing on-chip dei dati e bassi costi.

Il sensore è stato creato con maggiore enfasi sugli aspetti che consentono di superare le limitazione del chip precedente, il RAPS-02. In particolare, è stata introdotta la possibilità di effettuare una lettura del chip solo in seguito ad un evento e non per scansione continua, la possibilità di selezionare la durata delle fasi di reset ed integrazione e un'elettronica di lettura configurabile con maggiore facilità. Il chip prevede inoltre una matrice di dimensioni paragonabili a quelle di un chip operativo (256x256 pixel di 10um di pitch) e la possibilità di funzionare a elevate frequenze (~10MHz). Per aumentare il frame-rate, la matrice è divisa in 4 sottomatrici 128x128 operanti in parallelo. Queste matrici differiscono fra di loro in quanto due sperimentano un pixel dotato di fotodiodo di 2x2um (pixel 'small'), mentre le altre due impiegano un pixel il cui fotodiodo occupa tutta l'area libera da elettronica (pixel Large).

Il sensore è stato realizzato in una tecnologia CMOS standard, la UMC 0.18 e prevede assieme alla matrice principale una serie di sottomatrici e di

singoli pixel per studiare possibilità alternative (enclosed-gate, 4T, pixel a più fotodiodi). Sono state inoltre inserite strutture che aiutino nella caratterizzazione elettrica del sensore, permettendo di valutare le risposte della catena di lettura e di caratterizzare la capacità dei fotodiodi.

Per il test del chip è stato deciso l'impiego di soluzioni flessibili, che permettano un'agevole modifica dei settaggi in modo da testare il sensore in condizioni di funzionamento anche fortemente differenti fra loro. Questo è possibile mediante l'impiego di logiche programmabili (FPGA) e programmi per PC di acquisizione sia nella generazione dei segnali di controllo del chip e dell'elettronica di scheda che nella gestione dei dati provenienti dal chip.

Le prime prove hanno dimostrato la piena funzionalità del sensore, nonostante alcuni fattori che possono ridurre le prestazioni, come il rumore istantaneo leggermente più elevato rispetto al sensore precedente (RAPS-02), nell'ordine di 2.1mV per le matrici a fotodiodo small. Anche le nuove caratteristiche inserite nella logica hanno dimostrato un funzionamento corretto per frequenze fino a 120MHz.

Verranno presto eseguiti test operativi, per valutare la risposta del sensore in termini di segnale e sensibilità a vari tipi di radiazione, similmente ai test descritti sul RAPS-02:

- Laser (visibile, IR, UV)
- Particelle di varia natura (fotoni X, particelle α e β , etc...)

In questo modo si potrà ottenere anche una caratterizzazione funzionale del sensore, che permetterà di definire al meglio specifiche, ambiti e caratteristiche dei modelli successivi. Anche un impiego diretto del sensore RAPS-03 è ipotizzabile per alcuni ambiti in cui sia sufficiente una matrice 128x256 (circa 1.2x2.5mm di dimensioni).

Sviluppi futuri

Grazie alle caratteristiche di configurabilità, flessibilità ed elevata risoluzione e frequenza del sensore, sono allo studio applicazioni dei chip RAPS-02 e RAPS-03 anche in ambiti diversi dalla fisica delle alte energie.

Una possibile applicazione per esperimenti di fisica è l'impiego del sensore per la spettrografia X di diffrazione. Se un fascio X di cui si vuole sapere lo spettro viene fatto incidere su un reticolo cristallino, questo subirà diffrazione di De Bragg, cosicché le componenti a diversa lunghezza d'onda saranno deviate di angoli differenti. Mediante l'impiego di un imager sarebbe dunque possibile vedere una serie di righe la cui posizione spaziale corrisponderebbe ad una frequenza presente nel fascio. Si potrà impiegare il RAPS per effettuare misure in questo senso, osservando se l'impiego di un sensore nudo (meno fotoni rilevati, ma maggiore precisione spaziale) possa fornire risultati migliori rispetto un sensore dotato di strato scintillatore (migliore rivelazione di fotoni, ma peggior risoluzione) in un ambito in cui la risoluzione risulta più importante della minimizzazione della dose di raggi X impiegata.

Anche la buona risoluzione del pixel ($10 \times 10 \mu\text{m}$), assieme all'elevata frequenza e alla possibilità di selezionare tempo d'integrazione può essere sfruttata, addirittura per applicazioni nel visibile, come ambiti astronomici, in cui il sensore viene illuminato con la luce proveniente da una stella con tempi di esposizione differenti in modo da osservare eventuali variazioni di luminosità e diametro del disco.

Inoltre anche il setup sperimentale potrà essere migliorato, con l'eventuale progettazione di una seconda scheda che permetta il pile-up con controllo e contemporanea lettura di più sensori. In questo modo sarà possibile valutare le prestazioni del RAPS-03 anche come tracker, osservando il passaggio di una particella in più punti e potendone così ricostruire la traiettoria. Un secondo miglioramento alla scheda di acquisizione potrebbe essere la possibilità di effettuare conversione A/D in parallelo e per tutte e 4 le matrici, in modo da poter impiegare nei test l'intera matrice 256×256 come area sensibile. Ambedue questi miglioramenti, a causa dell'aumentato throughput, che risulterebbe almeno quadruplicato, richiederebbero modifiche sia ai programmi VHDL delle schede che ai programmi PC.

BIBLIOGRAFIA

- [1].R. Turchetta, J.D. Berst, B. Casadei, G. Claus, C. Colledani, W. Dulinski, Y. Hu, D. Husson, J.P. Le Normand, J.L. Riester, G. Deptuch, U. Goerlach, S. Higuere, M. Winter - A Monolithic Active Pixel Sensor for Charged Particle Tracking and Imaging Using Standard VLSI CMOS Technology - Nuclear Instruments and Methods in Physics Research A 458, 2001 -677-689
- [2].E.R. Fossum – CMOS Image Sensors: Electronic Camera-On-A-Chip - IEEE Transactions on Electron Devices, Vol. 44, No. 10, October 1997 - 1689-1698
- [3].S. Heini, C. Hu-Guo, M. Winter, Y. Hu - A New PhotoFET for Monolithic Active Pixel Sensors Using CMOS Submicronic Technology - Electronics, Circuits and Systems, 2006. ICECS '06. 13th IEEE International Conference on 10-13 Dec. 2006 - 1148-1151
- [4].W. Dulinski, D. Berst, A. Besson, G. Claus, C. Colledani, G. Deptuch, M. Deveaux, A. Gay, D. Grandjean, Y. Gornushkin, A. Himmi, C.Hu, J.L. Riester, I. Valin, M. Winter – CMOS Monolithic Active Pixel Sensor for Minimum Ionising Particle Tracking Using Non-Epitaxial Silicon Substrate - IEEE Transactions on Nuclear Science, Volume 51, Issue 4, Part 1, Aug. 2004 - 1613-1617
- [5].S. Kleinfelder, S. Lim, X. Liu, A. El Gamal – A 10000 Frames/s CMOS Digital Pixel Sensor - IEEE Journal OF Solid-State Circuits, Vol. 36, No. 12, Dec 2001 – 2049-2059
- [6].C. Xu, W. Zhang, W.-H. Ki, M. Chan - A 1.0-V VDD CMOS Active-Pixel Sensor With Complementary Pixel Architecture and Pulsewidth Modulation Fabricated With a 0.25-um CMOS Process - IEEE Journal of Solid-State Circuits, Vol. 37, No. 12, Dec 2002 – 398-400

- [7]. H. Tian, B. Fowler, A. El Gamal – Analysis of Temporal Noise in CMOS Photodiode Active Pixel Sensor - IEEE Journal of Solid-State Circuits, Vol. 36, No. 1, Jan 2001 – 92-101
- [8]. G. Deptuch, J.D. Berst, G. Claus, C. Colledani, W. Dulinski, Y. Gornushkin, D. Husson, J.L. Riester, M. Winter - Design and Testing of Monolithic Active Pixel Sensors for Charged Particle Tracking - IEEE Transactions on Nuclear Science, Vol. 49, No. 2, Apr 2002 – 601-610
- [9]. B. Mheen, M. Kim, Y.-J. Song, S. Hong - Operation Principles of 0.18- μm Four-Transistor CMOS Image Pixels With a Nonfully Depleted Pinned Photodiode - IEEE Transactions on Electron Devices, Vol. 53, No. 11, Nov 2006 - 2735-2740
- [10]. 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 Electron Devices, Vol. 47, No. 11, Nov 2000 – 2110-2122
- [11]. S. Kleinfelder, F. Bieser, Y. Chen, R. Gareus, H.S. Matis, M. Oldenburg, F. Retiere, H.G. Ritter, H.H. Wieman, E. Yamamoto. - Novel Integrated CMOS Sensor Circuits - IEEE Transactions on Nuclear Science, Vol. 51, No. 5, Oct 2004 – 2328-2336
- [12]. G. Rizzo - Recent development on CMOS monolithic active pixel sensors - Nuclear Instruments and Methods in Physics Research A 576 (2007) -103–108
- [13]. B. Pain, Member, G. Yang, T.J. Cunningham, C. Wrigley, B. Hancock - An Enhanced-Performance CMOS Imager With a Flushed-Reset Photodiode Pixel - IEEE Transactions on Electron Devices, Vol. 50, No. 1, Jan 2003 – 48-56
- [14]. R. Turchetta – Sensori Monolitici a Pixels Attivi – Scuola Nazionale "Rivelatori ed Elettronica per Fisica delle Alte Energie, Astrofisica e Applicazioni Spaziali" - 26-30 Marzo 2007, Legnaro, Italia
- [15]. T. Greig, C. Castelli, A. Holland, D. Burt - The Design of an Active Pixel Sensor Test Structure Optimised for the Read Out of

- Scintillator Screens - Nuclear Instruments and Methods in Physics Research A 573 (2007) - 30–33
- [16]. J. Wyss - Introduction to radiation damage: concepts, physical quantities, radiation environments - Scuola Nazionale "Rivelatori ed Elettronica per Fisica delle Alte Energie, Astrofisica e Applicazioni Spaziali" - 26-30 Marzo 2007, Legnaro, Italia
- [17]. R. Turchetta - CMOS monolithic active pixel sensors (MAPS) for scientific applications: Some notes about radiation hardness - Nuclear Instruments and Methods in Physics Research A 583 (2007) - 131–133
- [18]. F. Faccio – Metodologie di disegno di ASIC resistenti alle radiazioni - Scuola Nazionale "Rivelatori ed Elettronica per Fisica delle Alte Energie, Astrofisica ed Applicazioni Spaziali" - 4-8 Aprile 2005, Legnaro, Italia
- [19]. M. Bruzzi - Danno da Radiazione in Rivelatori al Silicio - Scuola Nazionale "Rivelatori ed Elettronica per Fisica delle Alte Energie, astrofisica ed Applicazioni Spaziali" 4-8 Aprile 2005, Legnaro, Italia
- [20]. V. Re - Effetti delle radiazioni su dispositivi elettronici in tecnologie CMOS, JFET, bipolari - Scuola Nazionale "Rivelatori ed Elettronica per Fisica delle Alte Energie, astrofisica ed Applicazioni Spaziali" 4-8 Aprile 2005, Legnaro, Italia
- [21]. C. Redmond - Winning the Battle Against Latch-up in CMOS Analog Switches - Analog Dialogue 35-05 (2001)
- [22]. E. Gatti, P. Rehak - Review of Semiconductor Drift Detectors - Nuclear Instruments and Methods in Physics Research A 541 (2005) 47–60
- [23]. M. Artuso - Silicon Sensors Implemented on p-type Substrates for High Radiation Resistance Application - Nuclear Instruments and Methods in Physics Research A 582 (2007) - 835–838
- [24]. N. Wermes, L. Andricek, P. Fischer, K. Heinzinger, S. Herrmann, M. Karagounis, R. Kohrs, H. Krüger, G. Lutz, P. Lechner, I. Peric, M. Porro, R. H. Richter, G. Schaller, M. Schnecke-Radau, F. Schopper,

- H. Soltau, L. Strüder, M. Trimpl, J. Ulrici, J. Treis - New Results on DEPFET Pixel Detectors for Radiation Imaging and High Energy Particle Detection - IEEE Transactions on Nuclear Science, Vol. 51, No. 3, Jun 2004 – 1121-1128
- [25]. Yu. Gornushkin, M. Deveaux, A. Gay, A. Himmi, Ch. Hu, I. Valin, M. Winter, C. Colledani, G. Claus, G. Deptuch, W. Dulinski - Tracking Performance and Radiation Tolerance of Monolithic Active Pixel Sensors - Nuclear Instruments and Methods in Physics Research A 513 (2003) 291–295
- [26]. G. Gaycken, A. Besson, A. Gay, Yu. Gornushkin, D. Grandjean, F. Guilloux, S. Heini, A. Himmi, Ch. Hu, K. Jaaskelainen, M. Pellicioli, I. Valin, M. Winter, G. Claus, C. Colledani, G. Deptuch, W. Dulinski, M. Szelezniak. - Monolithic Active Pixel Sensors for Fast and High Resolution Vertex Detectors - Nuclear Instruments and Methods in Physics Research A 560 (2006) 44–48
- [27]. A. Giraldo, A. Paccagnella, A. Minzoni - Aspect Ratio Calculation in N-Channel MOSFETs with a Gate-Enclosed Layout - Solid-State Electronics 44 (2000) 981-989
- [28]. M. Bigas, E. Cabruja, J. Forest, J. Salvi - Review of CMOS Image Sensors - Microelectronics Journal 37 (2006) 433–451
- [29]. www.xilinx.com
- [30]. www.opalkelly.com
- [31]. www.wikipedia.org
- [32]. www.analog.com
- [33]. AD9238 – Data sheet
- [34]. AD7274 – Data sheet
- [35]. Spartan-3 Starter Kit Board User Guide
- [36]. Calzolari, Graffi, "Elementi di elettronica", Zanichelli 1984

SOMMARIO

INTRODUZIONE.....	5
La Tecnologia CMOS nella rivelazione di radiazioni.....	6
Funzionamento fisico della tecnologia CMOS	7
Il pixel PPS.....	11
Il pixel APS	12
La matrice APS e l'elettronica di lettura.....	14
Problematiche relative al reset	18
Altre componenti di rumore	22
Metodologie di riduzione del rumore.....	27
Effetti dello scaling sui dispositivi APS.....	29
Approcci alternativi.....	30
Danno da radiazione.....	41
Danno da esposizione.....	41
Danno da singolo evento	48
Rivelazione di raggi X	52
Acceleratori di particelle	56
Vertex detector	60
Microstrip.....	62
Silicon drift detectors	63
CCD	65
DEPFET	67
Problematiche per un vertex detector CMOS	69
CAPITOLO 2: IL PROGETTO RAPS	75
Introduzione	75
Scelta della tecnologia.....	75
Opzioni tecnologiche dei pixel APS	78
Pixel WIPS	79
RAPS-01	81
RAPS-02	84
Pixel SHARPS	85
Setup sperimentale	87

Caratterizzazione elettrica.....	88
Caratterizzazione funzionale.....	91
Raggi X.....	92
Sostanze radioattive.....	94
Prove laser.....	97
CAPITOLO 3: IL RAPS03.....	101
Limiti del chip RAPS-02.....	101
Specifiche del chip RAPS-03.....	101
La matrice principale del RAPS-03.....	104
Strutture di test del chip RAPS03.....	106
Logica del chip.....	110
Test.....	112
Setup sperimentale: hardware.....	113
Setup sperimentale: firmware.....	116
Setup sperimentale: software.....	121
Misure sul sensore.....	127
Elettronica di lettura.....	128
Misure di leakage.....	131
Misure di rumore delle matrici.....	133
Calcolo del Dynamic Range.....	137
Problemi osservati.....	138
Prove di imaging.....	142
CONCLUSIONI E SVILUPPI FUTURI.....	147
Conclusione.....	147
Sviluppi futuri.....	148
BIBLIOGRAFIA.....	151
SOMMARIO.....	155
RINGRAZIAMENTI.....	157

RINGRAZIAMENTI

Finita la tesi di dottorato, è doveroso ringraziare innanzitutto la mia famiglia, mia mamma, mio papà e mio fratello Marco per il supporto e l'affetto che non mi hanno mai fatto mancare. Ringrazio anche le mie care nonne Lina e Rosa, così come i miei nonni Mario e Vincenzo che purtroppo non ci sono più, zii, zie e cugini e tutti i miei parenti che mi hanno sempre confortato e aiutato.

Un sentito ringraziamento anche al mio Tutor Prof. Paolo Ciampolini e ai miei colleghi ed amici del laboratorio, per essermi sempre stati d'aiuto durante il periodo di tesi. Grazie ad Alessandro Marras, Andrea Ricci, Guido Matrella, Ferdinando Grossi, Matteo Grisanti e Valentina Bianchi. Un sentito ringraziamento anche a tutti gli amici che ho conosciuto durante questi tre anni di Dottorato in Palazzina 4: Giovanna, Matteo, Eric, Pietro, Andrea, Lorenzo, Francesco. Senza dimenticare gli amici e colleghi dei laboratori di Automazione e Elettronica Analogica: in particolare ringrazio Antonio Affanni, Andrea Toscani, Cristiano Azzolini, Alessio Facen, Silvia Dondi, Alessandro Carboni.

Ringrazio inoltre tutti i professori, i maestri che ho avuto in tutto il mio periodo di studio, dall'asilo all'università e i don del mio paese per avermi sempre aiutato a crescere culturalmente e spiritualmente.

Ringrazio altresì tutti gli amici della mia compagnia (in particolare Mirco, Giuliano, Enrico, Matteo B., Nicola, Evandro, Claudio, Marco, Gianluca, Mauro, Beatrice, Erika, Linda, Matteo S., Elena, Vera, Sabrina),

dell'università (in particolare i due Giovanni, Claudio, Angelo, Rosario, Paolo, Devis, Marco, Andrea, Lorenzo, Luca, Luisa, Alessandro, Manuela, Matteo, Matteo detto Carletto, Stefano, Gabriele, Francesco, Landry, Francesca), di Parma e provincia (Giacomo, Angelo, Sergio, Andrea, Pier Paolo, Cristina, Laura) dei viaggi diocesani (in particolare Miriam, Federica, Cristina, Claudia, Elena, Lorenzo, Valeria, Ilaria, Matteo, Maria Chiara, Carlo, Paolo, Cecilia, Sara, Antonella, Stefano, Fabrizio, Lucia, Angelo, Michela, Sara, Monica, Gloria) della Palazzina 4 e delle altre palazzine.

Centinaia i motivi per ringraziare tutti, dall'aiuto nelle più disparate difficoltà alle ore di divertimento insieme, una lista ancor più estesa del numero delle persone da ringraziare!

Spero di non aver dimenticato nessuno, in caso contrario sono pronto a far penitenza. Ringrazio ancora tutti per tutto!

Paolo