UNIVERSITÀ DEGLI STUDI DI CATANIA
FACOLTÀ DI INGEGNERIA
Corso di Laurea in Ingegneria Elettronica

Gaetano L'Episcopo

Caratterizzazione elettrica di una matrice di rivelatori di singolo fotone

TESI DI LAUREA

Relatore:
Chiar.mo Prof. Gaetano Palumbo

Correlatori:
Dott. Delfo Sanfilippo
Dott. Paolo Finocchiaro

Anno Accademico 2005 - 2006

A coloro che hanno un obiettivo da raggiungere

Niente è impossibile...

SOMMARIO

CAPITOLO 3

CARATTERIZZAZIONE DELLA CARICA E DEL GUADAGNO DI UN DISPOSITIVO SINGOLO E DI UNA MATRICE PLANARE

INTRODUZIONE

Negli ultimi anni, il campo della fotorivelazione ha conseguito notevoli progressi, dovuti alla crescente richiesta, in diversi settori scientifico-tecnologici, di dispositivi in grado di fornire pregevoli prestazioni nella rivelazione di fotoni e di segnali luminosi molto deboli dall'infrarosso all'ultravioletto. Il rapido avanzamento nella ricerca e nello sviluppo di nuovi sensori ottici con prestazioni sempre crescenti, in termini di sensibilità, precisione temporale ed affidabilità, trova un'importante posizione all'interno di un nuovo settore scientifico e tecnologico noto come **Fotonica**.

Nella misura di segnali luminosi molto deboli, il limite sulla sensibilità è dato dalla rivelazione dei singoli quanti della radiazione ottica, cioè **i singoli fotoni**. Data la piccolissima energia associata ad un singolo fotone, è, dunque, evidente la difficoltà presente nella misura di questi segnali luminosi. La domanda legittima, allora, è se la rivelazione di singoli fotoni abbia un interesse applicativo, oppure sia solamente un "record" da realizzare.

Risulta, quindi, opportuno citare alcune applicazioni di questi dispositivi.

Le tecniche basate sul solo "conteggio dei fotoni" (**Photon Counting - PC**) e sul rilevamento dei "tempi d'arrivo dei fotoni" (**Time Correlated Photon Counting - TCPC**) trovano ampia applicazione in molteplici campi della scienza e della tecnica: nella caratterizzazione di diodi laser e fibre ottiche; nelle misure di emissioni e decadimenti fluorescenti in medicina, chimica, scienza dei materiali e biologia; nel laser-ranging di satelliti; nell'astronomia, in particolare nei telescopi ad ottica adattiva; nella microspettrofluorometria per analisi di piccole quantità di materiali; nella fisica nucleare, in esperimenti con fibre scintillanti ed infine nelle misure di luminescenza nei semiconduttori.

Purtroppo, i normali fototubi a vuoto (**PhotoMultiplier tube - PMT**), o i fotodiodi a semiconduttore (**Avalanche Photodiodes - APD**), pur essendo sensibili ai singoli fotoni, non hanno una sufficiente risoluzione ed amplificazione per una loro corretta rivelazione. Inoltre, la corrente che si crea in uscita, in altre parole la carica emessa dal dispositivo, è molto piccola per poter essere misurata, poiché assume circa lo stesso ordine di grandezza del rumore introdotto dai circuiti elettronici collegati al rivelatore. Si ha bisogno, dunque, di fotorivelatori che permettano di avere in uscita, attraverso un'opportuna amplificazione, un segnale

elettrico misurabile e distinguibile dal rumore. La soluzione si ottiene con una normale giunzione p-n polarizzata inversamente ad una tensione superiore in modulo a quella di breakdown (tipicamente del 10-20%). Il dispositivo alimentato in questo modo funziona in *Geiger mode* ed è conosciuto con l'acronimo di **SPAD** (**Single Photon Avalanche Diode**). Lo SPAD è un sensore di luce capace di rivelare singoli fotoni; ad un tale dispositivo si richiede alta sensitività ai fotoni, basso rumore di fondo e, spesso, anche tensioni d'alimentazione piuttosto basse (circa 25-30 V). La naturale evoluzione dello SPAD è il **SiPM** (**Silicon PhotoMultiplier**), una matrice planare di SPAD, ciascuno col proprio circuito di quenching in serie, che lavorano in parallelo su un carico comune. Tale dispositivo, sommando in se tutti i pregi dello SPAD, permette di ottenere prestazioni migliori nell'ambito della rivelazione e del conteggio di fotoni.

Lo scopo di questo lavoro di tesi è la caratterizzazione di una matrice di SPAD attraverso l'analisi della carica emessa da un singolo dispositivo SPAD e da un'intera matrice planare in configurazione SiPM.

Nel **primo capitolo** è proposta una piccola introduzione alla Fotonica, seguita da un'ampia panoramica sui fotorivelatori a vuoto ed a stato solido. Sono illustrati i parametri che ne definiscono le prestazioni ed è descritto il funzionamento dello SPAD (posto a confronto con i tubi fotomoltiplicatori, i fotodiodi p-i-n ed i fotodiodi APD in zona lineare) e dei circuiti di quenching utilizzati con esso; il capitolo si conclude con una trattazione sul SiPM e sui suoi parametri prestazionali.

Nel **secondo capitolo** sono descritte le principali problematiche e le soluzioni adottate nella progettazione e nella fabbricazione di dispositivi singoli e di matrici in tecnologia planare.

Nel **terzo capitolo** sono illustrati ed analizzati i risultati ottenuti dalla caratterizzazione elettrica eseguita su un singolo dispositivo SPAD e su una matrice planare di SPAD in configurazione SiPM, tramite gli spettri di carica. A tali misure di carica ne sono state integrate altre allo scopo di analizzare meglio i risultati ottenuti. Inoltre, un certo spazio è dedicato al setup sperimentale e di misura utilizzato in questa caratterizzazione.

CAPITOLO 1
RIVELARE SINGOLI FOTONI: DALLA FOTONICA AL SiPM

1.1 La Fotonica: storia e applicazioni

La **Fotonica** è la scienza riguardante la classe di dispositivi che utilizzano i *fotoni*. Tale termine è stato coniato in analogia col termine Elettronica in riferimento alla sostituzione del fotone all'elettrone nelle operazioni tipiche dell'Elettronica quali l'elaborazione, la trasmissione e la memorizzazione di informazioni.

Il principio fisico che ha permesso di concepire ed utilizzare tali dispositivi è il *Dualismo Onda-Particella*, proprietà posseduta sia dai fotoni sia dagli elettroni e alla base della Teoria Quantistica. Secondo l'ipotesi di Planck, infatti, la radiazione elettromagnetica (e quindi anche la luce) assume un comportamento discreto, in termini energetici, tramite la teoria dei *quanti*, chiamati *fotoni* per la radiazione luminosa. Infatti, un *fotone* è la particella elementare che trasporta l'energia associata alla radiazione elettromagnetica nella banda di frequenze del visibile (da circa $4{,}5 \cdot 10^{14}$ Hz a circa $7{,}5 \cdot 10^{14}$ Hz). Tale energia è proporzionale alla frequenza, secondo la nota relazione di Planck:

$$E = h \cdot \nu \tag{1.1}$$

dove h è la costante di Planck (uguale circa a $6{,}62618 \cdot 10^{-34}$ J·s) e ν è la frequenza del fenomeno ondulatorio. Nel campo del visibile, l'energia associata ad un singolo fotone va da circa 1,77 eV a circa 3 eV **[1]**.

Ovviamente, tutto ciò è solamente un'ipotesi che, tuttavia, permette di studiare con buona approssimazione i fenomeni legati alle radiazioni luminose senza contraddire l'approccio ondulatorio classico. Il fenomeno che sta alla base di questi dispositivi è l'*effetto fotoelettrico*, che è stato spiegato, nel caso di un metallo, da Einstein, nel 1905, assumendo che all'aumentare della frequenza ν l'energia associata ad un fotone interagente col metallo raggiungerà un valore tale da superare *l'energia d'estrazione*, in altre parole l'energia necessaria all'elettrone per rompere il suo legame col metallo e generare, così, dei portatori di carica.

L'ipotesi di De Broglie assume, invece, che una particella qualunque può avere un comportamento ondulatorio, associando così alla materia (in questo caso un

fascio di particelle in moto) una radiazione elettromagnetica la cui lunghezza d'onda λ dipende dalla relazione:

$$\lambda = \frac{h}{p} \tag{1.2}$$

in cui h è la costante di Planck e p è la quantità di moto del fascio di particelle **[2]**.

Le applicazioni dei dispositivi basati sulla Fotonica sono molteplici, anche se in alcuni casi a livello sperimentale e spaziano dalle Telecomunicazioni (trasmissione su fibra ottica e conversione ottico-elettrica dei segnali), all'Astronomia (rivelazione delle radiazioni cosmiche provenienti dai meandri più remoti dello spazio), alla Medicina (la PET, acronimo inglese di Tomografia ad Emissione di Positrone, *Positron Emission Tomography*), all'Optoelettronica e alla Sensoristica (costruzione dei sistemi *encoder*, delle applicazioni per *imaging* e del "cosiddetto" *ladar*, il radar ottico, nelle applicazioni militari). I principali pregi di questi dispositivi, rispetto a quelli elettronici, sono la velocità, l'insensibilità alle interferenze elettromagnetiche, la capacità di trasportare ed immagazzinare una quantità maggiore di informazioni e la riduzione dei problemi di dissipazione del calore e della fornitura di potenza. Dalla molteplicità delle applicazioni nasce l'esigenza di avere un *fotorivelatore* molto versatile ma nello stesso tempo efficiente e preciso. Infatti, a questi dispositivi è richiesto spesso di rivelare non un semplice segnale luminoso, sia pur d'intensità molto debole (come nel caso delle radiazioni cosmiche), bensì il singolo fotone associato alla radiazione luminosa incidente.

La storia dei dispositivi fotorivelatori ha circa 100 anni. Infatti, nascono agli inizi del XX secolo con gli sviluppi della Fisica Quantistica e si sono sviluppati e progrediti fino ad oggi passando dal "vecchio" **PMT** (acronimo inglese di Tubo Foto-Moltiplicatore, *Photo-Multiplier Tube*), ai più "recenti" *fotorivelatori a stato solido* come: il **diodo p-i-n** (dal drogaggio della sua struttura interna), l'**APD** (acronimo inglese di Foto-Diodi a Valanga, *Avalache PhotoDiode*), lo **SPAD** (acronimo inglese di Diodo a Valanga di Singolo Fotone, *Single Photon Avalache Diode*) e il **SiPM** (acronimo inglese di Foto-Moltiplicatore al Silicio, *Silicon Photo-Multiplier*).

Nei paragrafi successivi saranno illustrati il funzionamento, le caratteristiche, i pregi ed i limiti dei dispositivi sopra menzionati, ponendo una particolare attenzione sulla caratterizzazione degli ultimi due arrivati: lo SPAD e il SiPM.

1.2 Tubo Fotomoltiplicatore (PMT)

Il primo dispositivo utilizzato per la rivelazione di flussi luminosi di bassa intensità è il tubo fotomoltiplicatore. Sviluppato per la prima volta da Elster e Geiter nel 1913 e commercializzato dal 1936 **[3]**, si tratta di un tubo di vetro sottovuoto che, sfruttando l'effetto fotoelettrico, converte il segnale luminoso in ingresso in una corrente elettrica misurabile in uscita. Il suo utilizzo è particolarmente consigliato in applicazioni che richiedono elevate sensibilità e risoluzione temporale, valori non ottenibili con altri tipi di fotorivelatori.

L'interno di questo tubo sottovuoto si compone di tre parti:

- Un *fotocatodo,* costituito superficialmente da un materiale sensibile alla luce (strato emittente) che ricopre uno strato metallico sottostante (strato di supporto).
- Un gruppo di 9-12 elettrodi, chiamati *dinodi,* disposti in maniera tale da "accelerare" gli elettroni emessi dal fotocatodo sui dinodi susseguenti attraverso un campo elettrico applicato dall'alimentazione esterna. Il primo dinodo produce un'emissione secondaria di circa 3-5 elettroni che vanno a colpire, ciascuno, un altro dinodo, quest'ultimo produce una seconda emissione di altri 3-5 elettroni, che vanno a colpire, ciascuno, un altro dinodo e così via in una reazione a catena, in modo da amplificare il segnale.
- Un *anodo* posto alla fine del tubo elettronico ha il compito di "raccogliere" gli elettroni emessi per generare il segnale finale d'uscita **[4]**.

Il processo d'emissione di elettroni secondari da parte dei dinodi consente di ottenere un'amplificazione interna del segnale. La corrente totale, prodotta in uscita dal fotomoltiplicatore, è, infatti, legata al numero di elettroni che giungono all'anodo e quindi al numero di dinodi presenti nel tubo, i quali ci danno una moltiplicazione di tipo esponenziale degli elettroni. Si può, così, affermare che l'aumento del numero di dinodi incrementa anche l'amplificazione finale dell'apparato. Inoltre, l'area del fotocatodo deve essere mantenuta grande il più possibile per raccogliere il maggior numero di fotoni.

Il PMT è alimentato da una tensione applicata esternamente e crescente dall'anodo verso il catodo. Tale tensione, necessaria per il processo di moltiplicazione, ha un valore compreso fra 100 e 180 V per ogni coppia di dinodi, la polarizzazione totale del dispositivo risulta, quindi, normalmente superiore al migliaio di Volt.

Di seguito è riportato uno schema di un tubo fotomoltiplicatore in cui è possibile vedere tutti gli elementi presenti all'interno del tubo a vuoto.

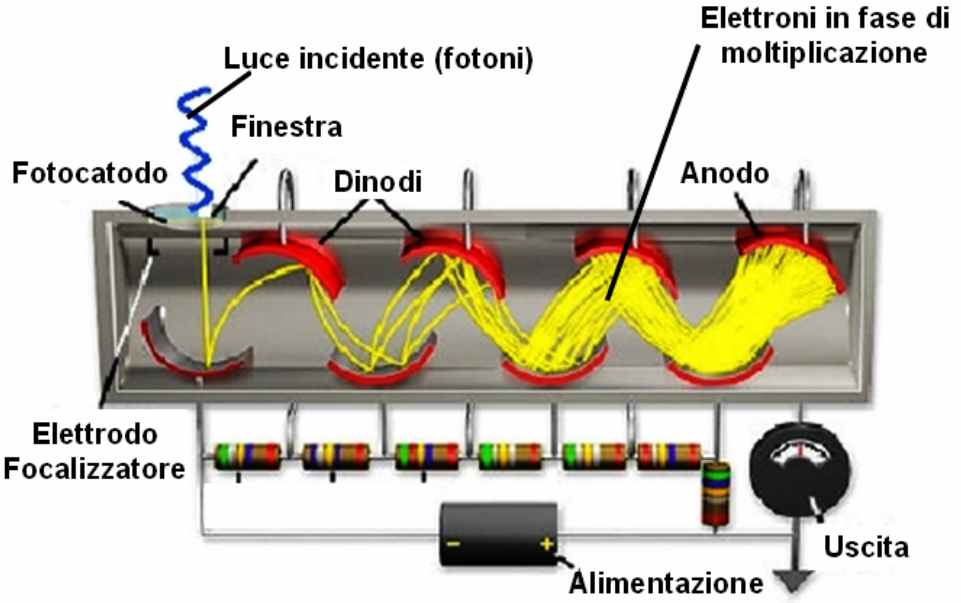

Figura 1.1
Schema di principio di un PMT

La Figura 1.1 mostra anche la circuiteria esterna utilizzata per applicare una differenza di potenziale costante ai vari dinodi ed uno strumento che rivela la corrente in uscita dal fotomoltiplicatore. Sempre osservando la figura precedente possiamo notare che, all'interno del tubo, le traiettorie degli elettroni sono rappresentate dalle curve gialle (la densità di queste traiettorie è ovviamente maggiore andando verso l'anodo). Inoltre, si hanno dei nuovi elettrodi focalizzanti che "guidano" gli elettroni emessi dal fotocatodo verso il primo dinodo.

Oltre ad un buon fattore d'amplificazione, i tubi fotomoltiplicatori hanno una risoluzione temporale variabile tra 1 ns, per i più comuni PMT e 30 ps per piastre moltiplicatrici ultra-veloci a micro-canale o micro-channel plate (MCP). Questi ultimi dispositivi sono dei tubi fotomoltiplicatori con una distribuzione continua di dinodi. I micro-channel plate (MCP) sono costituiti da un tubo di materiale isolante sulla cui superficie interna è depositato uno strato di materiale resistivo; applicando una differenza di potenziale alle estremità del dispositivo, si genera, lungo tutto il tubo, un gradiente di potenziale. Su un'estremità è montato un fotocatodo che, se colpito da un fascio di radiazione luminosa, emette elettroni. Tali portatori, incidendo sulle pareti del tubo, sono accelerati dal gradiente di tensione, subendo, così, varie amplificazioni prima di giungere all'altra estremità del dispositivo dove è posto l'anodo che raccoglierà il segnale amplificato per generare una corrente d'uscita.

Grazie all'elevata amplificazione, dovuta alla distribuzione continua di dinodi, le dimensioni del dispositivo possono essere molto ridotte, garantendo allo stesso tempo un fattore d'amplificazione pari a quello dei PMT tradizionali.

Tuttavia i PMT, sia tradizionali sia MCP, presentano parecchi svantaggi: sono dispositivi ingombranti, fragili, rumorosi, costosi (soprattutto i modelli più performanti) e limitati da un'elevata tensione d'alimentazione.

Fra i difetti sopra menzionati, il rumore è, senza dubbio, uno dei più importanti. Infatti, il tubo fotomoltiplicatore è sensibile ai disturbi elettromagnetici esterni ed all'instabilità dell'amplificazione, che, a sua volta, dipende dalle fluttuazioni nell'emissione degli elettroni da parte dei dinodi. In aggiunta a queste sorgenti di rumore, spesso trascurabili, si deve considerare che la corrente anodica di un PMT è diversa da zero anche in assenza di luce, questa è la "cosiddetta" *corrente di buio* (***"Dark Current"***). Inoltre, elettroni provenienti dai dinodi moltiplicatori (quindi non direttamente legati all'assorbimento di un fotone esterno) possono colpire le pareti del tubo e dare origine a fotoni per scintillazione. Se tali fotoni raggiungono il catodo, sono in grado, a loro volta, di produrre dei fotoelettroni "spuri" che vanno ad aumentare la corrente di buio e possono produrre anche degli impulsi secondari (***"Afterpulsing"***), che limitano di molto le prestazioni del tubo, quando è utilizzato per rivelare singoli fotoni. Per tutti questi motivi, intorno agli anni '80, si sono intensificati gli sforzi nel campo della Microelettronica per realizzare dei dispositivi a stato solido che potessero sostituire e superare i tubi fotomoltiplicatori. Infatti, sono innegabili ed allettanti i vantaggi offerti dai dispositivi a stato solido: ad esempio hanno dimensioni fisiche ridotte, tensioni d'alimentazione più basse, una maggiore robustezza ed affidabilità, insensibilità ai rumori elettromagnetici e, soprattutto, minori costi di produzione.

È, oltre a ciò, importante il vantaggio offerto in termini d'*efficienza quantica* di rivelazione dai dispositivi a semiconduttore; essa è superiore rispetto a quella che si riesce ad ottenere con i fotomoltiplicatori, soprattutto nelle regioni spettrali del rosso e del vicino infrarosso. Oggi i dispositivi a stato solido, in particolare i fotodiodi, grazie alle caratteristiche in precedenza segnalate, hanno sostituito i "vecchi" PMT in molte applicazioni.

Nei prossimi paragrafi saranno illustrate tre tipologie di fotorivelatori a stato solido:

1. Fotodiodi **p-i-n**;

2. *Avalanche Photodiode* - **APD**;

3. *Single Photon Avalanche Diode* - **SPAD**.

1.3 Fotorivelatori a stato solido

I fotorivelatori sono dispositivi che compiono la trasduzione del segnale ottico d'ingresso in un segnale elettrico d'uscita proporzionale all'intensità della radiazione luminosa che incide sulla loro area attiva.

Il principio fisico sfruttato per il suddetto scopo è il seguente. Ogni fotone, di lunghezza d'onda opportuna, incidendo su una porzione di semiconduttore intrinseco può essere assorbito e generare, grazie alla sua energia, una coppia elettrone-lacuna nel materiale che, sotto l'azione di un campo elettrico, viene separata e contribuisce alla corrente di fotoconduzione.

Ogni tipo di rivelatore fotoconduttivo ha bisogno di una polarizzazione esterna per poter rivelare la presenza di radiazione luminosa; si deve osservare che in assenza di radiazione incidente la corrente del rivelatore sia in sostanza nulla, visto anche il basso valore di conduttività del semiconduttore intrinseco. D'altra parte, in presenza di una radiazione luminosa con potenza ottica utile nota, il numero di portatori generati per assorbimento potrebbe essere rilevante e far assumere alla corrente in uscita valori apprezzabili. Il coefficiente d'assorbimento per un determinato tipo di materiale, dipende fortemente dalla lunghezza d'onda λ.

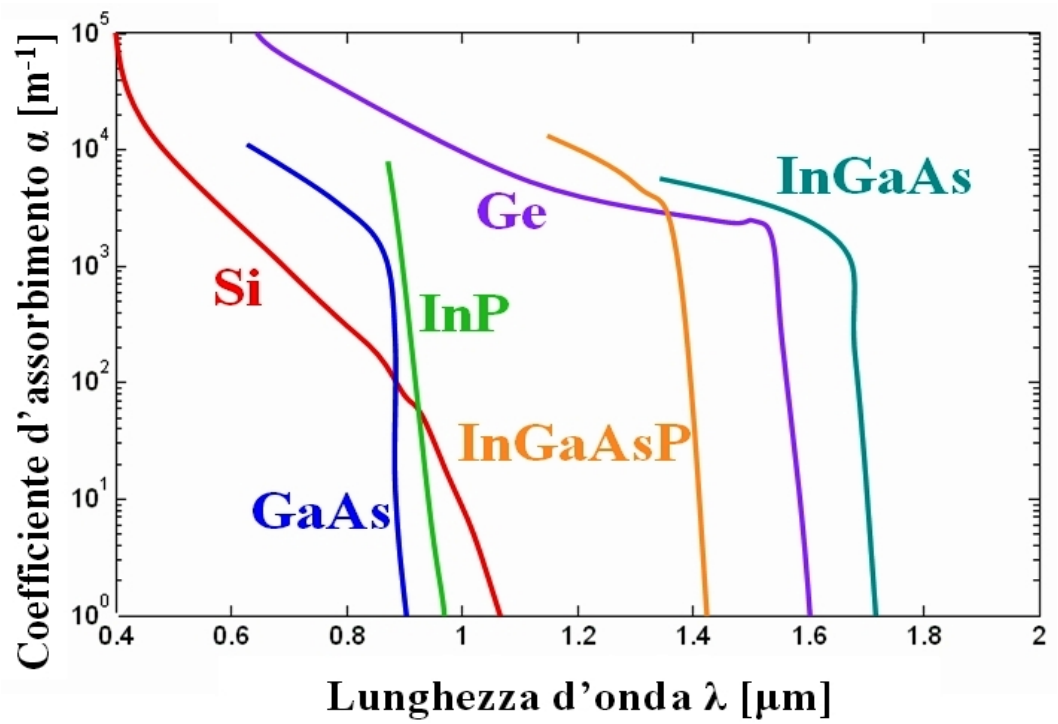

Figura 1.2

Coefficiente d'assorbimento di diversi materiali semiconduttori in funzione della lunghezza d'onda.

Dalla figura 1.2 si evince che esiste una lunghezza d'onda critica λ_c, oltre la quale l'assorbimento dei fotoni avviene troppo in profondità per consentire l'uso del dispositivo come fotorivelatore; questa lunghezza d'onda definisce il ***cut-off***.

Un fotorivelatore di tipo differente può essere ottenuto dotando il materiale di un campo elettrico intrinseco tale da provocare una corrente di fotoconduzione anche in assenza di una polarizzazione esterna: tale dispositivo prende il nome di ***fotodiodo***.

Un fotodiodo può, allora, essere rappresentato come una normale giunzione p-n polarizzata inversamente. Il principio di funzionamento è piuttosto semplice e molto simile a quello già descritto dei dispositivi a stato solido: quando un segnale ottico incide sulla giunzione, le coppie elettrone-lacuna generate dall'assorbimento dei fotoni, vengono separate ed accelerate dal campo elettrico presente nella zona di svuotamento in modo che le lacune si spostino verso la zona *p* e gli elettroni verso la regione *n*. Il moto di queste cariche elettriche darà, come segnale d'uscita, una corrente elettrica. È importante rilevare che, per produrre in uscita un segnale elettrico, i fotoni incidenti sul dispositivo devono avere un'energia maggiore dell'***energia di gap*** $\mathbf{E_g}$ (in altre parole sufficiente per far *"saltare"* i portatori dalla banda di conduzione a quella di valenza) ed essere rivelati all'interno della zona di svuotamento, poiché in questa zona è presente il campo elettrico che ha l'importante compito di separare le cariche elettriche. In realtà, anche fotoni assorbiti a meno di una lunghezza di diffusione dalla zona svuotata possono essere rivelati, in quanto è possibile la diffusione dei fotoelettroni fino alla zona svuotata.

1.3.1 Fotodiodo p-i-n

Per migliorare l'efficienza d'assorbimento di fotoni, si è pensato di creare una giunzione p-n con una zona di svuotamento più grande. Questi dispositivi sono chiamati ***fotodiodi p-i-n***.

Un *fotodiodo p-i-n* è realizzato mediante una normale giunzione p-n con, in mezzo, uno spessore di materiale poco drogato, detto *intrinseco*, rappresentato, appunto, dalla lettera *i*. Per avere la zona di svuotamento posta interamente nella regione intrinseca, lo strato *p* e lo strato *n* sono realizzati con un drogaggio maggiore rispetto ad una comune giunzione p-n.

Nella figura seguente (1.3) è rappresentata la cross-section di un fotodiodo p-i-n con rivestimento antiriflettente sopra la regione attiva.

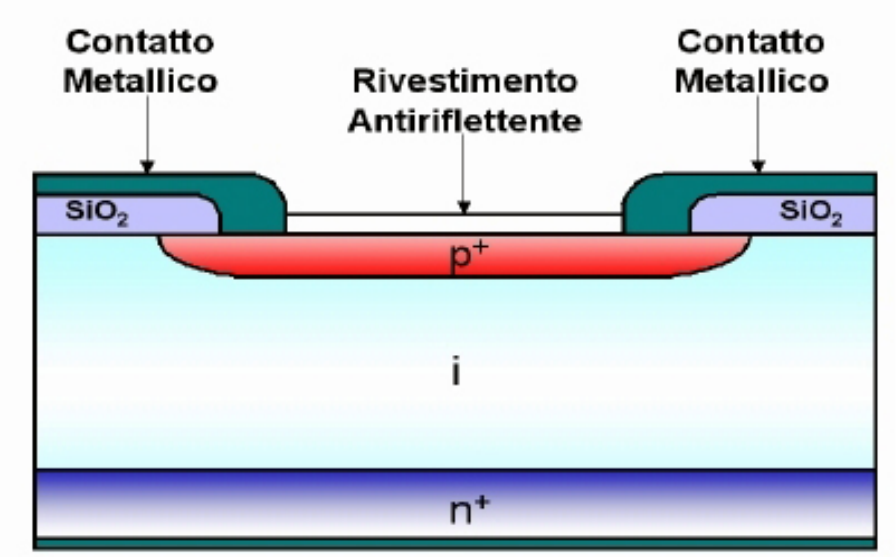

Figura 1.3
Cross-section di un fotodiodo p-i-n con rivestimento antiriflettente.

Le dimensioni di questo dispositivo, in particolare della sua zona di svuotamento o regione intrinseca, sono scelte in maniera tale da migliorare le specifiche di progetto in termini di *risposta spettrale*, *risposta in frequenza* e *rumore caratteristico*.

Per **risposta spettrale η** s'intende la relazione fra la potenza della luce incidente, per le varie lunghezze d'onda, e l'intensità della corrente generata all'uscita del fotodiodo:

$$\eta = \frac{P_{elettr.emessi}}{P_{fot.incidenti}} \tag{1.3}$$

Bisogna ricordare, però, che, se l'energia del fotone è più bassa dell'energia di gap E_g, non si avrà alcun effetto fotoelettrico. Considerando che l'energia di gap E_g del silicio è circa 1,12 eV a temperatura ambiente e che l'energia associata ad un fotone, per la relazione di Planck, è funzione della frequenza (e quindi della lunghezza d'onda λ) si può ricavare che nel vuoto **la massima lunghezza d'onda λ_{max}** che può essere assorbita dal silicio è data dalla seguente formula, ricavata dalla relazione 1.1:

$$\lambda_{\max} = \frac{h \cdot c}{E_g} \tag{1.4}$$

dove h è la già citata costante di Planck, c la velocità della luce nel vuoto (pari a circa $2{,}99792548 \cdot 10^{-34}$ m·s^{-1} **[1]**) ed E_g l'energia di gap del silicio (espressa in Joule).

In queste condizioni, si ricava che λ_{max} è circa 1,1 μm per il silicio

La **lunghezza d'onda di cut-off λ_c** rappresenta la minima lunghezza d'onda rivelabile dal dispositivo e dipende dallo spessore dello strato di diffusione.

La **risposta in frequenza** è generalmente caratterizzata dalla **frequenza di cut-off f_c** o dal **tempo di salita t_r**. La *frequenza di cut-off* f_c è la frequenza per cui il segnale d'uscita del fotodiodo diminuisce di 3 dB rispetto al suo valore a 100 kHz, mentre il *tempo di salita* t_r è il tempo necessario all'uscita per passare dal 10% al 90% del suo valore massimo. Questi due parametri sono, però, strettamente legati fra loro dalla seguente relazione:

$$t_r = \frac{0,35}{f_c} \tag{1.5}$$

I fattori che limitano la risposta in frequenza sono: i *tempi di diffusione* dei portatori per raggiungere la regione di svuotamento (quando sono assorbiti in una regione esterna ad essa), i *tempi di transito* dei portatori nella regione di svuotamento e, ad alte frequenze, la *capacità di giunzione* del fotodiodo. Questi tre parametri dipendono fortemente dalle dimensioni del dispositivo, in particolare della sua regione di svuotamento. Infatti, se si ampliano le sue dimensioni, si avrà una velocità di risposta sempre più lenta, ma, nello stesso tempo, si migliorerà l'efficienza di rivelazione del dispositivo, caratteristica che, al contrario, cresce con l'incremento dell'estensione dell'area attiva. Occorre, quindi, trovare il giusto compromesso nella scelta delle dimensioni per ottenere una buona risposta in frequenza senza intaccare l'efficienza **[5]**.

Il limite inferiore all'efficienza di rivelazione è fissato, invece, dal **rumore caratteristico**. Esso è dato dalla somma di due componenti: il *rumore termico* (dovuto alla generazione termica dei portatori, che aumenta le fluttuazioni statistiche del dispositivo) ed il *rumore shot* (dovuto alla granularità della radiazione incidente e della corrente elettrica).

Un ulteriore limite è dato dal *Nuclear Counter Effect*. Si tratta del segnale prodotto dalle particelle cariche in moto che attraversano il fotodiodo e determinano una ionizzazione del mezzo con la creazione di un segnale spurio che provoca una perdita di risoluzione nelle misure d'energia.

1.3.2 APD: fotodiodo a valanga polarizzato in zona lineare

Il fotodiodo a valanga o APD, consente di ottenere una buona amplificazione interna (circa 10^2) del segnale prodotto dall'assorbimento dei fotoni incidenti sulla sua regione di svuotamento.

Si tratta di una normale giunzione p-n che opera ad un'elevata tensione inversa di polarizzazione V_A, appena inferiore alla tensione di breakdown V_B.

Il funzionamento di un diodo a valanga APD, si basa sul fenomeno fisico di moltiplicazione a valanga dei portatori, prodotto dalla *ionizzazione per impatto*. Questo fenomeno si verifica quando il campo elettrico all'interno della regione di svuotamento è sufficientemente grande da accelerare un portatore, prodotto per l'assorbimento di un fotone, e dotarlo di un'elevata energia cinetica; in questo modo il portatore, urtando gli altri portatori della banda di valenza, trasferisce loro parte della propria energia cinetica, permettendo loro di *"saltare"* dalla banda di valenza a quella di conduzione e dando vita, così, ad una nuova coppia elettrone-lacuna che si somma a quella di partenza. Ovviamente, anche questa coppia sarà accelerata dal forte campo elettrico, acquisterà energia cinetica ed andrà a generare, sempre per impatto, nuove coppie di portatori, che si comporteranno allo stesso modo, dando origine, così, ad una valanga di portatori che procurerà un guadagno di corrente. Questo processo nei semiconduttori dipende fortemente dai *coefficienti di ionizzazione* α_p (per le lacune) ed α_n (per gli elettroni). Tali coefficienti rappresentano il numero medio di ionizzazioni che avvengono nel materiale semiconduttore per unità di lunghezza e per un fissato valore del campo elettrico locale (che provoca l'accelerazione dei portatori) **[6]**. La dipendenza dal campo elettrico di tali coefficienti è espressa dalla seguente relazione:

$$\alpha_{n,p} = A \cdot E \cdot e^{\frac{-B}{E}} \tag{1.6}$$

dove E rappresenta il campo elettrico applicato nella regione di svuotamento ed A e B sono dei coefficienti sperimentali dipendenti dalla minima energia necessaria per una collisione ionizzante e dal libero cammino medio. Il tasso di generazione elettrone-lacuna G da ionizzazione per impatto è dato, invece, da:

$$G = n\alpha_n v_n + p\alpha_p v_p \tag{1.7}$$

in cui n è la densità di elettroni in banda di conduzione, p la densità di lacune in banda di valenza, v_n e v_p sono le velocità di drift rispettivamente per gli elettroni e le lacune, α_n e α_p sono i coefficienti di ionizzazione rispettivamente per gli elettroni e le lacune. Esistono diversi approcci per il calcolo del coefficiente di ionizzazione; fra questi, il modello che riproduce meglio i dati sperimentali è denominato *"lucky electron model"*. Secondo tale modello, si assume che il guadagno energetico acquistato dalle cariche elettriche fra due collisioni successive sia mediamente inferiore all'energia necessaria per la generazione di una coppia di portatori ed esistano solamente alcuni elettroni statisticamente *"fortunati"*, che raggiungono un'energia sufficiente per la creazione di una coppia elettrone-lacuna **[7]**.

Un fotorivelatore APD è realizzato aggiungendo uno strato di semiconduttore *p*, detto *regione di moltiplicazione*, ad un normale diodo *p-i-n*.
Di seguito è mostrato un suo schema di principio.

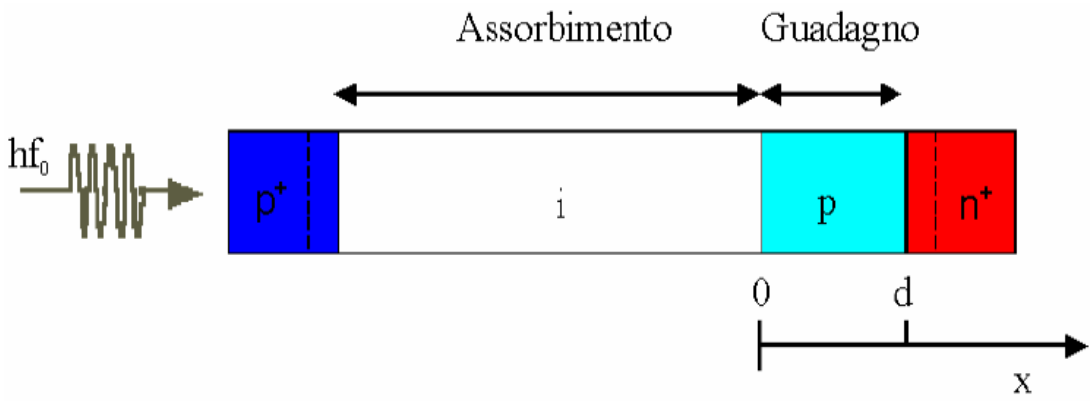

Figura 1.4
Schema di principio di un APD realizzato partendo da un diodo p-i-n.

Nello strato intrinseco *i* sono generati per assorbimento fotonico i portatori "primari", mentre nella susseguente regione *p* avviene la loro moltiplicazione a valanga. Questo tipo di fotorivelatori è indicato come *SAM-APD* (*Separate Absorption and Multiplication APD*).

Di seguito, è riprodotta la cross-section di un APD con strato antiriflettente ed anello di guardia per prevenire il breakdown e l'effetto tunnelling ai margini del diodo.

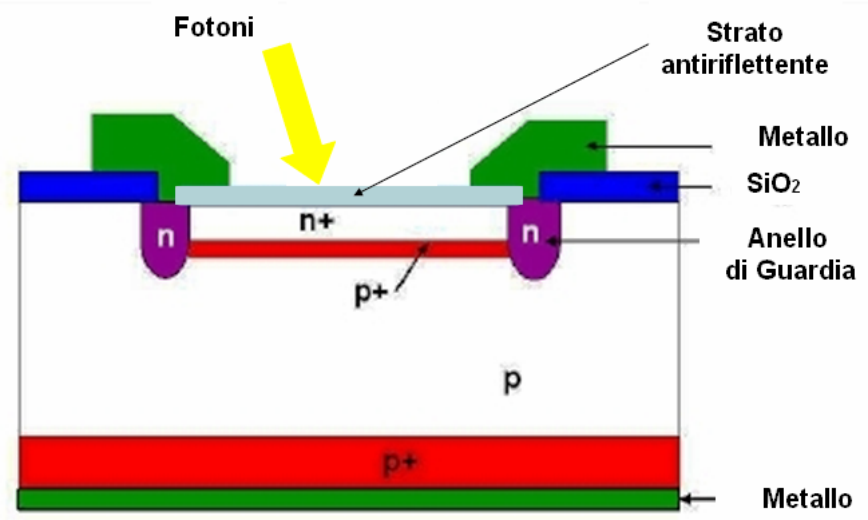

Figura 1.5
Cross-section di un APD con strato antiriflettente e anello di guardia.

Il **fattore di moltiplicazione M** è definito come il rapporto fra le densità di corrente nel punto x=d (dove d è la lunghezza della regione di moltiplicazione) e nel punto x=0, ed è espresso dalla seguente relazione:

$$M = \frac{J_e(d)}{J_e(0)} = \frac{1 - k_A}{e^{-(1-k_A)\cdot\alpha_n \cdot d} - k_A} \tag{1.8}$$

dove $k_A=\alpha_p/\alpha_n$ è il rapporto di ionizzazione **[7]**. Dalla formula (1.8) si vede che per d=0 otteniamo un fattore di moltiplicazione uguale ad uno (in altre parole non si ha nessuna zona di guadagno, ovvero M=1), mentre, per $d = - \ln(k_A) / [(1-k_A)\alpha_n]$, si ha $M \to +\infty$ (valanga innescata). La seguente formula, invece, lega la corrente prodotta nel dispositivo alla potenza ottica incidente:

$$I_{APD} = \frac{M \cdot \eta \cdot P \cdot q}{h \cdot f_0} \tag{1.9}$$

in cui f_0 è la frequenza della radiazione luminosa, η l'efficienza quantica, P la potenza ottica, q la carica elementare (pari a circa $1{,}6021892 \cdot 10^{-19}$ C **[1]**) e h la costante di Planck.

Si può affermare che la moltiplicazione a valanga negli APD è simile all'emissione secondaria di elettroni che avviene nei PMT ma il suo utilizzo, per l'amplificazione di segnali luminosi, presenta un'importante differenza: quando un elettrone (accelerato dal campo elettrico) è emesso dal reticolo cristallino del materiale, oltre a liberare un ulteriore elettrone, libera anche una lacuna.

Quest'ultima è accelerata dal campo elettrico in direzione opposta rispetto al percorso degli elettroni. Di conseguenza, questa lacuna può creare un'altra coppia elettrone-lacuna in una zona a monte rispetto al tragitto degli elettroni. Questo instaura una reazione positiva che, se da un lato aumenta l'amplificazione, dall'altro incrementa le fluttuazioni statistiche del processo di moltiplicazione, che, purtroppo, al crescere del campo elettrico, aumentano più rapidamente dell'amplificazione. Per cercare di ridurre questo svantaggio si utilizzano fotodiodi in silicio invece che al germanio, poiché nel silicio le lacune e gli elettroni non hanno la stessa efficienza di ionizzazione per urto e, se il campo elettrico è appena superiore al valore di soglia, l'efficienza di ionizzazione delle lacune è assai minore rispetto a quella degli elettroni e quindi il rumore associato diminuisce.

Un altro grave problema dell'APD in regione lineare è la forte dipendenza del fattore d'amplificazione del dispositivo con la temperatura. Infatti, quest'ultimo aumenta velocemente al diminuire della differenza:

$$V_B - V_A \tag{1.10}$$

dove V_B è la tensione di rottura (breakdown) e V_A è la tensione inversa applicata al dispositivo. Poiché il valore della tensione di rottura cambia al variare della temperatura, anche il guadagno dipende dalla temperatura e bastano variazioni anche di un solo grado centigrado per causare rilevanti cambiamenti sul valor medio del fattore d'amplificazione.

Per quanto riguarda la **risposta spettrale** degli APD, quando non c'è nessuna tensione applicata sul dispositivo, è simile a quella, già vista, dei fotodiodi p-i-n, invece, quando applichiamo una tensione inversa prossima a quella di rottura, la curva della risposta spettrale cambia leggermente, poiché l'efficienza di moltiplicazione delle cariche iniettate nella regione di valanga, e quindi il guadagno, dipende dalla lunghezza d'onda.

Per quanto riguarda la **risposta in frequenza**, anche per gli APD la *velocità di risposta* è influenzata dal *tempo di transito* delle cariche nella zona di svuotamento, dalla *capacità di giunzione* e dal *tempo di diffusione* delle cariche per giungere nella zona svuotata (se sono prodotte per assorbimento di fotoni nelle zone circostanti a quella di svuotamento). Anche in questo caso, si ha che la *frequenza di cut-off* f_c è legata al tempo di salita t_r dall'equazione:

$$t_r = \frac{0{,}35}{f_c} \tag{1.11}$$

Occorre, quindi, per aumentare la frequenza di cut-off, ridurre la capacità di giunzione con una regione di svuotamento più grande (che migliora anche l'efficienza quantica). Se, però, si allarga la regione di svuotamento, il tempo di transito impiegato dalle cariche in moto nella regione di svuotamento aumenterà. È necessario, quindi, dimensionare opportunamente lo spessore svuotato evitando lunghi tempi d'attraversamento di cariche e/o capacità parassite eccessivamente grandi.

Per quanto riguarda il **rumore caratteristico**, l'APD, operando con una tensione inversa di polarizzazione sufficiente a provocare la moltiplicazione di portatori per effetto valanga ed essendo la moltiplicazione a valanga un fenomeno di natura casuale (infatti, ogni coppia di portatori non è soggetta allo stesso processo di moltiplicazione), sarà soggetto al "cosiddetto" *rumore di valanga*.

Il *rumore shot* è maggiore negli APD rispetto ai fotodiodi p-i-n, ed è dato da:

$$I_n^2 = 2 \cdot q \cdot (I_1 + I_{dg}) \cdot M^2 \cdot B \cdot F + 2 \cdot q \cdot I_{sd} \cdot B \tag{1.12}$$

in cui q è la carica dell'elettrone, I_1 la foto-corrente a M=1, I_{dg} la componente della corrente di buio (dovuta al *rumore termico*) che è moltiplicata, I_{sd} la componente della corrente di buio che non è moltiplicata, B l'ampiezza di banda, M il guadagno e F il fattore di rumore (Excess Noise Factor) **[8]**, che è uguale a:

$$F = M \cdot K_A + (2 - \frac{1}{M}) \cdot (1 - K_A) \tag{1.13}$$

dove M è il guadagno e K_A il fattore di ionizzazione ($K_A=\alpha_p/\alpha_n$). Tale equazione (1.13) si riferisce al caso in cui gli elettroni sono iniettati nella regione di moltiplicazione. Per calcolare F nel caso delle lacune basta sostituire nella relazione 1.13 K_A con 1/ K_A. I casi limite si hanno per $\alpha_n = \alpha_p$ (cioè $K_A=1$), dove F=M, e per $\alpha_n \rightarrow 0$ (cioè K=0), dove, per M alti, F è uguale a 2; quest'ultimo caso rappresenta il limite inferiore di F poiché non è presente moltiplicazione di lacune. Questo implica che per ottenere un valore di F basso è necessario un largo rapporto tra i coefficienti di ionizzazione delle due specie di portatori di carica ed una valanga che parta dalle cariche con coefficiente di ionizzazione più alto. Nei fotodiodi al silicio si chiede che la valanga sia originata dagli elettroni. L'*Excess Noise Factor* è funzione della

lunghezza d'onda e dipende dal tipo di materiale usato, dal campo elettrico generato, dalla struttura e dalla forma della giunzione. La costruzione di un APD con regione di moltiplicazione larga comporta un valore di F maggiore, poiché un contributo rilevante alla moltiplicazione deriva dalle lacune. Nelle condizioni ideali, per minimizzare il rumore, occorre avere $K_A=0$ per gli elettroni e $K_A=\infty$ per le lacune. Solitamente sono utilizzati gli APD in cui le cariche iniettate nella regione di moltiplicazione sono gli elettroni, poiché in questo caso riusciamo ad avere $K_A \ll 1$.

Per quanto riguarda il **rapporto segnale-rumore S/N,** in un APD è dato da:

$$S/N = \frac{I^2 \cdot M^2}{\left[2q \cdot (I_l + I_{dg}) \cdot M^2 \cdot B \cdot F\right] + 2q \cdot I_{sd} \cdot B + \frac{4k \cdot T \cdot B}{R_L}} \qquad (1.14)$$

in cui il primo e il secondo termine del denominatore sono rumori shot, il terzo termine è il rumore termico, k è la costante di Boltzmann (pari a $1{,}38066 \cdot 10^{-23}$ $J \cdot K^{-1}$ **[1]**), T la temperatura espressa in gradi Kelvin e R_L la resistenza di carico. In questo caso il miglior rapporto S/N è ottenuto aumentando il guadagno fino a quando il rumore shot raggiunga un valore uguale a quello del rumore termico.

Per concludere la panoramica sull'APD, valutiamo il *Nuclear Counter Effect*, il segnale generato dal passaggio di particelle cariche in moto nel mezzo semiconduttore. Tale passaggio provoca la ionizzazione del mezzo con una produzione di coppie elettrone-lacuna rappresentanti un segnale di disturbo, specialmente se soggette al processo di moltiplicazione. In un APD, però, solo gli elettroni creati prima e le lacune generate dopo della regione d'amplificazione, sono amplificati dal campo elettrico; ciò significa che solamente una piccola parte della carica prodotta da particelle al minimo di ionizzazione produce un segnale comparabile con il segnale prodotto dal fotoelettrone, e solo tale parte è amplificata. Da ciò consegue che il *Nuclear Counter Effect* può considerarsi trascurabile per l'APD, mentre, come visto in precedenza, è decisivo per le prestazioni del fotodiodo p-i-n **[9]**.

1.3.3 SPAD: fotodiodo a valanga in Geiger mode

1.3.3.1 Funzionamento

Lo **SPAD** (Single Photon Avalanche Diode) è un normale fotodiodo APD polarizzato, però, con una tensione inversa V_A superiore (di circa il 10-20%) alla tensione di breakdown. In questo modo nella zona svuotata si ha un forte campo elettrico tale da trasferire al singolo portatore un grandissimo valore d'energia cinetica e rendere sufficiente solo una singola coppia elettrone-lacuna (generata per assorbimento di un fotone) ad innescare il processo di ionizzazione per impatto e la moltiplicazione a valanga dei portatori. L'elevata energia fornita dalla polarizzazione fa sì che il processo di moltiplicazione si autosostenga e che il guadagno sia elevato (nell'ordine di circa 10^6, contro i 10^2 degli APD in zona lineare). Di conseguenza, la corrente si porta velocemente e facilmente a livelli macroscopici dell'ordine dei milliampère (quindi facilmente misurabili dall'esterno) e con un tempo di salita nell'ordine delle centinaia di picosecondi: cosa che rende gli SPAD di gran lunga preferibili agli APD in regione lineare. La seguente figura illustra la differenza fra le due diverse regioni di funzionamento del fotodiodo: lineare e Geiger.

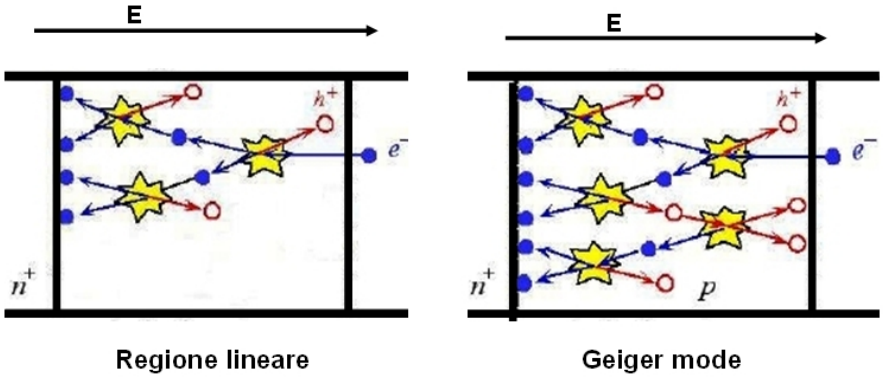

Figura 1.6
Rappresentazione di principio dell'andamento della ionizzazione da impatto nelle due regioni: lineare e Geiger.

La figura 1.6 rappresenta un'altra importante differenza fra le due regioni di funzionamento dell'APD: infatti, le lacune in regione lineare non hanno abbastanza energia per poter innescare delle valanghe e per questo il loro coefficiente di ionizzazione è molto basso rispetto a quello degli elettroni.

Poiché la corrente si autosostiene per effetto della valanga, essa fluirà nel sensore finché la tensione inversa di polarizzazione V_A sarà maggiore della tensione di breakdown V_B. Durante tale periodo di tempo, il dispositivo non è in grado di

distinguere l'arrivo di ulteriori fotoni e conterebbe, così, in uscita, un solo un fotone anche quando in ingresso ne è stato assorbito più di uno, essendo la valanga una sola. Per questo motivo, occorre ripristinare le condizioni iniziali "fermando" la valanga. L'unico modo per "spegnere" la corrente, eseguendo un controllo sulla valanga, è abbassare il campo elettrico ai capi della regione di svuotamento ad un valore tale da non permettere più la moltiplicazione per impatto dei portatori, riportando la tensione inversa di polarizzazione V_A sotto il valore di breakdown V_B ($V_A<V_B$) per un certo periodo, detto ***tempo di hold-off***, durante il quale il dispositivo non può rivelare l'arrivo di nessun fotone. A causa di questo, è preferibile che il tempo di hold-off sia il più piccolo possibile, in modo da tenere inattivo il sistema per pochissimo tempo. Per condurre a termine il tempo di hold-off e poter rivelare l'arrivo di un nuovo fotone, la tensione inversa di polarizzazione V_A deve essere riportata sopra il valore di breakdown V_B ($V_A>V_B$), ripristinando, così, le condizioni iniziali di funzionamento. Per ogni fotone rivelato osserveremo, così, in uscita un impulso di corrente, la cui durata è determinata da come si abbassa la tensione di polarizzazione e, in particolare, da cosa si usa per farlo. A tale fine sono stati progettati ed utilizzati diversi tipi di *circuiti di spegnimento*, detti, in inglese, ***quenching circuits***. Esistono due diverse famiglie di circuiti di quenching: i ***circuiti di quenching passivo (PQC)*** e i ***circuiti di quenching attivo (AQC)***; entrambe le famiglie, con i loro pregi e difetti, saranno trattate con più dettaglio nel seguito di questa tesi.
Segue il diagramma a blocchi del sistema di controllo in retroazione descritto.

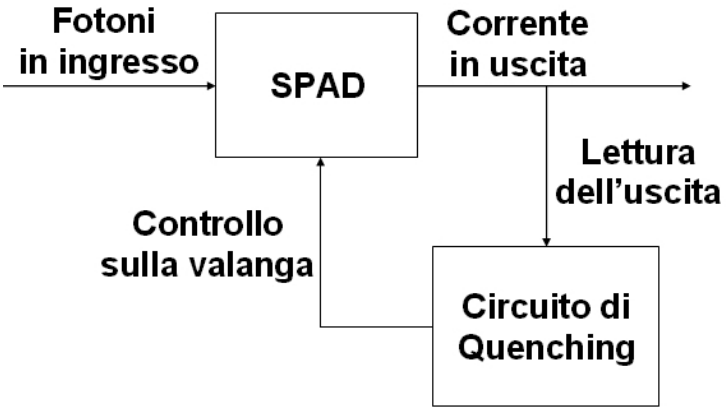

Figura 1.7
Diagramma a blocchi del sistema di controllo dello SPAD

Di seguito sono illustrati i fenomeni fisici che definiscono le prestazioni del dispositivo, quali il **Dark Count Rate**, l'**Afterpulsing** e l'**Efficienza quantica**.

1.3.3.2 Dark Count Rate

Come si è visto nel caso dei tubi fotomoltiplicatori, anche nelle misure di conteggio di fotoni con lo SPAD, è presente, purtroppo, un rumore interno dovuto alle **fluttuazioni Poissoniane di impulsi spontanei** provocati dalla generazione termica di coppie elettrone-lacuna nella regione svuotata. Poiché questi impulsi sono osservabili anche in condizioni di buio, il loro conteggio nell'unità di tempo è chiamato **dark count rate** (tasso di conteggi di buio).

Il reciproco del *dark count rate* è il valore medio dell'intervallo di tempo che trascorre dal momento in cui lo SPAD è polarizzato con una tensione inversa V_A maggiore di quella di breakdown V_B (è, quindi, pronto a rivelare l'arrivo di un fotone) sino all'istante in cui lo SPAD auto-innesca la valanga per generazione termica di una coppia elettrone-lacuna nella regione di svuotamento.

Misurando questi tempi, si valuta il rumore intrinseco di uno SPAD, ottenendo il limite di rivelazione del dispositivo, in termini di numero minimo di fotoni al secondo che è possibile discriminare dai conteggi di buio.

Il fenomeno dei conteggi di buio è spiegato dalla *teoria Shockley-Hall-Read* (SHR), secondo cui la generazione termica di coppie elettrone-lacuna, nella regione svuotata, è dovuta alla presenza di centri di generazione-ricombinazione (centri G-R) che hanno un livello energetico posto circa a metà gap fra la banda di valenza (VB) e la banda di conduzione (CB). Una transizione diretta avviene quando un elettrone si ricombina "saltando" dalla banda di conduzione a quella di valenza. Ora, sebbene queste transizioni dirette fra banda di valenza e banda di conduzione siano tipiche in tutti i semiconduttori, nel caso del silicio sono assai improbabili, tranne per elevate concentrazioni di elettroni e lacune. Nel silicio non si hanno transizioni dirette, perché sia l'elettrone sia la lacuna, nelle rispettive bande, hanno una certa energia cinetica ed una definita quantità di moto, che devono entrambe conservarsi nel processo di ricombinazione. É facile conservare l'energia cinetica tramite l'emissione di un adeguato fotone, mentre è difficile conservare la quantità di moto. Nel silicio, in particolare, gli elettroni meno energetici, liberi nella banda di conduzione, hanno una quantità di moto diversa da zero, mentre le lacune meno energetiche nella banda di valenza hanno una quantità di moto molto prossima a zero. La transizione diretta di un elettrone dalla banda di conduzione alla banda di valenza con emissione di fotone è, quindi, molto improbabile. Molto più probabile è, invece, la transizione attraverso un terzo stato, nella forma di una "trappola" localizzata posta (da un punto

di vista energetico) all'interno del gap fra le due bande di valenza e conduzione, in modo che faccia da gradino intermedio alla transizione e riesca ad assorbire la quantità di moto dell'elettrone libero; queste trappole sono chiamate *centri di ricombinazione* **[9]**.

Segue una rappresentazione schematica del modello SHR.

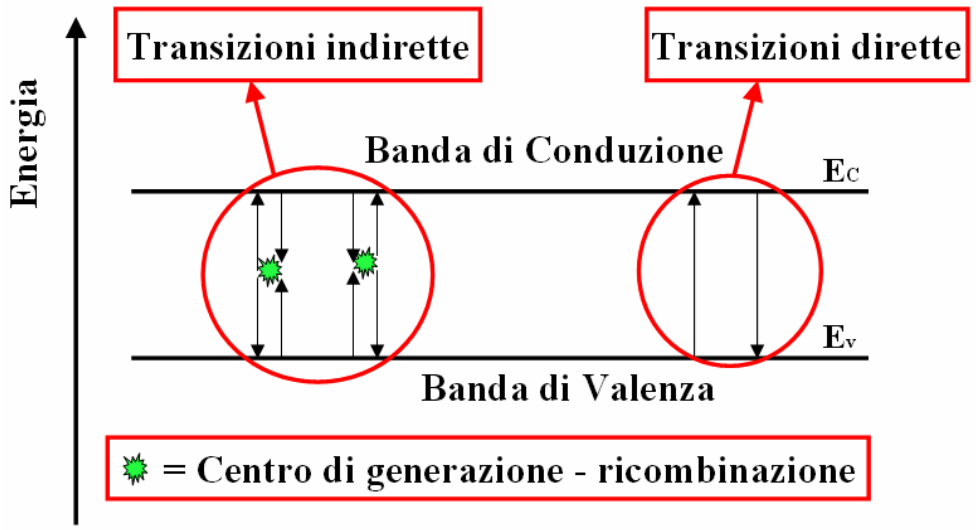

Figura 1.8
Rappresentazione schematica della teoria Shockley-Hall-Read

La presenza dei centri di ricombinazione è essenzialmente dovuta alle imperfezioni presenti nel reticolo cristallino che introducono livelli energetici all'interno del gap. Tali imperfezioni sono causate principalmente da due fattori: l'introduzione di *impurità* nel semiconduttore e l'esposizione a *radiazioni* ad alta energia.

Le impurità della III e V colonna della Tavola periodica degli elementi, se immesse in un campione di silicio, introducono livelli d'energia entro la banda di gap. Essendo tali elementi abbastanza simili al silicio (che appartiene alla IV colonna), i livelli di energia associati ad essi saranno poco profondi, ovvero vicini ai limiti della banda di valenza e di conduzione, diventando, quindi, accettori e donatori nel silicio. Altre impurità, invece, introducono livelli d'energia molto più vicini al centro del gap ed agiranno come centri di generazione perché avranno un tempo di rilascio degli elettroni maggiore rispetto ai centri poco profondi. L'altro modo per introdurre livelli d'energia nel gap è dato dall'esposizione del semiconduttore a radiazioni ad alta energia quali elettroni, protoni, raggi gamma e/o neutroni. Le particelle dotate d'alta energia possono spostare gli atomi dalla loro posizione nel

reticolo, provocando la formazione di difetti reticolari complessi (che si comportano come impurità introdotte nel semiconduttore) e di livelli energetici agenti da centri di ricombinazione. Si può notare che col procedere del bombardamento di radiazioni il tempo di vita comincia a diminuire.

In accordo al modello Shockley-Hall-Read, la popolazione dei centri di G-R, nella regione svuotata dello SPAD, occupata rispettivamente da elettroni e lacune è:

$$n_T = \frac{e_p}{e_n + e_p} \tag{1.15}$$

$$p_T = N_T - n_T \tag{1.16}$$

dove N_T la densità totale di centri G-R nella regione svuotata, n_T è la densità di centri occupati da elettroni, p_T è la densità di centri occupati da lacune ed infine, $e_{n,p}$ è la probabilità di emissione dai centri G-R rispettivamente degli elettroni e delle lacune, espressa dalla seguente formula:

$$e_{n,p} = \sigma_{n,p} \cdot v_{th} \cdot n_i \cdot e^{\pm\frac{(E_i - E_T)}{k \cdot T}} \tag{1.17}$$

in cui $\sigma_{n,p}$ è la sezione di cattura dei centri rispettivamente di elettroni e lacune, v_{th} la velocità termica, E_i il livello di Fermi intrinseco posto a metà gap (E_i=E_C-E_V), E_T il livello energetico del centro di ricombinazione-generazione considerato, k la costante di Boltzmann e T la temperatura espressa in gradi Kelvin.

Se si considera un singolo centro di ricombinazione, la probabilità che in un intervallo di tempo t non avvenga nessuna emissione è data dalla seguente equazione:

$$P_R(t) = e^{-t \cdot e_{n,p}} \tag{1.18}$$

Dalla relazione precedente, si può valutare la probabilità che nessuno degli N centri G-R emetta portatori nell'intervallo di tempo t e poiché tali centri sono statisticamente indipendenti si ha:

$$P_N = (e^{-t \cdot e_{n,p}})^N = e^{-t \cdot e_{n,p} \cdot N} = -\frac{N \cdot e_n \cdot e_p \cdot t}{e_n + e_p} \tag{1.19}$$

Ponendo nella precedente equazione $\tau=1/e_{n,p}\cdot N$ ed assumendo uguali, fra loro, le sezioni di cattura di elettroni e lacune, cioè ponendo $\sigma_n=\sigma_p=\sigma$, si ricava la seguente relazione:

$$\tau = \frac{e_n + e_p}{n_T \cdot A \cdot W \cdot e_n \cdot e_p} \approx \frac{2 \cdot \cosh\left[\frac{(E_i - E_T)}{k \cdot T}\right]}{N_T \cdot A \cdot W \cdot \sigma \cdot v_{th} \cdot n_i} \tag{1.20}$$

in cui A è l'area della regione svuotata e W il suo spessore.

Inoltre, in accordo con la teoria appena enunciata, si è dimostrato sperimentalmente che la distribuzione dei tempi di valanga del dispositivo è di tipo esponenziale con costante di tempo caratteristica data da τ.

Continuando la nostra panoramica sul dark count rate, si può osservare che in condizioni d'alto campo elettrico, l'emissività $e_{n,p}$ potrebbe essere condizionata dai fenomeni *trap assisted tunneling* (TAT) e *Poole-Frenkel*.

L'effetto *trap assisted tunneling* aumenta la probabilità d'emissione dei centri di ricombinazione G-R secondo la relazione:

$$\tau_{TAT} = \frac{\tau_{n,p}}{1 + 2\sqrt{3\pi} \cdot \frac{E}{E_\Gamma} \cdot e^{\left(\frac{E}{E_\Gamma}\right)^2}} \tag{1.21}$$

con $E_\Gamma \approx 5 \cdot 10^7$ V/m.

Tuttavia, dalle misure sperimentali effettuate, si osserva che il massimo campo elettrico E, ha un valore inferiore ad E_Γ e quindi possiamo considerare questi effetti del tutto trascurabili. Infatti, mentre l'equazione 1.21 prevede un aumento esponenziale del dark count al crescere della sovratensione applicata V_A-V_B; in realtà si è visto solamente un modesto aumento dei conteggi di buio. Ciò può essere spiegato con l'incremento dello strato della zona svuotata, dovuto alla sovratensione applicata, e con l'aumento della probabilità di triggering di una valanga, data da:

$$P_{trigger}(V_e) = 1 - e^{-\frac{V_e}{V_c}} \tag{1.22}$$

dove V_e è la sovratensione applicata alla giunzione e V_c è una tensione caratteristica connessa alla struttura **[10]**. Per l'effetto Poole-Frenkel si possono fare simili

considerazioni. Nel grafico seguente è riportato l'andamento del dark-count al crescere della sovratensione (V_A-V_B).

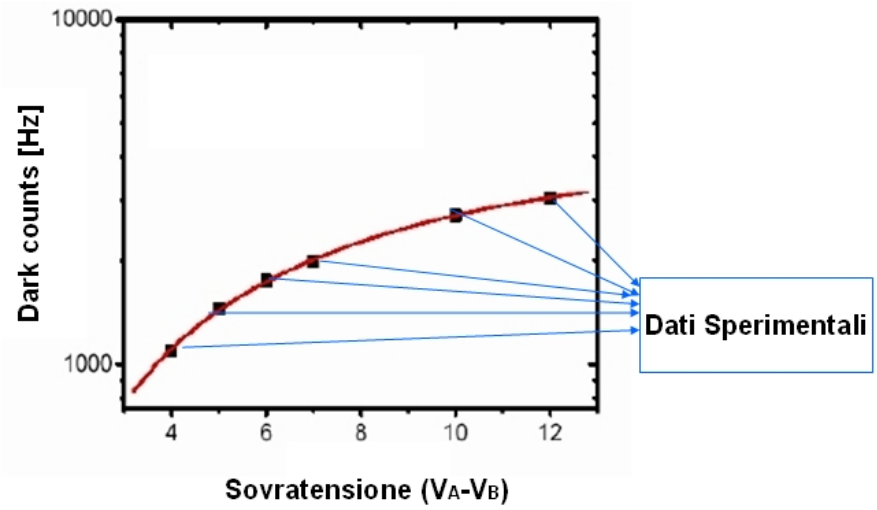

Figura 1.9
Andamento del dark count al variare della sovratensione (V_A-V_B).

Infine, è importante ricordare che la distribuzione statistica dei ritardi misurati ha un andamento di tipo esponenziale, poiché gli scatti del dispositivo sono eventi poissoniani. La figura seguente mostra l'istogramma degli eventi di valanga misurati per uno SPAD, usando la scala semilogaritmica.

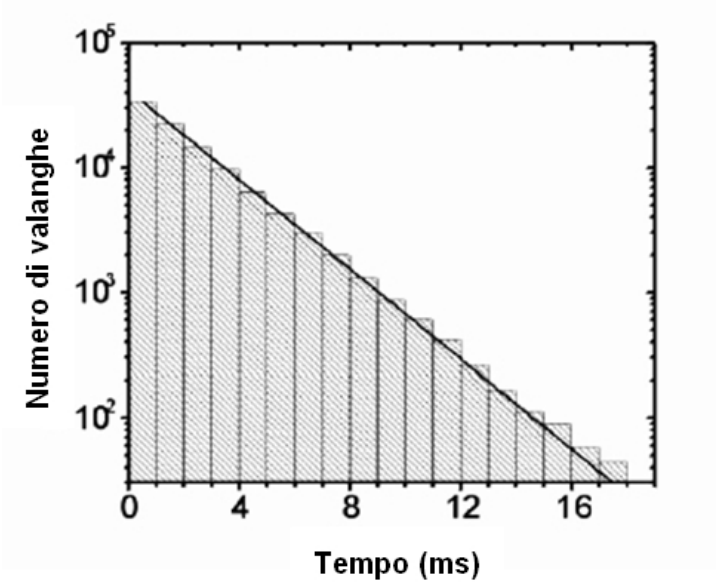

Figura 1.10
Distribuzione statistica dei ritardi per uno SPAD

1.3.3.3 Afterpulsing

L'**Afterpulsing** è un fenomeno di disturbo dovuto all'emissione secondaria di portatori per la presenza di difetti nella regione di svuotamento. Tali difetti possiedono dei livelli d'energia, intermedi fra metà gap ed il margine di banda, che possono catturare e rilasciare, con un certo tempo di ritardo, alcuni portatori della corrente di valanga: il comportamento di tali difetti è, dunque, assimilabile a quello di trappole.

Durante il tempo di hold-off, lo SPAD è insensibile all'emissione di queste cariche, poiché l'assenza di un forte campo elettrico ai capi della giunzione non permette ai portatori l'innesco di una valanga per impatto. Tuttavia, se il rilascio avviene più tardi, quando la tensione sullo SPAD è superiore a quella di rottura, la carica rilasciata può accendere una valanga "spuria", con un valore di carica inferiore, determinando una **correlazione di impulsi spuri**. Da quanto detto si evince che l'afterpulsing può far aumentare enormemente il dark count rate. Inoltre, questi impulsi secondari, non essendo distribuiti in maniera casuale (come un rumore bianco), ma come un impulso di dark count, provocano delle distorsioni rilevanti negli istogrammi di distribuzione dei tempi d'arrivo dei fotoni.

Un rimedio efficace per minimizzare questo problema è di avere dei tempi di hold-off lunghi in modo da spopolare la maggior parte delle trappole. Un'altra soluzione efficace potrebbe essere il raffreddamento del dispositivo per ridurre il dark count rate, ma in questo caso la diminuzione di temperatura farebbe accrescere il tempo di rilascio delle trappole e, quindi, aumenterebbe l'afterpulsing a meno di scegliere tempi di hold off più lunghi.

La probabilità di questo fenomeno è detta **"probabilità di afterpulsing" P_{ap}**. P_{ap} è nulla quando l'afterpulsing è eliminato, se, invece, P_{ap} è diverso da zero il sistema sarà affetto da questo problema e il dark totale $C_{m,dark}$ è maggiore del dark count primario C_{th} (cioè senza afterpulsing).
Tale variazione è espressa dalla relazione **[11]**:

$$C_{m,dark} = C_{th} + C_{th} \cdot P_{ap} + C_{th} \cdot (P_{ap})^2 + \ldots + C_{th} \cdot (P_{ap})^n + \ldots = C_{th} \cdot \sum_{n=0}^{+\infty} (P_{ap})^n = \frac{C_{th}}{1 - P_{ap}} \tag{1.23}$$

Uno dei tanti metodi per il calcolo della probabilità d'afterpulsing P_{ap} è qui di seguito esposto.

La probabilità che un elettrone generato da un evento di valanga sia intrappolato ed in seguito inneschi un evento di valanga è data da:

$$P_{ae} = N_{te} \cdot \sigma_{te} \cdot W_e \tag{1.24}$$

analogamente per una lacuna:

$$P_{ah} = N_{th} \cdot \sigma_{th} \cdot W_h \tag{1.25}$$

dove N_{te} ed N_{th} sono le concentrazioni di trappole rispettivamente per elettroni e lacune, σ_{te} e σ_{th} definiscono le sezioni di cattura delle trappole rispettivamente per elettroni e lacune e, infine, W_e e W_h rappresentano le larghezze effettive della regione di carica spaziale rispettivamente per elettroni e lacune (si calcolano dalla distribuzione spaziale di ionizzazione durante la scarica Geiger e variano con l'overvoltage).

Conoscendo P_{ae}, si può ricavare la probabilità totale d'afterpulsing P_{at}, in altre parole la probabilità che il numero totale n degli elettroni generati durante la valanga siano intrappolati e generino in seguito una valanga:

$$P_{at} = 1 - (1 - N_{te} \cdot \sigma_{te} \cdot W_e)^n \approx N_{te} \cdot \sigma_{te} \cdot W_e \cdot n \tag{1.26}$$

P_{at} rappresenta, dunque, la probabilità totale d'afterpulsing per gli elettroni, per le lacune, invece, esiste una formula analoga. Poiché, inoltre, la carica che passa attraverso la zona di svuotamento durante la valanga porta la tensione ai capi del dispositivo da V_b+V_e a V_b, possiamo esplicitare meglio n:

$$n = \frac{C \cdot V_e}{q} \tag{1.27}$$

dove C è la capacità della regione svuotata dello SPAD e q la carica dell'elettrone.
Sostituendo la 1.27 nella 1.26 otteniamo:

$$P_{at} \approx N_{te} \cdot \sigma_{te} \cdot W_e \cdot V_e \cdot \frac{C}{q} \tag{1.28}$$

Questa relazione fa notare che la probabilità d'afterpulsing cresce all'aumentare dell'area e della sovratensione di polarizzazione applicata. Per di più, la dipendenza dall'overvoltage non è lineare, essendo, come detto sopra, anche W_e funzione della

sovratensione V_e. Tuttavia, la relazione 1.28 offre un valido suggerimento per la riduzione dell'afterpulsing: occorre che la capacità della regione svuotata e le capacità della circuiteria di condizionamento siano di valore più piccolo possibile, in altre parole occorre ridurre le dimensioni del dispositivo.

La dipendenza della probabilità d'afterpulsing rispetto alla tensione d'overvoltage (V_e) è ben intuibile dal fatto che, durante il primo impulso di valanga, solo una frazione molto piccola di trappole è occupata da portatori, quindi la probabilità per un portatore di essere intrappolato è praticamente costante ed il numero di trappole occupate è proporzionale alla quantità totale di carica Q_{av}; quest'ultima è, ovviamente, legata alla corrente durante l'impulso di valanga, a sua volta funzione della sovratensione.

Bisogna, in ogni caso, osservare che nella precedente trattazione è stato considerato solo un unico tipo di trappole. Se, però, come comunemente accade, sono presenti differenti tipi di trappole (a diversi livelli energetici), ognuno di questi dovrà essere considerato separatamente nel calcolo; la probabilità d'afterpulsing totale sarà, infine, la somma delle probabilità d'afterpulsing ricavate per ogni singolo tipo di trappola.

È opportuno, oltre a ciò, considerare che, dopo l'impulso di valanga, alcune trappole potrebbero essere occupate da portatori, ci sarà, allora, una certa probabilità che il portatore intrappolato sia rilasciato e inneschi nuovamente la valanga. La suddetta densità di probabilità, rispetto al tempo, può essere espressa come:

$$P_a(t) = \sum_{i=1}^{n} A_i \cdot e^{\frac{-t}{\tau_i}} \tag{1.29}$$

Ogni termine di questa sommatoria riguarda un tipo differente di trappole con una propria concentrazione, sezione di cattura e tempo di vita τ_i.

Sperimentalmente, $P_a(t)$ può essere ricavata misurando la funzione d'autocorrelazione del segnale d'uscita dello SPAD.

L'equazione 1.29 fornisce la probabilità d'afterpulsing in funzione del tempo; la probabilità d'afterpulsing P_{ap} totale è l'integrale di questa quantità rispetto al tempo **[12]**.

1.3.3.4 Efficienza Quantica

La probabilità che un fotone incidente sia rivelato coincide con la probabilità che il singolo fotone inneschi la valanga.

Molte condizioni devono, però, verificarsi affinché il fotone sia rivelato.

Per prima cosa, il fotone può essere riflesso nell'interfaccia fra l'aria ed il silicio, oppure, più in generale, quando sono presenti diversi strati di materiale (come l'ossido di silicio SiO_2 che ricopre gran parte del dispositivo), il fotone può essere riflesso in qualche zona prima di raggiungere il silicio e, quindi, l'area attiva dello SPAD.

Si consideri, in questa trattazione, il sistema aria-ossido di silicio-silicio. La porzione di luce incidente trasmessa dal sistema è data dalla formula del coefficiente di trasmissione T per un sistema di lamine piane **[1]**:

$$T = \frac{4 \cdot n_1 \cdot n_{Si}}{\left| n_l \cdot B + C \right|^2} \tag{1.30}$$

con:

$$B = \cos(\delta) + i \cdot \frac{n_{Si}}{n_1} \cdot sen(\delta) \tag{1.31}$$

$$C = n_{Si} \cdot \cos(\delta) + i \cdot n_1 \cdot sen(\delta) \tag{1.32}$$

in cui δ è funzione della lunghezza d'onda e dello spessore dello strato intermedio secondo la formula:

$$\delta = \frac{2\pi}{\lambda} \cdot n_{OX} \cdot d \cdot \cos(\Phi) \tag{1.34}$$

dove T è, come detto, il coefficiente di trasmissione, n_1 e n_{ox} sono gli indici di rifrazione rispettivamente dell'aria e dello strato intermedio di ossido di silicio, n_{Si} è l'indice di rifrazione complesso del silicio, λ la lunghezza d'onda della luce relativa al vuoto, δ la variazione di fase, d lo spessore dello strato intermedio e Φ l'angolo misurato fra la normale alla superficie e la luce che colpisce l'interfaccia dello strato intermedio.

La seguente figura (1.11) illustra il sistema considerato in questa trattazione.

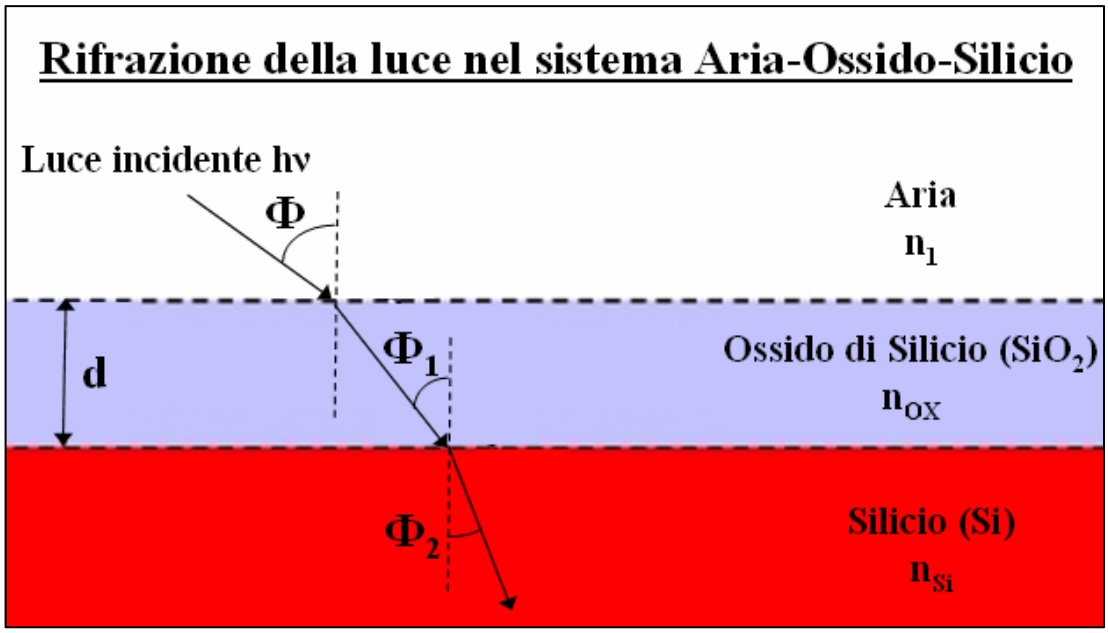

Figura 1.11

In figura non è stata considerata la luce riflessa. Le linee tratteggiate rappresentano le superfici di separazione dei mezzi, mentre le linee "tratto e punto" sono le normali a tali superfici. Inoltre, $\Phi > \Phi_1 > \Phi_2$ poiché la luce passa da un mezzo meno rifrangente ad uno più rifrangente [legge di Snell o della rifrazione: $n_1\, sen(\theta_1) = n_2\, sen(\theta_2)$]

A questo punto il fotone può essere assorbito, creando una coppia elettrone-lacuna, nella zona svuotata, oppure può arrivare fino alla zona quasi neutra. La probabilità d'assorbimento di un fotone in una regione dx, posta ad una certa distanza x dall'interfaccia, è data dal prodotto della probabilità che il fotone non sia stato assorbito viaggiando fino ad x e la probabilità che sia assorbito in dx:

$$P_a(x) = e^{-\alpha \cdot x} \cdot \alpha \cdot dx \tag{1.35}$$

in cui α è il coefficiente di assorbimento del silicio. Tale coefficiente varia in funzione della lunghezza d'onda della radiazione elettromagnetica incidente. Per la luce azzurra, ad esempio, il coefficiente d'assorbimento è maggiore rispetto alla luce rossa.

Nella figura seguente (1.12) è riportato il grafico del coefficiente d'assorbimento del silicio in funzione della lunghezza d'onda ad una temperatura di 300 K, in scala semilogaritmica.

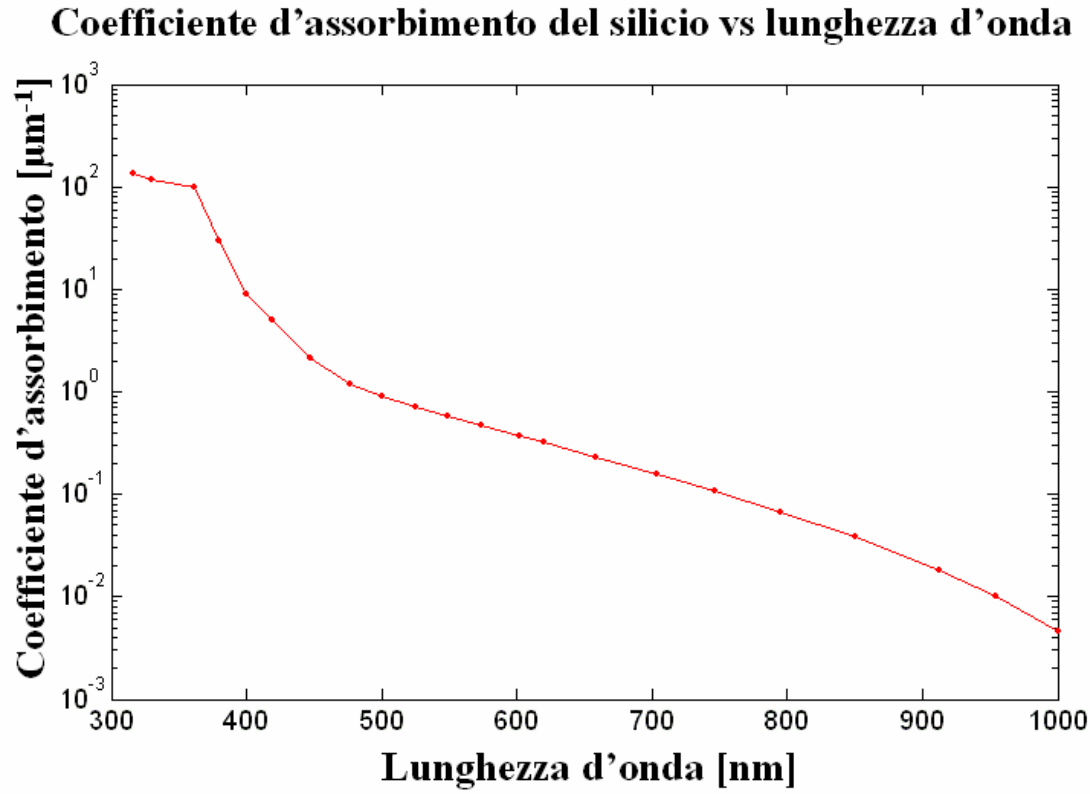

Figura 1.12

Coefficiente d'assorbimento del silicio in funzione della lunghezza d'onda.

Dalla precedente equazione, la (1.35), è possibile ricavare la posizione dove i fotoni, secondo la loro lunghezza d'onda, sono maggiormente assorbiti. Tuttavia, anche se i fotoni sono assorbiti nella regione di svuotamento, non è assicurata la generazione di coppie elettrone-lacuna, poiché essa dipende anche dalla sovratensione che è stata applicata allo SPAD, come si nota dall'equazione 1.22

L'altra possibilità è che i fotoni possano essere assorbiti nella regione quasi neutra. Tuttavia, poiché il nostro dispositivo è stato progettato per evitare l'assorbimento dei fotoni nella regione quasi neutra, possiamo trascurare questa seconda possibilità. La probabilità di rivelazione finale è

$$P_d(\lambda) = T(\lambda) \cdot P_{trigger}\left(V_{trigger}\right) \cdot \int_{x_n}^{x_p} P_a(x)dx \qquad (1.36)$$

dove x_n e x_p sono rispettivamente i bordi della regione svuotata n e p.

Nell'equazione precedente si è assunto che la probabilità di trigger rimanga la stessa nell'intera regione svuotata. Questo non è chiaramente vero in quanto tale probabilità è funzione della posizione, tuttavia, l'equazione può essere molto utile per eseguire, in maniera rapida, il calcolo dell'efficienza quantica e la stima che se ne ricava non è, poi, tanto lontana dal valore esatto.

1.4 Circuiti di quenching

Il circuito di spegnimento, o *quenching circuit*, ha il compito di spegnere la valanga che si genera nel dispositivo nel più breve tempo possibile, assicurando il corretto funzionamento del sensore. Per spegnere la valanga il circuito agisce sulla tensione inversa di polarizzazione dello SPAD, portandola sotto la tensione di breakdown (in valore assoluto) e ripristinandola dopo un certo intervallo di tempo, chiamato tempo di hold-off. Nella realizzazione del circuito si hanno delle specifiche stringenti sulla velocità di spegnimento, affinché la durata della corrente di valanga sia più piccola possibile. Questo requisito è necessario per tre principali motivi:

1. *Autoriscaldamento*.

2. *Intrappolamento dei portatori*.

3. *Cross-talk ottico*.

Il primo punto riguarda l'eccessiva dissipazione di potenza, che porta al surriscaldamento del diodo e alla conseguente variazione della tensione di breakdown, che introduce distorsioni non lineari nel photon counting.

Il secondo è legato al fatto che alcuni portatori di valanga che, intrappolati nei pressi della giunzione ed in seguito rilasciati, potrebbero reinnescare la valanga. Questi successivi impulsi correlati, che alterano il conteggio dei fotoni, possono essere ridotti minimizzando il numero totale di portatori che attraversano la giunzione durante la valanga, riducendo così la durata e l'intensità della valanga stessa.

Il terzo punto riguarda l'effetto di portatori "caldi" della valanga che emettono fotoni secondari dovuti a fenomeni di Bremsstrahlung; questi fotoni potrebbero essere assorbiti da SPAD "vicini", innescando in questi ultimi la valanga; tale fenomeno, denominato *"cross talk"*, sarà studiato ed approfondito più avanti parlando di SiPM e matrici di SPAD.

Come anticipato, esistono due diversi tipi di circuiti di quenching:

- Il circuito di quenching passivo (PQC)
- Il circuito di quenching attivo (AQC)

In realtà la tensione è portata sotto il breakdown solo con il quenching attivo: il quenching passivo abbassa la tensione, invece, al valore di breakdown.

Nei prossimi due paragrafi saranno approfonditi entrambi i circuiti di quenching.

1.4.1 Circuito di quenching passivo (PQC)

Il modo più semplice ed immediato per realizzare un circuito di spegnimento è inserire una resistenza R_L, di valore elevato (centinaia di kΩ), in serie alla sorgente di polarizzazione V_A dello SPAD, come illustrato nella seguente figura.

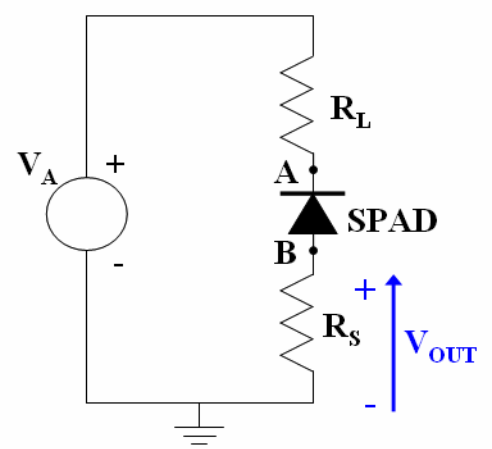

Figura 1.13
Circuito di quenching passivo.

La resistenza R_L permette che la tensione ai capi dello SPAD V_d (fra i nodi A e B) scenda sotto la tensione di alimentazione V_A, quando la valanga è innescata, fino al livello della tensione di breakdown V_B, autospegnendo la corrente di valanga. La resistenza R_S in serie verso massa serve ad osservare l'impulso generato dalla corrente di valanga. Ora saranno analizzati più in dettaglio il funzionamento e le prestazioni del circuito di quenching passivo. La seguente figura illustra il modello equivalente del circuito in esame; essendo lo SPAD un diodo che lavora nella regione di breakdown, il suo modello circuitale è quello di un diodo Zener **[13]**.

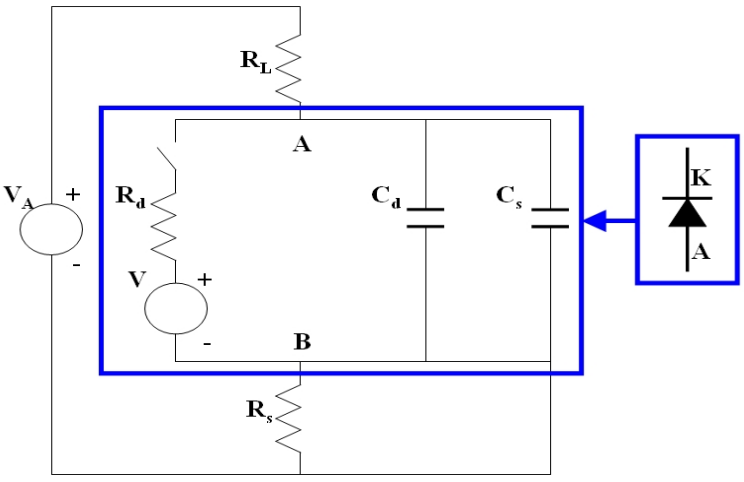

Figura 1.14
Modello equivalente del circuito di quenching passivo mostrato nella figura precedente; nel rettangolo si ha il circuito equivalente dello SPAD.

C_d è la capacità di giunzione dello SPAD (tipicamente vale qualche pF), C_s è la capacità parassita (pochi pF), R_d è data dalla serie della resistenza della carica spaziale e della resistenza della regione quasi neutra. Il valore di R_d dipende dalla struttura del dispositivo: va da un valore di circa 500 Ω per dispositivi con un'area dello strato svuotato ampia e molto spessa, fino ad alcuni kΩ per dispositivi con un'area piccola e giunzione molto stretta.

L'innesco del processo di moltiplicazione a valanga corrisponde alla chiusura dell'interruttore nel circuito equivalente; da quest'istante la corrente di valanga scarica rapidamente le capacità C_d e C_s facendo decrescere esponenzialmente la corrente I_d e la tensione V_d del diodo rispettivamente ad I_f e V_f.
Si hanno le seguenti relazioni:

$$I_d(t) = \frac{V_d(t) - V_B}{R_d} = \frac{V_E(t)}{R_d} \tag{1.37}$$

$$I_f = \frac{V_A - V_B}{R_d + R_L} \cong \frac{V_E}{R_L} \tag{1.38}$$

$$V_f = V_B + R_d \cdot I_f \tag{1.39}$$

Le approssimazioni sono giustificate dal fatto che $R_L >> R_d$.

Affinché lo SPAD si spenga, la corrente I_f deve essere minore di quella di latching I_q (che di solito vale 100 μA, negli SPAD sottili) **[14]**. La corrente di latching I_q rappresenta il valore della corrente I_d, sotto al quale la valanga non riesce ad autosostenersi, poiché pochi portatori attraversano la regione svuotata.

Così, quando I_d tende asintoticamente ad $I_f < I_q$ e raggiunge il valore I_q, il fenomeno di moltiplicazione a valanga mediamente si esaurisce spegnendo la valanga e la tensione che cade sul diodo (fra i nodi A e B) raggiunge il valore dato dalla seguente relazione:

$$V_q = V_B + R_d \cdot I_q \tag{1.40}$$

A questo l'interruttore si apre.
Da quanto appena detto si capisce che il valore di R_L non può essere troppo piccolo altrimenti I_f non potrà scendere sotto al valore di I_q e quindi la valanga non si auto-

spegnerà. Dalla 1.38 risulta che R_L deve aumentare di circa 10 kΩ per ogni volt di eccesso nella sovratensione $V_E=(V_A-V_B)$.

Osserviamo adesso il tempo di spegnimento della valanga: questo tempo è costante e dipende dalla capacità totale $C_d + C_s$ e dalla resistenza data dal parallelo fra R_L e R_d quindi in pratica semplicemente da Rd, poiché $R_L >> R_d$.

$$T_q = (C_d + C_s) \cdot \frac{R_d \cdot R_L}{(R_d + R_L)} \cong (C_d + C_s) \cdot R_d \qquad (1.41)$$

Quando il processo di moltiplicazione a valanga termina, nel circuito equivalente dello SPAD l'interruttore si apre, in questo modo le capacità si ricaricano lentamente per mezzo della piccola corrente che scorre nella resistenza di carico R_L e la tensione ai capi del diodo aumenta con costante di tempo data da:

$$T_r = R_L \cdot (C_d + C_s) \qquad (1.42)$$

Da questa relazione si osserva che R_L non può assumere valori troppo elevati, poiché ciò comporterebbe un tempo di recupero troppo lungo, durante il quale lo SPAD sarebbe insensibile all'arrivo dei fotoni. Di solito, dopo circa $5T_r$ si assume che la tensione ai capi del diodo riprende nuovamente il valore $V_A > V_B$ e, quindi, lo SPAD è pronto a rivelare l'arrivo di un nuovo fotone.

Occorre, così, trovare un compromesso nella scelta del valore di R_L tra la 1.40, che definisce la $I_f < I_q$, condizione necessaria per l'auto-spegnimento e la 1.42 che definisce il tempo di recupero T_r.

Nella figura seguente (1.15) è mostrato un grafico che illustra le forme d'onda dei segnali e di tensione V_d e di corrente I_d sul diodo quando la valanga si auto-spegne.

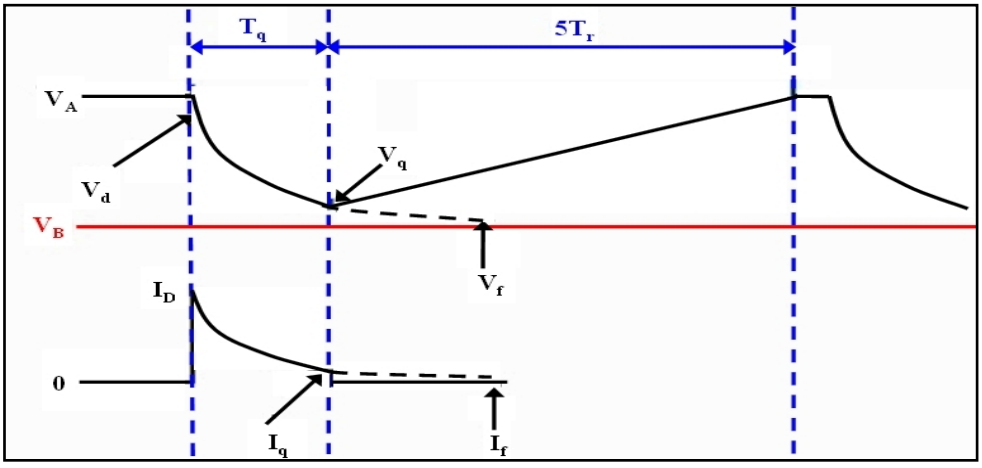

Figura 1.15

Andamento della tensione V_d e della corrente I_d quando la valanga si auto-spegne.

Nelle applicazioni di *photon counting*, i fotoni statisticamente distribuiti nel tempo che arrivano durante il tempo di recupero sono rivelati con minore efficienza, dato che l'overvoltage si riporta con grande lentezza al valore imposto dal circuito di polarizzazione esterno e produrranno un segnale di carica inferiore rispetto a quello standard peggiorando la risoluzione degli spettri di carica. Analogo discorso vale anche per i portatori generati termicamente e per afterpulsing durante il tempo di recupero. Inoltre, rispetto all'arrivo del fotone, il fronte di corrente di valanga generato presenta un ritardo con fluttuazioni attorno al valore medio nettamente decrescenti al crescere della tensione d'eccesso V_E. Nel rilevamento del tempo d'arrivo di fotoni, (photon timing) si hanno perciò gravi imprecisioni per gli eventi rivelati durante il recupero da eventi precedenti. Durante il tempo di recupero, la tensione d'eccesso V_E non è costante e quindi anche i ritardi del fronte di corrente non sono costanti. Quindi, il PQC riduce molto le prestazioni dello SPAD, limitando la loro applicazione pratica, in particolare nei casi in cui è richiesta un'elevata risoluzione temporale o si opera ad alte frequenze. L'unico vantaggio immediato del circuito di quenching passivo è la sua relativa semplicità di realizzazione.

La tensione ai capi di R_s dovuta al passaggio della corrente di valanga è data dalla seguente relazione:

$$V_s = V_E \cdot \frac{R_s}{R_d \cdot \left(1+\frac{C_d}{C_s}\right)} \cong I_f \cdot R_s \cdot \frac{R_L}{R_d \cdot \left(1+\frac{C_d}{C_s}\right)} \tag{1.43}$$

L'energia dissipata E_{pd} dallo SPAD durante un impulso di valanga è uguale alla diminuzione d'energia immagazzinata nella capacità C_s+C_d, quindi:

$$E_{pd} = \frac{1}{2}\cdot(C_d + C_s)\cdot(V_B + V_E)^2 - \frac{1}{2}\cdot(C_d + C_s)\cdot V_B^2 = (C_d + C_s)\cdot V_E \cdot \left(V_B + \frac{V_E}{2}\right) \cong$$

$$\cong (C_d + C_s)\cdot V_E \cdot V_B = Q_{pc}\cdot\left(V_B + \frac{V_E}{2}\right) \cong Q_{pc}\cdot V_B \tag{1.44}$$

dove Q_{pc} è la carica totale dell'impulso di valanga.

La potenza media dissipata P_d dallo SPAD è data da:

$$P_d = n_t \cdot E_{pd} \tag{1.45}$$

dove n_t è il numero di conteggi al secondo.

1.4.2 Circuiti di quenching attivi (AQC)

I limiti presenti nel sistema passivo sono superati con l'introduzione di un circuito di quenching attivo con cui è possibile l'applicazione degli SPAD su scala più ampia. Si tratta di un vero e proprio circuito integrato su silicio (quindi molto più grande, complesso e costoso della semplice resistenza in polisilicio del circuito di quenching passivo).

Il suo principio di funzionamento consiste nel polarizzare lo SPAD con una sorgente di tensione tramite una bassa resistenza, associata ad un circuito che ne "senta" la corrente di valanga e ne comandi, di conseguenza, (tramite un circuito driver di tensione) lo spegnimento, abbassando la tensione ai capi dello SPAD al di sotto della tensione di breakdown.

I principali vantaggi offerti dal circuito di quenching attivo sono: la veloce transizione dallo stato spento allo stato operativo e viceversa con una breve e ben definita durata del tempo della corrente di valanga e del tempo morto.

Esistono diverse varianti del circuito di quenching attivo **[15]**, in questa sede ne saranno trattate due differenti tipologie: una con il terminale di quenching (il terminale che opera sullo SPAD abbassandone la tensione) opposto al terminale sensing (il terminale che legge la corrente in uscita dalla SPAD), rappresentata nella figura 1.16, e l'altra con terminali di quenching e di sensing non opposti, illustrata nella figura 1.17.

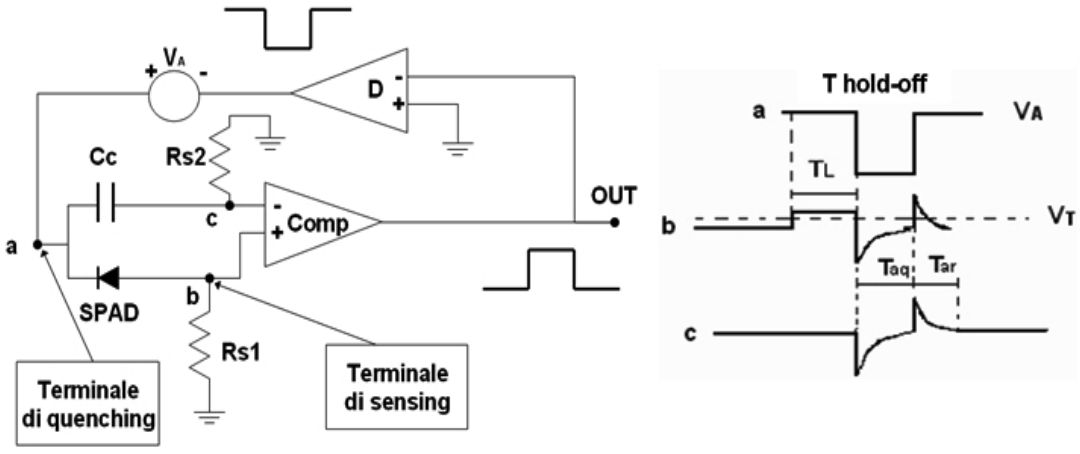

Figura 1.16

Schema semplificato di un AQC con terminali di quenching e sensing opposti.
Le forme d'onda delle tensioni sono relative ai nodi del circuito segnati con la stessa lettera.

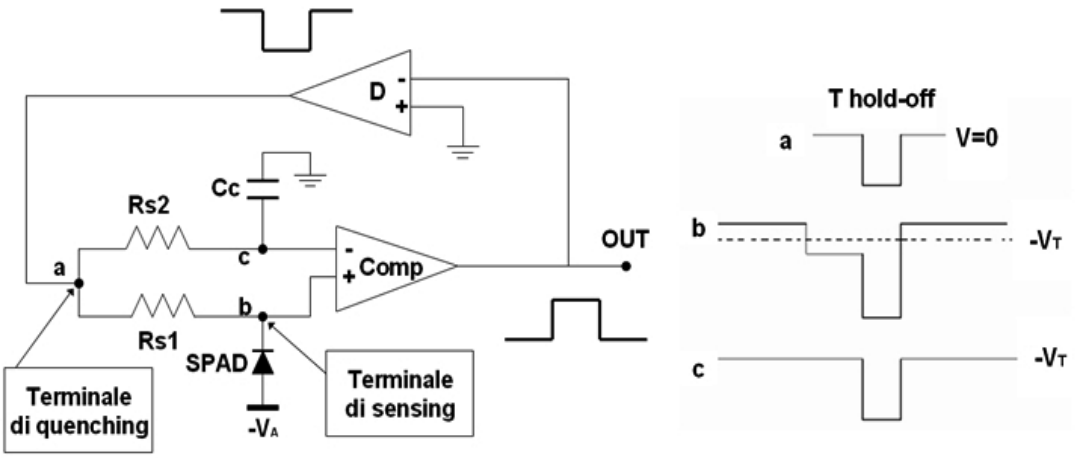

Figura 1.17

Schema semplificato di un AQC con terminali di quenching e sensing non opposti.
Le forme d'onda delle tensioni sono relative ai nodi del circuito segnati con la stessa lettera.

Nella configurazione con terminali di quenching e di sensing opposti (Figura 1.16), l'impulso di quenching è sovrapposto alla tensione V_A di polarizzazione dello SPAD. Tale configurazione non è adatta per SPAD con tensione di breakdown superiore ai 30 V per le difficoltà nella progettazione del circuito di drive che deve erogare l'impulso spegnente. Un ulteriore problema è dovuto all'iniezione di un impulso di corrente nell'ingresso dell'AQC da parte della capacità di giunzione del rivelatore quando lo SPAD viene riacceso dal circuito di quenching. Tale impulso ha la stessa polarità dell'impulso di valanga ed ampiezza comparabile, tanto da reinnescare il circuito di quenching e stabilizzarlo in una situazione d'oscillazione costante. Ad esempio, se C_d = 1 pF e V_E = 2 V con un tempo di transizione di 2 ns, la corrente iniettata è di circa 1 mA. Il problema è in parte superato dalla rete RC, formata dal resistore R_{s2} e dal capacitore C_c, connessa ad un terminale del comparatore, come mostrato in figura 1.17.

I problemi di adattabilità, in termini di tensione di breakdown, ad ogni SPAD del circuito di quenching sono superati con il circuito che ha i terminali di quenching e di sensing non opposti (Figura 1.17), poiché, in questa configurazione, uno dei terminali dello SPAD non è connesso a nessun punto del circuito di AQC ed è possibile applicare una qualsiasi tensione in continua richiesta dal dispositivo. Anche in questo caso la rete RC ha lo scopo di compensare gli impulsi di corrente introdotti dalla capacità di giunzione dello SPAD.

Sono ora descritte le principali caratteristiche dei circuiti di quenching attivo:

- T_{ac} è la durata dell'impulso di valanga: dipende dal *tempo di circolazione* T_L nel loop di quenching e dal *tempo di salita* T_{aq} dell'impulso di quenching:

$$T_{ac} = T_L + T_{aq} \tag{1.46}$$

  T_{ac} può essere minimizzato abbassando *la tensione di overvoltage* V_E. Infatti, T_{aq} cresce con l'ampiezza dell'impulso di quenching; il minimo valore T_{ac} raggiunto è di 5 ns con V_E più piccolo di 1 V e di 10 ns con attorno a 20 V.

- T_{ad} è il *tempo morto*, ovvero il minimo valore è dato dal doppio del tempo di circolazione $T_{L,}$ dal tempo di salita T_{aq} e dal tempo di discesa T_{ar} dell'impulso di quenching:

$$\left[T_{ad}\right]_{\min} = 2T_L + T_{aq} + T_{ar} \tag{1.47}$$

  Anche T_{ad} cresce con la sovratensione V_E: il minimo valore ottenuto è di circa 10 ns lavorando con una V_E sotto 1 V, mentre è di circa 40 ns con V_E vicino ai 20 V.

- T_{ar} é il *tempo di recupero*: deve essere necessariamente piccolo per avere una bassa probabilità di piccoli impulsi durante il tempo di recupero; questi impulsi devono aversi solo in corrispondenza del tempo di recupero, in modo tale da poter essere facilmente riconosciuti e scartati da un circuito apposito.

- Tempo di *hold-off*: è creato facilmente con un semplice circuito che mantiene bassa la tensione per un tempo abbastanza lungo; quest'intervallo permette il rilascio degli elettroni intrappolati nella regione svuotata, evitando così il fenomeno dell'afterpulsing.

- E_{ad} è l'*energia dissipata* dallo SPAD durante un impulso di valanga ed è data dalla formula seguente:

$$E_{ad} = Q_{ac} \cdot (V_B + V_E) \cong Q_{ac} \cdot V_B \tag{1.48}$$

  in cui Q_{ac} è la carica totale dell'impulso di valanga data da:

$$Q_{ac} = \frac{V_E}{R_d} \cdot \left(T_L + \frac{T_{aq}}{2} \right) \tag{1.49}$$

1.5 Il SiPM

Il **SiPM** o *Silicon Photo-Multplier* è, oggi, l'ultimo arrivato nella famiglia dei fotomoltiplicatori al silicio. Nasce, infatti, da un brevetto di Z. Sadygov nel 1996 ed è attualmente un dispositivo in fase di studio e sviluppo **[3]**.

Il SiPM consiste in una matrice planare di più SPAD, operanti in Geiger mode identici in forma, dimensioni e caratteristiche costruttive, connessi in parallelo (ovvero con i catodi e gli anodi rispettivamente in comune) ed operanti su un carico comune (R_{carico}). L'intento è coprire un'ampia area sensibile utilizzando tanti piccoli SPAD, aventi ognuno migliori prestazioni in termini di dark count e timing. Ciascuno SPAD della matrice è, poi, dotato di un resistore di quenching integrato R_L (avente valore identico per ogni singolo sensore), che, oltre a svolgere la funzione di spegnere la valanga portando la tensione sul singolo dispositivo al valore di breakdown, funge da disaccoppiamento elettrico fra uno SPAD e l'altro, permettendo loro di operare come se fossero indipendenti nonostante l'alimentazione ed il carico in comune. Il resistore di quenching è realizzato, di solito, in polisilicio o tramite strutture "più evolute", come la **MRS** (*Metal-Resistor-Semiconductor*). La scelta del PQC, nonostante i suoi limiti, già evidenziati, è giustificata dal fatto che costruire un circuito di quenching attivo per ogni singolo SPAD della matrice sarebbe un'operazione troppo complessa e costosa occupando, inoltre, troppo spazio sulla fetta di silicio a scapito dell'area attiva di ogni dispositivo e, come si vedrà in seguito, della sua efficienza di rivelazione. L'insieme costituito dalla serie SPAD-resistore di quenching è denominato, in letteratura, **"pixel"** o **"microcella" [16]**. Oggigiorno esistono SiPM con una densità pari a circa 1000 pixel per millimetro quadro **[17]**.
Segue lo schema circuitale del SiPM.

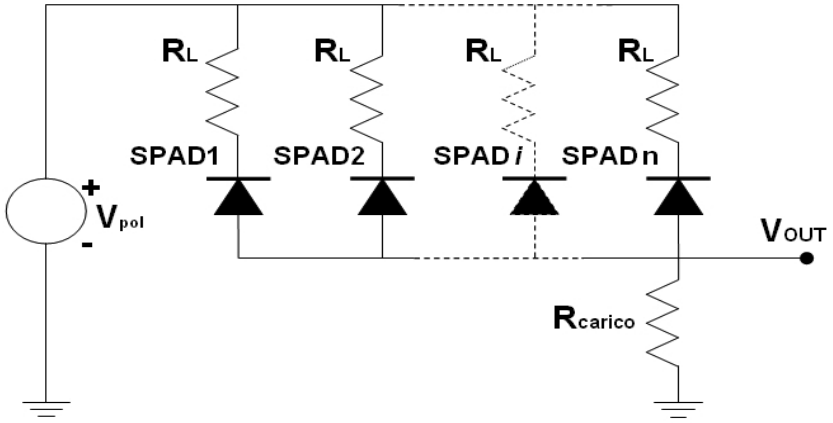

Figura 1.18
Schema circuitale di un SiPM

L'uscita è comune per tutti i pixel ed è costituita dalla somma delle cariche emesse dalle singole microcelle "accese" dall'assorbimento di un fotone. Se tutti i pixel sono identici ed emettono ciascuno la stessa quantità di carica quando assorbono un fotone, misurando la carica totale in uscita, si può risalire al numero di pixel accesi. Infatti, con l'ipotesi fatta d'uniformità di pixel, la carica in uscita è un multiplo, pressoché intero, della carica emessa dalla singola microcella accesa: da qui si ricava il numero di fotoni assorbiti. Da quanto detto, il singolo pixel si comporta come un dispositivo digitale (in quanto emette una quantità predefinita di carica quando rivela un fotone), mentre l'intero SiPM è un dispositivo analogico, poiché in uscita si vede la somma di questi impulsi di carica. Restano, purtroppo, limitanti problemi legati al dark count, all'afterpulsing, all'insensibilità dello SPAD durante l'Hold-off, alle interferenze ottiche ed elettriche fra pixel (il cosiddetto *"cross talk")*, alle tolleranze di processo, alla disuniformità locale fra pixel ed al fatto che una singola microcella in uscita emette la stessa quantità di carica, anche se in ingresso ha assorbito simultaneamente più di un fotone. É importante osservare, inoltre, che, essendo il carico comune, in uscita non si avrà nessuna informazione su quali pixel hanno rivelato i fotoni e, quindi, sulla posizione in cui è avvenuto l'assorbimento. A causa di ciò, non si può, purtroppo, usare il SiPM per imaging.

Anche l'alimentazione è in comune e semplifica di molto la struttura del SiPM: infatti, sono necessari solo due pad, uno per gli anodi e uno per i catodi di ogni microcella, per alimentare tutti quanti gli SPAD della matrice.
Segue un'immagine tridimensionale della struttura del SiPM.

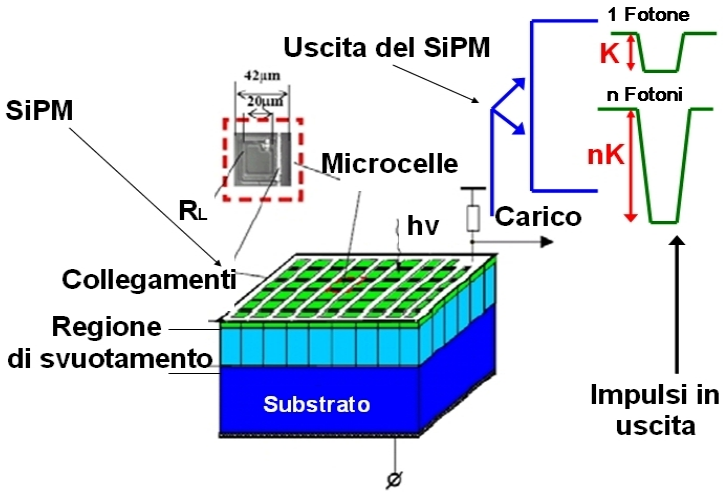

Figura 1.19
Struttura del SiPM

Di seguito saranno presentate le caratteristiche vincolanti sulle prestazioni del SiPM, come il guadagno, l'efficienza di rivelazione, la dinamica ed il rumore.

1.5.1 Il Guadagno

Il guadagno è un parametro molto importante per il SiPM. Per un singolo pixel, esso è definito come il rapporto fra la carica totale Q_{pixel} che attraversa la giunzione del pixel, quando viene rivelato un fotone, e la carica elementare q **[17]**.

$$G_{pixel} = \frac{Q_{pixel}}{q} = C_{pixel} \cdot \frac{(V_{pol} - V_B)}{q} \tag{1.50}$$

dove C_{pixel} è la capacità intrinseca di un singolo pixel, V_{pol} la tensione di alimentazione del SiPM e V_B la tensione di breakdown. Il guadagno di una microcella, quindi, dipende dall'overvoltage e dalla capacità intrinseca e rappresenta il numero medio di portatori che attraversano la giunzione durante una valanga.

É importante che tale guadagno sia uniforme per tutti i pixel al fine di poter facilmente "decifrare" il segnale in uscita dal SiPM nel photon-counting. Purtroppo, la tensione di rottura V_B e la capacità intrinseca C_{pixel} sono, spesso, parametri variabili con poca uniformità fra pixel e pixel. Inoltre, alla capacità intrinseca C_{pixel} potrebbero aggiungersi eventuali contributi capacitivi parassiti, dovuti, ad esempio, alle piste di metal o al bonding, che potrebbero influire sull'uniformità del guadagno fra pixel. In più, la tensione di breakdown V_B è variabile con la temperatura, quindi, dalla 1.50, anche il guadagno del singolo pixel è dipendente dalla temperatura.
Gli unici parametri, quindi, su cui è possibile agire concretamente dall'esterno, per variare il guadagno e determinarlo, sono la tensione di alimentazione V_{pol} e le dimensioni dei pixel (da cui dipende la capacità intrinseca C_{pixel}): aumentando tali parametri (o anche solo uno di essi) si aumenterà il guadagno e viceversa. Per il calcolo della carica totale in uscita dal SiPM occorre misurare la tensione in uscita sulla resistenza di carico R_{carico}, dividerla per il valore di tale resistenza ed integrare la corrente di uscita ottenuta rispetto al tempo per la durata del segnale.

$$Q_{tot} = \int_{t_1}^{t_2} \frac{V_{out}(t)}{R_{carico}} dt \tag{1.51}$$

in cui t_1 e t_2 sono gli estremi dell'intervallo di tempo in cui è applicato il segnale.

Dividendo la carica totale in uscita ricavata Q_{tot} per la carica emessa da un singolo pixel Q_{pixel} si ricava il numero di pixel "accesi" e, quindi, il numero dei fotoni rivelati, sempre nell'ipotesi che la carica Q_{pixel} sia uniforme **[17]**.

1.5.2 Efficienza di rivelazione di fotoni

L'*efficienza di rivelazione di fotoni* o *photon detection efficiency* (PDE) esprime la capacità del SiPM di rivelare fotoni. Essa è data dal prodotto di tre fattori:

$$PDE = QE \cdot \varepsilon_{geom} \cdot \varepsilon_{Geiger} \tag{1.52}$$

dove QE è l'*efficienza quantica*, ε_{geom} un fattore geometrico e ε_{Geiger} la probabilità che un fotoelettrone inneschi la valanga (*probabilità di trigger*) **[16]**.

L'*efficienza quantica* QE è le probabilità che un fotone incidente sia assorbito; essa è data da:

$$QE = (1 - R) \cdot (1 - e^{\alpha \cdot x}) \tag{1.53}$$

in cui 1-R sarebbe il coefficiente di trasmissione T per il sistema aria-ossido-silicio, mentre R è il coefficiente di riflessione per lo stesso sistema (si osservi che R+T=1 con $0 \leq R \leq 1$ e $0 \leq T \leq 1$), α è il coefficiente di assorbimento del silicio, che dipende fortemente, come già visto nel paragrafo 1.3.3.4, dalla lunghezza d'onda della radiazione elettromagnetica incidente, e x la variabile che identifica l'effettivo spessore della superficie di contatto. Per migliorare l'efficienza quantica occorre abbassare il valore del coefficiente di riflessione R utilizzando un rivestimento antiriflettente sulla superficie esterna del dispositivo ed aumentare lo spessore della superficie più esterna (soprattutto per migliorare l'assorbimento per la luce rossa, il cui coefficiente di assorbimento α è più basso e la cui lunghezza d'onda è maggiore).

Il fattore geometrico ε_{geom}, detto anche *fattore geometrico* o *fill factor*, è il rapporto fra l'area attiva (o area sensibile) A_{act} e l'area totale A_{tot} del dispositivo:

$$\varepsilon_{geom} = \frac{A_{act}}{A_{tot}} \tag{1.54}$$

Per aumentare tale fattore occorre incrementare l'area attiva del dispositivo **[16]**.

La probabilità che un fotoelettrone inneschi la valanga ε_{Geiger} è espressa dalla relazione 1.22 e, come detto, dipende principalmente dall'overvoltage applicato e dalla temperatura.

Attualmente, l'efficienza di rivelazione dei fotoni è di circa il 24% per la luce verde **[18]**.

1.5.3 Dinamica

La *dinamica* in un SiPM è definita come il massimo numero di fotoni che possono essere rivelati simultaneamente dal dispositivo. Tale caratteristica è fortemente limitata dal numero finito di pixel contenuti in un SiPM. Infatti, come anticipato in precedenza, un pixel emette la stessa quantità di carica anche se assorbe contemporaneamente più di un fotone, quindi, anche se ne arrivano nello stesso tempo più di uno ne rileveremo in uscita solamente uno. Inoltre, occorre osservare che una microcella, durante il "tempo morto" (il tempo in cui, durante il quenching, la sovratensione sul pixel aumenta progressivamente dal valore di breakdown a quello di polarizzazione), rivela fotoni con scarsa efficienza, come già osservato in precedenza, producendo un segnale di carica inferiore rispetto a quello standard e peggiorando, così, la risoluzione degli spettri di carica. Per questo motivo il SiPM funziona molto bene solo quando il numero medio di fotoni in ingresso per pixel è molto piccolo (di solito inferiore a 2, con una buona efficienza di rivelazione di fotoni). Se, invece, tale media è un numero molto alto, il segnale in uscita dal dispositivo (ovvero il numero di pixel accesi $N_{pixel\text{-}fired}$) saturerà al numero di pixel del SiPM. L'andamento di tale saturazione può essere calcolato supponendo, come poi del resto è in realtà, poissoniana la dispersione σ dei fotoni che arrivano sulla superficie del nostro dispositivo, cioè:

$$\sigma = \sqrt{N_{fotoni} \cdot PDE} \tag{1.55}$$

dove N_{fotoni} è il numero di fotoni che arrivano sulla superficie del SiPM e PDE è la sua efficienza di rivelazione di fotoni **[16]**.

La relazione che si ricava da ciò ed esprime la saturazione è:

$$N_{pixel-fired} = m \cdot \left(1 - e^{\frac{-N_{fotoni} \cdot PDE}{m}} \right) \tag{1.56}$$

in cui $N_{pixel\text{-}fired}$ è il numero di pixel accesi, m il numero di pixel, N_{fotoni} il numero di fotoni che arrivano sulla superficie del SiPM e PDE la sua efficienza di rivelazione di fotoni. In questa formula il rapporto N_{fotoni}/m esprime il numero medio di fotoni in ingresso per microcella. Inoltre si deve osservare che tale relazione non tiene conto degli effetti del rumore che alterano notevolmente il conteggio di fotoni **[17]**.

Per poter apprezzare meglio la saturazione del segnale in uscita, sono proposti due grafici che mettono in relazione il numero di pixel accesi con il numero medio di fotoni per pixel applicando la 1.55. Nel primo grafico (figura 1.20) si è supposto una PDE=100% (valore ideale), mentre nel secondo (figura 1.21) una PDE=20% (valore più attendibile con le attuali tecnologie). In entrambi i grafici si è poi posto m=500.

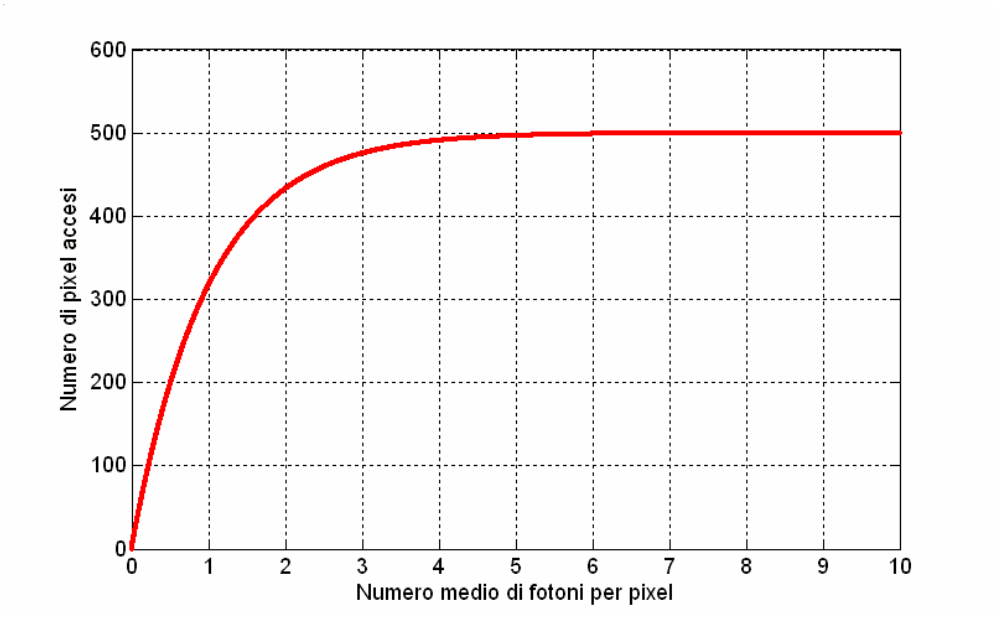

Figura 1.20
Curva di saturazione per PDE=100% e m=500

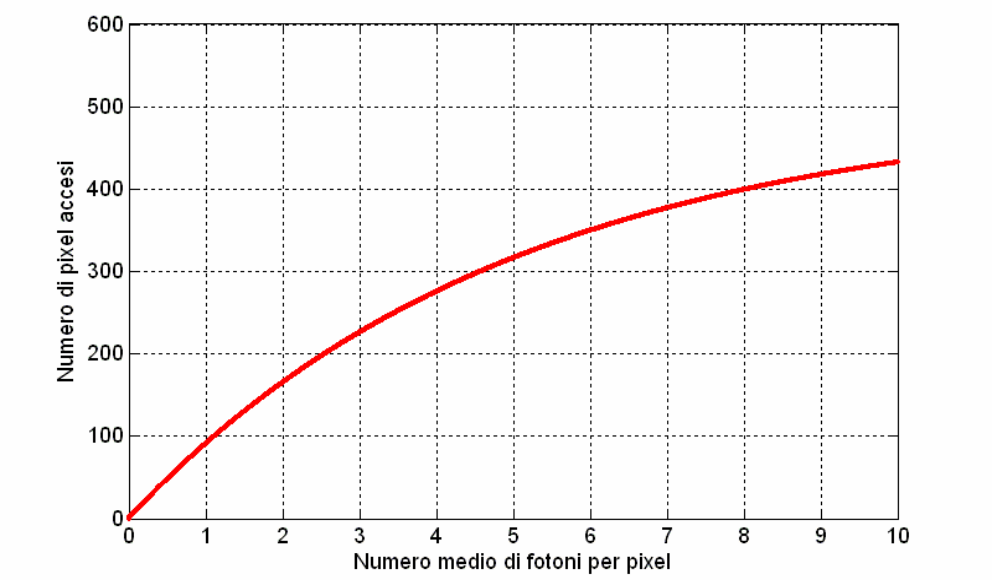

Figura 1.21
Curva di saturazione per PDE=20% e m=500

Dai due grafici si evince che la curva saturi più lentamente per valori di PDE più bassi e che tale curva possa essere linearizzata per basse medie di fotoni per pixel.

1.5.4 Rumore e cross talk

Essendo il SiPM costituito da SPAD, i problemi di rumore già trattati per lo SPAD, come il **dark count** e l'**afterpulsing**, valgono anche per il SiPM.

A questi problemi, purtroppo, se ne aggiunge un altro, che ancora non è stato trattato in questo lavoro di tesi, quello del **cross talk**. Tale rumore è tipico di tutte le matrici integrate di fotomoltiplicatori (quindi anche del SiPM) e si presenta quando due o più pixel, della stessa matrice, interferiscono (diafonia) fra loro. Tale interferenza può essere di due tipi: ottica, si parla, allora, di *cross talk ottico*, oppure elettrica, si parla, in questo caso, di *cross talk elettrico*.

Più dettagliatamente, il *cross talk ottico* si verifica quando, durante la rottura della giunzione dovuta alla rivelazione di un fotone da parte dell'area attiva, i portatori che formano la corrente inversa emettono fotoni, detti fotoni di Bremsstralhung, che possono propagarsi nel silicio come in una guida d'onda, raggiungere l'area attiva di un altro SPAD, venire assorbiti ed innescare, così, la moltiplicazione a valanga in un altro pixel del SiPM, generandovi un impulso spurio non correlato all'assorbimento del fotone.

Tale effetto è illustrato in figura 1.22.

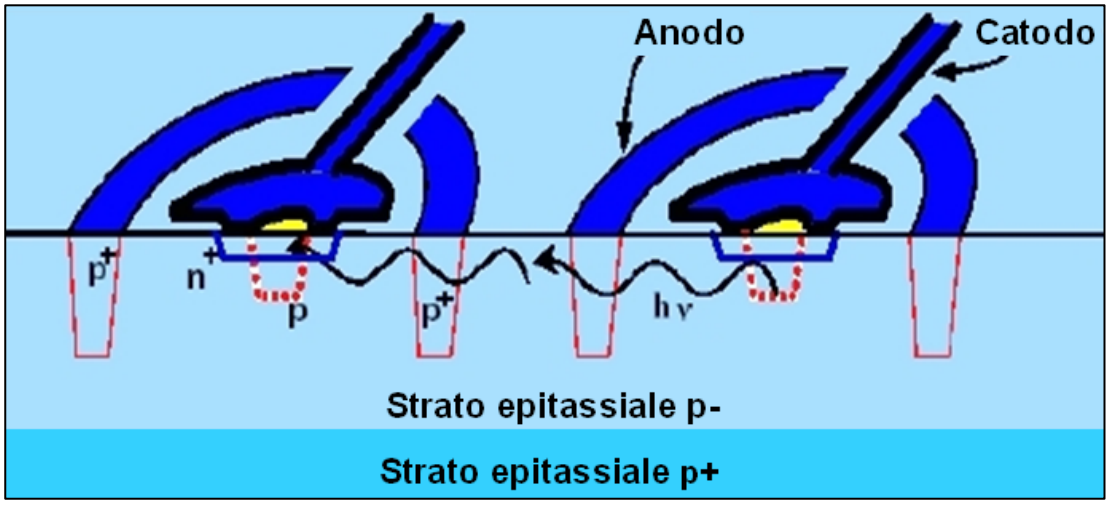

Figura 1.22
Illustrazione schematica del cross talk ottico fra due pixel vicini.

La seguente equazione descrive la relazione che si ha fra la corrente inversa generata durante la valanga e il flusso di fotoni emesso **[19]**:

$$I_{photon} = \left(\frac{I_{reverse}}{q} \right) \cdot \gamma \qquad (1.57)$$

in cui $I_{reverse}$ è la corrente inversa che attraversa la giunzione, γ il coefficiente di emissione (circa $2{,}9 \cdot 10^{-5}$ per ogni elettrone che attraversa la giunzione) e q la carica dell'elettrone. Anche se, dal valore di γ, si ha un'emissione di fotoni molto bassa (un fotone emesso circa ogni 10^5 elettroni), la probabilità del cross-talk ottico tra gli elementi vicini risulta significativa a causa della estrema sensibilità degli SPAD che costituiscono i vari i pixel.

La figura seguente mostra lo spettro della luce emessa da un pixel polarizzato con una corrente $I_{reverse}$ di 100 μA e un coefficiente di emissione γ pari a $2{,}9 \cdot 10^{-5}$. Nel grafico sono messi in relazione il numero di fotoni emessi, la lunghezza d'onda e la percentuale alle varie lunghezze d'onda, rispetto al totale della luce emessa.

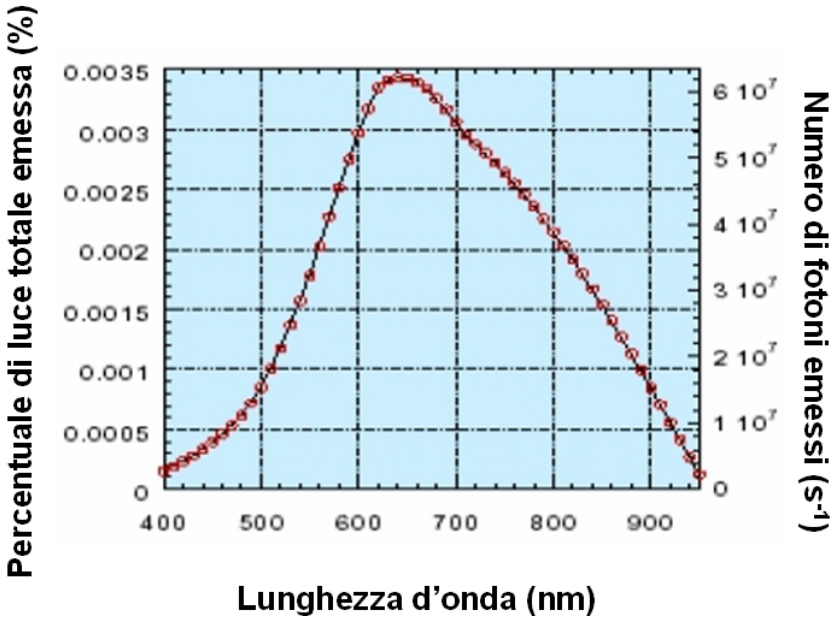

Figura 1.23

Grafico della luce emessa in percentuale e del numero di fotoni emesso in funzione della lunghezza d'onda, per un pixel polarizzato inversamente con una corrente di 100 μA.

Il pixel vicino assorbe il flusso di fotoni proveniente da un'altra microcella in base al coefficiente d'assorbimento α del silicio, riportato, precedentemente, in figura 1.12 in funzione della lunghezza d'onda.

Il flusso luminoso si attenua seguendo la relazione espressa dall'equazione:

$$I_{photon}(\lambda, d) = I_{photon}(\lambda, 0) \cdot e^{-\alpha(\lambda) \cdot d} \qquad (1.58)$$

dove λ è la lunghezza d'onda dei fotoni emessi dallo SPAD in valanga, α il coefficiente di assorbimento nel silicio e d la distanza percorsa dal flusso di fotoni.

Il *cross-talk elettrico*, invece, ha luogo quando dei portatori di carica emessi dalla giunzione del pixel sorgente diffondono attraverso la regione epitassiale di tipo p^+, comune a tutte le microcelle del SiPM, e, raggiungendo altri pixel vicini, innescano in essi il fenomeno di moltiplicazione a valanga generando impulsi spuri, non correlati con l'assorbimento di fotoni. Questo problema è così rilevante anche perché i pixel sono sensibili a basse quantità di corrente (nel nostro caso si tratta di qualche femtoampère), oltre a basse intensità di luce.

Esiste, infine, un terzo tipo di cross talk, definito come *de-trapping stimolato*, dovuto all'emissione di una radiazione elettromagnetica, come in un'antenna, da parte dei portatori di carica in moto durante una valanga in un pixel. Tale radiazione può propagarsi lungo il silicio, soprattutto lungo il substrato comune, raggiungendo i pixel vicini ed inducendo, se l'energia associata alla radiazione è sufficiente da far saltare un portatore dalla banda di valenza a quella di conduzione, la liberazione, in un pixel vicino, di un portatore intrappolato e potenzialmente in grado d'innescare un impulso di valanga spurio **[20]**.

Esistono due modi per cercare di eliminare, o almeno ridurre, il cross talk ottico ed elettrico. Il primo consiste nell'aumentare il **pitch** (la minima distanza fra i centri delle zone attive di due pixel contigui), il secondo, invece, prevede lo scavo fra un pixel e l'altro di **trincee** riempite di ossido e/o metallo in modo da realizzare un isolamento ottico-elettrico fra le microcelle.

Fra i due metodi, l'isolamento fra pixel tramite *trench* è, senza dubbio, il più efficace. Inoltre, l'uso delle trincee non riduce più di tanto l'area attiva dei pixel, per la relativa sottigliezza di tali trench, e, quindi, intacca pochissimo il *fill factor* e, di conseguenza, l'*efficienza di rivelazione di fotoni*. L'aumento del *pitch*, invece, fa scendere considerevolmente l'area attiva, il *fill factor* e, di conseguenza, anche l'*efficienza di rivelazione di fotoni.*

Nell'ambito del prossimo capitolo sarà illustrato un metodo per la realizzazione di queste trincee.

Fine del Capitolo 1

Bibliografia

[1] S. Barbarino - ***"Appunti di Campi Elettromagnetici - Anno Accademico 2003-2004"***.

[2] P. Mazzoldi - M. Nigro - C. Voci ***"Fisica Vol. II - Elettromagnetismo-Onde"*** - Seconda Edizione - Dipartimento di Fisica Galileo Galilei Padova.

[3] D. Renker ***"Presentation in Beaune 2005"***.

[4] S. Cova - F. Zappa - M. Ghioni ***"Fotorivelatori microelettronici ultrasensibili"*** - Politecnico di Milano Dipartimento di Elettronica e Informazione Milano, ALTA FREQUENZA - RIVISTA DI ELETTRONICA / VOL. 9 N. 1 / Gennaio-Febbraio 1997.

[5] Simon M. Sze ***"Dispositivi a semiconduttore"*** - *Comportamento fisico e tecnologia* - EDITORE ULRICO HOEPLI MILANO.

[6] A. S. Grove ***"Physics and technology of semiconductor devices"*** – J. Wiley and Sons, 1967.

[7] Simon M. Sze ***"Physics of semiconductor devices"*** – J. Wiley and Sons (1981).

[8] J. E. Bateman - S.R. Burge - R. Stephenson ***"Gain and Noise Measurement on Two Avalanche Photodiode Proposed for the CMS ECAL, CLRC"*** - Technical Report, RAL/TR-95-001 (1995)

[9] Richard S. Muller - Theodore I. Kamins ***"Dispositivi elettronici nei circuiti integrati" -*** nuova edizione BOLLATI BORINGHIERI.

[10] H. Dautet - P. Deschamps - B. Dion - Andrew D. MacGregor - D. MacSween - Robert J. McIntyre - C. Trottier and Paul P. Webb ***"Photon counting techniques with silicon avalanche photodiodes"*** - APPLIED OPTICS / Vol.32, No. 21 / 20 July 1993.

[11] E. Sciacca - A. C. Giudice - D. Sanfilippo - S. Lombardo - F. Zappa - R.Consentino - M. Mazzillo - M. Ghioni - G. Fallica - M. Belluso - G.Bonanno - S. Cova - E. Rimini ***"Silicon Planar Technology for Single-Photon Optical Detectors"*** - IEEE TRANSACTIONS ON ELECTRON DEVICES / VOL. 50 N. 4 / APRIL 2003.

[12] W.J. Kindt - H.W. van Zeijl ***"Modelling and Fabrication of Geiger mode Avalanche Photodiodes"*** - IEEE TRANSACTIOS ON NUCLEAR SCIENCE / VOL 45, N. 3, JUNE 1998.

[13] J. Millman - A. Grabel ***"Microelettronica"*** - McGraw-Hill (1987).

[14] R.H. Haitz, ***"Model for the electrical behavior of a microplasma"***, J. Appl.Phys.35, 1370-1376 (1964).

[15] S. Cova, M. Ghioni, A. Lacaita, C. Samori, and F. Zappa, ***"Avalance photodiodes and quenching circuits for single-photon detection"***- APPLIED OPTICS / Vol. 35 No. 12 / 20 April 1996.

[16] P. Buzhan, B. Dolgoshein, A. Ilyin, V. Kantserov, V. Kaplin, A. Karakash, A. Pleshko, E. Popova, S. Smirnov, Yu. Volkov, L. Filatov, S. Klemin and F. Kayumov ***"An advanced study of Silicon Photo-Multiplier"*** - ICFA Instrumentation Bulletin.

[17] J. Barral ***"Study of Silicon Photomultipliers"*** - Thesis - École Polytechnique, France - 13th April-2nd July 2004.

[18] V. Saveliev ***"The recent development and study of silicon phomultiplier"*** - Nuclear Instruments and Methods in Physics Research A 535 (2004) pag 528–532.

[19] J.C. Jackson - D. Phelan - A.P.Morrison - R.M. Redfern - A.Mathewson ***"Characterization of Geiger mode avalanche photodiodes for fluorescence***

decay measurements" - Proceedings of SPIE / Vol. 4650-07 / Photonics West, San Jose, CA, January 19-25, 2002.

[20] P. Finocchiaro - A. Campisi - L. Cosentino - A. Pappalardo - F. Musumeci - S. Privitera - A. Scordino - S. Tudisco - G. Fallica - D. Sanfilippo - M. Mazzillo - A. Piazza - J. Van Erps - S. Van Overmeiere - M. Vervaeke - B. Volckaerts - P. Vynck - A. Hermannes - H. Thienponts - S. Lombardo - E. Sciacca ***"SPAD arrays and micro-optics: towards a real single photon spectrometer"*** - in print on JOURNAL OF MODERN OPTICS.

CAPITOLO 2

PROGETTAZIONE E FABBRICAZIONE DI DISPOSITIVI SINGOLI E DI MATRICI PLANARI

2.1 Strategie di progettazione di SPAD

La letteratura riporta due differenti strategie di progettazione di rivelatori di singolo fotone. La prima di queste è proposta da *McIntyre, P. P. Webb* e *J. Conradi* e sviluppata nei laboratori di ricerca industriale della *EG&G Optoelectronics* (oggi *Perkin Elmer Optoelectronics*). Essi progettano dispositivi ad alta efficienza di rivelazione su un range spettrale molto ampio. La seconda di queste è stata, invece, proposta e sviluppata dal gruppo del *prof S. Cova* del Politecnico di Milano.

La principale differenza fra queste due strategie risiede nella progettazione dello spessore della regione di svuotamento. Infatti, i primi hanno una zona svuotata molto spessa, pari all'intero spessore del wafer (circa 120 μm), mentre i dispositivi del prof S. Cova hanno una regione svuotata molto più sottile (di pochi micron).

I dispositivi progettati da McIntyre, P. P. Webb e J. Conradi hanno lo spessore dello strato di svuotamento di circa 30-40 μm, l'area attiva con circa 500 μm e tensione di breakdown V_B superiore ai 100 V. Di conseguenza tali dispositivi hanno un'efficienza molto alta, poiché è legata allo spessore della regione svuotata del dispositivo, ma non sono compatibili con le attuali tecnologie planari, non sono integrabili in matrici, la loro alta tensione d'alimentazione comporta complicazioni alla circuiteria di controllo e, poiché dissipano troppa potenza (circa 10 W per ogni impulso di valanga), necessitano di un'apparecchiatura esterna per il raffreddamento.

Il secondo tipo di dispositivi (quelli del prof S. Cova) sono, invece, compatibili con la tecnologia planare CMOS: sono, infatti, molto sottili (lo spessore del loro strato di svuotamento va da 1 μm a 4 μm), con diametro d'area attiva compreso fra i 10 μm e i 100 μm e tensione di breakdown V_B inferiore ai 35 V. Tali dispositivi sono progettati per rivelare, con un'elevata precisione, il tempo d'arrivo dei fotoni con un esiguo valore di dark count rate, a temperatura ambiente, e bassa

dissipazione di potenza (pochi milliwatt per impulso rivelato), per quest'ultimo motivo non hanno bisogno di raffreddamento esterno **[1]**.

La figura seguente illustra l'efficienza di rivelazione di singolo fotone per SPAD spessi e sottili, confrontandoli con tubi fotomoltiplicatori PMT, al variare della lunghezza d'onda.

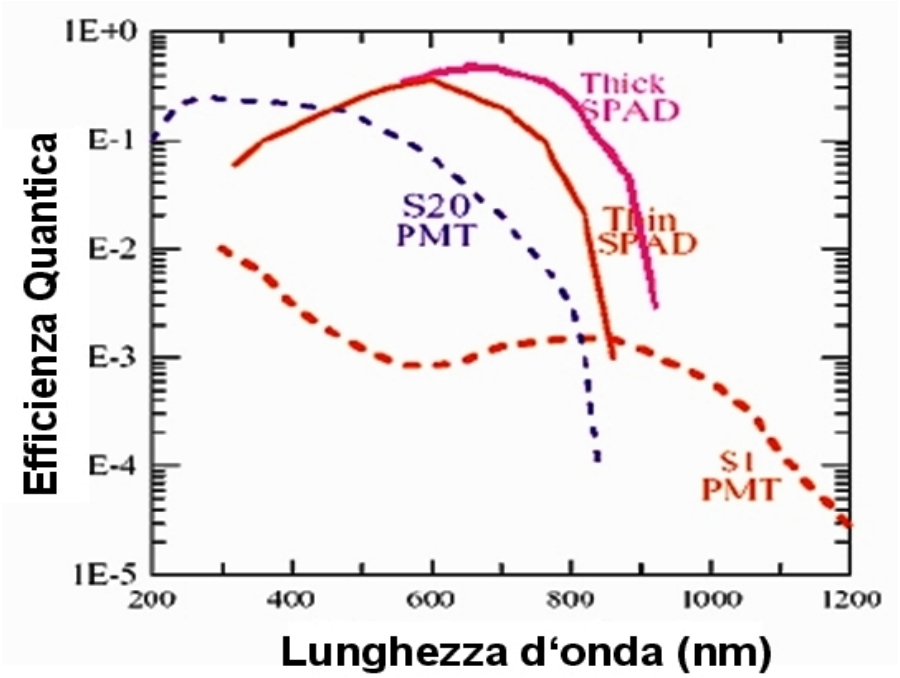

Figura 2.1
Efficienza di rivelazione di SPAD, spessi e sottili, in silicio e di fotomoltiplicatori impiegati nella regione spettrale del visibile (con fotocatodo S20) e nel vicino infrarosso (S1).

Come si nota l'efficienza di rivelazione per SPAD spessi si mantiene sopra il 50% fino a 900 nm di lunghezza d'onda per abbassarsi, poi, sotto il 3% intorno ai 1064 nm. La precisione che si raggiunge nel determinare il tempo d'arrivo di un fotone per SPAD spessi è abbastanza buona fra 150 ps e 350 ps, considerando il FWHM (acronimo dell'inglese *Full Width at Half Maximum*, *Massima Ampiezza a Metà Massimo*) a seconda del diametro del dispositivo.

Gli SPAD sottili hanno un'efficienza quantica nel visibile più bassa rispetto ai primi, ma in ogni caso migliore rispetto ai "vecchi" PMT. La precisione nel determinare il tempo d'arrivo d'un fotone per SPAD sottili è inferiore ai 20 ps (adatta al *photon timing*).

In questa tesi saranno trattati SPAD sottili (appartenenti, quindi, alla seconda tipologia) realizzati dal gruppo **R&D** della **STMicroelectronics** di **Catania**.

2.2 Problematiche e soluzioni tipiche nella fase di progettazione di uno SPAD singolo

Come già detto nel precedente capitolo, lo SPAD è una giunzione p-n polarizzata inversamente sopra la tensione di breakdown. Quando un fotone colpisce la giunzione si ha una certa probabilità che generi una coppia elettrone-lacuna **[2]**.

Occorre, però, sottolineare che vi sono due possibili zone del dispositivo in cui può essere generata la coppia elettrone-lacuna:

- *La regione neutra:* in questo caso, i portatori minoritari generati possono diffondere verso la regione svuotata ed innescare il processo di moltiplicazione a valanga solo quando raggiungono la regione svuotata. Infatti, solo nella regione di svuotamento l'intensità del campo elettrico è tale da permettere il processo di moltiplicazione a valanga. Tuttavia, è possibile che le cariche si ricombinino prima di raggiungere la zona svuotata ed il fotone assorbito, in questa situazione, non viene rivelato.
- *La zona di svuotamento*: in questa zona si ha un forte campo elettrico tale da permettere alle coppie di portatori di separarsi ed avere un'energia tale da innescare il processo di moltiplicazione a valanga.

Nel secondo caso si ha, dunque, una probabilità più alta d'innesco della valanga. Per questo motivo il dispositivo deve essere progettato in maniera tale che il fotone incidente, per essere rivelato, sia assorbito nella regione di svuotamento.

Un altro aspetto da tenere in considerazione nel progetto della regione svuotata è quello della tensione di breakdown ai bordi di tale regione. Infatti, per l'effetto della curvatura della giunzione, il forte campo elettrico che si genera nella zona di svuotamento risulta essere più grande ai suoi bordi, provocando, quindi, la rottura della giunzione a tensioni di polarizzazione più basse di quelle previste nella regione planare. Inoltre, la *"rottura ai bordi"* risulta essere più evidente e problematica in regioni molto sottili e con piccoli raggi di curvatura: proprio, purtroppo, il caso dei dispositivi qui trattati.

Di seguito sono riportate due figure: la prima (2.2) mostra una comune giunzione p-n con rottura ai bordi dovuti all'arrotondamento dei bordi della regione di svuotamento, mentre la seconda (2.3) illustra il legame fra tensione di breakdown, raggio di curvatura dei bordi, drogaggio e geometria del dispositivo **[3]**.

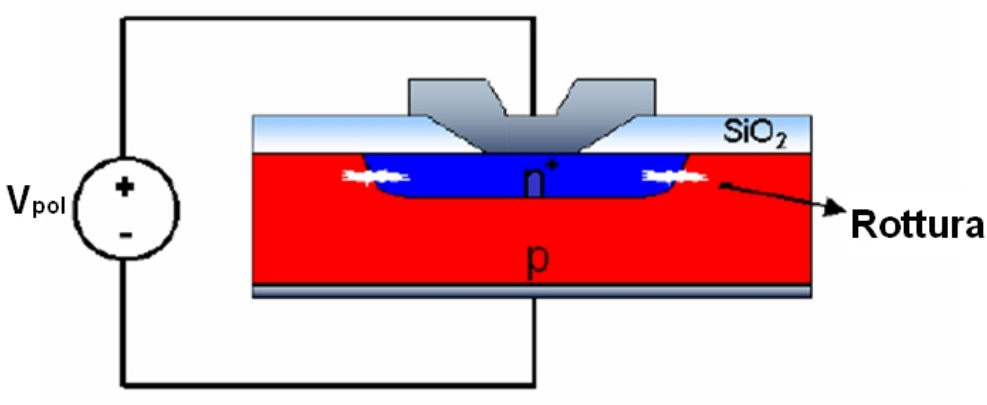

Figura 2.2
Giunzione planare p-n polarizzata inversamente con i bordi arrotondati.

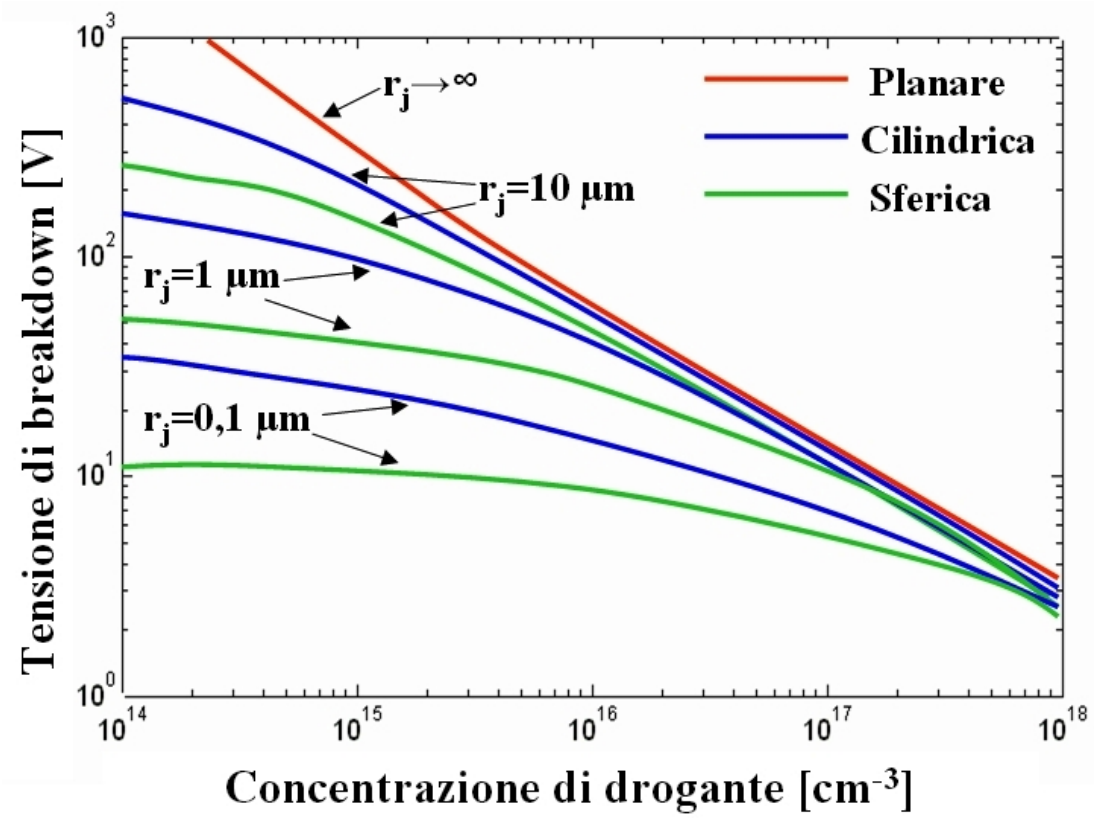

Figura 2.3
Tensione di breakdown in funzione della concentrazione al variare del raggio di curvatura r_j, per una giunzione brusca con geometria planare, cilindrica, e sferica.

La conseguenza più immediata del breakdown localizzato solo ai bordi della zona di svuotamento del dispositivo è che l'effettiva regione attiva risulterà solo nella zona di curvatura periferica. Ciò comporta che solo una piccola parte del fotodiodo sarà in grado di rivelare luce. Per avere l'intera zona svuotata sensibile al fotone incidente ed evitare il problema della rottura ai bordi, si realizza una struttura in cui una regione di tipo p è arricchita di drogante nella zona centrale dell'area attiva del

dispositivo (zona di tipo p^+). In questo modo si rende uniforme il campo elettrico in tutta la zona svuotata.
La seguente figura (2.4) mostra la struttura appena descritta.

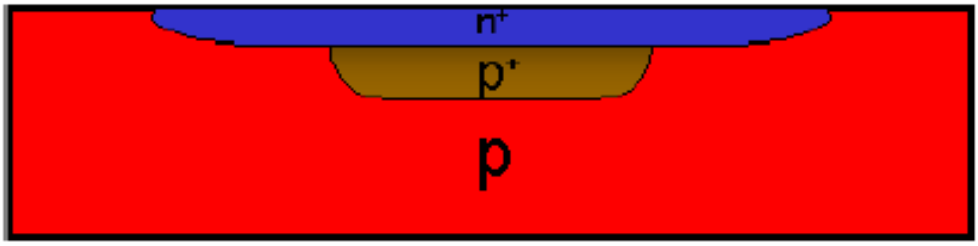

Figura 2.4
Giunzione p-n con zona d'arricchimento p^+ per evitare la rottura ai bordi.

È importante, inoltre, che il dispositivo sia sensibile all'ultravioletto, poiché la lunghezza d'assorbimento nel silicio per raggi ultravioletti (UV) è pari a 0,1 µm, il fotorivelatore deve essere realizzato in modo tale che la regione n^+ abbia uno spessore non superiore a qualche migliaio di Angstrom (1 Å =10^{-10} m).

Un altro parametro molto importante da prendere in considerazione in termini di progetto è il tempo di risposta dello SPAD. Infatti, quando un fotone incide sul dispositivo può accadere che sia assorbito nella regione quasi neutra sotto la zona attiva ad una distanza inferiore di una lunghezza di diffusione dal bordo della zona di svuotamento. Appena il fotone è assorbito, si crea una coppia elettrone-lacuna. L'elettrone, così generato, può diffondere verso la regione svuotata in un determinato intervallo di tempo ed innescare, così, una valanga. Tutto ciò porterà un ritardo nella rivelazione del fotone. Tale ritardo, poiché i portatori diffondono in modo casuale, è statisticamente distribuito con legge con *legge esponenziale* del tipo $e^{t/\tau}$, dove il valore di τ è dato da:

$$\tau = \frac{s^2}{D_s \cdot \pi^2} \tag{2.1}$$

in cui s è lo spessore della zona neutra e D_s la costante di diffusione. Poiché D_s assume valori intorno ai 25 cm^2/s e s intorno ai 500 µm, allora, la τ risulta uguale a 10 µs. Questo valore è molto grande, assolutamente non accettabile per un dispositivo che opera in photon timing. Per risolvere quest'ultimo problema si crea, fra l'area attiva dello SPAD ed il substrato, un'altra giunzione p-n che ha il compito di evitare la risalita degli elettroni generati dall'assorbimento del fotone sotto la zona attiva. É chiaro che, in questo modo, si abbassa l'efficienza quantica dello SPAD,

poiché sono rivelati meno fotoni, ma si migliora il suo tempo di risposta. Questa giunzione p-n è creata mediante la crescita epitassiale di uno strato p^- su un substrato di tipo n^-, creando, così, una barriera di potenziale che impedisce agli elettroni di risalire verso la zona attiva.

In realtà, la zona epitassiale è costituita da due strati di tipo p con concentrazione differente. Immediatamente sopra il substrato (di tipo n^-) si ha uno strato drogato p^+, in modo da creare un percorso a bassa resistenza per la corrente (nell'ordine dei milliampère), generata durante la valanga, rendendo, così, il dispositivo più veloce, grazie a questo cammino **[4]**.

Da ciò che è stato detto fino a questo momento, il dispositivo, attorno alla zona attiva, può essere schematizzato, come in figura 2.5.

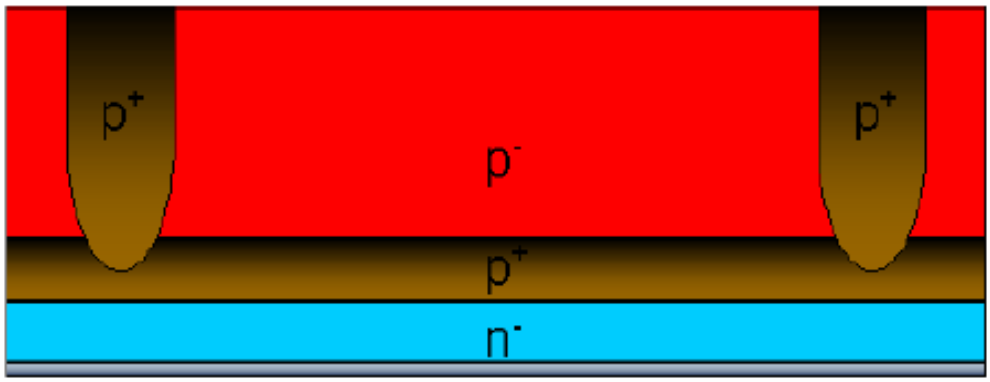

Figura 2.5
La figura illustra la struttura che circonda la regione attiva dello SPAD: substrato di tipo n^-, lo strato epitassiale p^+ (che permette insieme al substrato di migliorare la risposta del dispositivo) e il sinker p^+.

In figura (2.5), sono presenti anche due sinker (sacche) di tipo p^+ che realizzano un miglior contatto ohmico fra la zona p ed il metallo, utilizzato per applicare una tensione a tale regione, creando, inoltre, un percorso a bassa resistenza con lo strato p^+ di fondo.

Infine, per quanto riguarda la geometria tridimensionale del dispositivo, si è scelto di utilizzare una forma cilindrica per evitare la formazione di angoli che "abbassino" la tensione di breakdown della giunzione, "anticipandone" la rottura.

Di seguito è illustrato un esempio del processo di fabbricazione di uno SPAD singolo.

2.3 Processo di fabbricazione di uno SPAD singolo

Per ridurre fenomeni limitanti come dark-count ed afterpulsing, il materiale utilizzato nella costruzione del dispositivo deve essere quasi del tutto esente da difetti, presentando una bassissima dose d'impurità. A tale scopo si realizza un substrato poco drogato di tipo n^- con le seguenti caratteristiche:

1. Una resistenza pari a 4 kΩ·cm;

2. Una concentrazione d'impurità inferiore a 10^{12} cm^{-3};

3. Un tempo di vita dei portatori pari a 1 ms (abbastanza alto);

4. Una concentrazione di ossigeno sotto $5 \cdot 10^{16}$ cm^{-3}.

Il primo passo di processo è la realizzazione di una zona di silicio drogato di tipo p^+, con atomi di boro (B), spessa 2 µm, cresciuta mediante epitassia. Com'è stato già detto, lo scopo di questa regione è ridurre la resistenza incontrata dalle cariche in moto durante la valanga. In seguito, è cresciuta, sempre mediante epitassia, un'ulteriore zona di tipo p^- spessa 5 µm, creando, così, una giunzione p-n che, come già detto, evita la risalita dei fotoelettroni che diffondono dalla regione quasi neutra.. Entrambe le epitassie sono ottenute con la tecnica ***CVD*** (*chemical vapor deposition*). La figura seguente (2.6) rappresenta il dispositivo fino a questo momento del processo.

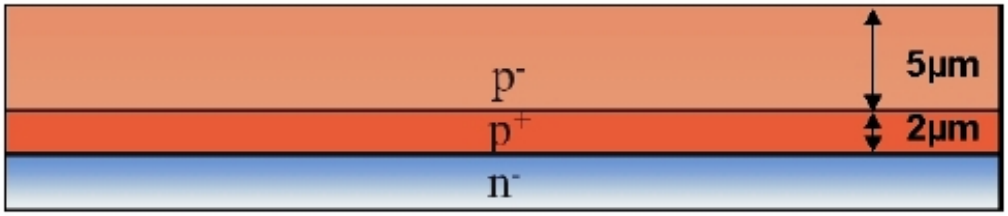

Figura 2.6
Substrato n^- con bassa dose d'impurità, su cui é stato cresciuto uno strato p^+, mediante epitassia, ed un successivo strato p^- per deposizione di Boro.

In seguito, è realizzato il sinker di tipo p^+, mediante impiantazione, per avere un buon contatto ohmico e, quindi, una bassa, in pratica nulla, caduta di tensione fra il metallo ed il semiconduttore, facendo cadere, così, tutta la tensione ai capi del dispositivo. La geometria del sinker sarà ad anello attorno all'area attiva del dispositivo. Prima di eseguire l'impiantazione del boro, è fatto crescere un nuovo sottile strato d'ossido spesso 500 Å (ossido di schermo di pre-impianto) per evitare la

contaminazione dello SPAD da parte di impurità metalliche residue del processo d'impiantazione (knock on). I processi compiuti in questa fase, quindi, sono:

1. L'*ossidazione termica*, che fa crescere uno strato d'ossido di 7000 Å per proteggere la parte del dispositivo che non deve subire l'impiantazione.

2. La *litografia* per definire la geometria del sinker.

3. L'*attacco in dry dell'ossido* per eliminarlo laddove si andrà ad impiantare.

La seguente figura (2.7) mostra l'aspetto del dispositivo prima di compiere l'impiantazione.

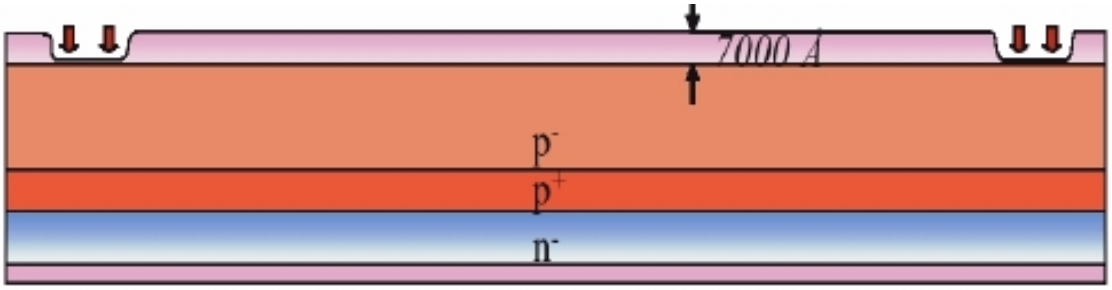

Figura 2.7
Prima della formazione del sinker, che sarà creato nelle zone indicate dalle frecce rosse, il dispositivo è coperto con uno strato di ossido di 7000 Å.

A questo punto la fetta è pronta per subire il processo d'impiantazione che, solitamente, è effettuato con una dose di boro di 10^{15} cm^{-2} e con un'energia d'impianto di 120 keV.

Com'è stato ripetuto più volte, occorre un dispositivo molto pulito per ridurre effetti limitanti come il dark count e l'afterpulsing; per questo motivo si realizza attorno allo SPAD un anello, poiché la forma del dispositivo è cilindrica, di *"getter"* di tipo n^+. Il processo di gettering ha il compito di catturare le possibili impurità metalliche che o sono presenti nella zona attiva del dispositivo o possono diffondere durante l'annealing finale (eseguito a 700°C per 8 ore) raggiungendo il fondo della fetta. Tale processo è realizzato mediante *predeposizione* di fosforo (sotto forma di $POCl_3$). Non si usa l'impiantazione in questo caso, poiché la predeposizione consente drogaggi più alti, con una diffusione su un'area più larga, un minore impatto e, quindi, meno danni e difettosità sulla struttura cristallina del silicio.

Per valutare il fenomeno della diffusione delle impurità, si consideri la loro *lunghezza di diffusione* L:

$$L = \sqrt{D \cdot t} \tag{2.2}$$

dove t è il *tempo di diffusione* delle impurità e D il *coefficiente di diffusione*, dato da:

$$D = D_0 \cdot e^{-\frac{E_A}{k \cdot T}} \tag{2.3}$$

in cui D_0 è il fattore pre-esponenziale, E_A l'energia di attivazione necessaria per la diffusione espressa in eV, k la costante di Boltzmann e T la temperatura espressa in gradi Kelvin. Un esempio può essere fatto utilizzando i valori mostrati nella tabella seguente **[5]**; si vede che, la lunghezza di diffusione del potassio (K), durante l'annealing, è circa 76 μm (D_0=1,1·10^{-3} cm^2/s, E_A=0,76 eV). Quindi, se nella regione attiva del dispositivo sono presenti degli atomi di potassio, questi non riusciranno mai a raggiungere il fondo della fetta poiché é distante circa 500 μm.

Elemento	D_0 [cm^2/s]	E_A [eV]	Solubilità [cm^{-3}]
Li	2,3·10^{-3} ÷ 9,4·10^{-4}	0,63 ÷ 0,78	7·10^{19} (1200ºC)
Na	1,6·10^{-3}	0,72	10^{18} ÷ 9·10^{18} (600ºC ÷ 1200ºC)
K	1,1·10^{-3}	0,76	9·10^{17} ÷ 7·10^{18} (600ºC ÷ 1300ºC)
Cu	4·10^{-2}	1	5·10^{15} ÷ 3·10^{18} (600ºC ÷ 1300ºC)
Ag	2·10^{-3}	1,6	6,5·10^{15} ÷ 2·10^{17} (600ºC ÷ 1350ºC)
Au	1,1·10^{-3}	1,12	5·10^{14} ÷ 5·10^{16} (900ºC ÷ 1300ºC)
(Au)$_i$	2,4·10^{-4}	3,9	-
(Au)$_s$	2,8·10^{-3}	2,04	-
Pt	1,5·10^2 ÷ 1,7·10^2	2,15 ÷ 2,22	4·10^{16} ÷ 5·10^{17} (800ºC ÷ 1200ºC)
Ni	0,1	1,9	6·10^{18} (1200ºC ÷ 1300ºC)
Cr	0,01	1	2·10^{13} ÷ 2,5·10^{15} (900ºC ÷ 1280ºC)
Co	9,2·10^4	2,8	2,5·10^{16} (1300ºC)
O_2	7·10^{-2}	2,44	1,5·10^{17} ÷ 2·10^{18} (1000ºC ÷ 1400ºC)
H_2	9,4·10^{-3}	0,48	2,4·10^{21}·exp(-1,86/kT) (P=1 atm)
Si	1,5·10^3	5,02	Auto-diffusione per T>1050ºC
Si	1,5	4,25	Auto-diffusione per T<900ºC

Tabella 2.1

Tabella con il fattore pre-esponenzialeD_0 e l'energia d'attivazione E_A dei coefficienti di diffusione delle più importanti impurità del silicio.

Il processo per realizzare l'anello di getter locale (n^+) è il seguente:

1. *Ossidazione termica*: per fare crescere un ossido di 4000 Å allo scopo di proteggere le zone del dispositivo in cui non si deve drogare.

2. *Litografia*: per definire l'area interessata al drogaggio.

3. *Attacco in dry dell'ossido*: per rimuovere l'ossido dalla zona in cui si vuole realizzare il getter;

4. *Predeposizione* del fosforo.

La figura seguente (2.8) illustra tale processo.

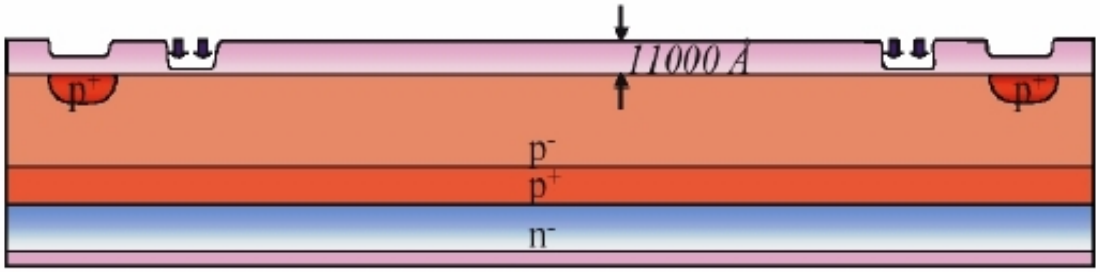

Figura 2.8
Step del processo utilizzato per la creazione dell'anello di getter; da notare la presenza del sinker p^+, che avrà, mediante i vari processi termici, una profondità sempre maggiore. Inoltre, lo spessore totale d'ossido è di 11000 Å.

A questo punto, occorre realizzare l'area attiva del dispositivo. In queste fasi è importantissimo evitare l'introduzione di impurità o difetti. Si dovrebbero quindi evitare attacchi in dry ed impiantazioni ioniche, poiché sono processi molto "aggressivi" nei confronti del dispositivo, in particolare della sua struttura cristallina.

Purtroppo, però, la creazione della zona p^+ nella parte centrale dell'area attiva (necessaria per evitare un elevato campo elettrico ai bordi della zona svuotata) richiede un'impiantazione a bassa energia di boro seguita da un annealing ad alta temperatura. Durante questa fase, per non utilizzare l'attacco in dry dell'ossido, che potrebbe danneggiare il dispositivo tramite le radiazioni usate, è eseguita l'apertura per l'impiantazione mediante una maschera di resist al posto di una maschera d'ossido. Inoltre la definizione dell'area attiva è effettuata attraverso un attacco wet, per evitare ogni possibile danno alla struttura cristallina dovuto a possibili radiazioni.

Di seguito, sono elencati i vari step per la realizzazione dell'area attiva e della zona di tipo p+ (sempre dell'area attiva) del dispositivo:

1. *Ossidazione termica*, per fare crescere un ossido di 1200 Å.

2. *Litografia*, per definire l'area attiva del dispositivo.

3. *Attacco in wet dell'ossido*.

4. *Ossido di pre-impianto* di circa 250 Å.

5. *Deposizione di uno strato di resist* di circa 1,9 μm.

6. *Litografia*, per definire l'impianto di enrichment.

7. *Cottura* del resist di pre-impianto.

8. *Impianto di enrichment*.

9. *Annealing* attraverso un processo di diffusione in forno.

L'annealing è un processo necessario dopo un drogaggio mediante impiantazione ionica perché il reticolo cristallino subisce danni considerevoli durante

il processo d'impiantazione, quindi occorre l'annealing per ripristinare l'ordine cristallino ed assicurare che gli atomi droganti impiantati si pongano in siti costituzionali, dove possono diventare donatori o accettori **[6]**.

Per formare la zona di tipo p^+ solitamente sono utilizzati:

- Una dose di Boro di circa $6 \cdot 10^{12}$ cm^{-2}.
- Un'energia d'impiantazione di 40 keV.
- Un annealing di circa 2 ore ad una temperatura di 1150°C.

Con questi valori si ottiene una tensione di breakdown di circa 30 V.

Le seguenti figure (2.9) e (2.10) mostrano lo stato del dispositivo dopo aver subito i precedenti processi; in particolare la figura (2.9) illustra lo stato del dispositivo prima dell'impianto di boro e dell'annealing, mentre la (2.10) mostra il dispositivo subito dopo l'annealing.

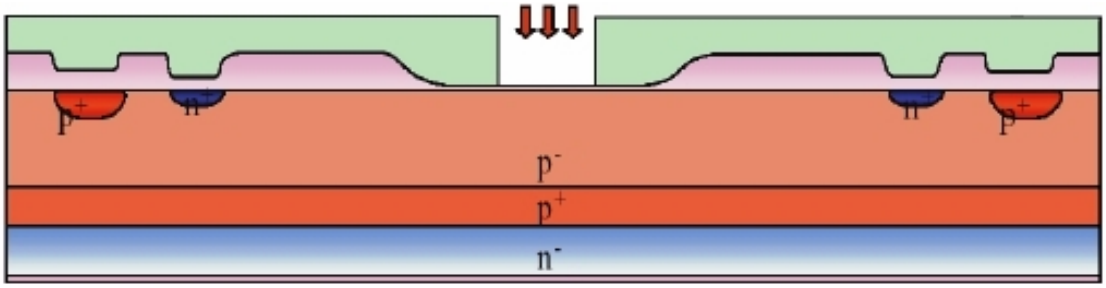

Figura 2.9

Drogaggio di enrichment: nella zona centrale del dispositivo si ha un sottilissimo strato di ossido che riduce la contaminazione da knock on, durante l'impianto, nell'area attiva. In questo momento il dispositivo non ha ancora subito l'annealing di 2 ore a 1150°C.

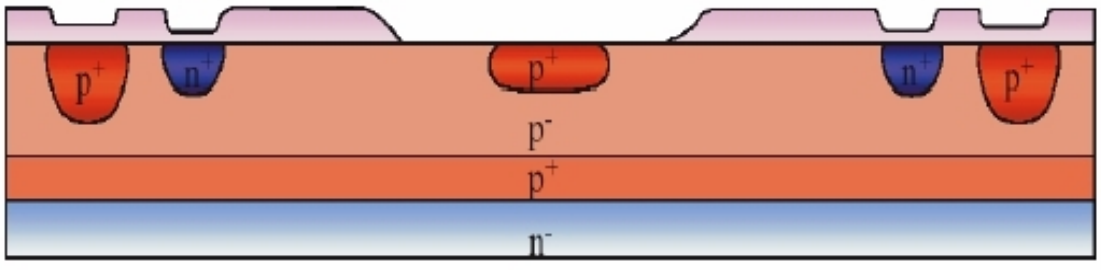

Figura 2.10

Diffusione di enrichment: dopo l'annealing, nella zona centrale si crea la sacca p^+. Da notare, inoltre, che dopo l'annealing le altre sacche diffondono più in profondità.

Per realizzare lo strato drogato di tipo n^+ dell'area attiva, non è utilizzata l'impiantazione, che peggiora le prestazioni dello SPAD, in termini soprattutto di dark count ed afterpulsing, aumentando la difettosità nella regione di svuotamento, ma una deposizione di polisilicio drogato "in situ". Inoltre, per ridurre ulteriormente

il numero di conteggi di buio (dark count rate), è stato eseguito il processo di RTA (Rapid Thermal Annealing) per "riordinare" la struttura cristallina dopo il drogaggio, riducendo, così, limitatamente alla breve durata di tale processo, le impurità, e per realizzare giunzioni molto sottili al fine di non ridurre troppo l'efficienza del fotorivelatore nella luce blu e nell'ultravioletto, formando in questo modo un basso profilo di arsenico (As) sotto il polisilicio dentro lo strato epitassiale p^-.

Gli step utilizzati per la creazione della zona n^+ sono:

1. *Rimozione* del precedente ossido di pre-impianto.

2. *Deposizione* di 1000 Å di silicio poli-cristallino drogato in situ.

3. *Litografia*.

4. *Rimozione* del polisilicio ove non è necessario.

A questo punto, prima del RTA (a 1100°C per 120 secondi), che ha il compito di far diffondere l'arsenico, è depositato, mediante CVD, un nuovo strato di ossido di 4000 Å, che ha il compito d'impedire l'evaporazione dell'arsenico durante il processo di RTA e soprattutto proteggere l'area attiva dello SPAD durante la successiva predeposizione di fosforo, che creerà uno step ulteriore di gettering.

Nella figura 2.11 è illustrato lo stato del dispositivo dopo la deposizione di silicio poli-cristallino, nella figura 2.12 è raffigurato lo stato del dispositivo dopo il processo di RTA, mentre nella figura 2.13 è rappresentata la concentrazione finale del drogaggio dell'area attiva creata rispetto alla profondità, misurata attraverso l'analisi *Spreading Resistance Profiling* (SRP).

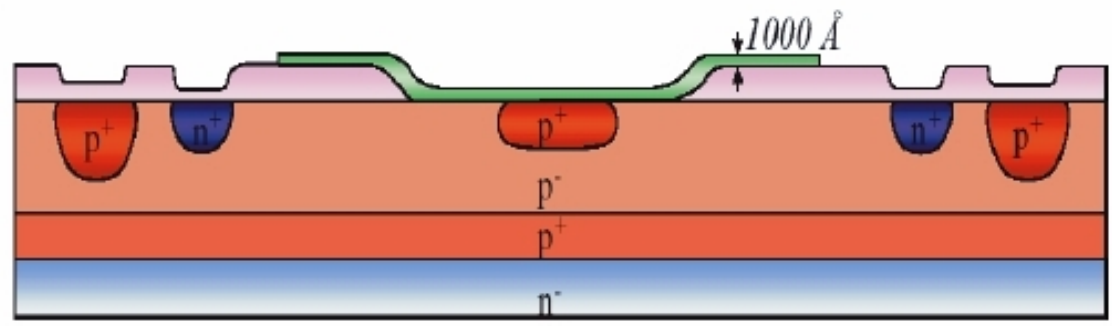

Figura 2.11
In figura è mostrata la deposizione di 1000 Å di silicio poli-cristallino sull'area attiva; da questo strato avverrà la diffusione dell'arsenico, mediante RTA, verso la zona p^+.

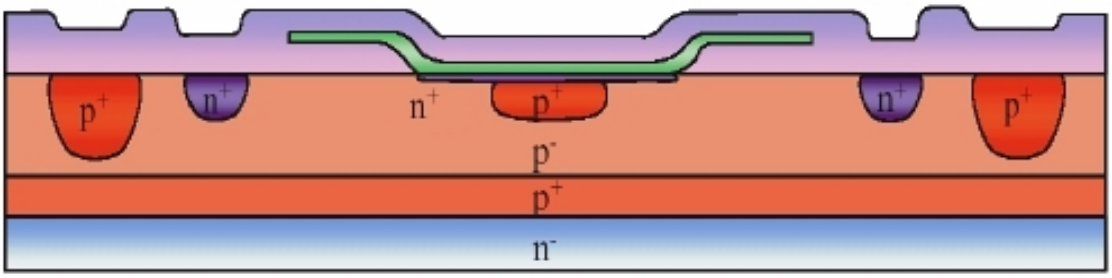

Figura 2.12

Dopo il processo di RTA, si forma il sottile strato n^+ che ha un drogaggio di circa 1020 cm^{-3}. Inoltre, sopra lo strato poli-cristallino si ha uno strato di ossido di 4000 Å realizzato mediante CVD, che ha il compito di non far evaporare l'arsenico durante il processo RTA.

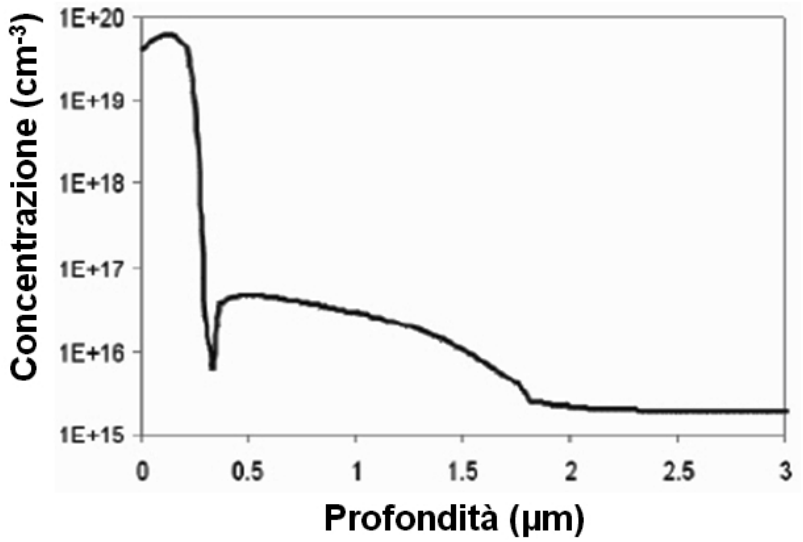

Figura 2.13

Profilo del drogaggio finale dell'area attiva dello SPAD misurato tramite SRP.

A questo punto, resta solo realizzare i contatti metallici che permettono di polarizzare il dispositivo. Gli step per la loro realizzazione sono:

1. *Litografia* per definire i contatti del dispositivo.

2. *Rimozione dell'ossido* dove si deve realizzare il contatto.

3. *Deposizione* di una lega metallica formata da Al-Si-Cu.

4. *Litografia* della metal e *rimozione* della metal ove non è necessaria.

Infine, si realizza un processo a bassa temperatura in *forming gas* (90% di N_2 e 10% di H_2) per favorire l'adesione fra la metal ed il silicio e disattivare i legami pendenti degli atomi di silicio posti sulla superficie del reticolo cristallino.

La seguente figura (2.14) mostra l'aspetto finale dello SPAD.

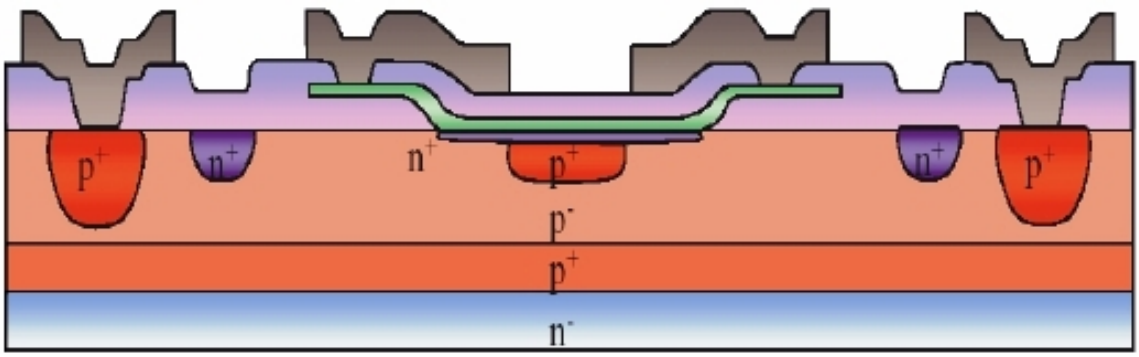

Figura 2.14
La figura illustra la cross section di uno SPAD completo.

Nelle immagini successive (2.15 e 2.16) sono mostrati il layout di uno SPAD da 100 µm e la rispettiva maschera di layout. Si può notare l'anello più esterno, la zona p^+, denominata con la A, anodo, mentre l'anello interno, il catodo (n^+), è denominato K.

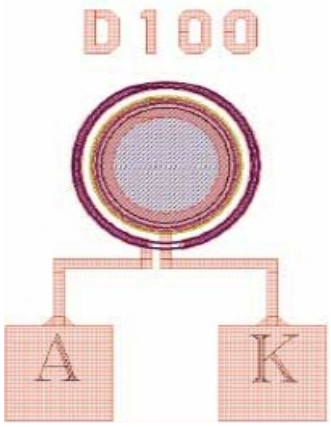

Figura 2.15
Layout di uno SPAD singolo da 100 µm.

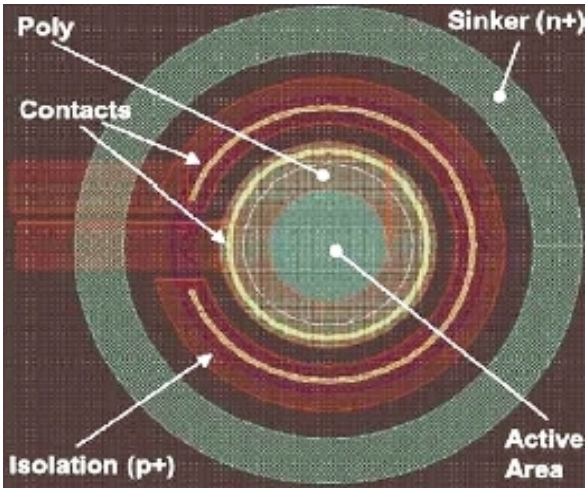

Figura 2.16
Maschera di layout di uno SPAD: si noti la forma cilindrica dello SPAD, in cui le zone drogate in modo diverso formano delle corone circolari attorno all'area attiva del dispositivo.

Di seguito sono mostrate due immagini (2.17 e 2.18) al microscopio elettronico a scansione (**SEM**) di uno SPAD, con un diametro dell'area attiva di 10 µm, in cui si può notare una diversa scala di grigio per le zone drogate in maniera differente e per l'ossido.

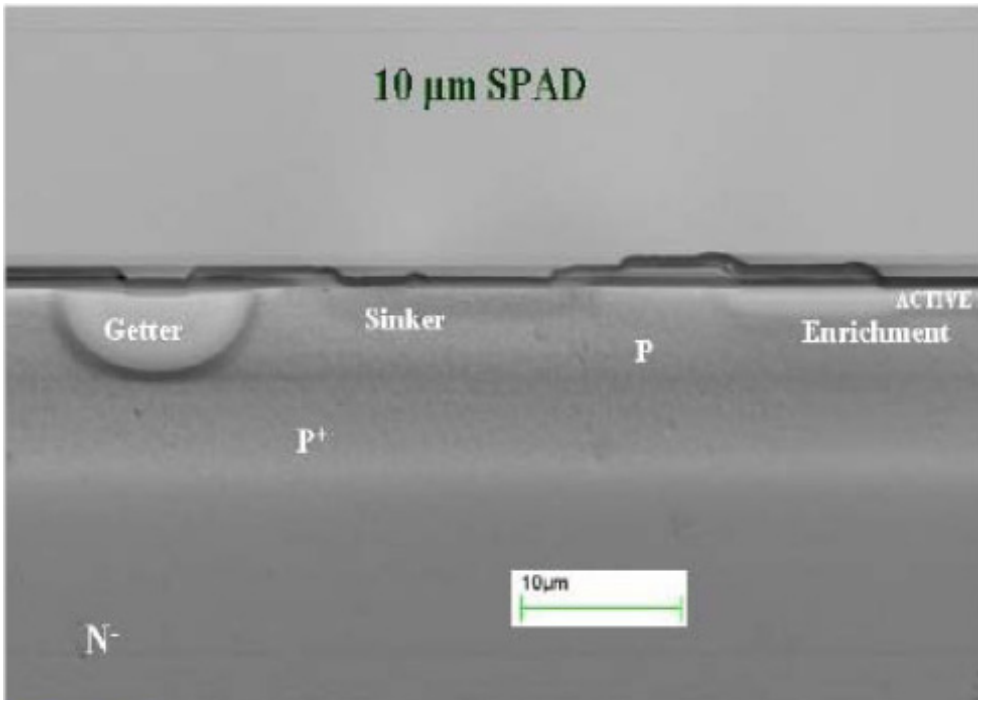

Figura 2.17
Immagine al SEM che raffigura la struttura finita di uno SPAD da 10 µm.

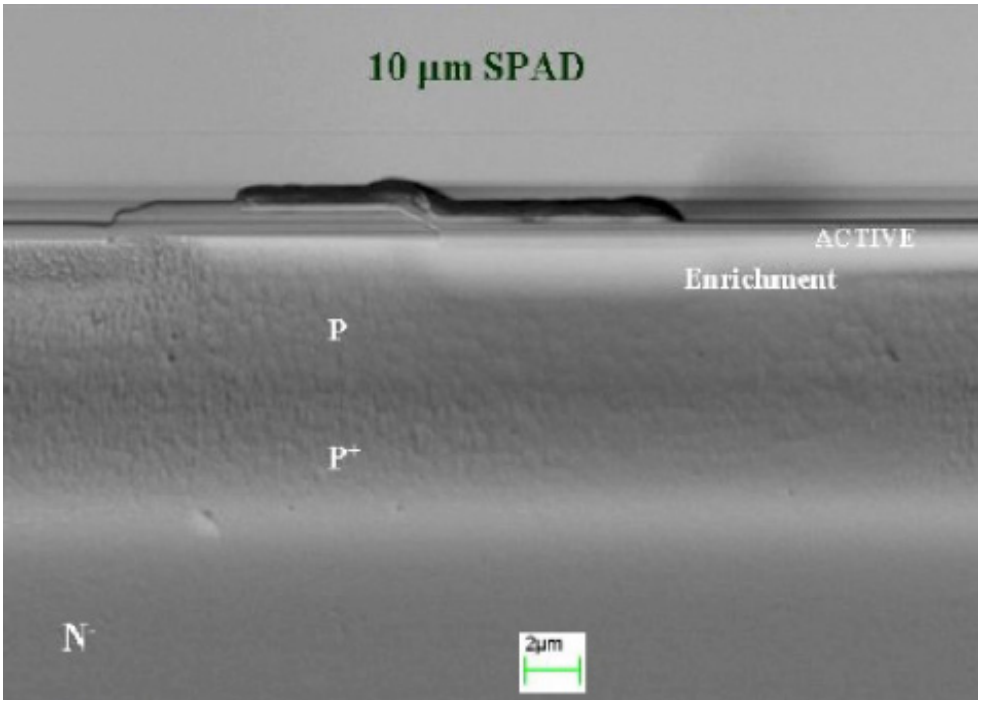

Figura 2.18
Immagine al SEM che riporta con elevato ingrandimento l'area attiva di uno SPAD da 10 µm.

2.4 Vantaggi e problematiche in una matrice di SPAD

Nei paragrafi precedenti è stata trattata la realizzazione di un singolo SPAD. Tuttavia, per rendere più versatile il dispositivo ed ampliare il numero delle sue possibili applicazioni, è preferibile costruire delle matrici integrate di SPAD, come, ad esempio, nel caso del SiPM.

I vantaggi delle matrici di SPAD sono numerosi: ad esempio, con le matrici integrate si possono realizzare misure di conteggio di fotoni (*photon counting*) considerando la distribuzione spaziale lungo una qualsiasi coordinata. Inoltre, le matrici di SPAD introducono un funzionamento nuovo, simile a quello di un fotomoltiplicatore con un'ampia zona sensibile paragonabile a quella dei PMT; in questo caso si possono, dunque, adoperare anche degli SPAD più piccoli di diametro che hanno prestazioni migliori (soprattutto in termini di dark-count, afterpulsing e timing): è questo il caso del SiPM **[7]**.

Purtroppo, le matrici integrate di SPAD sono influenzate dal problema del *cross talk* (*diafonia*), sia *ottico* sia *elettrico*, che, come visto e spiegato nel capitolo precedente a proposito dei SiPM, limita moltissimo le prestazioni delle matrici.

Di seguito, è spiegato il processo di fabbricazione di una matrice planare di SPAD, tenendo conto di tutte le procedure utilizzate per minimizzare il cross talk.

2.4.1 Processo di realizzazione di una matrice planare

I processi utilizzati per realizzare i singoli pixel di una matrice sono uguali a quelli utilizzati e visti in precedenza per il singolo SPAD. Esistono, tuttavia, alcune rilevanti differenze da prendere in considerazione.

La prima riguarda la creazione della zona di getter locale, che, negli SPAD singoli, è realizzata a forma d'anello attorno ad ogni dispositivo, mentre nella matrice è realizzata con una predeposizione "a tappeto" di fosforo su tutta la fetta in modo da circondare interamente tutte le singole microcelle.

La seconda, molto più rimarchevole della prima, è, invece, dovuta alla risoluzione del problema del *cross talk* (*diafonia*), ottico ed elettrico. A tal proposito, si è pensato di introdurre delle "barriere", le trincee (*trench*, in inglese), in grado di

schermare i singoli elementi, fotoni ed elettroni, evitando che "saltino" da un pixel all'altro. In letteratura sono proposte diverse strategie per ottenere tale isolamento: le più importanti sono quelle di *A. Mathewson* e *W. J. Kindt*. *Mathewson* suggerisce la costruzione di trench o scavi riempiti con ossido in degli array realizzati su wafer *SOI* (*Silicon On Insulator*), mentre *Kindt* consiglia la realizzazione di trench profondi scavati nel silicio e riempiti di lega metallica. Nelle matrici trattate in questo lavoro di tesi si è deciso di utilizzare una soluzione ibrida fra quelle proposte da *Mathewson* e *Kindt.* Infatti, l'*isolamento ottico* è effettuato con uno strato di metallo, che ha il compito di schermare la radiazione luminosa generata durante la valanga, mentre l'*isolamento elettrico* è realizzato a giunzione, per questo si parlerà di matrici *isolate a giunzione*. Lo stato del dispositivo, prima dei processi eseguiti per la costruzione del trench, è mostrato nella figura (2.14).

Per *isolare a giunzione* la matrice di SPAD, inizialmente, si effettua, prima di realizzare i contatti metallici fra pixel, un processo di fotomascheratura ed in seguito un processo di attacco in plasma (*in dry*) dei vari strati di ossido presenti sulla superficie del dispositivo. In seguito, si procede allo scavo nel silicio di trench, circolari (data la forma di ogni microcella), attorno all'area attiva di ogni pixel, come illustrato in figura 2.18. Tali trench sono profondi circa 10 µm nel silicio, mentre la loro larghezza è pari a 1 µm negli array con diametro d'area attiva del singolo pixel di 20 µm e 0,8 µm negli array da 40 µm e 60 µm.
Questi processi sono illustrati nella seguente figura (2.19):

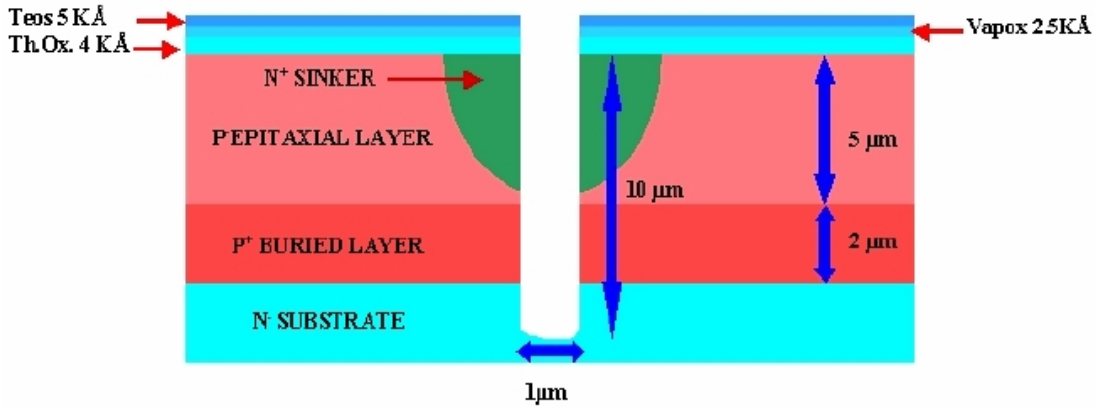

Figura 2.19
Zona di getter dopo l'attacco in dry.

Dopo aver realizzato la buca, gli step impiegati per ottenere l'isolamento elettrico, sono i seguenti:

1. Pr*edeposizione* di fosforo (sotto forma di $POCl_3$) per realizzare una sacca di tipo n^+, attorno ai trench, profonda almeno quanto l'epitassia del

dispositivo, ottenendo, così, un isolamento elettrico, mediante una giunzione p-n, nella zona di anodo p^+ che è comune a tutti i pixel della matrice.

2. *Rimozione* dei vetri di fosforo, generati dal precedente processo di predeposizione, dalla superficie dei dispositivi.

3. *Crescita* di circa 250 Å di ossido termico sulle pareti e sul fondo dei trench, per migliorare l'isolamento elettrico fra i pixel degli array

4. *Deposizione* di 1000 Å di TEOS (ossido) con la tecnica CVD.

5. *RTA* (a 1000°C per 60 s) per favorire l'adesione fra i vari strati di ossido presenti sulla superficie del dispositivo.

Nelle due seguenti figure (2.20 e 2.21) sono illustrati i processi prima esaminati per ottenere l'isolamento elettrico fra i pixel.

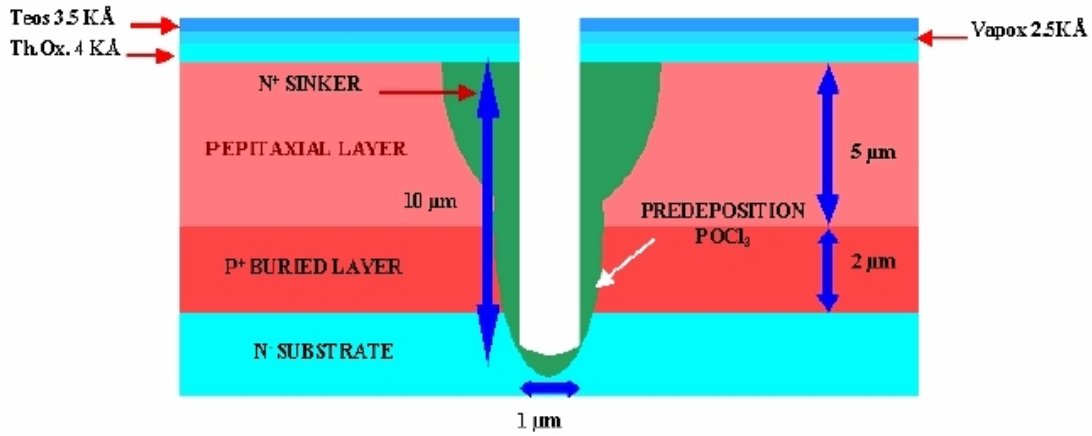

Figura 2.20

Realizzazione dell'isolamento elettrico a giunzione: predeposizione di $POCl_3$ nel trench e rimozione vetri di fosforo. La zona n^+ giunge fino al substrato n-, creando così, insieme alla zona p^+, una giunzione p-n che fa da ostacolo agli elettroni che si dirigono verso altri pixel.

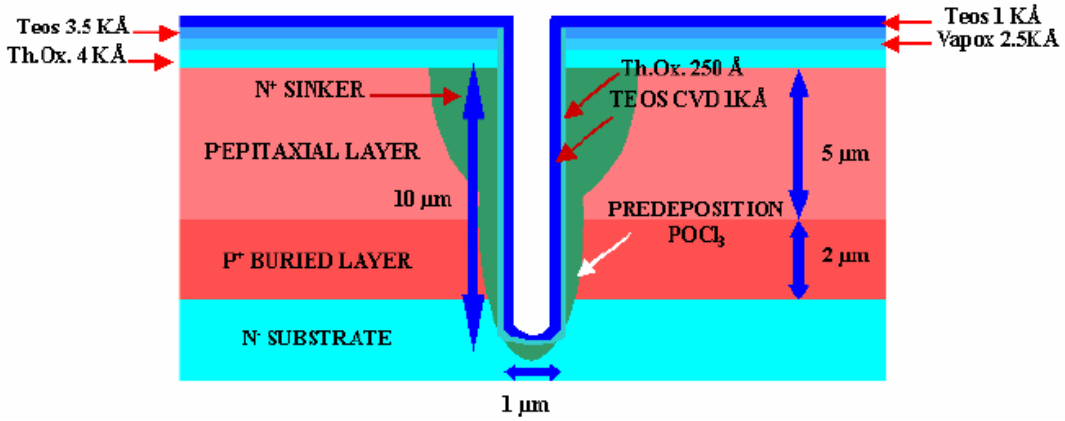

Figura 2.21

Isolamento elettrico a giunzione: ossidazione termica (~ 250 Å), TEOS CVD (~ 1000 Å), RTA (a 1000°C per 60 s, N_2).

La predeposizione di fosforo, sotto forma di $POCl_3$, è realizzata nel trench in seguito agli step di processo costruttivi dello SPAD e permette un'efficace pulizia dalle impurità all'interno dell'area attiva. Infatti, il $POCl_3$ si mantiene a più elevate concentrazioni rispetto al fosforo presente nel sinker n^+, non avendo subito tutti i processi costruttivi descritti in precedenza per la creazione dell'area attiva, e realizza, così, un gettering migliore rispetto al sinker n^+, creato, invece, durante tali processi (diffusione, RTA, CVD, ecc.).

Dopo aver completato l'isolamento elettrico, si passa alla realizzazione dell'isolamento ottico; in questo caso i processi utilizzati sono:

1. *Sputtering* della barriera di Ti/TiN (titanio e nitruro di titanio) per favorire una migliore adesione del tungsteno rispetto a quanto, invece, si avrebbe depositando direttamente il tungsteno su silicio. Lo strato di titanio depositato è di 450 Å, mentre quello di nitruro di titanio è di 750 Å;

2. *Deposizione* del tungsteno (Wolframio - W) mediante CVD: lo spessore di tungsteno depositato è di 8000 Å. É utilizzato il tungsteno poiché la tecnologia di deposizione di questo metallo è abbastanza consolidata e, inoltre, esso ha una temperatura di fusione superiore ai 3000°C.

Di seguito è proposta una figura (2.22) relativa agli step sopra descritti.

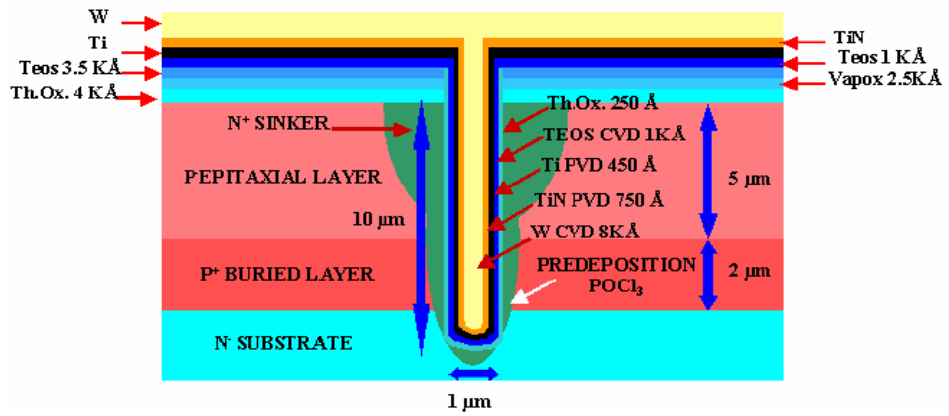

Figura 2.22

Isolamento ottico: in figura sono facilmente individuabili, attraverso tre diversi colori (nero, arancione e giallo), gli strati di Ti, TiN e W.

Dopo la creazione del doppio isolamento, elettrico ed ottico, si rimuove lo strato di tungsteno precedentemente depositato a tappeto su tutta la fetta mediante *etchback* e si procede all'attacco della barriera Ti/TiN depositata anch'essa a tappeto su tutta la fetta; si ricopre, poi, il dispositivo con uno strato di ossido di 3000 Å di vapox per la passivazione, affinché non ci siano contatti fra la metal (utilizzata per contattare l'anodo e il catodo) ed il tungsteno. Di seguito, la figura 2.23 illustra lo

stato del dispositivo fino a questo momento, mentre la figura 2.24 riporta la cross-section del dispositivo alla fine di questo processo.

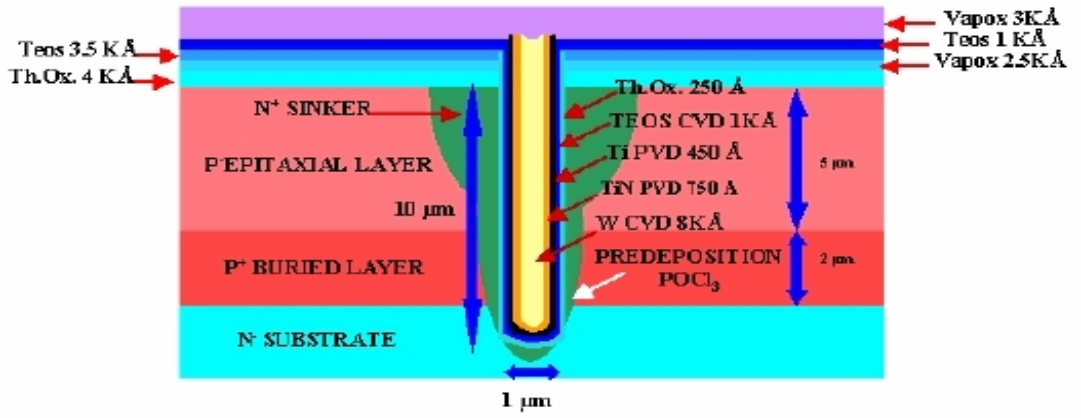

Figura 2.23
Immagine del trench alla fine del processo

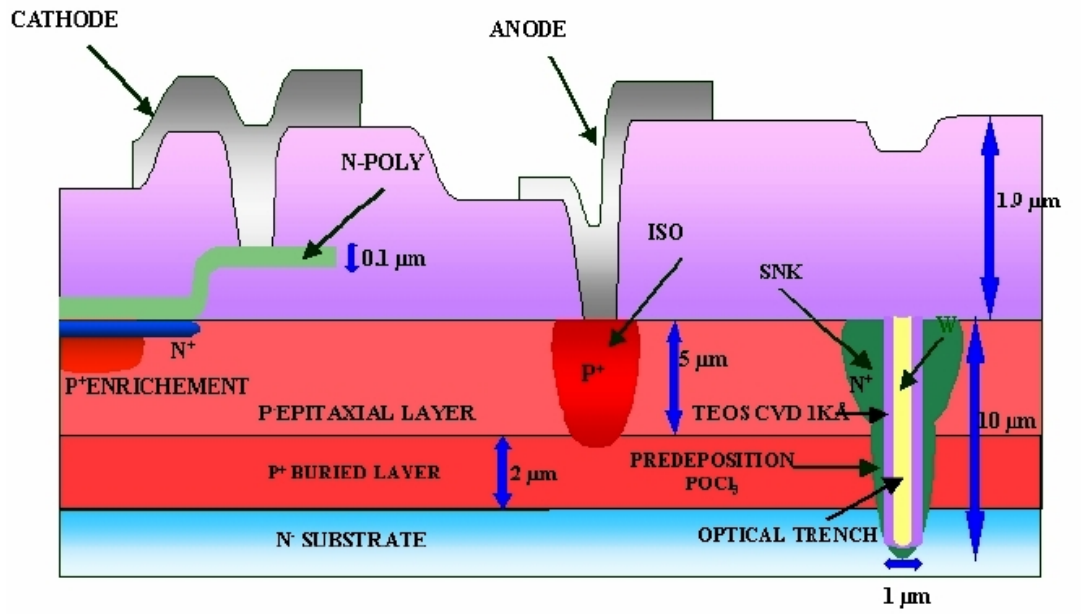

Figura 2.24
In figura è illustrato mezzo pixel, di diametro pari a 20 µm, in cui è stato realizzato l'isolamento elettrico ed ottico.

Poiché nel caso di un mezzo conduttore, la distanza percorsa dalla luce all'interno del conduttore è pari al coefficiente di penetrazione δ, dato dall'inverso del coefficiente d'attenuazione α **[8]**:

$$\delta = \frac{1}{\alpha} = \sqrt{\frac{2}{\omega \cdot \sigma \cdot \mu}} \qquad (2.4)$$

in cui ω è la pulsazione dell'onda elettromagnetica d'interesse (in questo caso, per la luce, è pari a circa $2\pi \cdot 10^{15}$ rad/s), σ è la conducibilità del mezzo (in questo caso, è

circa $18{,}2 \cdot 10^{8}$ S/m) e, infine, μ è permeabilità magnetica del mezzo (in questo caso, è $4\pi \cdot 10^{-7}$ H/m). Dalla formula e dai valori indicati si ottiene una profondità di penetrazione δ pari a circa 0,37 nm. Da questo valore ricavato di δ, si vede che lo spessore dello strato di tungsteno è sufficiente a riflettere possibili raggi luminosi generati dalla rottura per valanga di un pixel della matrice.

Esistono, tuttavia, matrici prive d'isolamento fra pixel. Esse hanno un'area attiva maggiore rispetto a quelle con isolamento, ma presentano maggiori problemi in termini di cross talk, sia ottico sia elettrico.

Nella figura seguente (2.25) è proposta, invece, la cross-section di una matrice priva d'isolamento fra pixel.

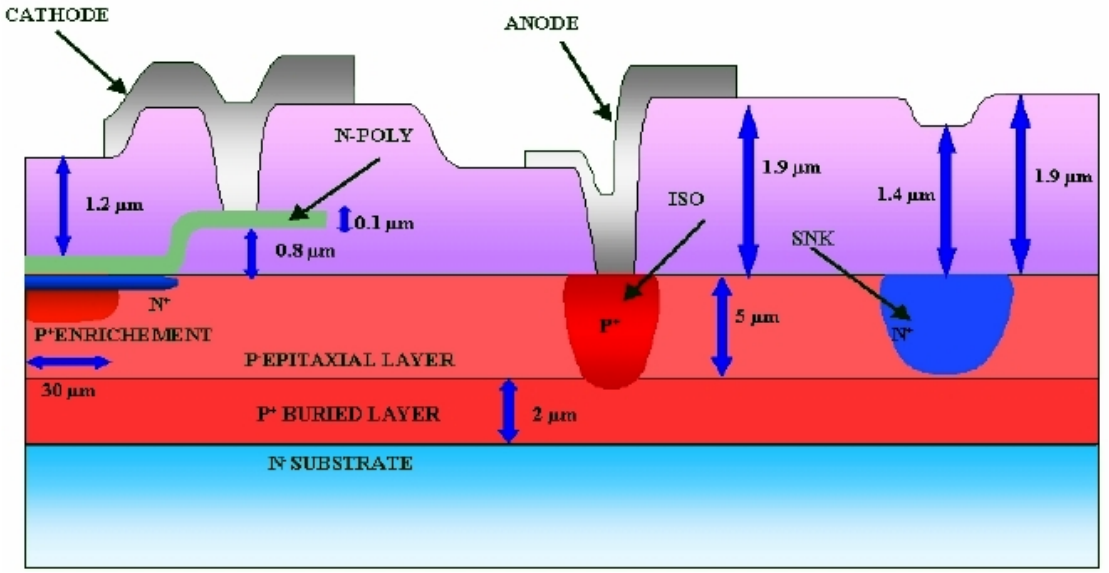

Figura 2.25
Sezione di un pixel con diametro di 60 µm senza isolamento.

Nelle figure 2.26 e 2.27 sono, invece, riportate due foto significative realizzate al SEM LEO (il microscopio elettronico a scansione) a fine processo e relative rispettivamente al top del trench e ad una coppia di pixel adiacenti, del lotto Y503012, in cui è visibile la sacca di tipo n realizzata sul fondo e sulle pareti dei trench mediante pre-deposizione di $POCl_3$.

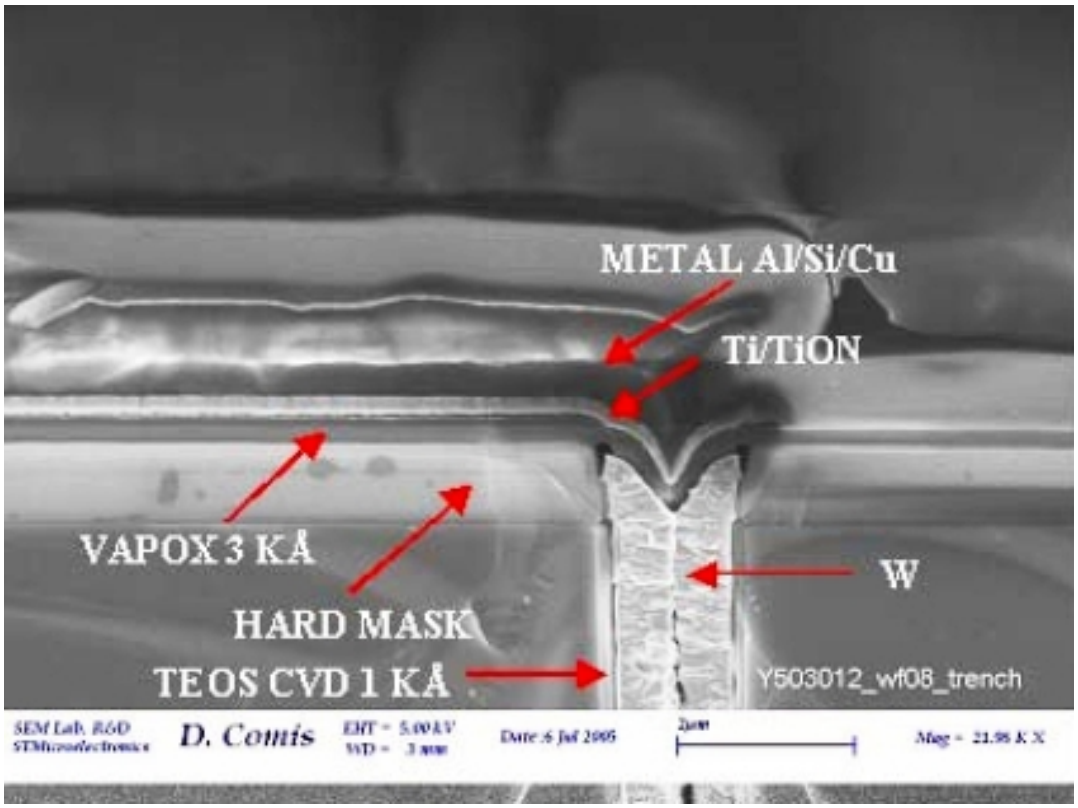

Figura 2.26
Foto SEM relativa al top del trench di un pixel di un array di SPAD.

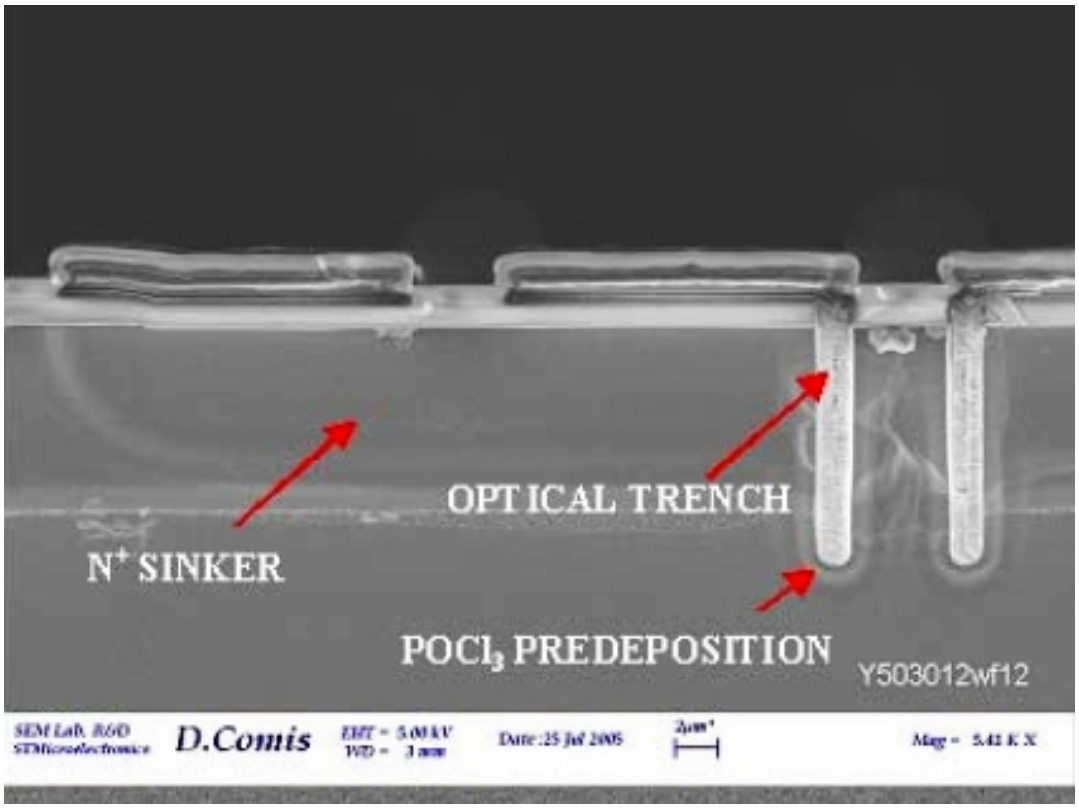

Figura 2.27
Foto SEM relativa a due pixel adiacenti di un array.

Dall'analisi SEM realizzata per questo lotto si osserva che il trench è chiuso al *top*; il tungsteno ricopre le pareti ed il fondo del trench, per cui è garantito uno strato di tungsteno sufficiente per l'isolamento ottico.

Di seguito è riportata un'immagine di layout (2.28) di una matrice 5x5 di SPAD con diametro di 20 µm del lotto Y503012.

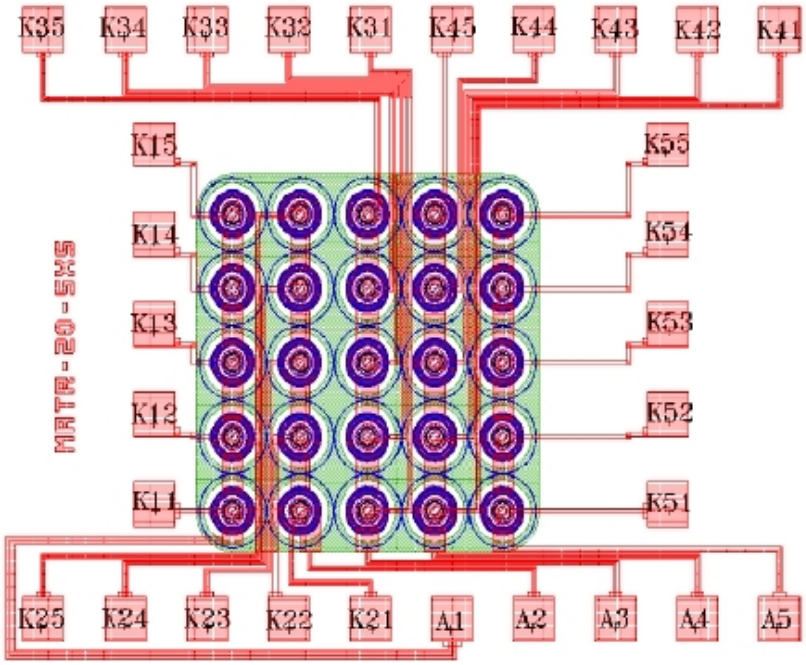

Figura 2.28

La figura illustra il layout di una matrice 5x5 da 20 µm del lottoY503012 . Si osservi che la matrice ha i pixel della stessa colonna con gli anodi (A1, A2, A3,A4, A5) a comune.

La figura 2.29 riporta, inoltre, due immagini al microscopio ottico: a sinistra è rappresentata la parte superficiale della matrice, in cui si distinguono le aree attive circolari dei pixel circondate dai contatti metallici per alimentazioni ed uscite, a destra sono visibili i fili di bonding utilizzati per la piedinatura della matrice nel chip.

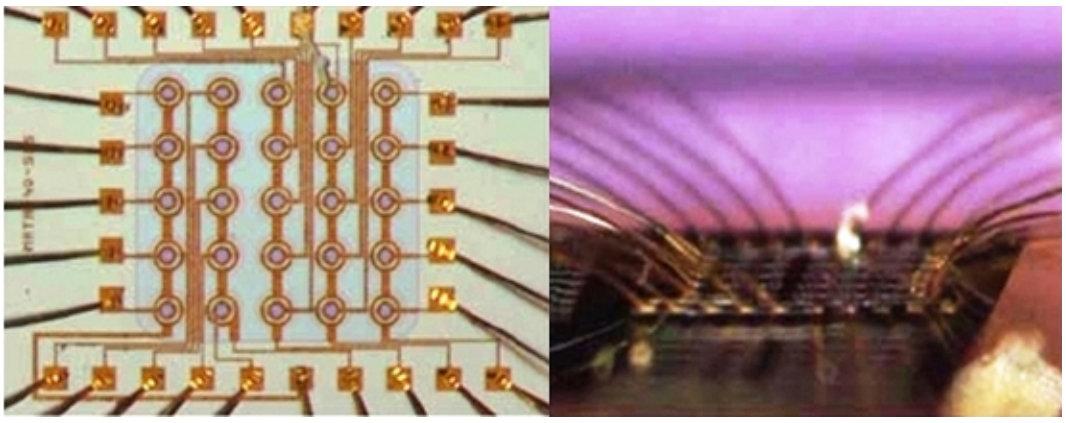

Figura 2.29

Immagini al microscopio ottico di una matrice di SPAD e dei fili di bonding.

Di seguito, infine, sono riportate due immagini (2.30) realizzate con un profilometro ad interferometria ottica in cui si evidenzia, in alto, il profilo tridimensionale di una matrice 5x5 di SPAD da 40 μm e, in basso, il profilo tridimensionale di un singolo pixel di tale matrice. La profondità è rappresentata dai colori, la cui legenda è riportata in alto a destra.

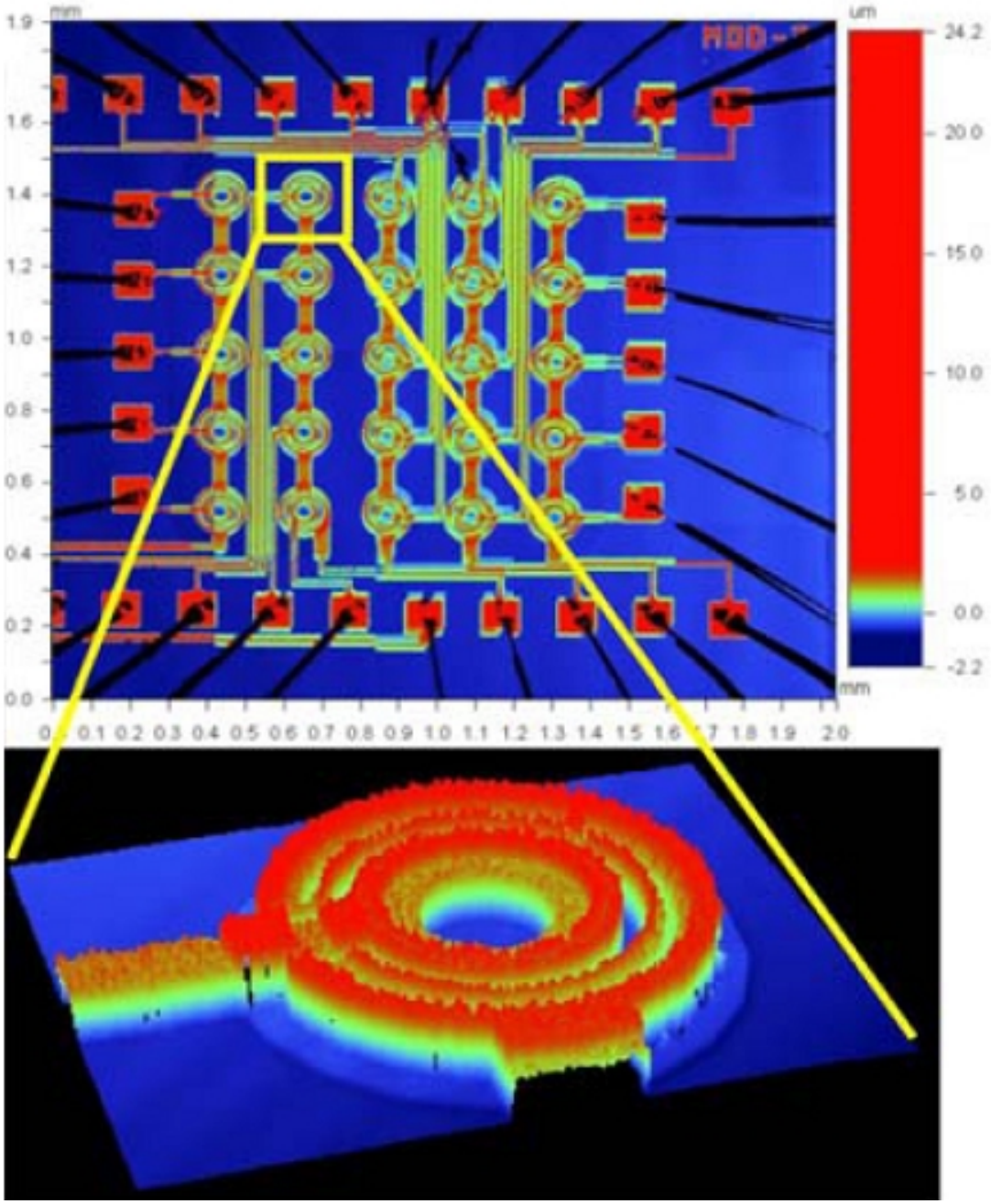

Figura 2.30
Immagini realizzate con un profilometro ad interferometria ottica di una matrice 5x5 di SPAD da 40 μm e di un singolo pixel.

Fine del Capitolo 2

Bibliografia

[1] S. Cova - F. Zappa - M. Ghioni "***Fotorivelatori microelettronici ultrasensibili***" - Politecnico di Milano Dipartimento di Elettronica e Informazione Milano, ALTA FREQUENZA - RIVISTA DI ELETTRONICA / Vol. 9 N. 1 / Gennaio-Febbraio 1997.

[2] H. Dautet - P. Deschamps - B. Dion - Andrew D. MacGregor - D. MacSween - Robert J. McIntyre - C. Trottier and Paul P. Webb "***Photon counting techniques with silicon avalanche photodiodes***" – APPLIED OPTICS / Vol. 32 No. 21 / 20 July 1993.

[3] P.P. Webb - R.J. McIntyre "***Large area reach through avalanche diodes for X-ray spectroscopy***" - IEEE Trans. Nucl. Sci., Vol. 23, pp. 138 – 144, 1976.

[4] E. Sciacca - A. C. Giudice - D. Sanfilippo - S. Lombardo - F. Zappa - R. Consentino - M. Mazzillo - M. Ghioni - G. Fallica - M. Belluso - G. Bonanno - S. Cova - E. Rimini "***Silicon Planar Technology for Single-Photon optical Detectors***" - IEEE TRANSACTIONS ON ELECTRON DEVICES / Vol. 50 N. 4 / April 2003.

[5] After Beadle et al. (1985), Fair (1981) and Tsai (1983).

[6] Richard S. Muller - Theodore I. - Kamins "***Dispositivi elettronici nei circuiti integrati***" - nuova edizione BOLLATI BORINGHIERI.

[7] V. Golovin - V. Saveliev "***Novel type of avalanche photodetector with Geiger mode operation***" - Nuclear Instruments and Methods in Physics Research A 518 (2004) 560-564.

[8] S. Barbarino - ***"Appunti di Campi Elettromagnetici - Anno Accademico 2003-2004"***.

CAPITOLO 3

CARATTERIZZAZIONE DELLA CARICA E DEL GUADAGNO DI UN DISPOSITIVO SINGOLO E DI UNA MATRICE PLANARE

In questo capitolo, sono presentati i risultati delle misure, eseguite su chip presso i *Laboratori Nazionali del Sud* dell'*Istituto Nazionale di Fisica Nucleare* di *Catania*, della carica di un singolo SPAD da 20 μm di diametro d'area attiva e di una matrice planare 5x5 di SPAD, sempre da 20 μm di diametro d'area attiva ciascuno, in configurazione SiPM. A tali misure si è cercato, inoltre, di dare delle spiegazioni e delle interpretazioni, facendo riferimento anche a successive misure, eseguite presso i laboratori della *STMicroelectronics* di *Catania*.

Il singolo dispositivo SPAD e la matrice planare oggetto di queste misure fanno parte del lotto Y503012 realizzato dal gruppo *R&D* della *STMicroelectronics* di *Catania*.

Prima della presentazione e dell'analisi delle misure è necessario, però, fare una piccola introduzione sul sistema di misura utilizzato sia per le misure sul dispositivo singolo sia per quelle sulla matrice planare.

3.1 Setup sperimentale

Le misure di carica sui dispositivi in esame sono state eseguite mantenendo il dispositivo "al buio", in maniera tale da non assorbire alcun fotone dall'esterno, e prelevando più volte il segnale di corrente in uscita dal singolo SPAD e dalla matrice, i cui pixel sono stati connessi in configurazione SiPM, su un carico da 50 Ω per diversi valori della tensione di polarizzazione. Il circuito di quenching utilizzato per queste misure è quello passivo con resistori di quenching da 100 kΩ di tipo SMD. Gli schemi circuitali utilizzati sono rappresentati in figura 1.13, per le misure sul singolo pixel, e in figura 1.18, per le misure in configurazione SiPM. Per le interconnessioni fra i pixel e le varie componenti discrete del circuito di quenching è

stata utilizzata una basetta di polarizzazione, sulla quale sono stati saldati i resistori di quenching ed inserito il chip, come mostrato nella seguente immagine.

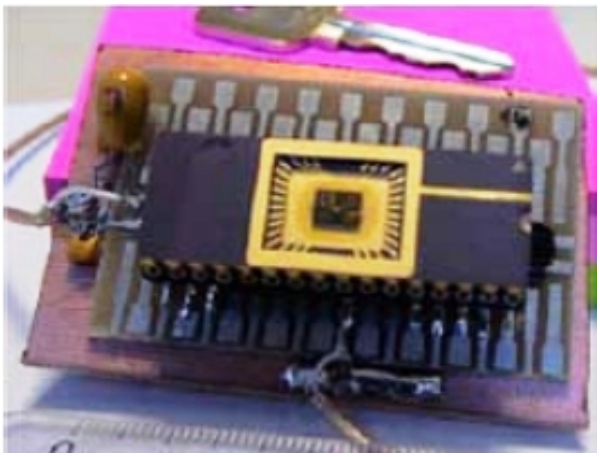

Figura 3.1

L'immagine mostra, in primo piano, il package, il chip con gli SPAD e la basetta di polarizzazione su cui sono saldati le componenti discrete del circuito di quenching passivo.

Lo schema utilizzato, invece, per processare e ricavare dalla corrente di uscita la carica che attraversa la giunzione del dispositivo durante una valanga è il seguente:

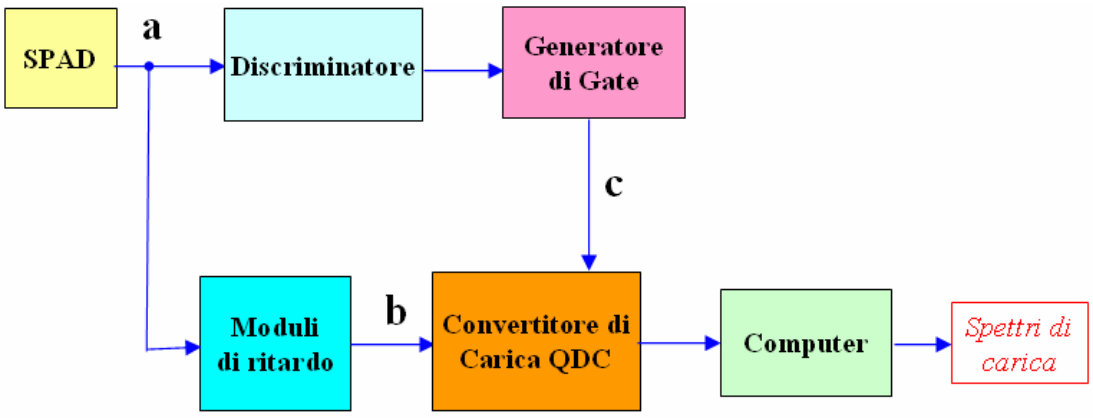

Figura 3.2

Schema utilizzato per processare il segnale in uscita dallo SPAD

Secondo tale schema, la tensione sul carico in uscita (a), dovuta alla corrente di valanga, viene letta ed inviata ad un *discriminatore a soglia* che emette un impulso non appena il segnale di tensione supera la soglia impostata (in queste misure è stata scelta una soglia di 20 mV, allo scopo di eliminare la possibile presenza di picchi di tensione dovuti a rumore di fondo e di attenuare l'incidenza sul segnale di eventuali impulsi spuri di bassa intensità). L'impulso in uscita dal discriminatore è una sorta di trigger per il *generatore di gate* che, una volta ricevuto il segnale dal discriminatore,

emette un altro impulso di durata finita, chiamato *gate di integrazione* (c), che servirà al *convertitore di carica QDC* (modello Silena 4418/Q) come tempo d'integrazione, appunto, per il segnale in uscita dal sensore, opportunamente ritardato (b). La scelta della durata del gate di integrazione e del ritardo del segnale in uscita dallo SPAD è stata fatta facendo riferimento ai segnali osservati all'oscilloscopio per assicurare un perfetto sincronismo fra il segnale in uscita dallo SPAD ed il gate di integrazione, senza il quale la misura sarebbe inficiata. La scelta della durata del gate d'integrazione è caduta sui 20 ns in modo da considerare solo il segnale strettamente legato alla fase di produzione della valanga e da trascurare, invece, la componente dovuta alla corrente in fase di scarica che tende a stabilizzarsi ad un valore, diverso da zero, pari ad I_f, come visto nel paragrafo 1.4.1.
La seguente figura riassume la sequenza dei segnali.

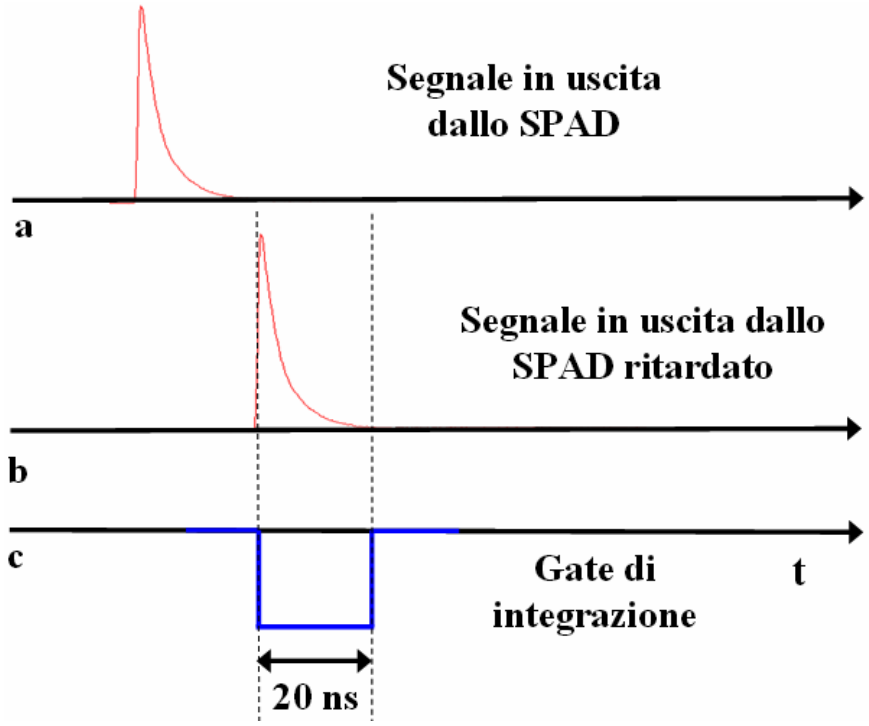

Figura 3.3
Sequenza dei segnali gestiti nella fase di acquisizione dati

Il convertitore di carica QDC esegue l'integrazione del segnale dello SPAD, opportunamente ritardato, sulla durata del segnale di gate (20 ns) e ne ricava, così, la quantità di carica che attraversa la giunzione durante la valanga. Tale informazione sulla carica è, inoltre, convertita dal convertitore in un segnale digitale tramite un sistema a 4096 canali (12 bit) con una calibrazione in carica di 62,5 fC: a ciascun canale corrisponde un dato valore di carica multiplo del passo di quantizzazione

(62,5 fC) ed il convertitore, dopo aver effettuato l'integrazione, assegna il risultato conseguito al canale corrispondente. L'operazione di acquisizione, fin qui vista, è ripetuta per ogni evento di valanga che supera la soglia del discriminatore durante la fase di acquisizione dati e per ciascuna di queste acquisizioni il segnale digitale in uscita dal convertitore è ricevuto dal computer, che, tramite il software di acquisizione e presentazione dati PAWX11, conta quanti eventi sono stati rilevati in ciascun canale e li rappresenta in un istogramma, chiamato **spettro di carica [1].**

Uno *spettro di carica* presenta alle ascisse i canali del convertitore di carica, di solito raggruppati in base 4 (si dice in gergo che il bin è 4 a 1) per velocizzare il controllo sugli spettri con basso conteggio, ed alle ordinate il numero di eventi contati per quel dato canale. La figura seguente (3.4) rappresenta uno spettro di carica dalla caratteristica forma segmentata, tipica degli istogrammi.

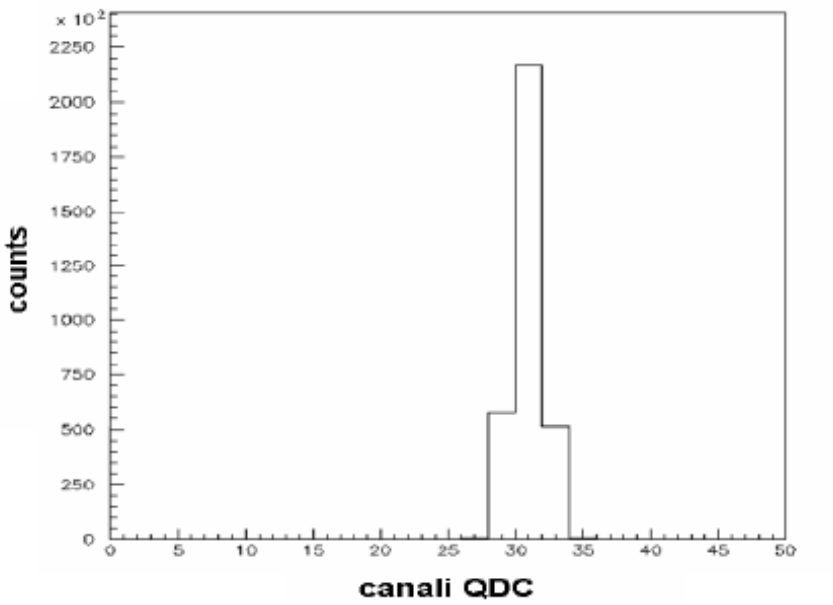

Figura 3.4

Esempio di uno spettro di carica osservato on line durante la fase di acquisizione dati.

A questo punto è possibile trattare e rielaborare questi dati in maniera statistica e ricavare le informazioni desiderate.

Lo scopo delle misure di questo capitolo è, appunto, caratterizzare la carica ed il guadagno di un dispositivo singolo, prima, e di una matrice in configurazione SiPM, dopo, a partire dagli spettri di carica.

Nei paragrafi seguenti sono riportati i risultati delle misure di carica, con le rispettive analisi e riflessioni, su singolo pixel e sulla matrice planare in configurazione SiPM.

3.2 Risultati ed analisi delle misure di carica e guadagno su singolo pixel

In questo paragrafo sono presentati ed analizzati i risultati delle misure di carica e guadagno su un singolo SPAD. Il dispositivo in questione è il pixel K11 della matrice da 20 μm di diametro d'area attiva del lotto Y503012. La misura, come detto in precedenza, è stata eseguita su chip ed è consistita nell'acquisizione "al buio" degli spettri di carica per quattro diverse tensioni (33 V, 35 V, 37 V e 39 V), tutte sopra la tensione di breakdown nominale del dispositivo, pari a circa 30,8 V, ed a temperatura ambiente. La durata di ciascuna acquisizione è stata di circa qualche minuto. Lo schema circuitale utilizzato è descritto nella figura 1.13. Il resistore di quenching utilizzato è pari a 100 kΩ, mentre il carico è di 50 Ω.

Gli spettri di carica ottenuti per le quattro diverse tensioni sono i seguenti, in scala lineare:

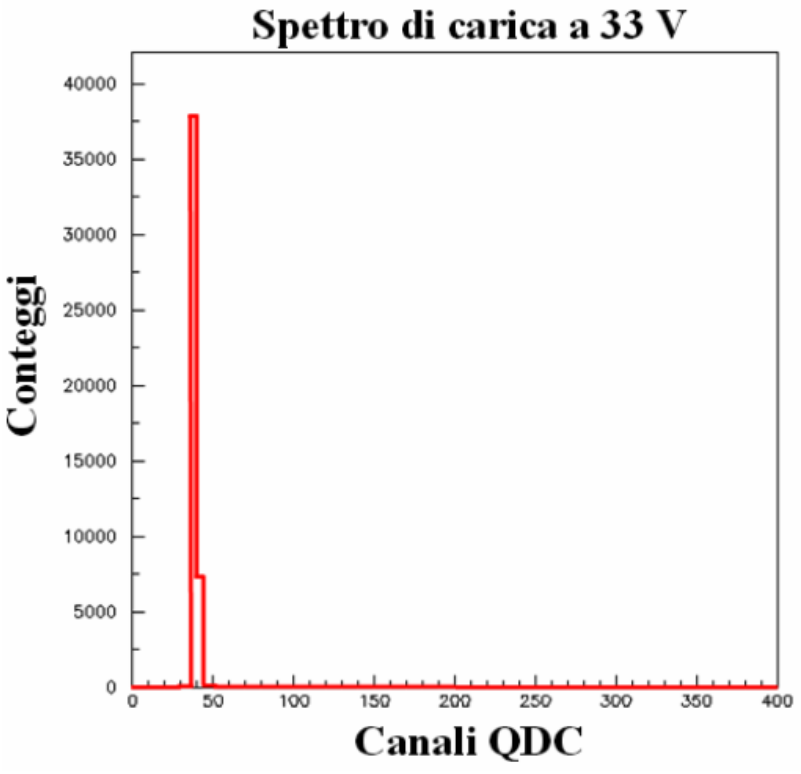

Figura 3.5
Spettro di carica del pixel K11 alla tensione di polarizzazione di 33 V. Si noti come l'andamento di tale spettro tenda ad una delta di Dirac.

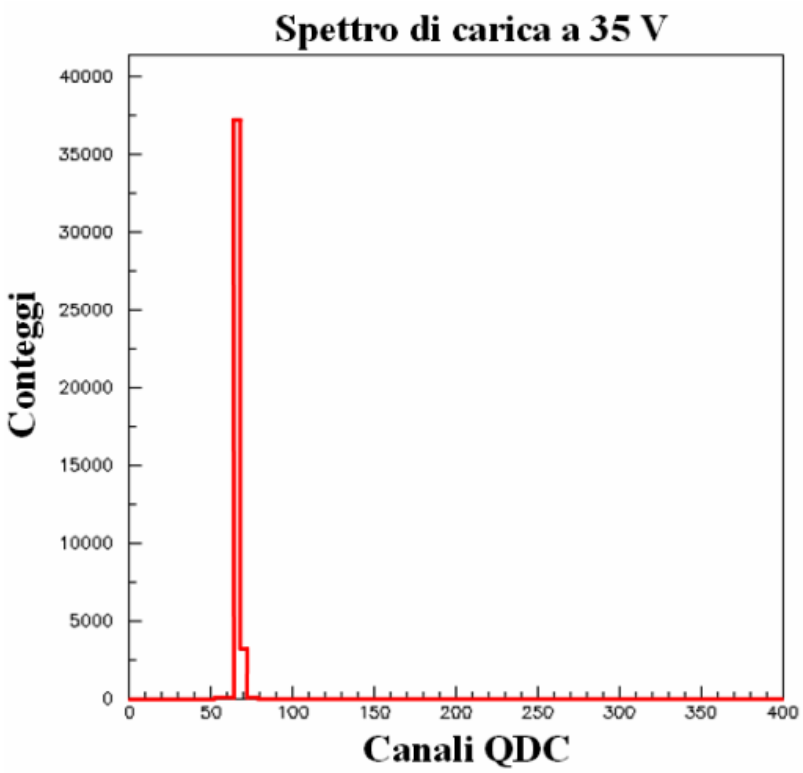

Figura 3.6

Spettro di carica del pixel K11 alla tensione di polarizzazione di 35 V.
Si noti come tale spettro assomigli ad una traslazione verso destra di quello a 33 V.

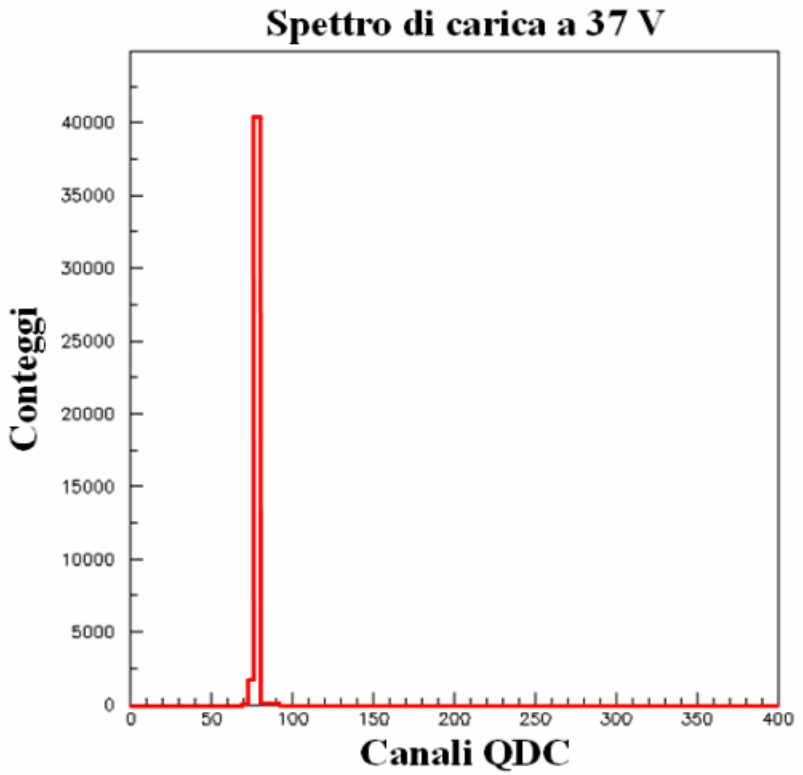

Figura 3.7

Spettro di carica del pixel K11 alla tensione di polarizzazione di 37 V.
Si noti l'incipit di un leggero allargamento verso sinistra.

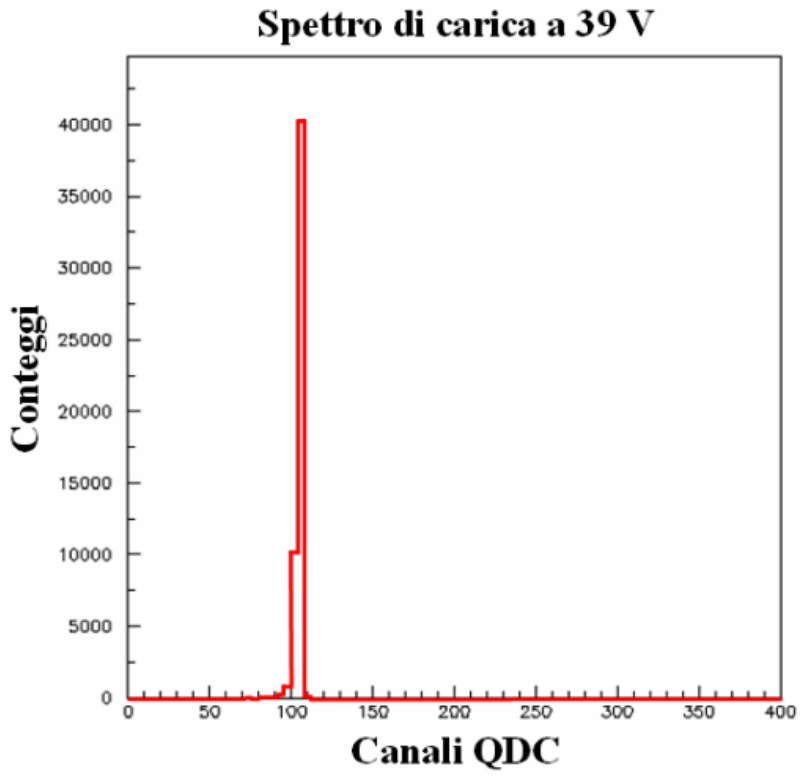

Figura 3.8

Spettro di carica del pixel K11 alla tensione di polarizzazione di 39V. Si osservi l'allargamento dello spettro, notevolmente maggiore, rispetto agli spettri ottenuti a tensioni di polarizzazione più basse.

I risultati ottenuti da questi spettri di carica sono riconducibili ad una funzione gaussiana con un valore medio (la carica media in una valanga per il pixel), un'ampiezza, una deviazione standard (σ) ed un'ampiezza a metà altezza (FWHM). Per tale motivo i dati ottenuti sono stati trattati ed interpolati con una funzione gaussiana, allo scopo di ricavare carica media, deviazione standard e FWHM.

La funzione matematica utilizzata per fare il fit di questi dati è la seguente:

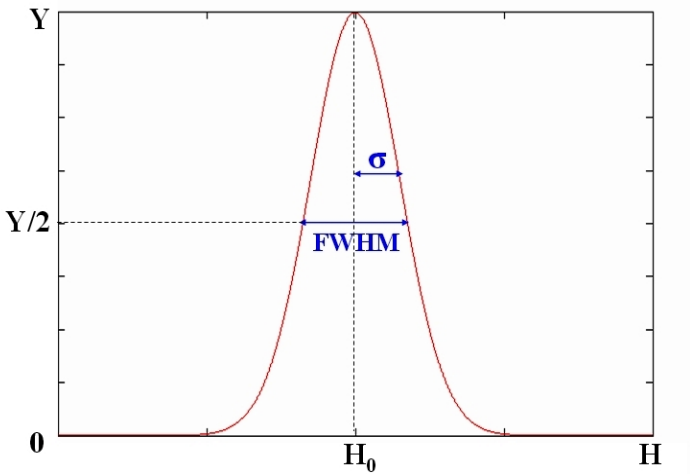

Figura 3.9

Funzione gaussiana utilizzata per il fit dei dati

L'espressione di questa funzione matematica è la seguente:

$$G(H) = \frac{A}{\sigma \cdot \sqrt{2\pi}} \cdot e^{-\left(\frac{(H-H_0)^2}{2\cdot\sigma^2}\right)} \tag{3.1}$$

in cui A è l'ampiezza, H_0 il centroide (l'ascissa per cui si ha il valore di picco in ampiezza e, in questo caso, corrisponde alla carica media del pixel) e σ la deviazione standard. La deviazione standard σ è relazionata con il FWHM dalla formula:

$$FWHM \cong 2{,}35 \cdot \sigma \tag{3.2}$$

Viene, inoltre, definita come *risoluzione in ampiezza* R dello spettro il rapporto fra il FWHM e il centroide H_0:

$$R = \frac{FWHM}{H_0} \tag{3.3}$$

R ha un valore adimensionale ed è espressa, di solito, in percentuale **[2]**. Tale valore può essere considerato, in alcune applicazioni (come la rivelazione di particelle in Fisica Nucleare), un parametro importante sulle prestazioni del sensore. Infatti, tanto più basso è il valore della risoluzione in ampiezza R, tanto più lo spettro tenderà ad una delta di Dirac ed il rivelatore, avendo la sua risposta meno fluttuazioni, riuscirà a distinguere ed a separare meglio gli impulsi di luce che contengono un differente numero di fotoni **[1]**.

In base a queste osservazioni è stato eseguito il fit dei dati ricavati, ottenendo:

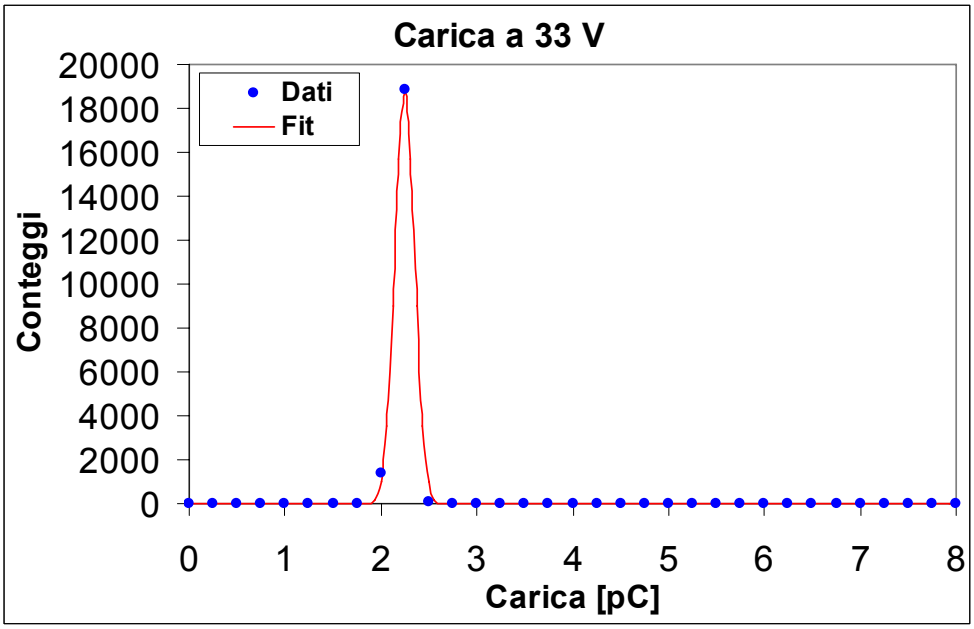

Figura 3.10
Fit gaussiano (in rosso) dei dati (in blu) ricavati a 33 V.
Si osservi come il fit coincida, quasi fedelmente, con i dati.

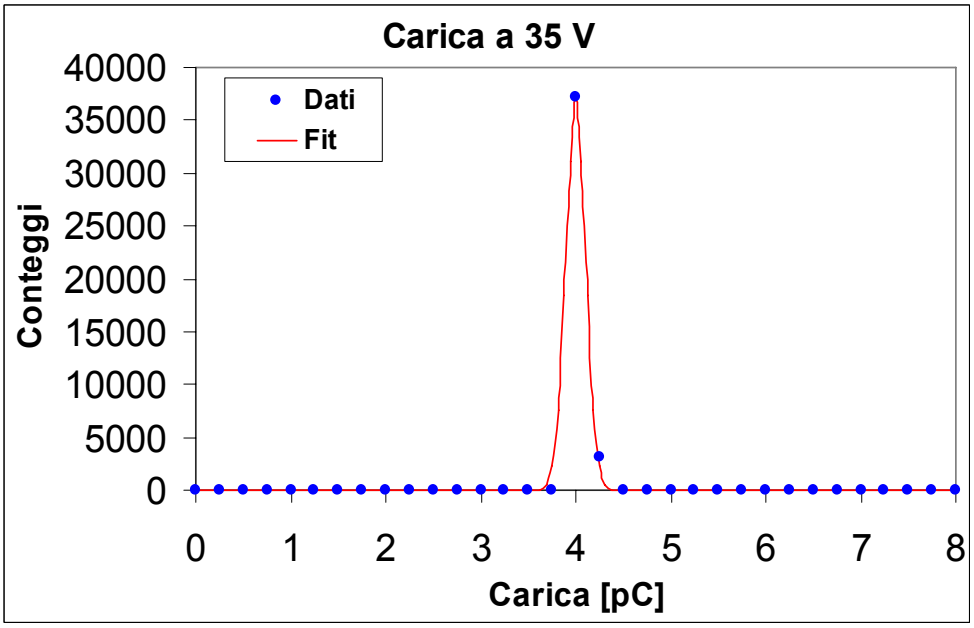

Figura 3.11
Fit gaussiano (in rosso) dei dati (in blu) ricavati a 35 V.
Si noti l'allargamento dei dati verso destra rispetto allo spettro ottenuto a 33 V.

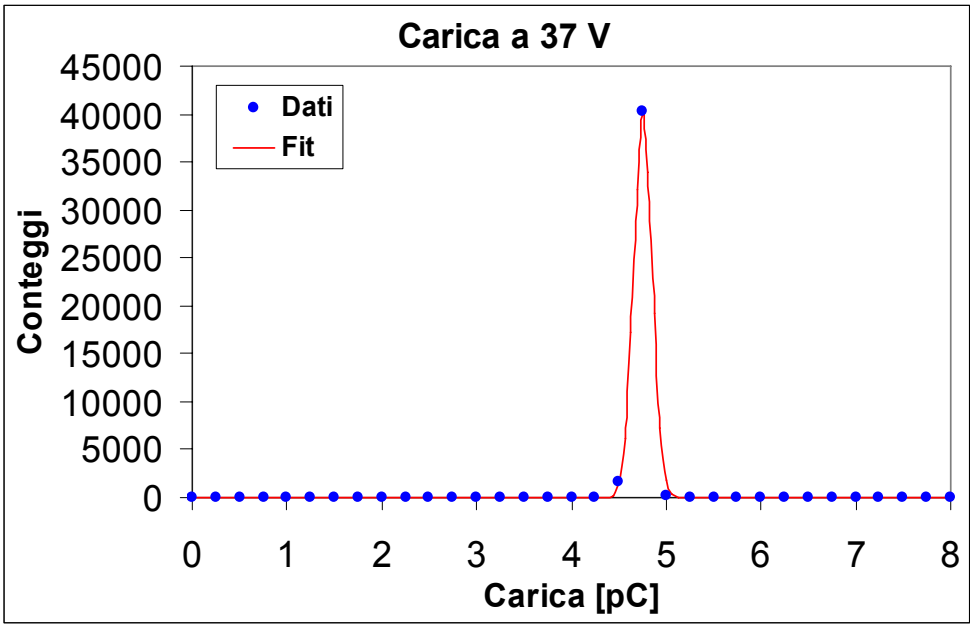

Figura 3.12
Fit gaussiano (in rosso) dei dati (in blu) ricavati a 37 V.
Si noti l'allargamento verso sinistra dei dati ricavati.

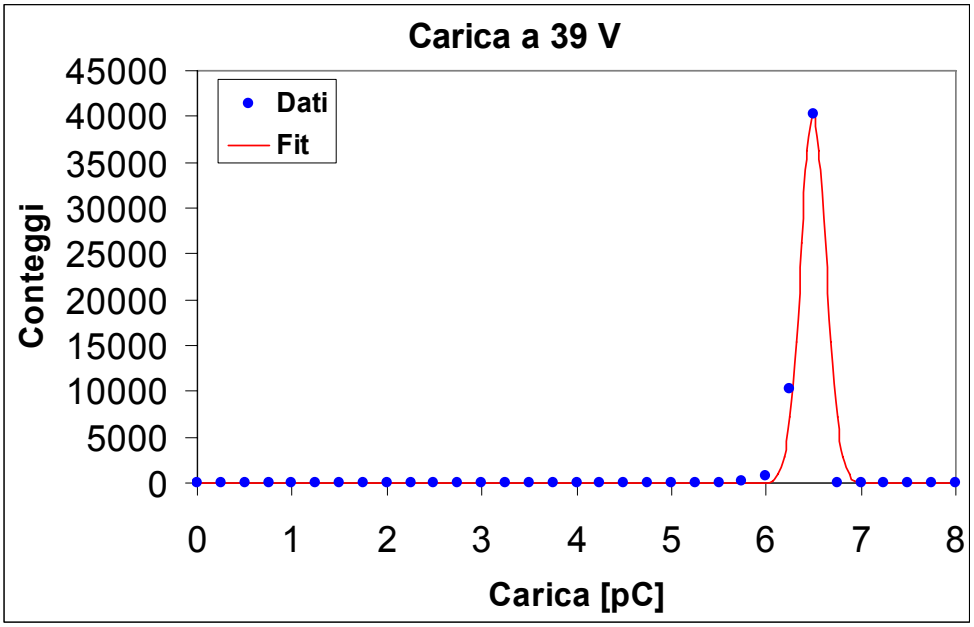

Figura 3.13
Fit gaussiano (in rosso) dei dati (in blu) ricavati a 39 V.
Si noti come il fit si discosti dai dati.

La seguente tabella, invece, riassume i risultati ottenuti alle quattro diverse tensioni applicate:

V_A	Centroide [pC]	σ	FWHM	Risoluzione	Guadagno
33V	2,25	$0{,}102 \cdot 10^{-12}$	$0{,}241 \cdot 10^{-12}$	10,71 %	$1{,}404 \cdot 10^{7}$
35V	4,00	$0{,}105 \cdot 10^{-12}$	$0{,}247 \cdot 10^{-12}$	6,18 %	$2{,}497 \cdot 10^{7}$
37V	4,75	$0{,}099 \cdot 10^{-12}$	$0{,}233 \cdot 10^{-12}$	4,9 %	$2{,}965 \cdot 10^{7}$
39V	6,50	$0{,}135 \cdot 10^{-12}$	$0{,}317 \cdot 10^{-12}$	4,88 %	$4{,}057 \cdot 10^{7}$

Tabella 3.1
Tabella riassuntiva dei dati ottenuti per il pixel K11

Bisogna, tuttavia, osservare che il fit è da considerarsi uno strumento puramente indicativo per la lettura dei dati, soprattutto per la presenza di pochissimi valori con ordinata diversa da zero, ma che si dimostra molto utile in questa trattazione. Come si può notare dai risultati ottenuti, la carica media (il centroide) e, di conseguenza, anche il guadagno aumentano con la tensione applicata, come già ampiamente previsto dalla formula 1.50.

La risoluzione percentuale in ampiezza si riduce e, di conseguenza, migliora al crescere della tensione applicata, questo perchè la risoluzione è inversamente proporzionale alla carica media (che, come visto, cresce con l'incremento della

tensione). Ciononostante, si osserva, a tensioni più alte, un progressivo aumento della deviazione standard σ e dell'ampiezza a metà altezza FWHM. Questo comporta un allargamento degli spettri ed una certa indeterminatezza nella risposta in carica del rivelatore.

L'allargamento degli spettri di carica ha molteplici spiegazioni. Una prima interpretazione sarebbe legata alle fluttuazioni statistiche tipiche del fenomeno della valanga, che incorporerebbe in se un numero differente di portatori per l'influenza di diversi fattori, uno di questi, ad esempio, potrebbe essere il punto in cui s'innescherebbe la valanga: ci sarà un valanga più "ricca" di elettroni quando il suo punto d'innesco è più vicino al centro dell'area attiva del dispositivo, più povera, invece, quando è prossimo ai bordi dell'area attiva.

Un'altra spiegazione sarebbe legata alla scarsa efficacia del quenching a tensioni elevate. Infatti, a tensioni più alte il valore del resistore di quenching scelto (100 kΩ) non è più in grado di provocare ai propri capi una caduta di tensione tale da spegnere completamente la valanga innescata nel dispositivo. Quindi, il valore I_f, cui tende esponenzialmente la corrente dello SPAD nella fase di scarica, non sarà inferiore alla corrente di latching I_q e, così, la corrente di valanga, ad elevate tensioni, non riuscirà ad annullarsi, com'è possibile osservare nelle seguenti forme d'onda medie (su 10000 campioni), acquisite con un oscilloscopio digitale, della corrente di valanga dello stesso dispositivo con un resistore di quenching di 100 kΩ.

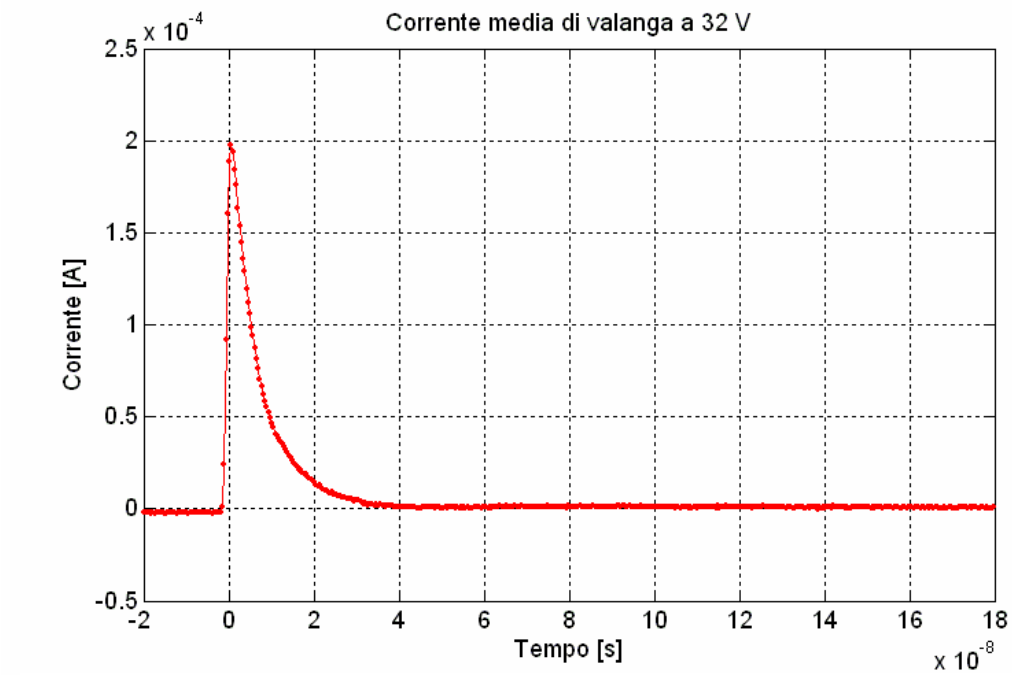

Figura 3.14

Forma d'onda media della corrente di valanga dello SPAD a 32 V. Si osservi come la corrente si annulli dopo poco più di 30 ns.

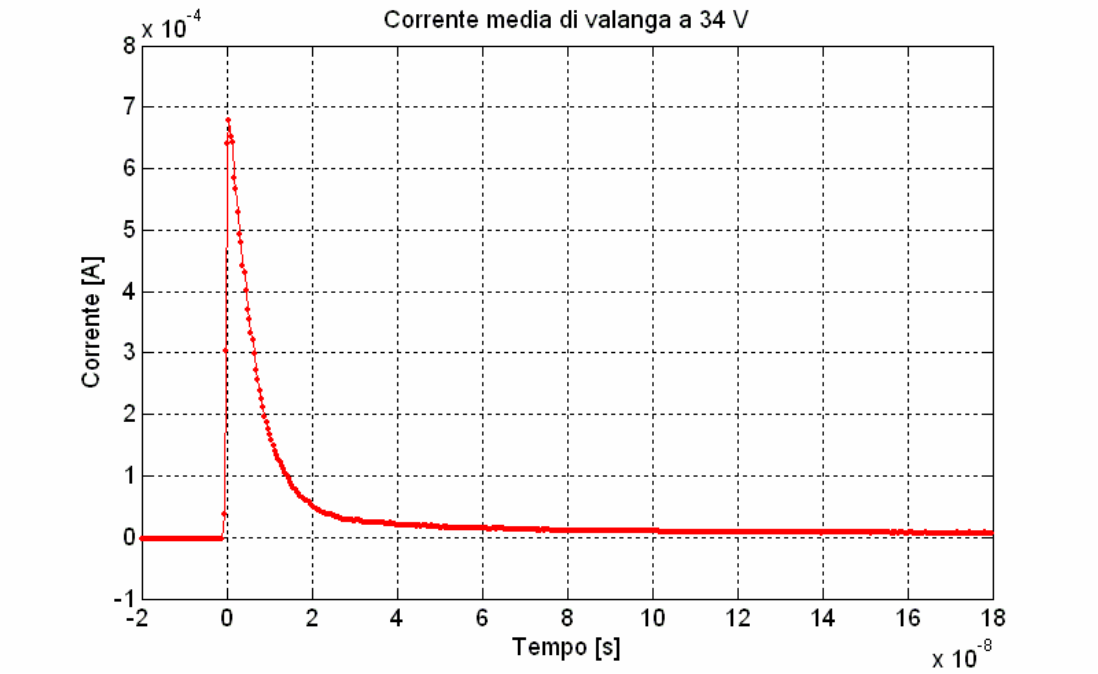

Figura 3.15

Forma d'onda media della corrente di valanga dello SPAD a 34 V.
Si noti come la corrente arrivi ad annullarsi dopo un tempo molto lungo.

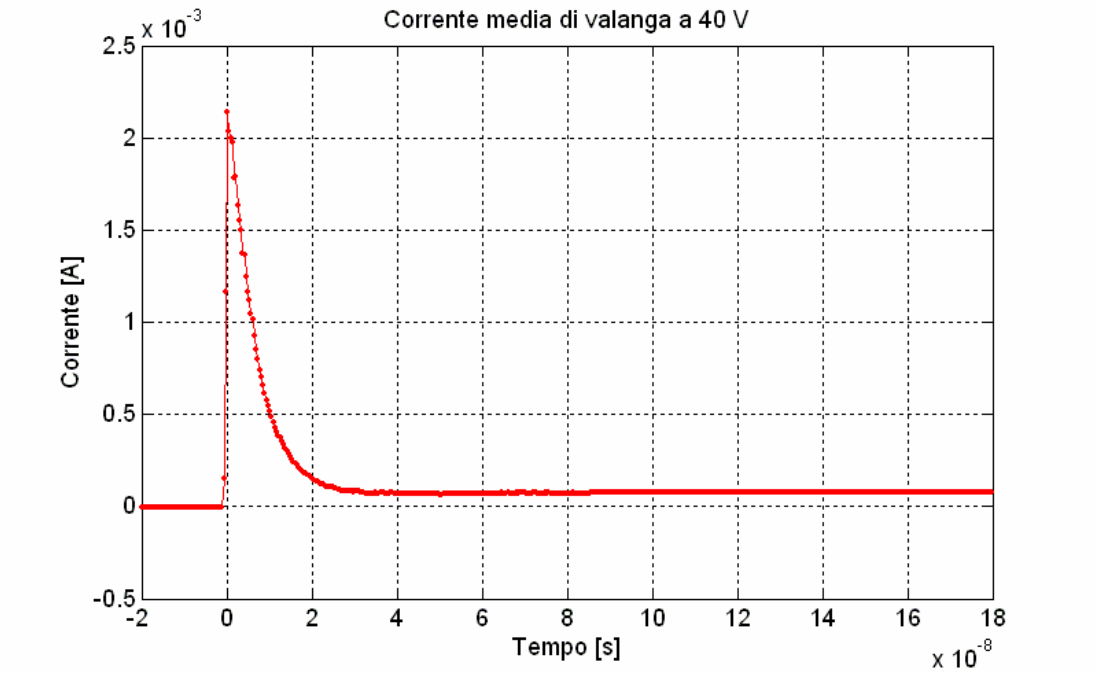

Figura 3.16

Forma d'onda media della corrente di valanga dello SPAD a 40 V.
Si osservi come la corrente non riesca ad annullarsi.

Da queste forme d'onda medie della corrente di valanga si può notare che solo a 32 V il resistore di quenching da 100 kΩ riesce a spegnere efficacemente la valanga, mentre già a 34 V il quenching diventa più problematico, richiedendo un tempo più lungo, ed a tensioni di polarizzazione ancora più alte (ad esempio a 40 V)

la corrente di valanga non viene assolutamente spenta dal resistore di quenching (ne occorrerebbe uno di dimensioni maggiori oppure si potrebbe ricorrere ad un circuito di quenching attivo). La mancanza di un quenching efficace e del conseguente annullamento della corrente di valanga porta il convertitore di carica QDC ad integrare un'area maggiore, sottesa alla forma d'onda dell'impulso della corrente di valanga, e, quindi, in uscita darà un valore di carica più alto, che, logicamente, allargherà lo spettro verso destra.

L'inefficacia del circuito di quenching passivo è complice, inoltre, di un ulteriore allargamento dello spettro di carica, stavolta, però, verso sinistra. Tale fenomeno, ben osservabile negli spettri ricavati a 37 V ed a 39 V, è provocato principalmente dagli eventi d'afterpulsing generati durante il lungo tempo di ricarica dello SPAD e di spegnimento della corrente di valanga. Infatti, immediatamente dopo un evento di valanga si ha un'elevata densità di portatori intrappolati nei livelli energetici intermedi fra la banda di valenza e quella di conduzione e, di conseguenza, un'elevata probabilità d'afterpulsing. Poiché il dispositivo, nella fase di spegnimento della valanga, continua ad essere polarizzato ancora poco sopra la propria tensione di breakdown ed il tempo di spegnimento della valanga è piuttosto lungo, tali portatori, essendo progressivamente rilasciati dalle trappole durante questo intervallo di tempo, potranno innescare delle piccole valanghe con una quantità di carica inferiore a quella media, che, nello spettro, si rifletterà con un allargamento verso sinistra ed un discostamento dei dati dal fit gaussiano **[3]**. Anche i portatori generati termicamente durante la fase di ricarica dello SPAD potranno innescare altre piccole valanghe, che finiranno anch'esse per allargare gli spettri di carica. Ovviamente, come osservato dagli spettri, la probabilità di generare tali valanghe cresce con l'incremento della tensione di polarizzazione, non solo in accordo alla formula 1.22 sulla probabilità d'innesco di una valanga, ma anche in accordo al fatto che la probabilità d'afterpulsing cresce all'aumentare della sovratensione di polarizzazione applicata, secondo la relazione 1.28. Il fatto che tali fenomeni siano stati osservati meglio a tensioni più elevate è dovuto, inoltre, alla presenza, nel sistema di misura utilizzato, di un discriminatore a soglia di tensione. Infatti, a tensioni più basse (33 V e 35 V) è assai più improbabile che la tensione prodotta sul carico di uscita (50 Ω) dalle correnti dovute a queste piccole valanghe, innescate durante la fase di ricarica, riesca a superare la soglia di sbarramento del discriminatore (posta, come riferito in precedenza, a 20 mV).

Per concludere, infine, la caratterizzazione della carica di valanga di un singolo pixel, si è pensato di linearizzare con una retta i singoli valori di carica media ottenuti alle diverse tensioni di polarizzazione, come si vede nella seguente figura:

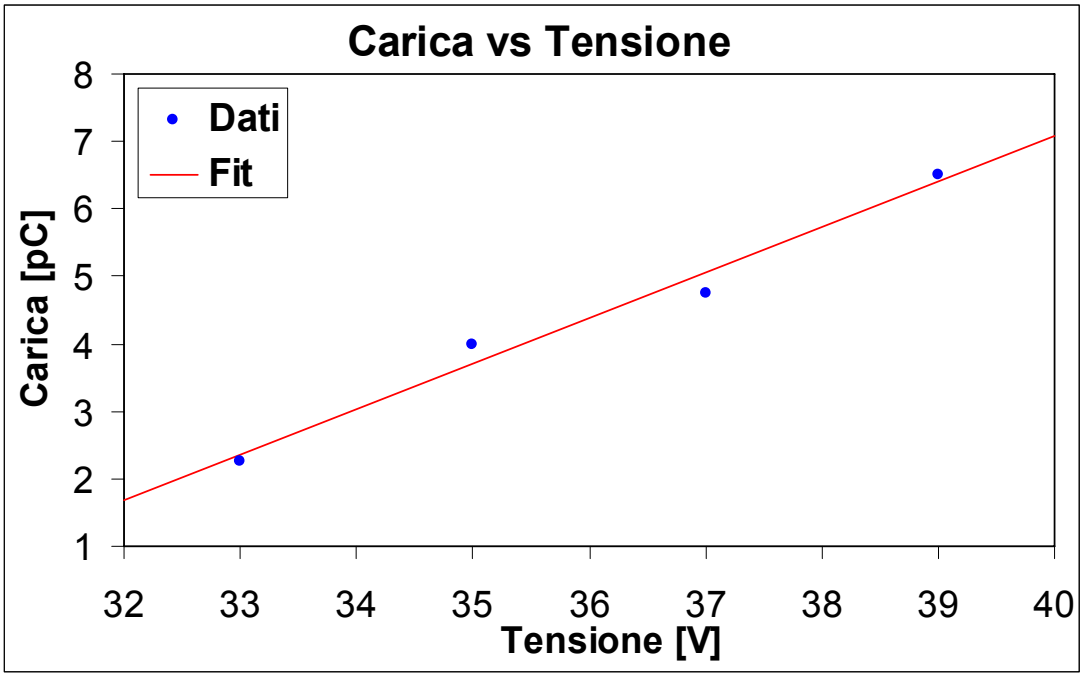

Figura 3.17
Rappresentazione della retta di linearizzazione carica-tensione (in rosso) dei dati sperimentali di carica ottenuti alle varie tensioni (puntini blu).

L'equazione della retta ottenuta e rappresentata in figura è:

$$Q = 0{,}625 \cdot 10^{-12} \cdot V - 1{,}9925 \cdot 10^{-11} \qquad (3.4)$$

dove V è la tensione applicata e Q è la carica media rivelata.

Confrontando l'equazione di questa retta con la relazione che esprime la carica media di un impulso di valanga **[4]**:

$$Q_{media} = C_{pixel} \cdot (V_{pol} - V_B) \qquad (3.5)$$

in cui Q_{media} è la carica media, C_{pixel} è la capacità intrinseca del pixel, V_{pol} è la tensione di polarizzazione e V_B è la tensione di breakdown, si ricavano i valori della capacità dello SPAD (pari a circa 0,625 pF) e della sua tensione di breakdown (pari a circa 29,52 V).

3.3 Risultati ed analisi delle misure di carica su una matrice planare in configurazione SiPM

In questo paragrafo sono presentati ed analizzati i risultati delle misure di carica e guadagno su una matrice da 20 μm di diametro d'area attiva del lotto Y503012, in configurazione SiPM. Lo schema circuitale utilizzato è, come detto in precedenza, rappresentato in figura 1.18; i resistori di quenching utilizzati sono da 100 kΩ, mentre il carico comune è di 50 Ω. Com'è già noto, la misura è stata eseguita su chip e per interconnettere i vari pixel della matrice con i rispettivi resistori di quenching e con il carico comune, simulando il comportamento del SiPM, si è fatto uso di una basetta di polarizzazione (rappresentata in figura 3.1). La misura è consistita nell'acquisizione, "al buio", degli spettri di carica per cinque diverse tensioni (32 V, 34 V, 36 V, 38 V e 40 V), tutte sopra la tensione di breakdown nominale del dispositivo, pari a circa 30,8 V, ed a temperatura ambiente. Si è, inoltre, misurata la durata di ogni acquisizione: 14 minuti a 32 V, 4 minuti a 34 V, 5 minuti a 36 V, 3 minuti a 38 V e 2 minuti a 40 V.

La scelta di una durata differente per ciascuna acquisizione non è casuale ed è legata al differente numero di eventi rilevabili nell'unità di tempo alle diverse tensioni di polarizzazione. Infatti, come ormai è risaputo dalla relazione 1.22, la probabilità di trigger di una valanga cresce con l'overvoltage applicato e, quindi, a tensioni più basse il numero di eventi rilevabile nell'unità di tempo è inferiore e, perciò, occorrerebbe un tempo più lungo per poterne osservare un numero statisticamente rilevante. L'unica eccezione a tale regola empirica è rappresentata dall'acquisizione eseguita a 34 V.

Agli spettri di carica ottenuti ed alle loro consuete elaborazioni sono state, in seguito, integrate delle misure di capacità, breakdown e dark count al fine di indagare e spiegare meglio l'andamento di tali spettri.

Seguono gli spettri di carica ottenuti alle cinque diverse tensioni di polarizzazione, rappresentati in scala semilogaritmica.

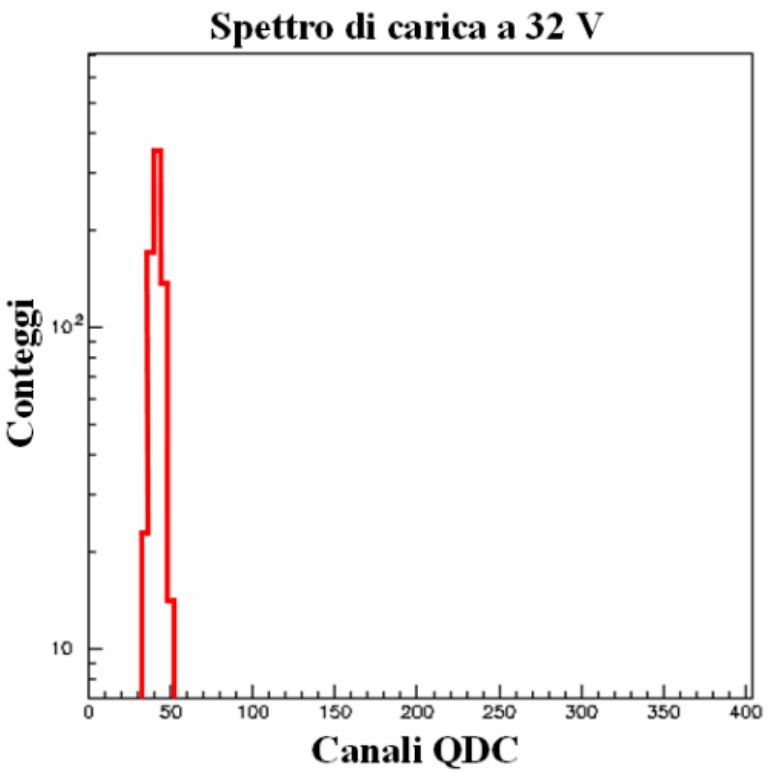

Figura 3.18

Spettro di carica ottenuto a 32 V in configurazione SiPM.
Si noti una leggerissima somiglianza con gli spettri ricavati per singolo pixel, con una sostanziale differenza nella larghezza dello spettro, maggiore nel caso SiPM.

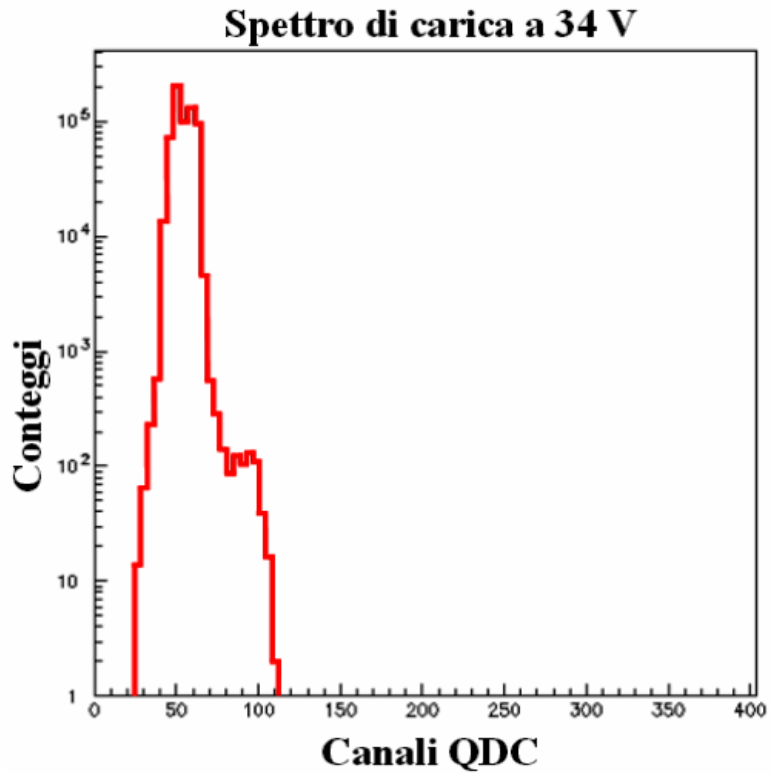

Figura 3.19

Spettro di carica ottenuto a 34 V in configurazione SiPM.
Si osservi la deformazione dello spettro, rispetto all'acquisizione a 32, con l'allargamento verso destra e la presenza di più picchi.

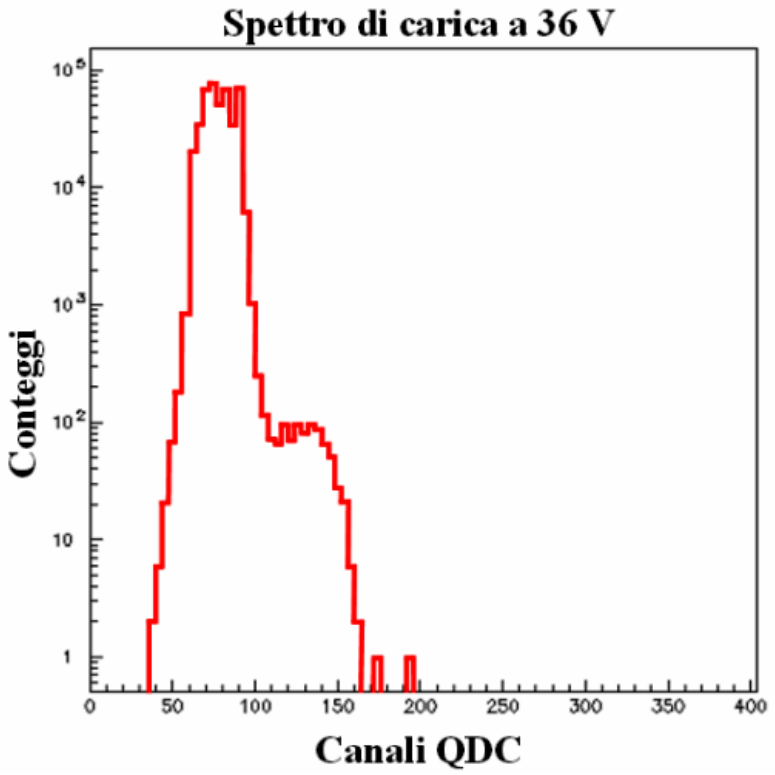

Figura 3.20

Spettro di carica ottenuto a 36 V in configurazione SiPM.
Si noti la maggiore deformazione rispetto all'acquisizione a 34 V: sembrerebbe quasi la sovrapposizione di due gaussiane con differenti valori medi, massimi e deviazioni standard.

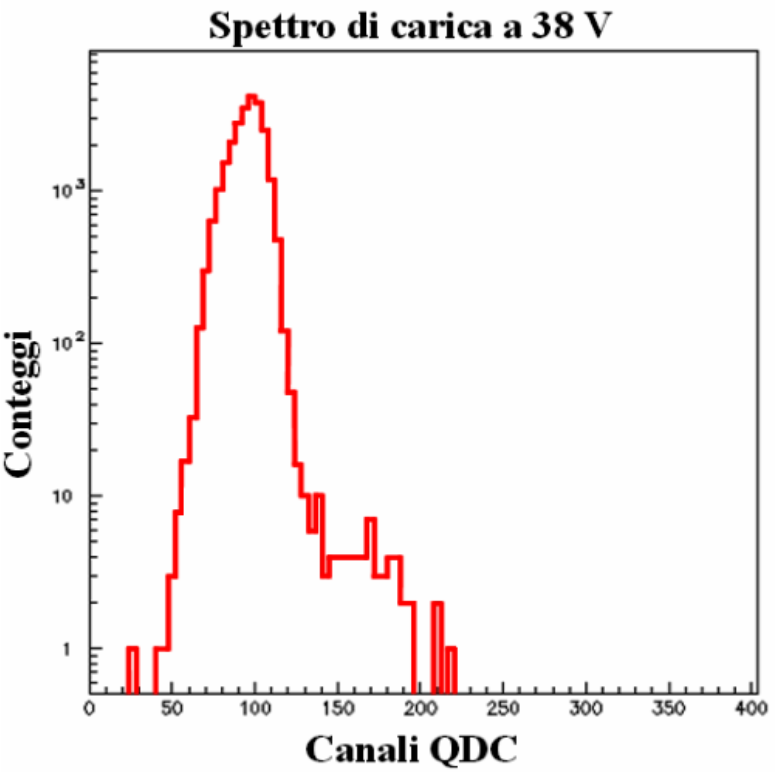

Figura 3.21

Spettro di carica ottenuto a 38 V in configurazione SiPM.
Si osservi l'allargamento dello spettro rispetto alle acquisizioni a tensioni più basse.

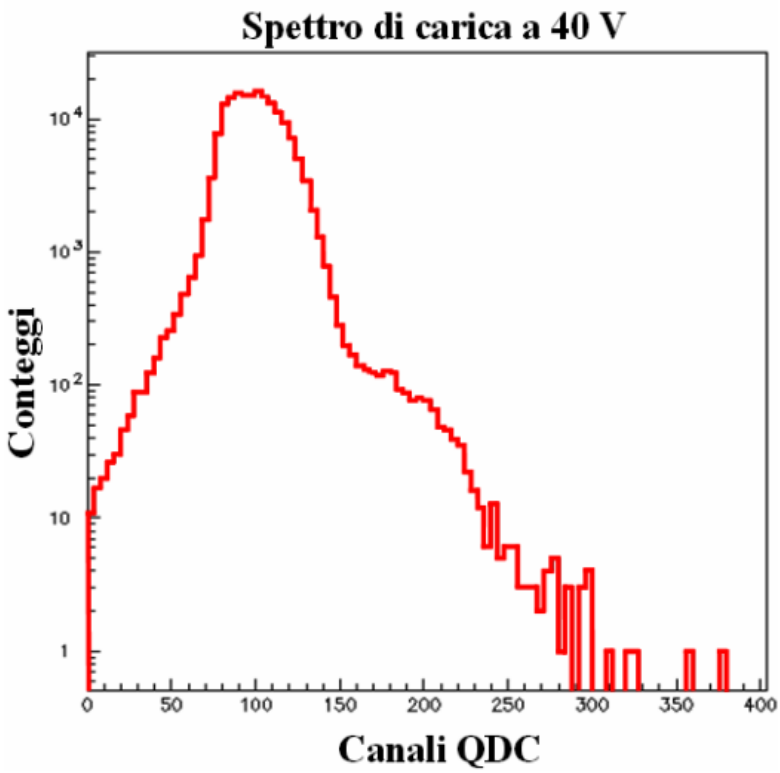

Figura 3.22
Spettro di carica ottenuto a 40 V in configurazione SiPM.
L'allargamento dello spettro è tale da coinvolgere quasi tutti i canali del convertitore QDC.

Seguono le elaborazioni degli spettri di carica acquisiti, sempre in scala semilogaritmica.

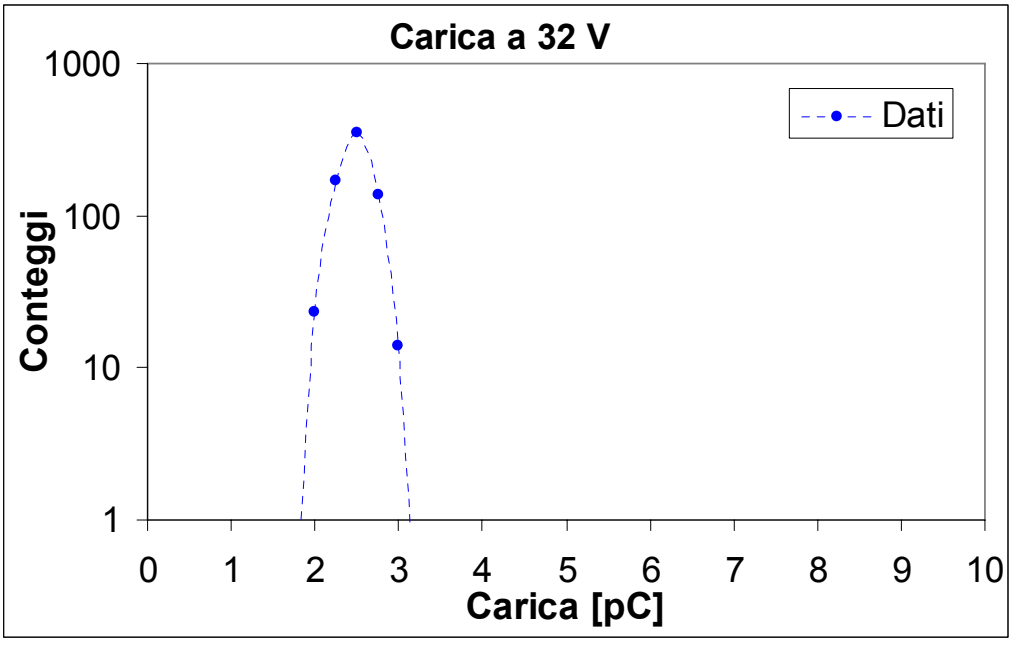

Figura 3.23
Elaborazione dello spettro di carica a 32 V.
La forma ricalca quella di una funzione gaussiana.

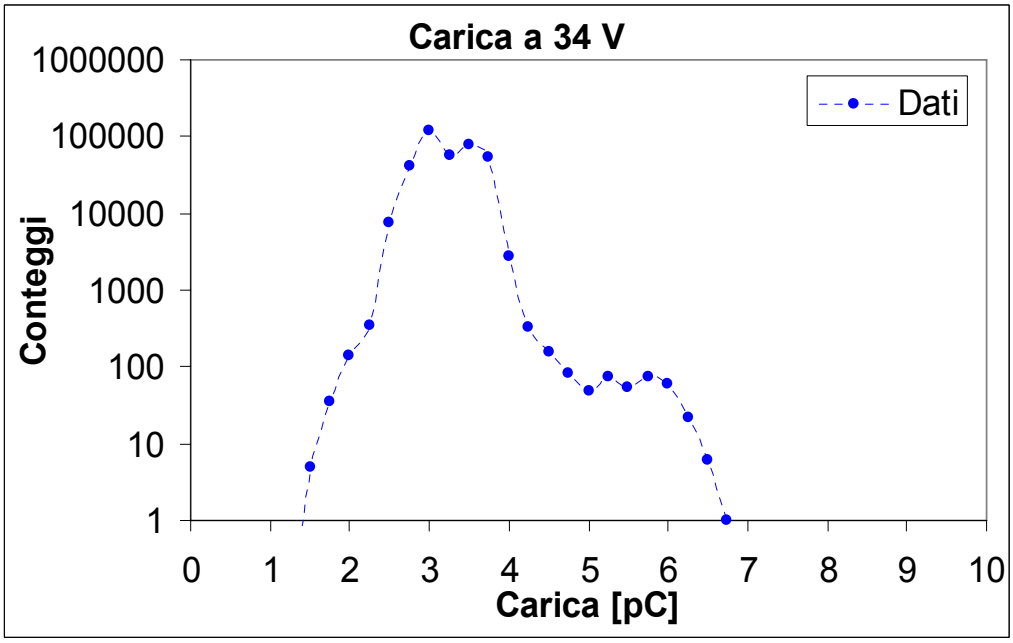

Figura 3.24
Elaborazione dello spettro di carica a 34 V.
La forma non è più riconducibile ad una funzione gaussiana.
Si assiste ad un allargamento verso destra.

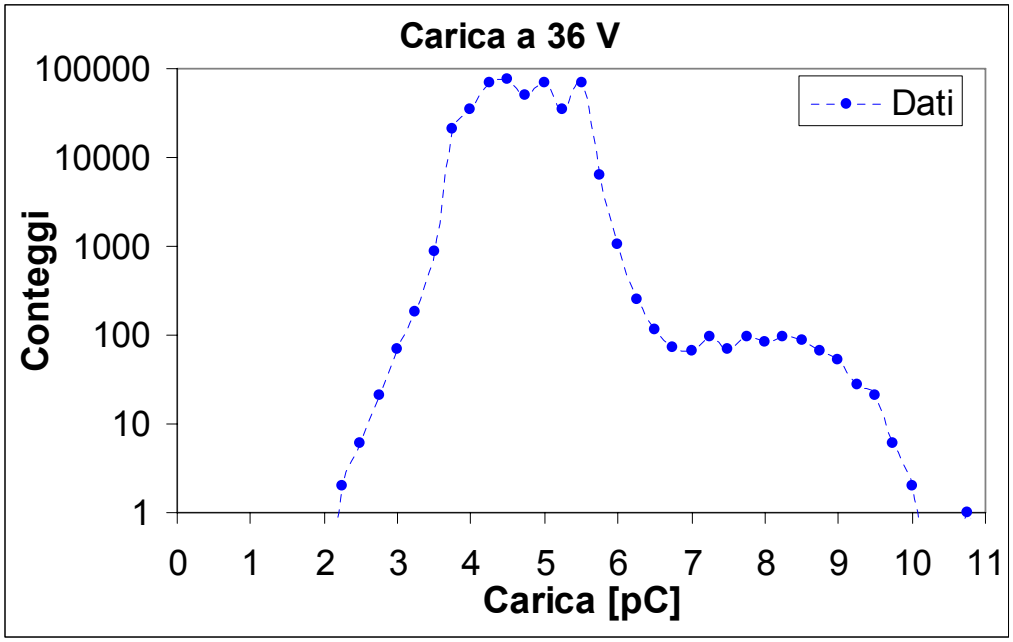

Figura 3.25
Elaborazione dello spettro di carica a 36 V.
L'allargamento è considerevole e la presenza di più valori di picco lascia riflettere.

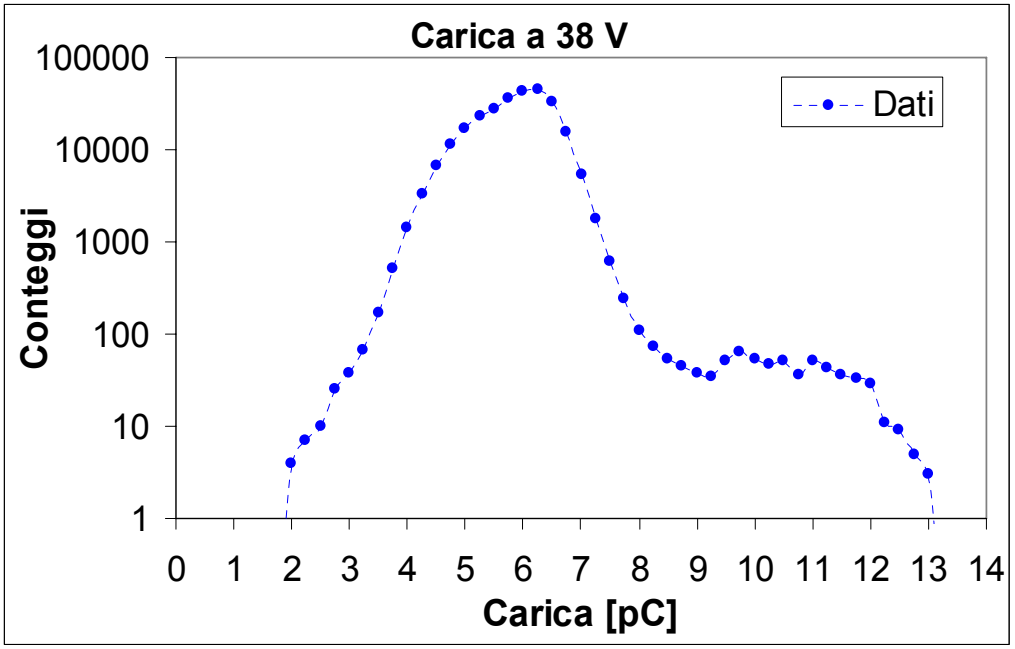

Figura 3.26
Elaborazione dello spettro di carica a 38 V.
La deformazione e l'allargamento sono in crescita rispetto all'acquisizione precedente.

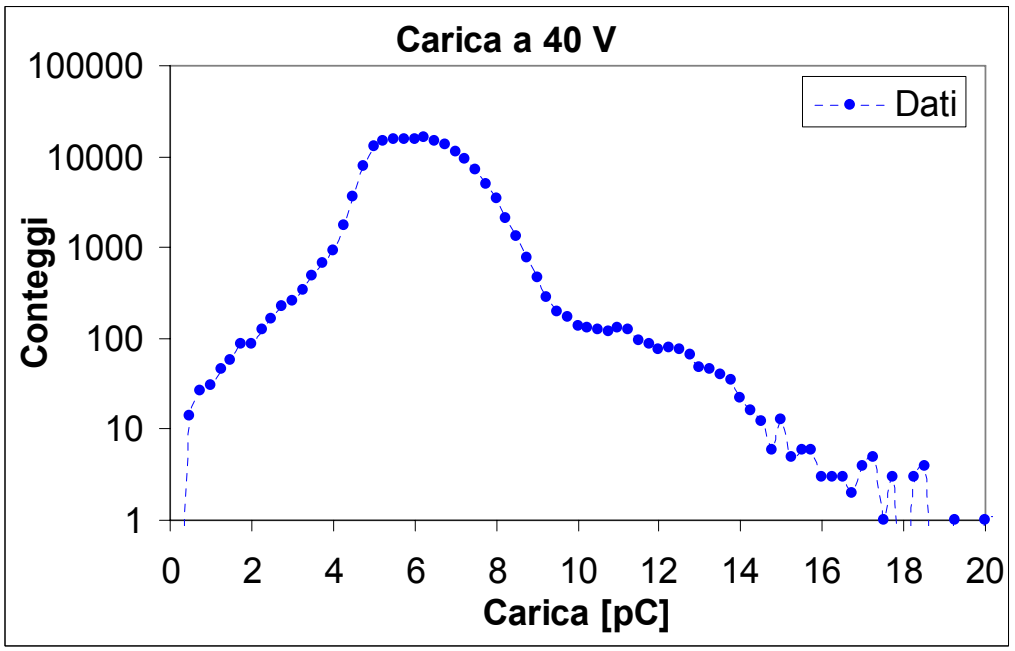

Figura 3.27
Elaborazione dello spettro di carica a 40 V.
La deformazione e l'allargamento sono superiori ad ogni altra acquisizione.

Osservando gli spettri di carica ottenuti e le loro rispettive elaborazioni, si nota che solo a 32 V si ha uno spettro riconducibile ad una funzione gaussiana. A tensioni superiori, invece, si assiste ad un allargamento e ad una progressiva deformazione dello spettro per la presenza di più valori di picco e di "code" più lunghe ed asimmetriche. Tale andamento ha differenti spiegazioni e necessità di ulteriori indagini.

Per prima cosa è conveniente suddividere i dati ricavati dagli spettri per tensioni maggiori di 32 V in *bande di carica*, a seconda se il valore di carica misurato nella singola finestra d'integrazione (della durata di 20 ns) sia, presumibilmente, riconducibile alla risposta di un singolo pixel (si parlerà, in questa trattazione, di *banda singola di carica*), o alla risposta di più pixel (si parlerà, in questo caso, di *banda multipla di carica*).

In base a questo, si ottiene, valutando approssimativamente le varie ampiezze di banda:

Tensione di polarizzazione	Ampiezza della banda singola di carica [pC]	Ampiezza della banda multipla di carica [pC]
34 V	2 ÷ 4,25	4,5 ÷ 6,75
36 V	3,25 ÷ 6,25	6,5 ÷ 10
38 V	3,5 ÷ 7,5	7,5 ÷ 13
40 V	4 ÷ 9	9,25 ÷ 18,5

Tabella 3.2
Tabella riassuntiva delle ampiezze di banda ricavate dagli spettri di carica

Occorre, tuttavia, osservare come tale suddivisione sia approssimativa ed arbitraria, soprattutto nella scelta dei valori di confine fra una banda e l'altra (anche se è stato fatto riferimento alle misure di carica sui singoli pixel), ma che si dimostra molto utile per poter capire e spiegare meglio l'andamento di tali spettri.

Nella seguente tabella sono, invece, riportati il numero di eventi rilevati in ciascuna banda di carica alle varie tensioni di polarizzazione, per le differenti durate di acquisizione.

Tensione di polarizzazione	Eventi nella banda singola di carica	Eventi nella banda multipla di carica	Durata dell'acquisizione
34 V	$3{,}6 \cdot 10^5$	233	4 minuti
36 V	$4{,}35 \cdot 10^5$	707	5 minuti
38 V	$2{,}71 \cdot 10^5$	562	3 minuti
40 V	$1{,}75 \cdot 10^5$	1391	2 minuti

Tabella 3.3
Tabella riassuntiva degli eventi rilevati nelle rispettive bande di carica

Per quanto riguarda il comportamento dello spettro nella banda singola di carica, valgono tutte le considerazioni fatte per l'analisi della carica e del guadagno del singolo pixel (fluttuazioni statistiche del processo di valanga, inefficacia del quenching, afterpulsing, ecc). Tuttavia, tali ipotesi, seppur molto plausibili e fondate, non bastano da sole a spiegare la presenza di più picchi e, soprattutto, l'eccessivo allargamento dello spettro di carica. Esistono, dunque, ulteriori spiegazioni per questi fenomeni.

Una di queste spiegazioni, molto probabile in una configurazione SiPM, è la possibile somma, in uscita, fra la carica emessa dall'impulso di valanga di un singolo pixel e la carica (notevolmente inferiore in quantità) delle "code" di ricarica (dovute ad un pessimo quenching, soprattutto al crescere della tensione, come già visto) degli impulsi di altri pixel. Il risultato di tale somma è, ovviamente, un valore di carica più grande e, di conseguenza, un allargamento dello spettro verso destra.

Un'altra spiegazione è legata alla presenza di *disuniformità* di guadagno fra un pixel e l'altro. A sostegno di questa ipotesi, sono state eseguite alcune misure di capacità e di breakdown su ogni singolo pixel della matrice in esame. Infatti, analizzando la relazione sulla carica media emessa in una valanga da uno SPAD **[4]**:

$$Q_{media} = C_{pixel} \cdot (V_{pol} - V_B) \tag{3.6}$$

in cui Q_{media} è la carica media, C_{pixel} la capacità intrinseca del pixel, V_{pol} la tensione di polarizzazione e V_B la tensione di breakdown, si osserva che delle differenze, anche piccole, di capacità (C_{pixel}) e di tensioni di breakdown (V_B) fra un pixel e l'altro possono introdurre delle differenze nell'emissione di carica apprezzabili negli spettri con un allargamento dello spettro e con la presenza di eventuali picchi negli eventi rilevati ad un determinato valore di carica a seconda se viene integrata in uscita la

corrente emessa da un pixel oppure di un altro. La possibile presenza di picchi nel conteggio degli eventi a determinati valori di carica può essere, inoltre, legata anche a differenze nel dark count dei singoli pixel.

Si è deciso, quindi, partendo dalla relazione 3.6, di eseguire delle misure di capacità e breakdown su tutti i pixel al fine di appurare l'eventuale presenza di disuniformità di guadagno fra un pixel e l'altro e confermare quanto detto al riguardo.

Le misure di capacità sono state eseguite a vuoto su chip sia ad una tensione di polarizzazione nulla (0 V), sia a 30 V (sotto la tensione nominale di breakdown) con un capacimetro (modello Boonton 7200) e tengono conto della capacità intrinseca del pixel e delle capacità parassite introdotte dalle piste di metal su silicio, dai fili usati per il bonding e dai piedini per le connessioni esterne del chip.

I risultati di queste misure di capacità sono qui di seguito riportati.

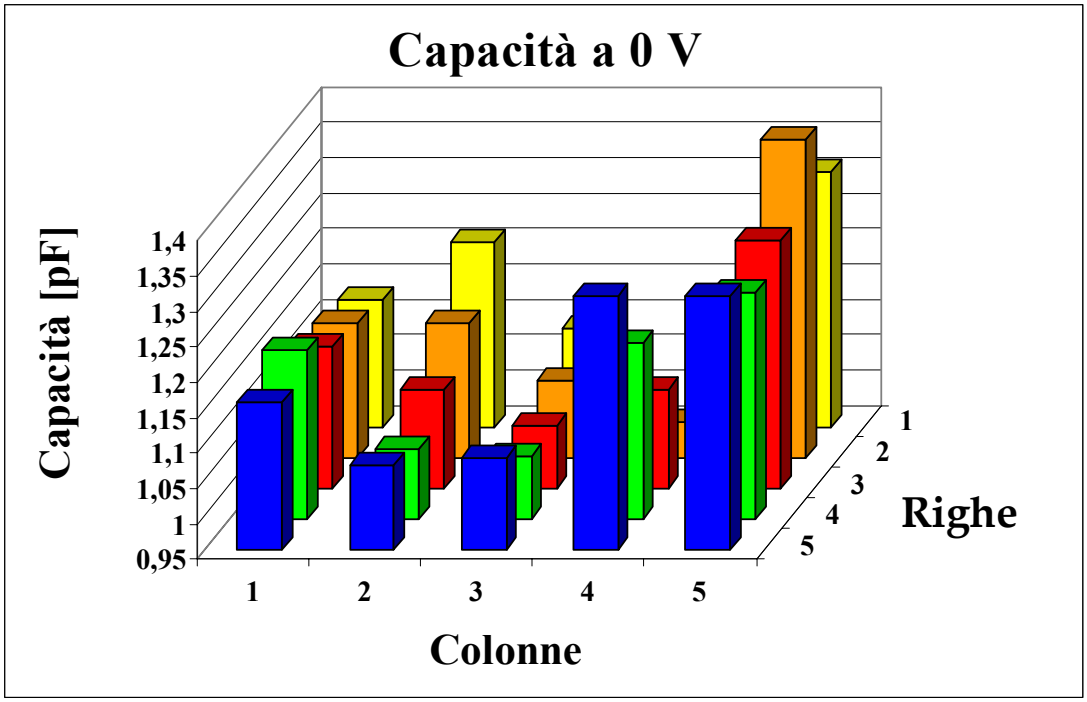

Figura 3.28
Istogramma con i valori di capacità misurati per ogni singolo pixel della matrice a 0 V

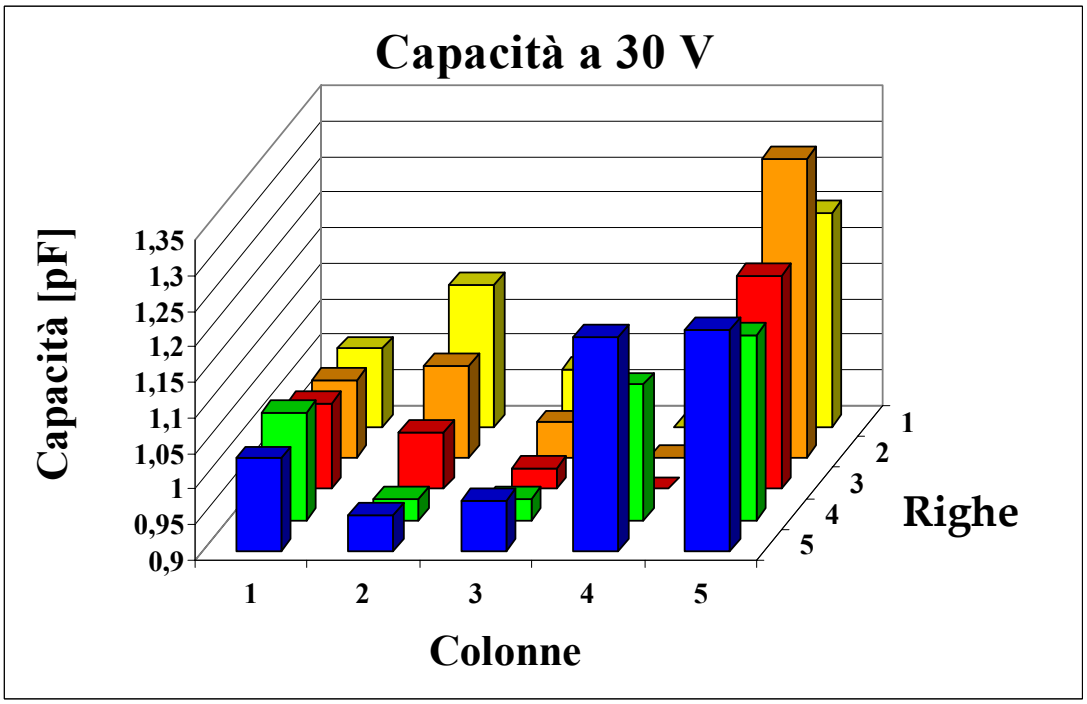

Figura 3.29
Istogramma con i valori di capacità misurati per ogni singolo pixel della matrice a 30 V

Da queste misure, sia a 0 V, sia a 30 V, si ricava che esiste una reale differenza di capacità fra un pixel e l'altro nell'ordine di qualche decimo di picofarad, imputabile, per lo più, alla differente lunghezza delle piste di metal su silicio (come si osserva nel layout in figura 2.27) e dei fili usati per il bonding, anzi si è osservata una certa proporzionalità fra la lunghezza di queste piste di metal ed i valori di capacità misurati. Le differenze, invece, fra le misure a 0 V ed a 30 V su uno stesso pixel (a 30 V si sono trovati valori più piccoli, di qualche decimo di picofarad, che a 0 V) sono imputabili al capacità di transizione del diodo, approssimabile con quella di un condensatore a facce piane e parallele:

$$C_{pixel} = \frac{\varepsilon \cdot A}{W} \tag{3.7}$$

dove C_{pixel} è la capacità di transizione del pixel, ε la costante dielettrica del semiconduttore, A l'area delle giunzione e W l'ampiezza della regione di svuotamento. Poiché W, essendo lo SPAD approssimabile ad una giunzione brusca, dipende dalla radice quadrata della tensione di polarizzazione ed aumenta al crescere della tensione inversa applicata, passando da 0 V a 30 V la capacità C_{pixel} si riduce di pochissimo **[5]**. Inoltre, ad acuire, di molto, le differenze di capacità fra un pixel e l'altro contribuisce anche la basetta di polarizzazione utilizzata nelle misure in

configurazione SiPM per le interconnessioni. Infatti, essendo le piste di metal della basetta di differente lunghezza per i vari pixel, sarà introdotta una capacità parassita C_S, di valore variabile da pixel a pixel, in parallelo verso massa alla capacità C_{pixel} (come si vede in figura 1.14). A causa di ciò occorre correggere la relazione sulla carica media emessa dal pixel come segue:

$$Q_{Media} = \left(C_{Pixel} + C_S\right) \cdot \left(V_{pol} - V_B\right) \tag{3.8}$$

in cui Q_{media} è la carica media, C_{pixel} la capacità intrinseca del pixel, C_S la capacità parassita introdotta dalla basetta, V_{pol} la tensione di polarizzazione e V_B la tensione di breakdown.

Anche le misure della tensione breakdown, ottenute ricavando la caratteristica tensione-corrente per ciascun pixel, hanno rivelato delle piccolissime differenze fra un pixel e l'altro nelle tensioni di breakdown nell'ordine di qualche centesimo (al massimo, in alcuni casi, un decimo) di Volt.

La presenza di disuniformità è, inoltre, comprovata da delle misure di carica e guadagno effettuate su alcuni singoli pixel campione della matrice, in cui sono notate delle differenze nelle cariche emesse dai pixel presi in considerazione. La seguente figura rappresenta le differenze riscontrate nelle misure su singolo pixel.

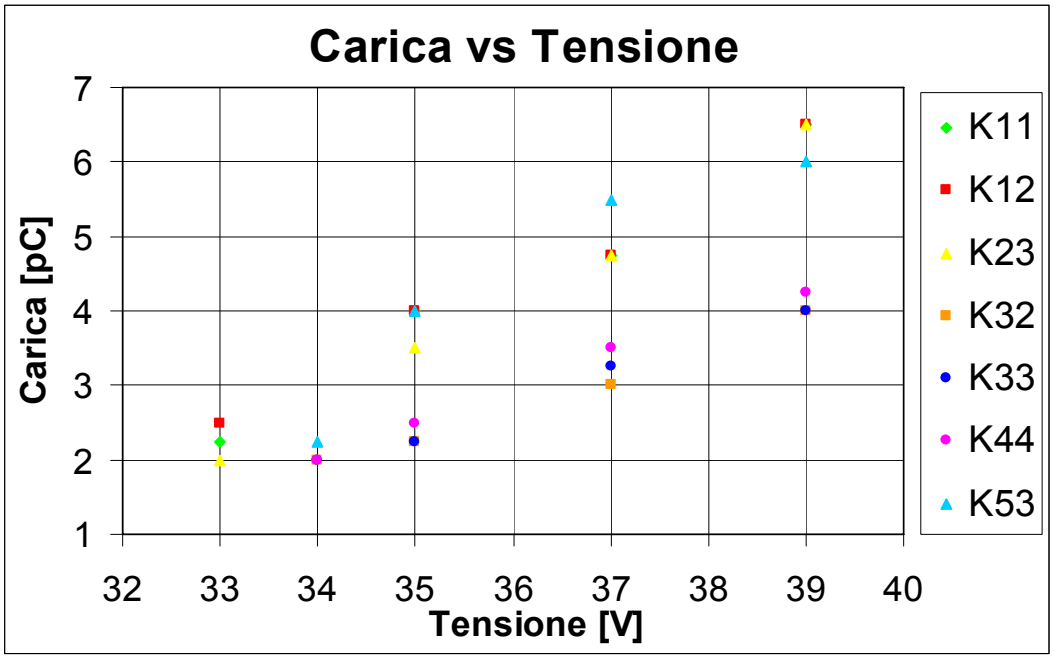

Figura 3.30
Rappresentazione delle misure di carica su alcuni singoli pixel campione. Da notare i differenti valori rilevati.

La presenza di disuniformità fra un pixel e l'altro è, infine, verificata anche dalle misure di dark count eseguite su tutti i pixel di questa matrice, come si può osservare dal seguente grafico:

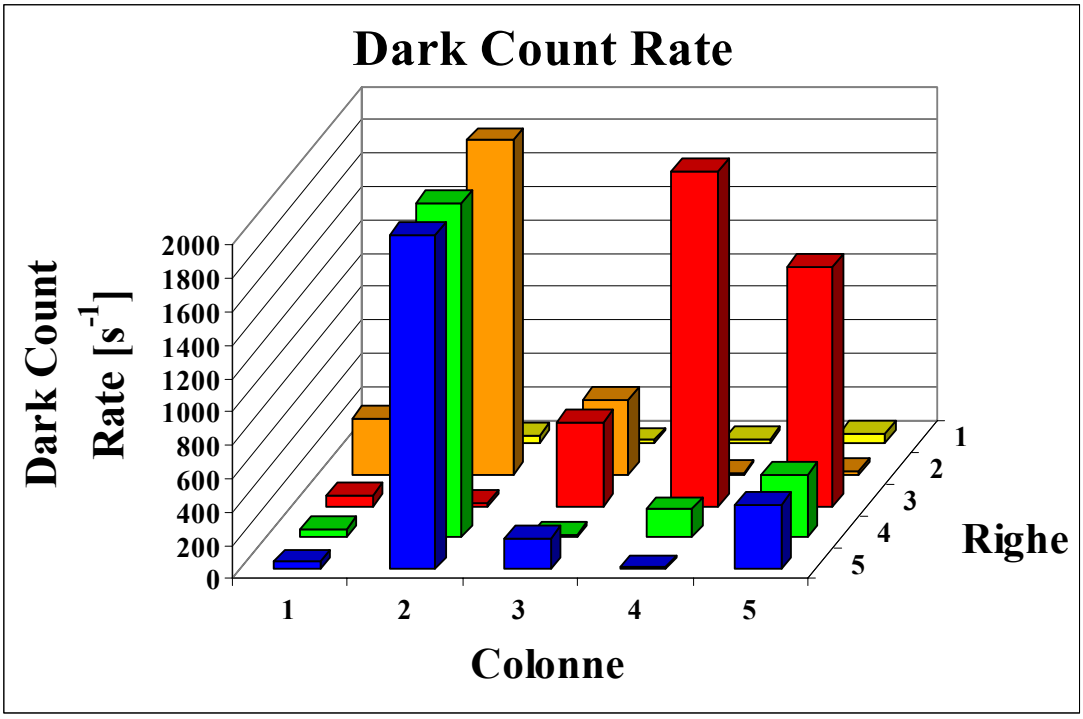

Figura 3.31
Mappatura del Dark Count Rate di tutti pixel della matrice.
Da notare le grandi differenze nei dark count dei vari pixel.

Tale misura è stata eseguita direttamente su fetta, mantenendo ciascun dispositivo "al buio" e polarizzandolo, tramite un impulsatore (a), per 50 ms ad una sovratensione di 5 V rispetto al valore di tensione di rottura del fotorivelatore. Viene acquisito con un oscilloscopio digitale (b), il ritardo, rispetto all'istante in cui è fornita la sovratensione, con cui una coppia elettrone-lacuna prodotta per effetto della generazione termica nella regione di svuotamento innesca una valanga. Al termine dei 50 ms la tensione ai capi del pixel è portata 5 Volt al di sotto della tensione di breakdown ed il rivelatore rimane in tali condizioni (dette condizioni di off) per altri 450 ms (duty cycle del 10%). Questo tempo così lungo è opportuno per garantire che tutti i centri intrappolanti, che potrebbero dare effetti d'afterpulsing, siano svuotati. Tale acquisizione è ripetuta iterativamente per 350 cicli consecutivi per ciascun pixel ed il dark count rate è dato dal reciproco della media dei ritardi d'innesco della valanga. Lo schema circuitale è un po' diverso da quelli visti finora ed è rappresentato nella figura seguente:

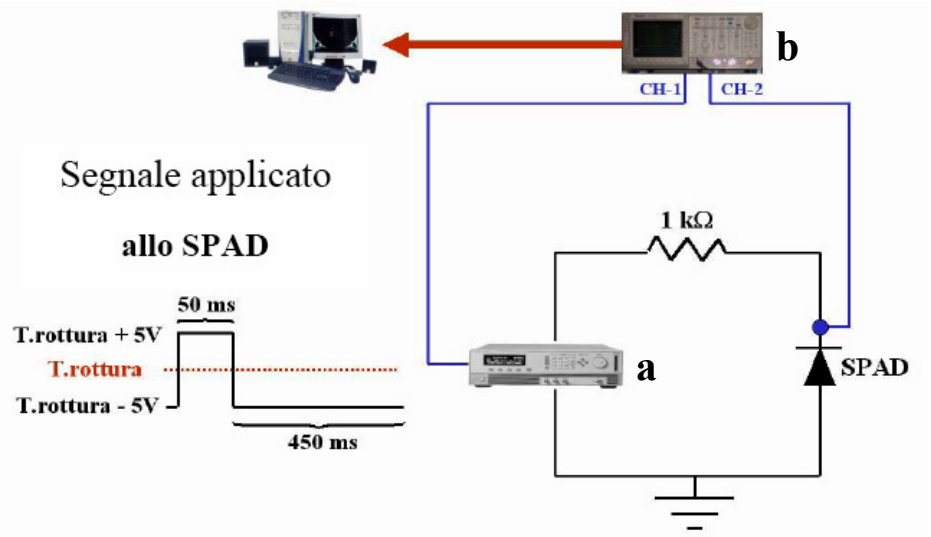

Figura 3.32

Setup utilizzato: l'impulsatore (a) manda il segnale alla serie formata dallo SPAD e dalla resistenza di 1 kΩ. L'oscilloscopio (b) misura il tempo impiegato dallo SPAD per andare in valanga e il PC ha il compito di raccogliere tutti i tempi di ritardo.

La resistenza in serie allo SPAD è stata scelta piccola (1 kΩ) per avere, mediante una corrente di valanga di qualche milliampère (limitata dalla resistenza interna del dispositivo), un abbassamento di tensione di pochi Volt sul dispositivo. Il trigger dell'oscilloscopio è settato sul segnale dell'impulsatore, in modo tale che la misura abbia inizio dall'istante in cui lo SPAD sia polarizzato sopra il breakdown; a questo punto l'oscilloscopio misura l'intervallo di tempo che scorre da quando viene fornito il segnale di overvoltage fino a quando s'innesca la valanga.

Per quanto riguarda, invece, gli eventi rilevati nella *banda multipla di carica* degli spettri ricavati in queste misure, occorre imputare la loro esistenza, con molta probabilità, a fenomeni di cross talk. Infatti, la presenza di eventi nella banda multipla di carica è legata all'emissione di corrente di più pixel (dai valori di carica riscontrati si tratta di due pixel) sul carico comune durante la finestra d'integrazione del segnale, fissata a 20 ns. Poiché è statisticamente poco probabile che due pixel emettano, "al buio", due valanghe in un intervallo inferiore a 20 ns senza alcuna correlazione fra tali eventi, i casi rilevati nella banda multipla di carica sono, senza dubbio, imputabili al cross talk. A sostegno di tale ipotesi, inoltre, occorre citare degli studi eseguiti in precedenza ai *Laboratori Nazionali del Sud* di *Catania* su tale fenomeno. Da tali studi è emerso che, nel caso di *de-trapping stimolato* (descritto nel paragrafo 1.5.4), sono necessari mediamente circa 2 ns affinché possa esserci un effetto d'induzione elettromagnetica da un pixel percorso dalla valanga ad un altro. Quindi, la finestra d'integrazione di 20 ns utilizzata in queste misure è ampiamente sensibile al de-trapping stimolato ed è molto probabile che ciò sia la causa degli

eventi rivelati in banda multipla, anche perchè in un oscilloscopio digitale si sono spesso osservati sul carico comune degli impulsi a circa 2 ns di distanza, cosa che escluderebbe sia il cross talk ottico, che richiede circa 1 ps, sia quello elettrico, che richiede circa 1 μs **[6]**. Tuttavia, nulla, in teoria, vieta che in tale banda possano esserci eventi da imputare al cross talk di tipo ottico, poiché tale tipo di cross talk si esaurisce mediamente in tempi dell'ordine di 1 ps, quindi ampiamente rivelabile all'interno della finestra di 20 ns, ed anche perchè la risoluzione dell'oscilloscopio utilizzato non è tale da raggiungere 1 ps. È, invece, da escludere categoricamente che tali eventi siano dovuti a cross talk elettrico, poiché la diffusione (con anche un piccolo contributo di drift) di un elettrone da un pixel all'altro richiede tempi nell'ordine di circa 1 μs, non rivelabile nella finestra d'integrazione di 20 ns.

È possibile, inoltre, calcolare, in maniera statistica, dai dati ricavati dagli spettri di carica la probabilità di cross talk di questa matrice, in configurazione SiPM, al variare della tensione di polarizzazione. Tale calcolo è effettuato contando gli eventi misurati nella banda multipla di carica, mediandoli nell'unità di tempo sulla durata dell'acquisizione e confrontandoli con la media degli eventi di dark count al secondo di tutti i pixel della matrice sul carico comune della configurazione SiPM, come riassunto dalla seguente relazione:

$$P_{cross-talk} = \frac{N_c}{T_{acq} \cdot N_{dark}} \tag{3.9}$$

in cui $P_{cross\text{-}talk}$ è la probabilità di cross talk, N_c il numero di eventi rilevati nella banda multipla di carica, T_{acq} il tempo di acquisizione, espresso in secondi, ed N_{dark} la media di eventi di dark count al secondo di tutti i pixel della matrice sul carico comune. Dalla tabella 3.3 sono noti sia T_{acq} sia N_{dark}, occorre, dunque, trovare N_{dark}. Tale valore è stato ricavato compiendo una media su 100 secondi del numero di eventi di dark count contati al secondo sul carico comune della matrice. Lo schema circuitale utilizzato è quello in figura 1.18, mentre lo schema per il conteggio degli eventi è il seguente:

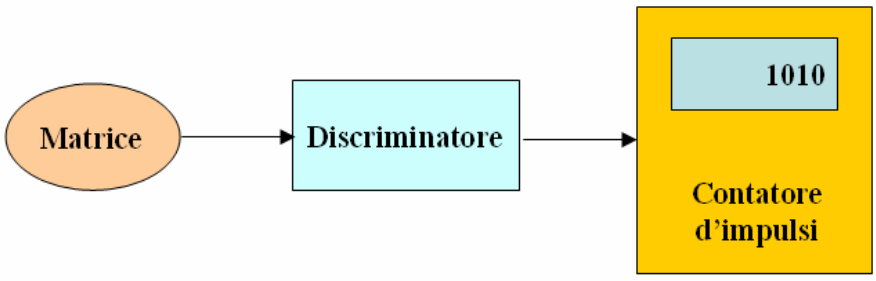

Figura 3.33
Schema utilizzato per il conteggio degli eventi di dark count.

Secondo tale schema, il segnale analogico in uscita sul carico comune della configurazione SiPM è processato dal discriminatore che emette un impulso ogni qual volta il segnale analogico in uscita supera la soglia di discriminazione, fissata a 20 mV (da tenere in considerazione come parametro, poiché vi può essere un insieme di eventi spuri di ampiezza inferiore persi nel conteggio complessivo). Un contatore di impulsi si preoccupa, poi, di contare gli impulsi emessi dal discriminatore in un secondo e di fare la loro media su 100 secondi.

I risultati ricavati da tali misure, per valori di tensione superiori a 32 V, sono:

Tensione applicata	Numero medio di eventi di dark count [s^{-1}]
34 V	$4{,}5 \cdot 10^{4}$
36 V	$6{,}4 \cdot 10^{4}$
38 V	$8{,}4 \cdot 10^{4}$
40 V	$9{,}5 \cdot 10^{4}$

Tabella 3.4
Tabella dei risultati delle misure di dark count della matrice in configurazione SiPM

A questo punto, sono disponibili tutti i dati per poter calcolare la probabilità di cross talk, applicando la relazione 3.8, ed i risultati ottenuti sono i seguenti:

Tensione applicata	Probabilità di cross talk
34 V	$2{,}1 \cdot 10^{-5}$
36 V	$3{,}6 \cdot 10^{-5}$
38 V	$3{,}7 \cdot 10^{-5}$
40 V	$1{,}2 \cdot 10^{-4}$

Tabella 3.5
Tabella della probabilità di cross talk della matrice in configurazione SiPM

Come prevedibile, la probabilità statistica di cross talk cresce con la tensione applicata, perchè con l'overvoltage aumenta la quantità di carica che fluisce nell'evento di valanga ed aumenterebbero, così, sia gli effetti induttivi di tipo elettromagnetico sia la probabilità che un elettrone o una lacuna possano essere liberati dal livello trappola alla banda di conduzione o di valenza dando luogo così ad un evento spurio di valanga (di cross talk) nel pixel in rivelazione.

Fine del Capitolo 3

Bibliografia

[1] A. Campisi - ***"Caratterizzazione di matrici di fotodiodi a valanga per la rivelazione di singoli fotoni"*** - Tesi di laurea - Politecnico di Milano - Dicembre 2005 - http://www.lns.infn.it/previous_theses/Tesi_A_Campisi.pdf

[2] W.R.Leo ***"Techiniques for Nuclear and Particles Physics Experiments"*** - Sprinter-Verlag.

[3] P. Finocchiaro - A. Campisi - D. Corso - L. Cosentino - G. Fallica - S. Lombardo - M. Mazzillo - F. Musumeci - A. Piazza - G. Privitera - S. Privitera - E. Rimini - D. Sanfilippo - E. Sciacca - A. Scordino and S. Tudisco ***"Test of Scintillator Readout With Single Photon Avalanche Photodiodes"*** - IEEE TRANSACTIONS ON NUCLEAR SCIENCE, VOL. 52, NO. 6, DECEMBER 2005, pag 3040-3046.

[4] J. Barral ***"Study of Silicon Photomultipliers"*** - Thesis - École Polytechnique, France - 13th April-2nd July 2004.

[5] J. Millman - A. Grabel ***"Microelettronica"*** - McGraw-Hill (1987).

[6] P. Finocchiaro - A. Campisi - L. Cosentino - A. Pappalardo - F. Musumeci - S. Privitera - A. Scordino - S. Tudisco - G. Fallica - D. Sanfilippo - M. Mazzillo - A. Piazza - J. Van Erps - S. Van Overmeiere - M. Vervaeke - B. Volckaerts - P. Vynck - A. Hermannes - H. Thienponts - S. Lombardo - E. Sciacca ***"SPAD arrays and micro-optics: towards a real single photon spectrometer"*** - in print on JOURNAL OF MODERN OPTICS.

CONCLUSIONI

Lo scopo della parte sperimentale di questo lavoro di tesi è stato caratterizzare, attraverso gli spettri di carica, sia la carica di valanga emessa "al buio" da un singolo dispositivo SPAD da 20 μm di diametro di area attiva al variare della tensione di polarizzazione, sia la carica emessa, sempre "al buio", su un carico comune da un'intera matrice 5x5 di dispositivi SPAD, anch'essi da 20 μm di diametro di area attiva, del lotto Y503012, in configurazione SiPM al variare della tensione di polarizzazione. Il pregio di tali spettri è che consentono, con un'ottima precisione, di poter misurare e trattare statisticamente la carica emessa in diversi impulsi di valanga, rappresentandola con degli istogrammi. Da questo è stato possibile ricavare importanti informazioni sul comportamento e sulle prestazioni dello SPAD e, con qualche approssimazione (visto l'utilizzo di resistori di quenching non integrati), del SiPM.

Per quanto riguarda la carica emessa da un singolo pixel, si sono ricavati degli spettri di carica molto "puliti", con poche fluttuazioni statistiche ed andamenti molto simili a quelli di una delta di Dirac (o almeno di una funzione gaussiana con una deviazione standard σ molto esigua). Solo a tensioni più elevate (oltre i 34 V) si è notato un piccolissimo allargamento dello spettro dovuto sia alla scarsa efficacia del resistore di quenching da 100 kΩ utilizzato (evidenziando, quindi, la necessità di scegliere un valore opportuno del resistore di quenching in base alla tensione di polarizzazione applicata) sia all'aumento della probabilità di afterpulsing. Nel complesso, tuttavia, si è visto che i singoli dispositivi SPAD da 20 μm realizzati dal gruppo *R&D* della *STMicroelectronics* di *Catania* abbiano una risposta in carica pressoché stabile (a parte qualche piccolo effetto parassita), requisito importante per poter essere utilizzati in alcune applicazioni della Fisica Nucleare come la rivelazione di particelle.

Per quanto riguarda la carica emessa da una matrice 5x5 di SPAD in configurazione SiPM, si è ottenuto uno spettro "pulito" solo a 32 V. A tensioni più elevate, invece, sono stati riscontrati alcuni problemi tipici del SiPM vero e proprio quali la disuniformità di guadagno ed il cross talk. Su tali problemi, in particolare la disuniformità, sono state eseguite nuove indagini tentando di risalire alle fonti di tale problema attraverso la formula del guadagno di un singolo pixel e si sono avuti degli ottimi riscontri. Infatti, sono state effettuate delle misure di capacità e breakdown su

tutti i pixel, rilevando delle piccole differenze nelle capacità dell'ordine di decimi di picofarad (dovute essenzialmente alla differente lunghezza delle piste di metal su silicio e dei fili usati per il bonding più che al pixel vero e proprio) e nelle tensioni di breakdown dell'ordine di centesimi, al massimo un decimo, di Volt. Sono state riscontrate, inoltre, anche delle grosse disuniformità nel dark count rate dei singoli pixel. Tuttavia, si è visto come questi fattori, da soli, non fossero sufficienti a poter spiegare le grosse differenze di guadagno riscontrate fra un pixel e l'altro, nelle misure su singolo pixel, e che fosse presente una fonte di disuniformità esterna dovuta alla capacità parassita introdotta dalla basetta di polarizzazione utilizzata, purtroppo, per le interconnessioni e per simulare il comportamento SiPM. La speranza è che in futuro si possa realizzare un dispositivo SiPM vero e proprio, con i resistori di quenching integrati nel silicio, e poter confrontare i risultati ottenuti oggi, simulando il dispositivo, con il vero SiPM. Per quanto riguarda, infine, il cross talk, si è cercato di individuare le cause possibili di tale fenomeno ed è stata calcolata, in maniera statistica la probabilità di cross talk (rivelata nell'ordine di $10^{-4} \div 10^{-5}$). Si è visto come tale problema sia dovuto, per lo più, ad effetti induttivi di tipo elettromagnetico (de-trapping stimolato) e, in teoria, anche ad interferenze di tipo ottico e cresca all'aumentare della tensione di polarizzazione applicata, come, del resto, era previsto, visto che a tensioni più alte aumenta la quantità di carica che fluisce nell'evento di valanga ed aumenterebbero, così, sia gli effetti induttivi di tipo elettromagnetico sia la probabilità che un elettrone o una lacuna possano essere liberati dal livello trappola alla banda di conduzione o di valenza dando luogo così ad un evento spurio di valanga (di cross talk) nel pixel in rivelazione.

RINGRAZIAMENTI

Molte persone mi sono state vicino sia in questi mesi di lavoro per la stesura di questa tesi sia in questi ultimi tre anni e mezzo di studi universitari. Tutte meritano un sentito ringraziamento per aver contribuito, in un modo o nell'altro, al raggiungimento di questa piccola ma importante tappa della mia vita e dei miei studi universitari.

In primo luogo, desidero ringraziare vivamente il Chiar.mo Prof. Gaetano Palumbo per essere stato relatore di questa tesi e, soprattutto, per avermi dato l'importante opportunità di poter svolgere questo lavoro, accettando peraltro una mia richiesta, in un'importante azienda quale la STMicroelectronics, permettendomi d'imparare cose nuove ed utili per il mio futuro, di conoscere un nuovo mondo e d'incontrare persone speciali, preparate e pronte ad aiutarmi in ogni momento.

Ringrazio sentitamente il Dott. Giorgio Fallica ed il Dott. Delfo Sanfilippo, mio tutor e correlatore di questa tesi, per avermi permesso di lavorare nel loro gruppo presso la STMicroelectronics, per essersi sempre preoccupati per il mio lavoro di tesi, facendo fronte ad alcuni problemi logistici che si sono presentati in questi mesi, per i loro preziosi insegnamenti, per le loro lucide spiegazioni e per la loro piena disponibilità e pazienza dimostrata nei miei confronti.

Ringrazio profondamente il Dott. Paolo Finocchiaro, correlatore di questa tesi, per avermi dato l'importante possibilità di poter lavorare ed eseguire le misure di carica ai Laboratori Nazionali del Sud di Catania, per la sua cordiale disponibilità, il suo prezioso aiuto, i suoi utili insegnamenti, i suoi validi consigli e le interessanti discussioni avute sui risultati delle misure effettuate in questi mesi.

Un ringraziamento speciale lo voglio dare a tre persone altrettanto speciali che, come degli angeli custodi, mi hanno quotidianamente guidato in questo lavoro di tesi con la loro leale amicizia, il loro fraterno affetto, il loro proficuo quanto indispensabile aiuto, la loro infinita pazienza, la loro sincera comprensione, i loro utili consigli e preziosi chiarimenti: si tratta del Dott. Massimo Mazzillo, dell'Ing. Giovanni Condorelli e dell'Ing. Angelo Campisi.

Desidero ringraziare tutti i componenti dell'Ic's Technology Development Group del settore R&D della STMicroelectronics di Catania per avermi ospitato in tutti questi mesi e fatto sentire come in una seconda grande famiglia, in particolare l'Ing. Alessandro Piazza, la Dott.ssa Barbara Caltabiano e l'Ing. Salvo Marchese.

Un grazie va anche al Dott. Alfio Pappalardo dell'INFN-LNS di Catania.

Un grosso e sincero ringraziamento voglio riservarlo ai miei preziosi genitori, Alfio e Pina: il loro prezioso sostegno, il loro quotidiano conforto, la loro innata fiducia nei miei mezzi e, soprattutto, il loro immenso amore hanno contribuito moltissimo a farmi raggiungere questo piccolo ma importante obiettivo.

Sono molto grato anche al mio fratellino Giuseppe per essere stato sempre al mio fianco e per farmi sentire spesso ancora un bambino.

Ringrazio sentitamente anche tutti i miei parenti (nonni, zii e cugini) per essermi stati anche loro vicini, premurosi e solidali in tutti questi anni. In particolare, voglio ricordare i miei nonni Gaetano, Giuseppe e Silvestra che non ci sono più e che sarebbero stati felici di vedermi raggiungere questo piccolo traguardo.

Ringrazio, infine, tutti i miei amici e colleghi che hanno condiviso con me tutte le gioie ed i dolori che la vita ci ha riservato finora: citarvi tutti sarebbe un po' lungo, ma uno per uno siete tutti nel mio cuore e nella mia memoria.

GRAZIE A TUTTI VOI!!!