



UNIVERSITÀ DEGLI STUDI DI PISA

FACOLTA' DI INGEGNERIA

Dipartimento di Ingegneria dell'informazione

TESI DI LAUREA IN INGEGNERIA ELETTRONICA

**Tecnologia della fotorivelazione basata su
dispositivi a semiconduttore**

IL CANDIDATO
Enrico MOLINARI

I RELATORI
Dr. Fabrizio CEI
Prof. Alessandro DILIGENTI
Dr.ssa Maria G. BISOGNI

ANNO ACCADEMICO 2009-2010

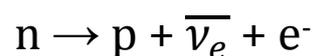
1) Introduzione

Lo scopo di questa tesi è di descrivere gli attuali dispositivi microelettronici a stato solido (a semiconduttore inorganico) di tipo bulk, ovvero dispositivi privi di nanostrutture confinanti per i portatori di carica mobile, come pozzi, fili e punti quantici (“quantum wells, quantum wires, quantum dots”), adatti a rivelare elettricamente un segnale luminoso incidente, caratterizzato da una certa intensità e frequenza. Presenteremo una panoramica sulle tecniche di fotorivelazione più utilizzate, riservando ampio spazio alla fisica che sta alla base dell’optoelettronica ed accennando brevemente alle applicazioni di maggior interesse, come ad esempio le telecomunicazioni su fibra ottica, i rilevamenti telemetrici o vari tipi di “mapping” per uso di diagnostica medica, terminando con i fotomoltiplicatori, di ultima generazione, candidati ad essere utilizzati nella tomografia ad emissione di positroni (in inglese PET – “Positron Emission Tomography”). Questi ultimi sono noti, in ambito tecnico – scientifico, con il nome di SiPMs (“Silicon Photon Multipliers” – fotomoltiplicatori al silicio). Dei dispositivi per fotorivelazione che prenderemo in esame descriveremo i principi di funzionamento, richiamando alcuni concetti fondamentali di fisica dei semiconduttori e alcune loro proprietà ottiche, le principali caratteristiche e le proprietà in base a cui vengono classificati (ad esempio l’efficienza quantica, l’efficienza di rivelazione dei fotoni, il rapporto segnale/rumore, il guadagno, il range dinamico, la responsività, la risoluzione temporale ecc...) e citeremo brevemente alcune tecnologie e passi di processo importanti che ne consentono l’integrazione a livello sub – micrometrico. Analizzeremo i dispositivi optoelettronici, ad omoginzioni ed eteroginzioni, più utilizzati, come i PIN (semiconduttore drogato p/semiconduttore intrinseco/semiconduttore drogato n), gli APDs (“Avalanche Photo Diodes” – fotodiodi a valanga), i SAM – APDs (“Separate Absorption and Multiplication APDs” - fotodiodi a valanga a regioni di assorbimento e moltiplicazione separate), i SAGM – APDs (“Separate Grading Absorption and Multiplication APDs” – SAM a variazione graduale del gap di energia), i GM – APDs (“Geiger Mode APDs” – APDs funzionanti in modalità Geiger) ed i SiPM. Vedremo come il loro impiego consenta di risolvere alcuni dei problemi che interessano i dispositivi a tubo (noti come PMTs – “Photon Multipliers Tubes” - fotomoltiplicatori a tubo), ampiamente usati in passato e tutt’oggi. Infine, una volta chiarite la fisica e la topologia dei suddetti fotorivelatori, illustreremo (in appendice D), a livello di principio, il contesto strumentale nel quale vanno inseriti, al fine di acquisire immagini computerizzate nell’ambito della PET.

2) La tomografia ad emissione di positroni (“Positron Emission Tomography” – PET)

2.1) Il principio fisico alla base della PET

Prima di discutere il funzionamento di un generico fotorivelatore, è opportuno descrivere sinteticamente il principio fisico, sfruttato nell’ambito della PET, per l’acquisizione di immagini computerizzate. In natura un neutrone isolato (la cui vita media τ è stimata intorno ai 15 minuti, con un’accuratezza dell’ordine del secondo) può decadere in un insieme di particelle comprendente un protone, un antineutrino elettronico ed un elettrone (quest’ultimo storicamente chiamato “raggio beta”), secondo la seguente reazione, conosciuta sotto il nome di “decadimento beta”:



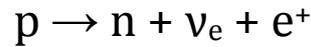
massa del protone = $m_p = 1,672\ 621\ 71(29) \times 10^{-27}$ kg

massa del neutrone = $m_n = 1,674\ 927\ 29(28) \times 10^{-27}$ kg

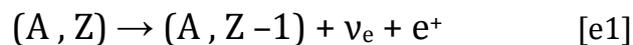
tempo di dimezzamento = $t_{1/2} = \tau \ln 2$

Il processo ora discusso è possibile grazie alla conservazione dell’energia: il neutrone ha massa a riposo maggiore di quella del protone, per cui la quantità di energia pari a $(m_n - m_p)c^2$ (l’energia “residua” della trasformazione del neutrone isolato in un protone) dà origine alle altre due particelle. Dunque, essendo $m_n > m_p$,

il decadimento beta avviene spontaneamente in natura. La reazione seguente si riferisce, invece, ad un decadimento, noto come “decadimento beta inverso”, che non può avvenire spontaneamente in natura, dal momento che la legge di conservazione dell’energia non sarebbe verificata.



Un protone isolato non può trasformarsi in un insieme di particelle comprendente un neutrone, un neutrino elettronico ed un positrone (l’antiparticella dell’elettrone, cioè una particella avente massa pari a quella dell’elettrone e carica uguale, ma di segno opposto), poiché $m_p < m_n$, per cui se non interviene un “agente esterno” che fornisce al protone un’ulteriore energia pari a circa 2MeV, tale reazione non avviene. Se invece il protone, anziché essere libero, risulta legato (ad esempio un protone all’interno di un nucleo), l’energia di legame nucleare (pari a circa 8 MeV) fornisce al protone “l’energia mancante”, e la reazione precedente può aver luogo con tempi caratteristici che dipendono dal nucleo nel quale si trova il protone. Può quindi accadere che un nucleo (A,Z) decada nel modo seguente:



Questa reazione è alla base della PET.

2.2) Il principio su cui si basa la rivelazione fotonica in ambito PET

Sappiamo che ogni sostanza o composto, se iniettato nel corpo umano, si concentra maggiormente su certi tessuti piuttosto che su altri (ad esempio lo Iodio 131 si deposita maggiormente sui tessuti tiroidei). Dunque se iniettiamo nel paziente (spesso si fa per via endovenosa) un certo isotopo legato ad una molecola metabolicamente attiva (in genere uno zucchero), tale molecola si deposita sul tessuto di interesse, veicolandovi l’isotopo ad essa associato. Per la scelta di quest’ultimo si ricorre frequentemente al ^{18}F , poiché ha un tempo di dimezzamento $t_{1/2}$ di 118 minuti, che è sufficientemente corto per poter essere impiegato ai fini di una visita medica, e abbastanza lungo per poter essere trasportato dal ciclotrone di produzione (acceleratore di particelle con il quale si producono gli isotopi di interesse) all’ospedale; tuttavia, solo a titolo di esempio, assumiamo di poter usare il sodio 22 (^{22}Na). Prima che una buona percentuale di campione iniettato si sia depositata sul tessuto di interesse è necessario attendere il lasso di tempo caratteristico della molecola metabolicamente attiva utilizzata per veicolare l’isotopo ^{22}Na . A quel punto l’isotopo decade seguendo la [e1], resa possibile dal fatto che il protone è legato all’interno di un nucleo. Nel caso del sodio 22:



Il tessuto sul quale si è concentrato il sodio, dopo il relativo tempo di dimezzamento, diviene una “sorgente” di positroni (detti anche raggi β^+): tali raggi costituiscono una “traccia” (motivo per cui gli elementi come ^{22}Na , ^{11}C , ^{13}N , ^{15}O , ^{18}F , ^{67}Ga e ^{131}I vengono chiamati “ isotopi traccianti β^+ ”) che, se indirettamente rilevata, ci consente di determinare la posizione che il campione iniettato ha assunto all’interno del corpo (in altri termini di mappare la densità dell’isotopo nel corpo). Un positrone, emesso a partire dal tessuto, viaggia lungo una certa traiettoria (circa due millimetri in acqua), finché non incontra un elettrone atomico: a questo punto si verifica un’annichilazione positrone/elettrone, in cui l’energia di massa $2m_e c^2$ della coppia elettrone/positrone si trasforma in una coppia di fotoni γ_1 e γ_2 , entrambi di energia pari a 511 keV, ovvero appartenenti allo spettro dei raggi gamma. Questi, a partire dal punto di annichilazione, viaggiano lungo la stessa direzione ma in versi opposti.

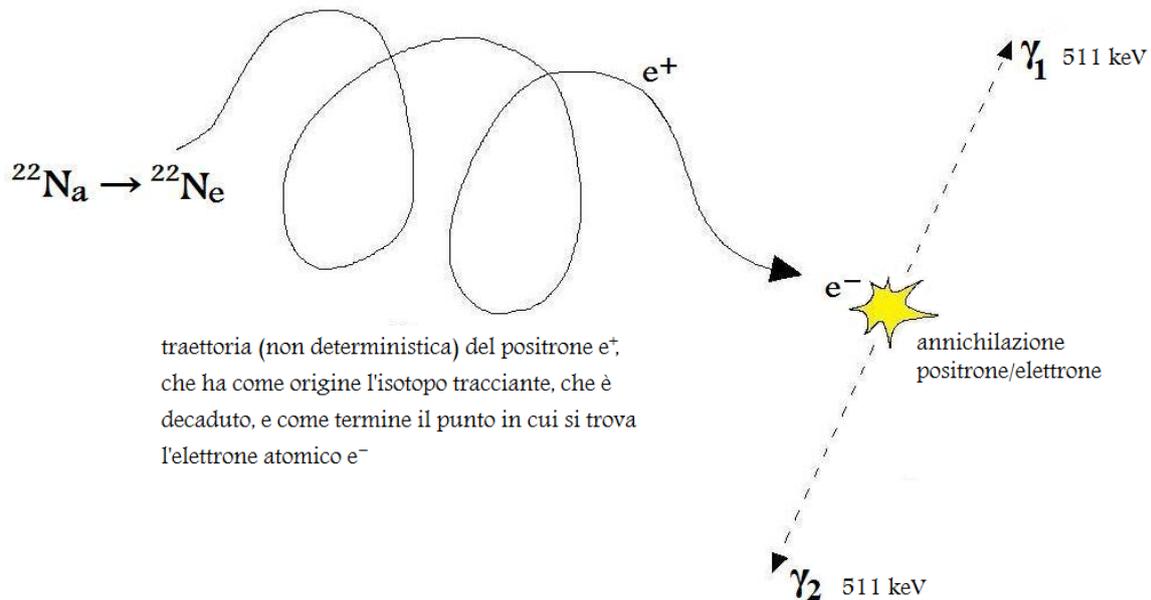


Fig. 1

Rappresentazione del fenomeno fisico alla base della rivelazione fotonica in ambito PET. Il positrone emesso dal decadimento beta inverso attraversa una piccola regione di tessuto e annichila con un elettrone atomico; l'energia di massa elettrone/positrone si trasforma nei due fotoni γ_1 e γ_2 . Ciascuno di questi può essere visto come un'onda elettromagnetica che viaggia lungo una traettoria rettilinea, in base alla prima legge dell'ottica lineare.

La direzione di propagazione dei due fotoni prende il nome di LOR ("Line of Response") ed è la linea di volo di γ_1 e γ_2 in un punto della quale si è verificato l'evento di annichilazione. Il tessuto, divenuto una sorgente di positroni a causa del decadimento beta inverso, grazie all'annichilazione materia/antimateria diviene ora una sorgente di fotoni, la cui frequenza rientra nel range delle onde elettromagnetiche gamma. Non ci resta che predisporre una strumentazione atta alla rivelazione di tali fotoni, per ricostruire un'immagine al computer della geometria del tessuto su cui si sono depositati gli isotopi traccianti. Saranno necessari almeno due fotorivelatori, uno che riveli l'arrivo del fotone γ_1 e l'altro che riveli l'arrivo di γ_2 . Occorrerà un'elettronica di "front-end" (elettronica di lettura) capace di estrarre l'informazione circa la posizione approssimata del punto di annichilazione, lungo la LOR congiungente i due fotorivelatori. Se poi, anziché avere due fotorivelatori, abbiamo (caso più realistico) due pannelli, ciascuno costituito da diversi fotorivelatori, allora è necessario che l'elettronica suddetta prenda in considerazione solo una coppia di fotoni incidenti (uno incidente sul pannello 1, l'altro sul pannello 2) in due istanti separati da pochi nanosecondi, ossia due fotoni quasi simultanei. In tal modo i due fotorivelatori così eccitati, posti ciascuno su un pannello, individueranno una possibile LOR. Se riusciamo a determinare tutte le possibili LOR, con opportuni algoritmi (di retroproiezione) si possono ricostruire i probabili punti di annichilazione e dunque la forma del tessuto da visualizzare. È questo il principio cardine della tecnica diagnostica PET.

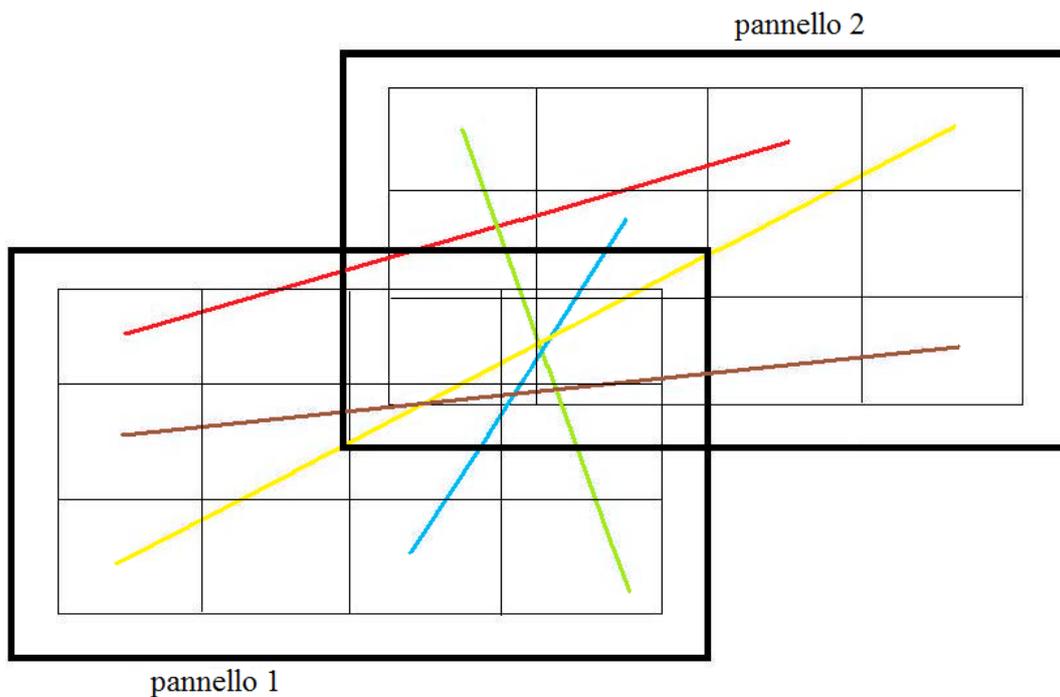


Fig. 2

Schema di principio dello stadio di fotorivelazione di un sistema PET. Nell'esempio ciascun pannello è costituito da 12 SiPMs. L'elettronica di lettura, a valle dei due pannelli, tiene conto solo di due fotoni incidenti quasi simultaneamente. Due SiPMs interessati da due eventi luminosi simultanei individuano una possibile LOR, lungo la quale si trova un punto di annichilazione. Le linee colorate rappresentano le LOR rilevate, dal cui insieme si deduce la forma del tessuto.

Esiste anche un'estensione della PET chiamata TOF PET ("Time of Flight PET"). L'elettronica, a valle del sistema di fotorivelazione, calcola la differenza temporale fra gli istanti di arrivo dei due fotoni γ_1 e γ_2 : se il punto di annichilazione si trova, lungo la LOR, più vicino al SiPM 1 e più lontano dal SiPM 2, assumendo che le traiettorie dei due fotoni siano rettilinee (e quindi trascurando le deviazioni rifrattive dovute alle discontinuità dell'indice di rifrazione n), e che la velocità di propagazione dell'onda elettromagnetica relativa al fotone γ_1 (c/n , con n dipendente dal punto del tessuto nel quale, ad un certo istante, si trova il fotone) sia approssimativamente uguale a quella dell'onda relativa al fotone γ_2 , allora γ_1 arriva prima di γ_2 . Dunque, calcolando la differenza fra i tempi di arrivo, si deduce la posizione approssimata del punto di annichilazione (si legga l'appendice D per ulteriori dettagli sulla TOF PET).

3) Il tubo fotomoltiplicatore (PMT – "Photon Multiplier Tube")

3.1) Concetti generali riguardanti un fotorivelatore

Prima dello sviluppo e del perfezionamento delle tecnologie che consentono, oggi, un'integrazione "spinta" dei dispositivi a stato solido, i fotomoltiplicatori fatti con tubi a vuoto costituivano le uniche apparecchiature usate per la rivelazione di eventi luminosi (al giorno d'oggi conservano ancora un ruolo centrale in diverse applicazioni, PET inclusa). In generale un fotorivelatore può essere visto come un sistema che riceve in ingresso un segnale luminoso, monocromatico o più in generale policromatico, e che rende disponibile in uscita un segnale elettrico la cui ampiezza e struttura temporale sono legate alle caratteristiche del segnale ottico incidente. Il segnale luminoso incidente sul fotorivelatore possiede uno spettro di intensità luminosa $I_v = \mu_v c$, dove μ_v è la densità spettrale volumetrica di energia, misurata in J/m^3Hz , per cui I_v , che si misura in W/m^2Hz , è proporzionale al numero di fotoni di cui consta il segnale luminoso, o meglio al numero di fotoni equienergetici (energia = $h\nu$) che attraversano una superficie unitaria nell'unità di tempo.

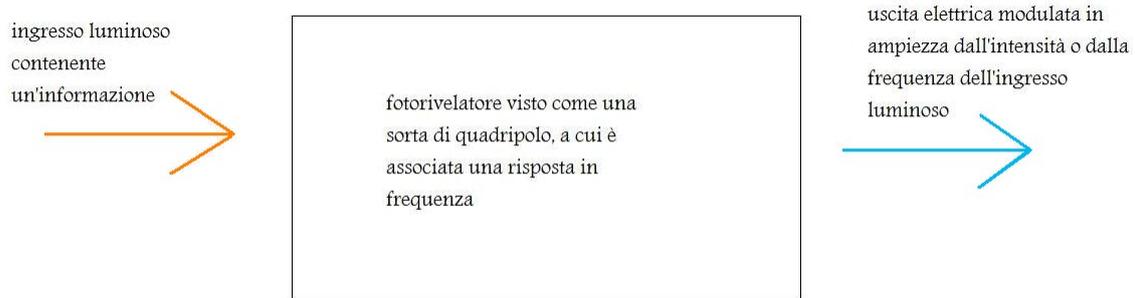


Fig. 3

Modello teorico di un qualunque dispositivo di fotorivelazione

Ad esempio se in ingresso entra un segnale luminoso monocromatico a frequenza ν ed intensità variabile, in uscita avremo un segnale elettrico la cui ampiezza è determinata proprio dal numero di fotoni che incidono sul fotorivelatore: se l'intensità I_ν è alta, la corrente di uscita del sistema sarà alta, se I_ν è bassa la corrente sarà bassa. Dunque la corrente d'uscita è un segnale elettrico la cui ampiezza è proporzionale all'energia luminosa incidente sull'ingresso. Oppure supponiamo che l'evento luminoso monocromatico sia ad intensità fissa e frequenza variabile: allora il modulo del segnale elettrico in uscita sarà proporzionale a tale frequenza.

Il caso ideale di frequenza ν fissata ed intensità I_ν variabile potrebbe essere riprodotto sperimentalmente illuminando un fotorivelatore con un laser DFB ("Distributed Feed Back"); ovviamente l'esperimento si discosta dalla situazione ideale descritta poiché la frequenza ν , durante la modulazione elettrica dell'intensità ottica uscente dal laser ed incidente sul fotorivelatore, varia leggermente a causa del ben noto "effetto Chirp" legato al funzionamento del laser. Inoltre il laser "tipo bulk" (= non basato su nanostrutture, ovvero nessun fenomeno di confinamento quantistico rilevante per i portatori di carica) DFB, così come qualunque altro dispositivo emettitore di luce, possiede uno spettro di emissione di larghezza finita, per cui il segnale incidente sul fotorivelatore non può essere monocromatico, per quanto la selettività (= coerenza spettrale) del laser DFB sia buona. Il caso ideale di intensità fissata e frequenza ν variabile potrebbe essere riprodotto sperimentalmente illuminando un fotorivelatore con un laser DBR ("Distributed Bragg Reflector"), meno selettivo del laser DFB ma accordabile in frequenza; l'esperimento si discosta dalla situazione ideale descritta poiché durante la modulazione elettrica della frequenza ν , uscente dal laser ed incidente sul fotorivelatore, l'intensità ottica I_ν varia, oltre alla solita non monocromaticità del segnale luminoso.

Quello che cambia, fondamentale, fra un tipo di fotorivelatore ed un altro è:

- la modalità con la quale l'energia luminosa interagisce col sistema, ovvero in base a quale principio fisico la luce in questione trasmette al sistema "la sua informazione": in altre parole come l'informazione passa dal "dominio fotonico" a quello "elettronico", cioè il principio fisico attraverso il quale si forma il "fotosegno" elettronico originale (primario)
- la modalità con la quale il fotosegno elettronico primario viene "trattato" al fine di renderne possibile o di migliorarne "l'intelligibilità" da parte di un sistema esterno di front - end

3.2.1) Il tubo fotomoltiplicatore classico (PMT): estrazione elettronica per effetto fotoelettrico

Nel caso dei PMT l'interazione fra la luce incidente ed il sistema di rivelazione è modellizzabile mediante il ben noto effetto fotoelettrico ed il postulato di Planck, entrambi riassunti nella figura 4.

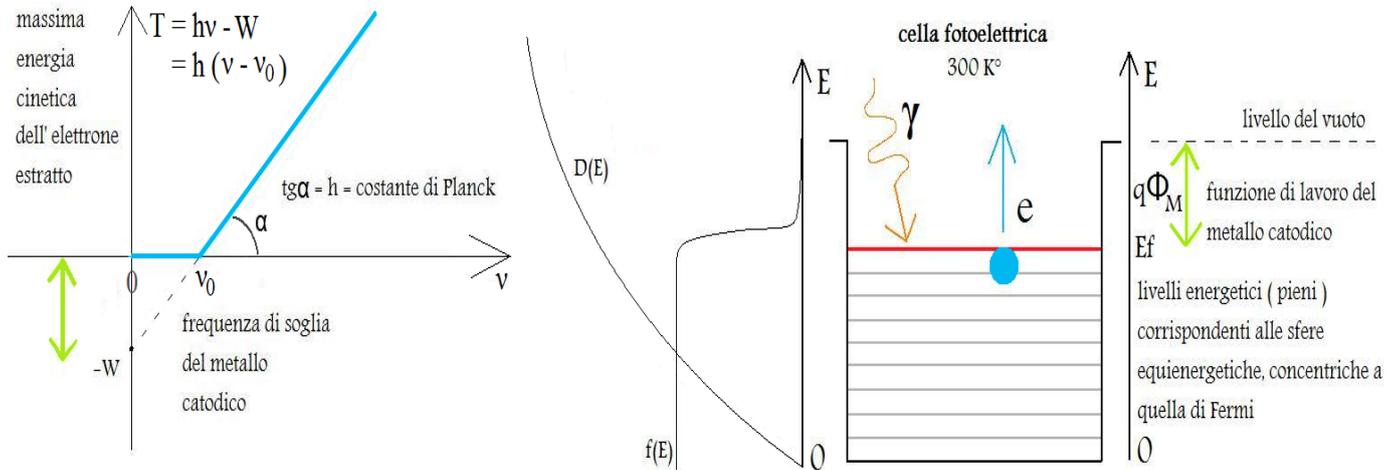


Fig. 4

Rappresentazione schematica dell'effetto fotoelettrico, con il quale è possibile modellizzare l'interazione fra i fotoni e la placca metallica catodica sulla quale questi incidono.

A sinistra è riportata la relazione fra la massima energia cinetica T dell'elettrone estratto da un metallo e la frequenza della radiazione luminosa incidente ν . ν_0 è la "frequenza di soglia" del metallo, sotto la quale la radiazione incidente non estrae alcun elettrone. A destra è riportata la buca di potenziale, prevista dal "modello dell'elettrone libero", con la quale si è soliti rappresentare un metallo dal punto di vista energetico. L'elettrone al suo interno può essere visto, da un punto di vista ondulatorio, come una singola onda piana monocromatica di De Broglie $\Psi_e(r,t)$, la cui espressione completa è:

$$\Psi_e(r,t) = A e^{i \mathbf{k} \cdot \mathbf{r}} e^{-i \frac{E}{\hbar} t}$$

\mathbf{k} è il vettore d'onda, il cui modulo è pari a $n2\pi/\lambda_0$, dove λ_0 è la lunghezza d'onda dell'onda elettromagnetica associata all'elettrone nel vuoto, n è l'indice di rifrazione del metallo in cui si trova l'elettrone, calcolato alla lunghezza d'onda λ_0 , \mathbf{r} è il vettore posizione, che moltiplica scalarmente \mathbf{k} , mentre E/\hbar coincide con la frequenza angolare ω_0 dell'elettrone nel metallo, che per convenzione può essere assunta pari a quella dell'elettrone nel vuoto (a rigore tale uguaglianza è falsa). Il vettore \mathbf{k} individua direzione e verso di propagazione dell'onda di De Broglie $\Psi_e(r,t)$ associata all'elettrone, mentre il suo versore $\hat{\mathbf{k}}$, moltiplicato per l'intensità ottica dell'onda I_ν , fornisce il vettore di Poynting associato a $\Psi_e(r,t)$. A rappresenta la costante di normalizzazione. È bene ricordare che un'onda di De Broglie, formalmente, non è normalizzabile, poiché questa descrive il comportamento spaziale e temporale di una particella massiva (un elettrone) libera, ossia delocalizzata in una regione infinita di spazio nella quale la densità di probabilità di trovare la particella è equidistribuita e di valore prossimo a zero. Nel caso in cui il cristallo metallico, nel quale si trova l'elettrone (quest'ultimo avente massa a riposo m_0 ed energia totale E), ha dimensioni molto maggiori della lunghezza d'onda λ_e dell'elettrone, la cui espressione è:

$$\lambda_e = \lambda_{\text{De Broglie}} = \frac{h}{\sqrt{2m_0 E}}$$

allora l'onda di De Broglie diventa normalizzabile e dalla definizione di densità di probabilità $|\Psi_e(r,t)|^2$ si ricava che:

$$A = \begin{cases} 0, & \text{fuori dal metallo} \\ \frac{1}{\sqrt{V}}, & \text{dentro il metallo} \end{cases}$$

dove V è il volume del cristallo metallico. Quindi l'elettrone è localizzato all'interno del metallo, all'interno della buca di potenziale del metallo, se quest'ultimo non è esposto alla luce. Se il cristallo viene colpito da un fotone di energia almeno pari alla funzione di lavoro $q\Phi_M$ tipica del metallo in questione, un elettrone al suo interno può assorbire l'energia $h\nu$ ($> h\nu_0$) fornita dal fotone ed uscire dalla buca di potenziale. In altre parole la densità di probabilità dell'elettrone può aumentare al di fuori della buca, al di fuori del metallo, azzerandosi, in sostanza, all'interno. Questa situazione corrisponde all'estrazione dell'elettrone per effetto fotoelettrico.

Nel grafico mostrato nella parte destra di figura 4, $f(E)$ è la ben nota distribuzione statistica quantistica di Fermi - Dirac (di cui ricorderemo alcuni elementi nel paragrafo 4.1), mentre $D(E) = 2\pi\left(\frac{2m_0}{h^2}\right)^{\frac{3}{2}}\sqrt{E}$ ($\text{eV}^{-1}\text{cm}^{-3}$) è la densità tridimensionale di stati per gli elettroni del metallo ed è proporzionale alla radice quadrata dell'energia totale degli elettroni (E); m_0 è la massa a riposo dell'elettrone.

3.2.2) Il tubo fotomoltiplicatore classico (PMT): struttura di base

Riferiamoci al PMT riportato in figura 5:

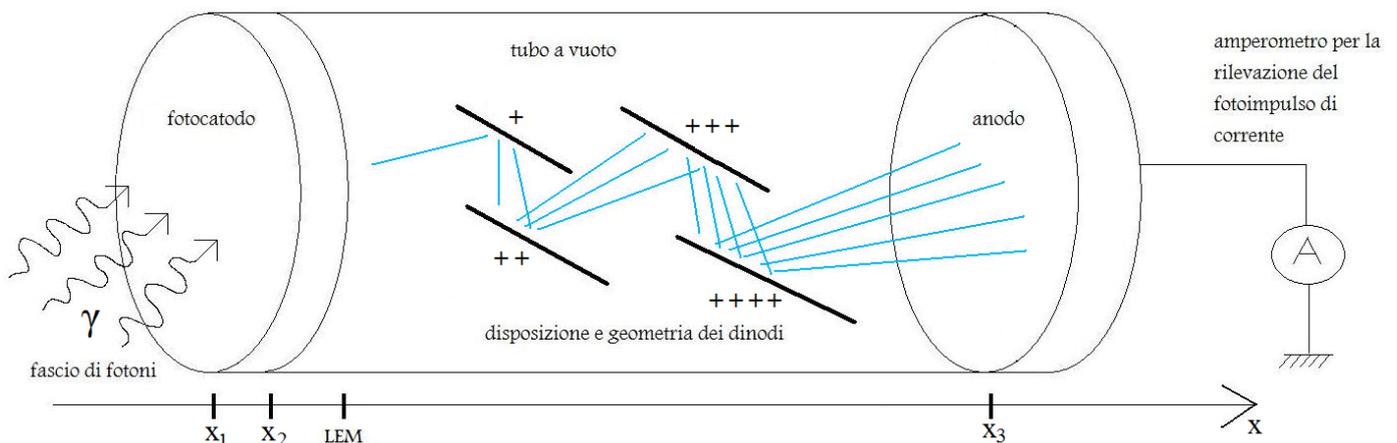


Fig. 5

Disegno schematico di un PMT. La luce irradia la superficie di un elettrodo, chiamato "fotocatodo", atto alla rivelazione di un evento luminoso di opportune frequenze. In virtù dell'effetto fotoelettrico gli elettroni estratti dalla placca, posizionata fra x_1 e x_2 , costituiscono un segnale (primario) il cui flusso viene collimato, presso la sezione indicata con "LEM", da un sistema di lensing elettromagnetico. Il segnale è progressivamente moltiplicato all'interno delle regioni di potenziale create dai dinodi. Infine giunge presso una placca metallica chiamata "anodo", che ha il compito di raccogliere le cariche elettriche.

il tubo fotomoltiplicatore raffigurato riceve la radiazione luminosa sulla superficie di un elettrodo a forma di piastra, chiamato "fotocatodo", attraverso una finestra antiriflettente (multistrato), al fine di minimizzare la frazione riflessa della potenza ottica incidente. I materiali con cui sono costruiti la finestra ed il fotocatodo sottostante dipendono dalla frequenza della luce che siamo interessati a rivelare. Fissata una particolare frequenza ν_γ , il materiale della finestra antiriflettente va scelto in modo da creare la minore discontinuità possibile fra l'indice di rifrazione dell'aria (leggermente maggiore di quello del vuoto, che è 1) e quello del materiale con cui è costruito il catodo, così da minimizzare il coefficiente di riflessione ottica R (in base alla legge di Snell, riportata nel glossario), e in modo che l'energia di gap del materiale della finestra sia sufficientemente maggiore dell'energia dei fotoni incidenti ($E_{\text{gap}} > h\nu_\gamma$), così che i fotoni non siano assorbiti dalla finestra (trasparenza passiva del materiale della finestra rispetto alla frequenza ν_γ). Il materiale con cui costruire il fotocatodo deve essere scelto in modo che l'energia fotonica $h\nu_\gamma$ sia maggiore della funzione di lavoro $q\Phi$, nel caso di un metallo, così da consentire l'estrazione di elettroni per effetto fotoelettrico, o maggiore dell'affinità

elettronica $q\phi$, nel caso di un semiconduttore a temperatura ambiente. Per rivelare luce ad alta energia, ovvero ad alta frequenza, ad esempio blu o ultravioletta, è indicato realizzare il fotocatodo con un metallo, poiché l'energia fotonica in questione è compatibile con la funzione di lavoro dei metalli ($q\Phi_{\text{metallo}} \approx 3 \div 6 \text{ eV}$). La finestra può essere realizzata con silice SiO_2 , dal momento che $h\nu \approx \text{qualche eV} < E_{\text{gapSiO}_2} (\approx 9 \text{ eV})$. Un fotocatodo realizzato in metallo non consentirebbe la fotorivelazione della luce visibile (blu $\approx 400\text{nm} < \lambda < 700\text{nm} \approx$ rosso), dato che questa è costituita da fotoni aventi energie di circa 2eV , insufficienti per l'estrazione di elettroni.

Il PMT è costituito da un tubo di vetro in cui è stato creato un vuoto spinto o molto spinto (fra qualche mbar e frazioni di μbar), al cui interno sono disposti dei dinodi (una catena di differenze di potenziale), ovvero delle placche metalliche la cui geometria e posizioni reciproche sono studiate per consentire, al flusso di elettroni, di fluire attraverso il tubo in modo da ricevere la maggiore amplificazione possibile. Ciascun dinodo è connesso ad un terminale positivo, la cui tensione è maggiore rispetto alla tensione del dinodo precedente. Ai fini della nostra esposizione è più semplice schematizzare l'effetto di trascinamento complessivo degli elettroni (lungo l'asse x , verso l'anodo) da parte dei campi generati dalle tensioni positive dei vari dinodi con l'effetto di trascinamento prodotto da un unico campo elettrico ϵ , generato da una differenza di potenziale anodo - catodo di circa 2000 V (valore tipico), come mostrato in figura 6.

3.2.3) Il tubo fotomoltiplicatore classico (PMT): funzionamento in condizione di buio

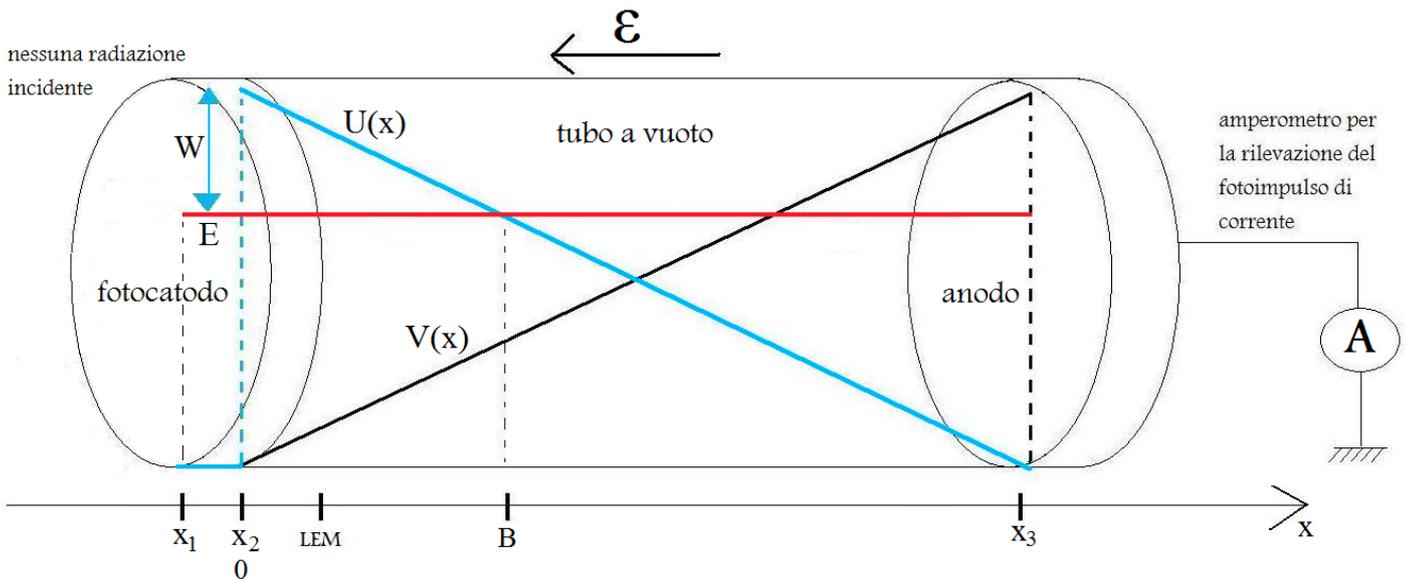


Fig. 6

Schema di un PMT in cui supponiamo, per ragioni di semplificazione analitica, che l'effetto di trascinamento del flusso elettronico sia garantito da un'unica differenza di potenziale fra anodo e fotocatodo (valore tipico di 2000 V). Il fotocatodo non è irradiato dalla luce.

Finché il fotocatodo non è colpito da alcuna luce, oppure finché la luce monocromatica che lo colpisce ha frequenza inferiore a quella di soglia del materiale con cui è costruito, ciascun elettrone del metallo si trova in una buca di potenziale dalla quale è poco probabile che esca. Tale probabilità è tanto inferiore quanto maggiore è il lavoro di estrazione $W (= U(x_2) - E$, dove E è l'energia totale di un singolo elettrone nel catodo, x_2 l'interfaccia catodo - vuoto, $U(x) = U(x_2) - q\epsilon x$ l'energia potenziale, rappresentata in blu nel disegno), tanto inferiore quanto maggiore è la larghezza media della barriera di potenziale B e quanto inferiore è il coefficiente di trasmissione T (effetto tunnel) relativo alla suddetta barriera. Tale coefficiente è calcolabile, approssimativamente, utilizzando il noto metodo WKB (Wentzel - Kramers - Brilluoin), secondo il quale:

$$T \cong e^{-\int_{x_2}^B 2k(x) dx} \quad \text{dove } k(x) = \sqrt{\frac{2m_0[U(x)-E]}{\hbar^2}} \quad [e2]$$

in cui m_0 è la massa a riposo del singolo elettrone del catodo, mentre $\hbar = h/2\pi$ è la costante di Planck ridotta.

In assenza di luce sul catodo un qualunque elettrone, localizzato all'interno dell'elettrodo, può essere descritto, come già osservato, da una funzione d'onda $\psi_e(x,t)$ a quadrato sommabile, la quale consta di una dipendenza spaziale e di una temporale; queste due funzioni fattorizzano, dal momento che gli operatori hamiltoniano e temporale commutano.

$$\psi_e(x,t) = \psi_e(x)e^{-i\frac{E}{\hbar}t}$$

La dipendenza spaziale $\psi_e(x)$ è ricavabile risolvendo l'equazione di Schrödinger, indipendente dal tempo, nel caso monodimensionale (si osservi la [g8]):

$$\frac{d^2\psi_e(x)}{dx^2} + \frac{2m_0}{\hbar^2} [E - U(x)]\psi_e(x) = 0 \quad [e3]$$

quindi:

$$\psi_e(x,t) = e^{-i\frac{E}{\hbar}t} \begin{cases} Ae^{ik_1x} + Be^{-ik_1x}, & \text{per } x_1 < x < x_2 \\ Ce^{k_2x} + De^{-k_2x}, & \text{per } x_2 < x < B \\ Fe^{ik_3x} + Ge^{-ik_3x}, & \text{per } B < x < x_3 \end{cases}$$

dove:

$$k_1 = \sqrt{\frac{2m_0E}{\hbar^2}} \quad k_2 = k_2(x) = \sqrt{\frac{2m_0[U(x)-E]}{\hbar^2}} \quad k_3 = k_3(x) = \sqrt{\frac{2m_0[E-U(x)]}{\hbar^2}}$$

si noti come le interfacce fotocatodo/vuoto e vuoto/anodo comportino la riflessione di un'onda di De Broglie progressiva, dando luogo ad una regressiva. Le costanti complesse A, B, C, D, F e G sono calcolabili imponendo le condizioni al contorno di continuità per $\psi_e(x)$ e per la sua derivata.

$\psi_e(x,t_B)$, ad un generico istante t_B precedente l'evento luminoso ed il conseguente assorbimento fotonico ($t_B =$ istante di buio), ha la rappresentazione grafica mostrata qualitativamente in figura 7.

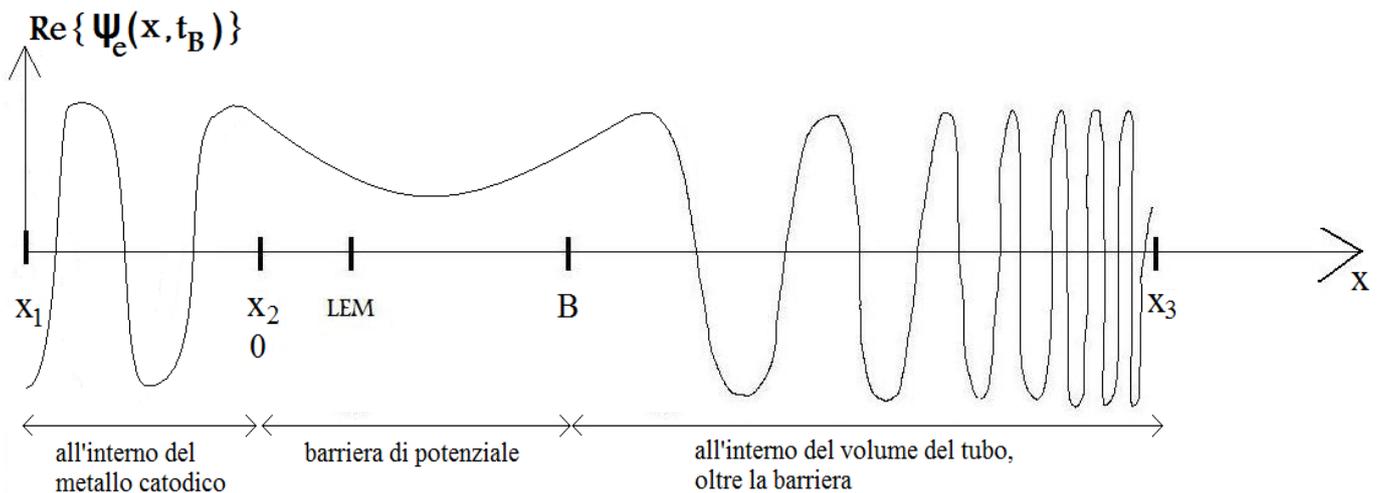


Fig. 7

Rappresentazione qualitativa della dipendenza spaziale della funzione d'onda associata ad un elettrone del metallo catodico, quando questo non è irradiato dalla luce.

A causa dell'effetto tunnel c'è una densità di probabilità non nulla, seppur abbastanza piccola (rappresentata da $|\psi_e(x,t_B)|^2$), che un elettrone, anche in assenza di luce incidente sul fotocatodo, attraversi la barriera compresa fra le coordinate x_2 e B e riesca a raggiungere l'anodo, dando luogo ad una corrente di uscita (un valore tipico, a

temperatura ambiente, potrebbe essere dell'ordine dei fA), rilevabile dall'amperometro posto in serie all'anodo stesso. Questa "dark current" (corrente di buio) è, ovviamente, non gradita, poiché costituisce un errore sistematico che degrada l'accuratezza di una qualunque misura dell'intensità luminosa incidente.

Richiamando il noto concetto di meccanica quantistica della "densità superficiale di corrente di probabilità di particella", possiamo definire la "densità di corrente di probabilità di elettrone" $j_e(x, t_B)$, monodimensionale, relativa ad un elettrone descritto da $\psi_e(x, t_B)$, nel modo seguente (riferiamoci alla figura 6):

$$j_e(x, t_B) = -\frac{i\hbar}{2m_0} \left[\psi_e(x, t_B)^* \frac{\partial \psi_e(x, t_B)}{\partial x} - \psi_e(x, t_B) \frac{\partial \psi_e(x, t_B)^*}{\partial x} \right] = \frac{\hbar}{m_0} \text{Im} \left\{ \psi_e(x, t_B)^* \frac{\partial \psi_e(x, t_B)}{\partial x} \right\} \quad [e4]$$

dove:

$$\psi_e(x, t_B) = \psi_e(x) e^{-i \frac{E}{\hbar} t_B} \quad \psi_e(x) \text{ è stata ricavata dalla [e3]}$$

e considerando l'espressione approssimata del coefficiente di trasmissione T di barriera di potenziale, fornita nella [e2], possiamo mettere in relazione la densità di corrente di probabilità di elettrone all'interno del fotocatodo (metallico) non investito dalla luce $j_e(x_1 < x < x_2, t_B)$, ovvero in ingresso alla barriera, con la densità di corrente di probabilità di elettrone presso la sezione indicata con la coordinata B $j_e(B, t_B)$, ossia in uscita alla barriera, utilizzando la definizione generale del coefficiente T:

$$|j_e(B, t_B)| \cong T |j_e(x_1 < x < x_2, t_B)|$$

$j_e(B, t_B)$ rappresenta una "densità di corrente di probabilità di rumore di buio" primaria. $j_e(B, t_B)$ dipende fortemente dalla temperatura a cui si trova il PMT, in particolar modo il fotocatodo (ad alte temperature l'estrazione calda degli elettroni incrementa il rumore di buio), e più debolmente dalle tensioni dinodiche: è bene osservare, infatti, che maggiore è la tensione risultante fra anodo e catodo, maggiore è la pendenza, lungo l'asse x del PMT, dell'energia potenziale U(x) (proporzionale al campo ϵ) e maggiore è il coefficiente di trasmissione T della barriera, dato che la larghezza media di quest'ultima è diminuita. Chiaramente $|j_e(B, t_B)|$ subirà un'amplificazione dovuta ai dinodi, prima che il flusso di elettroni primari e secondari raggiunga l'anodo, peggiorando ulteriormente il problema del rumore di buio.

3.2.4) Il tubo fotomoltiplicatore classico (PMT): funzionamento in condizione di illuminazione

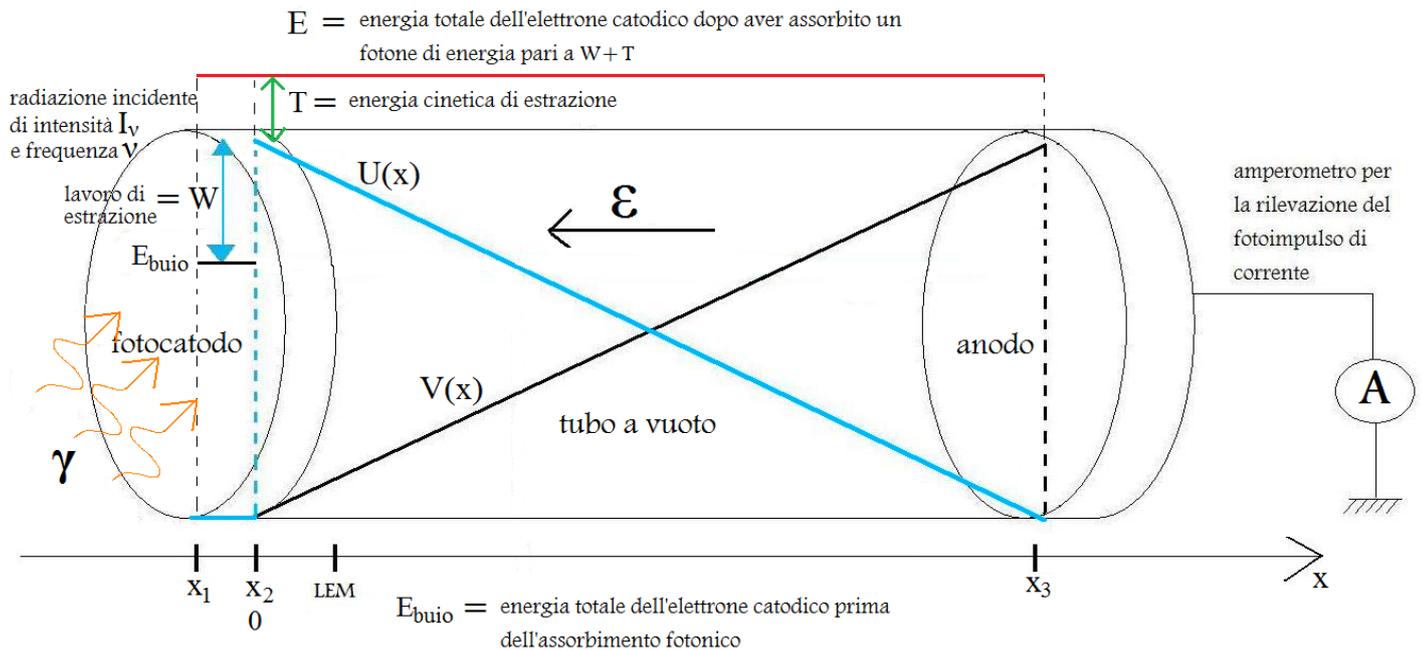


Fig. 8

Schema di un PMT in cui supponiamo, per ragioni di semplificazione analitica, che l'effetto di trascinamento del flusso elettronico sia garantito da un'unica differenza di potenziale fra anodo e fotocatodo (valore tipico di 2000 V). Il fotocatodo è irradiato da una luce di intensità I_v e frequenza ν . La barriera di potenziale mostrata in figura 6, compresa fra le coordinate x_2 e B, è scomparsa in seguito all'assorbimento fotonico da parte di un elettrone appartenente al fotocatodo metallico.

Adesso supponiamo che incida, sul catodo, una radiazione luminosa monocromatica avente una certa frequenza $\nu > \nu_0$ ed una certa intensità I_v : supponendo che "l'efficienza quantica" QE del PMT (si veda il glossario) sia pari ad 1, ovvero che per ogni fotone assorbito da parte del fotocatodo venga emesso un elettrone avente energia cinetica $T = h(\nu - \nu_0)$, la radiazione luminosa fornirà ad un certo numero di elettroni, il quale è proporzionale a I_v , l'energia $W = h\nu_0$ necessaria all'estrazione. Gli elettroni "primari", così eccitati, fuoriescono dal catodo. È bene considerare che la QE reale tipica di un buon PMT commerciale si aggira intorno al 25%, ragion per cui se arrivano sul fotocatodo 100 fotoni, ciascuno avente energia $h\nu$ maggiore di W , dal metallo non vedremo uscire 100 elettroni, ciascuno con energia cinetica $T = h(\nu - \nu_0)$, bensì mediamente 25. Il flusso di elettroni uscente dal metallo viene collimato da lenti elettromagnetiche, in modo da incidere con precisione sul primo dinodo; tale collisione provoca l'emissione di elettroni secondari da parte del metallo di cui è costituito il dinodo, oltre che di raggi X. Gli elettroni secondari vengono accelerati verso il secondo dinodo, posto a potenziale positivo rispetto al primo; pertanto gli elettroni secondari, e quello che resta dei primari, incidono sul secondo dinodo, il quale emetterà un numero ancora maggiore di elettroni secondari, e così via. Il fenomeno di "moltiplicazione" degli elettroni si ripete in cascata, finché il flusso di elettroni raggiunge l'anodo, formando una corrente J rilevabile dall'amperometro. Il guadagno del PMT, ovvero il numero di elettroni che arrivano sull'anodo per ciascun elettrone primario, è determinato dall'intensità e dalla geometria del campo elettrico all'interno del tubo ed è, in genere, compreso fra 10^6 e 10^8 . Aumentando l'intensità I_v della radiazione luminosa, oppure aumentando la frequenza ν , aumenta la corrente misurata J , poiché si verifica un incremento del numero di elettroni secondari prodotti dalla catena di dinodi. La differenza fra gli effetti prodotti dall'aumento di I_v a parità di ν e dall'aumento di ν a parità di I_v è riscontrabile a livello di elettroni primari (non secondari): infatti nel primo caso il numero di elettroni primari estratti dal fotocatodo aumenta, mentre la velocità che questi hanno appena dopo l'estrazione rimane costante, invece nel secondo caso è la velocità appena dopo l'estrazione ad aumentare, mentre il numero di elettroni fuoriuscenti dall'elettrodo rimane costante.

È possibile modellizzare l'estrazione "fredda" di elettroni dal fotocatodo di un PMT, che si verifica quando il materiale (il metallo) con cui l'elettrodo è realizzato viene irradiato dalla luce (trascurando l'effetto termoionico,

il quale diventa importante nel caso di fotocatodo ad alta temperatura, e non a temperatura ambiente), attraverso un formalismo ondulatorio analogo a quello riportato nel paragrafo 3.2.3, utilizzato nella condizione di buio. Consideriamo un elettrone catodico descritto dalla funzione d'onda $\psi_e(x, t_L)$, dove t_L è un istante appena successivo all'assorbimento, da parte dell'elettrone, di un fotone di energia $h\nu_\gamma = W + T$ (lavoro di estrazione del metallo + energia cinetica di fuoriuscita dell'elettrone dal fotocatodo). Abbiamo che:

$$\psi_e(x, t_L) = \psi_e(x) e^{-i \frac{E}{\hbar} t_L}$$

dove $\psi_e(x)$ è la dipendenza spaziale della funzione d'onda $\psi_e(x, t_L)$ che descrive l'elettrone "scatterato" dal fotone incidente sul catodo, E è somma dell'energia dell'elettrone catodico in assenza di radiazione luminosa (E_{buio}) con l'energia fotonica $h\nu_\gamma$ assorbita dall'elettrone. Risolvendo l'equazione di Schrödinger, indipendente dal tempo, sempre nel caso monodimensionale (si osservi la [e3]), ovviamente prendendo come nuovo valore dell'energia totale E dell'elettrone quello relativo alla nuova situazione di illuminazione (si osservi la figura 8), otteniamo un'espressione di $\psi_e(x)$ la cui rappresentazione grafica è quella mostrata, qualitativamente, in figura 9.

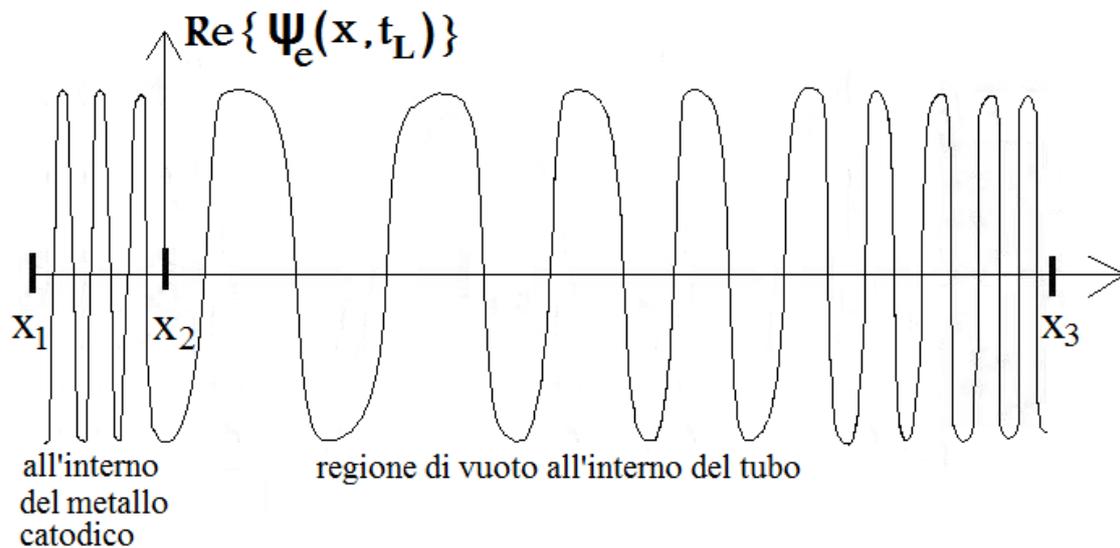


Fig. 9

Rappresentazione qualitativa della dipendenza spaziale della funzione d'onda associata ad un elettrone del metallo catodico, quando questo è irradiato dalla luce.

L'energia E dell'elettrone, successivamente all'assorbimento fotonico, è maggiore di $U(x)$ per qualunque x lungo l'asse del PMT, quindi la barriera di potenziale incontrata nel caso di assenza di luce incidente non esiste più; nella regione compresa fra le coordinate x_2 e B (figure 6 e 7), precedentemente interessata dalla barriera di potenziale, la sovrapposizione di esponenziali reali ha lasciato il posto ad una sovrapposizione di esponenziali complessi, cioè ad un'oscillazione, come ben visibile in figura 9. In quella regione la densità di probabilità di trovare l'elettrone primario è aumentata, rispetto al caso di buio, e ciò può essere dimostrato analiticamente calcolando il coefficiente di trasmissione $T_L(x_2)$, cioè il valore del coefficiente di trasmissione dell'elettrone primario (del "fotoelettrone") all'interfaccia fotocatodo illuminato/regione di vuoto. Si ha che:

$$T_L(x_2) = \left| \frac{j_e(x_2^+, t_L)}{j_e(x_2^-, t_L)} \right| = \left| \frac{j_{e\text{TRASMESSA}}}{j_{e\text{INCIDENTE}}} \right|$$

dove numeratore e denominatore si calcolano invocando la definizione riportata nella [e4]. Si osserverà che $T_L(x_2) < 1$, nonostante che $E > U(x) \forall x$, per cui l'estrazione dell'elettrone catodico primario non è sicura neanche in seguito all'assorbimento del fotone, e che $T_L(x_2)$ è maggiore del coefficiente di trasmissione T , riportato nella [e2], relativo alla barriera di potenziale nel caso di buio. Dunque è più probabile, in seguito all'assorbimento fotonico, che l'elettrone venga estratto dal fotocatodo (ossia che attraversi l'interfaccia posta presso la sezione $x=x_2$), rispetto alla situazione di buio.

3.3) Il fotodiode ibrido (HPD – “Hybrid Photodiode”)

Una valida alternativa al PMT classico è il dispositivo seguente, chiamato “fotodiode ibrido” (HPD – “Hybrid Photodiode”).

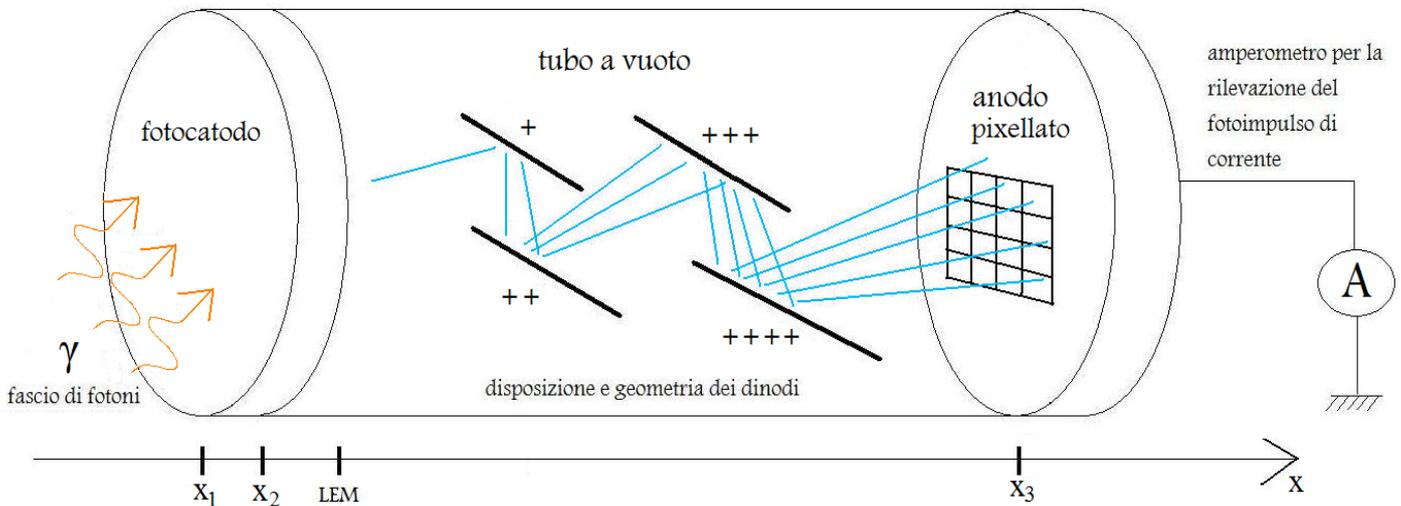


Fig. 10

Schema di un HPD, un dispositivo strutturalmente molto simile ad un PMT classico, ad eccezione dell’anodo, costituito da una matrice di fotodiode a semiconduttore, in parallelo fra loro, fortemente polarizzati in inversa. I fotodiode sfruttano la moltiplicazione, dovuta al breakdown a valanga, delle fotocariche provenienti dal volume del tubo. Il vantaggio in termini di guadagno, che è superiore a quello tipico di un PMT classico, grazie al contributo moltiplicativo fornito dai fotopixel, è accompagnato dall’introduzione di sorgenti di rumore tipiche della componentistica attiva a semiconduttore.

Il funzionamento è del tutto analogo a quello del PMT normale, con l’eccezione del sistema di raccolta delle cariche in prossimità dell’anodo: gli elettroni, arrivati nei pressi di x_3 , vengono raccolti da una matrice quadrata di microcelle (fotodiode a semiconduttore, di cui parleremo in seguito) montate in parallelo, polarizzate inversamente. All’interno di ciascuna delle microcelle si verifica un fenomeno di “breakdown a valanga” che porta alla generazione di un elevatissimo numero di cariche. Queste, in ciascuna microcella, formano una corrente j dovuta al fatto che la microcella in questione ha ricevuto uno o più elettroni dal tubo. La somma di tutte le correnti j dà origine alla corrente J , legata all’entità della luce incidente sul catodo. Nel HPD equipaggiato con dinodi il guadagno, che è maggiore di quello del PMT classico, è affidato sia ai campi dinodici all’interno del tubo, sia ai fenomeni di moltiplicazione interni alle microcelle dell’anodo pixellato. La QE per un buon HPD commerciale è sempre, circa, del 25%. Il voltaggio anodo – catodo tipico è di 15kV.

A fronte di un incremento del guadagno, dovuto al contributo moltiplicativo dell’anodo “speciale”, si ha un’accentuazione del rumore che affligge la misura di corrente J . Oltre alla corrente di buio prima citata, si aggiungono, ora, tutte le fluttuazioni statistiche ineliminabili:

- 1) il “rumore termico” (gaussiano bianco), ovvero la fluttuazione termica statistica degli elettroni che comporta variazioni non deterministiche (non prevedibili) del modulo e della polarità del campo, relativo alla presenza degli elettroni termici “liberi” all’interno del reticolo cristallino;
- 2) il “rumore shot” (anche questo gaussiano bianco), cioè il rumore di attraversamento di barriera di potenziale;
- 3) il “rumore flicker” (detto anche “rumore rosa”), ossia la variazione stocastica nel tempo, in un punto del cristallo, di una qualche popolazione di portatori di carica.

Questi rumori condizionano pesantemente le prestazioni di una microcella fatta a giunzione pn. Inoltre l’anodo pixellato soffre della deriva delle caratteristiche, rispetto alla variazione della temperatura e all’invecchiamento del dispositivo, che è tipica di tutta la componentistica attiva: ciò causa un ulteriore scostamento dalla situazione ideale. In un HPD in cui il guadagno è affidato unicamente ai pixels dell’anodo, cioè in un HPD privo dei dinodi

interni al tubo a vuoto, dove i fotoelettroni vengono accelerati per tutta la lunghezza del tubo senza incontrare ostacoli, il guadagno scende al di sotto di 10^3 .

I problemi legati all'uso dei PMT sono:

- ingombro eccessivo da parte del tubo (dipendente dal contesto di utilizzo e dal tipo di PMT);
- voltaggi operativi elevati;
- modalità e condizioni particolari di lavoro: per esempio un PMT, per evitare che campi magnetici esterni perturbino il normale flusso di elettroni, dovrebbe essere posto in una camera anecoica adeguatamente schermante.

Tali problemi sono risolti con l'utilizzo di fotorivelatori microelettronici a stato solido (a semiconduttore).

4) Concetti basilari del fotodiode ad omogiunzione pn

4.1.1) La omogiunzione pn in condizione di equilibrio

Riferiamoci all'omogiunzione brusca pn mostrata in figura 11:

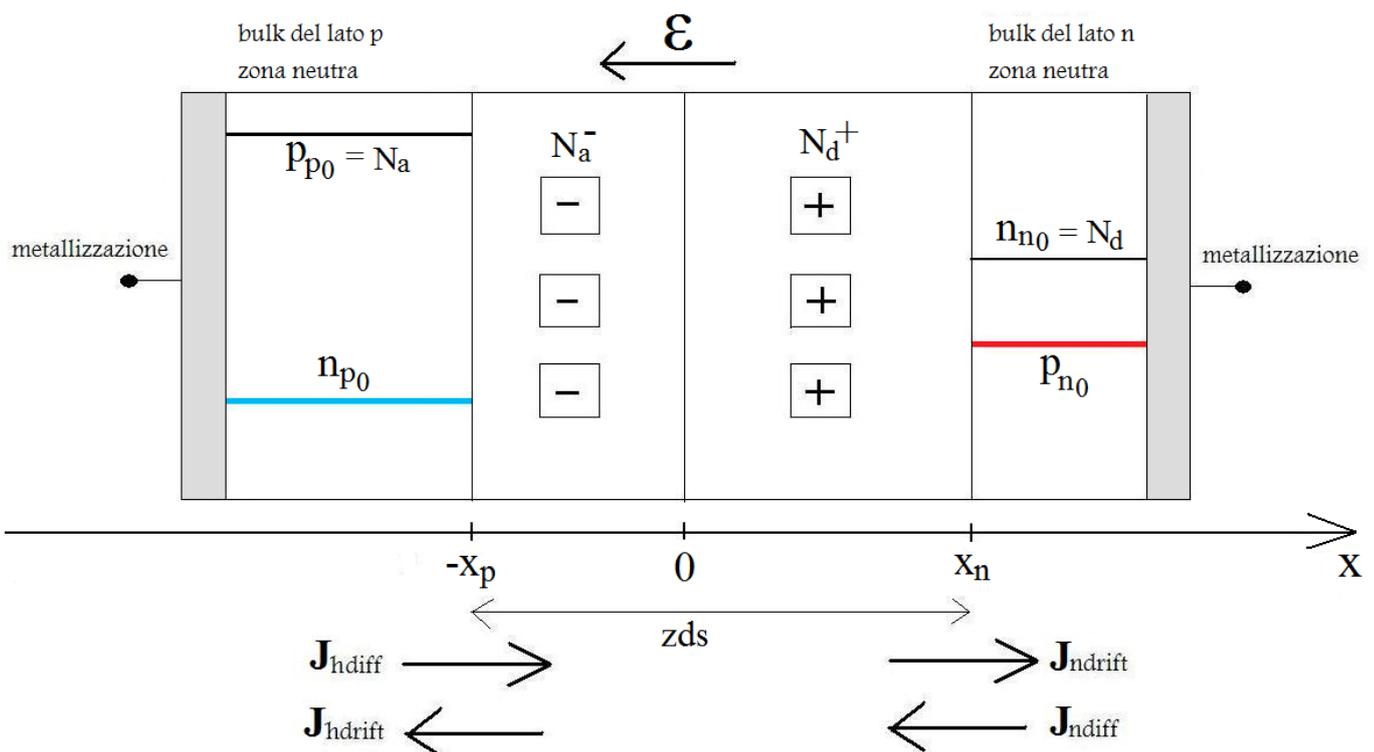


Fig. 11

Rappresentazione schematica di un'omogiunzione brusca pn con relative metallizzazioni di contatto. Sono presenti i profili delle concentrazioni dei droganti (coincidenti con quelli delle concentrazioni dei portatori di carica mobili maggioritari), i profili dei minoritari e la zona di svuotamento della carica mobile (regione di carica spaziale, responsabile del campo ϵ).

Nelle considerazioni analitiche che seguiranno (a livello concettuale non perdiamo in generalità) assumeremo le seguenti ipotesi:

- dispositivo monodimensionale
- svuotamento completo all'interfaccia silicio drogato p (Sip)/silicio drogato n (Sin), ovvero assenza di cariche mobili; sono presenti soltanto le regioni accettrici e donatrici di elettroni, ovvero la carica spaziale che crea il campo ε nella zona di svuotamento zds
- temperatura uniforme di svuotamento, ovvero gli elettroni nel Sin sono tutti e soli quelli dovuti allo svuotamento degli stati donatori localizzati presso i droganti del 5° gruppo, mentre le lacune nel Sip sono tutte e sole quelle dovute al riempimento degli stati accettori localizzati presso i droganti del 3° gruppo; approssimativamente $50K < T < 500K$
- validità della "legge di azione di massa" $np = n_i^2$ nei bulk dei lati p ed n, dove n_i è la concentrazione dei portatori nel Si intrinseco che, a temperatura ambiente, vale circa $1.45 \times 10^{10} \text{ cm}^{-3}$, ed aumenta esponenzialmente con la temperatura; i profili delle concentrazioni dei portatori minoritari (linea blu e rossa) sono costanti nei due bulk e valgono:

$$n_{po} = \frac{n_i^2}{N_a} \qquad p_{no} = \frac{n_i^2}{N_d}$$

- contatti ohmici fra ciascuna metallizzazione ed il silicio: la barriera di potenziale fra metallo e Si è talmente sottile che, per effetto tunnel, le cariche la attraversano con coefficiente di trasmissione T prossimo ad 1, per cui la resistenza presentata da tali interfacce è molto minore (trascurabile) rispetto alla resistenza, in serie, presentata dalla zds
- anche in virtù della precedente considerazione, "l'effetto diodo" è concentrato totalmente fra $-x_p$ e x_n

Mettiamoci in condizione di equilibrio, ovvero quando la giunzione non è polarizzata. Dal modello "drift - diffusion" (trascinamento - diffusione) si ha che:

$$J_{ntot} = qD_n \frac{dn(x)}{dx} + q\mu_n n(x)\varepsilon(x) = 0 \qquad J_{htot} = -qD_p \frac{dp(x)}{dx} + q\mu_p p(x)\varepsilon(x) = 0 \quad [e5]$$

dove $n(x)$ e $p(x)$ sono la concentrazione (cm^{-3}) degli elettroni e delle lacune, μ_n e μ_p le mobilità ($\text{cm}^2 \text{ V}^{-1} \text{ s}^{-1}$) degli elettroni e delle lacune, D_n e D_p i coefficienti di diffusione ($\text{cm}^2 \text{ s}^{-1}$) di elettroni e lacune, q la carica elementare, $\varepsilon(x)$ il campo elettrico prodotto dai donatori e dagli accettori ionizzati, all'interno della regione di svuotamento. I flussi dei due tipi di portatori di carica sono nulli, ovvero la componente diffusiva è perfettamente compensata da quella di trascinamento. Questo equilibrio (stazionario, termico ed elettrico) è spiegabile anche mediante il diagramma a bande di energia di figura 12.

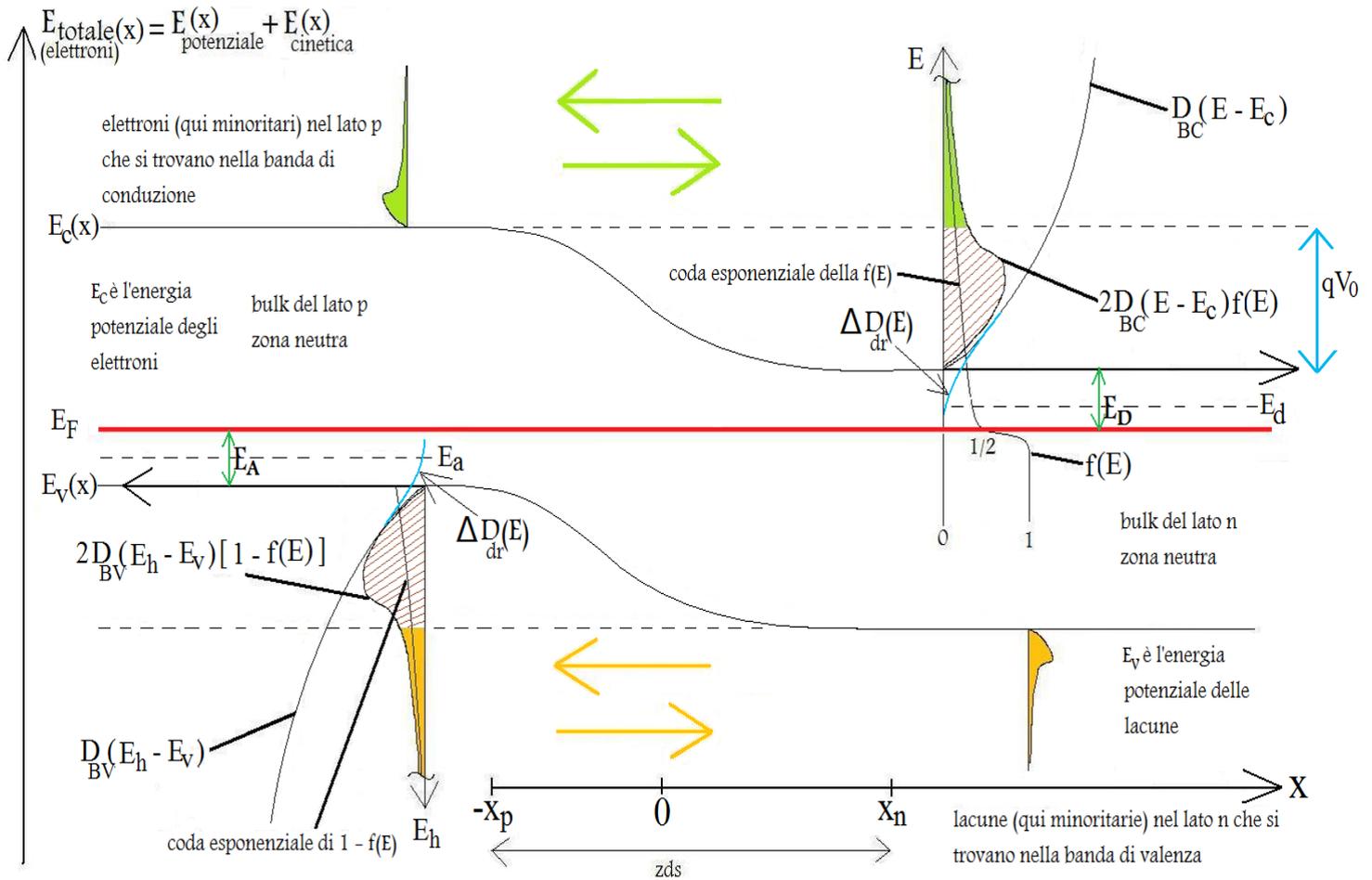


Fig. 12

Diagramma a bande di un'omogiunzione pn in condizioni di equilibrio stazionario, termico ed elettrico. Al di sopra di E_C sono rappresentate le concentrazioni $n(E) = 2D_{BC}(E - E_C)f(E)$, in unità di volume e di energia ($\text{eV}^{-1}\text{cm}^{-3}$), degli elettroni, maggioritari nel lato n, minoritari in quello p. Dualmente, al di sotto di E_V sono rappresentate le concentrazioni $p(E_h) = 2D_{BV}(E_h - E_V)[1 - f(E)]$, in unità di volume e di energia ($\text{eV}^{-1}\text{cm}^{-3}$), delle lacune. E_h è l'energia totale delle lacune, il cui vettore è contrario rispetto a quello relativo all'energia totale degli elettroni. $\Delta D_{dr}(E)$ rappresenta la concentrazione di stati permessi in BC ed in BV, in unità di volume e di energia ($\text{eV}^{-1}\text{cm}^{-3}$), individuati dal posizionamento casuale e disordinato delle impurezze droganti all'interno della matrice cristallina del semiconduttore: si noti come questi stati "spuri" siano localizzati sia all'interno delle bande di conduzione e di valenza che all'interno del gap, dove non ci dovrebbero essere stati occupabili da elettroni o lacune.

4.1.2) La omogiunzione pn in condizione di equilibrio: richiami di fisica dei semiconduttori e della giunzione pn

In riferimento al diagramma a bande riportato in figura 12, identifichiamo con $D_{BC}(E - E_C)$ la ben nota "densità tridimensionale di stati in banda di conduzione":

$$D_{BC}(E - E_C) = 2\pi \left(\frac{2m_{ed}^*}{h^2} \right)^{\frac{3}{2}} \sqrt{E - E_C} \quad (\text{eV}^{-1}\text{cm}^{-3})$$

$D_{BC}(E - E_C)$ è proporzionale alla radice quadrata dell'energia totale E degli elettroni. m_{ed}^* è la "massa efficace dell'elettrone per la densità di stati", che in un semiconduttore a gap diretto (GaAs, InAs, InP ecc...) coincide con qualunque altro tipo di massa efficace dell'elettrone, ad esempio con la "massa efficace per la mobilità" ricavabile dalla concavità, valutata in un particolare punto dello spazio reciproco dei vettori d'onda, del diagramma a bande $E - k$ relativo al semiconduttore, oppure con la "massa efficace dell'elettrone per la conducibilità". Tuttavia in un semiconduttore a gap indiretto (C, Si, Ge, GeSi, GaP, AlAs ecc...) m_{ed}^* ha una sua particolare definizione, sempre basata sul concetto di massa efficace per la mobilità; nel caso del silicio m_{ed}^* è data dalla seguente espressione:

$$m_{ed}^* = 6^{\frac{2}{3}} [m_{eL}^* (m_{eT}^*)^2]^{\frac{1}{3}} \approx 1.08m_0$$

dove il 6 corrisponde al numero di valli X equivalenti, rintracciabili sul diagramma a bande $E - k$ del silicio, ovvero al numero di superfici ellissoidali equienergetiche $E - k$ (si veda il glossario), m_{eT}^* è la massa efficace “trasversale” dell’elettrone per la mobilità, che nel silicio vale circa $0.19m_0$, mentre m_{eL}^* è la massa efficace “longitudinale” dell’elettrone per la mobilità, che nel silicio vale circa $0.98m_0$. m_{eL}^* e m_{eT}^* presentano questi valori se sono valutate presso il punto di simmetria X .

Sempre in riferimento al diagramma a bande mostrato in figura 12, identifichiamo con $f(E)$ la ben nota “distribuzione statistica quantistica di Fermi – Dirac”:

$$f(E) = \frac{1}{1 + e^{\frac{E-E_F}{kT}}}$$

alla quale obbediscono tutte le particelle aventi spin semi-intero, dette “fermioni”, quindi anche gli elettroni. Il suo “sfarfallamento”, dipendente dalla temperatura del semiconduttore, intorno al punto fisso ($f(E_F) = 1/2$ per qualunque temperatura) è stato esagerato, nel diagramma riportato in figura 12, per motivi di chiarezza.

E_d è il livello energetico individuato dagli atomi droganti del 5° gruppo (per il silicio le specie droganti tipo n comunemente utilizzate sono, come noto, P, As e Sb), ovvero i livelli donatori di elettroni. Questi risultano praticamente tutti svuotati: gli elettroni che li popolano alla temperatura di 0 K, avendo ipotizzato che la giunzione pn si trova a temperature ordinarie, ossia di svuotamento, sono quasi tutti andati in banda di conduzione, per cui $f(E_d) \approx 0$, ovvero la probabilità di occupazione di quei livelli è trascurabile.

E_a è il livello energetico individuato dagli atomi droganti del 3° gruppo (usualmente viene utilizzato B per il silicio, e talvolta, soprattutto in strutture non di silicio, anche Al, Ga, In), ovvero i livelli accettori di elettroni. Questi, per temperature ordinarie, risultano quasi tutti riempiti da elettroni provenienti dalla banda di valenza, per cui $f(E_a) \approx 1$.

Per livelli di drogaggio n non troppo alti ($10^{14} \text{ cm}^{-3} < N_d < 10^{18} \text{ cm}^{-3}$) l’energia E_{D0} ($=E_D$ per drogaggi bassi) necessaria alla ionizzazione del singolo atomo di impurezza appartenente al 5° gruppo dipende fortemente dal tipo di semiconduttore che è stato drogato, cioè dal cristallo che ospita l’impurezza (“host cristall”), e debolmente dalla natura dell’atomo drogante. Nell’ipotesi in cui “l’extraelettrone” (il quinto elettrone), che per $T = 0 \text{ K}$ è posizionato su uno stato all’interno del gap localizzato ad energia E_d , risenta unicamente dell’attrazione coulombiana prodotta dall’atomo pentavalente al quale appartiene, e nell’ipotesi in cui l’atomo pentavalente possa essere modellizzato mediante un sistema costituito dall’extraelettrone e da una carica positiva che funge da centro di forze elettriche del sistema (è evidente l’analogia con l’atomo di idrogeno), possiamo scrivere la seguente espressione approssimata dell’energia di ionizzazione per un’impurezza drogante tipo n :

$$E_{D0} \approx \frac{\mu e^4}{8\epsilon_0^2 \epsilon_r^2 h^2} \approx \frac{13.6 \text{ eV}}{\epsilon_r^2} \quad [e6]$$

E_{D0} è dell’ordine di qualche centesimo di eV, per cui E_d è vicinissimo a E_C . La forte dipendenza di E_{D0} dal cristallo semiconduttore che “ospita” l’elemento drogante è rappresentata dalla costante dielettrica relativa ϵ_r del semiconduttore (≈ 11.8 per il Si, ≈ 13.2 per il GaAs). μ è la massa ridotta associata al sistema “atom-like”, ovvero il modello “ad atomo di idrogeno” attraverso il quale possiamo descrivere l’atomo del 5° gruppo; il sistema “atom-like” è formato dall’extraelettrone (e^-), la cui massa è pari alla massa a riposo degli elettroni m_0 in quanto non appartenente, per $T \approx 0 \text{ K}$, né alla BV né alla BC, e dal protone (e^+), di massa m_p ($\approx 1840 m_0$). In formula:

$$\mu = \frac{m_0 m_p}{m_0 + m_p}$$

Più precisamente la [e6] può essere scritta nel modo seguente:

$$E_{D0} \approx \frac{13.6 \text{ eV}}{\epsilon_r^2} \frac{m_\sigma^*}{m_0} \quad [e7]$$

dove m_σ^* è la “massa efficace dell’elettrone per la conducibilità”, che quantifica la “reattività” dell’elettrone, in BC, nei confronti degli stimoli elettrostatici imposti dall’esterno (campi di polarizzazione). m_σ^* è la massa presente al denominatore dell’espressione della mobilità per gli elettroni ed è definita nel modo seguente per qualunque semiconduttore a gap indiretto:

$$m^*_\sigma = 3 \left(\frac{2}{m^*_{eT}} + \frac{1}{m^*_{eL}} \right) \cdot 1$$

m^*_σ vale circa $0.26m_0$ nella valle X del Si, mentre nel caso del GaAs m^*_σ coincide con m^*_e e con m^*_{ed} (essendo GaAs a gap diretto); nella valle Γ del GaAs m^*_σ vale circa $0.067m_0$. Forniamo alcuni valori di E_{D0} , ricavati dalla [e7], per tre materiali utilizzati in campo fotonico:

$$E_{D0(\text{Si})} \approx 25 \text{ meV}$$

$$E_{D0(\text{GaAs})} \approx 5.2 \text{ meV}$$

$$E_{D0(\text{Ge})} \approx 12.7 \text{ meV}$$

Dal modello appena discusso si evince la forte dipendenza, legata ad ϵ_r e m^*_σ , di E_{D0} dalla natura del cristallo “ospitante” (drogato). Gli errori nel calcolo di E_{D0} sono legati principalmente al fatto che il fittaggio del diagramma $E(k)$ del semiconduttore è il risultato di un’interpolazione fra punti (punti = quote energetiche = energie elettroniche totali E presso i punti (k_x, k_y, k_z)), pertanto le concavità calcolabili dal diagramma $E - k$ sono soggette ad errore; quest’ultimo si ripercuote sul calcolo delle masse efficaci per la mobilità e quindi sia sulle masse efficaci per le densità di stati che sulle masse efficaci per la conducibilità. Inoltre il modello discusso è molto semplice e non tiene conto di molti fattori presenti nella realtà, come ad esempio le interazioni dell’extraelettrone con i campi coulombiani prodotti dagli atomi vicini (primi vicini, secondi vicini ecc...), sia che questi siano gli atomi costituenti il semiconduttore che altri atomi droganti, o come le interazioni fra l’extraelettrone e gli altri extraelettroni appartenenti agli atomi pentavalenti vicini. Pertanto l’energia E_{D0} varia sensibilmente in funzione della specie drogante utilizzata, per quanto tale dipendenza non si evince dalla [e7]: sperimentalmente si osserva che nel Si con drogaggio di P E_{D0} vale 45 meV, con drogaggio di As vale 49 meV, con drogaggio di Sb 39 meV, benchè dal modello teorico riassunto nella [e7] si ottenga un valore di E_{D0} pari a circa 25 meV. Sempre sperimentalmente si osserva che nel GaAs il valore di E_{D0} fluttua intorno a 5.2 meV in funzione dell’elemento tetravalente (C, Si, Ge, St) che, sostituendosi ad atomi di Ga (trivalenti), drogano di tipo n il GaAs. Ragionamenti del tutto analoghi possono essere fatti per il livello energetico E_a e per l’energia E_{A0} che deve essere fornita ad un’impurezza drogante accettrice, ovvero un atomo trivalente, per consentirne la ionizzazione e quindi per promuovere la sua lacuna da E_a alla BV, nel caso di drogaggio p non elevato ($10^{14} \text{ cm}^{-3} < N_a < 10^{18} \text{ cm}^{-3}$). In formula:

$$E_{A0} \approx \frac{13.6 \text{ eV}}{\epsilon_r^2} \frac{m^*_h}{m_0} \approx \frac{13.6 \text{ eV}}{\epsilon_r^2} \frac{m^*_{HH}}{m_0} \quad [e8]$$

m^*_{HH} è la massa efficace della lacuna pesante (HH = “heavy hole” = lacuna pesante) per la mobilità e vale circa $0.5m_0$ nel Si e $0.45m_0$ nel GaAs, per cui E_{A0} vale circa 48 meV nel Si e 36 meV nel GaAs. m^*_h , la cui definizione formale è riportata nella [e9], è invece la “massa efficace della lacuna per la mobilità”, che coincide con la massa efficace della lacuna per la densità di stati in BV e con la massa efficace della lacuna per la conducibilità, sia nei semiconduttori a gap diretto che in quelli a gap indiretto. Sperimentalmente si osserva che nel Si con drogaggio di B E_{A0} vale 45 meV, mentre con drogaggio di Al 57 meV. Sempre sperimentalmente si osserva che nel GaAs il valore di E_{A0} fluttua intorno a 36 meV in funzione dell’elemento tetravalente (C, Si, Ge, St) che, sostituendosi ad atomi di As (pentavalenti), drogano di tipo p il GaAs. Per tenere conto di tali discrepanze dai valori teorici di E_{D0} ed E_{A0} è necessario complicare il semplice modello “atom – like” approntando delle modifiche alle formule [e7] ed [e8], aggiungendo cioè dei termini potenziali perturbativi alla semplice energia potenziale di Coulomb. Tali correzioni sono note come “central cell corrections”.

L’espressione di m^*_h è la seguente:

$$m^*_h = \left[(m^*_{HH})^{\frac{3}{2}} + (m^*_{LH})^{\frac{3}{2}} \right]^{\frac{2}{3}} \quad [e9]$$

m^*_{LH} è la massa efficace della lacuna leggera (LH = “light hole” = lacuna leggera) per la mobilità. Nel Si m^*_{LH} è circa $0.15m_0$, valutata presso il punto Γ della BV del diagramma $E(k)$ del Si, mentre nel GaAs vale circa $0.08m_0$, sempre valutata nello stesso punto del diagramma $E(k)$ del GaAs. m^*_h costituisce una sorta di massa fittizia che tiene conto del fatto che la banda di valenza BV di un qualunque cristallo semiconduttore “di tipo bulk”, ovvero cristallo non comprensivo di nanostrutture quali “quantum wells”, “quantum wires” o “quantum dots” (pozzi quantici, fili quantici, punti quantici), ovvero un semiconduttore privo di qualunque effetto di confinamento quantistico, cioè nanometrico, dei portatori di carica (bulk = struttura tridimensionale “classica” di un cristallo semiconduttore), come noto, non è unica. La BV, in molti semiconduttori inorganici di tipo bulk, degenera in tre sottobande $E - k$ distinte. Ciò è dovuto al fatto che un qualunque stato elettronico, sia in BV che in BC, può essere

rappresentato ondulatoriamente, applicando il ben noto teorema di Bloch, mediante una funzione d'onda di Bloch $\psi(r,k)$, la quale è data dal prodotto fra un'onda piana monocromatica e una funzione spaziale $u(r)$: quest'ultima è chiamata, in fisica dello stato solido, "central cell part" della funzione di Bloch (= "parte centrale di cella" della funzione di Bloch), la cui periodicità spaziale è la stessa del cristallo semiconduttore, vale a dire la costante reticolare a (A°). Come noto all'interno di ciascuna cella fondamentale del reticolo cristallino del semiconduttore, quest'ultimo sia a gap diretto che a gap indiretto, la dipendenza spaziale $u(r)$ della funzione relativa ad uno stato elettronico in BV, cioè la dipendenza spaziale della densità di probabilità $|u(r)|^2$ relativa alla funzione di Bloch di un elettrone o di una lacuna in BV, ha una forma molto simile a quella caratteristica degli orbitali elettronici p_x , p_y e p_z . Infatti gli elettroni in BV di un qualunque semiconduttore (strutturato a diamante, oppure strutturato a zincoblenda, ecc...) contribuiscono a tenere uniti gli atomi che lo formano partecipando ai legami covalenti di tipo σ fra atomi primi vicini, e tali legami sono costituiti da sovrapposizioni di orbitali di tipo p lungo le tre direzioni, per la precisione da sovrapposizioni di orbitali ibridi sp^3 . Un elettrone (o una lacuna) in BV può quindi avere tre diversi valori dell'energia totale E , a seconda che lo stato elettronico di Bloch $\psi(r,k)$, che descrive l'elettrone (o la lacuna) in BV, abbia la forma spaziale $u(r)$ di p_x , di p_y o di p_z , ovvero a seconda dell'orbitale p sul quale il portatore di carica si trova. Ciascuno dei tre valori dell'energia totale E appartiene ad una sottobanda diversa di BV. Si osservi in proposito la parte sinistra di figura 13.

In altri termini la BV può essere vista come la degenerazione del livello energetico atomico di un singolo orbitale p , il cui elettrone non è isolato all'interno del sistema atomico a cui appartiene, bensì interagisce con gli elettroni ed i nuclei di tutti gli atomi vicini, per cui maggiore è la vicinanza, e quindi l'interazione coulombiana, con gli atomi vicini, più l'energia dell'orbitale p , la quale sarebbe un singolo livello energetico nel caso di atomo isolato, ovvero non all'interno di un cristallo, degenera in una banda di valenza sempre più larga lungo l'asse dell'energia totale E . Se poi consideriamo non trascurabile la nota interazione spin - orbita ("spin - orbit coupling"), ovvero l'interazione fra il momento angolare intrinseco (= di spin) dell'elettrone che si trova in BV ed il campo magnetico prodotto dall'orbita del nucleo degli atomi del semiconduttore intorno ai centri di massa degli stessi atomi, in altri termini l'overlap (= prodotto scalare) fra il momento angolare di spin dell'elettrone \vec{S} ed il momento angolare orbitale totale dell'atomo \vec{L} a cui l'elettrone appartiene, che si può dimostrare essere, di fatto, coincidente con il momento angolare orbitale dello stesso elettrone orbitante intorno al centro di massa dell'atomo (il momento angolare orbitale del nucleo intorno al centro di massa è trascurabile, in quanto $m_0 \ll m_{\text{nucleo}}$), allora le tre sottobande di valenza si separano, dal punto di vista energetico, ancora più nettamente, come mostrato nella parte destra di figura 13.

Ragionamenti abbastanza simili possono essere fatti anche per la banda di conduzione BC. Applicando sempre il teorema di Bloch, un elettrone in BC può essere descritto da una funzione di Bloch $\psi(r,k)$, la quale è data dal prodotto fra un'onda piana monocromatica ed una funzione spaziale $u(r)$, chiamata sempre "central cell part" della funzione di Bloch, la cui periodicità spaziale è la stessa del cristallo semiconduttore, ovvero la costante reticolare a (A°). All'interno di ciascuna cella fondamentale del reticolo cristallino di un semiconduttore a gap diretto, la dipendenza spaziale $u(r)$ dello stato elettronico in BC, cioè la dipendenza spaziale della densità di probabilità $|u(r)|^2$ relativa alla funzione di Bloch dell'elettrone in BC, ha una forma molto simile a quella caratteristica dell'orbitale elettronico atomico s , il quale possiede simmetria sferica. All'interno di ciascuna cella fondamentale del reticolo cristallino di un semiconduttore a gap indiretto, la dipendenza spaziale $u(r)$ dello stato elettronico in BC ha, invece, una forma simile a quella relativa ad un "mescolamento" (= sovrapposizione) dell'orbitale atomico s con un orbitale atomico p . Quindi la BC di un semiconduttore a gap diretto è la degenerazione del livello energetico atomico di un singolo orbitale s , il cui elettrone, essendo all'interno di un cristallo e non all'interno del solo atomo a cui appartiene, risente dell'interazione con N atomi vicini che formano il cristallo, pertanto il suo livello energetico discreto degenera in un insieme, molto fitto, di N livelli energetici discreti, dando luogo alla banda di conduzione. Analogamente la BC di un semiconduttore a gap indiretto è la degenerazione del livello energetico atomico di un singolo orbitale atomico ibrido $s + p$, il cui elettrone, essendo all'interno di un cristallo, risente dell'interazione con N atomi, pertanto il suo livello energetico degenera in un insieme, molto fitto, di N livelli energetici discreti, formando la banda di conduzione. Il meccanismo di degenerazione di un livello energetico atomico discreto (p , s oppure $s + p$) in una banda (BV oppure BC), che rappresenta la struttura energetica tipica dei cristalli semiconduttori, costituisce il principio fisico che correlato alla tecnica di calcolo riassunta nell'espressione [g6] del glossario, fornisce la sintesi del metodo di calcolo delle bande di energia noto come TBM, ovvero "Tight Binding Method".

Ciascun punto della curva $E - k$ della BC rappresenta l'energia, presso un particolare vettore d'onda $k = (k_x, k_y, k_z)$, di una qualunque combinazione lineare dei due vettori ortonormali $|\sigma\uparrow\rangle$ e $|\sigma\downarrow\rangle$, che qui indichiamo, secondo la notazione astratta di Dirac (si veda il glossario), utilizzando i simboli di bra e ket. Nei semiconduttori a gap

diretto, dove $u(r)$, per gli elettroni in BC, ha simmetria sferica s, ma anche in quelli a gap indiretto, dove la simmetria di $u(r)$, che a rigore è data da un mescolamento degli orbitali s e p, è di fatto dominata da quella dell'orbitale s (la componente dovuta all'orbitale p può essere trascurata), la BC, al contrario di BV, non subisce evidenti alterazioni dovute all'interazione spin - orbita, poiché \vec{L} dell'atomo, di fatto uguale al momento angolare orbitale dell'elettrone, è praticamente nullo, quindi anche l'overlap fra \vec{L} e \vec{S} è, di fatto, nullo. Pertanto presso un certo vettore d'onda $k = (k_x, k_y, k_z)$ l'autovalore dell'energia dello stato elettronico $|\sigma\uparrow\rangle$ approssimativamente coincide con l'autovalore dell'energia dello stato elettronico $|\sigma\downarrow\rangle$, rendendo la doppia degenerazione della BC, al contrario della tripla degenerazione che interessa la BV, decisamente trascurabile (come mostrato in figura 13).

Per ricavare operativamente il diagramma a bande riportato nella parte sinistra di figura 13 sarebbe necessario utilizzare un programma computerizzato di fitting che implementa uno dei metodi di calcolo dei diagrammi E - k, come ad esempio il già citato TBM, che è il metodo più generale in quanto fornisce risultati accettabili in qualunque regione del diagramma E - k (storicamente il TBM è stato il primo), oppure il metodo dello "pseudopotenziale", oppure il metodo "k-p"; quest'ultimo fornisce buoni risultati solo presso i massimi di BV ed i minimi di BC, ovvero presso i cosiddetti "bandedges". Concettualmente qualunque approccio di calcolo deriva dalla risoluzione dell'equazione di Schrödinger per il "problema cristallino":

$$\hat{H}_{cristallo}|\psi_k\rangle = E(k) |\psi_k\rangle$$

dove:

$$\hat{H}_{cristallo} = \hat{H}_{atomico} + \Delta U \quad [e10]$$

$\hat{H}_{cristallo}$ è l'operatore hamiltoniano i cui N (= numero degli atomi vicini con cui l'elettrone $|\psi_k\rangle$ interagisce, compreso l'atomo di appartenenza) autovalori dell'energia rappresentano i punti dell'intera struttura a bande E - k (struttura "full bands") formata, complessivamente, da N sottobande, alcune delle quali sono sottobande di valenza e altre sottobande di conduzione (un autovalore per ciascuna sottobanda), presso un determinato vettore d'onda $k = (k_x, k_y, k_z)$. $|\psi_k\rangle$ è dato da una particolare sovrapposizione di N stati $|\psi_i\rangle$, ciascuno dei quali è un'orbitale atomico che supponiamo strettamente legato al nucleo dell'atomo di appartenenza (da cui il nome TBM - "Tight Binding Method"): è proprio questa ipotesi che ci consente di affermare che l'interazione fra due qualsiasi stati atomici $|\psi_i\rangle$ è trascurabile, e quindi che gli stati $|\psi_i\rangle$ formano un sistema completo sul quale poter sviluppare $|\psi_k\rangle$ (= sviluppo di uno stato cristallino $|\psi_k\rangle$ a partire da stati puramente atomici $|\psi_i\rangle$). $\hat{H}_{cristallo}$, ossia l'hamiltoniano risolvete il "problema cristallino" degli N autovalori dell'energia dell'autostato $|\psi_k\rangle$, è dato dall'operatore hamiltoniano risolvete il "problema atomico", ovvero l'hamiltoniano valido nel caso in cui l'atomo a cui appartiene l'elettrone sia isolato ($\hat{H}_{atomico}$ è l'hamiltoniano con il quale scriviamo l'equazione di Schrödinger valida per l'atomo idrogenoide), cioè non facente parte di un reticolo cristallino, sommato al termine potenziale ΔU , il quale tiene conto dell'interazione coulombiana cristallina responsabile della degenerazione di un livello energetico discreto in una banda.

Per ricavare concettualmente il diagramma a bande riportato nella parte destra di figura 13 è necessario tenere conto dell'interazione spin - orbita, risolvendo l'equazione di Schrödinger con un operatore hamiltoniano \hat{H}_{totale} la cui espressione astratta è:

$$\hat{H}_{totale} = \hat{H}_{cristallo} + \hat{H}_{SO}$$

\hat{H}_{SO} è l'operatore hamiltoniano che implementa l'interazione esistente fra lo spin dell'elettrone ed il campo magnetico prodotto dalla precessione del nucleo dell'atomo intorno al centro di massa di quest'ultimo. Assumendo che l'elettrone avente momento angolare intrinseco \vec{S} risenta unicamente del campo magnetico prodotto dal nucleo dell'atomo di appartenenza, in altri termini unicamente del momento angolare orbitale \vec{L} dell'atomo stesso, possiamo adottare la seguente espressione per l'hamiltoniano di spin - orbita:

$$\hat{H}_{SO} = \left[\frac{\mu_B}{\hbar c^2 e m_0} \frac{1}{r} \frac{\partial U_{Coulomb}(r)}{\partial r} \right] \vec{L} \cdot \vec{S} = \lambda \vec{L} \cdot \vec{S} \quad [e11]$$

dove r è, approssimativamente, il vettore posizione dell'elettrone rispetto al centro di massa dell'atomo, \cdot è ovviamente un prodotto scalare e μ_B è il ben noto "magnetone di Bohr", definito nel modo seguente:

$$\mu_B = \frac{e\hbar}{2m_0} \approx 0.927 \times 10^{-23} \text{ Am}^2$$

Nella trattazione che seguirà si renderà necessario utilizzare un formalismo di meccanica quantistica molto utilizzato per lo studio dei materiali adatti per scopi fotonici, dunque anche per fotorivelazione: il sistema completo di autofunzioni $\Phi_{j,j_z}(r, \theta, \varphi)$ di Clebsch – Gordan. Questa base ortonormale consta di sei vettori ed è utile per descrivere lo stato di un elettrone o di una lacuna in banda di valenza. I sei vettori $\Phi_{j,j_z}(r, \theta, \varphi)$ in questione sono: $\Phi_{3/2;3/2}, \Phi_{3/2;-3/2}, \Phi_{3/2;1/2}, \Phi_{3/2;-1/2}, \Phi_{1/2;1/2}, \Phi_{1/2;-1/2}$. Queste funzioni sono indicizzate dal “numero quantico J del momento angolare totale” (per momento angolare totale \vec{J} si intende il momento angolare orbitale \vec{L} sommato al momento angolare di spin \vec{S}) e dal “numero quantico J_z della terza componente del momento angolare totale”. L’esigenza di utilizzare i due numeri quantici appena citati deriva dal fatto che se adoperiamo la base ortonormale “consueta”, che chiameremo ρ , per descrivere lo stato di un elettrone o di una lacuna in BV, costituita dai sei vettori $p_x \uparrow, p_x \downarrow, p_y \uparrow, p_y \downarrow, p_z \uparrow, p_z \downarrow$ (detti anche “mixed states”, poiché ciascuno è formato dal “mescolamento” fra stati puri $Y_{l,m}(\theta, \varphi)$ relativi al momento angolare orbitale e stati puri χ_s relativi al momento angolare di spin), allora una generica funzione $\psi(r, \theta, \varphi, \uparrow \text{ o } \downarrow)$ sviluppata all’interno del sottospazio vettoriale (di L^2) individuato da ρ , non può essere espressa come il prodotto fra le seguenti funzioni:

- 1) funzione radiale $R_{n,l}(r)$, dove r è appunto la coordinata radiale, n è il numero quantico principale che indicizza gli autovalori dell’energia ed l è il numero quantico orbitale, che indicizza il tipo di orbitale (l’orbitale s è indicizzato da $l = 0$, quello p da $l = 1$ ecc...). L’espressione generale di $R_{n,l}(r)$ è:

$$R_{n,l}(r) = A_{n,l} r^l p(r) e^{-\frac{r}{na}}$$

dove $p(r)$ è un polinomio, nell’indeterminata r , di grado $n - l - 1$, mentre $A_{n,l}$ è una costante reale di normalizzazione

- 2) funzione angolare $Y_{l,m}(\theta, \varphi) = \theta_{l,m}(\theta)\phi_m(\varphi)$, dove φ è la coordinata angolare azimutale, θ la coordinata angolare di altezza, m il numero quantico magnetico; l’espressione completa di $Y_{l,m}(\theta, \varphi)$ è la seguente:

$$Y_{l,m}(\theta, \varphi) = \left\{ \sqrt{\frac{(l-|m|)!(2l+1)}{4\pi(l+|m|)!}} P_l^{|m|}(\cos \theta) \right\} e^{im\varphi}$$

dove $P_l^{|m|}(\cos \theta)$ sono i polinomi associati di Legendre, la cui espressione generale è:

$$P_l^{|m|}(\cos \theta) = [1 - \cos^2(\theta)]^{\frac{|m|}{2}} \frac{d^{|m|}}{d\cos^{|m|}(\theta)} \left\{ \left(\frac{1}{2^l l!} \right) \frac{d^l}{d\cos^l(\theta)} [(\cos^2(\theta) - 1)^l] \right\}$$

$\phi_m(\varphi)$ è autofunzione dell’operatore \hat{L}_z , mentre $Y_{l,m}(\theta, \varphi)$ è autofunzione degli operatori \hat{L}^2 ed \hat{L}_z . La funzione $\phi_m(\varphi)$ ha la seguente espressione normalizzata, tipica di un qualunque sistema oscillatorio:

$$\phi_m(\varphi) = \frac{1}{\sqrt{2\pi}} e^{im\varphi} = \frac{1}{\sqrt{2\pi}} e^{i\left(\frac{L_z}{\hbar}\right)\varphi}$$

- 3) funzionale di spin $\chi_s = \text{up “}\uparrow\text{” oppure down “}\downarrow\text{”}$, dove s è numero quantico di spin ($+/- \frac{1}{2}$). χ_s è autofunzione degli operatori \hat{S}^2 ed \hat{S}_z

Questo poiché gli operatori quantistici \hat{H} , \hat{L}^2 , \hat{L}_z ed \hat{S}^2 non commutano, quindi non esiste nessun sistema completo di autofunzioni condiviso dai suddetti operatori, in altri termini non esiste alcuna base ortonormale in cui ciascun vettore sia autofunzione di quegli operatori: data una generica funzione $\psi(r, \theta, \varphi, \uparrow \text{ o } \downarrow)$ costruita su un qualunque sistema completo, ad esempio ρ , le dipendenze da r, θ, φ e spin non fattorizzano. Se vogliamo che questa fattorizzazione sia possibile dobbiamo introdurre i nuovi operatori quantistici \hat{J} e \hat{J}_z , rispettivamente “operatore del momento angolare totale” e “operatore della terza componente del momento angolare totale”, i cui numeri quantici sono appunto $J = l + s = 1 +/- \frac{1}{2} = 3/2$ oppure $1/2$, e J_z , dove $-J < J_z < J$, pertanto $J_z = +/- 3/2$ e $+/- 1/2$ se $J = 3/2$, mentre $J_z = +/- 1/2$ se $J = 1/2$. Dunque il sistema completo $\Phi_{j,j_z}(r, \theta, \varphi)$ di Clebsch – Gordan è una base ortonormale di autofunzioni condivisa dagli operatori \hat{H} , \hat{J} e \hat{J}_z , rendendo così possibile la fattorizzazione appena discussa. Le funzioni $\Phi_{j,j_z}(r, \theta, \varphi)$ rappresentano stati puri relativi al momento angolare totale.

La banda di valenza BV consta di:

- una sottobanda di valenza indicata con l'acronimo HH, che significa appunto "heavy hole", cioè "lacuna pesante", ossia una sottobanda che ospita, presso gli stati permessi posizionati lungo la sua curva E - k, lacune aventi grandi masse efficaci per la mobilità. HH è una sottobanda caratterizzata, sul diagramma E - k, da una concavità poco pronunciata. Nella maggior parte dei semiconduttori HH ha la propria sommità localizzata presso il punto notevole di simmetria Γ ("high symmetry point" Γ), ovvero, all'interno del reticolo reciproco, presso il vettore d'onda nullo $k_0 = (k_{0x}, k_{0y}, k_{0z}) = (0, 0, 0)$. Ciascun punto della sottobanda HH rappresenta l'energia, presso un particolare vettore d'onda $k = (k_x, k_y, k_z)$, di una qualunque combinazione lineare dei due vettori ortonormali $\Phi_{j,jz}$ di Clebsch - Gordan $\Phi_{3/2;3/2}$ e $\Phi_{3/2;-3/2}$, spesso indicati, secondo la notazione astratta di Dirac (si veda il glossario), con i simboli di bra e ket nella forma seguente: $|3/2;3/2\rangle$ e $|3/2;-3/2\rangle$. In altre parole ciascun punto della sottobanda HH rappresenta l'autovalore dell'energia, presso un particolare vettore d'onda k, di un qualunque vettore appartenente al sottospazio vettoriale, definito sul campo complesso, individuato dai due vettori ortonormali $\Phi_{j,jz}$ di Clebsch - Gordan $\Phi_{3/2;3/2}$ e $\Phi_{3/2;-3/2}$. Queste due "central cell functions" sono così definite:

$$\Phi_{3/2;3/2} = -\frac{1}{\sqrt{2}} (|p_x \rangle + i |p_y \rangle) \uparrow$$

$$\Phi_{3/2;-3/2} = \frac{1}{\sqrt{2}} (|p_x \rangle - i |p_y \rangle) \downarrow$$

dove \uparrow rappresenta un funzionale, che moltiplica una combinazione lineare degli stati "classici" p_x e p_y , indicante il verso "up" dello spin dell'elettrone o, come nel caso in questione, della lacuna, mentre \downarrow un altro funzionale indicante il verso "down" dello spin della lacuna.

- una sottobanda di valenza indicata con l'acronimo LH, che significa appunto "light hole", cioè "lacuna leggera", ossia una sottobanda che ospita, presso gli stati permessi posizionati lungo la sua curva E - k, lacune aventi piccole masse efficaci per la mobilità. LH è una sottobanda caratterizzata, sul diagramma E - k, da una concavità più pronunciata di quella della sottobanda HH. Indipendentemente dal fatto che trascuriamo o meno l'interazione spin - orbita la sommità della sottobanda di valenza LH coincide esattamente con la sommità della sottobanda di valenza HH. Se consideriamo trascurabile l'interazione spin - orbita osserviamo, sul diagramma E - k, la sovrapposizione di HH con LH in più punti vicini a quello di sommità; se invece consideriamo non trascurabile l'interazione spin - orbita osserviamo, sul diagramma E - k, un'evidente separazione fra HH ed LH presso quegli stessi punti prossimi alla sommità. Ciascun punto della sottobanda LH rappresenta l'energia, presso un particolare vettore d'onda $k = (k_x, k_y, k_z)$, di una qualunque combinazione lineare dei due vettori ortonormali $\Phi_{j,jz}$ di Clebsch - Gordan $\Phi_{3/2;1/2}$ e $\Phi_{3/2;-1/2}$, spesso indicati, secondo la notazione astratta di Dirac, con i simboli di bra e ket nella forma seguente: $|3/2;1/2\rangle$ e $|3/2;-1/2\rangle$. Queste due "central cell functions" sono così definite:

$$\Phi_{3/2;1/2} = -\frac{1}{\sqrt{6}} [(|p_x \rangle + i |p_y \rangle) \downarrow - 2 |p_z \rangle \uparrow]$$

$$\Phi_{3/2;-1/2} = \frac{1}{\sqrt{6}} [(|p_x \rangle - i |p_y \rangle) \uparrow + 2 |p_z \rangle \downarrow]$$

- una sottobanda di valenza indicata con l'acronimo SO, che significa "split - off", cioè "tagliata fuori". Se trascuriamo l'interazione spin - orbita la sommità della sottobanda SO coincide con le sommità delle sottobande HH ed LH, se invece consideriamo non trascurabile tale interazione osserviamo, sul diagramma E - k, una traslazione verticale verso energie totali E inferiori, approssimativamente rigida, della sottobanda SO. Il punto di sommità di SO va a posizionarsi al di sotto delle sommità di HH ed LH, sempre presso il punto Γ , di un'energia pari a Δ_{SO} , quantità definita nel modo seguente:

$$\Delta_{SO} = \frac{3\lambda\hbar^2}{2}$$

dove λ è la quantità definita nella [e8]. Alcuni valori tipici di Δ_{SO} sono: 44meV per il Si (nel Si lo split – off della sottobanda SO è di fatto trascurabile), 0.14 eV per il InP, 0.29 eV per il Ge, 0.35 eV per il GaAs, 0.41 eV per il InAs, 0.82 eV per il InSb. Ciascun punto della sottobanda SO rappresenta l'energia, presso un particolare vettore d'onda $k = (k_x, k_y, k_z)$, di una qualunque combinazione lineare dei due vettori ortonormali $\Phi_{j;z}$ di Clebsch – Gordan $\Phi_{1/2;1/2}$ e $\Phi_{1/2;-1/2}$, spesso indicati, secondo la notazione astratta di Dirac, con i simboli di bra e ket nella forma seguente: $|1/2;1/2\rangle$ e $|1/2;-1/2\rangle$. Queste due “central cell functions” sono così definite:

$$\Phi_{1/2;1/2} = -\frac{1}{\sqrt{3}} \left[(|p_x\rangle + i|p_y\rangle) \downarrow + |p_z\rangle \uparrow \right]$$

$$\Phi_{1/2;-1/2} = -\frac{1}{\sqrt{3}} \left[(|p_x\rangle - i|p_y\rangle) \uparrow - |p_z\rangle \downarrow \right]$$

Si può dimostrare che il valor medio di \hat{H}_{SO} , ovvero la perturbazione energetica media che subisce uno stato elettronico $\Phi_{j;z}(r, \theta, \varphi)$ in BV, a causa dell'interazione spin – orbita, è dato dalla seguente espressione:

$$\langle \hat{H}_{SO} \rangle = \langle \Phi_{j;z} | \hat{H}_{SO} | \Phi_{j;z} \rangle = \frac{\lambda \hbar^2}{2} [J(J+1) - l(l+1) - s(s+1)] = \frac{\Delta_{SO}}{3} \left[J(J+1) - \frac{11}{4} \right]$$

Va notato che per ottenere un diagramma a bande E – k come quello mostrato in figura 13, dove è riportato il caso particolare dell'GaAs alla temperatura di 300 K, utilizzando un programma di fitting che implementi il metodo di calcolo delle bande TBM, è necessario lavorare con uno stato elettronico $|\psi_k\rangle$ che sia ortogonale rispetto agli stati che descrivono il comportamento degli elettroni appartenenti ai gusci energetici più interni, ovvero i gusci non di valenza. Ad esempio nel Si gli elettroni di valenza, cioè quelli del guscio più esterno, appartengono agli orbitali 3s e 3p, per cui il guscio energetico di valenza è quello individuato dal numero quantico principale $n = 3$; gli elettroni appartenenti agli orbitali 1s, 2s e 2p, ovvero quelli appartenenti ai gusci energetici individuati dai numeri quantici principali $n = 1$ ed $n = 2$, sono gli elettroni interni, o non di valenza. Tali stati elettronici “profondi” sono noti come “core states” $|\xi_c\rangle$, costituiscono una base ortonormale e generalmente non hanno alcuna importanza dal punto di vista fotonico. Con buona approssimazione possiamo affermare che i core states non interagiscono in alcun modo con gli atomi del cristallo vicini a quello di appartenenza, per cui i loro livelli energetici discreti E_{Cn} , chiamati “core levels” ($n =$ numero quantico principale = indice dell'autovalore dell'energia nel problema puramente atomico), che sono calcolabili nell'ambito del problema atomico, ovvero come se l'atomo di appartenenza non faccia parte di un cristallo semiconduttore, non degenerano in bande e rimangono approssimativamente invariati nel passaggio dal problema atomico a quello cristallino; in altre parole i core states $|\xi_c\rangle$ non si accorgono del mondo esterno, del cristallo, non risentono di alcuna interazione coulombiana esterna all'atomo di appartenenza. Dunque per ciascun atomo all'interno della matrice cristallina sia gli autovalori dell'energia E_{Cn} che i relativi core states $|\xi_c\rangle$ sono calcolabili risolvendo l'equazione di Schrödinger valida per l'atomo isolato:

$$\hat{H}_{atomico} |\xi_c\rangle = E_{Cn}(k) |\xi_c\rangle \quad [e12]$$

Se identifichiamo con ϕ_k un generico stato elettronico avente vettore d'onda k, non ortogonale rispetto ai core states, quindi dipendente da questi (i $|\xi_c\rangle$ rappresentano alcune delle componenti di ϕ_k), questo stato può essere ortogonalizzato rispetto ai core states, che sono noti dalla [e12], ottenendo così lo stato elettronico ortogonale $|\psi_k\rangle$ ottimo per descrivere un elettrone in BV o in BC.

$$|\psi_k\rangle = \phi_k - \sum_n \sum_j (\phi_k \cdot \xi_{Cn,j}) \xi_{Cn,j}$$

$\xi_{Cn,j}$ può essere visto come un versore, $\phi_k \cdot \xi_{Cn,j}$ la norma del vettore di sovrapposizione fra ϕ_k ed il versore $\xi_{Cn,j}$, di conseguenza $(\phi_k \cdot \xi_{Cn,j}) \xi_{Cn,j}$ è il vettore di overlap fra il generico stato elettronico ϕ_k ed il core state $\xi_{Cn,j}$. $j (\neq n)$ indica che in uno stesso guscio energetico interno, indicizzato dal numero quantico principale n, ci possono essere più core states, indicizzati da j.

Se per fittare un diagramma a bande, ad esempio per il GaAs, comprendente gli effetti dell'accoppiamento spin – orbita risolvessimo la seguente equazione di Schrödinger:

$$\hat{H}_{totale} \phi_k = E_{TR}(k) \phi_k$$

quello che troveremmo non è esattamente il diagramma $E(k)$ che vorremmo, riportato nella parte destra della figura 13, bensì una versione di quel diagramma traslata (TR = traslato) verso l'alto di un'energia pari a:

$$E_{\text{offset}}(\mathbf{k}) = \sum_n \sum_j (\phi_{\mathbf{k}} \cdot \xi_{\text{Cn},j}) E_{\text{Cn}}(\mathbf{k})$$

quindi:

$$E_{\text{TR}}(\mathbf{k}) = E(\mathbf{k}) + E_{\text{offset}}(\mathbf{k})$$

Ciascun core state $|\xi_{\text{Cn},j}\rangle$ determina una traslazione del diagramma a bande $E(k)$ di un'energia pari al core level E_{Cn} , autovalore relativo a $|\xi_{\text{Cn},j}\rangle$, moltiplicato per il modulo dell'overlap fra $\phi_{\mathbf{k}}$ e lo stesso $|\xi_{\text{Cn},j}\rangle$; in altre parole la traslazione determinata dal singolo core state $|\xi_{\text{Cn},j}\rangle$ è data dall'energia E_{Cn} dello stesso core state modulata (= pesata) dalla dipendenza dello stato elettronico generico $\phi_{\mathbf{k}}$ da $|\xi_{\text{Cn},j}\rangle$. Per cui le sommità delle sottobande di valenza HH ed LH non si troverebbero a quota 0 eV, come è conveniente che sia, specialmente se i diagrammi a bande devono essere utilizzati per scopi fotonici, bensì a quota:

$$E_{\text{offset}}(\Gamma) = \sum_n \sum_j (\phi_{\mathbf{k}} \cdot \xi_{\text{Cn},j}) E_{\text{Cn}}(\Gamma)$$

Questo poichè $\phi_{\mathbf{k}}$ dipende dai core states, pertanto se applichiamo \hat{H}_{totale} a $\phi_{\mathbf{k}}$ otteniamo un diagramma a bande $E_{\text{TR}}(\mathbf{k})$ che dipende anche dai core levels E_{Cn} , autovalori relativi ai core states $|\xi_{\text{C}}\rangle$. Se invece lavoriamo con uno stato elettronico ortogonalizzato, ossia $|\psi_{\mathbf{k}}\rangle$, appartenente alla BV o alla BC, l'equazione di Schrödinger diventa:

$$\hat{H}_{\text{totale}} |\psi_{\mathbf{k}}\rangle = E(\mathbf{k}) |\psi_{\mathbf{k}}\rangle$$

dove $E(k)$ è esattamente il diagramma desiderato, mostrato a destra di figura 13, privo di offset energetico.

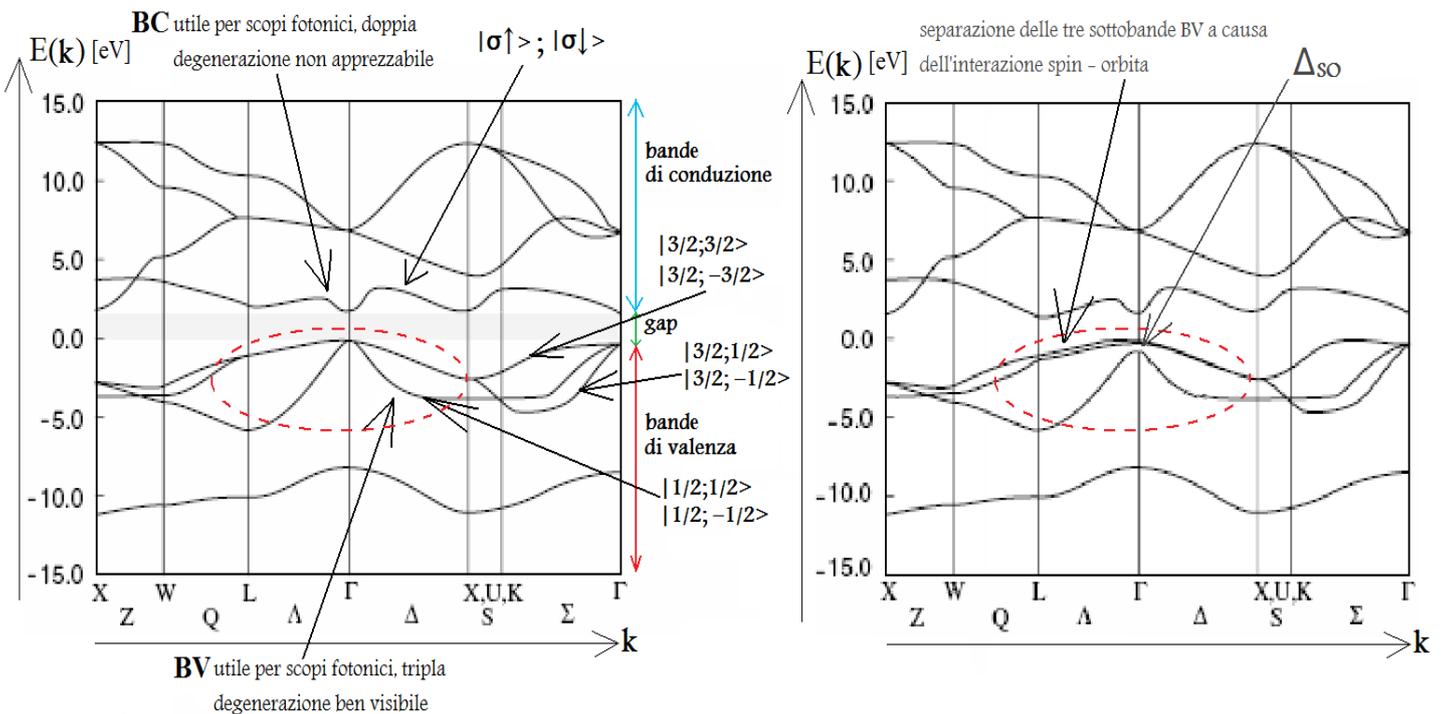


Fig. 13

Diagramma a bande $E - k$ per il GaAs alla temperatura di 300 K, ottenuto utilizzando un programma di fitting che implementa in metodo TBM per il calcolo delle bande di energia. A sinistra è mostrato il diagramma che non tiene conto dell'interazione spin-orbita, pertanto nella matrice hamiltoniana H_{totale} scritta in forma di Heisemberg, necessaria per il calcolo del "determinante secolare" (si veda il glossario) che fornisce gli autovalori dell'energia presso un vettore d'onda k fissato, non sono presenti elementi potenziali aggiuntivi ($h_{\text{SO}rc}$, $r = \text{riga}$, $c = \text{colonna}$) legati all'accoppiamento spin-orbita. A destra è mostrato il diagramma che tiene conto dell'interazione spin-orbita, pertanto nella matrice hamiltoniana H_{totale} in forma di Heisemberg sono presenti elementi potenziali aggiuntivi che tengono conto dell'accoppiamento spin-orbita. Si noti la regione di "bandedges" circondata dall'ellisse tratteggiata nei due diagrammi a bande: la separazione fra le sottobande di valenza HH ed LH e lo "splittamento" della sottobanda SO è visibile.

In figura 14 riportiamo un ingrandimento, molto qualitativo, della regione centrale del diagramma $E(k)$ mostrato a destra di figura 13, comprendente i “bandedges”, dove la separazione fra le sottobande di valenza HH ed LH e l’energia di split – off Δ_{SO} , dovute all’interazione spin – orbita, sono enfatizzate graficamente per ragioni di chiarezza, e dove è messa in risalto la differenza fra la curva $E - k$ della BC di un semiconduttore a gap diretto, che risulta dalla degenerazione di un orbitale atomico di tipo s, e la curva $E - k$ della BC di un semiconduttore a gap indiretto, che risulta dalla degenerazione di un orbitale atomico ibrido di tipo s + p. In proposito possiamo notare che la concavità presso il minimo della BC di un semiconduttore a gap indiretto (ad esempio C, Si, Ge, GeSi, GaP, AlAs ecc...), la cui matrice cristallina ha una struttura a diamante, è abbastanza meno pronunciata rispetto alla concavità presso il minimo della BC di un semiconduttore a gap diretto, cristallo la cui struttura è di tipo zincoblenda. In altre parole la BC, nel caso di gap indiretto, ospita elettroni aventi grande massa efficace per la mobilità e per questo motivo viene spesso indicata con HE (“heavy electron” – elettrone pesante), mentre nel caso di gap diretto la BC ospita elettroni aventi piccola massa efficace per la mobilità e quindi viene spesso indicata con LE (“light electron” – elettrone leggero). Questa differenza di concavità presso il minimo della BC, e quindi di mobilità per gli elettroni di conduzione, è spiegabile formalmente mediante la nota “teoria delle perturbazioni” (piccole perturbazioni). Gli elettroni che si trovano su stati posizionati, lungo la curva $E(k)$ della BC, appena al di sopra del minimo E_C , sono soggetti ad una perturbazione energetica dovuta ad agitazione termica oppure ad un campo elettrico $\vec{\varepsilon}$ imposto dall’esterno; supponiamo che le cause di tale perturbazione siano di piccola entità, pertanto che la temperatura del cristallo sia standard, cioè prossima a quella ambiente, e che i campi esterni $\vec{\varepsilon}$ di polarizzazione (del dispositivo) siano abbastanza piccoli, vale a dire di circa $10^4 \div 10^5$ V/cm. Nel caso di gap indiretto l’hamiltoniana H_{totale} di un elettrone in BC perturbato è data dalla somma fra l’hamiltoniana priva di perturbazione H_0 , la quale descrive un elettrone posizionato esattamente sul minimo E_C della BC, e l’hamiltoniana H_1 , che rappresenta la perturbazione cinetica subita dall’elettrone, nell’ipotesi di dispersione parabolica. Quest’ultima approssimazione è ragionevole, come ben noto, solo presso il minimo di BC ed il massimo di BV. In formule:

$$H_0 = \frac{p_0^2}{2m_0} + U(r)$$

$$H_1 = \frac{\hbar^2(k - k_0)^2}{2m_0}$$

$$H_{totale} = H_0 + H_1$$

Nel caso di gap diretto l’hamiltoniana di un elettrone in BC perturbato è data dalla somma fra l’hamiltoniana priva di perturbazione H_0 , l’hamiltoniana H_1 che descrive la perturbazione parabolica subita dall’elettrone e l’hamiltoniana H_2 , la quale costituisce un termine cinetico aggiuntivo la cui dipendenza dal vettore d’onda k è lineare. H_2 è chiamata anche hamiltoniana di “warping” (“deformazione”), poiché sommandosi ad H_0 ed H_1 produce, sul diagramma $E - k$, un’accentuazione della concavità della BC, consentendo a molti semiconduttori a gap diretto (GaAs, InAs, InP ecc...) di avere elettroni in BC più leggeri (più reattivi agli stimoli elettrostatici imposti dall’esterno, maggiore mobilità) rispetto agli stessi elettroni in semiconduttori a gap indiretto. Una buona parte dell’elettronica che si occupa di dispositivi HEMTs (“high electron mobility transistors” – transistori ad alta mobilità elettronica μ_n) è basata sulla realizzazione di eterostrutture ($Al_{0.3}Ga_{0.7}As$ tipo n/GaAs intrinseco, ad esempio) formate, in prevalenza, da semiconduttori a gap diretto, nei quali il “warping” della BC incrementa la conducibilità elettronica $\sigma_n = e\mu_n n$. In formule:

$$H_2 = \frac{\hbar}{m_0} (k - k_0) P_{k_0} = \frac{\hbar}{m_0} (k - k_0) \langle \sigma(r, k_0) | \hat{P} | \sigma(r, k_0) \rangle$$

$$H_{totale} = H_0 + H_1 + H_2$$

dove P_{k_0} è chiamato “momentum matrix element” (elemento matriciale della quantità di moto $p = \hbar k$) ed indica il valor medio della quantità di moto dell’elettrone localizzato presso lo stato individuato dal vettore d’onda $k = k_0$. In formule:

$$\langle \sigma(r, k_0) | \hat{P} | \sigma(r, k_0) \rangle = \int \sigma(r, k_0)^* [-i\hbar \nabla] \sigma(r, k_0) d^3r$$

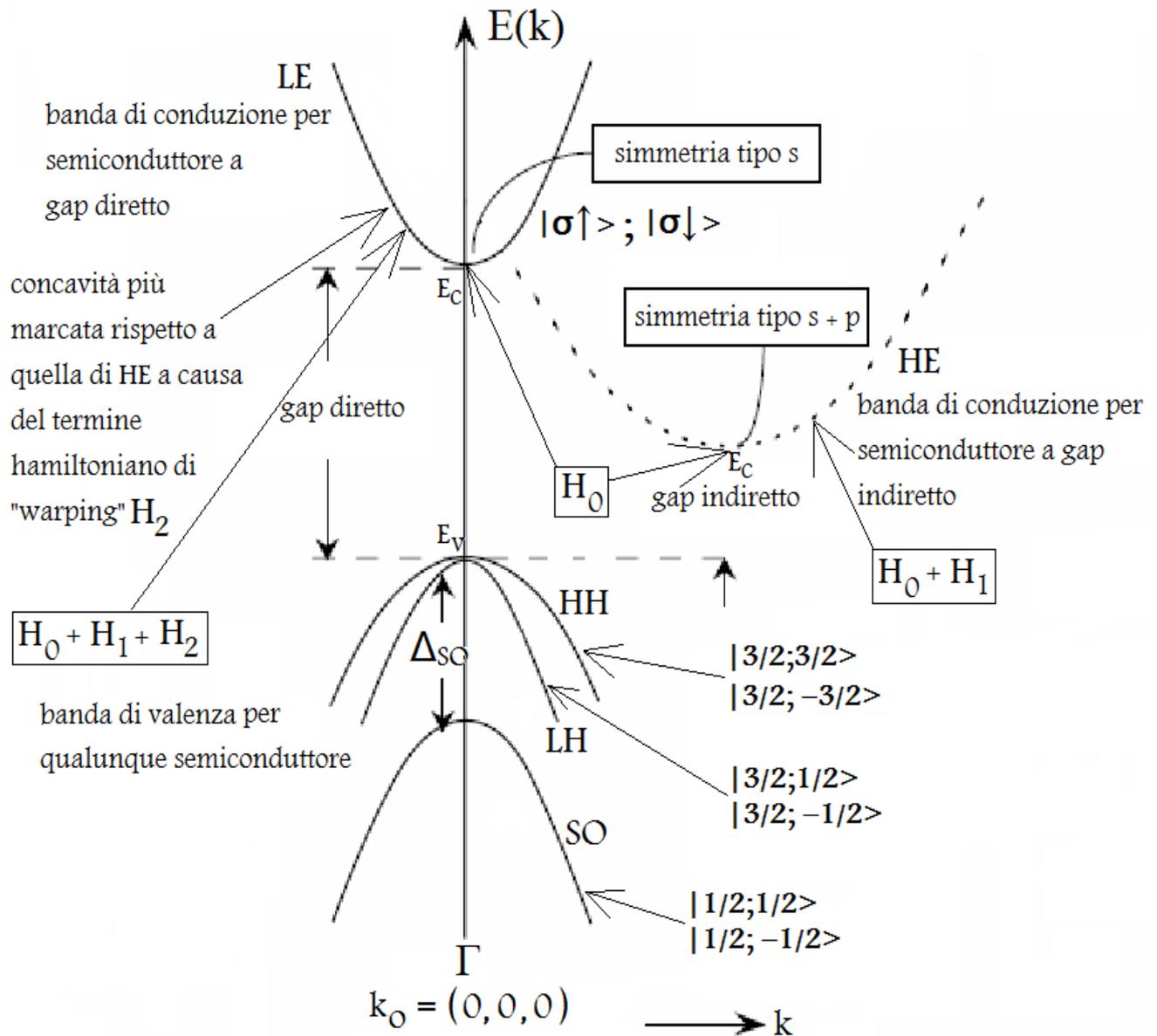


Fig. 14

Ingrandimento, molto qualitativo, della regione centrale del diagramma a bande $E - k$ riportato nella parte destra di figura 13, comprendente i "bandedges". A causa dell'interazione spin - orbita si verifica una separazione fra le sottobande di valenza HH ed LH ed uno "splittamento" della sottobanda di valenza SO di un'energia pari a Δ_{SO} . Queste alterazioni energetiche sono riportate graficamente in modo enfatizzato per ragioni di chiarezza. Inoltre è messa in risalto la curva $E - k$ della BC di un semiconduttore a gap diretto, chiamata LE, che è il risultato della degenerazione di un orbitale atomico di tipo s, e la curva $E - k$ della BC di un semiconduttore a gap indiretto, chiamata HE, che è il risultato della degenerazione di un orbitale atomico ibrido di tipo s + p. Si noti che presso il minimo E_c la concavità di LE è più marcata rispetto alla concavità di HE, a causa del termine cinetico aggiuntivo di "warping" ("deformazione") H_2 .

Osservando il diagramma a bande riportato in figura 12 possiamo notare una densità tridimensionale di stati in BC ed in BV, noti come "band tail states" a causa della forma a coda caratteristica di questa densità di stati presso E_c ed E_v , la quale viene indicata con $\Delta D_{dr}(E)$ ($eV^{-1}cm^{-3}$); $\Delta D_{dr}(E)$ si sovrappone alle densità tridimensionali di stati $D_{BC}(E)$ e $D_{BV}(E)$ teoricamente previste. Questi stati permessi sono localizzati sia all'interno delle bande BC e BV che all'interno del gap, dove non ci dovrebbero essere stati occupabili da elettroni o lacune, e si manifestano soprattutto in conseguenza di alti livelli di drogaggio. Tali stati "spuri" sono individuati dal posizionamento casuale e disordinato degli atomi droganti all'interno della matrice cristallina del semiconduttore. A causa della loro presenza il gap effettivo caratteristico di un particolare semiconduttore è decrementato, consentendo transizioni ottiche (anche in assorbimento, e quindi nell'ambito della fotorivelazione) teoricamente non possibili.

Anche l'energia di gap E_{gap} di un semiconduttore può subire fluttuazioni dovute a molteplici fattori, come ad esempio la concentrazione di impurezze casualmente distribuite nel reticolo cristallino, che abbiamo appena discusso, oppure la temperatura e la concentrazione di portatori di carica mobile. Se la temperatura varia anche la geometria cristallina del semiconduttore subisce dei cambiamenti, seppur poco significativi: ciò comporta la variazione delle caratteristiche geometriche, e di conseguenza di quelle elettriche, le quali parametrizzano il sistema di quattro equazioni (differenziali) a quattro incognite derivante dall'utilizzo del noto modello analitico di Kronig - Penney. Di conseguenza anche la funzione $F(E)$ dell'energia totale E degli elettroni, che deve soddisfare la condizione $-1 < F(E) < 1$ per poter localizzare livelli energetici E permessi (cioè stati permessi), varia con la temperatura, comportando fluttuazioni termiche sia del gap che della forma $E(k)$ delle bande. Invece se la temperatura del semiconduttore è molto elevata oppure nel caso in cui il segnale ottico incidente su un fotorecettore a semiconduttore è molto intenso (alta potenza ottica da rivelare), la concentrazione di portatori di carica mobile, ovvero di elettroni in BC (n) e di lacune in BV (p), può diventare grande, alterando leggermente la forma del diagramma a bande e lo spessore del gap. Infatti se per qualsivoglia causa n è elevata, un singolo elettrone in BC non può più essere visto come una particella quasi classica (approssimazione della statistica di Fermi - Dirac con quella di Boltzmann), ipotesi consentita se n è bassa, ovvero come una particella di fatto libera, quindi altamente delocalizzata nello spazio reale delle coordinate e altamente localizzata all'interno dello spazio reciproco dei vettori d'onda. A causa delle molte interazioni con gli altri elettroni in BC, abbiamo che l'elettrone preso in considerazione risulta altamente localizzato nello spazio reale delle coordinate e altamente delocalizzato all'interno dello spazio reciproco dei vettori d'onda; in altre parole non è più distinguibile dagli altri elettroni che si trovano in BC. L'intero insieme di N elettroni in BC deve essere descritto dalla seguente funzione d'onda antisimmetrica:

$$\Psi_{\text{antisim}}(\mathbf{r}, \mathbf{k}) = \frac{1}{\sqrt{N!}} \det \begin{pmatrix} \psi_{\alpha_1}^1 & \psi_{\alpha_1}^2 & \psi_{\alpha_1}^3 & \dots & \psi_{\alpha_1}^N \\ \psi_{\alpha_2}^1 & \psi_{\alpha_2}^2 & \vdots & & \psi_{\alpha_2}^N \\ \vdots & \dots & & & \psi_{\alpha_N}^N \end{pmatrix}$$

Nella prima riga del ben noto "determinante di Slater" sono allocate tutte e sole le funzioni ψ_1 , nella seconda riga sono allocate tutte e sole le funzioni ψ_2 , nella terza riga tutte e sole le funzioni ψ_3 e così via. Nella prima colonna del determinante di Slater sono allocati tutti e soli gli insiemi di numeri quantici α_1 , nella seconda colonna sono allocati tutti e soli gli insiemi di numeri quantici α_2 , nella terza colonna tutti e soli gli α_3 e così via. Nell'espressione della funzione d'onda antisimmetrica gli $\{\psi_i\} \forall i \in [1, N]$ rappresentano le funzioni d'onda degli N elettroni (in BC) del sistema considerati distinguibili, non interagenti, cioè trattabili classicamente, mentre gli $\{\alpha_i\} \forall i \in [1, N]$ rappresentano gli N insiemi di numeri quantici n, l, m, s ($\alpha_i = n_i, l_i, m_i, s_i$). In accordo con il principio di esclusione di Pauli, il determinante di Slater vale zero nel caso in cui due elettroni qualsiasi appartenenti al sistema (sistema = BC) condividano tutti e quattro i numeri quantici (ad esempio se α_1 fosse uguale ad α_2 le prime due righe sarebbero uguali), rendendo così nulla la funzione d'onda antisimmetrica che descrive il sistema (nonché la sua densità di probabilità), dunque rendendo impossibile una simile eventualità. Maggiore è la concentrazione n di elettroni in BC, maggiore è la schermatura che questi esercitano sul potenziale prodotto dagli atomi (=ioni positivi) del cristallo, minore è l'interazione coulombiana tra ciascuno degli N elettroni e gli atomi del semiconduttore, minore è l'energia potenziale di ciascuno degli N elettroni e quindi minore è l'energia totale E degli stessi. Un ragionamento analogo può essere fatto per un sistema di lacune in BV. Pertanto abbiamo un decremento di E_c ed un incremento di E_v , e quindi un decremento di E_{gap} . Per il silicio drogato n la relazione empirica che consente di calcolare il decremento dell'energia di gap dovuto sia alle fluttuazioni termiche che al livello di drogaggio, ovvero alla concentrazione n di elettroni in BC, è la seguente:

$$\Delta E_{\text{gap}} = - 22.5 \left(\frac{N_d}{10^{18}} \frac{300}{T(K)} \right)^{\frac{1}{2}} \text{ meV}$$

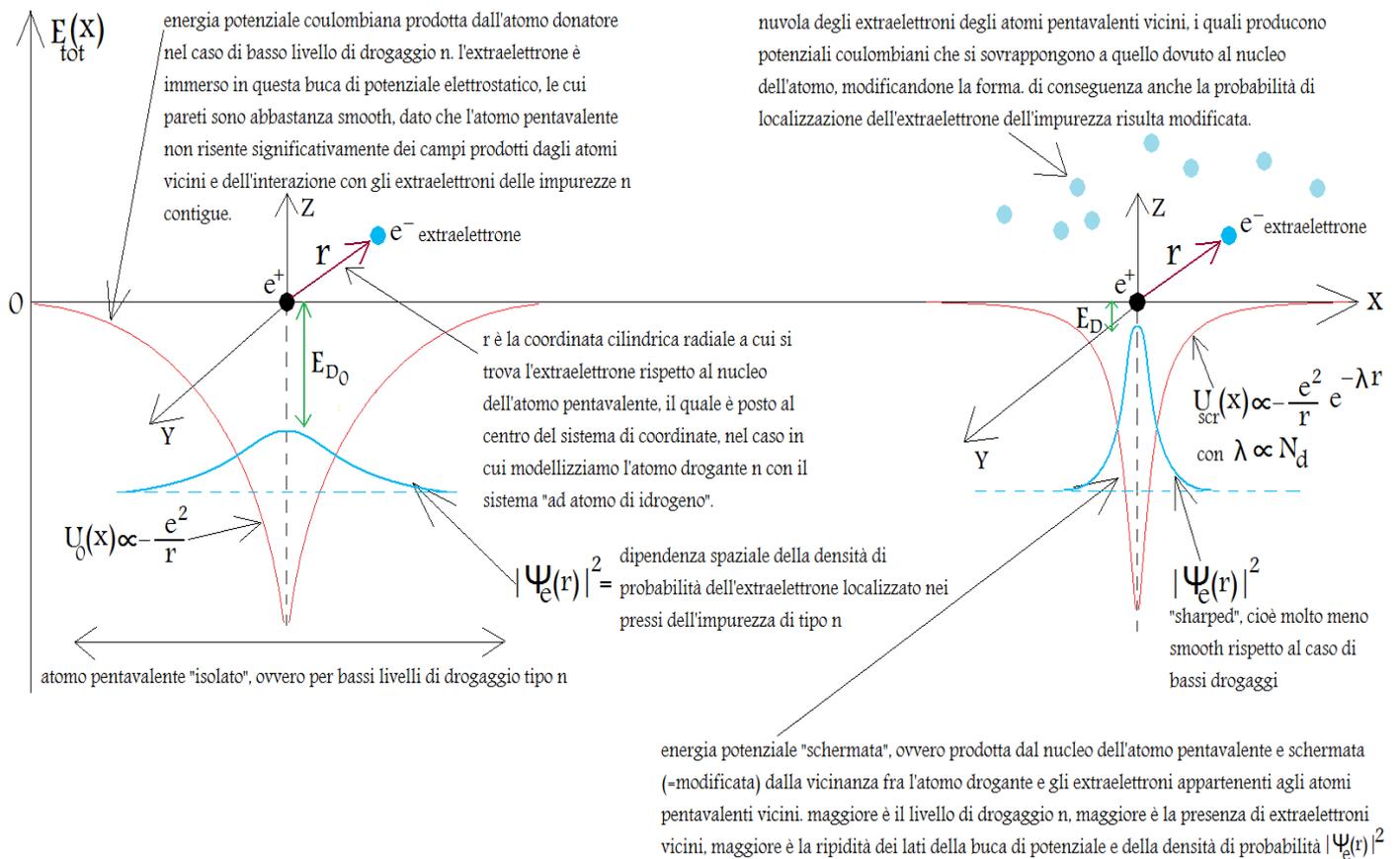
Per il GaAs (tipo bulk) è stata trovata la seguente relazione empirica per calcolare il decremento di E_{gap} dovuto alla concentrazione n di elettroni in BC e a quella p di lacune in BV (e indirettamente alle fluttuazioni termiche):

$$E_{\text{gap}} = E_{\text{gap}0} - 1.6 \times 10^{-8} \left(n^{\frac{1}{3}} + p^{\frac{1}{3}} \right) \text{ eV}$$

dove $E_{\text{gap}0}$ è il valore dell'energia di gap per il GaAs (tipo bulk), che è di circa 1.51 eV a $T = 0$ K e di circa 1.43 eV a $T = 300$ K.

Sempre riferendoci al diagramma a bande mostrato in figura 12 possiamo osservare che gli elettroni ψ_e in BC, all'interno del lato n dell'omogiunzione, appartenenti, dal punto di vista energetico, alla regione tratteggiata (nel disegno) della concentrazione $n(E) = 2D_{BC}(E - E_C)f(E)$, vedono lungo il verso negativo dell'asse X una barriera di energia potenziale $E_C(X)$ ben approssimabile con un gradino di potenziale (la ben nota funzione di "heavy side" $\sigma(X)$) di altezza qV_0 e di estensione spaziale infinita. Pertanto la densità superficiale di corrente di probabilità di particella $j_e(x)$ (la cui espressione è riportata nella [e4]), associata ad un qualunque stato elettronico ψ_e di energia E compresa fra E_C ed $E_C + qV_0$, è nulla; in altre parole ψ_e vede un sistema "chiuso" dalla barriera qV_0 , la quale impedisce qualunque corrente (densità superficiale di corrente) $qj_e(x)$ sostenuta da elettroni aventi energia cinetica minore di qV_0 . Ad un generico stato elettronico ψ_e , che si trova in BC, avente energia totale E , localizzato all'interno del lato n dell'omogiunzione e appartenente, dal punto di vista energetico, alla regione colorata di verde (nel disegno) della concentrazione $n(E)$, è invece associata una densità di corrente di probabilità di particella $j_e(x)$ non nulla, la quale viene perfettamente bilanciata da una corrente di particella $-j_e(x)$ associata ad uno stato elettronico ψ'_e , in BC, avente energia totale $E' = E$, il quale si trova all'interno del lato p dell'omogiunzione; abbiamo così formalizzato il bilanciamento delle due correnti $\pm qj_e(x)$ di elettroni maggioritari e minoritari, colorate di verde nella figura 12. Un discorso del tutto analogo può essere fatto per spiegare formalmente l'annullamento delle due correnti $\pm qj_h(x)$ di lacune maggioritarie e minoritarie, colorate di giallo nella figura 12.

Considerando sempre il diagramma a bande riportato in figura 12 va notato che sia E_a che E_d sono trattabili come due singoli livelli energetici (livelli discreti) solo per drogaggi non troppo alti: per drogaggi elevati (intorno a 10^{20} cm^{-3} , come ordine di grandezza) tali livelli degenerano, in realtà, in due bande di energia ΔE_a e ΔE_d , note come "impurity bands", ossia bande create dall'interazione reciproca fra impurezze (impurezze droganti, in questo caso, ovvero impurezze volute): infatti ciascuna lacuna e ciascun elettrone estrinseci risentono, oltre che del campo elettrico prodotto dal nucleo dell'atomo drogante a cui appartengono, anche dei campi elettrici prodotti dai nuclei degli atomi droganti vicini (primi vicini, secondi vicini e così via, man mano che consideriamo drogaggi sempre più elevati), con conseguente degenerazione dei livelli energetici discreti in bande. Si osservi in proposito il disegno qualitativo riportato in figura 15. Esiste un effetto di "screening" (schermatura elettrostatica) dell'energia potenziale coulombiana $U_0(r)$ di ciascun atomo drogante donatore, nell'ipotesi di drogaggio n elevato: tale schermatura è dovuta agli elettroni estrinseci, detti anche "extraelettroni", appartenenti agli atomi droganti pentavalenti vicini. Nel caso di alta concentrazione N_D di donatori gli extraelettroni formano una sorta di "nuvola elettronica" nei pressi di ciascun donatore, la quale produce una deformazione del potenziale $U_0(r)$, con conseguente deformazione della dipendenza spaziale della densità di probabilità $|\psi_e(r)|^2$ relativa all'extraelettrone fornito dal donatore. Tale deformazione fa assumere al potenziale un andamento spaziale $U_{scr}(r)$ tanto più "sharped" (ripido) quanto maggiore è il livello di drogaggio n. Conseguentemente anche $|\psi_e(r)|^2$ assume una forma più ripida, pertanto l'extraelettrone risulta più localizzato intorno all'impurezza pentavalente rispetto al caso di basso drogaggio n, quindi la probabilità di estrarlo dall'atomo donatore e promuoverlo in BC risulta maggiore. Ciò comporta, intuitivamente, una diminuzione dell'energia di ionizzazione E_D degli atomi donatori, così come una diminuzione dell'energia di ionizzazione E_A degli atomi accettori, nel caso di alti drogaggi.



livello di drogaggio N_D basso

livello di drogaggio N_D alto

Fig. 15

A sinistra è mostrata la dipendenza spaziale dell'energia potenziale coulombiana $U_0(r)$ di un singolo atomo donatore, nel caso di drogaggio n non elevato, con la densità di probabilità $|\psi_e(r)|^2$ dell'extraelettrone fornito dall'impurezza pentavalente e l'energia di ionizzazione E_{D0} degli atomi donatori. A destra è mostrato, concettualmente, il fenomeno di "screening" (schermatura elettrostatica) dell'energia potenziale coulombiana $U_0(r)$, nel caso di drogaggio n elevato, dovuto agli elettroni estrinseci appartenenti agli atomi droganti pentavalenti vicini. Si noti il potenziale $U_{scr}(r)$, che è tanto più deformato rispetto a $U_0(r)$ quanto maggiore è il livello di drogaggio N_D , nonché una localizzazione dell'extraelettrone più ristretta intorno all'atomo drogante di appartenenza, oltre alla diminuzione dell'energia di ionizzazione E_D degli atomi donatori.

La degenerazione di E_a ed E_d nelle rispettive "impurity bands" ΔE_a e ΔE_d comporta l'avvicinamento di E_{FV} (livello di Fermi nel Si tipo p) ad E_V , nel bulk p, e di E_{Fn} (livello di Fermi nel Si tipo n) ad E_C , nel bulk n, vale a dire una traslazione del livello di Fermi n verso energie E maggiori ed una traslazione analoga del livello di Fermi p verso energie E minori: in tal modo presso tutta la "impurity band" ΔE_d abbiamo $f(E) = 1$ per $T = 0$ K, pertanto nessun elettrone estrinseco in banda di conduzione, e presso tutta la "impurity band" ΔE_a abbiamo $f(E) = 0$ per $T = 0$ K, per cui nessuna lacuna estrinseca in banda di valenza.

Per temperature ordinarie l'espressione approssimata di E_{Fn} e di E_{FV} , in funzione di temperatura e drogaggi, è data dalla relazione di Joyce - Dixon:

$$E_{Fn} = E_C + kT \left[\ln \left(\frac{n}{N_C} \right) + \frac{1}{\sqrt{8}} \left(\frac{n}{N_C} \right) \right] \quad E_{FV} = E_V - kT \left[\ln \left(\frac{p}{N_V} \right) + \frac{1}{\sqrt{8}} \left(\frac{p}{N_V} \right) \right] \quad [e13]$$

dove k è costante di Boltzmann, T è la temperatura assoluta, n e p sono, rispettivamente, la concentrazione di elettroni in banda di conduzione BC e la concentrazione di lacune in banda di valenza BV. N_C ed N_V (cm^{-3}) rappresentano le ben note "densità di stati equivalenti", rispettivamente in BC ed in BV, e sono definite nel modo seguente:

$$N_C = 2 \left(\frac{2\pi kT m_{ed}^*}{h^2} \right)^{\frac{3}{2}} \quad N_V = 2 \left(\frac{2\pi kT m_{hd}^*}{h^2} \right)^{\frac{3}{2}}$$

Nel silicio, a temperatura ambiente, N_C vale circa $2.8 \times 10^{19} \text{ cm}^{-3}$, mentre N_V circa $1.04 \times 10^{19} \text{ cm}^{-3}$. m_{ed}^* è la massa efficace dell'elettrone per la densità di stati in BC, già definita, che per il silicio vale:

$$m_{ed}^* = 6^{\frac{2}{3}} [m_{eL}^* (m_{eT}^*)^2]^{\frac{1}{3}} \approx 1.08 m_0$$

mentre m_{hd}^* è la massa efficace della lacuna per la densità di stati in BV. Nel silicio m_{hd}^* è circa $0.55 m_0$, quindi è inferiore alla massa efficace m_{ed}^* dell'elettrone per la densità di stati in BC, pertanto la densità di stati in BC è più alta rispetto alla densità di stati in BV; in altri termini nel silicio una lacuna in BV avente energia cinetica E_{cin} ha a disposizione un numero di stati occupabili minore del numero di stati permessi a disposizione di un elettrone, in BC, che possiede la stessa energia cinetica E_{cin} . All'aumentare della temperatura questa asimmetria fra $D_{BC}(E)$ e $D_{BV}(E)$ costringe il livello di Fermi intrinseco E_{Fi} del silicio, di cui ricordiamo l'espressione nella [e14], ad una traslazione verso energie elettroniche E inferiori, al fine di garantire l'eguaglianza fra la concentrazione intrinseca n degli elettroni in BC e la concentrazione intrinseca p delle lacune in BV.

$$E_{Fi} = \frac{E_C + E_V}{2} + \frac{3}{4} kT \ln \left(\frac{m_{hd}^*}{m_{ed}^*} \right) \quad [e14]$$

In alcuni semiconduttori a gap diretto, ad esempio il GaAs oppure il InAs, si verifica la situazione opposta; $m_{ed}^* = m_e^* = m_\sigma^*$ è circa $0.067 m_0$ nella valle Γ di $E(k)$ di GaAs e circa $0.027 m_0$ nella valle Γ di $E(k)$ di InAs, mentre m_{hd}^* è circa $0.4 m_0$ per entrambi i semiconduttori. In entrambi $m_{hd}^* > m_{ed}^*$, per cui la densità di stati in BV è più alta rispetto alla densità di stati in BC. All'aumentare della temperatura l'asimmetria fra $D_{BC}(E)$ e $D_{BV}(E)$ costringe il livello di Fermi intrinseco E_{Fi} , in entrambi i semiconduttori a gap diretto (l'espressione è sempre quella in [e14]), ad una traslazione verso energie elettroniche E superiori, al fine di garantire l'eguaglianza fra la concentrazione intrinseca n degli elettroni in BC e la concentrazione intrinseca p delle lacune in BV.

Utilizzando il metodo k-p per il calcolo delle bande di energia $E(k)$ nei pressi dei "bandedges", e facendo ricorso all'approssimazione di dispersione parabolica, è possibile ottenere l'espressione analitica [e15] della massa efficace dell'elettrone m_e^* in funzione dell'energia di gap E_{gap} , cioè dell'energia di "bandgap" relativa alla valle Γ di $E(k)$ ($E_{gap\Gamma} = E_{gap}$), per qualunque semiconduttore a gap diretto avente una struttura cristallina a zincoblend. Come già spiegato in precedenza, nei semiconduttori a gap diretto la massa efficace dell'elettrone m_e^* ha un'unica definizione, ovvero $m_e^* = m_{ed}^* = m_\sigma^*$.

$$\frac{m_e^*}{m_0} = \left\{ 1 + \frac{\left[\frac{2P_{CV}^2}{m_0} \right]}{3} \left(\frac{2}{E_{gap}} + \frac{1}{E_{gap} + \Delta_{SO}} \right) \right\}^{-1} \quad [e15]$$

Nella [e15] m_e^* è presentata normalizzata rispetto alla massa a riposo dell'elettrone. Δ_{SO} è la già definita "energia di split - off" dovuta all'interazione spin - orbita. P_{CV} è un "momentum matrix element" (elemento matriciale della quantità di moto $p = \hbar k$) e rappresenta un fattore (un prodotto scalare) proporzionale all'interazione fra uno stato elettronico di tipo σ in BC e gli stati elettronici di tipo p in BV. P_{CV} implementa matematicamente il fenomeno fisico secondo il quale la massa efficace di un elettrone in BC, ovvero la sua inerzia nei confronti degli stimoli elettrici imposti dall'esterno (tensioni di polarizzazione del dispositivo), è legata non solo all'interazione fra l'elettrone di conduzione ed il campo periodico prodotto dagli atomi del reticolo cristallino, ma anche all'interazione fra l'elettrone e le nubi elettroniche relative agli stati in BV, vale a dire gli orbitali di tipo p . L'espressione di P_{CV} è la seguente:

$$P_{CV} = \langle \psi_{BC}(r, k_0) | \hat{P} | \psi_{BV}(r, k_0) \rangle = \langle \sigma(r, k_0) | \hat{P} | p(r, k_0) \rangle = \int \sigma(r, k_0)^* [-i\hbar \nabla] p(r, k_0) d^3 r$$

dove k_0 è il vettore d'onda relativo al punto di simmetria Γ . La quantità compresa fra parentesi quadre, nella [e15], spesso è identificata con E_p ed il suo valore numerico si trova tabulato per vari semiconduttori utilizzati in applicazioni fotoniche. Riportiamo alcuni di questi valori: $E_p(\text{GaAs}) = 25.7 \text{ eV}$, $E_p(\text{InP}) = 20.4 \text{ eV}$, $E_p(\text{InAs}) = 22.2 \text{ eV}$, $E_p(\text{CdTe}) = 20.7 \text{ eV}$. Dalla [e15] risulta evidente la proporzionalità diretta, di fatto lineare secondo il modello teorico utilizzato, fra E_{gap} e m_e^* : tale proporzionalità è confermata qualitativamente dalle evidenze sperimentali, come mostrato nel grafico riportato in figura 16.

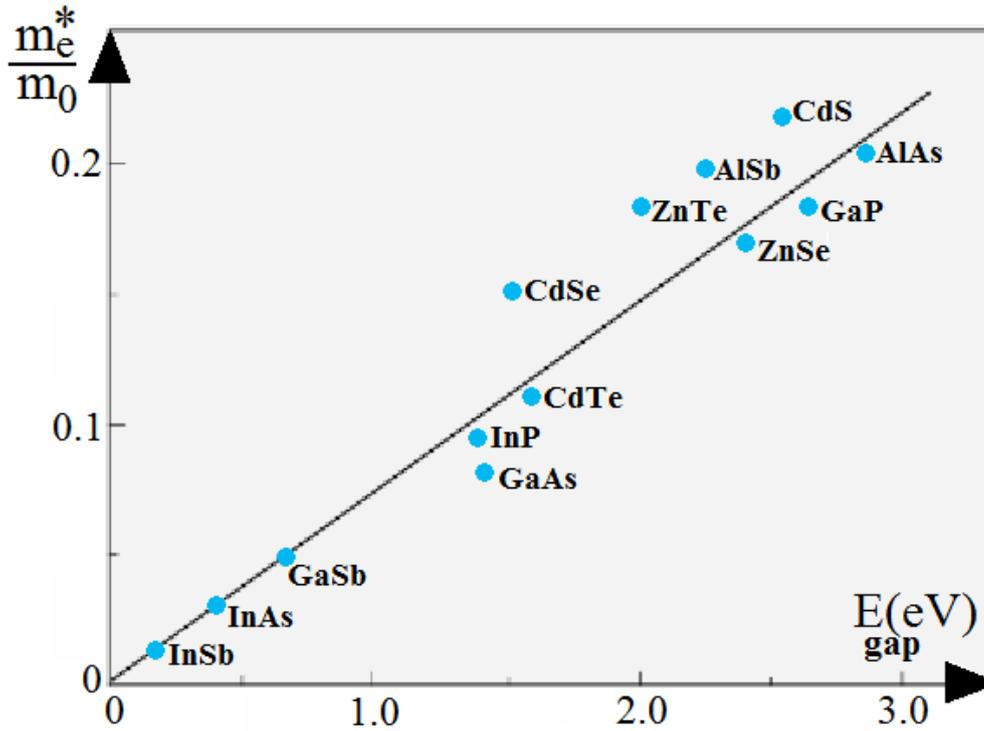


Fig. 16

Grafico sperimentale della dipendenza, approssimativamente lineare, della massa efficace dell'elettrone m_e^* dall'energia di gap E_{gap} ($E_{\text{gap}} = E_{\text{gap}(r)}$), per vari semiconduttori binari di interesse fotonico, a gap diretto, di tipo III - V, aventi struttura cristallina di tipo zincoblenda. L'evidenza sperimentale conferma, qualitativamente, la relazione fra m_e^* ed E_{gap} che si vince dall'espressione [e15].

Per temperature simili a quella ambiente la distribuzione di Fermi - Dirac in BC, nel Si tipo n, è approssimabile con quella classica di Boltzmann, cioè la coda della Fermi - Dirac presso le energie elettroniche E superiori ad E_c è ben approssimabile con una distribuzione ad esponenziale decrescente con E ; in altre parole ci sono pochissimi elettroni in BC e quasi tutti gli stati in BC sono vuoti ($n \ll N_c$), quindi gli elettroni sono quasi non interagenti fra loro, comportandosi come se fossero particelle classiche. Dualmente per temperature simili a quella ambiente, nel Si tipo p, la coda della funzione complementare di $f(E)$, ossia $1 - f(E)$, ovvero la distribuzione di probabilità di occupazione di stati in BV per le lacune, è approssimabile, per energie elettroniche E inferiori ad E_v , con la distribuzione di Boltzmann ad esponenziale decrescente con E_h ; in altre parole ci sono pochissime lacune in BV e quasi tutti gli stati in BV sono vuoti ($p \ll N_v$), pertanto le lacune sono quasi non interagenti fra loro, comportandosi come particelle classiche. Quindi per temperature e drogaggi ordinari la [e13] fornisce le relazioni approssimate, riportate in [e16], fra le concentrazioni n e p di portatori di carica mobili ed i livelli di Fermi E_{F_n} ed E_{F_p} , rispettivamente nel caso di semiconduttore di tipo n e di tipo p.

$$E_{F_n} \approx E_c + kT \ln\left(\frac{n}{N_c}\right) \quad E_{F_p} \approx E_v - kT \ln\left(\frac{p}{N_v}\right) \quad [e16]$$

Le relazioni fra n ed E_{F_n} e fra p ed E_{F_p} , mostrate nella [e16], potrebbero essere ricavate più rigorosamente (in modo meno approssimato), a temperatura e drogaggi fissati, per via numerica attraverso la [e17] e la [e18], oppure in modo iterativo trovando il livello di Fermi che soddisfa l'equazione della neutralità di carica.

Nel caso di drogaggi molto alti ($> 10^{20} \text{ cm}^{-3}$) ΔE_a può sovrapporsi alla banda di valenza nel bulk p, mentre ΔE_d può sovrapporsi alla banda di conduzione nel bulk n: in tal caso E_{F_p} è posizionata internamente alla BV nel bulk p, mentre E_{F_n} è posizionata internamente alla BC nel bulk n. Un simile dispositivo p^+/n^+ è noto come "giunzione degenera". Il lato n della giunzione degenera si comporta, di fatto, come un metallo (un conduttore, un materiale disparivalente), poichè possiede portatori di carica mobili, e quindi conduce corrente, persino a $T = 0 \text{ K}$.

La concentrazione n degli elettroni (maggioritari) di conduzione, cioè degli elettroni in BC, all'interno del bulk del Si tipo n, è data da:

$$n = \int_{E_C}^{\infty} 2D_{BC}(E - E_C)f(E)dE = N_C e^{-\frac{E_C - E_F}{kT}} \approx N_d \quad [e17]$$

Il 2 è aggiunto per tenere in considerazione la “degenerazione di spin”, che di fatto raddoppia il numero di stati, in BC, popolabili dagli elettroni. Infatti il ben noto esperimento di Stern e Gerlach ha messo in evidenza che l’elettrone possiede un “momento magnetico intrinseco” $\vec{\mu}$, detto anche “momento magnetico di spin”, la cui terza componente può avere modulo pari a $\mu_Z = \mu_B$ oppure $\mu_Z = -\mu_B$, e quindi che l’elettrone possiede un “momento angolare intrinseco” \vec{S} , detto anche “momento angolare di spin”, la cui terza componente può avere modulo pari a $S_Z = -\frac{1}{2}\hbar$, se $\mu_Z = \mu_B$, oppure $S_Z = +\frac{1}{2}\hbar$, se $\mu_Z = -\mu_B$. Quindi un singolo stato, individuato nello spazio reciproco (spazio di Fourier) dei numeri d’onda $k_x k_y k_z$, è occupabile da due elettroni aventi spin antiparalleli, ovvero da un elettrone caratterizzato da $\mu_Z = \mu_B$ e $S_Z = -\frac{1}{2}\hbar$, cioè caratterizzato dal numero quantico di spin $s = -\frac{1}{2}$, e da un elettrone caratterizzato da $\mu_Z = -\mu_B$ e $S_Z = +\frac{1}{2}\hbar$, cioè caratterizzato dal numero quantico di spin $s = +\frac{1}{2}$.

È possibile generalizzare la relazione [e16] scrivendone una valida in un qualunque punto della giunzione, purchè in condizioni di equilibrio, ovvero purchè il livello di Fermi E_F sia unico in tutta la giunzione:

$$n(x) = N_C e^{-\frac{E_C(x) - E_F}{kT}}$$

L’equilibrio, espresso dalle due relazioni riportate nella [e5], si instaura, come già spiegato, nel momento in cui la porzione di n con energia totale maggiore del livello E_C del bulk del lato p (nel disegno di figura 12 è colorata di verde), diventa uguale alla concentrazione complessiva degli elettroni di conduzione (minoritari) nel bulk del lato p (concentrazione sempre evidenziata in verde). In questa situazione il flusso netto di elettroni attraverso la giunzione è nullo ed i maggioritari vedono una barriera di energia potenziale pari a qV_0 , dove V_0 è il potenziale di contatto, detto anche (impropriamente) “potenziale chimico”, che tipicamente vale 0.6 – 0.7 V. V_0 è facilmente calcolabile considerando una delle due relazioni riportate nella [e5] e tenendo presente che:

$$\varepsilon(x) = -\frac{dV(x)}{dx}$$

oltre alla nota relazione di Einstein:

$$\frac{D_n}{\mu_n} = \frac{D_p}{\mu_p} = \frac{kT}{q} = V_T$$

dove V_T è nota “tensione equivalente alla temperatura”. Considerazioni equivalenti sono possibili per quanto riguarda l’equilibrio del flusso netto di lacune, in prossimità della giunzione. Infatti la concentrazione di lacune in banda di valenza nel bulk del Si tipo p è:

$$p = \int_{E_V}^{\infty} 2D_{BV}(E_h - E_V)[1 - f(E)]dE = N_V e^{-\frac{E_F - E_V}{kT}} \approx N_a \quad [e18]$$

dove $D_{BV}(E_h - E_V)$ è la densità tridimensionale degli stati in banda di valenza, la cui espressione è:

$$D_{BV}(E_h - E_V) = 2\pi \left(\frac{2m_{hd}^*}{h^2} \right)^{\frac{3}{2}} \sqrt{E_h - E_V}$$

È possibile generalizzare la [e18] scrivendo una relazione valida in un qualunque punto della giunzione, purchè in condizioni di equilibrio:

$$p(x) = N_V e^{-\frac{E_F - E_V(x)}{kT}}$$

L’equilibrio è tale per cui la porzione di p con energia E_h maggiore del livello E_V del bulk del lato n, ovvero quella colorata di giallo nella figura 12, è uguale alla concentrazione totale di lacune (minoritarie) in banda di valenza all’interno del bulk del lato n (anche questa concentrazione è sempre evidenziata in giallo). Le lacune maggioritarie vedono, dal lato p verso quello n, una barriera di potenziale alta qV_0 . Prendiamo in considerazione il grafico della densità di carica totale $\rho(x)$, di seguito riportato:

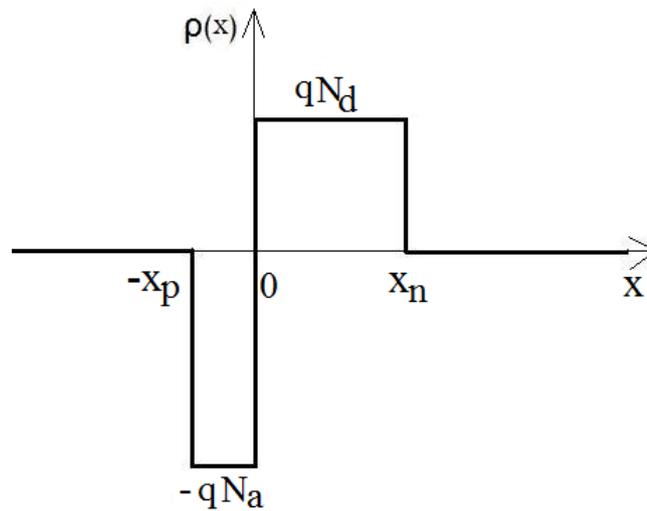


Fig. 17

Densità di carica totale $\rho(x)$.

Integrando una volta l'equazione di Poisson, la quale costituisce un caso particolare della più generica terza equazione di Maxwell (assumendo $\rho(x) = -qN_a$ fra $-x_p$ e 0, e $\rho(x) = qN_d$ fra 0 ed x_n):

$$\frac{d^2}{dx^2}V(x) = -\frac{\rho(x)}{\epsilon_S}$$

in cui ϵ_S è la costante dielettrica del semiconduttore utilizzato (11,8 ϵ_0 nel caso del Si), otteniamo l'andamento, lungo l'asse x , del campo elettrico $\epsilon(x)$. Per calcolare le due costanti, derivanti dall'operazione d'integrazione, basta imporre la condizione al contorno di campo nullo presso $-x_p$ ed x_n .

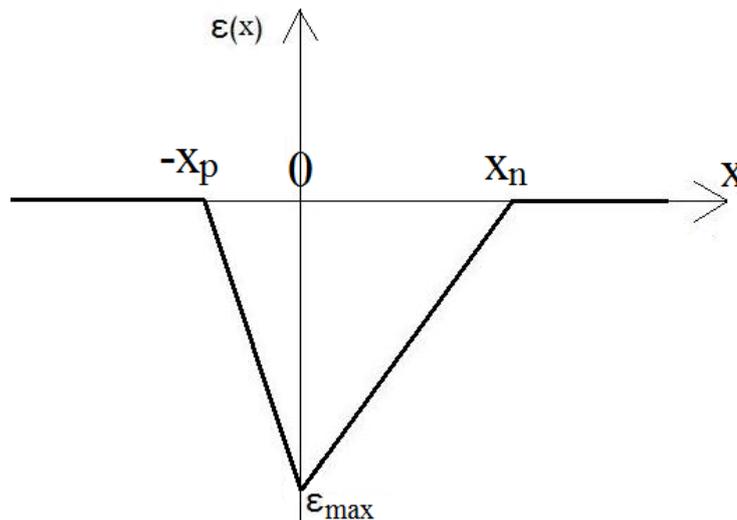


Fig. 18

Dipendenza spaziale del campo elettrico prodotto da donatori ed accettori ionizzati, in assenza di polarizzazione.

Otteniamo:

$$\varepsilon(x) = \begin{cases} \frac{-qN_a}{\varepsilon_S} x - \frac{qN_a x_p}{\varepsilon_S} & \text{per } -x_p < x < 0 \\ \frac{qN_d}{\varepsilon_S} x - \frac{qN_d x_n}{\varepsilon_S} & \text{per } 0 < x < x_n \end{cases}$$

Per la continuità del campo in $x = 0$ (dato che si tratta di un'omogiunzione) si ha che:

$$\varepsilon_{\max} = \frac{qN_d}{\varepsilon_S} x_n = \frac{qN_a}{\varepsilon_S} x_p$$

Integrando infine $\varepsilon(x)$ ricaviamo l'andamento del potenziale $V(x)$: nel nostro caso semplificato abbiamo due archi di parabola, uno con la concavità rivolta verso l'alto, l'altro verso il basso. Le due costanti di integrazione si calcolano imponendo, come condizioni al contorno, che $V(x)$ sia nullo presso $-x_p$ e pari a V_0 presso x_n :

$$V(x) = \begin{cases} \frac{qN_a}{2\varepsilon_S} (x + x_p)^2 & \text{per } -x_p < x < 0 \\ V_0 - \frac{qN_d}{2\varepsilon_S} (x - x_n)^2 & \text{per } 0 < x < x_n \end{cases}$$

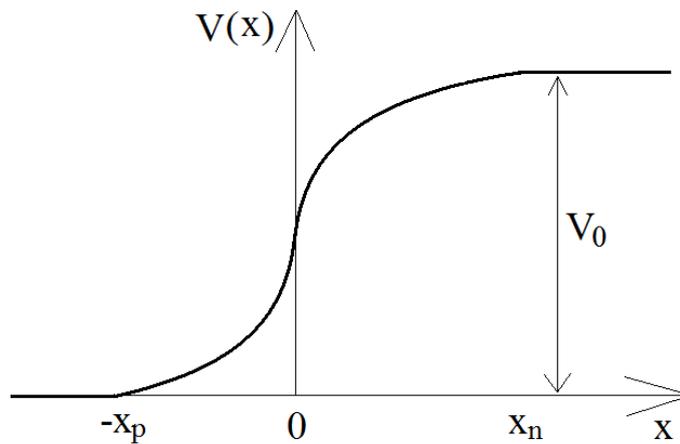


Fig. 19

Andamento spaziale del potenziale in assenza di polarizzazione.

4.2.1) La omogiunzione pn in condizione di polarizzazione inversa

Adesso polarizziamo la giunzione in maniera inversa, applicando ai terminali metallici del diodo una tensione $V < 0$. L'effetto su $V(x)$ è quello mostrato in figura 20.

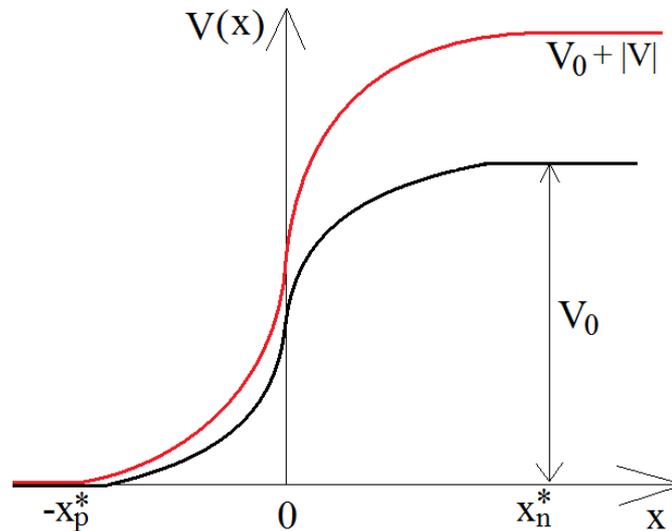


Fig. 20

Andamento spaziale (evidenziato in rosso) del potenziale in presenza di polarizzazione inversa.

I portatori maggioritari, all'atto dell'applicazione di una tensione inversa, vedono momentaneamente prevalere la componente di trascinamento, della corrente a loro associata, rispetto a quella di diffusione (il campo si è rafforzato), come si evince dalle due equazioni riportate nella [e5], ragion per cui le lacune vengono richiamate verso il bulk del Si tipo p, così come gli elettroni verso il bulk del Si tipo n. Mentre gli elettroni, nei pressi di x_n , dove sono maggioritari, sono trascinati verso il bulk n a causa del campo (x_n si sposta a destra, $x_n \rightarrow x_n^*$), nella stessa regione i minoritari, cioè le lacune, vengono estratte dal campo e trascinate attraverso la zds, finché non arrivano nel bulk p. In maniera duale mentre le lacune, nei pressi di $-x_p$, dove sono maggioritarie, vengono trascinate verso il bulk p a causa del campo ($-x_p$ si sposta a sinistra, $-x_p \rightarrow -x_p^*$), nella stessa regione i minoritari, cioè gli elettroni, vengono estratti dal campo e trascinati attraverso la zds, finché non arrivano nel bulk n. Ciò comporta un allargamento della zds ($zds \rightarrow zds^*$) ed un innalzamento della barriera di potenziale vista dai portatori maggioritari, nonché un'estrazione di minoritari, la quale provoca una corrente inversa di saturazione I_0 , generalmente dell'ordine del fA, nel caso di un'omogiunzione pn avente una superficie attiva S dell'ordine di $100 \times 100 \mu\text{m}^2$. La situazione dei profili delle concentrazioni dei portatori minoritari è mostrata in figura 21.

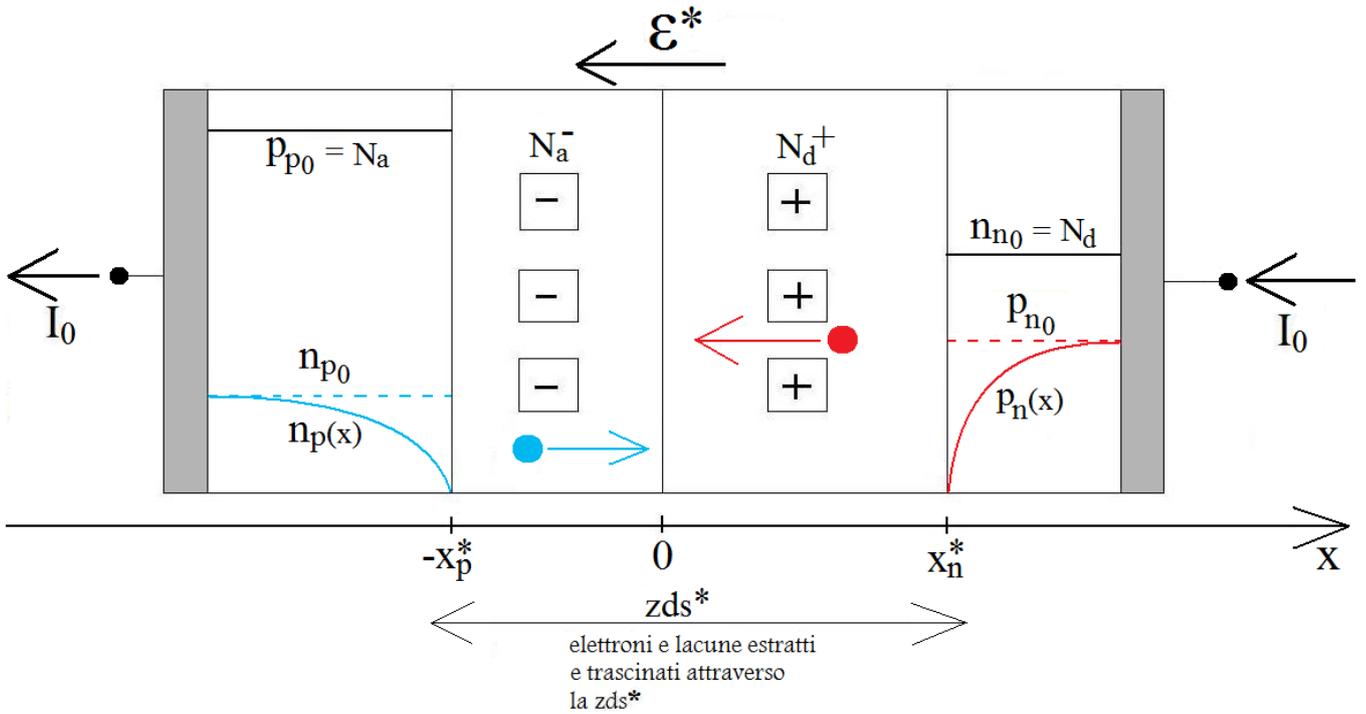


Fig. 21

Schema della giunzione pn polarizzata in inversa. È individuabile la zona di svuotamento zds^* , maggiore di quella (zds) che caratterizza la stessa giunzione priva di polarizzazione esterna. Sono presenti, inoltre, i profili delle concentrazioni dei portatori minoritari, la cui estrazione, causata dal campo elettrico, sostiene la corrente inversa di saturazione I_0 .

zds^* è la larghezza della zona di svuotamento incrementata dalla polarizzazione inversa. Nell'ipotesi di omogiunzione pn brusca l'espressione di zds^* è la seguente:

$$zds^* = \sqrt{\frac{2\epsilon_s}{q} \left(\frac{1}{N_a} + \frac{1}{N_d} \right) (V_0 + |V|)}$$

x_n^* e $-x_p^*$ sono gli estremi della zona di svuotamento allargata. In figura 22 è riportato l'andamento spaziale del campo $\epsilon(x)$, rafforzatosi a causa della polarizzazione inversa.

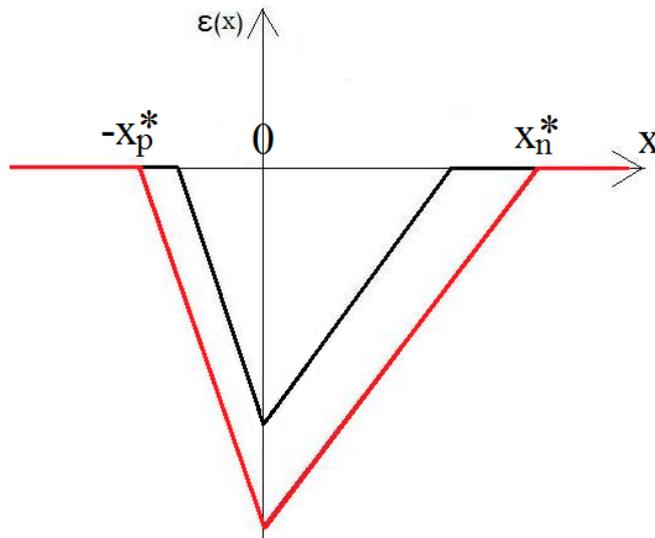


Fig. 22

Dipendenza spaziale del campo elettrico prodotto da donatori ed accettori ionizzati, in presenza di polarizzazione inversa.

Consideriamo l'estrazione delle lacune minoritarie: il nuovo campo (indicato, in figura 22, con la linea rossa) provoca uno svuotamento totale di lacune presso x_n^* , dove n_p diventa minore di n_i^2 . Si viene quindi a formare un considerevole gradiente di concentrazione di minoritari orientato lungo l'asse x , per cui, in osservanza alla legge della diffusione di Fick, s'instaura una diffusione orientata in modo opposto all'asse x , tale da attenuare la differenza di concentrazione di lacune. Questa diffusione servirebbe, in assenza di campo elettrico "svuotante", a ristabilire la legge dell'azione di massa presso x_n^* , ovvero a ristabilire l'equazione $n_p = n_i^2$. Tuttavia in una posizione contigua a x_n^* viene a mancare una frazione di concentrazione, che ha diffuso verso sinistra, per cui si forma un secondo gradiente (minore del primo), e quindi un'ulteriore diffusione verso sinistra. Iterando questo ragionamento per posizioni x contigue una all'altra, si arriva al bulk p , luogo in cui avvengono processi di generazione termica indispensabili al rifornimento di lacune; nel bulk p il profilo $p_n(x)$ assume un valore pari a p_{n0} . Considerando un volume di Si tipo n , interamente posizionato a destra di x_n^* , possiamo scrivere la ben nota "equazione di continuità" per le lacune minoritarie:

$$\frac{\partial p_n(x,t)}{\partial t} = -\frac{1}{q} \frac{\partial J_h(x,t)}{\partial x} + G - R \quad [e19]$$

Il primo membro è la velocità con cui la concentrazione di lacune varia nel volume, mentre nel secondo riconosciamo la variazione, lungo x , della densità di corrente totale di lacune e il rate di generazione netto ($G - R$) di lacune all'interno del volume. La corrente $J_h(x,t)$ è la somma dei due contributi di diffusione e di drift e può essere scritta nella forma:

$$J_h(x,t) = J_{hdiff}(x,t) + J_{hdrift}(x,t) = -qD_h \frac{dp_n(x,t)}{dx} + q\mu_h p_n(x,t)\varepsilon(x,t) \quad [e20]$$

Tenendo conto che:

- siamo a regime
- a destra di x_n^* il campo è nullo
- $G - R = -\frac{\delta p(x,t)}{\tau_h}$, dove δp è la concentrazione di lacune di eccesso (in questo caso eccesso negativo), mentre τ_h è il tempo di vita medio delle lacune in quella regione
- $p_n(x,t) = p_{n0} + \delta p(x,t)$

e sostituendo la [e20] nella [e19] si arriva ad una forma più semplice dell'equazione di continuità, qui di seguito riportata:

$$D_h \frac{d^2 \delta p(x)}{dx^2} = \frac{\delta p(x)}{\tau_h}$$

La soluzione fisicamente accettabile è:

$$\delta p(x) = \delta p(0) e^{-\frac{x}{L_h}}$$

dove L_h rappresenta la lunghezza di diffusione delle lacune, pari a $\sqrt{D_h \tau_h}$. La condizione al contorno dello svuotamento totale di lacune, presso x_n^* , impone che:

$$\delta p(0) = -p_{n0}$$

Possiamo concludere che:

$$p_n(x) = p_{n0} (1 - e^{-\frac{x}{L_h}}) \quad [e21]$$

Con considerazioni del tutto analoghe per gli elettroni svuotati nei pressi di $-x_p^*$ si ottiene:

$$n_p(x) = n_{p0} (1 - e^{-\frac{x}{L_n}}) \quad [e22]$$

dove L_n è la lunghezza di diffusione degli elettroni, pari a $\sqrt{D_n \tau_n}$. Sia nella [e21] che nella [e22] l'origine dell'asse x è presa solidale con la regione di estrazione di interesse, ovvero rispettivamente in x_n^* e $-x_p^*$. Valutando le

densità di corrente di diffusione ($J_{h\text{diff}}$ e $J_{n\text{diff}}$) in x_n^* e in $-x_p^*$ (completamente dipendenti dall'estrazione dei minoritari, e facili da ricavare, poichè si tratta di derivare, rispetto ad x , i profili [e21] ed [e22]), sommandole e moltiplicando il tutto per la superficie attiva S del diodo, otteniamo la corrente inversa di saturazione I_0 .

$$I_0 = qS \frac{D_h}{L_h} p_{no} + qS \frac{D_n}{L_n} n_{po}$$

Riportiamo la caratteristica (semiquantitativa) $I - V$ di una giunzione pn:

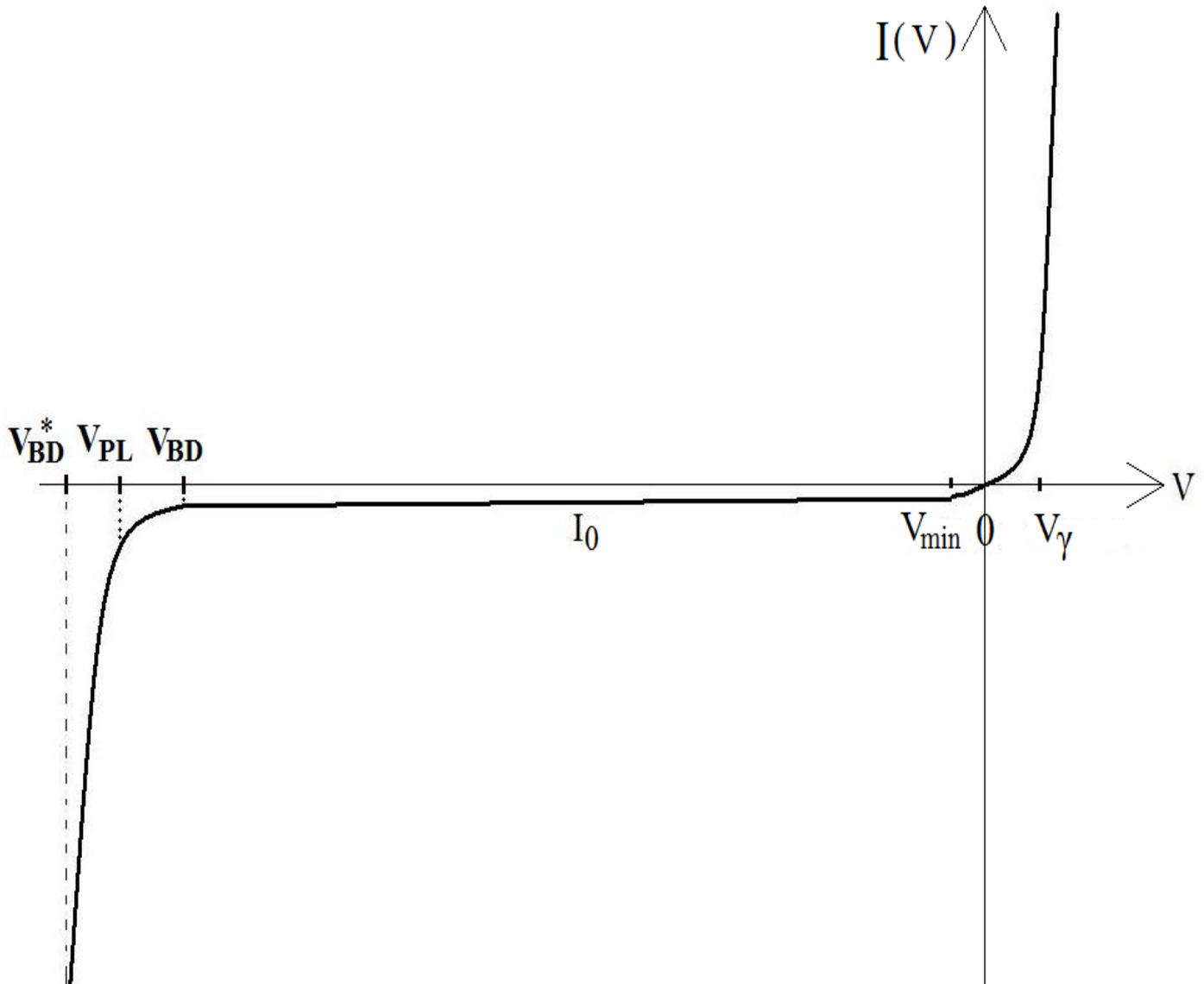


Fig. 23

Caratteristica $I - V$ semiquantitativa di una giunzione pn. V_γ è la tensione di polarizzazione diretta in corrispondenza della quale si crea un consistente flusso netto di lacune verso il lato n ed uno di elettroni verso il lato p. Questa iniezione di minoritari (permessa dall'abbassamento della barriera di potenziale vista dagli stessi) sostiene la corrente diretta. V_{\min} è la minima tensione di polarizzazione inversa necessaria all'instaurarsi dell'estrazione di minoritari. V_{BD} è la tensione superata la quale si instaura il fenomeno del breakdown a valanga (la probabilità di creare eventi ionizzanti mediante impattazione, da parte dei minoritari estratti, non è trascurabile). V_{PL} individua il punto di lavoro ottimo, presso cui il rapporto segnale/rumore della giunzione pn, utilizzata come fotorivelatore, è massimo. V_{BD}^* è la tensione asintotica presso cui la corrente inversa assume valori elevatissimi, essendo diventata una corrente di tunneling, a causa del fenomeno del breakdown Zener.

In polarizzazione inversa anche il diagramma a bande viene modificato, confermando le osservazioni precedenti:

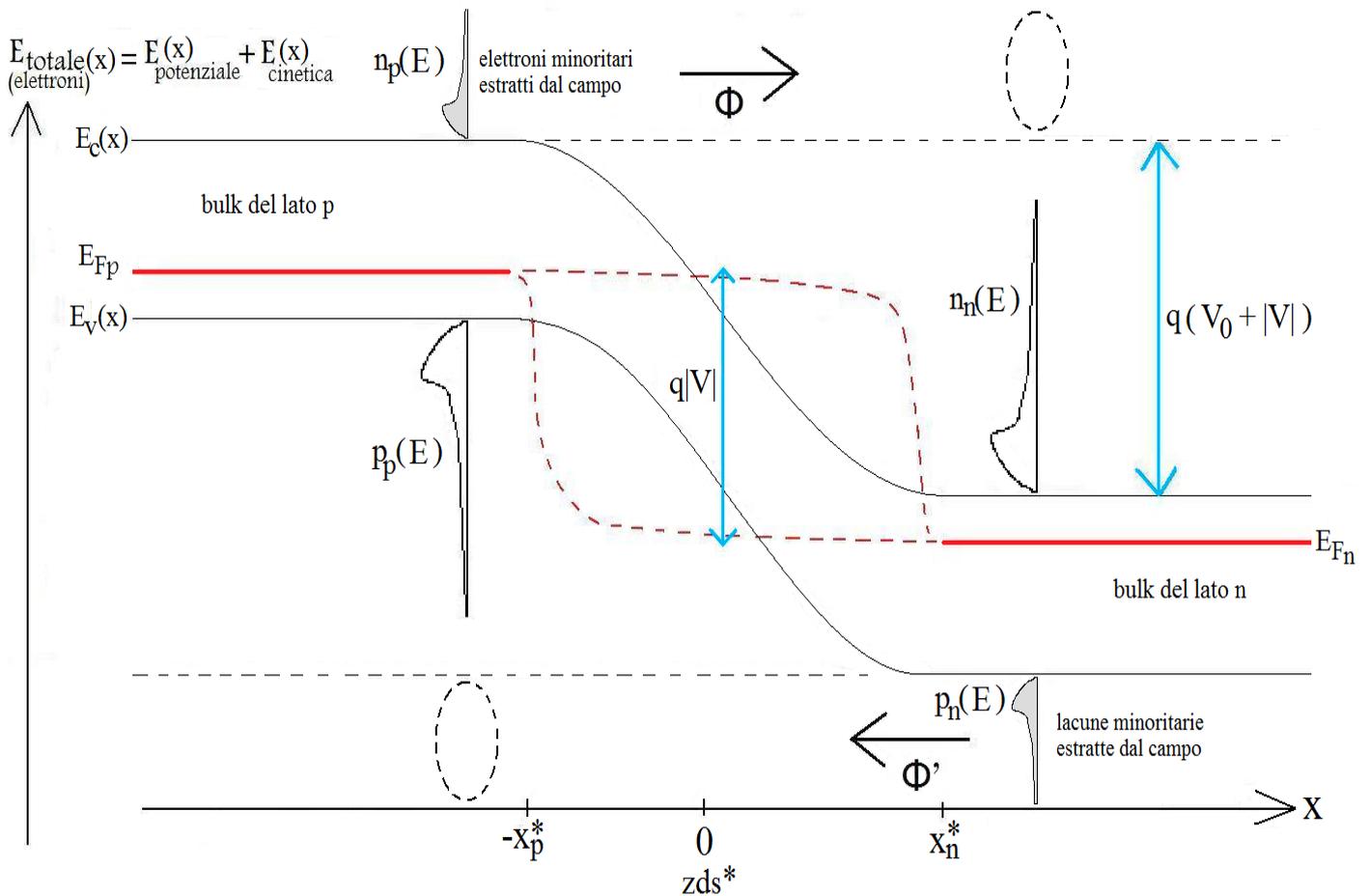


Fig. 24

Diagramma a bande dell'omogiunzione pn sottoposta a polarizzazione inversa, con $V_{\min} < |V| < V_{BD}$. Sono visibili i flussi di carica Φ e Φ' dovuti all'estrazione dei minoritari. È altresì evidente il disallineamento dell'energia di Fermi (la differenza fra il livello di Fermi nel bulk p ed il livello di Fermi nel bulk n è un'energia pari a $q|V|$) ed i quasi livelli di Fermi E_{Fn} ed E_{Fp} .

Se $|V|$ è maggiore di V_{\min} , il cui valore tipico è circa di 10mV (basta che la distribuzione $n_n(E)$ si trasli verso il basso di un'energia pari a qualche kT), allora $n_n(E)$ sottende, rispetto all'asse dell'energia totale degli elettroni, per $E > E_C$ del bulk p, una superficie praticamente nulla (l'ellisse tratteggiata in figura 24), per cui nessuno degli elettroni maggioritari presso x_n^* riesce a superare l'accresciuta barriera di potenziale. Gli elettroni della distribuzione $n_p(E)$ creano un flusso netto Φ , e quindi la loro estrazione contribuisce a sostenere I_0 (quest'ultima è dovuta anche all'estrazione delle lacune minoritarie di $p_n(E)$, che creano un flusso netto Φ'). Questa situazione si mantiene praticamente invariata (I_0 costante) per qualunque valore di $|V|$ maggiore di V_{\min} , fino ad un valore massimo pari a V_{BD} (in questo range Φ e Φ' rimangono praticamente invariati).

Nel contesto di disequilibrio elettrico appena descritto è comunque possibile calcolare i profili spaziali delle concentrazioni dei portatori di carica mobili nel modo consueto, ricorrendo ai "quasi livelli di Fermi E_{Fn} ed E_{Fp} ". Questi, come è noto, non hanno alcun significato fisico, bensì costituiscono dei meri strumenti matematici che rendono ragione della dipendenza spaziale delle concentrazioni di elettroni e lacune ($n(x)$ e $p(x)$) in presenza di un'alterazione delle condizioni di equilibrio, ad esempio dovuta ad una polarizzazione inversa. Dunque abbiamo:

$$n(x) = N_C e^{-\frac{E_C(x) - E_{Fn}(x)}{kT}} \quad p(x) = N_V e^{-\frac{E_{Fp}(x) - E_V(x)}{kT}}$$

4.2.2) I fenomeni di breakdown

Se la polarizzazione inversa è talmente alta che $|V|$ supera V_{BD} , si possono verificare due diversi fenomeni di "breakdown". Per tensioni poco superiori a V_{BD} si verifica il cosiddetto "breakdown a valanga" (sfruttabile, come vedremo, per moltiplicare correnti legate ad eventi luminosi); per tensioni ancora maggiori ($|V| \rightarrow V_{BD}^*$) si ha un breakdown assolutamente da evitare ai fini della fotorivelazione, noto come "breakdown Zener" (sfruttabile, in altri contesti, per stabilizzare la tensione che insiste ai terminali del diodo). Cominciamo a descrivere quest'ultimo, con riferimento al diagramma a bande di figura 25.

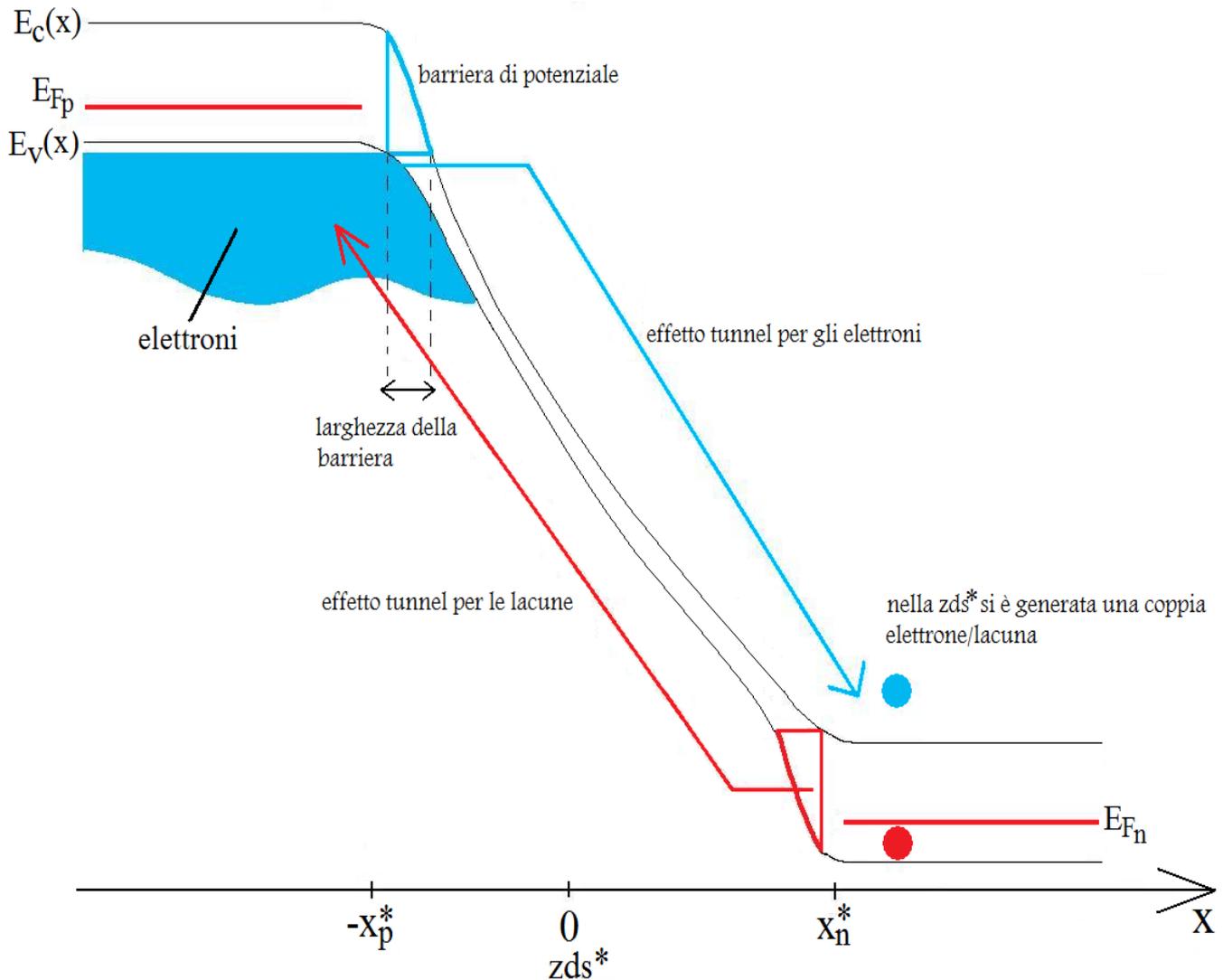


Fig. 25

Diagramma a bande della giunzione pn fortemente polarizzata in inversa, con $|V| \rightarrow V_{BD}^*$. È evidente come l'assottigliamento delle due barriere di potenziale provochi le correnti di tunneling degli elettroni e delle lacune.

Per tensione inversa prossima a V_{BD}^* , che è dell'ordine delle decine di V, si assiste ad un forte assottigliamento della barriera di potenziale vista dagli elettroni della banda di valenza del lato p, tale da permettere, per effetto tunnel, che questi attraversino in enorme quantità (rispetto agli elettroni estratti visti prima) la barriera e vadano nella banda di conduzione del lato n. Essendo nella banda di conduzione tali elettroni producono una corrente $\gg I_0$. Per giunta la loro presenza nella banda di conduzione del lato n implica la presenza di lacune nella banda di valenza. Queste vedono una barriera sottilissima, per cui anche loro, per effetto tunnel, la attraversano in gran numero, giungendo nella banda di valenza del lato p. Pertanto assistiamo ad una moltiplicazione

straordinaria della densità di corrente inversa, nota come “corrente di tunneling”, la quale sovrasterebbe una qualsiasi fotocorrente utile.

Adesso esponiamo il fenomeno del breakdown a valanga (“avalanche breakdown”), con riferimento alla figura 26.

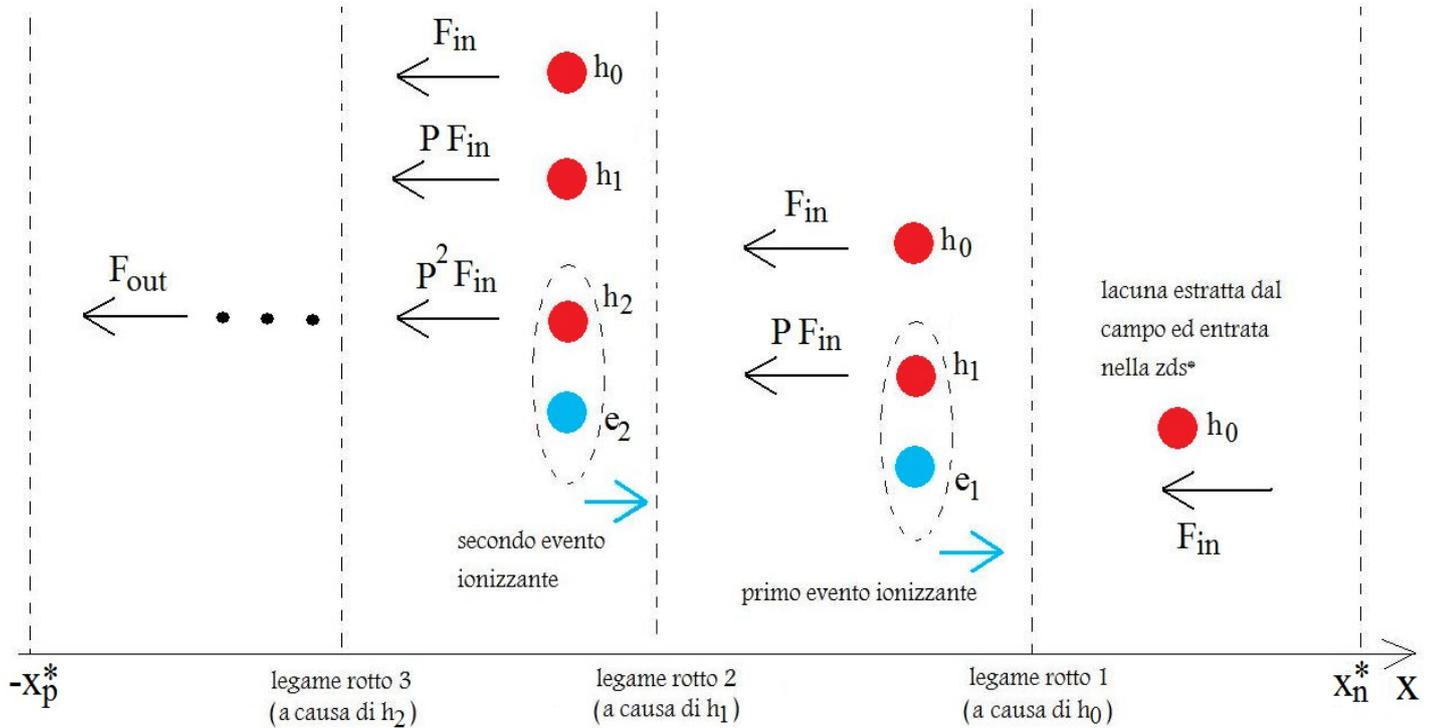


Fig. 26

Rappresentazione concettuale del breakdown a valanga ($|V|$ poco superiore a V_{BD}).

Supponiamo di seguire una singola lacuna (h_0) che viene estratta nei pressi di x_n^* ed immessa nella zds^* , a causa della polarizzazione inversa. La lacuna h_0 , a causa della presenza del campo estraente, è accelerata lungo il suo tragitto all'interno della zds^* , ricevendo un'energia cinetica E_e elevatissima in virtù del voltaggio inverso molto elevato (campo molto forte). L'energia E_e , che è maggiore dell'energia di ionizzazione del silicio E_i , consente alla lacuna, non appena “urta” contro un legame covalente Si - Si (formato dalla sovrapposizione di due orbitali ibridi sp^3 , ciascuno dei quali fornisce un elettrone di valenza), di romperlo, liberando un elettrone e_1 ed una lacuna h_1 ; l'elettrone secondario e_1 va in banda di conduzione e la lacuna secondaria h_1 in quella di valenza. Quindi mentre attraverso la superficie attiva S posta prima del legame rotto e ortogonale all'asse x fluiva un flusso F_{in} , associato alla lacuna h_0 , una superficie S identica e parallela, posizionata dopo il legame rotto, è attraversata dal flusso F_{in} e da un secondo flusso pari a $P F_{in}$, il quale è associato alla lacuna termogenerata h_1 , dove P è la probabilità che h_0 possa produrre la ionizzazione. P , compresa fra 0 ed 1 e tendente a 0 se $|V| < V_{BD}$, risulta tanto maggiore quanto maggiore è la V inversa. A sua volta h_1 romperà un legame, a breve distanza dal primo, provocando una seconda ionizzazione in cui si generano un elettrone e_2 ed una lacuna h_2 ; a quest'ultima è associato un flusso $P^2 F_{in}$. Gli elettroni come e_1 ed e_2 sono chiamati “raggi delta” ed il loro comportamento è analogo a quello delle lacune h_1, h_2 , ecc..., salvo alcune eventuali differenze nella probabilità di produrre ionizzazioni per impatto. Il flusso di lacune che attraversa la superficie attiva S in prossimità di $-x_p^*$ è dato da:

$$F_{out} = F_{in} + P F_{in} + P^2 F_{in} + P^3 F_{in} + \dots + P^n F_{in} + \dots$$

$$\frac{F_{out}}{F_{in}} = 1 + P + P^2 + P^3 + \dots = \sum_{k=0}^{\infty} P^k = \frac{1}{1-P} = \frac{1}{1 - \left(\frac{|V|}{V_{BD}}\right)^\alpha} = M \propto \frac{E_e}{E_i}$$

L'ultima relazione (quella di proporzionalità diretta) è valida solo nell'ipotesi in cui la carica primaria (h_0), durante tutto il suo percorso dal bulk di estrazione a quello di destinazione, non esca mai dal volume attivo della zds^* (volume di moltiplicazione, pari a $S \times zds^*$). M è il "fattore di moltiplicazione della densità di corrente" (più precisamente, in questo caso, è il fattore di moltiplicazione della densità di corrente d'estrazione, ovvero della corrente dovuta esclusivamente ai portatori minoritari estratti, termogenerati nei due bulk), ed è pari al numero di lacune che arrivano presso $-x_p^*$, prodotte nella zds^* , derivanti da una singola lacuna primaria estratta dal campo. Quindi alla "corrente di estrazione", la quale coincide esattamente con la corrente inversa I_0 solo nel caso in cui $P = 0$, ovvero se $|V| < V_{BD}$, si somma una "corrente di generazione" (generazione termica) formata dai portatori secondari, termogenerati lungo tutto il tragitto delle cariche primarie e raccolti "a fine corsa". La somma fra corrente di estrazione e corrente di generazione è chiamata "corrente di buio" oppure "corrente di dark - bulk"; quest'ultima costituisce un paramentro di fondamentale rilevanza per qualunque fotorivelatore a semiconduttore. A questa si sovrappone una "corrente di perdita" ("leakage current") legata alla ricombinazione di portatori secondari e primari all'interno della zds^* : tale ricombinazione è dovuta in prevalenza alle impurezze e ai difetti reticolari, concentrati soprattutto sulla superficie del dispositivo. Più $|V|$, partendo da V_{BD} , si avvicina al valore asintotico V_{BD}^* , e più M aumenta, fino a divergere, così che la corrente di generazione domina su quella di estrazione. La valanga è costituita dall'insieme delle cariche primarie e secondarie, quest'ultime generate nella zds^* , che vengono raccolte presso i due bulk di destinazione. Un discorso equivalente a quello svolto finora per le lacune vale per gli elettroni estratti e immessi nella zds^* : infatti ciascuno di loro innescherà, in linea teorica, una valanga proporzionale a $|V|$, pertanto per ciascun elettrone che entra nella zds^* da $-x_p^*$ ci saranno, in media, M elettroni raccolti nei pressi della sezione x_n^* . Affinchè si inneschi il breakdown a valanga è necessario che $|V|$ sia almeno pari a V_{BD} , ovvero che il valore massimo del campo elettrico (che si trova presso $x = 0$) sia:

$$\epsilon_{max} \geq \epsilon_{maxBD} = \sqrt{\frac{2qN_d(V_0 + V_{BD})}{\epsilon_S}}$$

4.3) Generalità sull'interazione fra luce e semiconduttore

La possibilità di utilizzare una giunzione pn come fotorivelatore si basa sul principio fisico secondo il quale la luce, una volta penetrata all'interno della matrice cristallina del semiconduttore con il quale la giunzione è realizzata, interagisce con gli elettroni atomici del cristallo. L'onda elettromagnetica incidente sul semiconduttore fornisce agli elettroni atomici, localizzati sugli orbitali leganti ibridi sp^3 mediante i quali gli atomi formano i legami covalenti, l'energia necessaria a promuoverli dalla banda di valenza a quella di conduzione. Gli elettroni così eccitati (fotoeccitati), chiamati anche "fotoelettroni", possono partecipare alla conduzione elettrica, se trascinati da un campo elettrico di polarizzazione imposto dall'esterno, ovvero ai terminali metallici della giunzione pn. Trattiamo brevemente il problema dell'interazione, detta anche "scattering", tra i fotoni incidenti su un materiale cristallino (un semiconduttore, nel nostro caso) e gli elettroni atomici; questo fenomeno è alla base delle proprietà ottiche dei materiali utilizzanti nel campo della fotonica, e nella fattispecie per la realizzazione dei fotorivelatori a semiconduttore.

La luce che si propaga all'interno di un cristallo può essere vista, in modo semplificato, come un'onda elettromagnetica piana. Questa è rappresentabile mediante una terna destrorsa di vettori formata dal vettore del campo elettrico \vec{E} , dal vettore del campo magnetico \vec{H} e dal versore d'onda \hat{k} , il quale individua direzione e verso di propagazione. Alla luce propagante, che noi supponiamo monocromatica, è associata una velocità di fase, la cui espressione è la seguente:

$$v_{fase} = \frac{c}{n_r} = \frac{c}{n_r' + in_r''} = \frac{c}{\sqrt{\epsilon_r + i\left(\frac{\sigma\mu_0}{\omega}\right)}}$$

n_r è l'indice di rifrazione del semiconduttore, complesso ($n_r' = \text{Re}\{n_r\}$, $n_r'' = \text{Im}\{n_r\}$) in quanto il cristallo è un mezzo dispersivo (v_{fase} è una funzione di ω) e dissipativo ($\sigma > 0$), ω è la pulsazione dell'onda, μ_0 è la costante di permeabilità magnetica del vuoto, ϵ_r è la costante dielettrica relativa del mezzo, la quale quantifica l'interazione fra la luce propagante ed il campo elettrico periodico prodotto dagli atomi nel reticolo cristallino, σ è la conducibilità del semiconduttore. La velocità di fase v_{fase} , a rigore, non coincide con la velocità v_{prop} con cui l'onda elettromagnetica si propaga all'interno di un generico cristallo; la velocità di propagazione, infatti, è definita nel modo seguente:

$$V_{\text{prop}} = \frac{c}{\sqrt{\epsilon_r \mu_r}}$$

dove μ_r è la costante di permeabilità magnetica relativa del mezzo. La velocità della luce attraverso un cristallo è decrementata di un fattore pari a $\sqrt{\epsilon_r \mu_r}$: tale rallentamento è dovuto all'interazione fra l'onda elettromagnetica propagante ed il campo elettrico periodico interno al reticolo. Se il cristallo nel quale si propaga la luce è un semiconduttore, la velocità di propagazione v_{prop} dell'onda è definita nel modo seguente:

$$V_{\text{prop}} \approx \frac{c}{\sqrt{\epsilon_r}}$$

poiché nei semiconduttori $\mu_r \approx 1$. Inoltre un semiconduttore, specialmente se intrinseco o leggermente drogato, ha una conducibilità σ compresa fra 10^{-8} e 10^3 S/cm, molto bassa rispetto a quella dei metalli come Ag, Cu, Au ed Al, che è dell'ordine di 10^6 S/cm, pertanto σ , nel caso di un cristallo semiconduttore, può essere considerata nulla all'interno dell'espressione di v_{fase} , consentendo di scrivere quanto segue:

$$V_{\text{fase}} \approx \frac{c}{\sqrt{\epsilon_r}}$$

Dunque la velocità di fase V_{fase} coincide, di fatto, con la velocità v_{prop} con cui la luce attraversa il semiconduttore.

Un'onda elettromagnetica piana è modellizzata univocamente dai vettori \vec{E} , \vec{H} e \hat{k} , ovvero le caratteristiche di questi vettori, come ad esempio la pulsazione, la polarizzazione, l'ampiezza ecc..., descrivono un determinato campo elettromagnetico propagante attraverso una matrice cristallina. In altri termini ad una luce monocromatica che attraversa un cristallo, ovvero ad un segnale ottico, è associabile una sola terna destroza \vec{E} , \vec{H} , \hat{k} . La stessa onda elettromagnetica può essere descritta, inoltre, da molte coppie (\vec{A}, Φ) di potenziali, dove \vec{A} è il "potenziale vettore" e Φ il "potenziale scalare". \vec{A} e Φ possono essere trattati come dei meri strumenti matematici adatti a risolvere più agevolmente le equazioni di Maxwell (\vec{E} , \vec{H} , \hat{k} corrispondono, invece, a reali osservabili fisici). Quindi ad una terna destroza \vec{E} , \vec{H} , \hat{k} corrispondono un certo numero di coppie (\vec{A}, Φ) , (\vec{A}', Φ') , (\vec{A}'', Φ'') , (\vec{A}''', Φ''') e così via. Ciascun set di potenziali possiede delle proprietà particolari e delle simmetrie, dette "gauges", e rappresenta compiutamente il campo elettrico \vec{E} , \vec{H} , \hat{k} . Pertanto conviene scegliere un'opportuna coppia (\vec{A}, Φ) , rappresentativa del campo elettromagnetico, le cui simmetrie consentano di semplificare la risoluzione del sistema di equazioni di Maxwell, che di seguito riportiamo in forma differenziale:

$$\begin{cases} \nabla \times \vec{E} + \frac{\partial \vec{B}}{\partial t} = 0 \\ \nabla \times \vec{H} - \frac{\partial \vec{D}}{\partial t} = J \\ \nabla \cdot \vec{D} = \rho \\ \nabla \cdot \vec{B} = 0 \end{cases}$$

dove $\nabla \times$ rappresenta il rotore, $\nabla \cdot$ la divergenza, \vec{B} = vettore di induzione magnetica = $\mu_0 \mu_r \vec{H}$, \vec{D} = vettore di induzione elettrica = $\epsilon_0 \epsilon_r \vec{E}$, J è la densità superficiale di corrente elettrica, ρ è la concentrazione volumetrica di carica elettrica. Abbiamo supposto che la densità superficiale di corrente magnetica J_m e la concentrazione volumetrica di carica magnetica ρ_m siano nulle.

Per un qualunque set di potenziali (\vec{A}, Φ) scelto, per definizione di potenziale vettore \vec{A} e potenziale scalare Φ , abbiamo che:

$$\vec{E} = - \frac{\partial \vec{A}}{\partial t} - \nabla \Phi$$

$$\vec{H} = \frac{\nabla \times \vec{A}}{\mu_0 \mu_r}$$

dove ∇ rappresenta il gradiente. Il set (\vec{A}, Φ) con cui lavoriamo deve avere la seguente proprietà:

$$\vec{A} = \vec{A}' - \nabla\chi$$

$$\Phi = \Phi' + \frac{\partial\chi}{\partial t}$$

dove χ è un funzionale fissato, mentre \vec{A}' e Φ' sono due potenziali che soddisfano la seguente simmetria, nota come “gauge di Lorentz”:

$$\nabla \cdot \vec{A}' + \frac{\partial\Phi'}{\partial t} = 0$$

Utilizziamo, oltre alla “gauge di Lorentz”, anche quella di Coulomb, spesso chiamata “radiation gauge”, secondo cui nelle equazioni di Maxwell \mathbf{J} e ρ sono nulle: tale scelta comporta che $\Phi = 0$. Inoltre la gauge di Coulomb prevede che $\nabla \cdot \vec{A} = 0$. Tenendo conto degli aspetti fin qui discussi possiamo rielaborare le equazioni di Maxwell in funzione del potenziale vettore e del potenziale scalare ed arrivare alla forma di D’Alambert per la seconda equazione:

$$\frac{1}{\mu_0} \nabla^2 \vec{A} = \varepsilon_0 \varepsilon_r \frac{\partial^2 \vec{A}}{\partial t^2}$$

dove ∇^2 rappresenta l’operatore laplaciano. Risolvendola otteniamo l’espressione del potenziale vettore:

$$A(\vec{r}, t) = A_0 e^{i(\vec{k} \cdot \vec{r} - \omega t)} \quad [\text{e23}]$$

dove \vec{r} è il vettore posizione all’interno del reticolo diretto di Bravais, mentre A_0 è il modulo del potenziale vettore \vec{A} . Noto il potenziale vettore è immediato calcolare le espressioni dei vettori campo elettrico e campo magnetico:

$$\vec{E} = -\frac{\partial \vec{A}}{\partial t} = -2\omega A_0 \sin(\vec{k} \cdot \vec{r} - \omega t) \hat{e}$$

$$\vec{H} = \frac{\nabla \times \vec{A}}{\mu_0 \mu_r} = \vec{k} \times \left[-2\omega \frac{A_0}{\mu_0 \mu_r} \sin(\vec{k} \cdot \vec{r} - \omega t) \hat{e} \right]$$

dove \times indica un prodotto vettoriale, \hat{e} rappresenta il versore che individua la direzione ed il verso del vettore campo elettrico \vec{E} . Quest’ultimo può essere espresso anche nel modo seguente, supponendo che la luce attraversi il semiconduttore lungo una direzione ed un verso individuati da un ipotetico asse x (perpendicolare ad \hat{e}):

$$\vec{E} = E_0 e^{i\omega \left(\frac{n_r' x}{c} - t \right)} e^{-\left(\frac{n_r'' \omega}{c} \right) x} \hat{e} \quad [\text{e24}]$$

dove:

$$\alpha(\omega) = \frac{2\omega}{c} n_r''(\omega) \quad [\text{e25}]$$

è il “coefficiente di assorbimento ottico specifico” (cm^{-1}), detto anche “coefficiente di assorbimento dell’energia elettromagnetica”, che rappresenta il numero medio di fotoni assorbiti dal materiale in un cm. Dalla [e24] si evince l’evanescenza lungo x della luce quando questa attraversa un cristallo, dispersivo e dissipativo, opaco nei confronti dell’onda elettromagnetica; il tasso di assorbimento dei fotoni, che comporta il passaggio di elettroni dalla BV alla BC, dipende dalla frequenza dell’onda propagante.

Conoscendo le espressioni di \vec{E} ed \vec{H} è possibile calcolare quella del vettore di Poynting \vec{S} ; dal modulo di \vec{S} si può ricavare l’intensità ottica media $\langle S \rangle_{\text{time}}$ (W/cm^2), dove la media è eseguita nel tempo. Se dividiamo $\langle S \rangle_{\text{time}}$ per la velocità di propagazione v_{prop} della luce nel semiconduttore otteniamo la seguente densità volumetrica di energia:

$$\frac{\langle S \rangle_{\text{time}}}{v_{\text{prop}}} = \frac{2\varepsilon_0 \varepsilon_r \omega^2 A_0^2}{c^2} \quad [\text{J}/\text{cm}^3]$$

Dato che la luce propagante è monocromatica per ipotesi, possiamo affermare che:

$$\frac{2\varepsilon_0\varepsilon_r\omega^2 A_0^2}{c^2} = \frac{n_{ph}\hbar\omega}{Vol} = n_\gamma(\omega) \hbar\omega$$

dove Vol è il volume unitario di semiconduttore (1cm³), n_{ph} è il numero di fotoni aventi energia $\hbar\omega$ nell'unità di volume, n_γ(ω) è la concentrazione di fotoni di energia $\hbar\omega$ e dipende dal valore assunto, presso l'energia $\hbar\omega$, dalla ben nota probabilità di occupazione di stati per i fotoni (bosoni con spin S_z = +/- \hbar) chiamata "distribuzione statistica quantistica di Bose - Einstein". Dall'ultima relazione si ricava che:

$$A_0 = \sqrt{\frac{\hbar c^2}{2\varepsilon_0\varepsilon_r\omega} n_\gamma(\omega)} \propto E_0 \quad [e26]$$

La [e26] conferma il concetto intuitivo secondo il quale maggiore è il numero di fotoni che condividono lo stesso insieme di numeri quantici, ovvero maggiore è il numero di fotoni coerenti (gemelli, cioè che hanno stessa energia, stessa direzione e verso di propagazione, stessa fase iniziale e stessa polarizzazione), e maggiore è l'intensità del campo elettromagnetico ed il modulo del potenziale vettore a questo associato.

Supponiamo che un fascio coerente di n_{ph} fotoni, tutti di energia $\hbar\omega$, investano una regione del cristallo di semiconduttore, e che in questa regione sia presente un elettrone atomico. L'elettrone può trovarsi in BC oppure in BV, per cui è descritto da un'onda di Bloch la cui componente "central cell function" u(r) ha simmetria di tipo s (oppure s + p) o di tipo p. Il sistema "pre - scattering", ovvero il sistema formato dall'elettrone e dai n_{ph} fotoni prima dell'interazione luce - materia, può essere descritto dal seguente stato quantistico iniziale di ket:

$$| i \rangle = | n_{ph}, k_i \rangle$$

Questo stato iniziale "totale" | i > di pre - scattering consta di un sottostato "fotonico" | n_{ph} > di pre - scattering e di un sottostato "elettronico" | k_i > di pre - scattering; k_i è il vettore d'onda "iniziale" dell'elettrone, che localizza la posizione dell'elettrone all'interno del diagramma a bande E - k prima dello scattering con la luce. L'onda elettromagnetica, scatterante con la materia, ha un potenziale vettore $\vec{A} = A_0 \hat{a}$, dove \hat{a} è il versore di polarizzazione del fascio coerente di fotoni. Ipotizziamo che l'interazione fra l'elettrone e la luce abbia effettivamente luogo: l'elettrone passerà dal suo stato iniziale k_i ad uno finale k_f, passando ad esempio dalla BV alla BC nel caso di assorbimento di uno degli n_{ph} fotoni iniziali, oppure passando dalla BC alla BV nel caso di emissione stimolata di un fotone, il quale si aggiungerà agli n_{ph} fotoni iniziali. Il sistema "post - scattering", ovvero il sistema formato dall'elettrone nel nuovo stato k_f e dai n_{ph} +/- 1 fotoni dopo l'interazione luce - materia, può essere descritto dal seguente stato quantistico finale (di ket) "totale" di post - scattering:

$$| f \rangle = | n_{ph} \mp 1, k_f \rangle$$

L'hamiltoniana dell'elettrone un attimo prima dello scattering con la luce (elettrone al buio) può essere espressa nel modo seguente:

$$H_{dark} = \frac{p^2}{2m_0} + U(r)$$

Durante lo scattering fra la luce e l'elettrone l'hamiltoniana di quest'ultimo si modifica nel modo seguente:

$$H_{light} = \frac{(p - eA)^2}{2m_0} + U(r) + e\Phi \quad [e27]$$

dove A è l'operatore potenziale vettore associato alla luce che scattera con l'elettrone, e è la carica elementare, Φ è il potenziale scalare associato alla luce. Espandendo la [e27] possiamo scrivere l'equazione di Schrödinger dipendente dal tempo per un elettrone atomico in procinto di scatterare con la luce:

$$i\hbar \frac{\partial \psi(k,t)}{\partial t} = \hat{H}_{light} \psi(k,t) = \left[\frac{p^2}{2m_0} - \frac{e}{2m_0} \mathbf{p} \cdot \mathbf{A} - \frac{e}{2m_0} \mathbf{A} \cdot \mathbf{p} + \frac{e^2 A^2}{2m_0} + U(r) + e\Phi \right] \psi(k,t)$$

Portando l'operatore della quantità di moto \hat{p} dallo spazio reciproco dei vettori d'onda a quello reale delle coordinate, possiamo scrivere l'equazione di Schrödinger nel modo seguente:

$$i\hbar \frac{\partial \psi(r,t)}{\partial t} = \hat{H}_{light} \psi(r,t) = \left[-\frac{\hbar^2 \nabla^2}{2m_0} + \frac{ie\hbar}{2m_0} \nabla \cdot \mathbf{A} + \frac{ie\hbar}{2m_0} \mathbf{A} \cdot \nabla + \frac{e^2 A^2}{2m_0} + U(r) + e\Phi \right] \psi(r,t)$$

È possibile dimostrare che nell'ipotesi in cui il cristallo del semiconduttore sia attraversato da piccoli segnali ottici (bassa potenza ottica) è concesso scrivere quanto segue:

$$i\hbar \frac{\partial \psi(r,t)}{\partial t} = \hat{H}_{light} \psi(r,t) = \left[-\frac{\hbar^2 \nabla^2}{2m_0} + \frac{ie\hbar}{2m_0} \nabla \cdot \mathbf{A} + \frac{ie\hbar}{m_0} \mathbf{A} \cdot \nabla + \frac{e^2 A^2}{2m_0} + U(r) + e\Phi \right] \psi(r,t)$$

La terza versione dell'equazione di Schrödinger differisce dalla seconda unicamente per il coefficiente complesso che moltiplica $\mathbf{A} \cdot \nabla$: al denominatore del coefficiente il 2 è sparito. Avendo adottato la gauge di Coulomb, oltre a quella di Lorentz, abbiamo che $\nabla \cdot \mathbf{A} = 0$ e che $\Phi = 0$ ($\mathbf{J} = 0$ e $\rho = 0$); inoltre il potenziale vettore \mathbf{A} , associato al piccolo segnale ottico propagante, ha modulo piccolo, pertanto A^2 , a maggior ragione, rende il termine potenziale a cui appartiene del tutto trascurabile rispetto agli altri addendi compresi tra le parentesi quadre. Dunque possiamo scrivere la versione finale dell'equazione di Schrödinger, dipendente dal tempo, per un elettrone atomico in procinto di scatterare con la luce:

$$i\hbar \frac{\partial \psi(r,t)}{\partial t} = \hat{H}_{light} \psi(r,t) = \left[-\frac{\hbar^2 \nabla^2}{2m_0} + U(r) + \frac{ie\hbar}{m_0} \mathbf{A} \cdot \nabla \right] \psi(r,t) = [H_0 + H_F] \psi(r,t)$$

$$H_0 = -\frac{\hbar^2 \nabla^2}{2m_0} + U(r) = H_{dark}$$

$$H_F = \frac{ie\hbar}{m_0} \mathbf{A} \cdot \nabla$$

H_0 è l'hamiltoniana dell'elettrone non scatterato dalla luce, ovvero l'energia totale dell'elettrone non stimolato da alcun fotone (elettrone "a riposo"), mentre $H_0 + H_F$ è l'hamiltoniana dell'elettrone scatterato dalla luce; H_F rappresenta il termine hamiltoniano "fotoperturbativo" per l'elettrone. Il prodotto scalare fra \mathbf{A} e ∇ , presente nell'espressione di H_F , è equivalente al prodotto scalare fra \hat{a} e \vec{p} , ricordando che $\vec{A} = A_0 \hat{a}$ e che $\vec{p} = -i\hbar \nabla$. Quindi il valore di H_F fornisce informazioni sui casi in cui un fotone scattera effettivamente con un elettrone del semiconduttore e sui tipi di scattering che possono aver luogo (assorbimento oppure emissione stimolata) in funzione dell'overlap fra \hat{a} e \vec{p} . Si presenta il seguente ventaglio di situazioni:

- affinché ci sia scattering tra fotone ed elettrone \mathbf{A} deve essere diverso da zero, ovvero $A_0 > 0$, e quindi $n_{ph} > 0$; se $\mathbf{A} = 0$, cioè $n_{ph} = 0$, allora anche $H_F = 0$, pertanto l'elettrone è in condizione di buio e non può né essere promosso dalla BV alla BC, mediante assorbimento di un fotone, né decadere dalla BC alla BV, con conseguente emissione di un fotone identico (gemello) a quello stimolatore. Se $\mathbf{A} = 0$ lo stato totale iniziale $|n_{ph}, k_i\rangle$ rimane inalterato, o comunque non subisce perturbazioni riconducibili a cause ottiche.
- affinché ci sia scattering tra fotone ed elettrone l'overlap fra \hat{a} e \vec{p} , cioè il prodotto scalare fra il vettore unitario di polarizzazione della luce ed il vettore impulso dell'elettrone atomico, deve essere diverso da zero. Se \hat{a} e \vec{p} sono ortogonali l'elettrone attraversa il piano su cui giace il campo elettrico, che è individuato da \hat{a} , per un lasso di tempo troppo breve affinché il campo elettrico possa accelerare o decelerare l'elettrone stesso. Se invece $\hat{a} \cdot \vec{p} \neq 0$ l'onda elettromagnetica che scattera con l'elettrone appartenente al cristallo ha il tempo di fornirgli energia, promuovendolo dalla BV alla BC, o di frenarlo, sottraendogli energia e quindi provocando il suo decadimento dalla BC alla BV.
- evidentemente se $\hat{a} \cdot \vec{p} = -p$ la probabilità di accelerazione dell'elettrone mediante l'assorbimento di uno degli n_{ph} fotoni iniziali è massima, ovvero la trasformazione dello stato totale iniziale (pre - scattering) $|n_{ph}, k_i\rangle$ in uno finale (post - scattering) $|n_{ph} - 1, k_f\rangle$ è massima. Se invece $\hat{a} \cdot \vec{p} = p$ la probabilità di decelerazione dell'elettrone mediante emissione di un fotone è massima, ovvero la trasformazione dello stato totale iniziale (pre - scattering) $|n_{ph}, k_i\rangle$ in uno finale (post - scattering) $|n_{ph} + 1, k_f\rangle$ è massima.

Chiaramente la transizione dallo stato totale pre - scattering a quello totale post - scattering ($|i\rangle \rightarrow |f\rangle$) deve soddisfare il principio di conservazione dell'energia, pertanto il fotone assorbito o il fotone emesso devono avere un'energia $\hbar\omega$ pari al modulo della differenza tra l'energia dello stato finale e l'energia dello stato iniziale ($|E_f - E_i| = \hbar\omega$), ossia $\hbar\omega$ deve essere uguale all'energia acquistata dall'elettrone durante l'assorbimento o all'energia persa dall'elettrone durante l'emissione; inoltre la transizione $|i\rangle \rightarrow |f\rangle$ deve soddisfare il principio di conservazione della quantità di moto $p = \hbar k$. A proposito di quest'ultima condizione, se il sistema consta solo di

un fotone ed un elettrone che scatterano (sistema a due corpi, cioè la situazione discussa in questo paragrafo), ovvero se non vi è alcun intervento di vibrazioni fononiche, la transizione dell'elettrone dalla BV alla BC o viceversa è, con buona approssimazione, rettilinea all'interno del diagramma a bande $E(k)$; si parla infatti di "approssimazione di dipolo elettrico", dove le due cariche uguali e di segno opposto rappresentano un elettrone ed una lacuna, mentre la distanza (energetica) fra le due cariche è proprio $|E_f - E_i| = \hbar\omega$. In altre parole $k_i \approx k_f \gg k_\gamma$, ovvero i numeri d'onda (gli impulsi) dell'elettrone prima e dopo lo scattering sono di fatto uguali, in quanto il fotone assorbito oppure emesso non ha contribuito in modo rilevante all'incremento o al decremento della quantità di moto dell'elettrone (nel paragrafo 4.4.2 stimeremo, qualitativamente, il numero d'onda k_γ di un fotone di energia $\hbar\omega$).

Possiamo quindi riportare in [e28] la legge che regola (quantifica) la probabilità di scattering fotone - elettrone, nota come "Fermi golden rule" (regola d'oro di Fermi):

$$W(|i\rangle) = \frac{2\pi}{\hbar} \sum_{|f\rangle} |\langle f | H_F | i \rangle|^2 \delta(|E_f - E_i| -/+ \hbar\omega) \quad [\text{cm}^{-3}\text{s}^{-1}] \quad [\text{e28}]$$

$W(|i\rangle)$ è il rate, cioè un indice di probabilità, con cui un elettrone localizzato presso uno stato iniziale (pre - scattering) k_i , e stimolato da un fotone di energia $\hbar\omega$ avente determinate caratteristiche (fase, energia, direzione e verso di propagazione, polarizzazione), cioè da un fotone localizzato presso uno stato iniziale γ_i (k_i è lo stato elettronico iniziale, γ_i lo stato fotonico iniziale), transita da un livello energetico (elettronico) iniziale E_i ad uno finale (post - scattering) E_f ; quest'ultimo è caratterizzato (individuato) da un certo numero di stati elettronici k_f . k_i e γ_i sono i sottostati dello stato totale iniziale $|i\rangle$. La legge di conservazione dell'energia è implementata nella [e28] dalla delta di Dirac, che assume valore unitario solo se il suo argomento è nullo, cioè solo se $\hbar\omega = |E_f - E_i|$. Il segno negativo davanti ad $\hbar\omega$ si riferisce ad una transizione di assorbimento, mentre quello positivo ad un'emissione. Il numero di fotoni di cui consta l'onda elettromagnetica che scatterà con l'elettrone, e quindi il modulo A_0 del potenziale vettore \vec{A} (\propto modulo del campo elettrico scatterante), è contenuto, come già visto, nell'hamiltoniana di fotoperturbazione H_F . In questa è contenuto anche il prodotto scalare $\hat{a} \cdot \vec{p}$, che rappresenta l'overlap fra il vettore dell'impulso dell'elettrone ed il versore del campo elettrico del segnale ottico, ovvero l'attitudine che ha la luce in questione, durante il processo di scattering, ad alterare il livello energetico iniziale E_i dell'elettrone. La sommatoria è eseguita su tutti gli stati finali totali $|f\rangle$ che afferiscono alla conclusione della transizione: tale sommatoria è, di fatto, la sommatoria su tutti gli stati elettronici finali k_f che afferiscono al livello energetico finale E_f su cui si posiziona l'elettrone a scattering avvenuto. In altre parole fissato $|i\rangle$, la sommatoria è proporzionale alla densità di stati per l'elettrone in BC o in BV, a seconda che la transizione descritta dalla regola d'oro di Fermi sia un assorbimento o un'emissione stimolata. In generale possiamo esprimere l'elemento matriciale all'interno del modulo quadro, nella sommatoria della [e28], nel modo seguente:

$$\langle f | H_F | i \rangle = \langle \psi_f(r,k) | \hat{H}_F | \psi_i(r,k) \rangle = \int \psi_f(r,k)^* e^{+/- i(k_\gamma \cdot r)} \left[\frac{i\hbar}{m_0} \mathbf{A} \cdot \nabla \right] \psi_i(r,k) d^3r \quad [\text{e29}]$$

dove ψ_f è lo stato elettronico finale, di arrivo per l'elettrone, ψ_i è lo stato elettronico iniziale, di partenza per l'elettrone, mentre $\exp\{+/- i(k_\gamma \cdot r)\}$ indica la presenza del fotone durante il processo di scattering e coincide con la forma "granulare", cioè di singolo fotone, dell'espressione del potenziale vettore riportata nella [e23]: + se la regola d'oro di Fermi descrive un assorbimento, - per un'emissione. La natura di ψ_f e ψ_i dipende dal tipo di struttura (bulk, pozzi quantici, fili quantici o punti quantici) nella quale si trovano gli elettroni atomici: nel caso di semiconduttore di tipo bulk (nessun confinamento quantistico) ψ_f e ψ_i sono delle onde di Bloch.

4.4.1) Un esempio di omogiunzione pn al silicio integrata, utilizzata come fotorivelatore, vista in sezione

In figura 27 è rappresentata la sezione, molto semplificata, di un'omogiunzione pn realizzata in silicio, usata come fotodiiodo. La luce incide su uno strato di materiale antiriflesso, per esempio teflon oppure vari tipi di silice SiO_2 , così da convogliare la maggior quantità possibile di energia luminosa sulla superficie attiva del dispositivo. Le metallizzazioni superiori devono essere le più sottili possibili, così da massimizzare la superficie illuminata. Gli angoli di riflessione e trasmissione alle interfacce seguono la nota legge di Snell (si veda il glossario), per cui è importante, in fase di progetto, curare la realizzazione dello strato antiriflesso in modo da ottenere il maggior rendimento possibile. I fotoni trasmessi (rifratti), superato anche lo strato di ossido di passivazione, entrano nel silicio. Supponiamo che la maggior parte dei fotoni arrivi fino all'interno della z_{ds}^* (approfondiremo nel capitolo 5 la questione dell'assorbimento della luce in regioni del fotodiiodo "non gradite", vale a dire nei bulk p ed n).

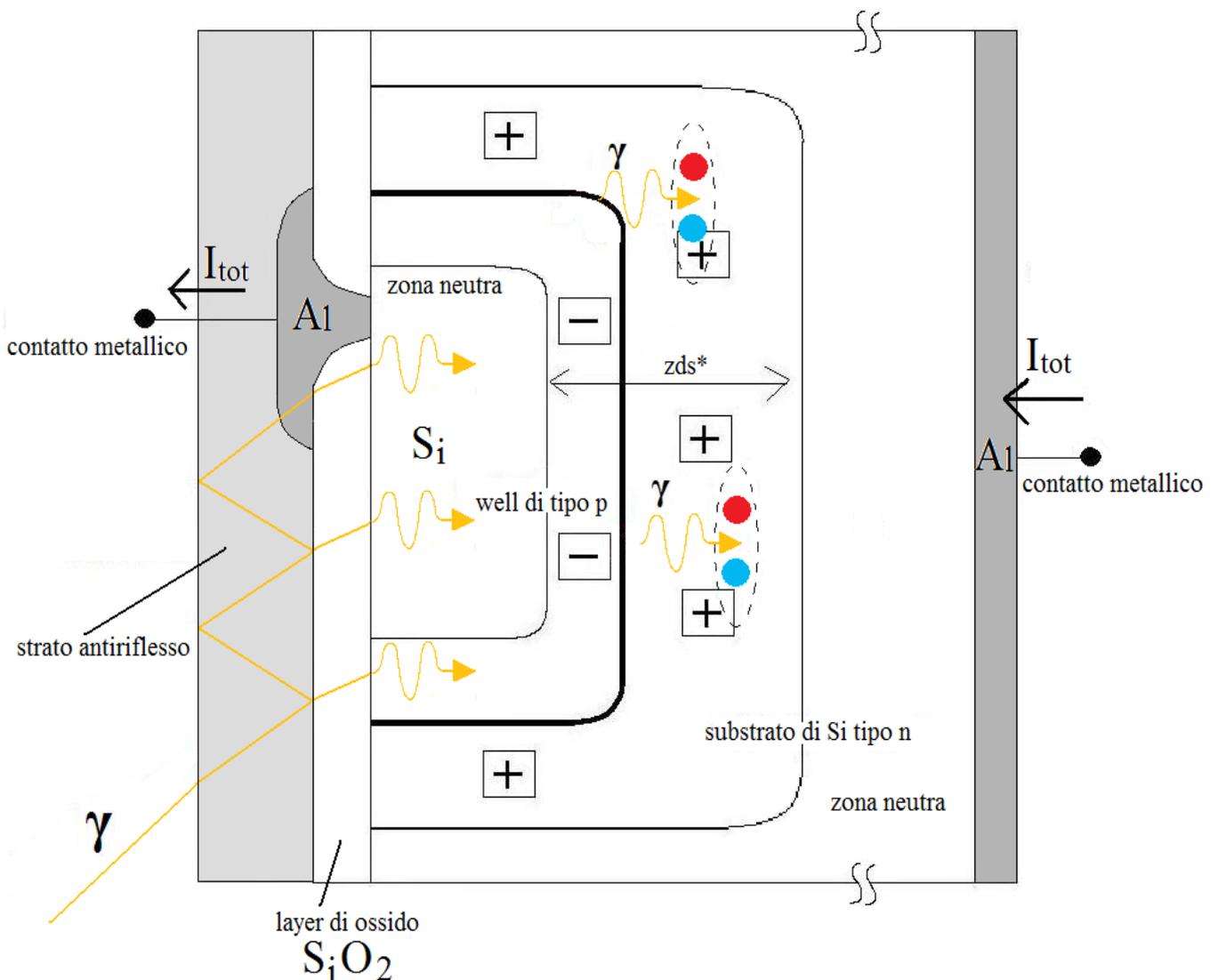


Fig. 27

Omogiunzione pn al silicio utilizzata come fotodiiodo (polarizzata in inversa), vista in sezione. Sono ben distinguibili il substrato n, nel quale è stato ricavato un pozzetto ("well") di semiconduttore drogato p (ad esempio per diffusione termica), l'ossido di silicio di passivazione, i contatti metallici e lo strato antiriflesso, utile per massimizzare la potenza ottica trasmessa all'interno della regione di svuotamento.

4.4.2) Un caso particolare di interazione fra luce e silicio

I fotoni possono interagire con il silicio all'interno della zds^* della giunzione pn polarizzata inversamente; nella zds^* , come è noto, lo svuotamento (completo) di portatori di carica mobili comporta, in assenza di radiazione luminosa incidente e a temperature ordinarie, la mancanza di elettroni in BC e di lacune in BV. I fotoni possono così rilasciare tutta la loro energia $E = \hbar\omega$ agli elettroni atomici del silicio, nella zds^* , provocando pertanto la generazione di coppie elettrone/lacuna (cariche primarie fotogenerate), come mostrato in figura 28.

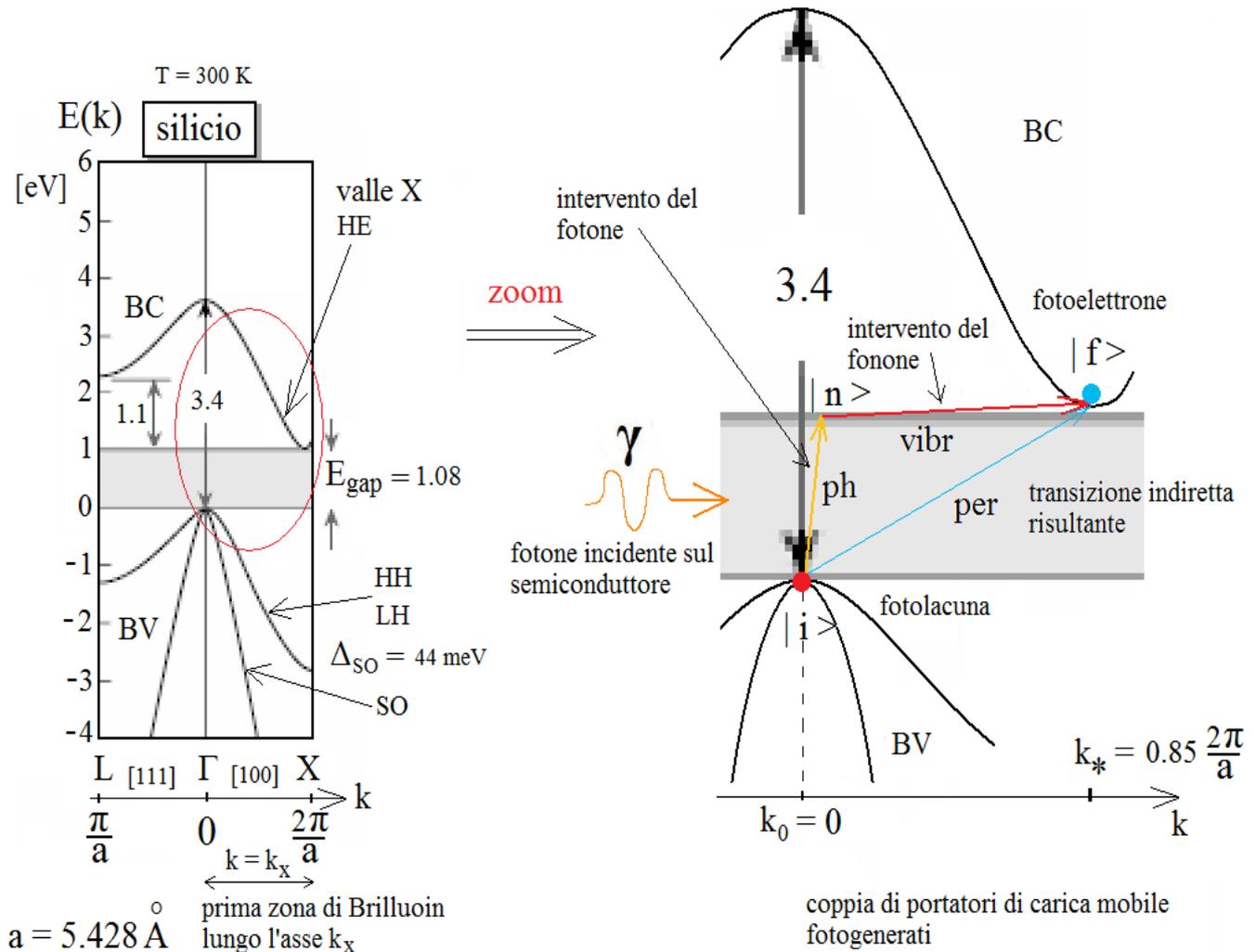


Fig. 28

Sulla sinistra è riportata una porzione del diagramma $E - k$ del silicio monocristallino a temperatura ambiente: si tratta della regione di $E(k)$ di maggiore interesse per applicazioni fotoniche. Sulla destra è mostrato un ingrandimento della parte del diagramma a bande del silicio circondata dall'ellisse rossa. Tale ingrandimento è finalizzato alla trattazione del problema a tre corpi, ovvero fotone, elettrone e fonone, che consente di spiegare l'assorbimento ottico, da parte del silicio, di lunghezze d'onda infrarosse prossime ai 1150 nm, ovvero di energie fotoniche prossime a quella di gap ($\approx 1.08 \text{ eV}$, gap della valle X).

Va tuttavia puntualizzato il ruolo importante che possono assumere le vibrazioni termiche, dette "fononi", del reticolo cristallino del silicio, ovvero le vibrazioni termiche di ciascun atomo del cristallo intorno alla propria posizione di equilibrio, cioè di minima energia. Supponiamo che arrivino, nella zds^* , dei fotoni "trasversali" (da non confondere con i fotoni "virtuali" che consentono molte interazioni fra particelle e decadimenti delle stesse), ovvero fotoni costituenti un segnale ottico monocromatico (da rivelare) caratterizzato da una polarizzazione (lineare, circolare, ellittica ecc...) perpendicolare rispetto alla direzione di propagazione della perturbazione elettromagnetica. Supponiamo che i fotoni abbiano lunghezza d'onda λ_γ pari a circa 1150 nm, cioè che appartengano al range NIR ("near infra red" - "luce infrarossa vicina"), cioè al range ottico le cui lunghezze d'onda non sono molto superiori ai 700 nm, i quali approssimativamente corrispondono alla luce rossa. Supponiamo che la generazione di una coppia e/h sia imputabile unicamente ad un singolo fotone incidente γ . Il

silicio è un semiconduttore a gap indiretto ($k_* \neq 0$), ovvero l'energia di gap più piccola E_{gap} rintracciabile nel diagramma $E(k)$ si trova in prossimità di una valle diversa da quella Γ ; nel caso particolare del silicio l' E_{gap} è localizzata presso la valle X. Il fotone γ incide sul semiconduttore con una quantità di moto data dalla seguente relazione:

$$\vec{p}_\gamma = \frac{h\nu_\gamma}{c} \hat{k}$$

dove $\nu_\gamma = \omega/2\pi$ è la frequenza del fotone, \hat{k} il versore d'onda. Il fotone γ , affinché il modello delle bande di energia $E - k$ sia congruente, dovrebbe garantire la conservazione dell'energia e della quantità di moto del sistema che γ stesso forma insieme all'elettrone con il quale scatterà. La conservazione dell'energia è soddisfatta, in quanto l'energia di γ , sommata all'energia dell'elettrone prima del salto energetico (prima della transizione, pre - scattering), fornisce l'energia dell'elettrone dopo il salto, ad assorbimento concluso.

$$E_{\text{fotone}} = E_\gamma = E_{\text{gapSi}} = 1,08 \text{ eV} = h\nu_\gamma$$

Tuttavia la conservazione della quantità di moto, considerando un sistema a soli due corpi (fotone - elettrone), non sarebbe soddisfatta, pertanto la transizione elettronica indiretta, mostrata nella parte destra della figura 28, non può avvenire unicamente per intervento del fotone. Come noto nel silicio il minimo della BC è posizionato lungo k_x presso il numero d'onda $k_* \approx 0.85(2\pi/a)$, ovvero quasi al limite superiore della prima zona di Brillouin lungo k_x (si veda il glossario). a rappresenta il passo reticolare, o costante reticolare ("lattice constant"), del silicio. La geometria cristallina del silicio è una struttura costituita, come noto, da due cubi FCC, dove uno è traslato rispetto all'altro lungo la direzione $[111]$ (inclinata di 55° rispetto a $[001]$) individuata dalla diagonale principale d_B ("body diagonal"), di un quarto della lunghezza di d_B ; a è il lato della struttura FCC, pari a 5.428 \AA nel silicio.

Adesso dimostriamo numericamente che il fotone, in virtù della sua lunghezza d'onda λ_γ , e quindi del suo vettore d'onda k_γ , non riesce a conferire all'elettrone, inizialmente localizzato vicino al massimo E_V della BV, una quantità di moto tale da consentirgli di passare da $k = k_0 = 0$ a $k = k_*$.

$$p_\gamma = \hbar k_\gamma = \hbar \left(\frac{2\pi}{\lambda_\gamma} \right) \approx 3.6 \cdot 10^{-9} \text{ eVs/m}$$

$$p_* = \hbar k_* = \hbar \left(0.85 \frac{2\pi}{a} \right) \approx 6.48 \cdot 10^{-6} \text{ eVs/m}$$

Pertanto il fotone γ fornirebbe all'elettrone una quantità di moto p_γ circa 1802 volte più piccola di quella necessaria ($\hbar k_*$) alla transizione elettronica dallo stato iniziale k_0 a quello finale k_* , e ciò è intuibile anche considerando che il fotone è una particella avente un'energia cinetica finita, ma di fatto una quantità di moto trascurabile ($k_\gamma \ll k_*$), essendo nulla la sua massa m_γ . Dunque un fotone consente transizioni elettroniche quasi verticali nel diagramma $E - k$. Grazie al fonone, al suo impulso, l'elettrone, inizialmente sul massimo della banda di valenza, può andare sul minimo della banda di conduzione, lasciando la lacuna al suo posto. Il fonone, ovvero il quanto dell'energia vibrazionale del reticolo cristallino, può essere visto anche come una particella avente bassissima energia cinetica ed elevata quantità di moto. Quindi un fonone consente transizioni elettroniche quasi orizzontali nel diagramma $E - k$. Dunque l'eventuale assorbimento della luce da parte di un semiconduttore a gap indiretto è un processo che per alcune frequenze, cioè per energie fotoniche prossime a quella di gap ($\hbar\omega \approx E_{\text{gap}}$), può avvenire solo con un processo di "ternalizzazione", cioè mediante l'intervento di tre corpi: l'elettrone deve infatti assorbire l'energia del fotone e l'impulso di un modo vibrazionale fononico "favorevole", per poter compiere la transizione indiretta.

Riferendoci alla parte destra di figura 28 possiamo affermare che la transizione indiretta "per", che porta l'elettrone dallo stato iniziale $|i\rangle$ a quello finale $|f\rangle$, è data dalla combinazione della transizione fotonica "ph", che porta l'elettrone dallo stato iniziale $|i\rangle$ a quello intermedio "virtuale" $|n\rangle$, con la transizione fononica "vibr", che porta l'elettrone dallo stato intermedio $|n\rangle$ a quello finale $|f\rangle$; pertanto l'hamiltoniana H_0 dell'elettrone posizionato sul massimo della BV, cioè l'hamiltoniana "di riposo", è perturbata dall'hamiltoniana H_{ph} legata all'assorbimento del fotone e da quella H_{vibr} legata all'assorbimento del fonone. In formule:

$$H_0 = \frac{p^2}{2m_0} + U(\mathbf{r})$$

$$H_0 + H_{ph} = H_{light} = \frac{(p - eA)^2}{2m_0} + U(\mathbf{r}) + e\Phi$$

$$H_{vibr} = -\frac{\hbar^2}{2m_0} (k_* - k_0)^2$$

$$H_{per} = H_0 + H_{ph} + H_{vibr}$$

dove l'espressione di H_{light} è stata ripresa dalla [e27], mentre H_{per} è l'hamiltoniana totale dell'elettrone, il quale, dopo aver scatterato con il fotone e con il fonone, è giunto sul minimo della BC, nel suo stato finale $|f\rangle$. Lo stato elettronico $|n\rangle$ è definito "virtuale" in quanto è posizionato all'interno del gap, cioè in un punto del diagramma $E - k$ in cui un elettrone non può collocarsi; quindi $|n\rangle$ dovrebbe essere uno stato non permesso all'elettrone, eppure nella nostra trattazione $|n\rangle$ deve ospitare l'elettrone quando questo assorbe il fotone e si appresta ad assorbire anche il fonone. Il problema è risolto pensando che l'elettrone si posiziona, nel diagramma $E(k)$, presso lo stato $|n\rangle$, ma per un intervallo di tempo $\Delta t \rightarrow 0$, cioè di fatto l'elettrone fotoperturbato non è mai collocato presso quello stato, ossia gli assorbimenti fotonico e fononico sono praticamente simultanei.

Modificando opportunamente l'espressione della regola d'oro di Fermi, riportata nella [e28], è possibile calcolare il rate W_{absIND} di assorbimento "indiretto" dei fotoni da parte di un semiconduttore a gap indiretto, cioè il rate di assorbimento legato alla transizione elettronica indiretta "per", nel modo seguente:

$$W_{absIND}(|i\rangle) = \frac{2\pi}{\hbar} \iiint \left| \sum_{\beta, \gamma} \frac{\langle f | H_{vibr(\beta)} | n \rangle \langle n | H_{ph(\gamma)} | i \rangle}{E_{|n\rangle} - E_{|i\rangle}} \right|^2 \delta(E_{gap} - \hbar\omega) \frac{d^3k}{(2\pi)^3} \quad [cm^{-3}s^{-1}] \quad [e30]$$

dove nell'argomento della delta di Dirac $|E_f - E_i| = E_{gap}$. Rispetto alla [e28] osserviamo che la sommatoria sugli stati finali $|f\rangle$ è sostituita da un'equivalente e più generica operazione d'integrazione sugli stati elettronici di arrivo $k = (k_x, k_y, k_z)$ (stati finali $k = k_f$), cioè sugli stati, localizzati presso il minimo della BC, disponibili ad accogliere l'elettrone (stati vuoti) che ha appena scatterato con il fotone e con il fonone. Il termine matriciale $\langle n | H_{ph(\gamma)} | i \rangle$ rappresenta la probabilità con la quale l'elettrone transita dallo stato iniziale $|i\rangle$ a quello intermedio $|n\rangle$ a causa della fotoperturbazione relativa ad un singolo fotone incidente γ , mentre il termine matriciale $\langle f | H_{vibr(\beta)} | n \rangle$ rappresenta la probabilità con la quale l'elettrone transita dallo stato intermedio virtuale $|n\rangle$ a quello finale $|f\rangle$ a causa dell'assorbimento di un singolo fonone β . La sommatoria indicizzata da β e γ rappresenta la molteplicità delle cause di scattering, sia ottiche (γ) che termiche (β), che possono consentire all'elettrone di transitare da $|i\rangle$ a $|f\rangle$. La sommatoria parziale su γ indica che la causa della transizione "ph" è riconducibile non solo ad un singolo fotone γ , ma anche ad un intero fascio coerente di fotoni γ , e la probabilità di avere la transizione "ph" è tanto maggiore quanto maggiore è il numero Σ_γ di modi fotonici γ che concorrono al passaggio dell'elettrone da $|i\rangle$ a $|n\rangle$. Analogamente la sommatoria parziale su β indica che la causa della transizione "vibr" è riconducibile non solo ad un singolo fonone β , ma anche ad un insieme di fononi β , e la probabilità di avere la transizione "vibr" è tanto maggiore quanto maggiore è il numero Σ_β di modi fononici β che concorrono al passaggio dell'elettrone da $|n\rangle$ a $|f\rangle$. Il termine $E_{|n\rangle} - E_{|i\rangle}$ rappresenta ovviamente la differenza fra l'energia totale dello stato intermedio $|n\rangle$ e quella dello stato iniziale $|i\rangle$.

La [e30] può essere resa più completa e formale se si tiene conto della natura virtuale, citata prima, dello stato intermedio $|n\rangle$. Come già detto $|n\rangle$ è uno stato virtuale, ossia è uno stato solo virtualmente permesso agli elettroni, quindi l'elettrone può collocarsi in $|n\rangle$ solo per un istante, ovvero per un lasso di tempo $\Delta t \rightarrow 0$. Ciò implica, per il principio di indeterminazione di Heisenberg, che l'energia totale E a cui si trova lo stato $|n\rangle$ sia interessata da un'incertezza $\Delta E \rightarrow \infty$, nella pratica diciamo da un'incertezza ΔE molto grande. Inviando sul semiconduttore a gap indiretto un fascio coerente di fotoni di energia $\hbar\omega$, conosciamo a priori la distanza energetica che separa, nel diagramma $E - k$, lo stato elettronico iniziale $|i\rangle$ da quello finale $|f\rangle$, e tale differenza energetica è pari proprio ad $\hbar\omega$ per la conservazione dell'energia. La conservazione dell'energia è garantita, nella formula [e30] del rate di assorbimento indiretto ed in generale nella regola d'oro di Fermi, dalla delta di Dirac. Pertanto sappiamo a priori dove andrà a collocarsi l'elettrone a transizione conclusa, ma non conosciamo altrettanto esattamente dove l'elettrone andrà a collocarsi un istante dopo lo scattering con il fotone, ossia un istante prima dello scattering con il fonone; in altri termini non conosciamo l'energia totale E di $|n\rangle$. In altre parole la conservazione dell'energia deve essere soddisfatta solo nella transizione elettronica fra i due stati reali $|i\rangle$ e $|f\rangle$ (stati permessi per un Δt finito), cioè nella transizione totale "per"; non è richiesta la conservazione dell'energia nella transizione fra lo stato reale $|i\rangle$ e quello virtuale $|n\rangle$. Quindi se un semiconduttore a gap indiretto assorbe un fotone incidente di energia $\hbar\omega$, la transizione indiretta "per" = "ph + vibr", che porta da $|i\rangle$

ad $|f\rangle$, può avvenire non solo attraverso un'unica combinazione fra una transizione ottica "ph" ed una transizione termica (vibrazionale) "vibr", come sembrerebbe osservando la parte destra di figura 28, bensì attraverso un numero elevatissimo (teoricamente infinito, essendo $\Delta E \rightarrow \infty$) di transizioni combinate "ph + vibr", tutte congiungenti $|i\rangle$ con $|f\rangle$. Questa molteplicità di "percorsi" $|i\rangle \rightarrow |f\rangle$ è dovuta alla grande incertezza ΔE sull'energia E di $|n\rangle$, ovvero alla molteplicità degli stati intermedi virtuali $|n_\alpha\rangle$, come mostrato in figura 28. Possiamo affermare altresì che lo stato finale $|f\rangle$ è deterministico, cioè imposto dalla conservazione dell'energia, mentre lo stato intermedio $|n_\alpha\rangle$ è aleatorio; tale stocasticità è dovuta all'incertezza ΔE .

$$W_{absIND}(|i\rangle) = \frac{2\pi}{\hbar} \iiint \left| \sum_{\alpha,\beta,\gamma} \frac{\langle f | H_{vibr(\beta)} | n_\alpha \rangle \langle n_\alpha | H_{ph(\gamma)} | i \rangle}{E_{|n_\alpha\rangle} - E_{|i\rangle}} \right|^2 \delta(E_{gap} - \hbar\omega) \frac{d^3k}{(2\pi)^3} \quad [\text{cm}^{-3}\text{s}^{-1}] \quad [\text{e31}]$$

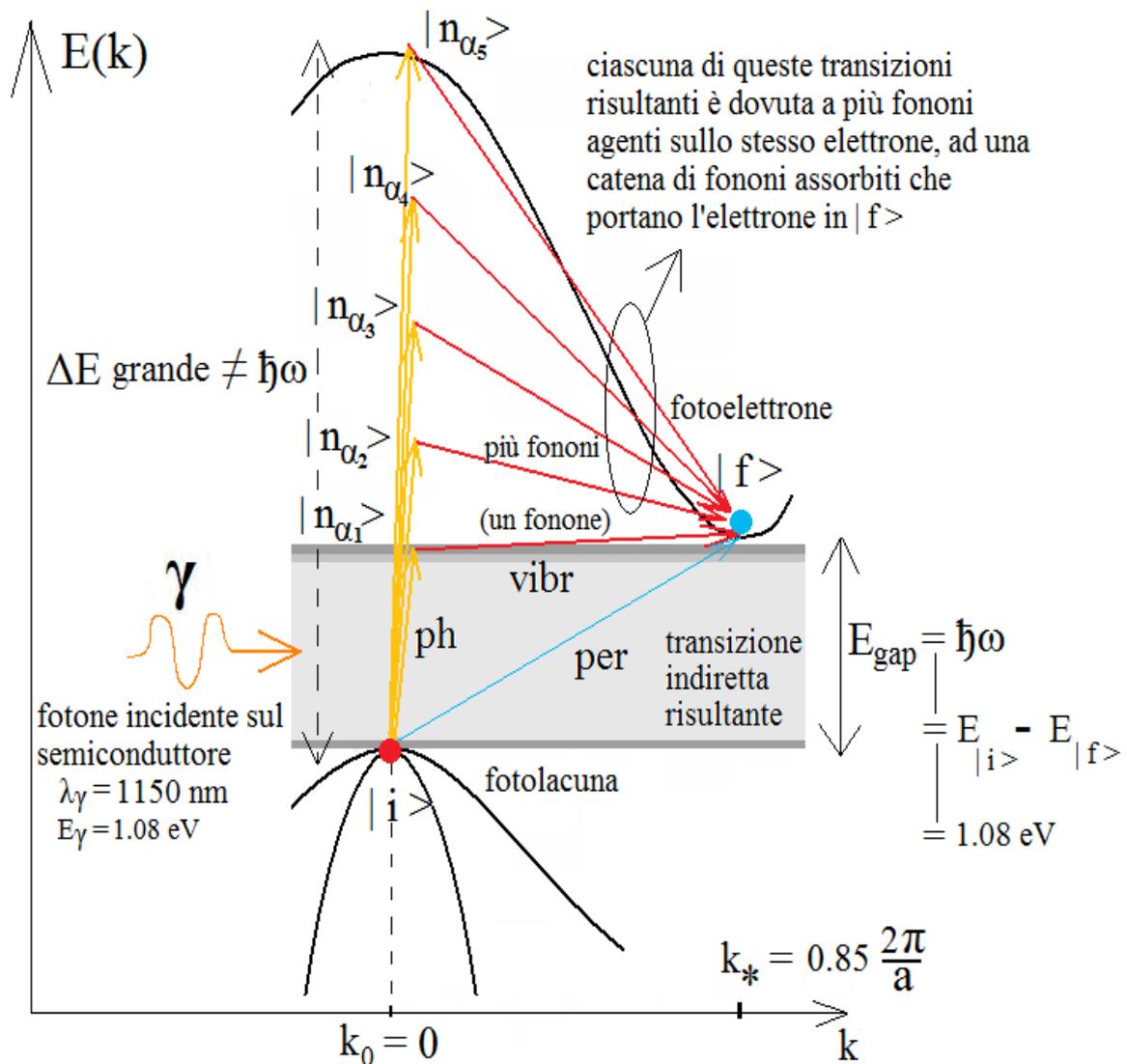


Fig. 29

Rappresentazione semiquantitativa delle molte transizioni $|i\rangle \rightarrow |f\rangle$ composte "ph + vibr" con le quali il silicio, come altri semiconduttori a gap indiretto, può assorbire energie fotoniche prossime a quella di gap (nel caso del silicio $\lambda_\gamma \approx 1150 \text{ nm}$, vicino infrarosso). La molteplicità degli stati intermedi virtuali $|n_\alpha\rangle$ è legata al fatto che un generico stato intermedio $|n\rangle$ può ospitare solo istantaneamente un elettrone proveniente dalla BV, pertanto $\Delta t \rightarrow 0$, quindi $\Delta E \rightarrow \infty$, perciò la conservazione dell'energia non è richiesta per la transizione $|i\rangle \rightarrow |n_\alpha\rangle$. Maggiore è l'energia totale E alla quale si trova lo stato virtuale $|n_\alpha\rangle$ e minore è la probabilità che la transizione totale $|i\rangle \rightarrow |f\rangle$ passi per quel particolare stato intermedio $|n_\alpha\rangle$. Ciascuna transizione $|n_\alpha\rangle \rightarrow |f\rangle$ che non sia approssimativamente orizzontale (nell'esempio tutte quelle circondate dall'ellisse) è legata all'intervento di una "catena" (una successione) di fononi, il cui assorbimento congiunto porta il fotoelettrone sul minimo della BC.

La [e31] differisce dalla [e30] unicamente per un ulteriore indice di sommatoria α , in aggiunta agli indici β e γ già presenti nella [e30]; α indicizza i numerosissimi stati elettronici virtuali $|n_\alpha\rangle$ (teoricamente infiniti) che possono intermediare la transizione indiretta totale “per” $|i\rangle \rightarrow |f\rangle$. Nella sommatoria parziale Σ_α ciascun termine matriciale composto $\langle f | H_{vibr(\beta)} | n_\alpha \rangle \langle n_\alpha | H_{ph(\gamma)} | i \rangle$ è “pesato” da una sorta di “coefficiente di probabilità” P_α , definito nel modo seguente:

$$P_\alpha = \frac{1}{E_{|n_\alpha\rangle} - E_{|i\rangle}}$$

Per un determinato valore dell’energia fotonica $\hbar\omega \geq E_{gap}$, incidente sul semiconduttore a gap indiretto, (nell’esempio di figura 29 abbiamo posto $\hbar\omega = E_{gap}$), possiamo affermare che maggiore è l’energia totale E a cui si trova $|n_\alpha\rangle$ e minore è la probabilità P_α che la transizione indiretta totale “per” $|i\rangle \rightarrow |f\rangle$ sia mediata da quel particolare stato intermedio virtuale $|n_\alpha\rangle$, ossia minore è il contributo della transizione composta $|i\rangle \rightarrow |n_\alpha\rangle \rightarrow |f\rangle$, cioè del rate parziale $W_{absIND(\alpha)}(|i\rangle)$, al rate totale $W_{absIND}(|i\rangle)$. Il rate totale $W_{absIND}(|i\rangle)$ può essere visto, infatti, come la sommatoria Σ_α dei rates parziali $W_{absIND(\alpha)}(|i\rangle)$, ciascuno relativo ad una singola transizione (percorso “ph + vibr”) $|i\rangle \rightarrow |n_\alpha\rangle \rightarrow |f\rangle$, pesati con i coefficienti P_α . In formule:

$$W_{absIND}(|i\rangle) = \sum_{\alpha=1}^{\infty} P_\alpha W_{absIND(\alpha)}(|i\rangle) \quad [e32]$$

La [e32], che rappresenta una forma compatta della [e31], sintetizza il concetto che sta alla base della situazione di scattering mostrata in figura 28; tale situazione di scattering può essere modellizzata utilizzando una semplice analogia elettrica, mostrata in figura 30.

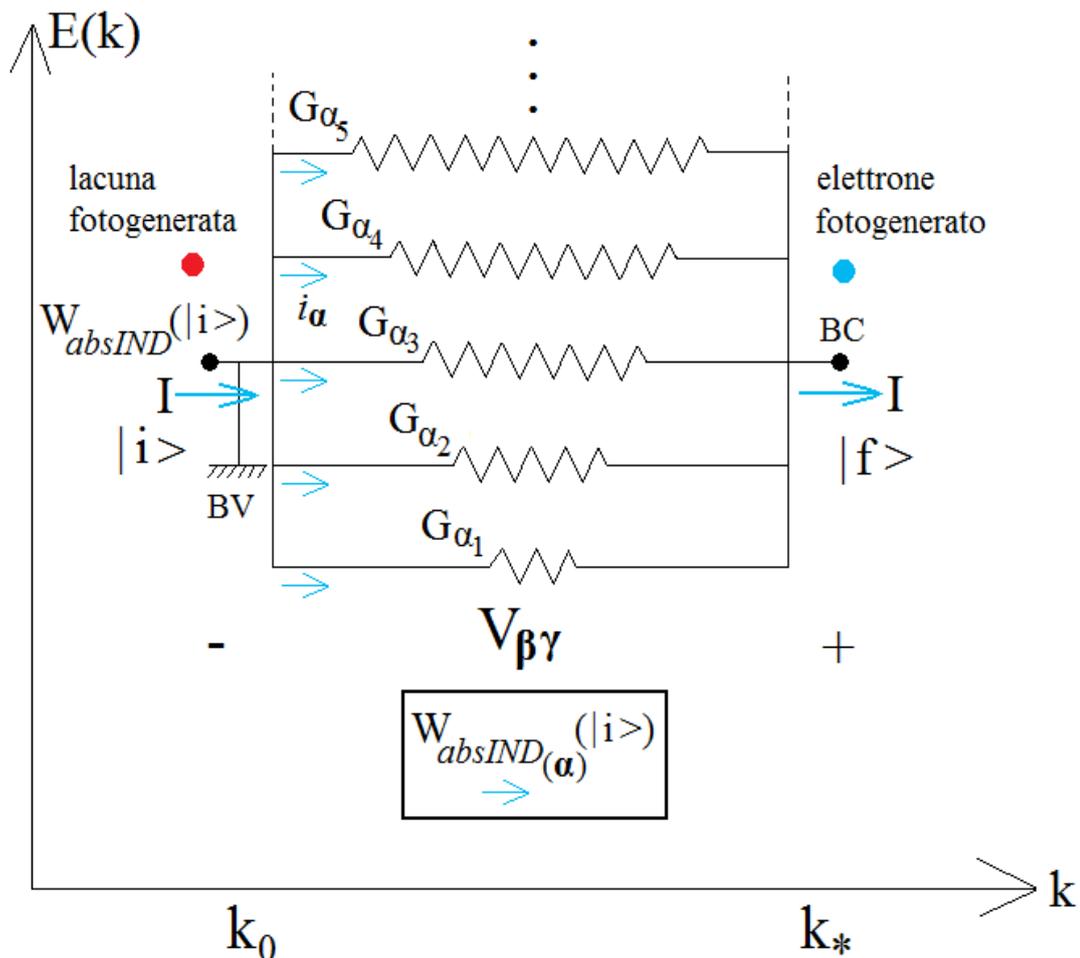


Fig. 30

Rappresentazione schematica del macrobipolo resistivo, utilizzato come partitore di corrente, che consente di modellizzare elettricamente le moltissime varianti con le quali può verificarsi l’assorbimento di un fotone, di energia $\hbar\omega \geq E_{gap}$, da parte di un semiconduttore a gap indiretto. L’esempio è riferito al silicio; l’energia fotonica considerata è $\hbar\omega = E_{gap}$.

$$W_{absIND}(|i\rangle) = I = \sum_{\alpha=1}^{\infty} i_{\alpha} = \sum_{\alpha=1}^{\infty} G_{\alpha} V_{\beta\gamma}$$

In figura 30 è mostrato un macrobipolo elettrico puramente resistivo, costituito da tanti bipoli resistivi montati in parallelo, i cui due nodi indipendenti sono quello di massa BV e quello BC a cui applichiamo la tensione $V_{\beta\gamma}$; il macrobipolo viene utilizzato come partitore della corrente I , prodotta dalla sollecitazione $V_{\beta\gamma}$. La corrente I si diparte nei vari bipoli resistivi proporzionalmente alla grandezza di ciascuna conduttanza G_{α} , ovvero in modo inversamente proporzionale alla grandezza della resistenza $R_{\alpha} = 1/G_{\alpha}$. Nel disegno di figura 30 G_{α} è tanto minore quanto maggiore è la lunghezza del simbolo elettrico convenzionale della resistenza, lunghezza che è proporzionale ad $R_{\alpha} = 1/G_{\alpha}$. Man mano che scorriamo, lungo il suo verso positivo, il vettore $E(k)$ delle energie elettroniche totali, ovvero man mano che consideriamo stati intermedi virtuali $|n_{\alpha}\rangle$ ad energie sempre maggiori, osserviamo una diminuzione di P_{α} e quindi una diminuzione del “contributo” delle componenti $W_{absIND(\alpha)}(|i\rangle)$ al rate totale $W_{absIND}(|i\rangle)$. Ciascun contributo $W_{absIND(\alpha)}(|i\rangle)$, pesato da P_{α} , è rappresentato, nell’analogia elettrica, dalla corrente i_{α} che attraversa la conduttanza G_{α} , mentre $W_{absIND}(|i\rangle)$ è la corrente totale I che attraversa il macrobipolo. Il passaggio della corrente elettronica I dal nodo di massa BV, che potrebbe rappresentare la banda di valenza, al nodo indipendente BC, che potrebbe rappresentare la banda di conduzione, corrisponde, nell’analogia elettrica, al passaggio dell’elettrone dallo stato iniziale $|i\rangle$ a quello finale $|f\rangle$, ovvero alla transizione indiretta $|i\rangle \rightarrow \sum_{\alpha} P_{\alpha} |n_{\alpha}\rangle \rightarrow |f\rangle$. Il termine $\sum_{\alpha} P_{\alpha} |n_{\alpha}\rangle$ può essere visto come la somma delle varie correnti i_{α} pesate con le conduttanze G_{α} . Quindi ciascuna G_{α} rappresenta, nell’analogia elettrica, la probabilità P_{α} , cioè l’inverso di $E_{|n_{\alpha}\rangle} - E_{|i\rangle}$. Il contributo più importante al rate totale $W_{absIND}(|i\rangle)$, ovvero alla corrente totale I , è fornito dal rate parziale $W_{absIND(\alpha1)}(|i\rangle)$, ovvero dalla corrente parziale $i_{\alpha1}$, relativa alla transizione composta $|i\rangle \rightarrow |n_{\alpha1}\rangle \rightarrow |f\rangle$ che è mostrata nella parte destra di figura 28. L’espressione di $W_{absIND(\alpha1)}(|i\rangle)$ è la seguente:

$$W_{absIND(\alpha1)}(|i\rangle) = \frac{2\pi}{\hbar} \iiint \left| \sum_{\beta,\gamma} \frac{\langle f | H_{vibr(\beta)} | n_{\alpha1} \rangle \langle n_{\alpha1} | H_{ph(\gamma)} | i \rangle}{\{ \hbar\omega \}} \right|^2 \delta(E_{gap} - \hbar\omega) \frac{d^3k}{(2\pi)^3} \quad [e33]$$

con $\hbar\omega = E_{gap}$. I rates parziali $W_{absIND(\alpha2)}(|i\rangle)$, $W_{absIND(\alpha3)}(|i\rangle)$, $W_{absIND(\alpha4)}(|i\rangle)$ ecc..., così come le correnti parziali $i_{\alpha2}$, $i_{\alpha3}$, $i_{\alpha4}$ ecc..., forniscono un contributo via via minore al rate totale $W_{absIND}(|i\rangle)$, ossia alla corrente totale I , poiché l’energia dello stato intermedio virtuale $E = E_{|n_{\alpha}\rangle}$ va aumentando, il denominatore fra parentesi graffe nella [e33] diventa sempre più grande di $\hbar\omega$ e quindi P_{α} e G_{α} decrescono. Così come nel circuito utilizzato nell’analogia elettrica è necessario applicare una sollecitazione, ovvero un potenziale $V_{\beta\gamma}$ positivo sul nodo BC, così da produrre una corrente elettronica dal nodo di massa BV al nodo BC, analogamente nel semiconduttore per promuovere un elettrone dallo stato iniziale $|i\rangle$ a quello finale $|f\rangle$ è necessario fornirgli una sollecitazione $V_{\beta\gamma}$. Questa consta di una componente fononica γ e di una fononica β . Va notato un aspetto riguardante la componente fononica β , che produce le transizioni $|n_{\alpha}\rangle \rightarrow |f\rangle$ colorate di rosso in figura 29. La transizione fononica “vibr”, che fa parte della transizione $|i\rangle \rightarrow |f\rangle$ mediata dallo stato virtuale $|n_{\alpha1}\rangle$, che è quello a più bassa energia, è dovuta ad un singolo fonone assorbito dall’elettrone, il quale ha appena scatterato con un fotone. La transizione “vibr” è associabile ad un solo fonone poiché nel diagramma $E - k$ si presenta praticamente orizzontale. Tutte le altre transizioni fononiche $|n_{\alpha}\rangle \rightarrow |f\rangle$, che nel disegno di figura 28 sono state evidenziate dall’ellisse, sono tutt’altro che orizzontali; ciascuna di queste copre un salto energetico $E_{|n_{\alpha}\rangle} - E_C$ abbastanza grande, che non può essere attribuito all’assorbimento di un singolo fonone da parte dell’elettrone appena fotostimolato. Pertanto ciascuna di quelle transizioni fononiche $|n_{\alpha}\rangle \rightarrow |f\rangle$ è da attribuirsi all’assorbimento di una “catena” di fononi, cioè una successione di modi vibrazionali favorevoli; questa successione è tanto più lunga quanto maggiore è il salto energetico $E_{|n_{\alpha}\rangle} - E_C$, ovvero quanto maggiore è l’energia dello stato virtuale intermedio $|n_{\alpha}\rangle$. Tale spiegazione conferma il fatto che maggiore è l’energia $E = E_{|n_{\alpha}\rangle}$ e minore è la probabilità che la transizione indiretta totale $|i\rangle \rightarrow |f\rangle$ passi attraverso $|n_{\alpha}\rangle$. Infatti maggiore è $E_{|n_{\alpha}\rangle}$ e maggiore è il numero di fononi che congiuntamente devono garantire il passaggio dell’elettrone dallo stato $|n_{\alpha}\rangle$ al minimo della banda di conduzione. Chiaramente maggiore è il numero di fononi che devono essere coinvolti nella transizione $|n_{\alpha}\rangle \rightarrow |f\rangle$ e minore è la probabilità P_{α} che la transizione $|i\rangle \rightarrow |n_{\alpha}\rangle \rightarrow |f\rangle$ si verifichi.

4.5) Fotodiodi classici e fotodiodi a valanga

Irradiando con un segnale ottico di opportuna frequenza la superficie attiva di un fotorivelatore a semiconduttore, polarizzato inversamente, otteniamo un segnale elettrico, formato dai portatori primari di carica fotogenerati nella zds^* , che è proporzionale all'intensità luminosa incidente; pertanto l'ampiezza del segnale elettrico è portatrice dell'informazione circa l'intensità della luce. Questo è il segnale elettrico "di base": le fotocariche primarie vengono trascinate dal campo della zds^* e raccolte presso x_n^* e $-x_p^*$, e da lì nei rispettivi bulk. A seconda di come "trattiamo" il fotosegnale di base possiamo distinguere due fondamentali famiglie di fotodiodi:

- fotodiodi classici
- fotodiodi a valanga

Nei primi il punto di riposo si trova, con riferimento alla figura 23, a destra di V_{BD} , per cui P tende a 0 e non si ha moltiplicazione di corrente (guadagno interno unitario); le cariche primarie fotogenerate vengono quindi convogliate dal campo senza produrre ionizzazioni. Il segnale elettrico di uscita è talmente basso che occorrono stadi amplificatori (esterni) in cascata, atti a trasformarlo in un segnale macroscopico intelligibile, ovvero processabile da un'elettronica. I secondi (di ultima generazione) amplificano internamente il fotosegnale di base: alcuni tra questi potrebbero essere impiegati nella PET. Il loro punto di riposo si trova a sinistra di V_{BD} , pertanto la zds^* è interessata da un campo sufficientemente elevato da consentire alle fotocariche primarie di produrre ionizzazioni a valanga. In questo caso si ha $P \neq 0$, M abbastanza alto e quindi guadagno interno elevato, per cui il fotosegnale di base, ricevendo un'amplificazione intrinseca, non deve passare necessariamente attraverso amplificatori esterni, i quali degraderebbero parametri come, ad esempio, la risoluzione temporale. Osservando la regione della caratteristica $I - V$ compresa fra V_{BD}^* e V_{BD} possiamo individuare un particolare valore di voltaggio inverso (V^*), il quale costituisce una sorta di "discriminante" fra due tipi di fotodiodi appartenenti alla seconda famiglia: quelli polarizzati con tensioni $|V| < V^*$ (cioè $V_{BD} < |V| < V^*$) e quelli con tensioni $|V| > V^*$. I primi sono noti come APDs ("Avalanche Photo Diodes"), i secondi come GM - APDs ("Geiger-Mode - Avalanche Photo Diodes"). Gli APDs sfruttano l'effetto valanga per moltiplicare la corrente di fotocariche primarie, ma il fattore moltiplicativo M non è altissimo (va da qualche decina a qualche centinaia, ovvero fra 10 e 20 dB), ragion per cui il fotosegnale utile, costituito dalla valanga dovuta agli eventi ionizzanti, è proporzionale al numero di fotoni interagenti col semiconduttore. Dunque gli APDs forniscono un'informazione elettrica proporzionale all'intensità luminosa incidente: misurando la corrente di uscita è possibile risalire al numero di fotoni assorbiti nella zds^* , e ciò costituisce un indiscutibile pregio. Un grosso difetto degli APDs può essere rappresentato dal basso rapporto segnale/rumore (SNR) presente ai terminali metallici di uscita, specialmente per potenze ottiche incidenti elevate: il campo elettrico nella zds^* , oltre ad accelerare i fotoportatori primari ed amplificare il segnale di base, presentando in uscita una corrente di ionizzazione elevata (che è quella di interesse, poiché è proporzionale all'intensità luminosa), estrae i portatori minoritari dalle zone di confine fra zds^* e zone neutre, e li accelera, permettendo a questa corrente di estrazione di intensificarsi. Dunque in uscita abbiamo sia la fotocorrente che costituisce il segnale di interesse, sia una corrente di dark - bulk, formata da cariche, primarie e (soprattutto) secondarie, che non hanno nulla a che fare con la luce entrata nel fotodiodo. Questa è una corrente di rumore, dal momento che non è determinata dal numero di fotoni assorbiti (è una corrente originata da portatori minoritari termogenerati nei bulk e diffusi fino a x_n^* e $-x_p^*$), la quale si sovrappone al segnale utile.

Nel capitolo seguente descriveremo il funzionamento dei fotodiodi "classici", ovvero dei fotorivelatori privi di guadagno interno ($M = 1$, nessun fenomeno di moltiplicazione a valanga), i quali costituiscono una sorta di "contatori di fotoni". In particolare ci soffermeremo sui fotodiodi, largamente usati nell'ambito delle comunicazioni ottiche su fibra, noti sotto il nome di "PIN" (semiconduttore drogato p/semiconduttore intrinseco/semiconduttore drogato n), sia ad omostruttura che ad eterostruttura, i quali consentono, anche in virtù del loro ottimo valore di efficienza quantica QE, di avere un buon SNR, soprattutto per alte potenze ottiche incidenti.

5) I fotodiodi PIN

5.1.1) Le principali modalità di interazione fra luce e semiconduttore: descrizione qualitativa

Per capire l'utilità, nonché i limiti, del fotorivelatore PIN è necessario prima discutere le modalità con le quali la potenza ottica incidente viene assorbita dal materiale costituente un generico fotodiode ed individuare le zone del dispositivo in cui l'assorbimento fotonico è vantaggioso e quelle in cui lo è meno. Prima ancora è opportuno ricordare le quattro modalità principali con le quali la luce interagisce con la materia (un semiconduttore, nel nostro caso), ovvero le quattro possibili interazioni (scatterings) tra i fotoni e gli elettroni atomici del cristallo investito dalla radiazione luminosa; tali interazioni comportano, come già visto genericamente nel paragrafo 4.3, delle transizioni elettroniche fra stati diversi ($|i\rangle \rightarrow |f\rangle$). Due di queste transizioni elettroniche sono spontanee, puramente quantistiche, ossia avvengono in assenza di perturbazioni luminose provenienti dall'esterno, mentre le altre avvengono solo in seguito ad una "stimolazione" da parte della luce incidente.

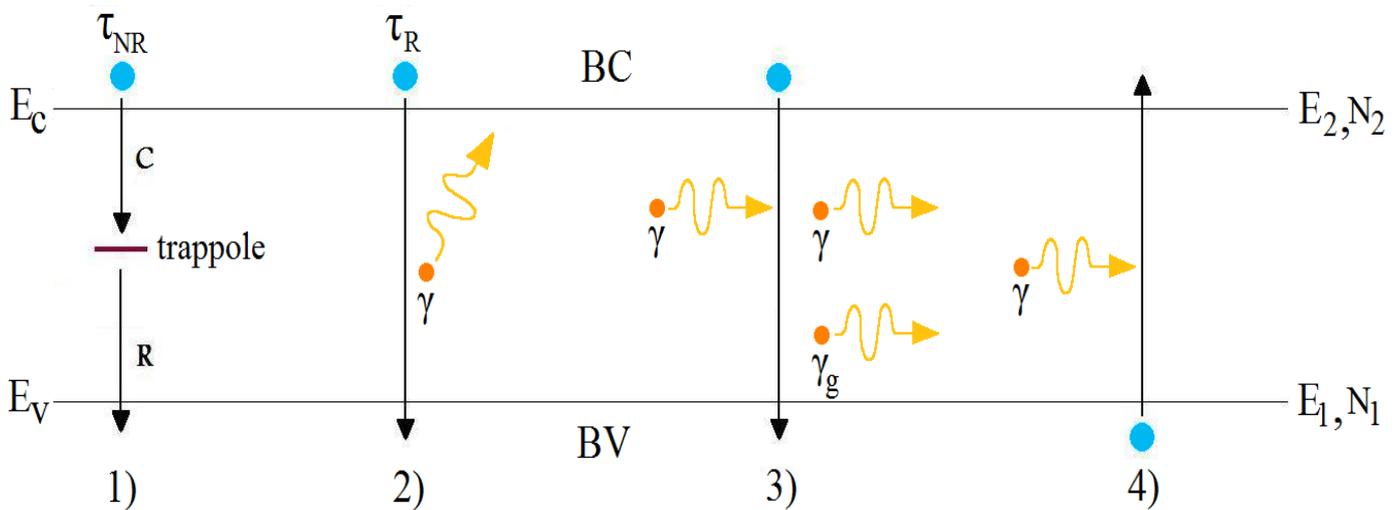


Fig. 31

Rappresentazione, mediante un sistema energetico a due livelli facilmente adattabile al più complesso modello a bande $E(k)$ valido per i semiconduttori, delle quattro interazioni possibili fra luce e materia: 1) decadimento spontaneo non radiativo, 2) decadimento spontaneo radiativo, 3) decadimento stimolato radiativo, 4) assorbimento fotonico.

1) Decadimento spontaneo non radiativo

Gli elettroni che si trovano su un livello energetico E_2 (E_2 è un livello appartenente alla BC, nel caso in cui il materiale sia un semiconduttore) hanno un tempo di vita medio τ_{NR} (NR = non radiativo), per cui gli elettroni stazionano su quel livello energetico, ovvero possiedono quel valore di energia totale, mediamente per un intervallo di tempo pari a τ_{NR} , durante il quale gli elettroni rappresentano le N_2 particelle che popolano il livello E_2 (nel nostro modello N_2 ed N_1 costituiscono delle concentrazioni di elettroni). Al termine di τ_{NR} gli elettroni decadono spontaneamente ad un livello energetico inferiore E_1 , più stabile di E_2 . Gli elettroni non raggiungono il livello stabile E_1 direttamente, ossia con un unico salto energetico, cioè con un'unica transizione $|i\rangle \rightarrow |f\rangle$, bensì attraverso processi di "cattura (C) e rilascio (R)" da parte di stati energetici profondi ("deep levels"), ovvero trappole di varia natura, alle quali gli elettroni cedono progressivamente la differenza di energia $E_2 - E_1$. La differenza di energia viene dissipata non mediante emissione di fotoni, bensì attraverso dei processi termici, quali ad esempio le vibrazioni modali delle impurezze (ad esempio le vibrazioni di metalli pesanti come Au, Cu, Fe, Cr, Cd e Co, posizionati nel cristallo semiconduttore in modo interstiziale) e dei vari difetti reticolari (di punto o di linea) che determinano le trappole energetiche profonde, ovvero stati permessi all'interno del gap. Nel caso dei semiconduttori si parla di "ricombinazione (elettrone/lacuna) non radiativa banda \rightarrow trappole

→ banda” o anche di “ricombinazione SRH (Shockley – Read – Hall)”. Nel caso in cui ci sia un solo “livello energetico intrappolatore” E_T a mediare il decadimento degli elettroni da E_2 a E_1 , è possibile scrivere la seguente formula per il rate di cattura (R_C) di elettroni da parte delle trappole:

$$R_C = N_2 \sigma_T v_{th} N_T (1 - f(E_T)) \quad [\text{cm}^{-3} \text{s}^{-1}]$$

Si noti come la probabilità con cui le trappole catturano gli elettroni è proporzionale:

- alla concentrazione N_2 di elettroni sul livello energetico instabile E_2 , che nei semiconduttori coincide con la concentrazione di elettroni in BC
- alla sezione di cattura σ_T , per gli elettroni, dei centri di intrappolamento
- alla velocità termica v_{th} degli elettroni (più questi sono veloci, maggiore è il numero di trappole “viste”, “visitare”, da ciascun elettrone nell’unità di tempo)
- alla concentrazione di trappole libere $N_T(1 - f(E_T))$: N_T è la concentrazione totale di trappole, ipotizzando che siano tutte equienergetiche, mentre $f(E_T)$ è il valore assunto dalla funzione di distribuzione di Fermi – Dirac all’energia del livello intrappolatore E_T .

In modo del tutto analogo è possibile scrivere la seguente formula per il rate di rilascio (R_R) di elettroni da parte delle trappole, assumendo sempre che ci sia un solo livello energetico intrappolatore E_T a mediare il decadimento. Tuttavia è preferibile pensare al rilascio di elettroni come all’intrappolamento di lacune provenienti da E_1 (dalla BV), pertanto:

$$R_R = p \sigma_T' v_{th} N_T f(E_T) \quad [\text{cm}^{-3} \text{s}^{-1}]$$

Si noti come la probabilità con cui le trappole catturano le lacune, ovvero rilasciano gli elettroni, è proporzionale:

- alla concentrazione p di lacune sul livello energetico E_1 , ossia in BV
- alla sezione di cattura σ_T' , per le lacune, dei centri di intrappolamento
- alla velocità termica v_{th} delle lacune del livello energetico E_1 , ovvero della BV (più le lacune sono veloci, ossia più gli N_1 elettroni di E_1 sono veloci, maggiore è il numero di trappole “visitare” da ciascuna lacuna nell’unità di tempo)
- alla concentrazione di trappole occupate (da elettroni) $N_T f(E_T)$: N_T è la concentrazione totale di trappole, sempre ipotizzando che siano tutte equienergetiche, mentre $f(E_T)$ è il valore assunto dalla Fermi – Dirac all’energia del livello intrappolatore E_T .

Il decadimento spontaneo non radiativo, fin qui descritto, è molto più frequente nei semiconduttori a gap indiretto piuttosto che in quelli a gap diretto. Ad esempio per valori standard delle concentrazioni di elettroni iniettati in BC e di lacune iniettate in BV ($10^{-14} < n = p < 10^{-16}$), all’interno della zona di svuotamento zds di molti tipi di LEDs, abbiamo che mediamente nel silicio (gap indiretto) solo una coppia elettrone/lacuna su centomila da luogo, ricombinandosi, ad emissione spontanea (tutte le altre si ricombinano in modo non radiativo), mentre nell’arseniuro di gallio (gap diretto) mediamente una coppia su due decade spontaneamente in modo radiativo. Quando la transizione (non radiativa) di un elettrone da E_2 a E_1 è conclusa, la popolazione N_2 è decrementata di un’unità, mentre la popolazione N_1 è incrementata di un’unità. È intuitivo scrivere il contributo della transizione 1) all’equazione di continuità del sistema, a due livelli energetici, da noi utilizzato:

$$\frac{dN_2(t)}{dt} = - C_{21} N_2(t) = - \frac{N_2(t)}{\tau_{NR}} \quad [\text{cm}^{-3} \text{s}^{-1}]$$

dove C_{21} è uno dei noti “coefficienti di Einstein”. La ricombinazione elettrone/lacuna spontanea non radiativa è uno dei maggiori problemi riguardanti il corretto funzionamento dei dispositivi emettitori di

luce, quali LEDs e LASERS. Un LED, almeno in condizioni normali (ad esempio a temperatura ambiente), non può funzionare se costruito con un semiconduttore a gap indiretto: le ricombinazioni non radiative banda → trappole → banda, molto più frequenti di quelle radiative (come già detto di un fattore pari a circa 10^5 , nel caso del silicio e di transizioni $E_c \rightarrow E_v$), fra elettroni in BC e lacune in BV iniettati, nella zds della giunzione pn, dalla corrente di polarizzazione diretta, non produrrebbero fotoni, dunque la giunzione non emetterebbe luce con intensità I_v sufficiente per poter essere percepita dall'occhio umano. Il decadimento spontaneo non radiativo va dunque ad incidere negativamente sull'efficienza interna η_{int} di un LED. Le ricombinazioni non radiative sono altresì una delle cause principali della bassa efficienza di emissione (in genere < 50%) che può interessare un LASER realizzato con delle tecniche non efficaci nel minimizzare la concentrazione d'impurezze all'interno della zona attiva appartenente al "core" (nucleo) del dispositivo. I decadimenti SRH implicano, infatti, un inutile consumo di inversione di popolazione, ovvero uno "spreco" del pompaggio elettronico, a cui deve essere sottoposto un LASER a semiconduttore, per poter "guadagnare" in termini di intensità luminosa e quindi emettere luce coerente apprezzabile all'esterno.

Esiste un'altra forma di decadimento spontaneo non radiativo, abbastanza rara, nota come "ricombinazione Auger" elettrone/lacuna; per descriverla è opportuno richiamare brevemente il seguente concetto di fisica dei semiconduttori. Per prima cosa consideriamo un volume macroscopico di semiconduttore a gap indiretto (silicio, ad esempio), le cui dimensioni sono $L_x = a N_x$, $L_y = a N_y$ ed $L_z = a N_z$, a rappresenta la costante reticolare del semiconduttore scelto (pari a qualche \AA), mentre N_j ($j = x, y, z$) è il numero di celle fondamentali di cui consta la lunghezza del cristallo semiconduttore lungo l'asse j considerato (N_j determina la lunghezza del cristallo lungo l'asse j). Come noto la prima zona di Brillouin lungo gli assi k_x , k_y e k_z del reticolo reciproco di Fourier (si veda il glossario), ovvero lungo le direzioni [100], [010] e [001], per un semiconduttore avente struttura cristallina diamantata (semiconduttori elementari) oppure a zincoblenda (semiconduttori composti III - V), ha un'estensione pari a $2\pi/a$. Quindi l'estensione della prima zona di Brillouin nello spazio dei vettori d'onda è una caratteristica intrinseca di ciascun semiconduttore, poiché lo è la costante reticolare a del cristallo. Il passo di quantizzazione Δk_{qj} ($= 2\pi/L_j$) all'interno del reticolo reciproco è, invece, determinato dalle dimensioni macroscopiche L_x , L_y ed L_z del cristallo in cui sono confinati gli elettroni, pertanto la sagomatura del cristallo determina la concentrazione degli stati elettronici permessi lungo le linee $E(k)$ del diagramma a bande, come mostrato qualitativamente in figura 32 (gli stati permessi in BC sono indicati con dei punti neri). La quantizzazione dei numeri d'onda lungo uno dei tre assi è ricavabile imponendo le note condizioni periodiche al contorno di Born - Von Karman, e ciò è possibile in virtù della natura non scatterante degli elettroni, all'interno di un cristallo "perfetto" (non realmente esistente); questa considerazione è una diretta conseguenza del teorema di Bloch, che nella teoria garantisce assenza di scattering, cioè elevata mobilità (cioè la delocalizzazione degli stati elettronici lungo tutto il volume del cristallo), per i portatori di carica all'interno di una matrice cristallina periodica perfettamente ordinata. Supponiamo quindi di aver sagomato il cristallo di silicio in modo che questo abbia una lunghezza, lungo l'asse x , pari a L_x . In tal modo abbiamo fissato anche la distanza Δk_{qx} , lungo l'asse k_x , tra due stati elettronici contigui ($\Delta k_{qx} = 2\pi/L_x$).

La ricombinazione Auger prevede la partecipazione di tre corpi, ad esempio due elettroni ed una lacuna oppure un elettrone e due lacune. Supponiamo di avere due elettroni in BC ed una lacuna in BV: se la posizione degli elettroni sulla curva $E(k)$ lo consente, esiste la possibilità che uno dei due elettroni (e_1) ceda all'altro (e_2) una certa quantità di energia ΔE ed una certa quantità di moto $\Delta P = \hbar\Delta k$, raggiungendo così un nuovo punto della curva $E(k)$; questo punto individua uno stato elettronico dal quale è possibile la ricombinazione (generalmente non radiativa) con la lacuna che si trova in BV. L'elettrone e_2 raggiunge un nuovo punto di $E(k)$ individuato dagli incrementi ΔE e Δk . La particolare situazione energetica ed impulsiva in cui devono trovarsi simultaneamente e_1 ed e_2 , nonché la quantizzazione Δk_{qj} del numero d'onda k_j (i diagrammi $E - k$ sono disegnati con tratti continui solo per semplicità grafica), rendono di fatto trascurabile il rate di ricombinazione Auger, in condizioni di temperatura e drogaggio ordinarie.

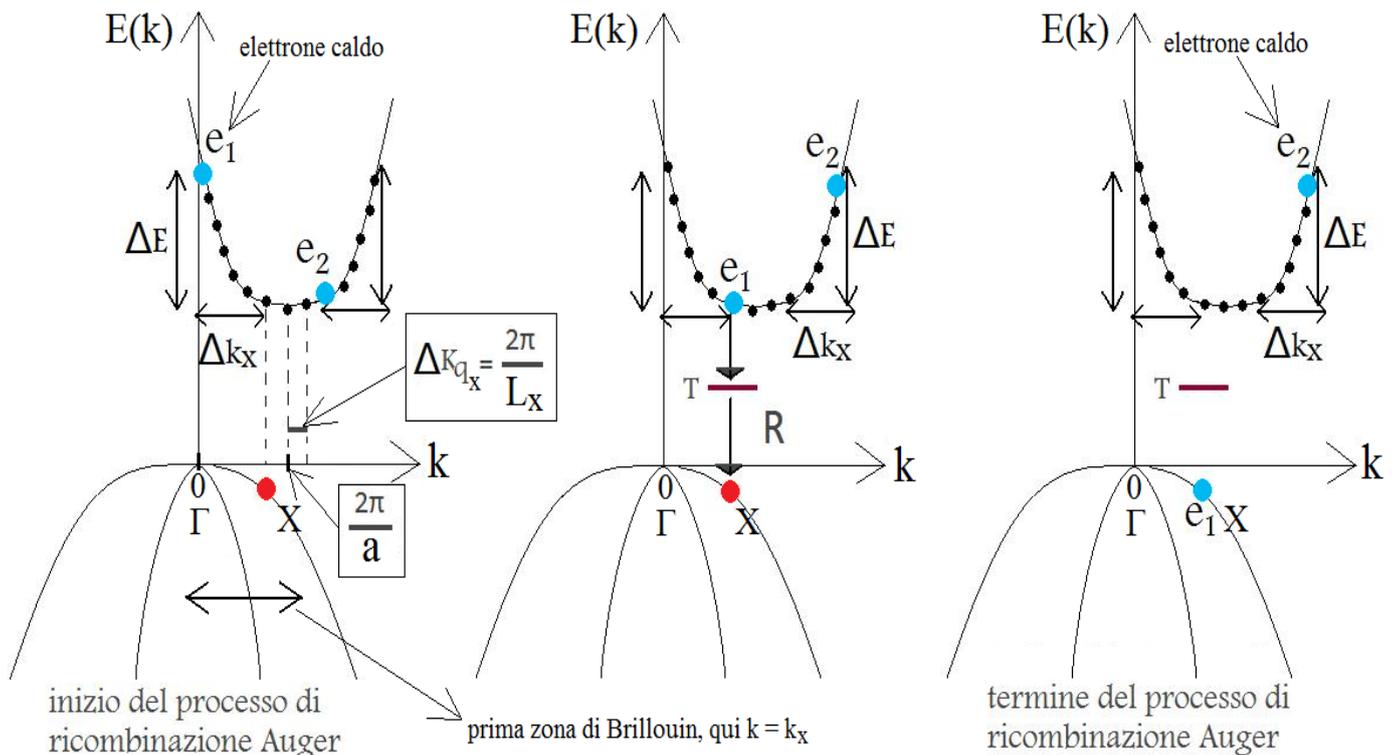


Fig. 32

Rappresentazione del processo, molto poco probabile, di ricombinazione Auger in un semiconduttore a gap indiretto. Si noti la dinamica energetico - impulsiva che coinvolge i tre corpi, nonché la quantizzazione Δk_{qx} del vettore d'onda k ($= k_x$) presso la prima zona di Brillouin lungo l'asse k_x , a è la costante reticolare del semiconduttore.

2) Decadimento spontaneo radiativo

Gli elettroni che si trovano sul livello energetico E_2 hanno un tempo di vita medio τ_R (R = radiativo), per cui questi stazionano su quel livello mediamente per un intervallo di tempo pari a τ_R , durante il quale gli elettroni sono le N_2 particelle che popolano il livello energetico E_2 . Al termine di τ_R gli elettroni decadono spontaneamente ad un livello energetico inferiore E_1 . Gli elettroni raggiungono il livello stabile E_1 direttamente, ovvero con un unico salto energetico, cedendo l'energia $E_2 - E_1$ sotto forma di un fotone γ , la cui energia è proprio $h\nu_\gamma = E_2 - E_1$. Il fotone, emesso spontaneamente, si allontana dal punto in cui è avvenuta la ricombinazione elettrone/lacuna. La sua frequenza ν_γ è nota, almeno teoricamente, con una certa accuratezza, in virtù della conservazione dell'energia, ma lo stesso non si può dire per la fase, la polarizzazione, la direzione ed il verso di propagazione dell'onda elettromagnetica associata al fotone (c'è una grande incertezza sul valore dei numeri quantici assunti dal fotone emesso spontaneamente), motivo per cui la luce prodotta dal decadimento spontaneo radiativo è ritenuta di una forma "poco pregiata" (questa è la luce prodotta dai LEDs, ad esempio, dispositivi molto più economici dei LASERS). La ricombinazione radiativa, nota anche come "ricombinazione banda \rightarrow banda", è molto più frequente nei semiconduttori a gap diretto: in quelli a gap indiretto è necessaria l'intermediazione di un modo vibrazionale fononico, per la conservazione della quantità di moto. È intuitivo scrivere il contributo della transizione 2) all'equazione di continuità del sistema, a due livelli, da noi utilizzato:

$$\frac{dN_2(t)}{dt} = -A_{21}N_2(t) = -\frac{N_2(t)}{\tau_R} \quad [\text{cm}^{-3} \text{s}^{-1}]$$

dove A_{21} è uno dei noti "coefficienti di Einstein". Nei semiconduttori a gap indiretto $\tau_{NR} \ll \tau_R$, mentre in quelli a gap diretto $\tau_{NR} \gg \tau_R$. Il tempo medio di decadimento spontaneo radiativo τ_R , per molti dei materiali (semiconduttori e non) usati in ambito optoelettronico, è abbastanza lungo, se confrontato con i tempi di commutazione caratteristici degli attuali sistemi di comunicazione ottici, per i quali il "modulation rate" elettronico può raggiungere l'ordine dei GHz. Quindi un valore abbastanza alto di τ_R comporta una forte limitazione della banda del dispositivo emettitore. Basti pensare che τ_R per un

LASER al rubino Al_2O_3 , drogato con una concentrazione di cromo pari a circa $1.58 \cdot 10^{19} \text{ cm}^{-3}$, è di circa 3ms, che è un tempo medio di decadimento spontaneo assolutamente incompatibile con il modulation rate sopra citato. Per il InGaAsP, che costituisce il core dei LASERs più utilizzati, attualmente, nell'ambito delle comunicazioni ottiche su grande distanza (terza finestra di trasmissione su fibra ottica di silice), τ_R è generalmente compreso fra 1 e 5 ns.

3) Emissione stimolata (Decadimento stimolato radiativo)

Se un fotone γ di energia $h\nu_\gamma = E_2 - E_1$, nell'attraversare il materiale cristallino, entra all'interno della sezione di cattura di un elettrone atomico che si trova sul livello energetico eccitato E_2 , questo elettrone può essere stimolato a decadere sul livello stabile E_1 , cedendo l'energia $E_2 - E_1$ sotto forma di un fotone γ_g . Come già osservato nel paragrafo 4.3 la probabilità di tale decadimento fotostimolato è legata al prodotto scalare fra il versore \hat{a} di polarizzazione del fotone incidente γ ed il vettore \vec{p} della quantità di moto dell'elettrone che si trova, prima dello scattering, sul livello energetico E_2 (in BC). In altri termini, indicando con ψ_2 lo stato elettronico prima del decadimento stimolato radiativo e con ψ_1 lo stato elettronico dopo il decadimento, l'elemento di matrice $|\langle \psi_1 | \hat{a} \cdot \vec{p} | \psi_2 \rangle|^2$ costituisce una "selection rule" (regola di selezione, discussa nel paragrafo 4.3) per l'emissione stimolata, ovvero una regola di "coupling" (accoppiamento) fotonico fra lo stato elettronico iniziale ψ_2 e lo stato elettronico finale ψ_1 . γ_g è un fotone "gemello" di γ , poiché i due bosoni hanno non solo la stessa frequenza, ma anche la stessa fase, la stessa polarizzazione, la stessa direzione e lo stesso verso di propagazione: γ e γ_g sono "identici". Alla fine del decadimento stimolato l'elettrone si trova sul livello E_1 (la popolazione N_2 è decrementata, quella N_1 è incrementata), mentre i due fotoni γ e γ_g , che formano un fascio di luce coerente, si allontanano. γ e γ_g costituiscono una forma "pregiata" di energia luminosa (la luce emessa dai LASERs si basa su questo fenomeno). Il contributo della transizione 3) all'equazione di continuità è:

$$\frac{dN_2(t)}{dt} = - B_{21}N_2(t)\rho(\nu) \quad [\text{cm}^{-3} \text{ s}^{-1}]$$

dove B_{21} è uno dei noti "coefficienti di Einstein", mentre $\rho(\nu)$ [$\text{J}/\text{m}^3\text{Hz}$] è la densità spettrale volumetrica di energia, legata alla presenza dei fotoni che attraversano il cristallo (semiconduttore). In prima approssimazione possiamo pensare di associare $\rho(\nu)$ alla densità spettrale volumetrica di energia di corpo nero:

$$\rho(\nu) = \frac{8\pi n^3 h\nu^3}{c^3} \frac{1}{e^{kT} - 1}$$

dove k è la costante di Boltzmann, mentre n (più precisamente $n(\nu)$) rappresenta l'indice di rifrazione del materiale investito da una radiazione di frequenza ν ; ancora più esattamente n costituisce la parte reale $n_r(\nu)$ dell'indice di rifrazione complesso $n_c(\nu)$ (come spiegato nel paragrafo 4.3). Va puntualizzato che associare $\rho(\nu)$ alla densità spettrale volumetrica di energia di corpo nero non è completamente corretto: infatti un corpo nero possiede uno spettro di emissione di fatto dipendente unicamente dalla temperatura T al suo interno, e decisamente largo in frequenza, mentre lo spettro di emissione di un semiconduttore, oltre a dipendere dalla propria struttura a bande (la quale dipende da T , ma anche da altri fattori), è in genere molto più selettivo in frequenza. Tale associazione è finalizzata unicamente al calcolo del coefficiente B_{21} .

4) Assorbimento fotonico

Un fotone di energia $h\nu_\gamma = E_2 - E_1$, se assorbito, consente ad un elettrone del livello stabile E_1 (ovvero la BV) di superare la barriera di potenziale dovuta al gap energetico, promuovendolo al livello E_2 (ossia in BC). Questa transizione, stimolata dal fotone incidente γ , causa un depauperamento della popolazione elettronica N_1 ed un arricchimento di quella N_2 . La probabilità di assorbimento è legata al prodotto scalare fra il versore \hat{a} di polarizzazione del fotone incidente γ ed il vettore \vec{p} della quantità di moto dell'elettrone che si trova, prima dello scattering, sul livello energetico stabile E_1 (in BV). In altri termini, indicando con ψ_1 e ψ_2 , rispettivamente, lo stato elettronico prima e dopo l'assorbimento fotonico,

l'elemento di matrice $|\langle \psi_2 | \hat{a} \cdot \vec{p} | \psi_1 \rangle|^2$ costituisce una regola di selezione per l'assorbimento dei fotoni γ , discussa nel paragrafo 4.3, ovvero una regola di accoppiamento ottico fra lo stato elettronico iniziale ψ_1 e lo stato elettronico finale ψ_2 . Il contributo della transizione 4) all'equazione di continuità del sistema a due livelli è:

$$\frac{dN_2(t)}{dt} = -\frac{dN_1(t)}{dt} = B_{12}N_1(t)\rho(\nu) \quad [\text{cm}^{-3} \text{s}^{-1}]$$

dove B_{12} è uno dei noti "coefficienti di Einstein". Nel caso particolare dei semiconduttori abbiamo che $B_{12} = B_{21}$: ciò è vero in quanto i numeri degli stati disponibili sia su un generico livello di energia della BV sia su un generico livello di energia della BC sono talmente alti che possono essere ritenuti uguali fra loro. In altre parole il numero di degenerazione g_1 , relativo al livello energetico appartenente alla BV, è uguale al numero di degenerazione g_2 , relativo al livello energetico appartenente alla BC. $\rho(\nu)$ è sempre la densità spettrale volumetrica di energia di corpo nero. Anche adesso va puntualizzato che non è completamente corretto modellizzare l'assorbimento dei fotoni, da parte di un semiconduttore, attraverso l'analogia con il corpo nero, poiché il corpo nero è, per definizione, un assorbitore perfetto di energia elettromagnetica, ovvero il suo spettro di assorbimento è costante per qualunque frequenza ν , mentre un semiconduttore, così come qualunque altro materiale in natura, possiede uno spettro di assorbimento ben limitato in frequenza e dipendente dalla struttura dei suoi livelli energetici (dalla struttura a bande, nel caso dei semiconduttori). L'analogia col corpo nero è finalizzata unicamente al calcolo del coefficiente B_{12} .

5.1.2) Le principali modalità di interazione fra luce e semiconduttore: calcolo dei rispettivi rates

Per stimare il rate di decadimento degli elettroni dalla BC alla BV è necessario applicare la regola d'oro di Fermi, riportata nella [e28], al caso particolare di transizione elettronica BC \rightarrow BV. Riscriviamo quindi la regola d'oro di Fermi nel caso di decadimento elettronico radiativo BC \rightarrow BV, senza specificarne il carattere spontaneo oppure stimolato:

$$W_{EM}(|i\rangle) = W_{EMsp}(|i\rangle) + W_{EMst}(|i\rangle) = \frac{2\pi}{\hbar} \sum_{|f\rangle} |\langle f | H_F | i \rangle|^2 \delta(|E_f - E_i| + \hbar\omega) \quad [e34]$$

$W_{EM}(|i\rangle)$ rappresenta il rate ($\text{cm}^{-3}\text{s}^{-1}$) di decadimento elettronico radiativo totale BC \rightarrow BV, ovvero la somma fra il rate $W_{EMsp}(|i\rangle)$ di decadimento elettronico radiativo spontaneo ed il rate $W_{EMst}(|i\rangle)$ di decadimento elettronico radiativo stimolato. Sappiamo inoltre che:

$$H_F = \frac{ie\hbar}{m_0} \vec{A} \cdot \nabla$$

H_F , come spiegato nel paragrafo 4.3, è il termine hamiltoniano fotoperturbativo. Come già osservato il potenziale vettore \vec{A} è dato da:

$$\vec{A} = A_0 \hat{a}$$

dove A_0 è il modulo del potenziale vettore. Nella teoria quantistica dei campi A_0 può essere altresì visto come un'applicazione lineare matriciale esprimibile nella forma seguente, nel caso di emissione ottica da parte del cristallo semiconduttore:

$$A_0 = \sqrt{\frac{\hbar}{2\varepsilon_0\varepsilon_r\omega V}} \hat{b}^\dagger$$

V è l'elemento di volume, all'interno dell'intero cristallo semiconduttore, nel quale avviene lo scattering tra i fotoni e gli elettroni atomici che si trovano in BC; in altri termini V è il volume della regione del cristallo soggetta ad illuminazione. \hat{b}^\dagger rappresenta "l'operatore quantistico di creazione" per i fotoni (che sono bosoni, da cui la lettera "b" con la quale solitamente si indica tale operatore) all'interno di un cristallo semiconduttore, il quale è stato "pompat" da una corrente elettrica imposta dall'esterno, come avviene per i LASERs a semiconduttore (LASERs "a quattro livelli") e per quelli non a stato solido (ad esempio costruiti con un tubo a vuoto riempito con una miscela costituita per il 15% di elio e per l'85% di neon), oppure da una sorgente ottica esterna, come per

i LASERS al rubino e per quelli in fibra di silice drogata con erbio (LASERS “a tre livelli”), oppure ancora da una sorgente termica. L'operatore quantistico di creazione \hat{b}^\dagger è uno strumento matematico che implementa, nella regola d'oro di Fermi applicata all'emissione ottica, il fenomeno di creazione di un fotone in seguito ad una transizione che porta un elettrone dalla BC alla BV spontaneamente oppure in seguito ad una sollecitazione (stimolazione) fotonica incidente. Se $|\psi_{0\gamma}\rangle = |0\rangle$ rappresenta lo stato fotonico corrispondente alla situazione di assenza di luce nel cristallo, cioè assenza di fotoni, il fenomeno dell'emissione di un fotone in seguito ad una ricombinazione spontanea elettrone/lacuna (in un semiconduttore) è implementato dall'applicazione di \hat{b}^\dagger sullo stato “di buio” $|\psi_{0\gamma}\rangle = |0\rangle$, la quale fornisce, come risultato, lo stato fotonico $|\psi_{1\gamma}\rangle = |1\rangle$ corrispondente alla presenza di un fotone emesso spontaneamente.

$$\hat{b}^\dagger |\psi_{0\gamma}\rangle = |\psi_{1\gamma}\rangle$$

$$\hat{b}^\dagger |0\rangle = |1\rangle$$

In generale se $|\psi_{n\gamma}\rangle = |n\rangle$ rappresenta lo stato fotonico corrispondente alla propagazione, attraverso il cristallo, di un fascio di luce formato da n fotoni, il fenomeno dell'emissione di un ulteriore fotone, causata da una ricombinazione spontanea elettrone/lacuna oppure da una stimolata, è implementato dall'applicazione di \hat{b}^\dagger sullo stato iniziale $|\psi_{n\gamma}\rangle = |n\rangle$, la quale fornisce, come risultato, lo stato fotonico finale $|\psi_{(n+1)\gamma}\rangle = |n+1\rangle$. In altre parole l'azione dell'operatore di creazione su uno stato quantistico $|\psi\rangle$ incrementa di un'unità il numero di particelle di $|\psi\rangle$.

$$\hat{b}^\dagger |\psi_{n\gamma}\rangle = |\psi_{(n+1)\gamma}\rangle$$

$$\hat{b}^\dagger |n\rangle = |n+1\rangle$$

Lo stato fotonico $|\psi\rangle$, la cui norma $\|\psi\|$ aumenta al susseguirsi delle emissioni ottiche spontanee e/o stimolate attraverso il materiale, può essere espresso, come è ben noto, attraverso una funzione d'onda simmetrica (tipica dei bosoni, così come quella antisimmetrica è caratteristica dei fermioni) oppure, utilizzando la notazione di Heisenberg, con un vettore colonna $\mathbb{C}^{N \times 1}$ sviluppato su un'opportuna base ortonormale. Quindi l'operatore di creazione \hat{b}^\dagger è esprimibile con una matrice $\mathbb{C}^{N \times N}$; tale matrice costituisce un'applicazione lineare che ha come dominio lo stato fotonico $|\psi\rangle$ e come immagine lo stato fotonico $\hat{b}^\dagger |\psi\rangle$, il quale consta di un fotone in più rispetto a $|\psi\rangle$, ragion per cui abbiamo che $\|\hat{b}^\dagger |\psi\rangle\| > \|\psi\rangle\|$. Proponiamo di seguito la matrice di creazione fotonica \hat{b}^\dagger :

$$\hat{b}^\dagger = \begin{pmatrix} 0 & 0 & 0 & \dots & \dots & 0 \\ \sqrt{1} & 0 & 0 & \vdots & \ddots & \vdots \\ 0 & \sqrt{2} & 0 & \dots & \dots & 0 \\ 0 & 0 & \sqrt{3} & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & 0 & \vdots \\ 0 & 0 & 0 & \dots & \sqrt{N} & 0 \end{pmatrix}$$

Abbiamo tutti gli elementi per esprimere la [e34] in una forma più esplicita:

$$W_{EM}(|\mathbf{i}\rangle) = \frac{2\pi}{\hbar} \sum_{|\mathbf{f}\rangle} \left| \langle k_f, n_{ph} + 1 \left| \frac{ie\hbar}{m_0} \sqrt{\frac{\hbar}{2\varepsilon_0\varepsilon_r\omega V}} \hat{b}^\dagger \cdot \hat{\mathbf{a}} \cdot \nabla \right| k_i, n_{ph} \rangle \right|^2 \delta(|E_f - E_i| + \hbar\omega) \quad [e35]$$

dove l'operatore di creazione \hat{b}^\dagger trasforma lo stato totale iniziale $|k_i, n_{ph}\rangle$ nello stato totale finale $|k_f, n_{ph} + 1\rangle$. Svolgendo alcuni calcoli nella [e35] otteniamo la seguente espressione:

$$W_{EM}(|\mathbf{i}\rangle) = \frac{2\pi}{\hbar} \frac{e^2}{m_0^2} \frac{\hbar\{n_{ph}+1\}}{2\omega\varepsilon_0\varepsilon_r} \sum_{|\mathbf{f}\rangle} \left| \int \psi_f(r)^* [\hat{\mathbf{a}} \cdot \vec{\mathbf{p}}] e^{-i(k_f \cdot r)} \psi_i(r) d^3r \right|^2 \delta(|E_f - E_i| + \hbar\omega) \quad [e36]$$

dove $\psi_i(r)$ è lo stato dell'elettrone quando questo si trova in BC, ovvero prima dell'emissione ottica prodotta dal suo decadimento in BV, $\psi_f(r)$ è lo stato dell'elettrone dopo che questo ha raggiunto la BV, cioè ad emissione

fotonica avvenuta. Il termine $\exp\{-i(\mathbf{k}_\gamma \cdot \mathbf{r})\}$ rappresenta l'onda piana, di numero d'onda k_γ , associata al fotone generato dalla ricombinazione elettrone/lacuna, mentre il termine compreso tra parentesi graffe, al numeratore della [e36], indica il subentramento di un ulteriore fotone nel fascio luminoso che inizialmente (cioè prima della ricombinazione elettrone/lacuna) constava di n_{ph} fotoni propaganti. Lo stesso termine tra parentesi graffe consente di separare, nel rate totale $W_{EM}(|\mathbf{i}\rangle)$ di emissione ottica, la componente $W_{EMsp}(|\mathbf{i}\rangle)$ dovuta alla ricombinazione spontanea elettrone/lacuna da quella $W_{EMst}(|\mathbf{i}\rangle)$ dovuta alla ricombinazione stimolata. In formule:

$$W_{EMsp}(|\mathbf{i}\rangle) = \frac{2\pi}{\hbar} \frac{e^2}{m_0^2} \frac{\hbar}{2\omega\epsilon_0\epsilon_r} \sum_{|f\rangle} \left| \int \psi_f(\mathbf{r})^* e^{-i(\mathbf{k}_\gamma \cdot \mathbf{r})} \psi_i(\mathbf{r}) d^3r \right|^2 \delta(|E_f - E_i| + \hbar\omega) \quad [e37]$$

$$W_{EMst}(|\mathbf{i}\rangle) = \frac{2\pi}{\hbar} \frac{e^2}{m_0^2} \frac{\hbar n_{ph}}{2\omega\epsilon_0\epsilon_r} \sum_{|f\rangle} \left| \int \psi_f(\mathbf{r})^* [\hat{\mathbf{a}} \cdot \vec{\mathbf{p}}] e^{-i(\mathbf{k}_\gamma \cdot \mathbf{r})} \psi_i(\mathbf{r}) d^3r \right|^2 \delta(|E_f - E_i| + \hbar\omega) \quad [e38]$$

Nella [e37], rispetto alla [e38], manca il numero n_{ph} di fotoni propaganti nel cristallo prima dell'emissione ottica spontanea, poiché quest'ultima non dipende dalla luce che attraversa il semiconduttore (la luce infatti non funge da stimolo per il decadimento spontaneo radiativo). Questo è il motivo per cui nella [e37] manca, sempre rispetto alla [e38], anche l'overlap $[\hat{\mathbf{a}} \cdot \vec{\mathbf{p}}]$ tra il vettore unitario di polarizzazione della luce incidente (la quale, infatti, è come se non ci fosse, dato che il cristallo può emettere spontaneamente fotoni anche partendo da una condizione di buio, cioè con $n_{ph} = 0$) ed il vettore impulso dell'elettrone in BC. La [e38] dipende invece sia dal numero totale di fotoni "potenziali stimolatori" n_{ph} che dalla regola di selezione $[\hat{\mathbf{a}} \cdot \vec{\mathbf{p}}]$; quest'ultima infatti seleziona quali fotoni, tra gli n_{ph} complessivi, presentano una probabilità non nulla ($[\hat{\mathbf{a}} \cdot \vec{\mathbf{p}}] > 0$) di stimolare il decadimento di un elettrone di conduzione, avente impulso $\vec{\mathbf{p}}$, verso la BV.

Osservando la [e38] si evince che maggiore è il pompaggio al quale è sottoposto il semiconduttore con cui è realizzato un LASER e maggiore è il numero di stati elettronici $|i\rangle$ pieni in BC (occupati dagli elettroni minoritari iniettati) e di stati elettronici $|f\rangle$ vuoti in BV (stati finali permessi agli elettroni, ovvero occupati dalle lacune minoritarie iniettate); questi ultimi possono essere occupati dagli elettroni che decadono dalla BC alla BV. Dunque maggiore è il pompaggio e maggiore è la sommatoria Σ sugli stati di arrivo $|f\rangle$ degli elettroni che sono in procinto di essere stimolati a decadere in BV. Osservando la [e37] si evince altresì che maggiore è l'iniezione di portatori minoritari all'interno della zds di un LED e maggiore è il numero di stati elettronici $|f\rangle$ vuoti in BV (sono sempre stati finali permessi), che possono essere riempiti dagli elettroni che decadono spontaneamente dalla BC alla BV, cioè stati che consentono la transizione elettronica spontanea $|i\rangle \rightarrow |f\rangle$. Quindi maggiore è la corrente d'iniezione I_{BIAS} di elettroni in BC e di lacune in BV (che può andare da qualche mA a qualche decina di mA), a cui sottoponiamo un semiconduttore, e maggiore è $W_{EMsp}(|\mathbf{i}\rangle)$, cioè minore è il tempo di vita medio τ_R radiativo ("radiative lifetime") degli elettroni in BC e delle lacune in BV, come si può notare dal grafico sperimentale mostrato in figura 33, dove riportiamo l'andamento di τ_R in funzione della concentrazione $n = p$ delle coppie elettrone/lacuna iniettate nella zds di un LED costruito con GaAs.

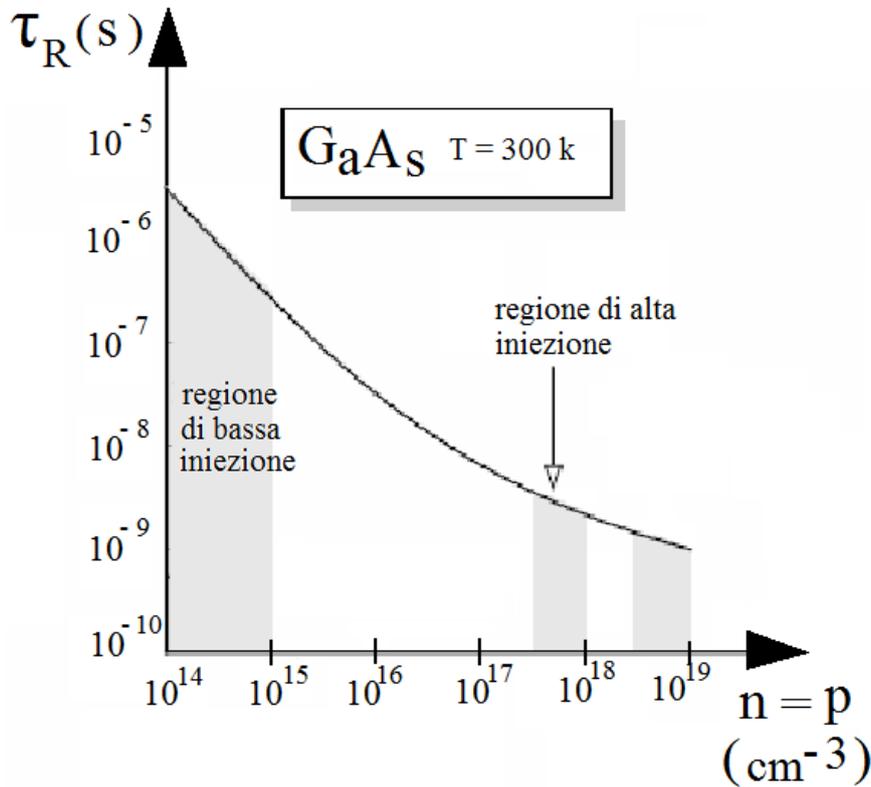


Fig. 33

Grafico sperimentale del tempo di vita medio τ_R radiativo (“radiative lifetime”) degli elettroni in BC e delle lacune in BV in funzione della concentrazione $n = p$ delle coppie elettrone/lacuna, iniettate nella zds di un LED costruito con GaAs. Per correnti d’iniezione I_{BIAS} molto deboli (qualche mA, come ordine di grandezza), cioè per $n = p \approx 10^{14} \div 10^{15} \text{ cm}^{-3}$, siamo in presenza di bassa iniezione di minoritari, pertanto il tempo di vita medio degli elettroni in BC e delle lacune in BV è alto (dell’ordine dei μs). Per correnti d’iniezione I_{BIAS} adeguate (qualche decina di mA, come ordine di grandezza), cioè per $n = p \approx 10^{17} \div 10^{18} \text{ cm}^{-3}$, siamo in presenza di alta iniezione di minoritari, pertanto il tempo di vita medio degli elettroni in BC e delle lacune in BV è piccolo (dell’ordine dei ns).

Per stimare il rate di assorbimento dei fotoni è necessario applicare sempre la regola d’oro di Fermi, però stavolta al caso particolare di transizione elettronica $BV \rightarrow BC$. Riscriviamo quindi la regola d’oro di Fermi nel caso di assorbimento ottico:

$$W_{ABS}(|\mathbf{i}\rangle) = \frac{2\pi}{\hbar} \sum_{|f\rangle} |\langle f| H_F |i\rangle|^2 \delta(|E_f - E_i| - \hbar\omega) \quad [e39]$$

sempre con:

$$H_F = \frac{ie\hbar}{m_0} \mathbf{A} \cdot \nabla$$

$$\vec{A} = A_0 \hat{a}$$

Nella teoria quantistica dei campi A_0 può essere visto come un’applicazione lineare matriciale esprimibile nella forma seguente, nel caso di assorbimento fotonico da parte del cristallo semiconduttore:

$$A_0 = \sqrt{\frac{\hbar}{2\varepsilon_0\varepsilon_r\omega V}} \hat{b}$$

V è l’elemento di volume, all’interno dell’intero cristallo semiconduttore, nel quale avviene lo scattering tra i fotoni e gli elettroni atomici che si trovano in BV; V è sempre il volume della regione del cristallo soggetta ad illuminazione. \hat{b} rappresenta “l’operatore quantistico di distruzione” per i fotoni all’interno di un cristallo semiconduttore. La fotorelevazione basata su dispositivi (giunzioni) a semiconduttore sfrutta il trascinarsi, imposto da una polarizzazione inversa applicata dall’esterno, di coppie elettrone/lacuna la cui generazione è causata proprio dall’assorbimento dell’energia luminosa da parte degli elettroni atomici in BV; tale assorbimento comporta la distruzione dei fotoni stessi. L’operatore quantistico di distruzione \hat{b} è uno strumento matematico

che implementa, nella regola d'oro di Fermi applicata all'assorbimento ottico, il fenomeno di distruzione di un fotone grazie al quale un elettrone passa dalla BV alla BC. In generale se $|\psi_{n\gamma}\rangle = |n\rangle$ rappresenta lo stato fotonico corrispondente alla propagazione, attraverso il cristallo, di un fascio di luce formato da n fotoni, il fenomeno dell'assorbimento di uno degli n fotoni è implementato dall'applicazione di \hat{b} sullo stato iniziale $|\psi_{n\gamma}\rangle = |n\rangle$; questa applicazione fornisce, come risultato, lo stato fotonico finale $|\psi_{(n-1)\gamma}\rangle = |n-1\rangle$. In altre parole l'azione dell'operatore di distruzione su uno stato quantistico $|\psi\rangle$ decrementa di un'unità il numero di particelle di $|\psi\rangle$.

$$\hat{b} |\psi_{n\gamma}\rangle = |\psi_{(n-1)\gamma}\rangle$$

$$\hat{b} |n\rangle = |n-1\rangle$$

Lo stato fotonico $|\psi\rangle$, la cui norma $\|\psi\|$ si depaupera al susseguirsi degli assorbimenti di fotoni, può essere espresso attraverso una funzione d'onda simmetrica oppure con l'utilizzo di un vettore colonna $\mathbb{C}^{N \times 1}$, secondo la notazione di Heisenberg, sviluppato su un'opportuna base ortonormale. Utilizzando la notazione di Heisenberg l'operatore di distruzione \hat{b} è esprimibile da una matrice $\mathbb{C}^{N \times N}$; tale matrice costituisce un'applicazione lineare che ha come dominio lo stato fotonico $|\psi\rangle$ e come immagine lo stato fotonico $\hat{b}|\psi\rangle$, il quale consta di un fotone in meno rispetto a $|\psi\rangle$, ragion per cui abbiamo che $\|\hat{b}|\psi\rangle\| < \|\psi\rangle\|$. Proponiamo di seguito la matrice di distruzione fotonica \hat{b} :

$$\hat{b} = \begin{pmatrix} 0 & \sqrt{1} & 0 & 0 & \dots & 0 \\ 0 & 0 & \sqrt{2} & 0 & \ddots & \vdots \\ 0 & 0 & 0 & \sqrt{3} & \dots & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & 0 & \sqrt{N} \\ 0 & 0 & 0 & \dots & 0 & 0 \end{pmatrix} = \hat{b}^{\dagger T}$$

Abbiamo tutti gli elementi per scrivere la [e39] in una forma più esplicita:

$$W_{\text{ABS}}(|\mathbf{i}\rangle) = \frac{2\pi}{\hbar} \sum_{|\mathbf{f}\rangle} \left| \langle k_f, n_{ph} - 1 \left| \frac{ie\hbar}{m_0} \sqrt{\frac{\hbar}{2\varepsilon_0\varepsilon_r\omega V}} \hat{b} \cdot \hat{\mathbf{a}} \cdot \nabla \right| k_i, n_{ph} \rangle \right|^2 \delta(|E_f - E_i| - \hbar\omega) \quad [\text{e40}]$$

dove l'operatore di distruzione \hat{b} trasforma lo stato totale iniziale $|k_i, n_{ph}\rangle$ nello stato totale finale $|k_f, n_{ph} - 1\rangle$. Svolgendo alcuni calcoli nella [e40] otteniamo la seguente espressione:

$$W_{\text{ABS}}(|\mathbf{i}\rangle) = \frac{2\pi}{\hbar} \frac{e^2}{m_0^2} \frac{\hbar n_{ph}}{2\omega\varepsilon_0\varepsilon_r} \sum_{|\mathbf{f}\rangle} \left| \int \psi_f(r)^* [\hat{\mathbf{a}} \cdot \vec{\mathbf{p}}] e^{+i(\mathbf{k}_\gamma \cdot r)} \psi_i(r) d^3r \right|^2 \delta(|E_f - E_i| - \hbar\omega) \quad [\text{e41}]$$

dove $\psi_i(r)$ è lo stato dell'elettrone quando questo si trova in BV, ovvero prima dell'assorbimento ottico, $\psi_f(r)$ è lo stato dell'elettrone dopo che questo è stato promosso in BC, cioè ad assorbimento avvenuto. Il termine $\exp\{+i(\mathbf{k}_\gamma \cdot r)\}$ rappresenta l'onda piana associata al fotone assorbito dall'elettrone atomico. Maggiore è il numero n_{ph} di fotoni che attraversano il cristallo e maggiore è il rate $W_{\text{ABS}}(|\mathbf{i}\rangle)$ di assorbimento, cioè maggiore è la probabilità di assorbimento della luce. Il prodotto scalare $[\hat{\mathbf{a}} \cdot \vec{\mathbf{p}}]$ funge sempre da regola di selezione.

5.2) L'equazione di continuità completa per un sistema di elettroni (semplificato) in un semiconduttore

Ricaviamo l'equazione di continuità completa, sommando i quattro contributi descritti nel paragrafo 5.1.1:

$$\frac{dN_2(t)}{dt} = -\frac{N_2(t)}{\tau_{NR}} - \frac{N_2(t)}{\tau_R} + B_{12}N_1(t)\rho(\nu) - B_{21}N_2(t)\rho(\nu) \quad [e42]$$

Se supponiamo che i decadimenti spontanei non radiativi siano improbabili, rispetto a quelli radiativi (ovvero $\tau_{NR} \gg \tau_R$), che il materiale per temperature ordinarie si trovi in condizione di equilibrio termico (le popolazioni N_1 ed N_2 , a meno di fluttuazioni granulari stocastiche a media nulla, possono essere considerate costanti nel tempo), e che il comportamento degli elettroni non sia molto dissimile da quello di particelle classiche, cioè assumendo che la statistica di Maxwell - Boltzmann sia valida per il nostro sistema a due livelli E_1 ed E_2 , possiamo scrivere quanto segue:

$$\left\{ \begin{array}{l} \rho(\nu) = \frac{\frac{N_2}{\tau_R}}{[N_1 - N_2] B_{12}} = \frac{8\pi n^3 h\nu^3}{c^3} \frac{1}{e^{\frac{h\nu}{kT}} - 1} \\ \frac{N_2}{N_1} = \frac{g_2}{g_1} \frac{e^{-\frac{E_2}{kT}}}{e^{-\frac{E_1}{kT}}} = e^{-\frac{h\nu}{kT}} \end{array} \right.$$

Possiamo quindi ricavare il coefficiente B_{12} :

$$B_{12} = B_{21} = \frac{c^3}{8\pi n^3 h\nu^3 \tau_R}$$

Come già detto B_{12} e B_{21} , insieme al già noto A_{21} ($= 1/\tau_R$), sono noti come "coefficienti di Einstein".

5.3.1) Un modello semiclassico per l'interazione fra luce e semiconduttore

Supponiamo adesso di avere un dispositivo costruito con un semiconduttore a gap indiretto, drogato oppure intrinseco, e di rompere l'equilibrio termico inviando su di esso una radiazione luminosa pluricromatica.

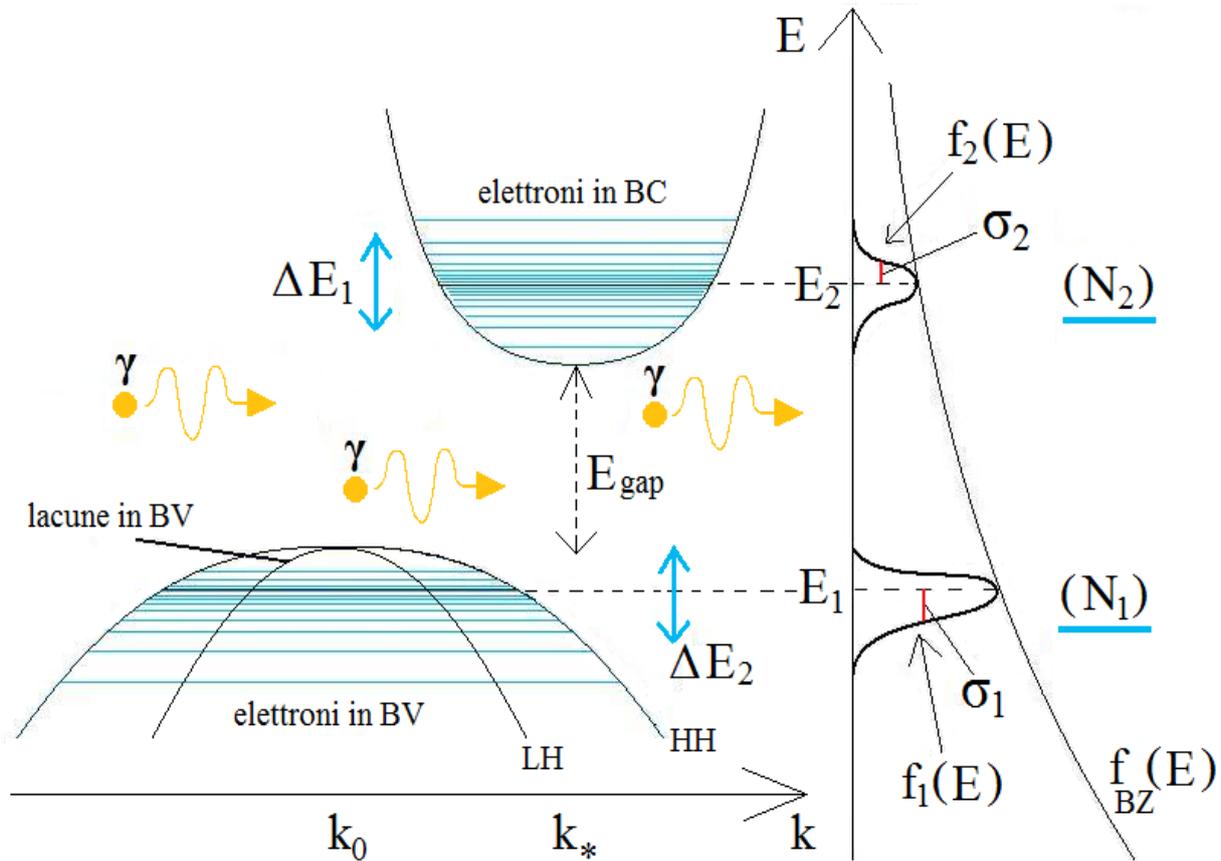


Fig. 34

Diagramma a bande qualitativo di un semiconduttore a gap indiretto (drogato oppure intrinseco), con il quale può essere realizzato un fotorelizzatore. L'equilibrio termico è alterato da una radiazione luminosa pluricromatica. Si noti l'adattamento del modello a bande a quello, più semplice, costituito da due livelli energetici (E_1 ed E_2), ciascuno visto come il livello medio di un ristretto range di livelli energetici (ΔE_1 e ΔE_2). N_1 è la popolazione di ΔE_1 , N_2 quella di ΔE_2 (più precisamente sono concentrazioni di popolazioni). Le funzioni di probabilità di occupazione $f_1(E)$ ed $f_2(E)$ indicano che i livelli medi hanno la maggiore probabilità di essere popolati dagli elettroni. Sui range energetici ΔE_1 e ΔE_2 sono distribuiti elettroni che supponiamo di poter trattare come particelle classiche ($f_{BZ}(E)$ indica la distribuzione statistica di Maxwell - Boltzman).

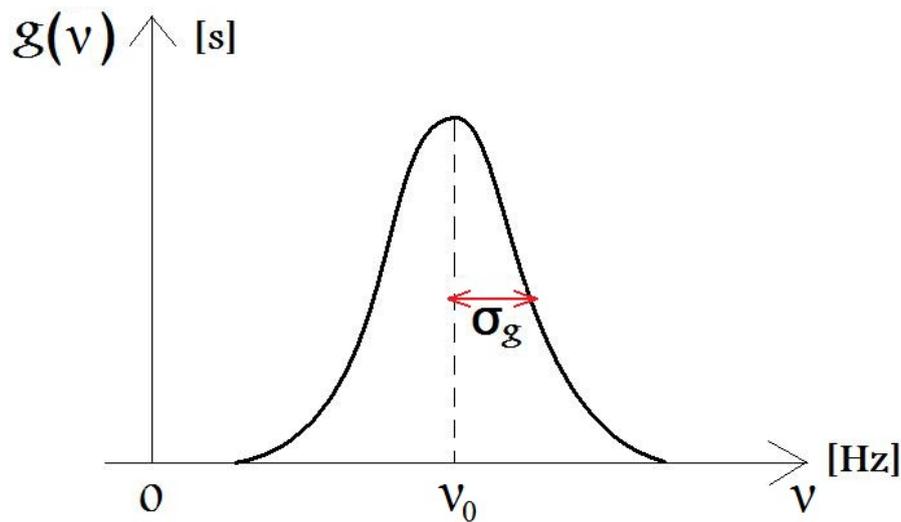
Riferiamoci alla figura 34: gli elettroni in BC, così come le lacune in BV, sono riconducibili sia all'attività ordinaria di generazione e ricombinazione termiche, sia all'eventuale drogaggio. Supponiamo di poter trattare il sistema di elettroni come un insieme di particelle classiche, ovvero non interagenti fra loro: stiamo quindi ipotizzando che la lunghezza d'onda λ di De Broglie degli elettroni in BV sia abbastanza più piccola della distanza media di separazione degli elettroni, ovvero della costante reticolare "a" del semiconduttore, e che gli elettroni in BC siano pochi (temperature e livelli di drogaggio ordinari), così che la precedente relazione fra λ ed "a" sia verificata. In altre parole stiamo ipotizzando che dato l'insieme degli elettroni $\{e_k\}$ in BV ed in BC, cioè l'insieme di tutti gli elettroni atomici di valenza del cristallo, gli operatori hamiltoniani \hat{H}_k commutano fra loro, dove ciascun \hat{H}_k è l'hamiltoniano che descrive l'energia totale del singolo elettrone e_k . Ciascun elettrone e_k risulta distinguibile dagli altri elettroni del sistema sia nello spazio reale delle coordinate x, y, z , poiché $\lambda \ll "a"$, pertanto la densità di probabilità $|\psi_k|^2$ dell'elettrone e_k non si sovrappone a quella degli altri elettroni atomici in modo significativo, sia nello spazio reciproco dei vettori d'onda k , poiché lo spettro di Fourier $F\{\psi_k\}$ dell'elettrone e_k , in quanto approssimabile con una delta di Dirac, non si sovrappone agli spettri degli altri elettroni. Assumiamo pertanto

che tutti gli elettroni, sia in BV che in BC, obbediscano alla statistica classica di Maxwell - Boltzman $f_{BZ}(E) = C \exp\{-E/kT\}$.

Adattiamo il modello a bande a quello, più semplice, costituito dai due livelli energetici E_1 ed E_2 . E_1 è un livello appena al di sotto del massimo E_V della BV, intorno al quale è distribuita la funzione $f_1(E)$ di probabilità di occupazione (da parte degli elettroni) dei livelli. E_1 può essere visto come il valor medio di un certo intervallo di livelli energetici ΔE_1 , tutti collocati ad di sotto del massimo della BV: al valor medio E_1 è associata la massima probabilità di occupazione, mentre allontanandosi da E_1 la probabilità diminuisce. σ_1 è la deviazione standard di $f_1(E)$. N_1 è la popolazione di ΔE_1 . E_2 è un livello appena al di sopra del minimo E_C della BC, intorno al quale è distribuita la funzione $f_2(E)$ di probabilità di occupazione (sempre da parte degli elettroni) dei livelli. E_2 può essere visto come il valor medio di un certo intervallo di livelli energetici ΔE_2 , tutti collocati ad di sopra del minimo della BC: al valor medio E_2 è associata la massima probabilità di occupazione, mentre allontanandosi da E_2 la probabilità diminuisce. σ_2 è la deviazione standard di $f_2(E)$. N_2 è la popolazione di ΔE_2 . In base a quanto detto finora possiamo scrivere:

$$E_{\text{gap}} \approx E_2 - E_1 = h\nu_0$$

dove ν_0 è la frequenza della luce incidente in corrispondenza della quale si ha la massima probabilità di assorbimento fotonico, così come la massima probabilità di emissione di fotoni, mediante la ricombinazione dei portatori. Spostandosi da ν_0 lungo l'asse delle frequenze ν osserviamo una diminuzione della probabilità di transizione degli elettroni fra ΔE_1 e ΔE_2 . Possiamo definire una funzione di "probabilità di transizione" $g(\nu)$ (misurata in secondi), nota come "Line Shape", che per un insieme di elettroni riconducibile ad un sistema a due livelli energetici è approssimabile con una funzione gaussiana.



$$g(\nu) \in \mathcal{N}(\nu_0, \sigma_g^2)$$

$$g(\nu) = \frac{1}{\sqrt{2\pi\sigma_g^2}} e^{-\frac{(\nu-\nu_0)^2}{2\sigma_g^2}}$$

Fig. 35

Grafico qualitativo della distribuzione di "probabilità di transizione" $g(\nu)$, nota come "Line Shape": si noti che per un insieme di elettroni, riconducibile ad un semplice sistema a due livelli energetici, $g(\nu)$ è approssimabile con una funzione gaussiana. ν_0 è il "centro banda" della Line Shape e σ_g la sua deviazione standard.

ν è una “variabile aleatoria gaussiana”, la frequenza ν_0 è nota come “centro banda” della Line Shape e costituisce il valor medio (detto anche “intensità”, oppure “valore atteso”) della distribuzione gaussiana, mentre σ_g ($= \sqrt{\sigma_1^2 + \sigma_2^2}$) è la sua deviazione standard; σ_g^2 è la varianza di ν . Avendo ipotizzato che il semiconduttore ha gap indiretto è ragionevole attendersi che la ricombinazione spontanea radiativa banda \rightarrow banda sia decisamente improbabile, in quanto tale transizione spontanea necessiterebbe dell'intervento di un modo vibrazionale fononico “favorevole”, al fine della conservazione della quantità di moto. Inoltre ipotizziamo che le concentrazioni di trappole energetiche all'interno del gap siano trascurabili e che la popolazione elettronica N_2 non sia trascurabile rispetto a quella N_1 . Pertanto la ricombinazione banda \rightarrow trappole \rightarrow banda è trascurabile rispetto a quella stimolata. Possiamo altresì affermare che le uniche transizioni ottiche significative sono quelle relative all'assorbimento fotonico e all'emissione stimolata. Dunque l'equazione di continuità diventa:

$$\frac{dN_2(t)}{dt} = - B_{12} [N_2(t) - N_1(t)] \rho_{\text{col}}(\nu)$$

Si noti la correzione apportata al termine che costituisce lo spettro di assorbimento e di emissione stimolata dei fotoni da parte del semiconduttore, che è assolutamente diverso dagli spettri di assorbimento e di emissione di corpo nero. $\rho_{\text{col}}(\nu)$ è infatti una densità spettrale volumetrica di energia “colorata”, ovvero uno spettro che “concentra” la densità volumetrica di energia di corpo nero (μ) presso quelle sole frequenze ν per le quali si ha, effettivamente, una probabilità di transizione (assorbimento/emissione stimolata) non nulla, ovvero in quelle frequenze presso le quali si ha un valore non nullo della Line Shape $g(\nu)$. In formule:

$$\rho_{\text{col}}(\nu) = g(\nu)\mu = g(\nu) \int_{-\infty}^{+\infty} \rho(\nu) d\nu \quad [\text{J}/\text{m}^3\text{Hz}]$$

Ricordando che, all'interno di un fascio di luce pluricromatica incidente, è possibile definire l'intensità luminosa I_ν come il numero di fotoni N_ν , tutti aventi energia $h\nu$, che attraversano una superficie unitaria nell'unità di tempo, è immediato esprimere lo spettro di assorbimento e di emissione stimolata del semiconduttore come:

$$\rho_{\text{col}}(\nu) = \frac{I_\nu}{c} g(\nu)$$

da cui si ha che:

$$\frac{dN_2(t)}{dt} = - B_{12} [N_2(t) - N_1(t)] \frac{I_\nu}{c} g(\nu) \quad [\text{e43}]$$

$N_2(t) - N_1(t)$ è la ben nota “inversione di popolazione”, che è negativa nel caso dei materiali con cui sono costruiti i fotorivelatori ($N_2 < N_1$, come è logico aspettarsi per un sistema a temperature ordinarie e all'equilibrio termico, e come è indicato anche dalla statistica di Maxwell - Boltzman). Il fatto che $N_2 < N_1$ comporta il prevalere del rate di assorbimento di fotoni, da parte del materiale, rispetto al rate di emissione stimolata, e infatti il segno positivo del differenziale di $N_2(t)$ indica proprio l'incremento, nel tempo, della popolazione del livello energetico E_2 , ossia l'aumento della concentrazione di elettroni in BC (elettroni promossi in BC dall'assorbimento fotonico netto). La Line Shape $g(\nu)$ rappresenta, qualitativamente, lo spettro di assorbimento e di emissione stimolata, dipendente dal semiconduttore che viene irradiato dalla luce, dal suo drogaggio e dalla sua temperatura.

5.3.2) Calcolo del profilo spaziale dell'intensità luminosa all'interno di un fotorivelatore a semiconduttore

Dalla relazione [e43] è intuitivo ricavare che la probabilità P_{transiz} con cui un fotone incidente, di energia $h\nu$, può stimolare un assorbimento o un'emissione è pari a:

$$P_{\text{transiz}} = B_{12} \frac{I_\nu}{c} \quad [s^{-2}]$$

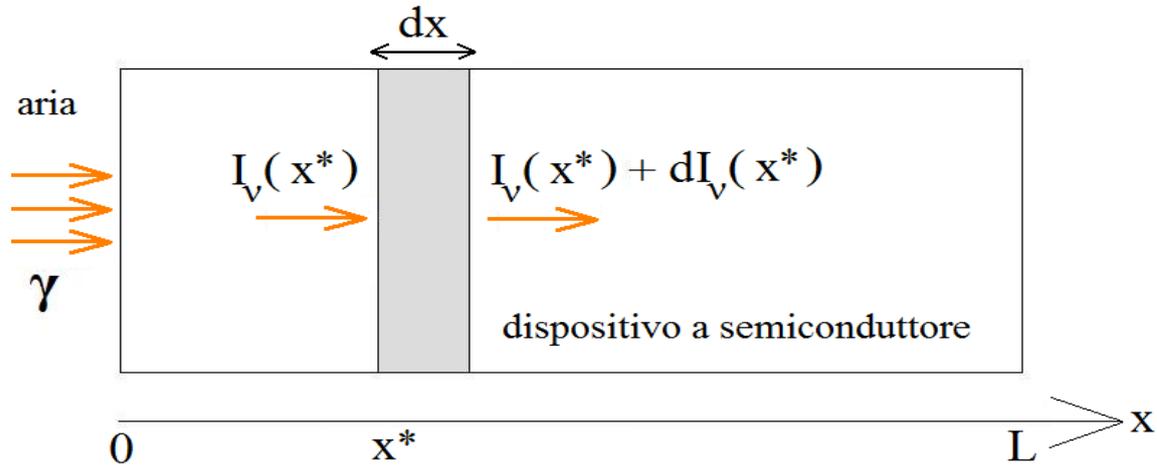


Fig. 36

Rappresentazione qualitativa dell'attraversamento di un generico dispositivo (fotorivelatore ad esempio) a semiconduttore da parte di una radiazione luminosa policromatica: $I_\nu(x^*)$ è lo spettro dell'intensità ottica presso la sezione del dispositivo posizionata alla generica coordinata x^* . dx è l'elemento di lunghezza del dispositivo presso il quale valutiamo l'incremento o il decremento infinitesimo della suddetta intensità luminosa. Partendo dall'analisi di questa situazione è possibile ricavare l'andamento dell'intensità ottica $I_\nu(x)$ che attraversa il dispositivo.

Consideriamo un fotorivelatore realizzato con il semiconduttore fin qui esaminato e modellizzato mediante il sistema a due livelli di energia E_1 ed E_2 . La luce policromatica, che investe il dispositivo, attraversa il materiale lungo la sua profondità. Le interazioni dei fotoni con i portatori di carica sono, prevalentemente, l'assorbimento e l'emissione stimolata, con una netta prevalenza della prima, a causa della relazione esistente fra la popolazione del livello energetico stabile (N_1) e quella del livello instabile (N_2). Consideriamo la sezione del dispositivo presso la generica coordinata x^* ($\forall x^*$ tale che $0 < x^* < L$). Presso questa sezione è possibile misurare un'intensità ottica $I_\nu(x^*)$, relativa ad una certa frequenza ν , mentre presso la sezione posizionata alla coordinata $x^* + dx$ misuriamo un'intensità ottica $I_\nu(x^*) + dI_\nu(x^*)$. Il differenziale di $I_\nu(x^*)$ ovviamente è negativo a causa del depauperamento del flusso fotonico causato dall'assorbimento netto da parte del semiconduttore. Se consideriamo un incremento spaziale elementare dx , di lunghezza infinitesima, è possibile attribuire, sia all'assorbimento che all'emissione stimolata, un contributo elementare, granulato, al flusso fotonico che attraversa il volume che ha come superficie la sezione del dispositivo e come altezza dx . In altre parole sarà possibile avere, lungo dx , l'assorbimento di , al più, un fotone, e l'emissione stimolata de , al più, un fotone. Scomponiamo quindi $dI_\nu(x^*)$ nei suoi due contributi: $dI_\nu(x^*)_{\text{ABS}} (< 0)$, che è il differenziale di $I_\nu(x^*)$ dovuto esclusivamente all'assorbimento fotonico, e $dI_\nu(x^*)_{\text{EM}} (> 0)$, che è il differenziale di $I_\nu(x^*)$ dovuto esclusivamente all'emissione stimolata. Possiamo scrivere quanto segue:

$$dI_\nu(x^*)_{\text{ABS}} = - h\nu \left[B_{12} \frac{I_\nu}{c} \right] N_1 dx g(\nu) \quad [e44]$$

– $h\nu$ è il contributo energetico dell'assorbimento nei confronti del flusso fotonico (ribadiamo che la decurtazione energetica $h\nu$ pari ad un singolo fotone è conseguenza della natura infinitesima di dx). La quantità compresa fra parentesi quadre è la probabilità di transizione (P_{transiz}), ossia la probabilità che l'assorbimento del fotone si verifichi. $N_1 dx$ è la concentrazione superficiale degli elettroni che potrebbero essere coinvolti dall'assorbimento

ottico, ovvero che potrebbero essere promossi in BC. In altre parole è il numero di elettroni per unità di superficie che possono prendere parte all'assorbimento. $g(\nu)$ è lo spettro di assorbimento del materiale.

Possiamo altresì scrivere che:

$$dI_\nu(x^*)_{EM} = + h\nu \left[B_{12} \frac{I_\nu}{c} \right] N_2 dx g(\nu) \quad [e45]$$

+ $h\nu$ è il contributo energetico dell'emissione stimolata al flusso fotonico. La quantità compresa fra parentesi quadre è la probabilità di transizione ($P_{transiz}$), ossia la probabilità che l'emissione stimolata del fotone si verifichi. $N_2 dx$ è la concentrazione superficiale degli elettroni che potrebbero essere coinvolti dall'emissione stimolata, ovvero che potrebbero essere stimolati a decadere in BV. $g(\nu)$ è lo spettro di emissione stimolata del materiale. Conseguentemente il differenziale totale $dI_\nu(x^*)$ dell'intensità luminosa alla frequenza ν , presso la sezione del dispositivo $x^* + dx$, è dato dalla somma dei differenziali parziali [e44] ed [e45]:

$$dI_\nu(x^*) = dI_\nu(x^*)_{ABS} + dI_\nu(x^*)_{EM} = \left\{ \frac{h\nu}{c} B_{12} g(\nu) \right\} (N_2 - N_1) I_\nu dx \quad [e46]$$

Il termine tra parentesi graffe è noto come "sezione di cattura" $\sigma(\nu)$ e può essere interpretato, da un punto di vista semiclassico, come la superficie S dentro la quale un fotone deve entrare affinché un elettrone che si trova energeticamente sul livello E_1 , e spazialmente al centro di S , possa "saltare" sul livello E_2 , decurtando il fotone dal flusso luminoso. Analogamente $\sigma(\nu)$ può essere vista come la superficie S dentro la quale un fotone γ deve entrare affinché un elettrone che si trova energeticamente sul livello E_2 , e spazialmente al centro di S , possa decadere sul livello E_1 , emettendo un fotone γ_g gemello di γ . Possiamo quindi evincere dalla [e46] la seguente equazione differenziale:

$$\frac{dI_\nu(x^*)}{dx} = \sigma(\nu) (N_2 - N_1) I_\nu(x^*) \quad [e47]$$

che può essere presentata anche nel modo seguente:

$$\frac{1}{I_\nu(x^*)} \frac{dI_\nu(x^*)}{dx} = \sigma(\nu) (N_2 - N_1) = -\alpha_0(\nu, x^*) \quad [cm^{-1}] \quad [e48]$$

$\alpha_0(\nu, x^*)$, una cui formulazione è già stata proposta nella [e25], è la "perdita ottica specifica", altrimenti nota come "coefficiente di assorbimento ottico specifico" (cm^{-1}) del materiale, presso la generica coordinata x^* , relativamente ad un fascio di fotoni monocromatico di frequenza ν , nell'ipotesi di "piccolo segnale", ovvero di bassa potenza ottica incidente. La soluzione dell'equazione differenziale, a variabili separabili, riportata nella [e47] è la seguente:

$$I_\nu(x) = I_\nu(0^+) e^{-\int_0^x \alpha_0(\nu, x) dx} \quad \forall x \text{ tale che } 0^+ < x < L$$

Si noti la presenza di $I_\nu(0^+)$ in luogo di $I_\nu(0)$: le due intensità ottiche sono, infatti, tanto più diverse ($I_\nu(0^+) < I_\nu(0)$) quanto maggiore è la discontinuità dell'indice di rifrazione all'interfaccia aria/semiconduttore con cui è costruito il dispositivo. Assumendo che α_0 dipenda solo da ν e non da x , possiamo concludere che:

$$I_\nu(x) = I_\nu(0^+) e^{-\alpha_0(\nu) x} \quad [e49]$$

5.3.3.1) Calcolo dell'espressione generale del coefficiente di assorbimento ottico specifico $\alpha(\nu, x)$ e dell'andamento spaziale dell'intensità luminosa $I_\nu(x)$ per alte potenze ottiche incidenti

La [e42] può essere scritta nella forma seguente:

$$\frac{dN_2(t)}{dt} = R_{pomp} - \frac{N_2(t)}{\tau_{NR}} - \frac{N_2(t)}{\tau_R} - \sigma(\nu) [N_2(t) - N_1(t)] \frac{I_\nu}{h\nu} \quad [e50]$$

dove R_{pomp} rappresenta il rate ($cm^{-3}s^{-1}$) di pompaggio termico, cioè il rate di generazione termica di coppie elettrone/lacuna, a cui il dispositivo è sottoposto dal momento che si trova a temperatura ambiente ($T = 300K$);

pertanto si hanno elettroni atomici che superano la barriera di gap unicamente per motivi termici. Supponiamo che il livello energetico E_1 coincida con il massimo della BV ed il livello energetico E_2 con il minimo della BC. E_1 è un livello instabile, ovvero gli elettroni che vi stazionano tendono a decadere su un livello energetico inferiore, all'interno della BV, molto rapidamente; il tempo di vita medio τ_1 degli elettroni su E_1 è praticamente nullo. E_2 è invece un livello metastabile, ovvero gli elettroni che vi stazionano tendono a decadere sul massimo della BV abbastanza lentamente; il tempo di vita medio τ_{21} degli elettroni su E_2 è abbastanza grande e coincide con il tempo di vita medio τ_R radiativo, del quale abbiamo riportato un grafico (in funzione dell'iniezione di coppie di portatori) in figura 33.

Supponiamo che il fotorelevatore si trovi nella condizione di:

- equilibrio termico: non c'è alcuna corrente di polarizzazione, imposta dall'esterno, che inietta elettroni in BC e lacune in BV, ossia gli elettroni che si trovano in BC sono unicamente quelli termogenerati, intrinseci e/o dovuti al drogaggio, mentre le lacune che si trovano in BV sono unicamente quelle dovute al passaggio degli elettroni suddetti dalla BV alla BC, per motivi termici o di drogaggio
- equilibrio stazionario: le concentrazioni delle popolazioni di portatori, a meno di fluttuazioni granulari stocastiche a media nulla, sono costanti nel tempo
- $\tau_{NR} \gg \tau_R$: il semiconduttore con cui è costruito il fotorelevatore è a gap diretto
- $N_1(t) \rightarrow 0 \forall t$: il livello E_1 è instabile, quindi, approssimativamente, sempre vuoto ($\tau_1 \rightarrow 0$)

La [e50] diventa:

$$0 = R_{pomp} - \frac{N_2}{\tau_R} - \sigma(\nu) N_2 \frac{I_\nu}{h\nu}$$

Se consideriamo il seguente sistema:

$$\begin{cases} 0 = R_{pomp} - \frac{N_2}{\tau_R} - \sigma(\nu) N_2 \frac{I_\nu}{h\nu} \\ \sigma(\nu) N_2 = \alpha(\nu, x) \end{cases}$$

dapprima sotto l'ipotesi di $I_\nu \rightarrow 0$, ovvero per basse intensità ottiche incidenti, e successivamente sotto l'ipotesi di $I_\nu \neq 0$, ovvero per qualsiasi valore dell'intensità ottica incidente (anche elevato), si ricava un'espressione più generale del coefficiente di assorbimento ottico specifico $\alpha(\nu, x)$, valida sia in regime di piccoli che di grandi segnali.

$$\alpha(\nu, x) = \frac{\alpha_0(\nu, x)}{1 + \frac{I_\nu(x)}{I_s} \overline{g(\nu)}} \quad [e51]$$

con:

$$\overline{g(\nu)} = \frac{g(\nu)}{g(\nu_0)}$$

$\overline{g(\nu)}$ rappresenta la Line Shape normalizzata rispetto al proprio valore massimo, il quale si trova in corrispondenza della frequenza di centro banda ν_0 . I_s rappresenta invece la minima intensità ottica di saturazione, ed è una quantità definita nel modo seguente:

$$I_s = \frac{h\nu_0}{\sigma(\nu_0) \tau_{eq}}$$

dove τ_{eq} costituisce una media dei tempi caratteristici di tutte le transizioni possibili all'interno del complesso sistema energetico formato dalla BV e dalla BC, ovvero una media dei tempi $\tau_{m \rightarrow n}$, dove E_m ed E_n sono due generici livelli energetici all'interno del diagramma a bande E - k: possono essere entrambi in BV, entrambi in BC oppure uno in BV e l'altro in BC.

La [e51] mette in evidenza che in regime di piccoli segnali, ovvero per $I_\nu(x) \ll I_s$, il coefficiente di assorbimento ottico specifico $\alpha(\nu, x)$ tende ad $\alpha_0(\nu, x)$, mentre in regime di grandi segnali, ovvero per $I_\nu(x)$ confrontabile con I_s , $\alpha(\nu, x) < \alpha_0(\nu, x)$. Per avere una conferma, seppur intuitiva, dell'ultima affermazione basti pensare che un flusso fotonico molto intenso ($I_\nu(x)$ alto) viene depauperato dal processo di assorbimento, quindi consente ad un cospicuo numero di elettroni atomici di compiere la transizione $BV \rightarrow BC$, ma al tempo stesso presenta buone probabilità di stimolare la transizione contraria $BC \rightarrow BV$ per una parte degli elettroni suddetti (appena giunti in

BC), dato che i fotoni non assorbiti sono ancora tanti e al tempo stesso gli elettroni promossi in BC sono in numero altrettanto elevato. Quindi ad un buon rate di assorbimento si accompagna un altrettanto elevato rate di emissione stimolata, con conseguente diminuzione dell'assorbimento ottico netto e del coefficiente $\alpha(v,x)$ rispetto ad $\alpha_0(v,x)$. Tanto più $I_v(x)$ è non trascurabile rispetto a I_s , tanto più $\alpha(v,x) < \alpha_0(v,x)$.

Dalla [e51] e dalla definizione di assorbimento ottico specifico, la cui forma valida per piccoli segnali è riportata nella [e48], ricaviamo l'equazione differenziale completa per l'intensità luminosa $I_v(x)$ all'interno di un mezzo fotoassorbente con il quale è realizzabile un fotorivelatore:

$$\frac{1}{I_v(x)} \frac{dI_v(x)}{dx} = \frac{\alpha_0(v,x)}{1 + \frac{I_v(x)}{I_s} g(v)} \quad [e52]$$

Nel caso di $I_v(x) \ll I_s$ la soluzione della [e52] è la già calcolata [e49], mentre per $I_v(x)$ confrontabile con I_s la soluzione è la seguente:

$$I_v(x) = I_v(0^+) - \frac{I_s \alpha_0(v)}{g(v)} x \quad [e53]$$

Dunque sotto l'ipotesi di potenza ottica incidente non bassa il profilo spaziale dell'intensità luminosa, attraverso il semiconduttore, non assume la forma di un esponenziale decrescente, come avviene invece nel caso di potenza ottica incidente bassa, bensì assume un andamento rettilineo con coefficiente angolare negativo.

Tuttavia in numerose applicazioni il segnale ottico da rivelare è molto debole, ovvero molto poco intenso e di lunghezze d'onda grandi, come ad esempio:

- la trasmissione di segnali ottici in terza finestra (l'unica finestra equalizzabile mediante l'opportuno dimensionamento del raggio del core della fibra stessa) su fibre "di transito", quindi monomodali, lunghe anche centinaia o migliaia di chilometri, lungo le quali la tratta fra un ripetitore ed il successivo ha una lunghezza media di 80 km (86 km, lungo la linea internazionale FLAG - "Fiber Optical Link Across the Globe")
- le misure telemetriche, nelle quali un impulso di luce, coerente (se l'apparecchiatura utilizzata per la fotometria implementa il metodo di rivelamento "Pulse Delay" - "ritardo dell'impulso") o meno (se l'apparecchiatura consiste in un meometro), deve essere inviato su un obiettivo catarifrangente lontano anche decine o centinaia di chilometri, il quale riflette la luce verso il dispositivo da cui questa è partita; sull'apparecchiatura è montato, oltre al fotoemettitore, anche un fotorivelatore
- il mapping di tessuti organici in ambito diagnostico PET
- la visione notturna attraverso l'utilizzo di un'apparecchiatura binoculare capace di rivelare ed elaborare il calore emesso dai corpi, sottoforma di luce FIR ("far infra red" - luce del lontano infrarosso), per riprodurre immagini degli stessi

Quindi, da qui in avanti, quando considereremo il coefficiente di assorbimento ottico specifico $\alpha(v) = \alpha(\hbar\omega)$ (spesso espresso anche come $\alpha(\lambda)$, per ragioni storiche), ci riferiremo sempre a quello valido per piccoli segnali $\alpha_0(\hbar\omega)$. Il profilo spaziale dell'intensità ottica $I_v(x)$ assumerà, ai fini della nostra trattazione, sempre una forma ad esponenziale decrescente, per qualunque tipo di fotorivelatore, PIN incluso.

5.3.3.2) Calcolo del profilo spettrale del coefficiente di assorbimento ottico specifico $\alpha_0(\hbar\omega)$

Adesso ricaviamo lo spettro del coefficiente di assorbimento ottico specifico $\alpha_0(\hbar\omega)$, riferendoci alla figura 37.

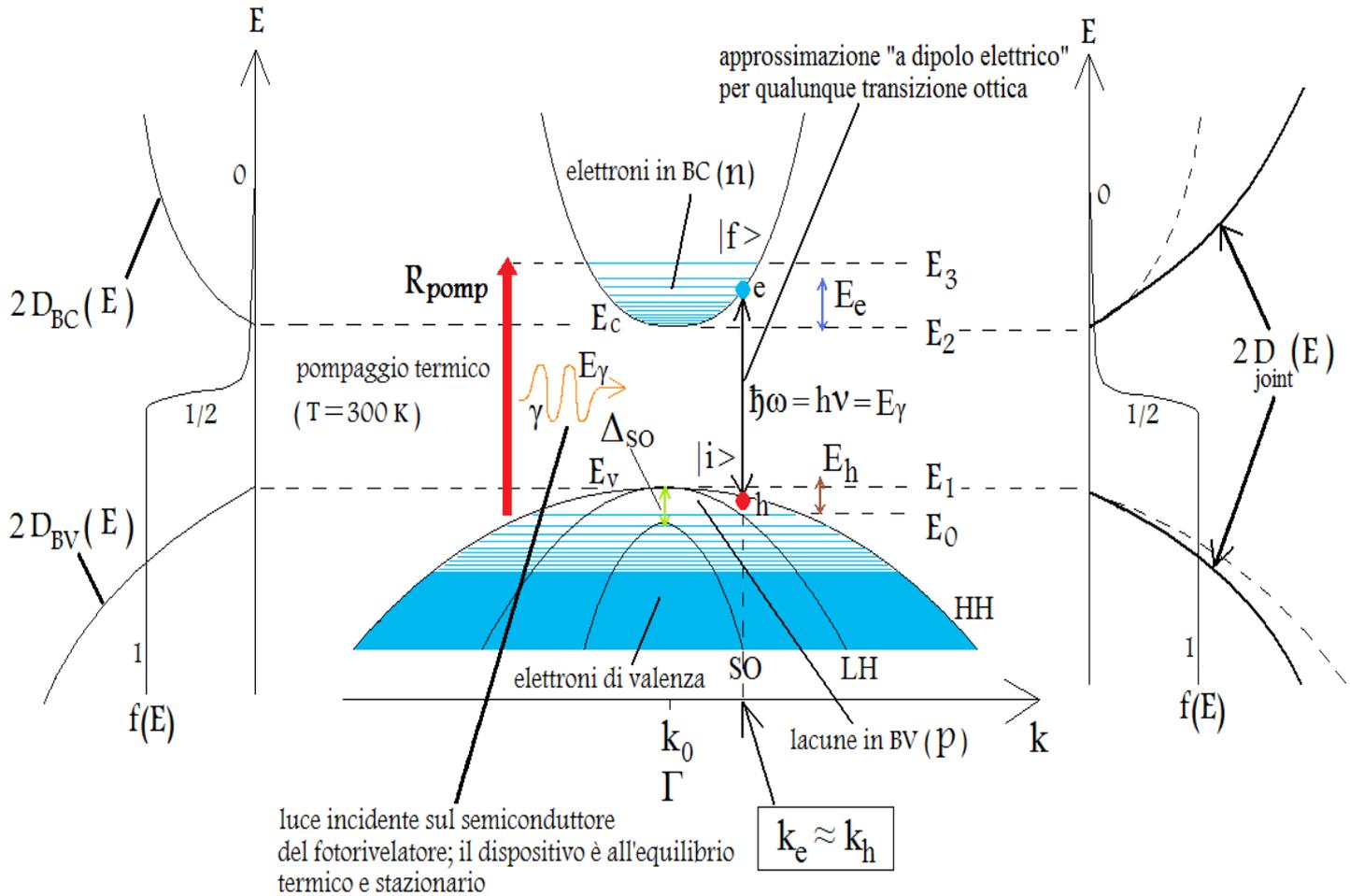


Fig. 37

Adattamento del modello a bande, valido per un semiconduttore (ad esempio a gap diretto, come il GaAs), ad un sistema energetico a quattro livelli, indicati con E_0, E_1, E_2 ed E_3 . E_0 è il livello fondamentale, E_1 è un livello instabile, E_2 è il livello metastabile, E_3 è un altro livello instabile. R_{pomp} costituisce il rate ($\text{cm}^{-3}\text{s}^{-1}$) del pompaggio termico (da E_0 a E_3) a cui il dispositivo è sottoposto, poiché quest'ultimo si trova a temperatura ambiente. $D_{\text{BC}}(E)$ e $D_{\text{BV}}(E)$ sono, rispettivamente, la densità di stati tridimensionale ($\text{cm}^{-3}\text{eV}^{-1}$) in BC ed in BV. $D_{\text{joint}}(E)$ è la densità di stati tridimensionale "congiunta" nel caso in cui il dispositivo sia di tipo bulk. L'utilità di $D_{\text{joint}}(E)$ consiste nel simmetrizzare una distribuzione di stati asimmetrica fra BC e BV, quindi nel semplificare la modellizzazione del profilo spettrale di $\alpha_0(\hbar\omega)$. $\hbar\omega$ è l'energia fotonica assorbita che provoca una transizione verticale (approssimazione a dipolo elettrico) di un elettrone atomico dalla BV alla BC. In seguito a tale assorbimento ottico E_e è l'energia cinetica dell'elettrone in BC, mentre E_h è l'energia cinetica della lacuna in BV.

È possibile adattare il modello a bande, valido per un semiconduttore, ad un sistema energetico (di elettroni) a quattro livelli, indicati con E_0, E_1, E_2 ed E_3 , come mostrato in figura 37. E_0 è il livello fondamentale, $E_1 (= E_V)$ è un livello instabile, $E_2 (= E_C)$ è il livello metastabile, E_3 è un altro livello instabile. R_{pomp} rappresenta il rate ($\text{cm}^{-3}\text{s}^{-1}$) del pompaggio termico (da E_0 a E_3), cioè il rate di generazione termica di coppie elettrone/lacuna a cui il dispositivo è sottoposto, trovandosi a temperatura ambiente. $D_{\text{BC}}(E)$ e $D_{\text{BV}}(E)$ sono, rispettivamente, la densità di stati tridimensionale ($\text{cm}^{-3}\text{eV}^{-1}$) in BC ed in BV. $f(E)$ è la distribuzione statistica quantistica di Fermi - Dirac. $D_{\text{joint}}(E)$ rappresenta la ben nota densità di stati tridimensionale "congiunta" ("joint") per un dispositivo di tipo bulk; l'utilità di $D_{\text{joint}}(E)$ consiste nel simmetrizzare una distribuzione di stati permessi asimmetrica fra BC e BV. L'asimmetria è dovuta alla disegualianza fra la massa efficace dell'elettrone per la densità di stati in BC (m_e^*) e la massa efficace della lacuna per la densità di stati in BV (m_h^*), ovvero è dovuta alla disegualianza fra la densità di stati equivalenti in BC (N_C) e la densità di stati equivalenti in BV (N_V). Pertanto $D_{\text{joint}}(E)$ è utile nel semplificare la modellizzazione del profilo spettrale qualitativo del coefficiente di assorbimento ottico specifico $\alpha_0(\hbar\omega)$.

Prima di tutto riprendiamo il rate di assorbimento riportato nella [e41] e applichiamo al caso specifico di fotorivelatore a semiconduttore di tipo bulk:

$$W_{\text{ABS}}(|\mathbf{i}\rangle) = \frac{2\pi}{\hbar} \frac{e^2}{m_0^2} \frac{\hbar n_{ph}}{2\omega \epsilon_0 \epsilon_r} \sum_{|\mathbf{f}\rangle} \left| \int \psi_f(\mathbf{r})^* [\hat{\mathbf{a}} \cdot \vec{\mathbf{p}}] e^{+i(\mathbf{k}_\gamma \cdot \mathbf{r})} \psi_i(\mathbf{r}) d^3r \right|^2 \delta(|E_f - E_i| - \hbar\omega)$$

dove $\psi_i(\mathbf{r})$ è lo stato elettronico iniziale, nell'esempio di figura 36 posizionato sulla sottobanda di valenza HH, di energia E_i , $\psi_f(\mathbf{r})$ è lo stato elettronico finale di energia E_f , $\vec{\mathbf{p}}$ è l'impulso dell'elettrone prima dell'assorbimento, cioè quando si trova ancora nello stato $\psi_i(\mathbf{r})$, $\hat{\mathbf{a}}$ è il versore di polarizzazione del fotone incidente γ (il quale fa parte di un fascio coerente di n_{ph} fotoni). Considerando di utilizzare un fotorivelatore di tipo bulk, sappiamo che i due stati $\psi_i(\mathbf{r})$ e $\psi_f(\mathbf{r})$ sono funzioni di Bloch, ovvero:

$$\psi_i(\mathbf{r}) = u_p(\mathbf{r}) e^{i(\mathbf{k}_h \cdot \mathbf{r})}$$

$$\psi_f(\mathbf{r}) = u_s(\mathbf{r}) e^{i(\mathbf{k}_e \cdot \mathbf{r})}$$

$u_p(\mathbf{r})$ è la "central cell function" avente simmetria di tipo p, essendo lo stato elettronico $\psi_i(\mathbf{r})$ iniziale (di pre - assorbimento) appartenente alla BV; $u_p(\mathbf{r})$ è un'opportuna combinazione lineare delle funzioni del sistema completo di Clebsch - Gordan. \mathbf{k}_h è il vettore d'onda della lacuna fotogenerata, ovvero dell'elettrone nel suo stato iniziale. $u_s(\mathbf{r})$ è la "central cell function" avente simmetria di tipo s (poiché il gap è diretto), essendo lo stato elettronico $\psi_f(\mathbf{r})$ finale (di post - assorbimento) appartenente alla BC; $u_s(\mathbf{r})$ è un'opportuna combinazione lineare delle funzioni $|\sigma \uparrow\rangle$ e $|\sigma \downarrow\rangle$. \mathbf{k}_e è il vettore d'onda dell'elettrone fotogenerato, ovvero dell'elettrone nel suo stato finale. Utilizzando l'approssimazione di dipolo elettrico abbiamo che:

$$\mathbf{k}_e \approx \mathbf{k}_h \approx \mathbf{k} \gg \mathbf{k}_\gamma \approx 0$$

ovvero la transizione ottica di assorbimento è di fatto verticale nel diagramma E - k. Ciò consente di esprimere il rate di assorbimento fotonico nel modo seguente:

$$W_{\text{ABS}}(|\mathbf{i}\rangle) = \frac{2\pi}{\hbar} \frac{e^2}{m_0^2} \frac{\hbar n_{ph}}{2\omega \epsilon_0 \epsilon_r} \sum_{|\mathbf{f}\rangle} \left| \int u_s(\mathbf{r}) [\hat{\mathbf{a}} \cdot \vec{\mathbf{p}}] u_p(\mathbf{r}) d^3r \right|^2 \delta(|E_f - E_i| - \hbar\omega) \quad [\text{e54}]$$

dove:

$$\int u_s(\mathbf{r}) [\hat{\mathbf{a}} \cdot \vec{\mathbf{p}}] u_p(\mathbf{r}) d^3r = \langle u_s | \hat{\mathbf{a}} \cdot \vec{\mathbf{p}} | u_p \rangle \quad [\text{e55}]$$

è l'elemento di matrice dell'impulso dell'elettrone, ovvero una regola di selezione, che determina (quantifica) l'accoppiamento ottico, in questo caso di assorbimento, fra la "central cell function p - like" dello stato iniziale in BV e la "central cell function s - like" dello stato finale in BC. Nei fotorivelatori comprendenti nanostrutture attive (di cui non ci occupiamo), come "quantum wells", "quantum wires" e "quantum dots", nelle quali avviene la fotogenerazione di portatori di carica, le regole di selezione sono più complicate di quella mostrata in [e55], che ribadiamo vale solo per un fotorivelatore di tipo bulk.

Valgono inoltre le seguenti relazioni, assumendo valida l'approssimazione parabolica per la dispersione E(k) intorno ai "bandedges" E_C ed E_V :

$$\hbar\omega - E_{\text{gap}} = E_e + E_h = \frac{\hbar^2 k^2}{2m_e^*} + \frac{\hbar^2 k^2}{2m_h^*} = \frac{\hbar^2 k^2}{2m_r^*}$$

dove m_r^* è la "massa efficace ridotta", definita nel modo seguente:

$$m_r^* = \frac{m_e^* m_h^*}{m_e^* + m_h^*}$$

m_r^* è quindi leggermente inferiore a m_e^* : ad esempio per il GaAs m_e^* , nella valle Γ , vale circa $0.067m_0$, mentre m_r^* , nella stessa valle, vale circa $0.058m_0$. Pertanto abbiamo che:

$$E_{|\mathbf{f}\rangle} = E_f = E_C + \frac{m_r^*}{m_e^*} (\hbar\omega - E_{\text{gap}})$$

$$E_{|i\rangle} = E_i = E_V - \frac{m_r^*}{m_h^*} (\hbar\omega - E_{\text{gap}})$$

La densità di stati tridimensionale congiunta $D_{\text{joint}}(E)$, per un fotorelevatore di tipo bulk, ha la stessa forma delle densità di stati tridimensionali in BC ed in BV, ad eccezione della la massa efficace: troviamo infatti, al posto delle masse efficaci degli elettroni e delle lacune rispettivamente, la massa efficace ridotta.

$$D_{\text{joint}}(E - E_{\text{gap}}) = 2\pi \left(\frac{2m_r^*}{h^2} \right)^{\frac{3}{2}} \sqrt{E - E_{\text{gap}}} \quad (\text{eV}^{-1}\text{cm}^{-3})$$

dove E è riferita all'energia fotonica $E_\gamma = \hbar\omega$ che investe il semiconduttore e che consente l'accoppiamento ottico fra stati in BC e stati in BV; i primi sono separati dai secondi, nel diagramma $E - k$, di un'energia pari proprio ad $\hbar\omega$, la quale è rappresentata dalla lunghezza del dipolo elettrico che ha come cariche i portatori fotogenerati, come si può notare nella figura 37. Pertanto è opportuno esprimere la densità di stati tridimensionale congiunta nel modo seguente:

$$D_{\text{joint}}(\hbar\omega - E_{\text{gap}}) = 2\pi \left(\frac{2m_r^*}{h^2} \right)^{\frac{3}{2}} \sqrt{\hbar\omega - E_{\text{gap}}} \quad [\text{e56}]$$

Supponiamo, come mostrato sempre in figura 37, che un fotone di energia $E_\gamma = \hbar\omega$ incida sul semiconduttore a gap diretto, ad esempio GaAs, con cui è costruito il fotorelevatore. Se $\hbar\omega < E_{\text{gap}}$ il fotone non ha energia sufficiente per essere assorbito dal materiale, quindi lo attraversa senza interagire con questo: si parla di "trasparenza passiva" del materiale nei confronti del fotone. In altre parole la sommatoria $\Sigma_{|f\rangle}$ sugli stati elettronici finali $|f\rangle$ (di arrivo), cioè la sommatoria sugli stati che possono ospitare un elettrone dopo che questo ha assorbito un fotone di energia $\hbar\omega$, è nulla: infatti nel gap non ci sono, almeno nella teoria, stati permessi che possono consentire assorbimenti al di sotto dell' E_{gap} . Pertanto il rate di assorbimento $W_{\text{ABS}}(|i\rangle)$ nel caso di $\hbar\omega < E_{\text{gap}}$ è nullo ($\Sigma_{|f\rangle} \rightarrow 0$). Se $\hbar\omega = E_{\text{gap}}$ siamo sempre in presenza di trasparenza passiva, dal momento che lo stato individuato sul massimo della BV (ad energia E_V) è sicuramente vuoto, privo di elettrone: infatti è altamente improbabile, in condizioni normali, che ci sia un elettrone atomico di valenza avente velocità nulla (quindi quantità di moto nulla, quindi numero d'onda nullo) lungo le tre direzioni del cristallo semiconduttore (nel punto di simmetria Γ abbiamo $(k_x, k_y, k_z) = (0, 0, 0)$). Inoltre anche lo stato localizzato presso il minimo della BC (ad energia E_C) deve rimanere vuoto (nessun elettrone proveniente dalla BV può occuparlo), dato che è impossibile trovare un elettrone di conduzione avente velocità nulla lungo le tre direzioni suddette. Anche nel caso di $\hbar\omega = E_{\text{gap}}$ la sommatoria $\Sigma_{|f\rangle}$ è di fatto nulla, rendendo nullo il rate $W_{\text{ABS}}(|i\rangle)$. Se $\hbar\omega > E_{\text{gap}}$ abbiamo, in BV, un certo numero di elettroni "candidati" ad essere promossi in BC, a causa dell'assorbimento del fotone di energia $E_\gamma = \hbar\omega$, ed un certo numero di stati liberi in BC, che costituiscono dei possibili stati accettori di elettroni provenienti dalla BV. Stavolta $\Sigma_{|f\rangle} \neq 0$ e quindi $W_{\text{ABS}}(|i\rangle) > 0$: l'assorbimento ottico è possibile e soggetto alla regola di selezione riportata nella [e55]. Se $\hbar\omega$ aumenta abbiamo che il numero di elettroni atomici di valenza "candidati" alla promozione in BC aumenta, ed aumenta anche il numero di stati liberi, in BC, potenziali accettori di elettroni, dato che aumenta, secondo la radice quadrata di $E = \hbar\omega$, la densità di stati congiunta $D_{\text{joint}}(E)$; infatti aumentano, congiuntamente, la densità di stati in BV e quella in BC. Più $\hbar\omega$ diventa maggiore di E_{gap} e più la presenza di elettroni in BV e la "capacità di accettazione elettronica" degli stati in BC aumentano, ovvero aumenta la sommatoria $\Sigma_{|f\rangle}$ e quindi il rate $W_{\text{ABS}}(|i\rangle)$. Adesso risulta evidente che nella [e54] la sommatoria $\Sigma_{|f\rangle}$ sugli stati finali $|f\rangle$ può essere sostituita con la densità di stati congiunta espressa nella [e56]:

$$W_{\text{ABS}}(|i\rangle) = \frac{2\pi}{\hbar} \frac{e^2}{m_0^2} \frac{\hbar n_{ph}}{2\omega \epsilon_0 \epsilon_r} \left| \langle u_s | \hat{a} \cdot \vec{p} | u_p \rangle \right|^2 \{ 2D_{\text{joint}}(\hbar\omega - E_{\text{gap}}) \} \quad [\text{e57}]$$

Il 2 è aggiunto, ovviamente, per tenere conto della degenerazione di spin. Dalla [e57] è immediato definire il coefficiente di assorbimento ottico specifico per piccoli segnali $\alpha_0(\hbar\omega)$: è sufficiente dividere il rate $W_{\text{ABS}}(|i\rangle)$ di assorbimento per il numero n_{ph} di fotoni che possono essere coinvolti dalla transizione ottica di assorbimento e per la velocità $v_{ph} = c/n_r''(\omega)$ con la quale i fotoni viaggiano nel cristallo semiconduttore, dove $n_r''(\omega)$ è la parte immaginaria dell'indice di rifrazione del semiconduttore. In altri termini al numeratore della [e57] troviamo la parte immaginaria $\epsilon_{r2}(\omega)$ della costante dielettrica relativa del mezzo permeato dalla luce, mentre al denominatore troviamo la sua parte reale $\epsilon_{r1}(\omega)$. In formule:

$$\alpha_0(\hbar\omega) = \frac{2\pi}{\hbar} \frac{e^2}{m_0^2} \frac{\hbar n_r''(\omega)}{2\omega c \epsilon_0 \epsilon_{r1}(\omega)} \left| \langle u_s | \hat{a} \cdot \vec{p} | u_p \rangle \right|^2 \{ 2D_{\text{joint}}(\hbar\omega - E_{\text{gap}}) \} \quad [\text{e58}]$$

Va tenuto presente che il rate di assorbimento ottico $W_{\text{ABS}}(|\mathbf{i}\rangle)$ definisce rigorosamente l'assorbimento $\alpha_0(\hbar\omega)$ per piccoli segnali, e non per qualunque potenza ottica incidente, poiché l'espressione [e57] è stata calcolata partendo dalla regola d'oro di Fermi, nella quale il termine hamiltoniano H_F fotoperturbativo ha una forma che è diretta conseguenza sia delle gauges di Lorentz e di Coulomb che del piccolo modulo A_0 del potenziale vettore \vec{A} : affinché quest'ultima condizione risulti verificata è necessario che la potenza ottica propagante attraverso il cristallo sia sufficientemente bassa (condizione di piccoli segnali ottici). Tuttavia la presenza della parte reale $\varepsilon_{r1}(\omega)$ della costante dielettrica relativa del semiconduttore ci consente di osservare che effettivamente, come mostrato dalla formula riportata nella [e51], all'aumentare del flusso fotonico propagante si ha un decremento del coefficiente di assorbimento ottico. Infatti maggiore è la potenza ottica incidente, maggiore è il numero di transizioni elettroniche $BV \rightarrow BC$, maggiore è la parte reale $n_r(\omega)$ dell'indice di rifrazione, cioè maggiore è $\varepsilon_{r1}(\omega)$, e minore è il coefficiente di assorbimento $\alpha(\hbar\omega)$.

Sempre dalla [e58] si può evincere un altro risultato di grande importanza pratica: un fotorivelatore presenta un coefficiente di assorbimento ottico specifico $\alpha_0(\hbar\omega)$ che varia in funzione della polarizzazione della luce incidente. Ciò è dovuto alla fluttuazione, nel tempo, della direzione sulla quale giace il versore \hat{a} di polarizzazione della luce incidente, che comporta la fluttuazione, nel tempo, del valore dell'elemento di matrice $\langle u_s | \hat{a} \cdot \vec{p} | u_p \rangle$. Più precisamente, ipotizziamo che una luce incidente polarizzata linearmente lungo un ipotetico asse j ($= x, y, z$) consenta, se assorbita, il salto energetico del gap unicamente ad elettroni atomici localizzati presso gli orbitali di tipo p_j , ovvero un fotone polarizzato lungo x può, se assorbito, promuovere un elettrone di valenza da uno stato iniziale in BV , la cui "central cell function" è di tipo p_x , ad uno stato finale in BC , la cui "central cell function" è di tipo s . Allo stesso modo un fotone polarizzato lungo y può, se assorbito, promuovere un elettrone di valenza da uno stato iniziale in BV , la cui "central cell function" è di tipo p_y , ad uno stato finale in BC , la cui "central cell function" è di tipo s ; inoltre un fotone polarizzato lungo z può, sempre se assorbito, promuovere un elettrone di valenza da uno stato iniziale in BV , la cui "central cell function" è di tipo p_z , ad uno stato finale in BC , la cui "central cell function" è sempre di tipo s . Inoltre ipotizziamo che la probabilità che un fotone assorbito, polarizzato lungo x , consenta la transizione elettronica $u_{p_x} \rightarrow u_s$ è pari alla probabilità che un fotone assorbito, polarizzato lungo y , consenta la transizione elettronica $u_{p_y} \rightarrow u_s$, la quale è pari alla probabilità che un fotone assorbito, polarizzato lungo z , consenta la transizione elettronica $u_{p_z} \rightarrow u_s$. In formule:

$$p_{cv} = \langle u_s | \hat{P}_x | u_{p_x} \rangle = \langle u_s | \hat{P}_y | u_{p_y} \rangle = \langle u_s | \hat{P}_z | u_{p_z} \rangle \quad [e59]$$

Ipotizziamo pertanto che qualunque altro accoppiamento ottico sia trascurabile, ovvero:

$$\langle u_s | \hat{P}_j | u_{p_i} \rangle = 0 \quad \forall i \neq j \quad [e60]$$

p_{cv} è un elemento di matrice il cui valore cambia sensibilmente da semiconduttore a semiconduttore; generalmente non si trovano tabulati i valori di p_{cv} , bensì quelli di una quantità legata univocamente a p_{cv} , ovvero l'energia E_P ("momentum matrix element energy" - energia dell'elemento di matrice della quantità di moto p), che abbiamo già citato in riferimento alla [e15]. Ricordiamo a titolo di esempio che per il GaAs si ha che:

$$E_P(\text{GaAs}) = \frac{2}{m_0} p_{cv}^2 \approx 25.7 \text{ eV}$$

Utilizzando la [e59] e la [e60] possiamo calcolare i momenti di matrice dei seguenti assorbimenti ottici:

- accoppiamento ottico fra la sottobanda di valenza HH e la banda di conduzione dovuto ai tre differenti stati di polarizzazione (lineare) della luce incidente sul fotorivelatore:

$$\langle u_s | \hat{P}_x | u_{HH} \rangle = \frac{p_{cv}}{\sqrt{2}}$$

$$\langle u_s | \hat{P}_y | u_{HH} \rangle = \frac{p_{cv}}{\sqrt{2}}$$

$$\langle u_s | \hat{P}_z | u_{HH} \rangle = 0$$

- accoppiamento ottico fra la sottobanda di valenza LH e la banda di conduzione dovuto ai tre differenti stati di polarizzazione (lineare) della luce incidente sul fotorivelatore:

$$\langle u_s | \hat{P}_x | u_{LH} \rangle = \frac{p_{cv}}{\sqrt{6}}$$

$$\langle u_s | \hat{P}_y | u_{LH} \rangle = \frac{p_{cv}}{\sqrt{6}}$$

$$\langle u_s | \hat{P}_z | u_{LH} \rangle = 2 \frac{p_{cv}}{\sqrt{6}}$$

dove u_{HH} e u_{LH} sono due “central cell functions” date da opportune combinazioni lineari degli stati ortonormali di Clebsch – Gordan $|j; j_z\rangle$, come qualitativamente mostrato qui di seguito ($\alpha, \beta, \eta, \delta \in \mathbb{C}$):

$$|u_{HH}\rangle = \alpha |3/2; 3/2\rangle + \beta |3/2; -3/2\rangle$$

$$|u_{LH}\rangle = \eta |3/2; 1/2\rangle + \delta |3/2; -1/2\rangle$$

Dai sei momenti di matrice precedenti si evince in modo formale la dipendenza del coefficiente di assorbimento ottico specifico $\alpha_0(\hbar\omega)$ dalla polarizzazione della luce. Fissato un opportuno sistema di riferimento cartesiano, solidale con il fotorivelatore, abbiamo che se la luce incidente sulla finestra antiriflettente del fotorivelatore è un’onda piana polarizzata linearmente lungo il piano xy, l’accoppiamento ottico fra la sottobanda di valenza HH e la banda di conduzione è tre volte maggiore dell’accoppiamento ottico fra la sottobanda di valenza LH e la stessa banda di conduzione. In altri termini con polarizzazione ottica lungo il piano xy il numero di elettroni atomici, promossi in BC a causa dell’assorbimento fotonico, provenienti dalla sottobanda di valenza HH è tre volte maggiore del numero di elettroni atomici, promossi in BC dall’assorbimento, provenienti dalla sottobanda di valenza LH. Pertanto la sottobanda di valenza HH contribuisce tre volte di più, rispetto a quella LH, nel fornire i fotoelettroni e quindi la fotocorrente uscente dal fotorivelatore irradiato. Tuttavia se la polarizzazione \hat{a} della luce incidente subisce variazioni (nelle fibre ottiche ciò avviene, benchè non rapidissimamente) anche $\alpha_0(\hbar\omega)$ varia, e ciò comporta una fluttuazione della fotocorrente in uscita dal fotodiodo. Questa fluttuazione potrebbe essere interpretata come una variazione dell’intensità ottica incidente e non come una rotazione del versore \hat{a} di polarizzazione del segnale ottico assorbito dal semiconduttore, pertanto le fluttuazioni della polarizzazione ottica costituiscono una sorgente di rumore. Se poi il versore \hat{a} di polarizzazione assume, in un certo intervallo di tempo Δt , la direzione individuata dall’asse z, la sottobanda di valenza HH e la banda di conduzione risultano totalmente disaccoppiate, dal punto di vista ottico. Ciò significa che se la luce incidente è polarizzata lungo l’asse z i fotoni non possono essere assorbiti da elettroni localizzati su HH: la sottobanda di valenza HH non contribuisce, durante il lasso di tempo Δt , alla fotocorrente. Questa è formata unicamente dagli elettroni atomici provenienti dalla sottobanda di valenza LH: con polarizzazione ottica lungo z il contributo di LH alla fotocorrente è raddoppiato, rispetto al suo contributo nel caso di polarizzazione ottica lungo il piano xy. Quindi è grazie agli stati elettronici posizionati lungo la curva $E(k)$ della sottobanda di valenza LH che il fotorivelatore di tipo bulk risponde alla luce incidente, con una fotocorrente rilevabile dal circuito esterno di polarizzazione (inversa) e lettura, anche se la radiazione è polarizzata completamente lungo z.

In molti semiconduttori a gap diretto l’energia di split – off Δ_{SO} è piuttosto alta (nel GaAs è di circa 0.35 eV), pertanto un qualunque stato elettronico u_{SO} dato dalla combinazione lineare delle funzioni di Clebsch – Gordan $|1/2; 1/2\rangle$ e $|1/2; -1/2\rangle$ è, approssimativamente, ortogonale rispetto agli stati elettronici appartenenti alle sottobande di valenza HH ed LH, e a maggior ragione rispetto agli stati elettronici appartenenti alla banda di conduzione. Pertanto l’accoppiamento ottico fra la sottobanda di valenza SO (“split – off subband” – sottobanda di valenza splittata) e la BC è di fatto nullo: gli elettroni della sottobanda SO non sono promuovibili in BC dallo scattering con i fotoni, qualunque polarizzazione questi abbiano. In formule:

$$\langle u_s | \hat{P}_j | u_{SO} \rangle \approx 0 \quad \forall j \in \{x, y, z\}$$

In figura 38 mostriamo lo spettro qualitativo del coefficiente di assorbimento ottico specifico $\alpha_0(\hbar\omega)$, la cui espressione analitica è riportata nella [e58]. Si noti che la famiglia di curve $\alpha_0(\hbar\omega)$ è parametrizzata dalla temperatura T alla quale si trova il fotorivelatore. Fissata una certa energia ottica incidente $\hbar\omega > E_{gap}$, si verifica che maggiore è T e minore è la concentrazione di elettroni in BV (cioè di stati iniziali $|i\rangle$) e di stati accettori liberi in BC, ovvero minore è la sommatoria $\Sigma_{|f\rangle}$, nella [e54], sugli stati elettronici finali e quindi minore è la probabilità di assorbimento dei fotoni aventi energia $E_f = \hbar\omega$.

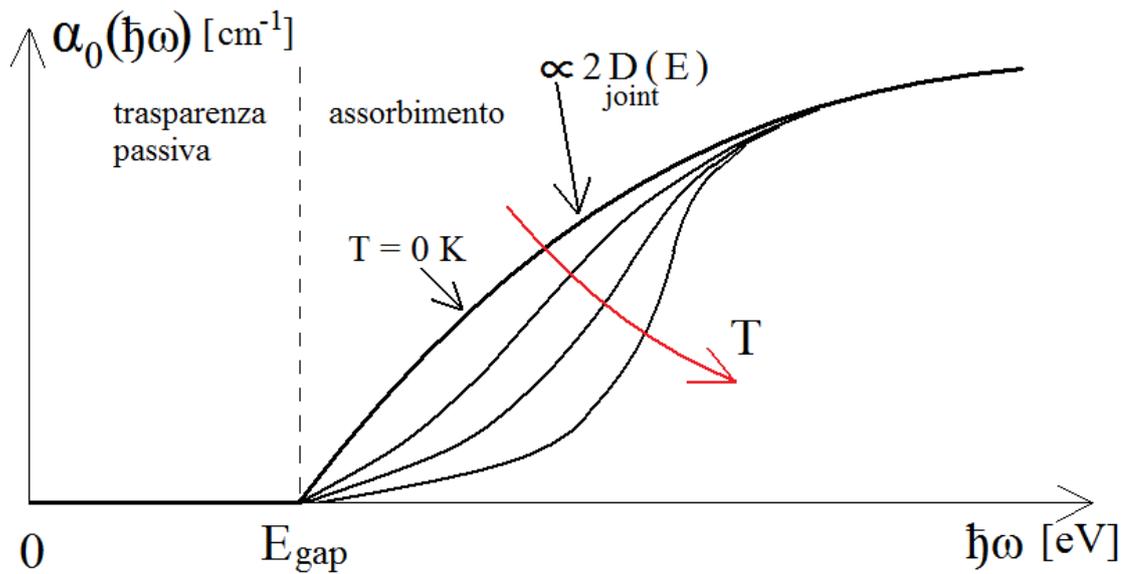


Fig. 38

Spettro qualitativo del coefficiente di assorbimento ottico specifico $\alpha_0(\hbar\omega)$ per un fotorecettore di tipo bulk. Per $\hbar\omega \leq E_{\text{gap}}$ i fotoni incidenti sul fotorecettore non interagiscono con gli elettroni atomici di valenza del semiconduttore, per cui il coefficiente di assorbimento è nullo (i fotoni non hanno energia sufficiente per essere assorbiti): si parla di “trasparenza passiva” del materiale nei confronti della luce incidente. Per $\hbar\omega > E_{\text{gap}}$ i fotoni possono essere assorbiti, e la probabilità di assorbimento degli stessi aumenta all’aumentare della loro energia $\hbar\omega$ rispetto ad E_{gap} : l’incremento di $\alpha_0(\hbar\omega)$ è proporzionale, fissato un certo valore di T (temperatura del fotorecettore), alla densità di stati congiunta tridimensionale $2D_{\text{joint}}(E)$. Fissata una certa energia ottica incidente $\hbar\omega > E_{\text{gap}}$, si verifica che maggiore è T e minore è la concentrazione di elettroni in BV e di stati accettori liberi in BC, quindi minore è la probabilità di assorbimento dei fotoni aventi energia $E_\gamma = \hbar\omega$. Ciò spiega la parametrizzazione della famiglia di curve $\alpha_0(\hbar\omega)$ rispetto a T .

5.3.3.3) Stima delle variazioni spettrali del coefficiente di assorbimento ottico specifico $\alpha_0(\hbar\omega)$ causate dagli eccitoni

Per rendere ancora più accurata la trattazione del coefficiente di assorbimento ottico specifico $\alpha_0(\hbar\omega)$, nonché per disegnarne uno spettro più vicino alla realtà, è opportuno tenere in considerazione il fenomeno spiegato brevemente qui di seguito. Nel disegnare i diagrammi a bande $E - k$ sfruttando uno dei noti metodi di calcolo delle bande di energia (TBM ad esempio), si considera sempre il semiconduttore in una situazione elettronica “di riposo”, nella quale quasi tutti gli elettroni atomici di valenza si trovano nella BV; in altre parole la BV è quasi del tutto occupata dagli elettroni mentre la BC è quasi interamente vuota. Consideriamo adesso una regione del cristallo nella quale abbiamo un elettrone in BC ed una lacuna in BV. L’elettrone di conduzione presenta una “situazione energetica”, ovvero una dispersione $E(k)$, che è determinata principalmente dalla propria interazione coulombiana sia con il nucleo dell’atomo a cui l’elettrone appartiene sia con i nuclei degli atomi primi vicini (oltre che con gli altri elettroni di valenza del cristallo). Queste interazioni, formalmente modellizzabili con l’approccio matematico noto come LCAO (“Linear Combination of Atomic Orbitals” – combinazione lineare di orbitali atomici), sono responsabili della struttura a bande $E(k)$ e delle densità di stati $D_{\text{BC}}(E)$ e $D_{\text{BV}}(E)$ tipiche di quel particolare semiconduttore; in altri termini tali interazioni comportano la degenerazione in bande dei livelli energetici atomici. Tuttavia, nella regione del cristallo prima citata, l’elettrone di conduzione risente anche dell’attrazione elettrostatica esercitata dalla lacuna in BV. L’elettrone in BC e la lacuna in BV costituiscono una sorta di sistema “atom – like”, ovvero possiamo descrivere il sistema a due corpi costituito dalla coppia elettrone/lacuna mediante l’analogia con l’atomo di idrogeno H_1^+ : la lacuna si comporta come il protone nell’atomo di idrogeno, ovvero come il centro delle forze elettriche, cioè come il centro dell’energia potenziale elettrica $U(r)$ nella quale è immerso l’elettrone di conduzione. Questa attrazione elettrostatica è trattabile come una perturbazione coulombiana $\Delta H_{\text{Coulomb}}$ della situazione energetica H_0 dell’elettrone espressa dal diagramma a bande $E - k$. Il diagramma $E(k)$, la densità di stati congiunta $D_{\text{joint}}(E)$ e, conseguentemente, il coefficiente di assorbimento ottico specifico $\alpha_0(\hbar\omega)$ vengono perturbati dall’attrazione elettrostatica aggiuntiva che lega l’elettrone di conduzione alla lacuna. Questo sistema a due corpi “atom – like”, chiamato “eccitone”, introduce degli stati elettronici, non previsti dalla teoria fin qui esposta, sia dentro che fuori dal gap, che possono accogliere elettroni provenienti dalla BV, consentendo ulteriori assorbimenti ottici. Questi ultimi si aggiungono agli assorbimenti banda – banda, cioè agli assorbimenti dipendenti dalle densità di stati $D_{\text{BC}}(E)$ e $D_{\text{BV}}(E)$ e dalla distribuzione statistica di Fermi – Dirac, la quale rivela quali stati sono effettivamente liberi, rispettivamente da elettroni e da lacune.

L'eccitone, che quantisticamente è descritto da una funzione d'onda a quadrato sommabile $\psi_{ex}(x,y,z)$ molto simile ad un'onda di Bloch, può essere rappresentato classicamente, all'interno di uno spazio di coordinate, come un sistema formato da un elettrone, individuato dal vettore posizione r_e , e da una lacuna, individuata dal vettore posizione r_h . Supponendo che l'elettrone si trovi nei pressi del minimo della BC e che la lacuna si trovi nei pressi del massimo della BV, cioè che le rispettive masse efficaci per la mobilità m_e^* e m_h^* siano positive e che si possa utilizzare l'approssimazione parabolica della dispersione $E(k)$, possiamo scrivere l'equazione di Schrödinger per l'eccitone nel modo seguente:

$$\left[-\frac{\hbar^2}{2m_e^*} \nabla_e^2 - \frac{\hbar^2}{2m_h^*} \nabla_h^2 - \frac{e^2}{4\pi\epsilon_0\epsilon_r|r_e - r_h|} \right] \psi_{ex}(x,y,z) = E_{ex} \psi_{ex}(x,y,z) \quad [e61]$$

dove ∇_e^2 è l'operatore laplaciano che agisce sulla dipendenza spaziale dell'elettrone, ∇_h^2 è l'operatore laplaciano che agisce sulla dipendenza spaziale della lacuna, mentre $r = r_e - r_h$ è il vettore posizione dell'elettrone rispetto alla lacuna. Adesso indichiamo con k_e il vettore d'onda dell'elettrone, con k_h il vettore d'onda della lacuna, con R il vettore posizione del centro di massa dell'intero sistema a due corpi, cioè il vettore posizione dell'eccitone: l'elettrone e la lacuna ruotano infatti intorno ad un centro di massa mobile la cui posizione è individuata da R . Supponiamo inoltre che l'eccitone costituisca un sistema isolato, ovvero che la quantità di moto del suo centro di massa rimanga costante nel tempo, ovvero:

$$\frac{\partial^2 R}{\partial t^2} = 0$$

Pertanto l'eccitone può essere visto come una particella massiva libera all'interno della matrice cristallina del semiconduttore, approssimativamente allo stesso modo degli elettroni in BC e delle lacune in BV. Il problema eccitonico a due corpi, fin qui trattato, può essere convertito in un più semplice problema a singolo corpo, considerando il sistema ridotto equivalente formato da una singola particella fittizia, avente una massa pari alla massa ridotta m_r^* del sistema a due corpi, che ruota intorno ad un centro di forze fisso (non più mobile come il centro di massa dell'eccitone) coincidente con l'origine degli assi cartesiani. La velocità di questa particella è la stessa con cui l'elettrone ruota intorno al centro di massa dell'eccitone, il raggio dell'orbita della particella intorno al centro di forze fisso ha la stessa lunghezza r della distanza che separa l'elettrone dalla lacuna, all'interno dell'eccitone. Indichiamo con k il vettore d'onda della singola particella fittizia del sistema ridotto e con k_{ex} il vettore d'onda dell'eccitone. In formule abbiamo:

$$r = r_e - r_h$$

$$k = \frac{m_e^* k_e + m_h^* k_h}{m_e^* + m_h^*}$$

$$R = \frac{m_e^* r_e + m_h^* r_h}{m_e^* + m_h^*}$$

$$k_{ex} = k_e - k_h$$

$$m_r^* = \frac{m_e^* m_h^*}{m_e^* + m_h^*}$$

Utilizzando queste trasformazioni possiamo scrivere la [e61] nel modo seguente:

$$\left[\frac{\hbar^2 k_{ex}^2}{2(m_e^* + m_h^*)} + \left\{ \frac{\hbar^2 k^2}{2m_r^*} - \frac{e^2}{4\pi\epsilon_0\epsilon_r|r|} \right\} \right] \psi_{ex}(x,y,z) = E_{ex} \psi_{ex}(x,y,z) \quad [e62]$$

$$[H_{cm} + \{ H_r \}] \psi_{ex}(x,y,z) = E_{ex} \psi_{ex}(x,y,z)$$

H_{cm} è l'operatore hamiltoniano che descrive l'energia cinetica dell'eccitone, ossia l'energia cinetica del suo centro di massa; l'eccitone è visto, nella fattispecie, come una particella libera avente vettore d'onda k_{ex} e massa $m_e^* + m_h^*$, che è la massa efficace totale del sistema a due corpi. Pertanto H_{cm} descrive il moto dell'eccitone attraverso il cristallo. H_r è l'operatore hamiltoniano che descrive l'energia cinetica e l'energia potenziale della particella fittizia avente massa m_r^* e posta a distanza orbitale r dal centro di forze fisso. Quindi H_r descrive il moto relativo del sistema "atom - like" elettrone - lacuna. È possibile dimostrare che H_{cm} ed H_r commutano:

$$[\hat{H}_{cm}, \hat{H}_r] = \hat{H}_{cm}\hat{H}_r - \hat{H}_r\hat{H}_{cm} = 0$$

Quindi le funzioni che rappresentano le dipendenze di $\psi_{ex}(x,y,z)$ dal moto dell'eccitone all'interno del cristallo e dal moto relativo del sistema elettrone - lacuna fattorizzano:

$$\psi_{ex}(x,y,z) = \varphi_{cm}(x,y,z)\chi_r(x,y,z)$$

dove φ_{cm} è la funzione che descrive il moto del centro di massa dell'eccitone attraverso il cristallo, mentre χ_r è la funzione di involuppo che descrive il problema "atom - like" elettrone - lacuna: χ_r è una sorta di "funzione atomica". Pertanto il problema della propagazione dell'eccitone attraverso il semiconduttore può essere separato dal problema dell'interazione elettrostatica fra elettrone in BC e lacuna in BV: quest'ultimo è formalmente identico al problema del calcolo degli autostati e degli autovalori dell'energia nell'atomo d'idrogeno. φ_{cm} è soluzione della seguente porzione dell'equazione di Schrödinger, per l'eccitone, riportata nella [e62]:

$$\left[\frac{\hbar^2 k_{ex}^2}{2(m_e^* + m_h^*)} \right] \varphi_{cm}(x,y,z) = E_{ex,cin} \varphi_{cm}(x,y,z)$$

dalla quale si ricava che:

$$\varphi_{cm}(x,y,z) = e^{i(k_{ex} \cdot R)}$$

χ_r è soluzione della seguente porzione "atom - like" dell'equazione di Schrödinger riportata nella [e62]:

$$\left[\frac{\hbar^2 k^2}{2m_r^*} - \frac{e^2}{4\pi\epsilon_0\epsilon_r|r|} \right] \chi_r(x,y,z) = E_{ex,n} \chi_r(x,y,z)$$

dalla quale si ricava che:

$$\chi_r(x,y,z) = R_n(r) \Phi_C(r_e) \Phi_V(r_h)$$

dove $R_n(r)$ rappresenta la dipendenza radiale della funzione di involuppo dell'eccitone, formalmente simile alla funzione radiale $R_{nl}(r)$, citata a pagina 21, calcolabile risolvendo l'equazione di Schrödinger nel caso dell'atomo d'idrogeno, o più generalmente nel caso dell'atomo idrogenoide. n è il numero quantico che indicizza l'energia totale dell'elettrone all'interno dell'eccitone. r è il vettore posizione dell'elettrone rispetto alla lacuna, ovvero della particella fittizia, di massa ridotta, rispetto all'origine degli assi. $\Phi_C(r_e)$ è la "central cell function" dell'elettrone in BC, individuato dal vettore posizione r_e : la simmetria di $\Phi_C(r_e)$ è di tipo s oppure di tipo $s + p$ a seconda che il semiconduttore sia a gap diretto o indiretto. $\Phi_V(r_h)$ è la "central cell function" della lacuna in BV, individuata dal vettore posizione r_h : la simmetria di $\Phi_V(r_h)$ è di tipo p per qualunque semiconduttore. Ad un certo istante fissato la funzione involuppo χ_r può essere localizzata, all'interno del cristallo, in una regione che si estende lungo poche celle fondamentali (cioè per $15 \div 30 \text{ \AA}$, dato che la lunghezza "a" della cella fondamentale di un semiconduttore convenzionale è dell'ordine di $5 \div 6 \text{ \AA}$), oppure in una regione del materiale più estesa, ovvero lungo una ventina di celle fondamentali (cioè per circa 100 \AA , approssimativamente). Nel primo caso l'eccitone, più localizzato, consta di un elettrone e di una lacuna ed è chiamato "eccitone di Frenkel", ovvero l'eccitone discusso finora, mentre nel secondo caso l'eccitone, più delocalizzato, consta di diversi elettroni e diverse lacune ed è chiamato "eccitone di Mott".

La soluzione completa della [e62] è pertanto la seguente:

$$\psi_{ex}(x,y,z) = e^{i(k_{ex} \cdot R)} R_n(r) \Phi_C(r_e) \Phi_V(r_h) \quad [e63]$$

A questo autostato eccitonico è associato il seguente autovalore dell'energia totale, data dal contributo $E_{ex,cin}$ dell'energia cinetica del centro di massa dell'eccitone e dal contributo $E_{ex,n}$ dell'energia totale dell'elettrone all'interno del sistema "atom - like".

$$E_{ex} = E_{ex,cin} + E_{ex,n} = \frac{\hbar^2 k_{ex}^2}{2(m_e^* + m_h^*)} - \frac{m_r^* e^4}{2(4\pi\epsilon_0\epsilon_r)^2 \hbar^2 n^2}$$

Assumendo che la transizione dell'elettrone dalla BV alla BC, dovuta a motivi termici o fotonici, si presenti praticamente verticale nel diagramma $E - k$ (approssimazione di dipolo elettrico), possiamo affermare che:

$$k_e \approx k_h$$

pertanto:

$$k_{ex} = k_e - k_h \approx 0$$

Quindi gli autovalori E_{ex} dell'energia totale eccitonica coincidono, di fatto, con gli autovalori $E_{ex,n}$ dell'energia del problema atomico.

$$E_{ex} = E_{ex,n} \approx -\frac{m_r^* e^4}{2(4\pi\epsilon_0\epsilon_r)^2\hbar^2 n^2} = -\frac{R_{ex}}{n^2} \approx -\frac{13.6 \text{ eV}}{\epsilon_r^2 n^2} \left(\frac{m_r^*}{m_0}\right)$$

dove R_{ex} è la costante di Rydberg per l'eccitone. Per la maggior parte dei semiconduttori utilizzati nella realizzazione dei fotorecettori R_{ex} è compresa fra 2 e 6 meV: ad esempio per il GaAs $R_{ex} \approx 4$ meV. Questi livelli energetici discreti "H₁¹ - like", individuati da valori interi positivi del numero quantico n ($n = 1, 2, 3, \dots$) e quadraticamente spazati, individuano degli stati elettronici permessi, i quali si sovrappongono (si aggiungono) agli stati elettronici legati alla BC e alla BV, ovvero alle densità di stati $D_{BC}(E)$ e $D_{BV}(E)$. I livelli eccitonici discreti indicizzati da n consentono delle transizioni ottiche discrete. In particolare avremo degli assorbimenti fotonici discreti, ovvero il semiconduttore assorbirà delle righe spettrali aventi energia:

$$\hbar\omega = E_{gap} - \frac{R_{ex}}{n^2} = E_n$$

cioè righe spettrali di energia inferiore a quella di gap. Tali assorbimenti si osservano sperimentalmente e non potrebbero essere spiegati senza il modello eccitonico. Questo spettro di assorbimento ottico al di sotto del gap $\alpha_{SUBg}(\hbar\omega)$ ($\alpha(\hbar\omega)$ "sub - gap") è discreto solo teoricamente: nella realtà le impurezze del cristallo e le vibrazioni fononiche degli atomi individuano livelli energetici "spuri" attorno a quelli E_n previsti dalla teoria eccitonica. Pertanto i livelli energetici discreti E_n , detti anche "risonanze eccitoniche discrete sub - gap", subiscono un allargamento ("broadening"), eventualmente modellizzabile mediante la sostituzione delle linee E_n con delle funzioni di distribuzione gaussiane.

Possiamo calcolare il rate $W_{ABS/sub-gap}$ con il quale un elettrone, inizialmente descritto da un "ground state" (stato di terra, stato fondamentale) ψ_0 , di energia E_0 , avente una funzione di involuppo χ_0 di simmetria p, localizzato presso il massimo della BV ($E_0 \rightarrow E_V$), può assorbire un fotone incidente di energia $\hbar\omega$ e "saltare" su uno stato eccitonico $\psi_{ex,n}$ "sub - gap", di energia E_n , indicizzato da un certo valore di n .

$$W_{ABS/sub-gap}(\psi_0 \rightarrow \psi_{ex,n}) = \frac{2\pi}{\hbar} \left(\frac{eA_0}{m_0}\right)^2 \left| \langle \chi_{ex,n} | \hat{a} \cdot \vec{p}(k_0) | \chi_0 \rangle \right|^2 \frac{\delta(E_{ex,n} + E_C - E_0 - \hbar\omega)}{\pi a_{ex}^3 n^3} \quad [e64]$$

L'espressione del modulo A_0 del potenziale vettore associato alla luce incidente dipende dalla radice quadrata del numero di fotoni n_{ph} , come mostrato nella [e26].

Nello scrivere la [e64] abbiamo fatto le seguenti ipotesi:

- l'elettrone si trova, mediamente, ad una distanza a_{ex} dalla lacuna pari a:

$$a_{ex} = a_B \epsilon_r \left(\frac{m_0}{m_r^*}\right) \approx 100 \text{ \AA}$$

dove a_B (≈ 52.9 pm) è il raggio classico di Bohr. L'eccitone in questione possiede una funzione di involuppo $\chi_r(x,y,z)$ abbastanza delocalizzata (lungo una ventina di celle fondamentali) e quindi è un eccitone di Mott. Essendo formato da più elettroni e più lacune, l'eccitone di Mott è descritto da una funzione $\psi_{ex}(x,y,z)$ più complicata di quella riportata nella [e63] (è descritto da una particolare sovrapposizione Σ_k di stati, di cui non ci occupiamo). Assumiamo per semplicità che uno stato eccitonico di Mott $\psi_{ex,n}(x,y,z)$ "sub - gap", indicizzato da n , abbia una funzione di involuppo $\chi_{ex,n}(x,y,z)$ di simmetria sferica. In formule:

$$\chi_{ex,n}(x,y,z) = \chi_{ex,n}(\mathbf{r}) = \frac{1}{\sqrt{\pi a_{ex}^3 n^3}} e^{-\frac{r}{a_{ex} n}}$$

- gli elettroni e le lacune appartenenti all'eccitone di Mott in questione sono tutti localizzati su stati k nei pressi del vettore d'onda k_0 , ovvero in un intorno stretto del punto di simmetria Γ
- ciascun elettrone di conduzione, appartenente all'eccitone di Mott, è posizionato nell'origine degli assi cartesiani: pertanto $r = 0$. Questa semplificazione analitica comporta l'esigenza di considerare stati eccitonici $\psi_{ex,n}(x,y,z)$ aventi funzioni di involuppo $\chi_{ex,n}(x,y,z)$ di simmetria sferica: infatti le funzioni $\chi_{ex,n}(x,y,z)$ di tipo s sono le uniche a presentare una densità di probabilità non nulla presso il loro centro, ovvero presso $r = 0$, cioè le uniche a poter accogliere un elettrone fotoeccitato proveniente dallo stato fondamentale ψ_0 (le funzioni $\chi_{ex,n}(x,y,z)$ di tipo p sono nulle in $r = 0$ e perciò non consentirebbero l'assorbimento ottico).

Se dividiamo il rate $W_{ABS/sub-gap}$, mostrato nella [e64], per il numero n_{ph} di fotoni incidenti e per la velocità v_{ph} di propagazione della luce nel semiconduttore, otteniamo il coefficiente di assorbimento ottico specifico "sub - gap" $\alpha_{SUBg}(\hbar\omega)$, che coincide con il coefficiente di assorbimento ottico totale specifico $\alpha_{TOT}(\hbar\omega < E_{gap})$. In formule:

$$\alpha_{TOT}(\hbar\omega < E_{gap}) = \alpha_{SUBg}(\hbar\omega) = \frac{W_{ABS/sub-gap}(\psi_0 \rightarrow \psi_{ex,n})}{n_{ph} v_{np}} \propto \frac{1}{n^3} \quad [e65]$$

L'altezza delle linee spettrali "sub - gap" posizionate presso i livelli energetici E_n , ovvero $\alpha_{SUBg}(\hbar\omega)$, come si può evincere partendo dalla [e64], decresce con andamento $1/n^3$, pertanto la probabilità che un elettrone atomico di valenza scatteri con un fotone incidente di energia $\hbar\omega = E_{gap} - R_{ex}/n^2 = E_C + E_{ex,n} - E_0$ e compia la transizione verticale $\psi_0 \rightarrow \psi_{ex,n}$ diminuisce all'aumentare di n , ovvero all'aumentare di $\hbar\omega$. Nei pressi di $\hbar\omega = E_{gap}$ l'assorbimento dovuto alla densità di stati tridimensionale congiunta $D_{joint}(\hbar\omega - E_{gap})$, ovvero l'assorbimento banda - banda, è ancora molto debole, pertanto gli effetti eccitonici continuano a rivestire un ruolo dominante nell'assorbimento ottico specifico: infatti $D_{joint}(\hbar\omega - E_{gap})$ è di fatto nulla, per cui anche il coefficiente di assorbimento ottico $\alpha_0(\hbar\omega)$ privo di effetti eccitonici lo è, mentre i livelli E_n si infittiscono quadraticamente. A cavallo di E_{gap} la densità dei livelli discreti E_n è talmente alta, quasi a formare un "continuum", che diventa ragionevole parlare di "densità tridimensionale degli stati eccitonici" $D_{ex}(E)$, contrariamente alla situazione "sub - gap", nella quale i livelli E_n erano troppo distanti gli uni dagli altri per potervi associare una densità di stati approssimativamente continua. $D_{ex}(E)$ è definita nel modo seguente:

$$D_{ex}(E) = 2 \frac{\partial n}{\partial E} = \frac{n^3}{R_{ex}}$$

dove il 2 è legato alla degenerazione di spin ed n è il numero quantico che indicizza gli autovalori E_n dell'energia (n è il numero quantico principale, utilizzando la nomenclatura tipica del problema dell'atomo idrogenoide). Possiamo pertanto calcolare il rate $W_{ABS/gap}$ con il quale un elettrone, inizialmente descritto da un "ground state" ψ_0 , di energia E_0 , avente una funzione di involuppo χ_0 di simmetria p, localizzato presso il massimo della BV ($E_0 \rightarrow E_V$), può assorbire un fotone incidente di energia $\hbar\omega \approx E_{gap}$ e "saltare" su uno stato eccitonico $\psi_{ex,n}$ localizzato nei pressi del minimo di BC, ovvero di energia $E_{ex,n} \approx 0$, indicizzato cioè da un valore piuttosto alto di n . Il rate $W_{ABS/gap}$ differisce dal rate $W_{ABS/sub-gap}$ per un termine di sommatoria $\sum_{|f_{ex}\rangle}$ sugli stati finali eccitonici: infatti la probabilità di transizione elettronica $\psi_0 \rightarrow \psi_{ex,n}$ decresce all'aumentare di n , secondo $1/n^3$, per $\hbar\omega < E_{gap}$, ma arrivati ad $\hbar\omega \approx E_{gap}$ la probabilità di transizione risente anche dell'incremento, secondo n^3 , della densità di stati eccitonici $D_{ex}(E)$. In formule:

$$W_{ABS/gap}(\psi_0 \rightarrow \psi_{ex,n}) = \frac{2\pi}{\hbar} \left(\frac{eA_0}{m_0}\right)^2 \sum_{|f_{ex}\rangle} |\langle \chi_{ex,n} | \hat{a} \cdot \vec{p}(k_0) | \chi_0 \rangle|^2 \frac{\delta(E_{ex,n} + E_C - E_0 - \hbar\omega)}{\pi a_{ex}^3 n^3} \quad [e66]$$

dove $E_{ex,n} + E_C - E_0 \approx E_{gap}$. Sostituendo la sommatoria sugli stati finali eccitonici $|f_{ex}\rangle$ con la densità di stati eccitonici $D_{ex}(E)$ otteniamo:

$$W_{\text{ABS/gap}}(\psi_0 \rightarrow \psi_{ex,n}) = \frac{2\pi}{\hbar} \left(\frac{eA_0}{m_0} \right)^2 \left| \langle \chi_{ex,n} | \hat{a} \cdot \vec{p}(\mathbf{k}_0) | \chi_0 \rangle \right|^2 \frac{1}{R_{ex}} \frac{\delta(E_{ex,n} + E_C - E_0 - \hbar\omega)}{\pi a_{ex}^3} \quad [\text{e67}]$$

Pertanto il coefficiente di assorbimento ottico specifico “a cavallo di E_{gap} ” $\alpha_g(\hbar\omega)$ non dipende da n , ovvero il suo spettro assume un andamento costante $\forall \hbar\omega$ appartenente ad un intorno ristretto di E_{gap} . In formule:

$$\alpha_{\text{TOT}}(\hbar\omega \approx E_{\text{gap}}) = \alpha_g(\hbar\omega) = \frac{W_{\text{ABS/gap}}(\psi_0 \rightarrow \psi_{ex,n})}{n_{ph} v_{np}} = \alpha_0(\hbar\omega) \left\{ \frac{2\pi\sqrt{R_{ex}}}{\sqrt{\hbar\omega - E_{\text{gap}}}} \right\} \quad [\text{e68}]$$

dove $\alpha_0(\hbar\omega)$, la cui espressione, mostrata nella [e58], dipende da $\sqrt{\hbar\omega - E_{\text{gap}}}$, è il coefficiente di assorbimento ottico specifico privo degli effetti eccitonici. Il termine compreso tra parentesi graffe è strettamente maggiore di 1. Per $\hbar\omega$ leggermente maggiore di E_{gap} ($\hbar\omega \gtrsim E_{\text{gap}}$) abbiamo un coefficiente di assorbimento ottico specifico di “over – gap” $\alpha_{\text{OVERg}}(\hbar\omega)$, definito nel modo seguente:

$$\alpha_{\text{TOT}}(\hbar\omega \gtrsim E_{\text{gap}}) = \alpha_{\text{OVERg}}(\hbar\omega) = \alpha_0(\hbar\omega) \left\{ \frac{\pi\zeta e^{\pi\zeta}}{\sinh(\pi\zeta)} \right\} \quad [\text{e69}]$$

con:

$$\zeta = \frac{R_{ex}}{\sqrt{\hbar\omega - E_{\text{gap}}}}$$

Maggiore è l'energia fotonica $\hbar\omega$ rispetto ad E_{gap} , minore è ζ , minore è il termine tra parentesi graffe, il quale risulta modificato rispetto a quello della [e68], e più $\alpha_{\text{OVERg}}(\hbar\omega)$ decresce, avvicinandosi ad $\alpha_0(\hbar\omega)$. Per $\hbar\omega \gg E_{\text{gap}}$, cioè per $\zeta \rightarrow 0$, il termine tra parentesi graffe decresce esponenzialmente fino a raggiungere il valore unitario: così $\alpha_{\text{TOT}}(\hbar\omega \gg E_{\text{gap}}) \rightarrow \alpha_0(\hbar\omega)$ dall'alto. Da un punto di vista prettamente matematico 1 è il valore asintotico del termine tra parentesi graffe, tuttavia nella realtà è sufficiente che $\hbar\omega$ superi E_{gap} di pochi meV, affinché $\alpha_{\text{TOT}}(\hbar\omega > E_{\text{gap}}) \rightarrow \alpha_0(\hbar\omega)$. In formule:

$$\alpha_{\text{TOT}}(\hbar\omega > E_{\text{gap}}) \approx \alpha_0(\hbar\omega) \quad [\text{e70}]$$

In figura 39 riportiamo lo spettro semiquantitativo del coefficiente di assorbimento ottico totale specifico $\alpha_{\text{TOT}}(\hbar\omega)$, per un generico semiconduttore (a gap diretto o indiretto) termostato ad una certa temperatura T fissata. Sul range delle energie fotoniche $\hbar\omega$ sono messe in evidenza, mediante le enumerazioni adottate nel testo, le quattro regioni spettrali nelle quali è possibile ottenere un assorbimento ottico dal semiconduttore con cui è costruito il fotorivelatore, quando questo è irradiato da una luce avente pulsazione ω . Si possono notare le righe spettrali “sub – gap”, spaziate quadraticamente e di altezza decrescente come $1/n^3$, la zona piatta intorno a E_{gap} e le due regioni “over – gap” nelle quali $\alpha_{\text{TOT}}(\hbar\omega)$ converge verso $\alpha_0(\hbar\omega)$; quest'ultimo, come già detto, rappresenta il coefficiente di assorbimento ottico specifico nel caso di assenza di eccitoni.

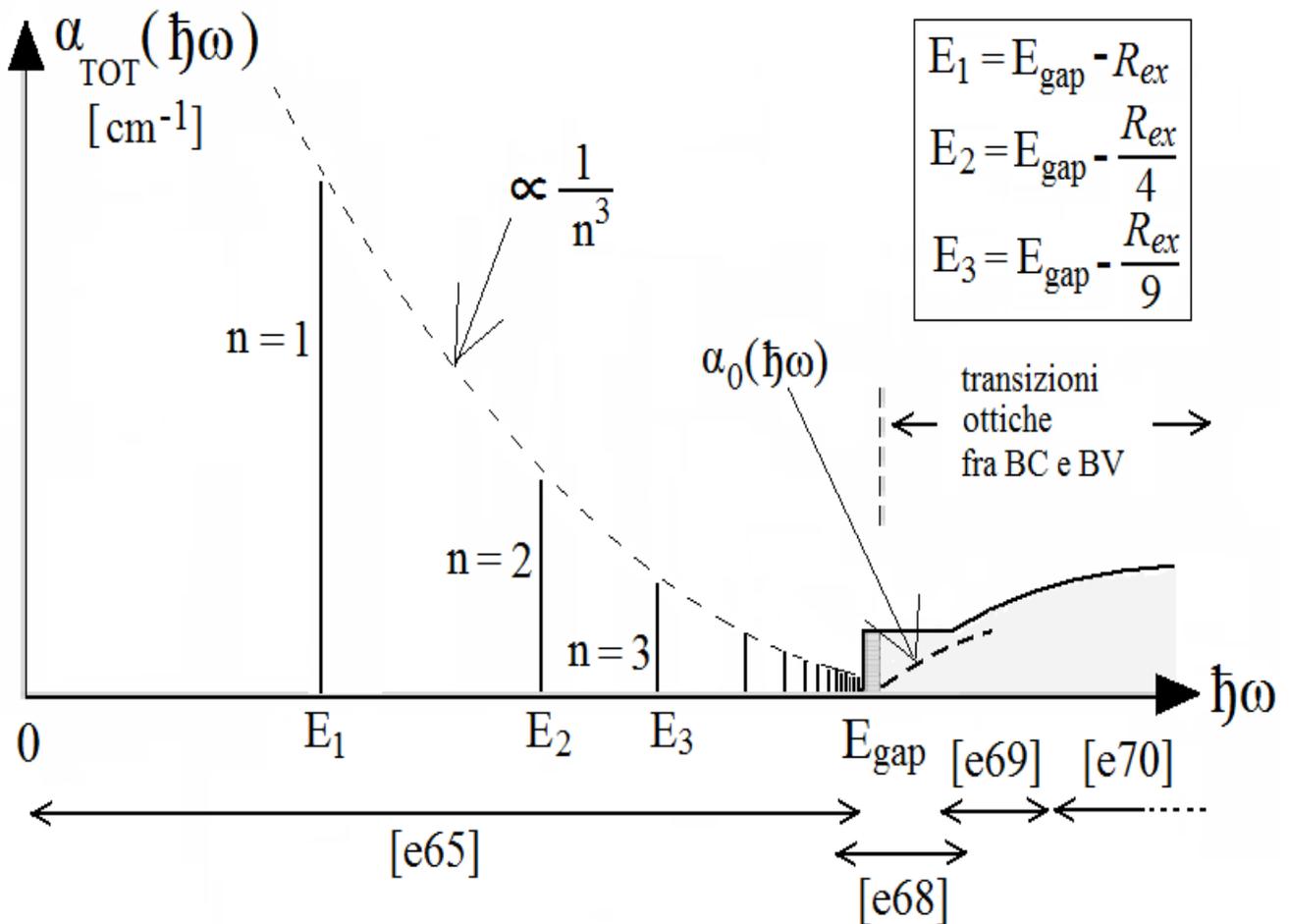


Fig. 39

Rappresentazione qualitativa dello spettro del coefficiente di assorbimento ottico totale specifico $\alpha_{TOT}(\hbar\omega)$, per un generico semiconduttore (a gap diretto o indiretto) termostato ad una certa temperatura T fissata. Sul range delle energie fotoniche $\hbar\omega$ sono messe in evidenza, mediante le enumerazioni adottate nel testo, le quattro regioni spettrali nelle quali è possibile ottenere un assorbimento ottico dal semiconduttore con cui è costruito il fotorivelatore, quando questo è irradiato da una luce avente pulsazione ω . Si possono notare le righe spettrali “sub - gap”, spaziate quadraticamente e di altezza decrescente come $1/n^3$, la zona piatta intorno a E_{gap} e le due regioni “over - gap” nelle quali $\alpha_{TOT}(\hbar\omega)$ converge verso $\alpha_0(\hbar\omega)$; quest’ultimo, come già detto, rappresenta il coefficiente di assorbimento ottico specifico nel caso di assenza di eccitoni.

È opportuno ribadire il carattere qualitativo degli spettri $\alpha_0(\hbar\omega)$, riportato in figura 38, e $\alpha_{TOT}(\hbar\omega)$, riportato in figura 39. Nel paragrafo 5.7, in figura 53, riporteremo lo spettro $\alpha_0(\lambda)$, ottenuto sperimentalmente, per alcuni dei semiconduttori (Si, Ge, GaAs, InP ecc...) maggiormente utilizzati per la costruzione dei fotorivelatori.

L’effetto eccitonico sull’assorbimento ottico di un semiconduttore dipende, in modo non trascurabile, anche dalla temperatura T alla quale il cristallo semiconduttore, con cui è realizzato il fotorivelatore, è termostato. È possibile rendersene conto osservando il grafico riportato in figura 40. Possiamo notare come le caratteristiche più importanti del coefficiente di assorbimento ottico totale specifico $\alpha_{TOT}(\hbar\omega)$, ovvero la soglia dell’energia di gap E_{gap} e la forma di un suo intorno ristretto, varino al variare della temperatura. Nella trattazione teorica appena conclusa abbiamo dimostrato come nell’intorno di E_{gap} il coefficiente $\alpha_{TOT}(\hbar\omega)$ fosse indipendente dal numero quantico n , ovvero piatto. Nella realtà generalmente ciò non accade: l’intorno di E_{gap} presenta un picco che emerge dallo spettro “non eccitonico” $\alpha_0(\hbar\omega)$. Il semiconduttore utilizzato per effettuare le misure di $\alpha_{TOT}(\hbar\omega)$, a diverse temperature T , è il GaAs. Si noti come all’aumentare di T si ha un decremento sia della soglia dell’energia E_{gap} che dell’ampiezza del picco eccitonico. L’effetto eccitonico sull’assorbimento non è molto visibile, nel caso del GaAs, se il fotorivelatore è termostato a temperatura ambiente. Nel grafico sperimentale, mostrato in figura 40, le risonanze “sub - gap” non sono state riportate.

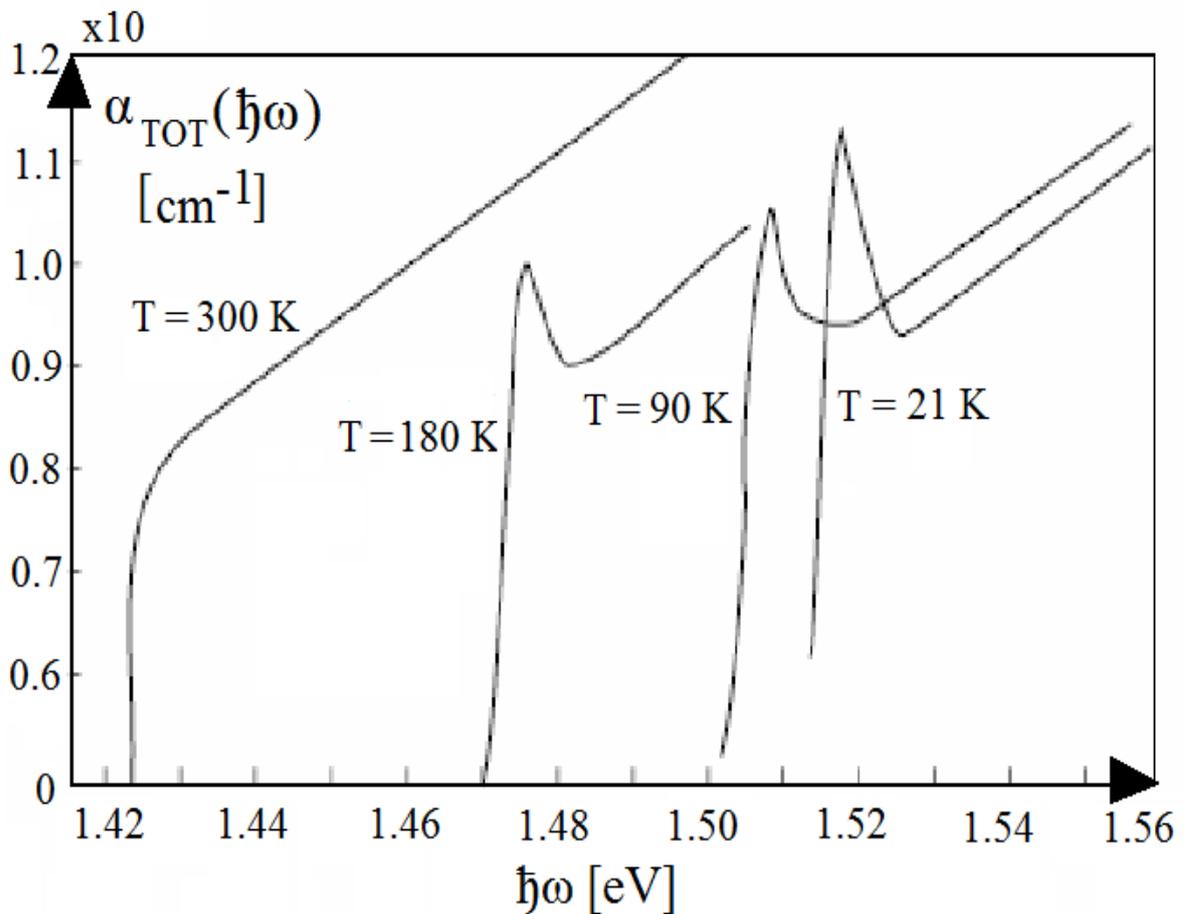


Fig. 40

Grafico sperimentale della dipendenza dalla temperatura T del coefficiente di assorbimento ottico specifico totale $\alpha_{TOT}(\hbar\omega)$ del cristallo di GaAs, con il quale può essere realizzato un fotorelevatore. Possiamo notare come le caratteristiche più importanti di $\alpha_{TOT}(\hbar\omega)$, ovvero la soglia dell'energia di gap E_{gap} e la forma di un suo intorno ristretto, varino al variare della temperatura. La soglia dell'energia di gap costituisce il discriminante fra un assorbimento ottico trascurabile (le righe spettrali E_n "sub - gap" non sono state rilevate nell'esperimento) ed un assorbimento ottico non trascurabile. Nella trattazione teorica abbiamo dimostrato come nell'intorno di E_{gap} il coefficiente $\alpha_{TOT}(\hbar\omega)$ fosse indipendente dal numero quantico n , ovvero piatto. Nella realtà ciò non accade: l'intorno di E_{gap} presenta un picco che emerge dallo spettro "non eccitonico" $\alpha_0(\hbar\omega)$. Si noti come all'aumentare di T si ha un decremento sia della soglia di E_{gap} che dell'ampiezza del picco. L'effetto eccitonico sull'assorbimento non è molto visibile, nel caso del GaAs, se il fotorelevatore è termostato a temperatura ambiente.

Gli effetti eccitonici sull'assorbimento ottico non sono molto tangibili nel caso di semiconduttori a gap piccolo, come ad esempio il InAs ($E_{gap} \approx 0.34$ eV a $T = 300$ K, $E_{gap} \approx 0.39$ eV a $T = 0$ K), anche se manteniamo il cristallo a temperatura T abbastanza bassa. Infatti abbiamo spiegato che la costante di Rydberg R_{ex} per l'eccitone è caratterizzata dalla seguente proporzionalità inversa:

$$R_{ex} \propto \frac{1}{\epsilon_r^2}$$

Pertanto se il gap è molto piccolo, cioè se ϵ_r è molto grande, abbiamo che R_{ex} è molto piccola, quindi gli effetti eccitonici sono concentrati tutti a ridosso della soglia di E_{gap} , da dove, aumentando $\hbar\omega$, cominciano a farsi sentire in modo non trascurabile, seppur gradualmente, le transizioni banda - banda (BC - BV). In altre parole le sottobande eccitoniche di conduzione risultano difficilmente distinguibili, dal punto di vista delle transizioni ottiche, dalla banda di conduzione BC originaria.

5.4) I problemi di efficienza quantica e di risoluzione temporale, di un fotorelevatore a giunzione pn, legati alla larghezza della zona di svuotamento

Consideriamo il diagramma a bande di figura 41, riferito ad una generica omogiunzione pn polarizzata in inversa, sottoposta ad illuminazione.

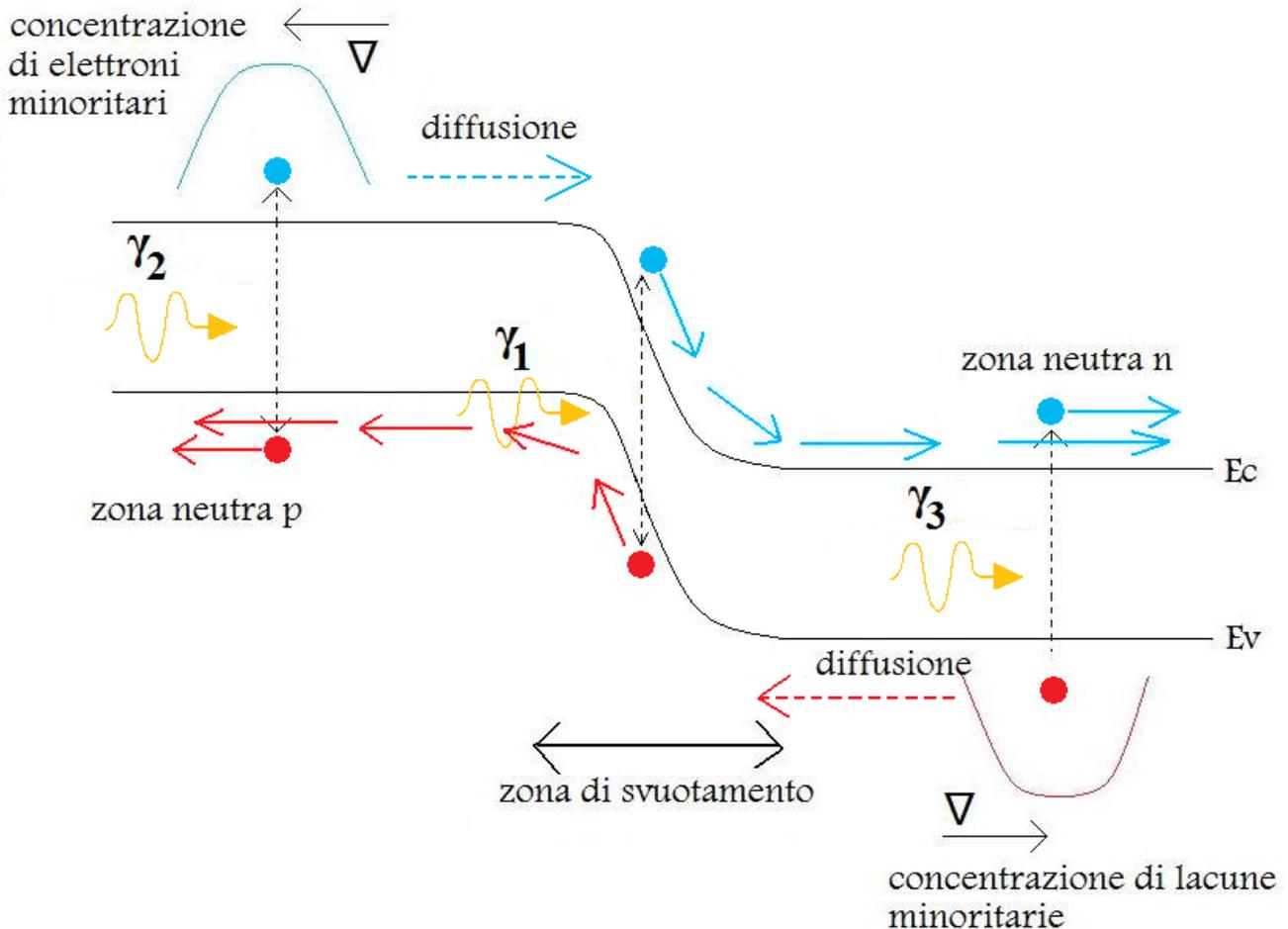


Fig. 41

Diagramma a bande di un'omogiunzione pn, polarizzata in inversa, su cui incide della luce. L'energia fotonica può essere assorbita in parte nella zona di svuotamento, in parte nelle zone neutre p ed n: questa ripartizione ha conseguenze sull'efficienza quantica e sulla risoluzione temporale del dispositivo.

Quando un pacchetto di fotoni entra nel fotodiodo, i fotoni possono essere assorbiti in differenti punti del dispositivo: in particolare alcuni saranno assorbiti nella zds^* , altri nelle zone neutre p ed n. Un fotone (γ_1) che viene assorbito nella zds^* , ovvero nella regione soggetta a campo, produce una fotocoppia che, non appena generata, risente istantaneamente dell'azione trascinante del campo: il fotoelettrone vede uno "scivolo" di energia potenziale E_c che lo conduce verso la zona neutra n. Questo moto, permesso dall'andamento spaziale di E_c , cioè dal campo, costituisce la J_{ndrift} di fotoconduzione, che è rilevata ai terminali metallici non appena l'elettrone arriva in prossimità del contatto ohmico fra lato n ed alluminio. La lacuna fotogenerata nella zds^* vede lo scivolo E_v , per cui è trascinata verso il bulk p (questo moto produce la J_{pdrift} di fotoconduzione), dove la fotolacuna dà il suo contributo alla fotocorrente totale. Affinchè queste due componenti di corrente siano effettivamente rilevate dal circuito esterno è necessario che i rispettivi portatori non si ricombinino durante il tragitto. L'elettrone deve attraversare la parte della zds^* che lo separa dall'interfaccia $zds^*/$ zona neutra n, e la distanza da quest'ultima interfaccia fino al contatto metallico. Al fine di garantire all'elettrone un'alta probabilità di attraversare la prima parte del percorso, è necessario che il Si sia sufficientemente puro, ovvero povero di trappole metalliche (cobalto, ferro, nichel ecc...oro soprattutto, il quale è, per motivi di matching reticolare, altamente inquinante per il silicio). Queste, infatti, introducono dei "deep levels" energetici, all'interno del gap, che possono catturare portatori, facilitandone la ricombinazione. Nel Si il tempo di vita medio dei portatori è dato dalla seguente relazione:

$$\tau = \frac{1}{N_t \sigma_c v_{th}}$$

dove N_t è la concentrazione di impurezze metalliche interstiziali (se consideriamo l'oro, un valore tipico per il Si non trattato con processi di gettering può essere 10^{14} cm^{-3}), σ_c è la sezione di cattura di queste impurezze, v_{th} la velocità termica degli elettroni in BC (un valore tipico può essere $2 \cdot 10^5 \text{ ms}^{-1}$). Quest'ultima può essere stimata considerando che i fotoelettroni in BC sono, tutto sommato, abbastanza pochi ($T \approx 300 \text{ K}$), rispetto a quelli rimasti in BV (la distribuzione $2D_{BC}(E - E_c)f(E)$ è abbastanza più bassa di $2D_{BC}(E - E_c)$, per $E > E_c$), per cui questi elettroni possono essere considerati come particelle non interagenti fra loro, e con comportamento "classico": pertanto la distribuzione di Fermi - Dirac è, in quella zona del diagramma a bande, praticamente un'esponenziale decrescente, e quindi è approssimabile con la distribuzione di Boltzmann. Dunque:

$$\frac{3kT}{2} = \frac{m_e^* v_{th}^2}{2}$$

Il seguente criterio di progetto è atto a garantire ai fotoportatori, generati nella zds^* , un attraversamento di quest'ultima privo di catture indesiderate:

$$\tau \gg T_{\max_{\text{transito}}} = \frac{zds^*}{v_{tr}}$$

dove v_{tr} è la velocità di transito. Un valore tipico per il tempo di transito è $T_{\max_{\text{transito}}} = 100 \text{ ps}$: tanto più è piccolo, tanto più il fotodiodo è veloce nel rispondere ad impulsi luminosi. Per garantire all'elettrone un'alta probabilità di attraversare, indenne, anche la seconda parte di tragitto, è opportuno che il tempo di vita medio dei minoritari τ_{eff} (le lacune, in questo caso) sia molto maggiore del tempo di transito dei maggioritari (gli elettroni) attraverso l'intero percorso. τ_{eff} è un tempo di vita che tiene conto della ricombinazione dei minoritari nel bulk, all'interfaccia fra bulk e zds^* e a quella fra bulk e contatto metallico (dunque una media pesata fra tre tempi di vita media diversi). Il guadagno fotoconduttivo di un rivelatore può essere espresso in funzione di τ_{eff} , tramite la formula:

$$G = \frac{\tau_{\text{eff}}(v_e - v_h)}{l}$$

in cui compaiono le velocità dei portatori e la distanza fra l'interfaccia zds^*/bulk e quella $\text{bulk}/\text{metallo}$. Anche la banda del fotorivelatore è legata a τ_{eff} :

$$B = \frac{1}{2\pi\tau_{\text{eff}}}$$

Da questa formula si può comprendere il problema fondamentale di un fotodiodo costituito da una semplice giunzione pn. Se i minoritari (le lacune) si ricombinano lentamente all'interno del bulk n (τ_{eff} alto), allora pochissimi fotomaggioritari (fotoelettroni) transitanti verranno ricombinati, per cui la fotocorrente sarà alta (G alto). Tuttavia i minoritari avranno il tempo necessario per diffondere verso la zds^* , e sono proprio queste correnti di diffusione (contrarie ai gradienti delle concentrazioni dei minoritari nei due bulk, secondo la legge di Fick) a rallentare la risposta del rivelatore all'impulso di luce che lo irradia (B è piccola).

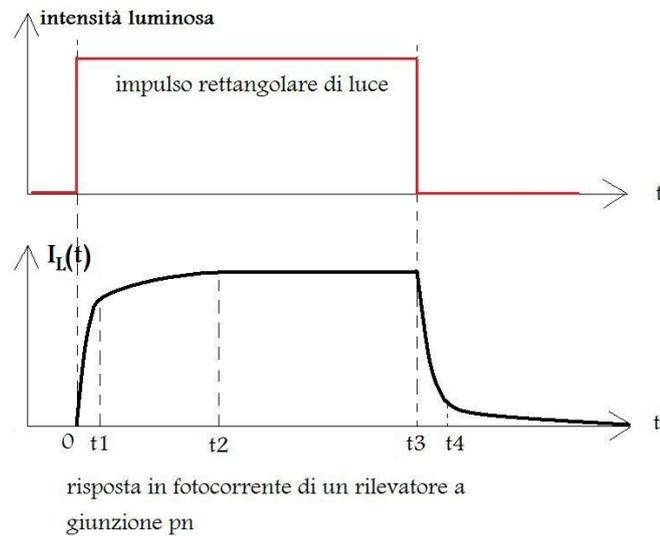


Fig. 42

Grafico temporale di un impulso rettangolare di luce, che potrebbe rappresentare la trasmissione di un bit di valore 1, incidente su un fotodiode a giunzione pn, e della conseguente risposta elettrica in corrente.

Riferiamoci alla figura 42. Supponiamo che sul fotodiode, all'istante 0, arrivi un segnale luminoso rettangolare: in questo istante alcuni fotoni verranno assorbiti nella zds^* (come γ_1 di figura 41), altri nelle zone neutre (come γ_2 e γ_3 , sempre di figura 41). I primi generano fotocoppie che vengono immediatamente trascinate dal campo e rapidamente raccolte nei rispettivi bulk di destinazione: tali coppie formano la fotocorrente che si instaura fra gli istanti 0 e t_1 . Tale corrente cresce rapidamente in virtù della rapidità del meccanismo di drift, tanto più rapidamente quanto minore è il tempo di transito attraverso la zds^* . All'istante t_1 la fotocorrente dovuta ai portatori generati nella zds^* è arrivata a regime. I secondi (γ_2 e γ_3) fotogenerano coppie, ciascuna delle quali contiene un maggioritario ed un minoritario: il maggioritario dà un contributo immediato alla fotocorrente (è nato già nel bulk di destinazione), e questo contributo si somma, fra 0 e t_1 , alla corrente dovuta ai fotoni come γ_1 . Il minoritario, se il suo tempo di vita medio lo permette, diffonde verso la zds^* (non essendoci campo nei bulk, la corrente è solo diffusiva). La diffusione è un processo di Arrhenius (attivato termicamente), per cui è molto lento (la diffusione su una distanza di $1 \mu m$ richiede un tempo di almeno 1 ns), specialmente per le lacune, essendo la loro mobilità circa tre o quattro volte inferiore a quella degli elettroni. All'interno dei bulk Si tipo n e Si tipo p, rispettivamente, la mobilità delle lacune e degli elettroni è:

$$\mu_h \approx 400 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1} \quad \mu_e \approx 1100 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$$

Tali valori diminuiscono ulteriormente quando i minoritari, diffondendosi, si avvicinano alle interfacce bulk/ zds^* . Fra gli istanti t_1 e t_2 queste due diffusioni iniziano a contribuire alla fotocorrente totale, dato che sempre più minoritari riescono ad arrivare alle interfacce sopracitate e ad oltrepassarle, arrivando nella zona di campo, dove vengono raccolti e trascinati verso i bulk di destinazione. All'istante t_2 anche i profili di diffusione sono a regime, per cui la corrente ha raggiunto il suo valor massimo. La lentezza con cui cresce la coda fra t_1 e t_2 è dovuta alla lentezza con la quale i portatori che diffondono vengono catturati dal campo e giungono al bulk di destinazione. Più la zds^* è piccola (a parità di radiazione luminosa), rispetto alle lunghezze dei due bulk, più la percentuale di fotoni incidenti assorbita nelle zone neutre sarà alta, più le correnti diffusive saranno importanti, più la coda fra t_1 e t_2 sarà lenta a raggiungere la condizione di regime e più il rivelatore sarà lento nel rispondere.

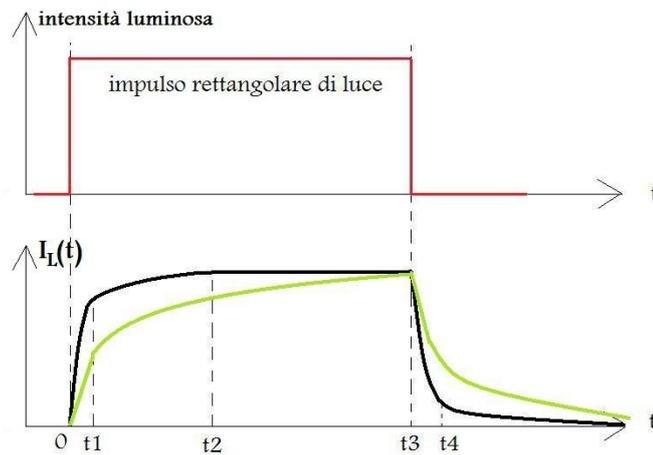


Fig. 43

Grafico temporale di un impulso rettangolare di luce, incidente su un fotodiodo a giunzione pn, e della conseguente risposta elettrica in corrente $I_L(t)$. La fotocorrente $I_L(t)$ disegnata di verde costituisce la risposta, a parità di potenza ottica incidente, di un fotodiodo la cui zona di svuotamento zds^* ha una larghezza inferiore a quella della zds^* del fotodiodo avente, come risposta elettrica, la $I_L(t)$ colorata di nero. Il primo dispositivo ha una risoluzione temporale peggiore di quella del secondo.

Cessata l'illuminazione all'istante t_3 , la zds^* smette di ospitare la fotogenerazione di coppie, il suo campo trascina a destinazione gli ultimi portatori rimasti, e all'istante t_4 il campo già non trasporta più fotoportatori nati nella zds^* , mentre continua a trascinare portatori fotogenerati nei bulk. Da t_4 in poi i "fotoeccessi" di concentrazione di minoritari, nei bulk, continuano a "sgonfiarsi" (lentamente, a causa della lentezza della diffusione verso la zds^*). La parte terminale di quella coda è imputabile all'ultimo cenno di diffusione delle lacune ($L_h < L_e$). Nel caso di zds^* eccessivamente piccole la maggior parte dei fotoni viene assorbita nelle zone neutre, pertanto i tempi di transito saranno dominati da quelli di diffusione; dunque il processo di raggiungimento della condizione di regime della fotocorrente totale sarà dominato da quello dei profili di diffusione, con la conseguenza riportata in figura 44:

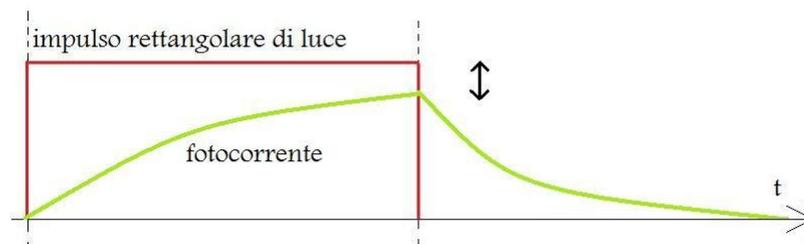


Fig. 44

Grafico temporale della risposta elettrica $I_L(t)$ (colorata di verde) di un fotodiodo, la cui zds^* ha una larghezza nettamente minore di quella delle regioni neutre p ed n. Una consistente frazione dell'energia fotonica incidente è assorbita nei bulk, con conseguente decremento della banda operativa del dispositivo, dovuto alla lentezza dei processi di diffusione termica dei fotoportatori.

L'attenuazione della fotocorrente massima è dovuta al fatto che, all'aumentare dei fotoportatori minoritari, aumenta la probabilità che questi si ricombinino prima di uscire dalle loro zone neutre, non riuscendo, dunque, a contribuire alla corrente (QE più piccola). L'andamento della corrente, sia nella parte di salita sia in quella di discesa, è regolato dalla costante di tempo $r_d C_j$ (grande), dove r_d è la resistenza equivalente del diodo, mentre C_j è la capacità della giunzione (importante in quanto la zds^* è piccola). Per evitare un simile inconveniente è necessario aumentare la larghezza della zds^* , facendo in modo che questa sia abbastanza maggiore delle lunghezze delle zone neutre p ed n. In tal modo solo una piccola frazione del pacchetto di fotoni incidenti verrà assorbita fuori dalla zds^* , pertanto le correnti di diffusione saranno trascurabili e non più limitanti per la banda: l'inerzia elettrica del fotodiodo sarà minore.

Un modo per determinare l'allargamento della zds^* , all'interno di un'omogiunzione pn, potrebbe essere quello di drogare leggermente i due lati p ed n: tuttavia le fotocorrenti di diffusione dei minoritari, all'interno dei due bulk, diverrebbero importanti, in quanto i minoritari subirebbero una ricombinazione moderata, per cui la banda

operativa del fotodiodo risulterebbe piccola. Per di più i bulk p ed n si rivelerebbero assai poco conduttivi. In vista delle applicazioni alle quali desideriamo destinare i fotorivelatori a semiconduttore, è chiaro che un dispositivo simile non potrebbe funzionare correttamente.



Fig. 45

Grafico della risposta elettrica $I_L(t)$ (colorata di verde) di un fotodiodo, la cui z_{ds}^* ha una larghezza nettamente maggiore di quella delle regioni neutre p ed n. L'energia fotonica è in gran parte assorbita nella z_{ds}^* , pertanto le correnti di diffusione sono trascurabili e non più limitanti per la banda, e la risoluzione temporale del fotodiodo è migliore.

5.5) L'andamento spaziale del campo elettrico e del potenziale all'interno di un fotorivelatore PIN asimmetrico

Un modo per allargare considerevolmente la z_{ds}^* è quello di aggiungere uno strato (i) di semiconduttore intrinseco, o leggermente drogato, fra i lati p ed n. La struttura così ottenuta prende il nome di "PIN"; vedremo nel paragrafo 5.11 come integrarla, a livello di passi di processo. Per disegnarne l'andamento spaziale del campo è utile pensare, inizialmente, la giunzione p/i come svincolata da quella i/n. Di seguito riportiamo i diagrammi a bande delle due giunzioni, entrambe con ciascun lato immaginato come non ancora interagente con l'altro (nessuno scambio di cariche).

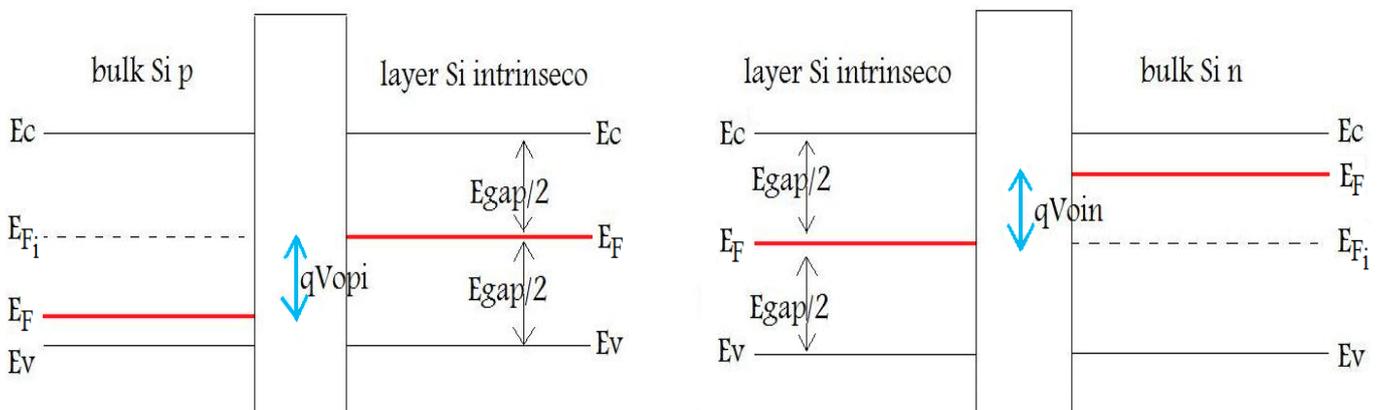


Fig. 46

A sinistra riportiamo il diagramma a bande della giunzione silicio p/silicio intrinseco, con ciascun lato pensato non ancora interagente con l'altro, ovvero prima dell'allineamento del livello di Fermi, come se i due bulk fossero fisicamente separati da una regione di spazio vuoto sufficientemente spessa (la probabilità con la quale le cariche passano da un lato ad un altro, per effetto tunnel, è trascurabile). V_{0pi} è il potenziale di contatto della giunzione p/i. A destra riportiamo il diagramma a bande della giunzione silicio intrinseco/silicio n, con ciascun lato pensato non ancora interagente con l'altro. V_{0in} è il potenziale di contatto della giunzione i/n.

$$qV_{0pi} = kT \ln \left(\frac{N_a}{n_i} \right)$$

$$qV_{0in} = kT \ln \left(\frac{N_d}{n_i} \right)$$

Immaginando di unire il lato p con lo strato intrinseco e quest'ultimo con il lato n, otteniamo il diagramma mostrato in figura 47:

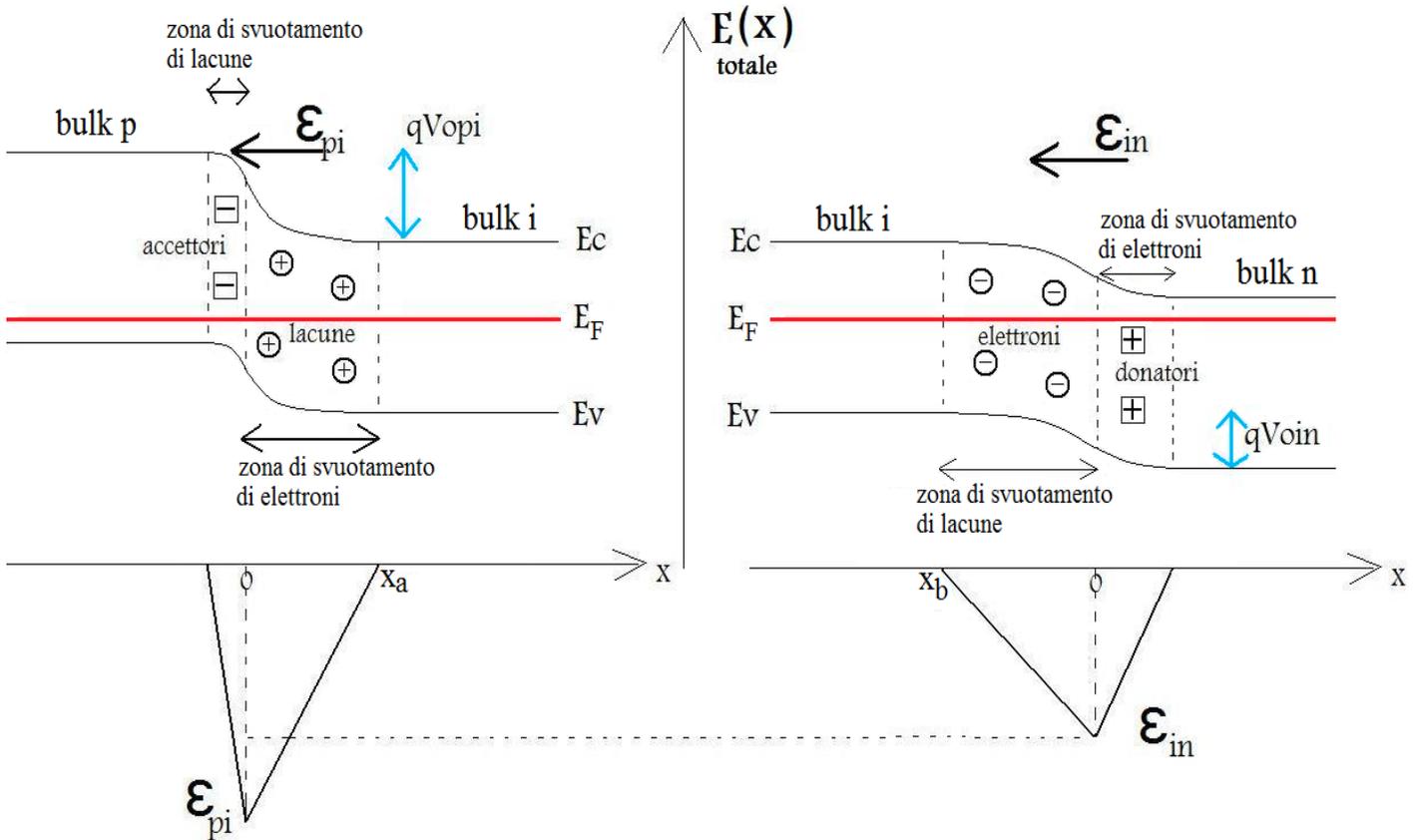


Fig. 47

A sinistra riportiamo il diagramma a bande della giunzione silicio p/silicio intrinseco in condizione di equilibrio, ovvero successivamente all'allineamento del livello di Fermi. A destra riportiamo il diagramma a bande della giunzione silicio intrinseco/silicio n, sempre all'equilibrio.

La pendenza di ϵ_{pi} in funzione di x , nella zona di svuotamento di lacune, è pari a $-qN_a/\epsilon_s$, quella dello stesso campo nella zona di svuotamento di elettroni è $+qN_i/\epsilon_s$, quella di ϵ_{in} nella zona di svuotamento di lacune $-qN_i/\epsilon_s$, quella dello stesso campo nella zona di svuotamento degli elettroni $+qN_d/\epsilon_s$. Se la luce irradia il dispositivo proprio sulla superficie dello strato p, quest'ultimo dovrebbe essere drogato più pesantemente del layer n: infatti più la concentrazione di accettori è elevata, più il layer p può essere sottile, e di conseguenza maggiore è la frazione della potenza ottica, incidente su p, che riesce a raggiungere lo strato intrinseco, e maggiore è l'efficienza quantica del PIN, come si vedrà fra breve. A causa di questa asimmetria fra i livelli di drogaggio dei due lati estrinseci, e in base alle pendenze citate e alle larghezze delle regioni di deplezione sul versante di p ed n (inversamente proporzionali alle concentrazioni dei drogaggi), è lecito attendersi un picco di ϵ_{pi} maggiore di quello di ϵ_{in} . L'effetto dello svuotamento di elettroni (pochi elettroni intrinseci, cioè termici) da parte delle lacune, provenienti dal bulk p, si estende per tutta la larghezza del layer intrinseco, fino all'interfaccia fra questo e il lato n. Pertanto immaginando di unire la giunzione p/i con quella i/n dobbiamo congiungere la coordinata x_a alla coordinata 0 della giunzione i/n. Per simmetria dobbiamo congiungere la coordinata x_b alla coordinata 0 della giunzione p/i. I risultati sono il campo $\epsilon(x)_{totale}$, che è tutto concentrato nella zds^* (avendo questa un'altissima resistenza), ed il potenziale $V_{pin}(x)$.

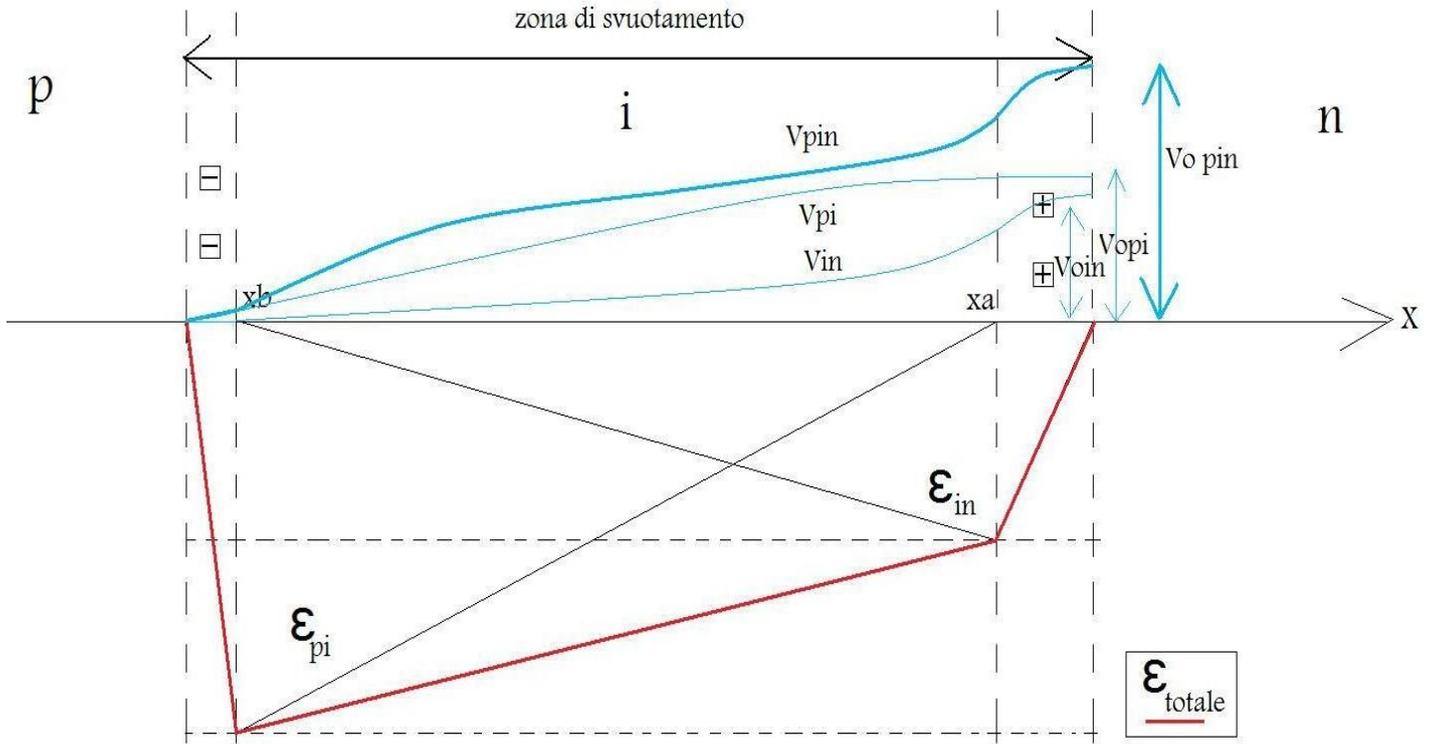


Fig. 48

Grafico spaziale del campo elettrico nella zona di svuotamento zds (sotto l'ipotesi di svuotamento totale), la quale comprende tutto lo strato intrinseco e le regioni, ad esso limitrofe, dei lati p ed n. Il campo $\varepsilon(x)_{\text{totale}}$ è dato dalla sovrapposizione dei contributi di campo $\varepsilon_{\text{pi}}(x)$ e $\varepsilon_{\text{in}}(x)$, il primo dovuto all'interazione elettrica fra silicio p e silicio intrinseco, il secondo dovuto all'interazione elettrica fra silicio intrinseco e silicio n. Il valore assoluto più elevato di $\varepsilon(x)_{\text{totale}}$ nei pressi dell'interfaccia p/i, rispetto a quello presso l'interfaccia i/n, è dovuto alla maggiore concentrazione di drogante nel lato p rispetto al lato n ($N_a > N_d$, la luce infatti entra dal lato p). Si noti anche la dipendenza spaziale del potenziale totale $V_{\text{pin}}(x)$: $V_{0\text{pin}}$ è il potenziale di contatto relativo all'intera struttura PIN.

La forma del campo elettrico all'interno dello strato intrinseco, il quale è interamente compreso all'interno della zds, è spiegabile anche nel modo seguente: essendo il lato p più pesantemente drogato di quello n, "l'efficacia svuotante" (ricombinante) delle lacune provenienti dal lato p, nei confronti degli elettroni intrinseci (termici) del layer i, è superiore all'efficacia svuotante degli elettroni, provenienti dal lato n, nei confronti delle lacune intrinseche dello stesso strato i, per cui all'interno di quest'ultimo si ha, all'equilibrio, una leggera prevalenza di lacune termiche rispetto agli elettroni. Quindi, riferendoci al diagramma a bande di figura 48, per qualunque coordinata x appartenente allo strato intrinseco, abbiamo che:

$$\text{concentrazione di carica totale nel layer intrinseco} = \rho_i(x) > 0$$

per cui, applicando l'equazione di Poisson, si ha che:

$$\frac{d^2}{dx^2} V_{\text{pin}}(x) = -\frac{\rho_i(x)}{\varepsilon_S} < 0$$

da cui si ricava che:

$$\frac{d\varepsilon_{\text{totale}}(x)}{dx} > 0$$

come era lecito aspettarsi dall'asimmetria fra i massimi dei campi $\varepsilon_{\text{pi}}(x)$ e $\varepsilon_{\text{in}}(x)$.

Applicando poi una tensione inversa, il cui modulo è in genere inferiore a 50 V, otteniamo la distribuzione di campo, mostrata in figura 49, per il PIN asimmetrico in polarizzazione inversa, la quale è condizione essenziale alla fotorivelazione:

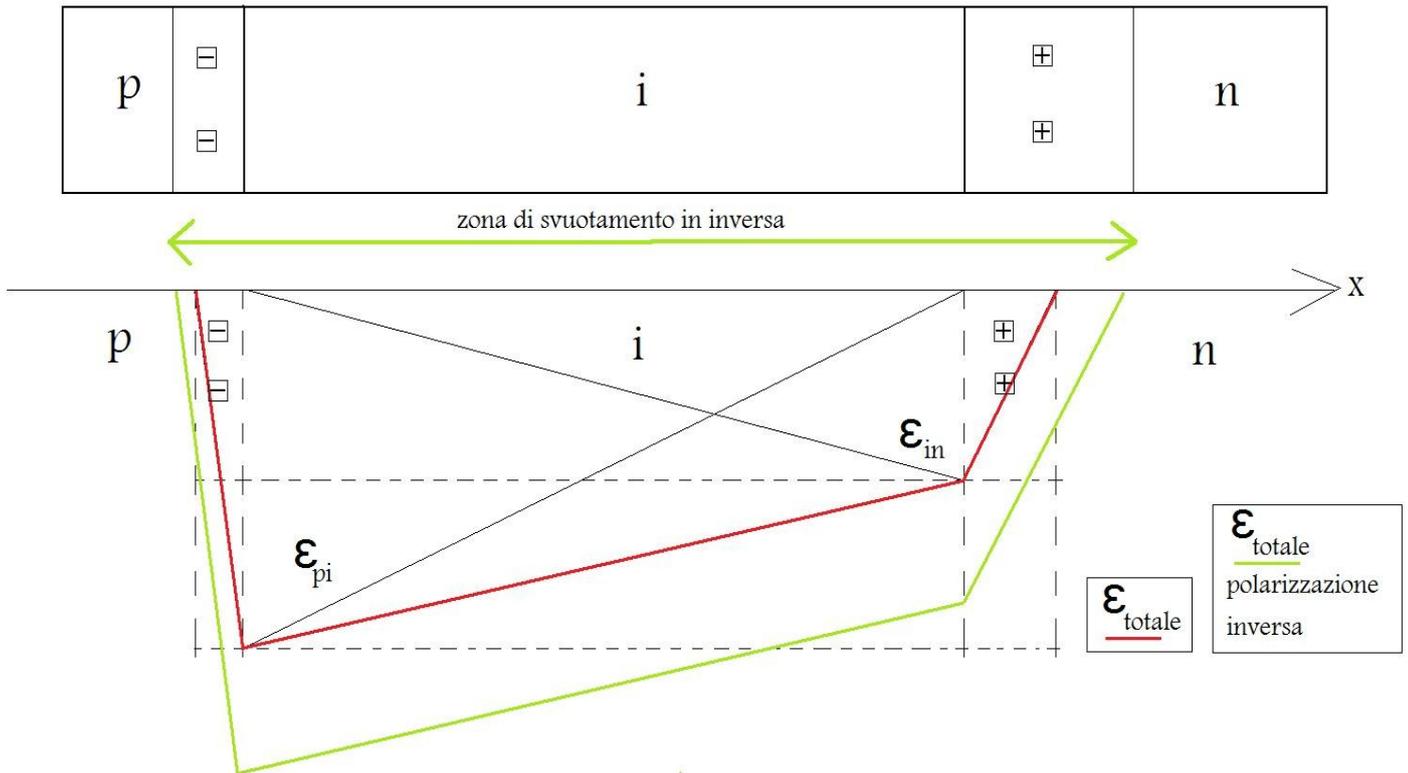


Fig. 49

Grafico spaziale del campo elettrico $\epsilon(x)_{totale}$ nella zona di svuotamento zds^* , nel caso di polarizzazione inversa. La larghezza della zds^* risulta incrementata, rispetto a quella in condizione di equilibrio, e lo stesso vale per il modulo del campo.

5.6) L'andamento spaziale del campo elettrico, del potenziale e delle bande di energia all'interno di un fotorivelatore PIN simmetrico

È possibile integrare un PIN nel quale il livello di drogaggio del lato p risulta uguale, approssimativamente, al livello di drogaggio del lato n (c'è simmetria fra la concentrazione degli accettori e quella dei donatori): in tal caso il rate della ricombinazione degli elettroni termici, all'interno dello strato intrinseco, causata dalle lacune del layer p, è uguale al rate della ricombinazione delle lacune termiche, sempre all'interno dello strato intrinseco, causata dagli elettroni del layer n. Quindi la concentrazione di carica totale, all'interno dello strato intrinseco, è nulla. Applicando l'equazione di Poisson, si ha che:

$$\frac{d^2}{dx^2} V_{pin}(x) = -\frac{\rho_i(x)}{\epsilon_S} = 0$$

da cui si ricava che:

$$\frac{d\epsilon_{totale}(x)}{dx} = 0$$

Il campo elettrico del PIN simmetrico, in condizioni di equilibrio, è costante nel layer i, che è interamente svuotato, mentre assume andamenti rettilinei (con le già citate pendenze) all'interno delle regioni di carica spaziale. Di conseguenza il potenziale $V_{pin}(x)$ presenta andamento rettilineo, come pure le bande di energia. Sia il potenziale che l'energia potenziale assumono le consuete forme paraboliche all'interno delle regioni di carica spaziale. Al fine di rendere molto sottili gli strati p ed n, nonché di aumentarne la conducibilità, può essere conveniente drogare pesantemente i due lati ($p^+/i/n^+$). Riportiamo in figura 50 il grafico spaziale delle bande di energia, del campo elettrico e del potenziale per un PIN simmetrico non polarizzato.

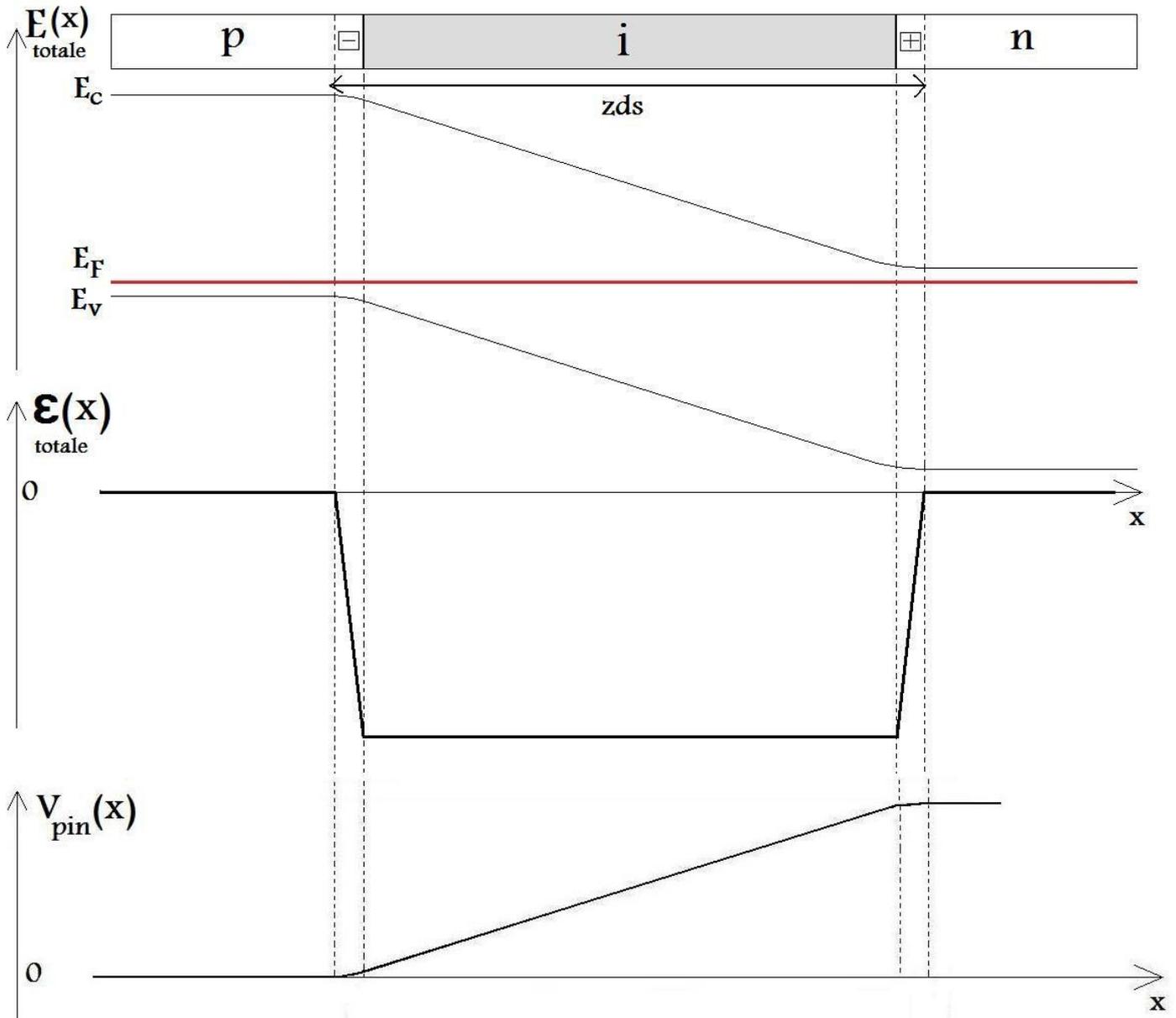


Fig. 50

Grafico spaziale delle bande di energia, del campo elettrico e del potenziale per un PIN simmetrico non polarizzato: si noti l'andamento costante del campo all'interno dello strato intrinseco, dovuto alla mancanza sia di carica fissa sia di carica mobile.

Applicando una tensione inversa al PIN simmetrico otteniamo il grafico spaziale delle bande di energia e del campo elettrico, mostrato in figura 51; si osservi come il layer intrinseco i sia particolarmente adatto all'assorbimento dei fotoni ed alla successiva accelerazione delle fotocariche primarie ($V_{\text{pin}}(x)$ è stato omesso, essendo questo semplicemente pari a $-E_c(x)/q$).

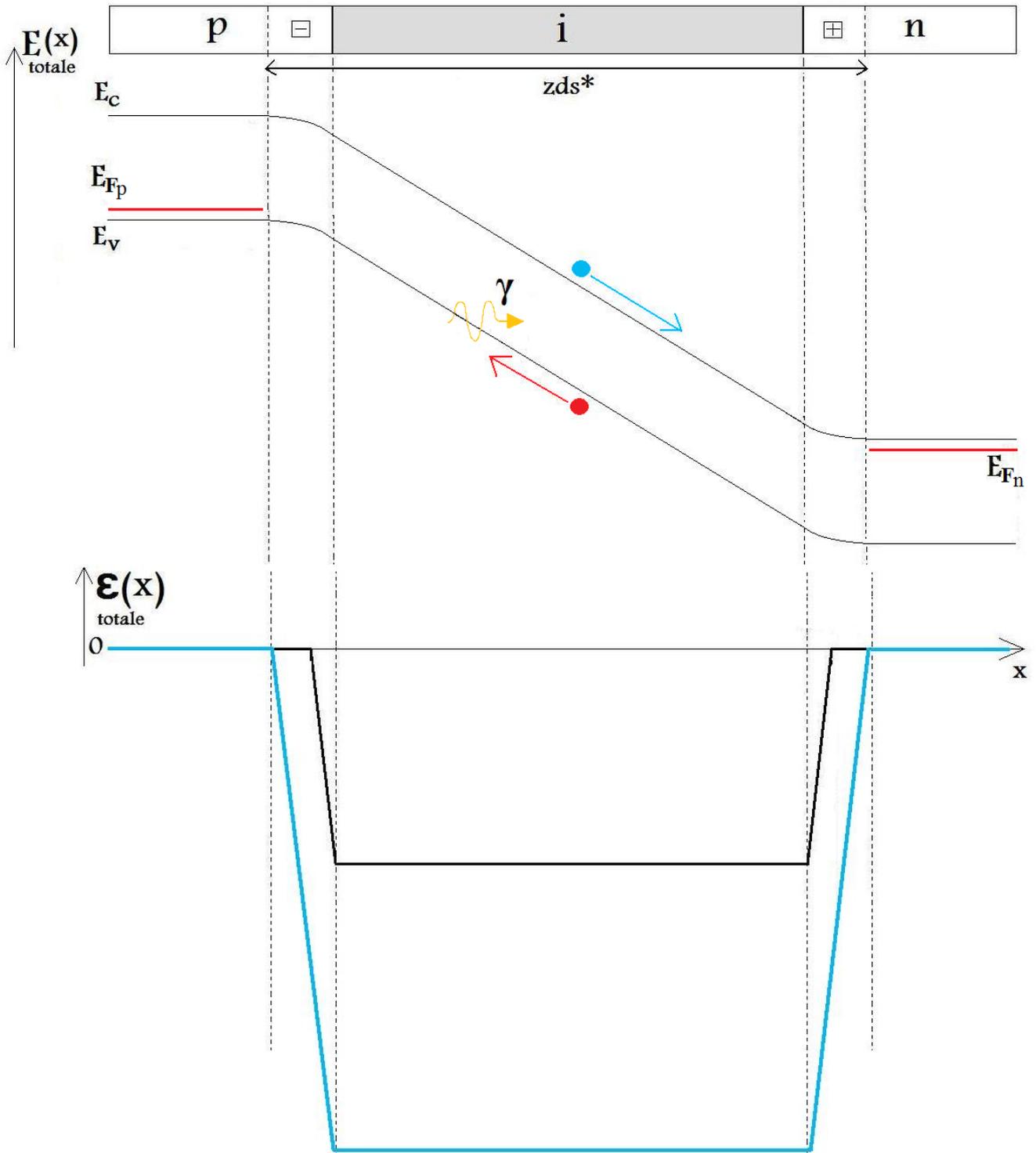


Fig. 51

Grafico spaziale delle bande di energia e del campo elettrico ($V_{pin}(x)$ è semplicemente pari a $-E_c(x)/q$) per un PIN simmetrico polarizzato inversamente: si noti l'andamento costante del campo (quello colorato di blu, che è notevolmente incrementato, rispetto alla situazione di equilibrio) all'interno dello strato intrinseco, dovuto alla mancanza sia di carica fissa che di carica mobile. Si noti come il layer intrinseco sia adatto all'assorbimento dei fotoni ed alla successiva accelerazione delle fotocariche primarie.

5.7) Potenza assorbita ed efficienza quantica per un PIN

Supponiamo, come mostrato in figura 52, di illuminare il lato p del PIN con una radiazione luminosa di frequenza opportuna ($\hbar\omega > E_{\text{gap}}$, come condizione generale), avente bassa intensità ottica I_0 (W/m^2), pertanto bassa potenza incidente P_0 (siamo in condizione di piccolo segnale).

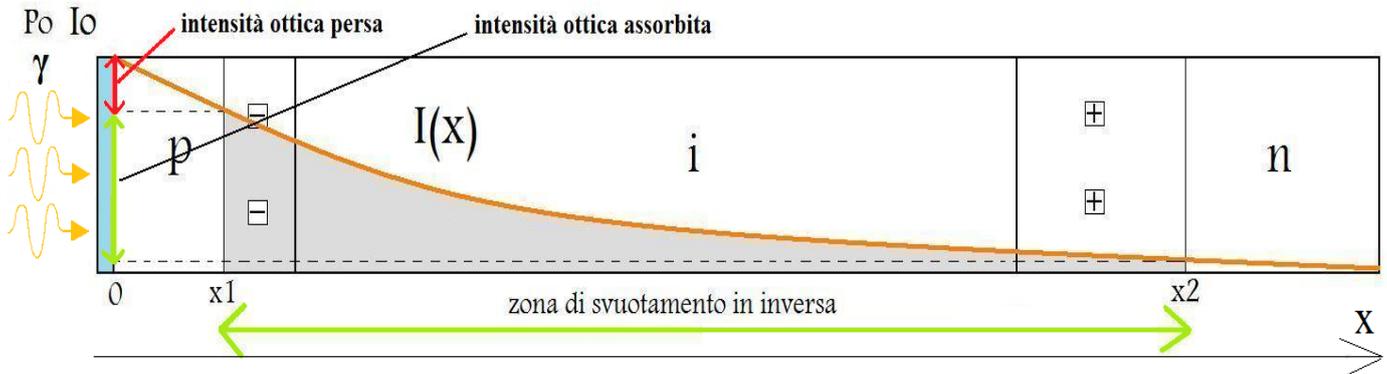


Fig. 52

Profilo dell'intensità luminosa $I(x)$ all'interno del PIN polarizzato inversamente (in condizione di piccolo segnale). L'andamento di $I(x)$, a forma di esponenziale decrescente, è causato dall'assorbimento dei fotoni da parte del semiconduttore. L'intensità ottica effettivamente assorbita dal PIN è pari a $I(x_1) - I(x_2)$, ovvero alla frazione dell'intensità ottica $I(0^+)$ che ha interagito con il materiale all'interno della zona di svuotamento.

Supponiamo, per il momento, che il dispositivo sia costruito con un semiconduttore a gap indiretto, come silicio o germanio (i primi PIN erano realizzati così). L'interfaccia aria/semiconduttore avrebbe una "riflettività" abbastanza alta, dovuta alla grande discontinuità dell'indice di rifrazione ($n_{\text{aria}} \approx 1$, mentre per una radiazione avente lunghezza d'onda $\lambda \approx 600$ nm si ha $n_{\text{Si}} \approx 3,4 \div 3,5$), nonché alla grande differenza fra l'impedenza caratteristica dell'aria (\approx impedenza caratteristica del vuoto $\zeta_0 = \sqrt{\mu_0/\epsilon_0} \approx 377\Omega$) e quella del semiconduttore. In particolare quest'ultima osservazione porta a concludere che i coefficienti di riflessione del campo elettrico, noti come "coefficienti di Fresnel", sono abbastanza alti. Pertanto una discreta frazione dell'energia luminosa non sarebbe trasmessa all'interno del PIN: a ciò si ovvia interponendo una finestra antiriflettente tale da attenuare le discontinuità sopra citate, caratterizzata da un indice di rifrazione intermedio ($n_{\text{vetro}}(\lambda \approx 600 \text{ nm}) \approx 1,5 \div 1,9$). Chiamiamo R la bassa riflettività dello strato antiriflesso. Dunque alla coordinata $x = 0^+$ abbiamo un'intensità trasmessa pari a $I_0(1 - R)$ ed una potenza $P_0(1 - R)$. Chiamiamo A la superficie attiva del PIN. L'intensità ottica nell'attraversare il materiale, che è prevalentemente assorbente, diminuisce in modo esponenziale, a causa dell'assorbimento dei fotoni. Dunque:

$$I(x) = I(0^+) e^{-\alpha x} = I_0(1 - R) e^{-\alpha x}$$

dove α è il coefficiente di assorbimento ottico specifico (cm^{-1}), nell'ipotesi di piccolo segnale, dipendente dal semiconduttore e dalla lunghezza d'onda della luce, discusso nei paragrafi 5.3.2, 5.3.3.1, 5.3.3.2 e 5.3.3.3. In figura 53 riportiamo $\alpha(\lambda)$ per i semiconduttori maggiormente utilizzati nella costruzione dei fotorivelatori.

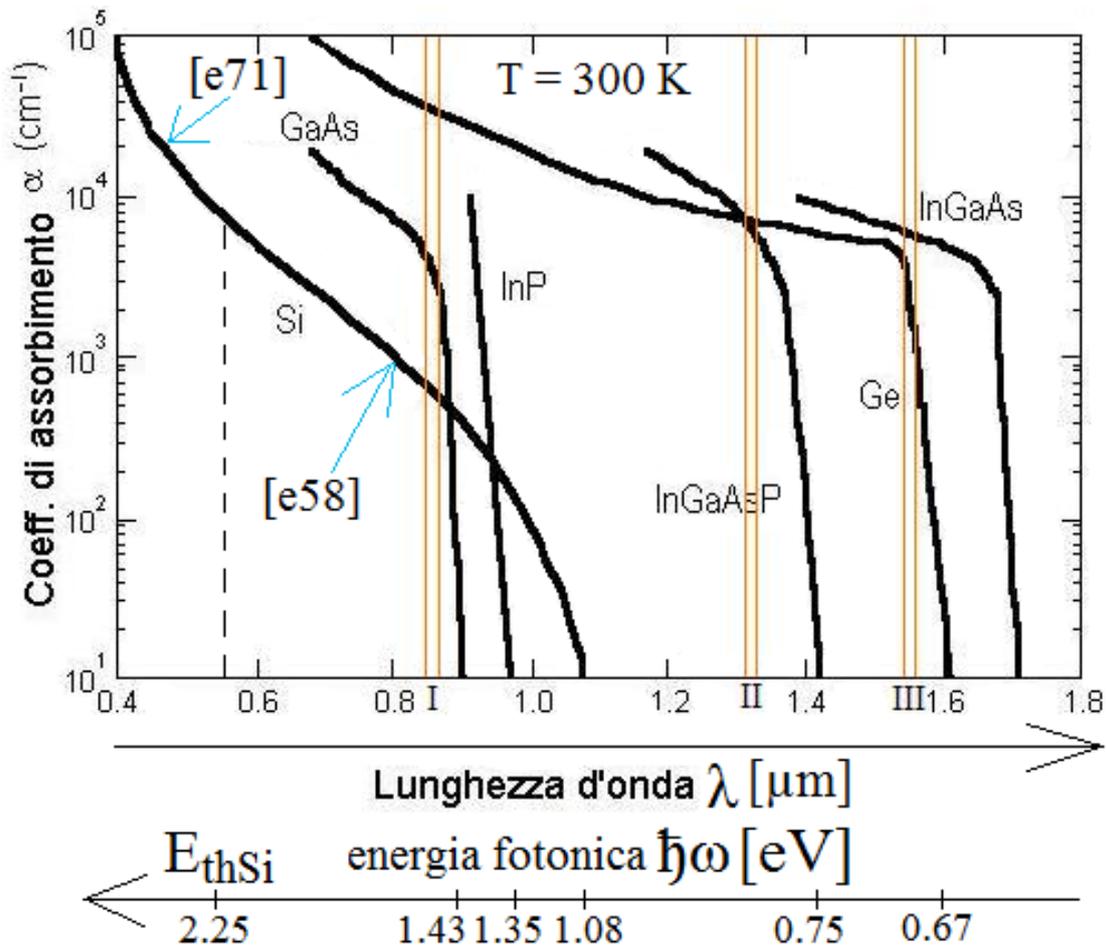


Fig. 53

Grafico, ottenuto sperimentalmente, del coefficiente di assorbimento ottico specifico in funzione della lunghezza d'onda della radiazione luminosa incidente, relativo ad alcuni semiconduttori, di tipo bulk, comunemente usati nella fabbricazione dei fotorelettori. Si notino le tre finestre di attenuazione (I, II, III) delle fibre ottiche costruite in silice SiO_2 , le cui frequenze centrali sono, rispettivamente, $0.85 \mu\text{m}$, $1.33 \mu\text{m}$, $1.55 \mu\text{m}$. Si noti, relativamente alla curva $\alpha(\lambda)$ del silicio, l'energia di soglia ("threshold") $E_{\text{thSi}} \approx 2.25 \text{ eV}$, la quale costituisce, per lo spettro del coefficiente di assorbimento ottico specifico $\alpha(\hbar\omega)$, l'energia fotonica incidente $\hbar\omega$ discriminante fra l'andamento a radice quadrata $\sqrt{\hbar\omega}$, riscontrabile per $\hbar\omega < E_{\text{thSi}}$, e l'andamento parabolico $(\hbar\omega)^2$, riscontrabile per $\hbar\omega > E_{\text{thSi}}$. Gli andamenti spettrali a radice quadrata e parabolico di $\alpha(\hbar\omega)$ sono collegati alle rispettive espressioni analitiche ([e58] ed [e71]) mediante la nomenclatura utilizzata nel testo.

La larga zds^* assicura che la maggior parte dei fotoni sarà assorbita nella zona di campo e solo un'esigua minoranza nelle zone neutre p ed n. Questo, come già detto, consente di considerare le correnti di diffusione non più limitanti per la velocità di risposta del dispositivo. Essendo tali correnti poco rilevanti (poiché dovute a pochi fotoni), assumiamo che i portatori fotogenerati nelle zone neutre non apportino un significativo contributo alla fotocorrente totale, cosicché la sua componente diffusiva, in linea di massima, non si fa sentire sul circuito esterno. Dunque la potenza assorbita nei bulk, ovvero per x compresa fra 0 e x_1 e per x maggiore di x_2 , è considerata potenza ottica persa: l'unica potenza utile, e quindi convertita in un segnale elettrico, è quella assorbita nella zds^* , cioè fra x_1 e x_2 . In formule:

$$I_{\text{assorbita}} = I(x_1) - I(x_2) = I_0(1 - R) e^{-\alpha x_1} (1 - e^{-\alpha(x_2 - x_1)}) = I_0(1 - R) e^{-\alpha x_1} (1 - e^{-\alpha zds^*})$$

$$P_{\text{assorbita}} = A (I(x_1) - I(x_2)) = P_0(1 - R) e^{-\alpha x_1} (1 - e^{-\alpha zds^*})$$

dove $P_0 = AI_0$

Nel PIN l'efficienza quantica QE, la cui definizione generale è riportata nel glossario, vale pertanto:

$$QE = \frac{\text{numero medio di coppie elettrone - lacuna fotogenerate e rilevate dal circuito esterno}}{\text{numero di fotoni incidenti}} = \frac{I_L}{\frac{P_0}{hf}} = \frac{P_{\text{assorbita}}}{P_0}$$

$$= (1 - R) e^{-\alpha x_1} (1 - e^{-\alpha z ds^*})$$

$$\text{con } I_L = \frac{q P_{\text{assorbita}}}{hf}$$

Rimandiamo l'analisi dettagliata del Rapporto Segnale - Rumore per un PIN ai paragrafi 6.3.2 e 6.6. Tuttavia possiamo già individuare una possibile causa intrinseca del rumore di fotorivelazione, all'interno di un PIN sottoposto ad illuminazione di opportuna frequenza ν , nel fatto che la QE a cui noi ci riferiamo è un "numero medio" e non esatto. Assumiamo, infatti, di illuminare un PIN con un impulso di luce avente una certa frequenza (un impulso monocromatico) ed una certa intensità ottica: per comodità di calcolo ipotizziamo che questa radiazione consti di 10 fotoni. Sui terminali metallici del PIN, all'atto dell'assorbimento della radiazione, osserviamo 8 fotocoppie elettrone/lacuna, le quali costituiscono la fotocorrente $I_L(t)$ di risposta del PIN: pertanto la QE in questa specifica fotorivelazione, in questo primo esperimento, è di circa l'80%. In altre parole la prima realizzazione della "variabile aleatoria" QE è pari a 0.8. Se ripetiamo l'esperimento nel modo più fedele possibile (stessa frequenza della luce, stessa intensità ottica, stesso PIN, stessa temperatura del PIN, ecc...) ci accorgiamo che il numero di fotocoppie raccolte dal campo e rilevate ai terminali sono: 7 nel secondo esperimento, 9 nel terzo, 6 nel quarto, 8 nel quinto, ecc... In altre parole la seconda realizzazione della QE è pari a 0.7, la terza realizzazione della QE è pari a 0.9, la quarta realizzazione è pari a 0.6, la quinta a 0.8 ecc... Pertanto QE, a rigore, è una variabile aleatoria discreta, a causa della natura granulare della carica elettrica, con un proprio valore di aspettazione, o valor medio, $E\{QE\}$, caratterizzata da una funzione di distribuzione $F_{QE}(QE)$, da una massa di probabilità $p_{QE}(QE)$ ecc... Possiamo quindi affermare che l'emissione di un fotoelettrone è un processo statistico, ovvero ha un'incertezza intrinseca che è alla base delle fluttuazioni aleatorie associate alla fotocorrente $I_L(t)$ in uscita da un PIN (o da un APD), in risposta ad un qualunque evento luminoso.

Osservando la figura 53 possiamo notare una differenza importante fra l'assorbimento ottico specifico di un semiconduttore a gap diretto, come ad esempio il GaAs, e l'assorbimento ottico specifico di un semiconduttore a gap indiretto, come ad esempio il Si. Per energie fotoniche incidenti $\hbar\omega$ leggermente superiori a quella di gap E_{gap} , il GaAs presenta uno spettro del coefficiente di assorbimento ottico specifico $\alpha(\hbar\omega)$ che cresce molto rapidamente all'aumentare di $\hbar\omega$, con una dipendenza molto approssimativamente riconducibile alla radice quadrata $\sqrt{\hbar\omega}$, imposta dalla densità tridimensionale congiunta degli stati $D_{\text{joint}}(\hbar\omega - E_{\text{gap}})$, tipica dei dispositivi di tipo bulk. Tale andamento, crescente con $\hbar\omega$, è reso particolarmente più ripido della normale dipendenza a radice quadrata dal fatto che il GaAs, così come il InP, il InGaAs ed il InGaAsP, è a gap diretto, pertanto le transizioni elettroniche $BV \rightarrow BC$ ($|i\rangle \rightarrow |f\rangle$), dovute allo scattering fra elettroni atomici e fotoni incidenti, sono dirette, pertanto non necessitano dell'intermediazione delle vibrazioni fononiche. Quindi la facilità (l'alta probabilità) con cui le energie fotoniche, leggermente superiori a E_{gap} , possono essere assorbite dagli elettroni atomici di valenza del semiconduttore a gap diretto comporta un aumento della ripidità di $\alpha(\hbar\omega) \propto \sqrt{\hbar\omega}$. Per energie fotoniche incidenti $\hbar\omega$ leggermente superiori a quella di gap E_{gap} , il Si presenta un coefficiente di assorbimento ottico specifico $\alpha(\hbar\omega)$ che cresce abbastanza lentamente all'aumentare di $\hbar\omega$, con una dipendenza a radice quadrata $\sqrt{\hbar\omega}$ imposta sempre dalla densità tridimensionale congiunta degli stati $D_{\text{joint}}(E - E_{\text{gap}})$. Tale andamento, crescente con $\hbar\omega$, è reso particolarmente meno ripido dello spettro $\alpha(\hbar\omega)$ dei semiconduttori a gap diretto, in quanto le transizioni elettroniche $BV \rightarrow BC$, dovute allo scattering fra elettroni di valenza e fotoni incidenti, sono indirette, ovvero intermedie da stati virtuali $|n\rangle$ ($|i\rangle \rightarrow |n\rangle \rightarrow |f\rangle$, come spiegato nel paragrafo 4.4.2), quindi necessitano dell'intermediazione di modi fononici favorevoli. Pertanto nel Si l'assorbimento ottico delle energie fotoniche $\hbar\omega$ leggermente superiori a E_{gap} è molto meno probabile che nei semiconduttori a gap diretto, con conseguente decremento della ripidità di $\alpha(\hbar\omega) \propto \sqrt{\hbar\omega}$. Quindi per energie ottiche $\hbar\omega < E_{\text{thSi}} \approx 2.25$ eV (la "silicon threshold energy" E_{thSi} , cioè "l'energia di soglia del silicio", appartiene al range ottico visibile ed è approssimativamente localizzata nei pressi della luce arancione) abbiamo uno spettro di assorbimento fotonico $\alpha(\hbar\omega)$ la cui espressione analitica è riportata nella [e58]. Per energie ottiche $\hbar\omega > E_{\text{thSi}}$ crescenti, come si può osservare in figura 54, la transizione ottica $BV \rightarrow BC$ ha sempre meno bisogno dell'intervento dei fononi, ovvero all'aumentare dell'energia fotonica incidente $\hbar\omega$ l'assorbimento ottico, che comporta il passaggio di un elettrone atomico del cristallo di silicio dal massimo della BV al minimo della BC, è sempre meno dipendente dalle transizioni fononiche parziali $|n\rangle \rightarrow |f\rangle$, ovvero è sempre più probabile. Infatti all'aumentare di $\hbar\omega$, partendo da E_{thSi} , la minima distanza "impulsiva" Δ_i , cioè la distanza, definita nello spazio dei vettori d'onda k , fra lo stato virtuale intermedio $|n\rangle$ e lo stato elettronico permesso $|f\rangle$ (appartenente alla BC) impulsivamente più vicino ad $|n\rangle$, diminuisce. Pertanto per energie fotoniche incidenti $\hbar\omega > E_{\text{thSi}}$ il coefficiente di assorbimento ottico specifico $\alpha(\hbar\omega)$ cresce parabolicamente all'aumentare di $\hbar\omega$. In formule:

$$\alpha(\hbar\omega > E_{\text{thSi}}) = \frac{M_\gamma^2 D_p^2 J_V (m_e^* m_h^*)^{3/2}}{8\pi^2 n_{ph} v_{ph} (E_{\text{gap}\Gamma} - \hbar\omega)^2 \hbar^6 \rho \omega_p} \eta_p(\omega_p) (\hbar\omega - E_{\text{thSi}})^2 \quad [\text{e71}]$$

dove:

- J_V è il numero di valli equivalenti all'interno del diagramma a bande $E - k$; nel silicio le valli X equivalenti sono 6, pertanto $J_V = 6$
- m_e^* è la massa efficace dell'elettrone per la mobilità, m_h^* è la massa efficace della lacuna per la mobilità, entrambe valutate nella valle in cui sono localizzate le coppie fotoelettrone/fotolacune, cioè in una delle valli X, nel silicio
- $E_{\text{gap}\Gamma}$ è l'energia del gap relativo al punto di simmetria Γ (dove $k = k_0 = (0,0,0)$), cioè il gap diretto; nel silicio $E_{\text{gap}\Gamma} \approx 3.4$ eV
- ρ [g/cm³] è la densità di massa del cristallo, che nel silicio è pari a circa 2.329 g/cm³
- $\omega_p = 2\pi/T_p$ è la pulsazione della vibrazione di un generico atomo del cristallo semiconduttore intorno alla propria posizione di equilibrio ed è proporzionale all'energia termica dell'atomo, ovvero alla temperatura del cristallo; fissato un punto (x,y,z) intorno alla posizione di equilibrio dell'atomo e un certo istante t_{xyz} nel quale l'atomo si trova localizzato nel punto (x,y,z) , T_p è l'intervallo di tempo medio tale per cui nell'istante $t_{xyz} + T_p$ l'atomo, a causa della vibrazione termica, ritorna a posizionarsi nel punto (x,y,z) . In altre parole ω_p rappresenta la pulsazione dei fononi che consentono una determinata transizione elettronica parziale $|n\rangle \rightarrow |f\rangle$, cioè la pulsazione fononica che "collega", nel silicio, gli stati elettronici localizzati nei pressi del punto Γ con quelli localizzati nei pressi del punto X.
- $\eta_p(\omega_p)$ è la probabilità di occupazione di stato dei fononi presso l'energia termica vibrazionale $\hbar\omega_p$, ovvero il valore della distribuzione statistica $\eta_p(E)$ dei fononi presso l'energia termica E legata ad un determinato valore di ω_p .
- D_p è il "potenziale di deformazione", ovvero un termine di potenziale elettrostatico proporzionale all'energia termica, cioè all'entità della vibrazione, all'energia dei fononi (alla pulsazione fononica ω_p), il quale deforma (perturba) la dipendenza spaziale dell'energia potenziale coulombiana $U(r)$ prodotta dagli atomi del cristallo; la deformazione D_p è causata dalla vibrazione termica degli atomi intorno ai rispettivi punti di equilibrio.
- M_γ è il "photon matrix element", ovvero l'elemento di matrice relativo ai fotoni γ incidenti sul semiconduttore del fotorivelatore, definito nel modo seguente:

$$M_\gamma^2 = \frac{e^2 \hbar n_{ph}}{2m_0^2 \epsilon_0 \epsilon_r \omega} \left| \langle u_{s+p} | \hat{a} \cdot \vec{p} | u_p \rangle \right|^2$$

dove riconosciamo l'elemento di matrice della quantità di moto dell'elettrone atomico, che è in procinto di scatterare con il fotone incidente. u_p è la "central cell function p - like" dello stato elettronico iniziale $|i\rangle$, posizionato sul massimo della BV, mentre u_{s+p} è la "central cell function" dello stato elettronico finale $|f\rangle$, posizionato sul minimo della BC, la cui simmetria ibrida $s + p$ è dovuta al gap indiretto.

- n_{ph} è il numero di fotoni incidenti sul semiconduttore, mentre v_{ph} è la velocità di propagazione della luce all'interno del cristallo

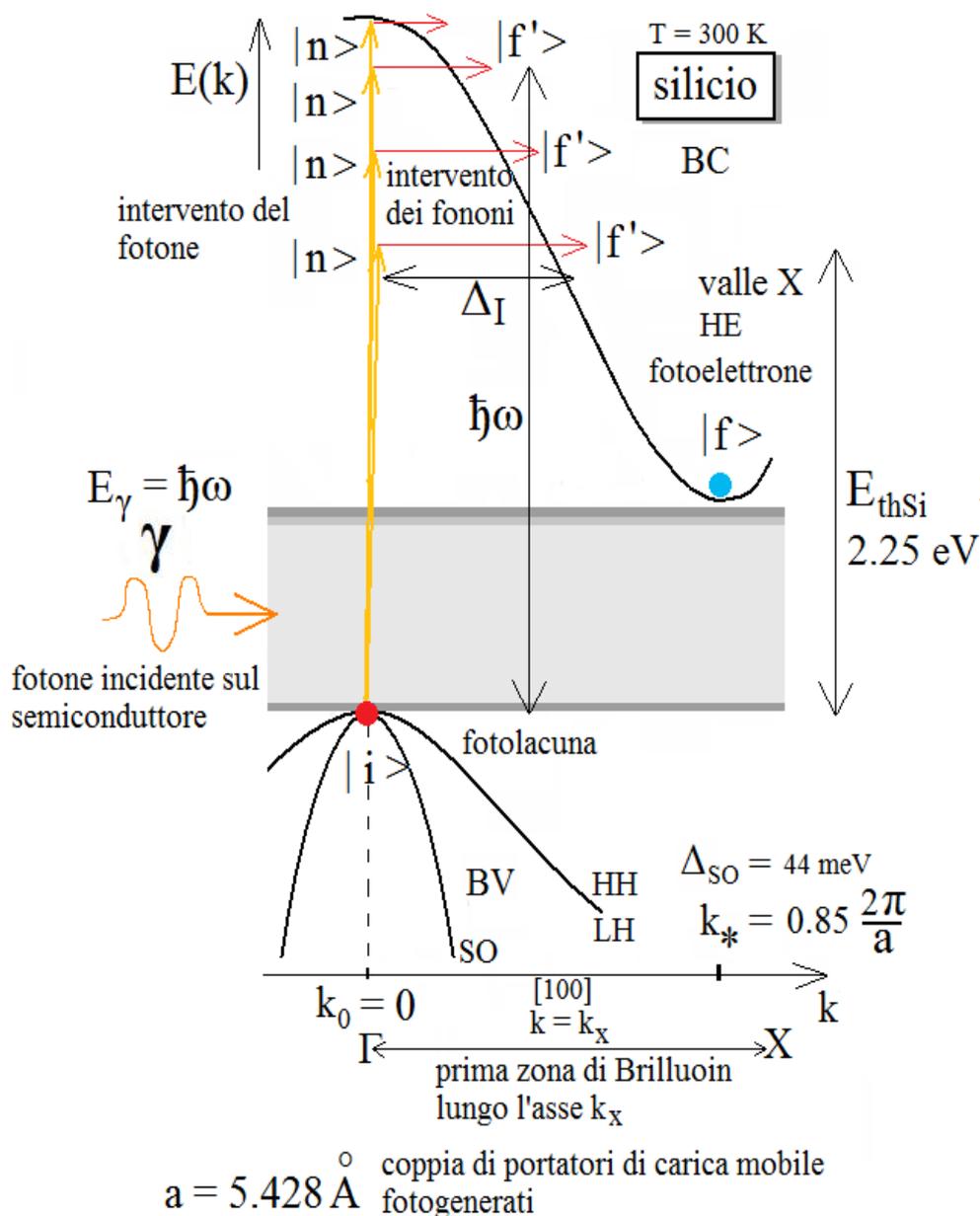


Fig. 54

Interpretazione intuitiva, mediante il diagramma a bande $E - k$ del silicio, dell'andamento parabolico $(\hbar\omega)^2$ del coefficiente di assorbimento ottico specifico $\alpha(\hbar\omega)$, riscontrabile nel caso di energie fotoniche incidenti $\hbar\omega > E_{thSi} \approx 2.25 \text{ eV}$. La distanza impulsiva Δ_i , cioè la distanza, definita nello spazio dei vettori d'onda k , fra lo stato virtuale intermedio $|n\rangle$ e lo stato elettronico permesso $|f\rangle$ (appartenente alla BC) impulsivamente più vicino ad $|n\rangle$, diminuisce all'aumentare di $\hbar\omega$. Pertanto per energie fotoniche incidenti $\hbar\omega > E_{thSi}$ crescenti la probabilità di transizione $|i\rangle \rightarrow |n\rangle \rightarrow |f\rangle$, in seguito allo scattering fra elettrone di valenza e fotone incidente, aumenta quadraticamente con $\hbar\omega$, dunque il coefficiente di assorbimento ottico specifico $\alpha(\hbar\omega)$ cresce parabolicamente all'aumentare di $\hbar\omega$.

5.8) Un possibile criterio di progetto per la larghezza della zona di svuotamento di un PIN

Per ottenere quindi una buona risposta in termini di fotocorrente, per una determinata lunghezza d'onda della luce incidente, è necessario che $QE(\lambda)$ sia adeguata, per cui la zds^* deve essere abbastanza larga, specialmente per semiconduttori a gap indiretto. Abbiamo visto nel paragrafo 4.4.2, infatti, come il gap indiretto possa introdurre delle "difficoltà" per quanto riguarda l'assorbimento di fotoni: a parità di potenza assorbita (e per range di frequenze opportuno, ad esempio $0.65 - 0.85 \mu\text{m}$), il rivelatore di Si (gap indiretto) deve avere una zds^* più lunga di quella del rivelatore di $GaAs$ (gap diretto, per cui assorbe con più facilità ed α è maggiore). Esiste un criterio geometrico per valutare un intervallo di lunghezze opportuno per la zds^* , considerando l'andamento della potenza ottica $P(x)$:

$$P(x) = A I(x) = P(0^+) e^{-\alpha x} = P_0(1 - R) e^{-\alpha x}$$

$$\operatorname{tg}\theta = - \left. \frac{dP(x)}{dx} \right|_{x=0} = P_0(1 - R) \alpha$$

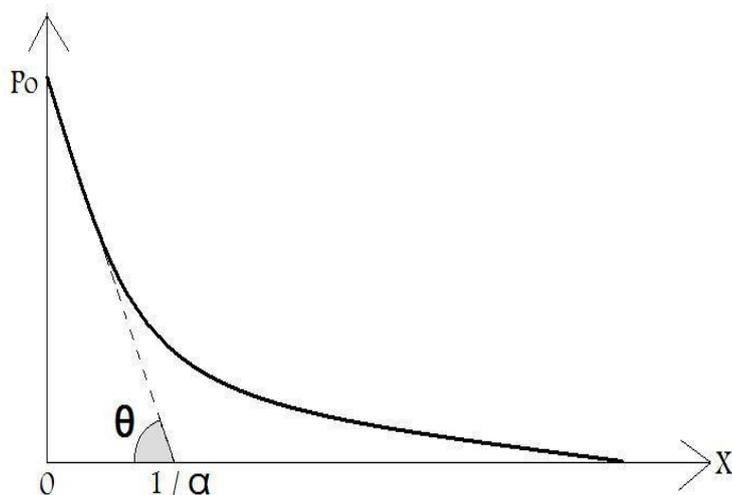


Fig. 55

Profilo $P(x)$ della potenza ottica all'interno di un fotorecettore PIN, dal quale, mediante un'osservazione geometrica, è possibile estrarre un utile criterio di progetto per la larghezza della zona di svuotamento del dispositivo.

Generalmente si costruisce una z_{ds} , agendo sullo spessore del layer intrinseco, avente una lunghezza z_{ds}^* (in presenza di polarizzazione) compresa fra $2/\alpha$ e $10/\alpha$. Il suo valore ottimo deriva da un compromesso tra efficienza quantica QE e tempo di risposta. All'aumentare di z_{ds}^* , infatti, aumenta QE, in quanto una frazione crescente della potenza incidente P_0 viene assorbita nella z_{ds}^* , ma aumenta anche il tempo di transito, a spese della velocità di risposta. Dunque se eccediamo nella lunghezza della z_{ds}^* ($>10/\alpha$) la fotocorrente $I_L(t)$, in risposta ad un impulso luminoso rettangolare, è quella colorata in verde nella figura 56.

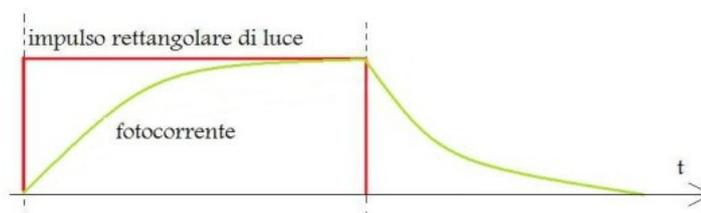


Fig. 56

Grafico della risposta elettrica $I_L(t)$ (colorata di verde) di un fotodiodo PIN, la cui z_{ds}^* è eccessivamente larga ($z_{ds}^* > 10/\alpha$). L'efficienza quantica tende ad 1, cosicchè solo una piccolissima frazione della potenza ottica incidente viene assorbita nelle regioni neutre p ed n, ovvero persa, ma la fotocorrente di uscita segue con molta lentezza il segnale ottico pilotante. La scarsa risoluzione temporale, ossia la banda operativa piccola, è dovuta al lungo tempo di transito delle fotocariche attraverso la z_{ds}^* .

Se invece scegliamo una lunghezza z_{ds}^* troppo piccola ($< 2/\alpha$) la risposta elettrica $I_L(t)$, ad un impulso luminoso rettangolare, è quella colorata in verde nella figura 57.

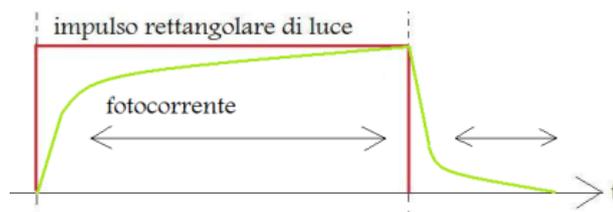


Fig. 57

Grafico della risposta elettrica $I_L(t)$ (colorata di verde) di un fotodiode PIN, la cui zds^* è eccessivamente stretta ($zds^* < 2/\alpha$). L'efficienza quantica è piccola, cosicchè solo una piccola frazione della potenza ottica incidente viene assorbita nella zds^* , e la fotocorrente di uscita segue con lentezza il segnale ottico pilotante, a causa dell'inerzia legata alla diffusione dei fotoportatori minoritari.

Per spiegare le code evidenziate nelle figure 56 e 57 è necessaria una precisazione. Con riferimento alla figura 51, abbiamo notato che la potenza assorbita è quella fra x_1 e x_2 (nella zds^*), sia perché la zona di campo è molto larga, tale da assorbire quasi tutti i fotoni incidenti, sia perché è proprio nella zds^* che a un fotone assorbito corrisponde la generazione di una fotocoppia. La potenza assorbita dal materiale nelle zone neutre è considerata "persa", perché le correnti diffusive non sono rilevabili dal circuito esterno. Questa affermazione è basata sul fatto che i fotoni assorbiti in quelle regioni sono pochi e che se un fotone dà origine ad una coppia nelle zone neutre, il portatore maggioritario della coppia dà contributo alla fotocorrente (in quanto carica già arrivata nel bulk di destinazione), mentre il corrispondente minoritario viene quasi certamente ricombinato (specialmente nel bulk p, che generalmente viene drogato in modo da essere p+, in modo da risultare molto sottile e quasi trasparente nei confronti dell'intensità ottica). Pertanto i pochi fotoni assorbiti nei bulk non danno luogo a coppie di fotoportatori, bensì a singoli fotomaggioritari, avvalorando la tesi della potenza persa nelle zone neutre. In realtà questo ragionamento vale solo nella parte più interna delle zone neutre, mentre in quella compresa entro una lunghezza di diffusione, a partire dall'interfaccia lato intrinseco/lato drogato, i fotominoritari hanno una probabilità non trascurabile di diffondere con successo verso la zds^* , e quindi di essere raccolti nel loro bulk di destinazione. Quindi la potenza, in questa zona di confine delle regioni neutre, non è persa, ma occorre attendere la diffusione, per quanto il tragitto sia abbastanza breve. Questa è la causa delle code, tanto più marcate quanto più $zds^* < 2/\alpha$. Per i PIN al silicio o al germanio la zds deve essere almeno di $20 \div 50 \mu\text{m}$, affinché QE sia accettabile, ed il tempo di transito ($\approx 200 \text{ ps}$), in tal caso, limita la banda della risposta in frequenza. Uno strato intrinseco così spesso non può essere deposto (cresciuto) mediante epitassia VPE - "Vapor Phase Epitaxy" - "epitassia da fase vapore", ovvero il tipo di epitassia che generalmente viene utilizzata per realizzare omostrutture (la VPE, infatti, si ferma a layers di massimo $15 \mu\text{m}$ di spessore), per cui si deve ricorrere ad una CVD. Nel caso di un PIN al GaAs, in cui il layer intrinseco può essere decisamente più corto (qualche μm), in virtù del gap diretto, una tecnica di crescita dello strato intrinseco largamente utilizzata in passato è la LPE - "Liquid Phase Epitaxy" - "epitassia da fase liquida". Attualmente, per PIN al GaAs, è molto più utilizzata la MOCVD - "Metal - Organic CVD", nella quale il precursore del gallio, da diffondere nella camera di lavoro insieme all'arsina AsH_3 , è una molecola organica, come ad esempio $\text{Ga}(\text{CH}_3)_3$. Nel caso di PIN all'InGaAs, semiconduttore a gap diretto, la zds può essere molto piccola ($\approx 3 \div 5 \mu\text{m}$), senza rinunciare ad una buona QE: la sua banda tipica è nell'intervallo di 3 - 5 GHz. Sono stati sperimentati anche diodi PIN con banda dell'ordine dei 20 GHz (layer intrinseco molto stretto, tempo di transito assai ridotto), ovviamente a spese dell'efficienza quantica. È possibile realizzare un PIN utilizzando un'eterostruttura in cui lo strato intrinseco è fatto di InGaAs, mentre quelli p ed n di InP.

5.9.1) Il grafico della compatibilità reticolare fra semiconduttori: utilità, descrizione ed interfacciamento tra semiconduttori di costante reticolare diversa

Supponiamo di voler rivelare una radiazione luminosa nell'intervallo di lunghezze d'onda che va da 1.3 a 1.6 μm (vicino - medio infrarosso), al quale appartengono la seconda (centrata in 1.33 μm) e la terza (centrata in 1.55 μm) finestra di attenuazione delle fibre ottiche di silice.

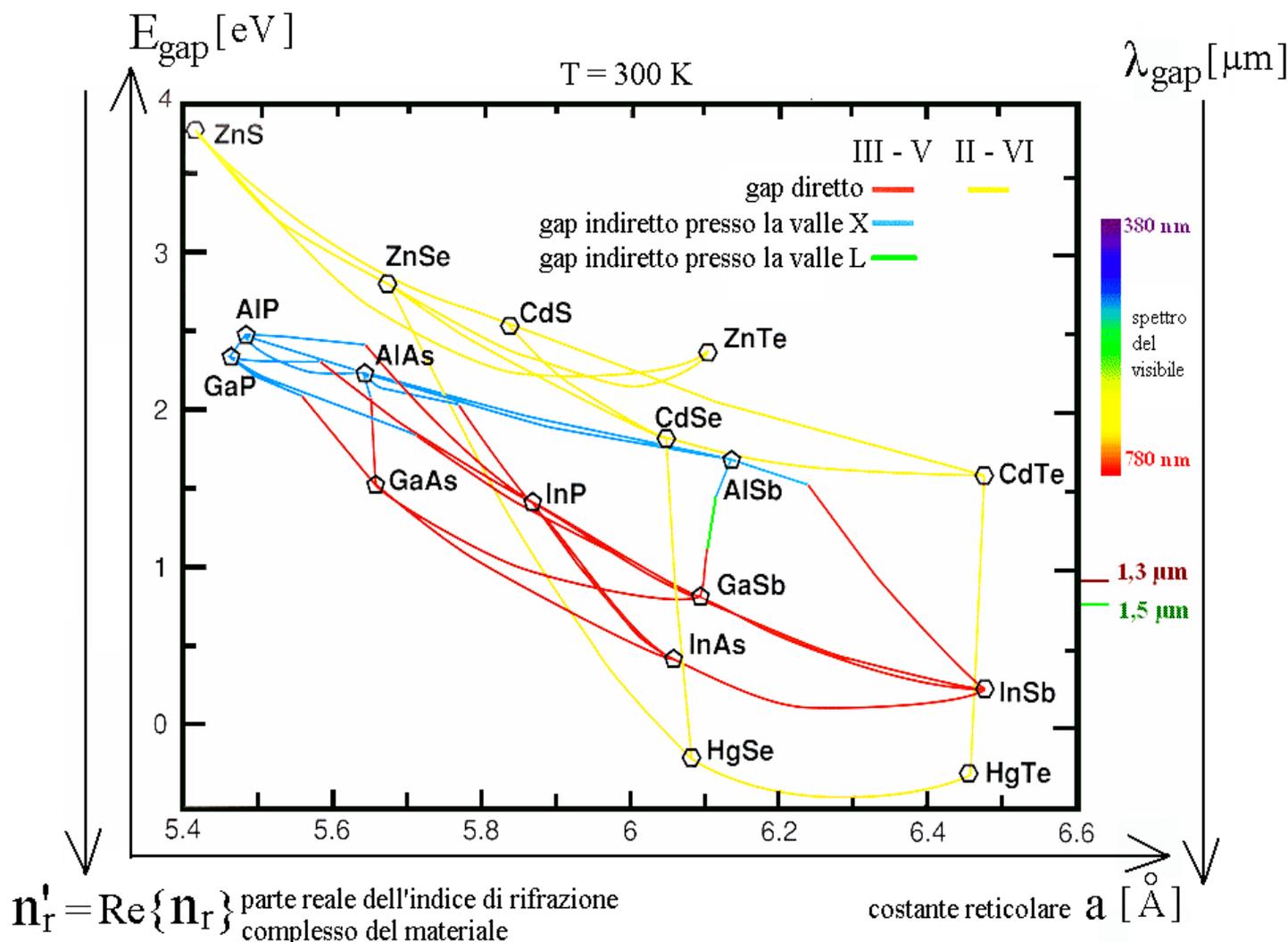


Fig. 58

Grafico della "compatibilità reticolare" ("lattice matching") fra diversi semiconduttori composti (III - V e II - VI), alcuni di questi largamente utilizzati nella fabbricazione di dispositivi optoelettronici, quali emettitori di luce (LEDs e LASERS) e fotorivelatori.

Per progettare un fotorivelatore PIN che soddisfi la suddetta specifica, ma più in generale per costruire un qualunque dispositivo optoelettronico ad eterogiunzioni (LEDs e LASERS, ad esempio), è necessario tener conto del grafico riportato in figura 58, noto come grafico della "compatibilità reticolare" ("lattice matching") fra semiconduttori. Questo grafico ha in ascissa il passo reticolare "a" (misurato in \AA), che caratterizza, da un punto di vista spaziale, la struttura e le proprietà cristalline dei materiali raffigurati, e in ordinata l'energia di gap E_{gap} (misurata in eV) e, nel verso opposto, sia la parte reale n'_r dell'indice di rifrazione complesso dei semiconduttori che la lunghezza d'onda λ_{gap} della radiazione luminosa del fotone che, per un dato semiconduttore, può produrre la transizione di un elettrone dal livello energetico E_V a quello E_C . Quest'ultima scala di grandezza, che è del tutto equivalente a quella dell'energia di gap, è usata soprattutto nell'ambito delle telecomunicazioni su fibra ottica, settore in cui è consuetudine, per motivi storici, riferirsi alle lunghezze d'onda λ .

Il grafico contiene dei punti, collegati fra loro da delle linee: i punti rappresentano dei semiconduttori binari, strutturati secondo la geometria zincoblenda, ciascuno ottenibile mediante un opportuno processo di crescita epitassiale realizzata in condizioni di ultra - alto vuoto (MBE - “Molecular Beam Epitaxy” - “epitassia da fasci molecolari”, descritta brevemente nell’appendice A). Ciascun semiconduttore binario è costituito da un elemento del V° gruppo (l’elemento scritto a destra) e da uno del III° gruppo (l’elemento scritto a sinistra), oppure da uno del VI° gruppo (scritto a destra) ed uno del II° gruppo (scritto a sinistra). Ciascun punto, ossia ciascun semiconduttore binario, ha come ascissa la propria costante reticolare, e come ordinata il valore dell’energia del gap più piccolo (che è rintracciabile, nel diagramma a bande E - k, nel punto Γ nel caso di gap diretto, nei punti X o L nel caso di gap indiretto) tipico della propria struttura a bande. Ciascuna linea collega due semiconduttori che differiscono, fra loro, unicamente per uno dei due elementi che li costituiscono. Ad esempio il G_aP è collegato, mediante una linea, all’ G_aA_s : in questo caso, percorrendo la linea in un senso o nell’altro, varia solo l’elemento del V° gruppo. Oppure l’ G_aA_s è collegato, mediante una linea, all’ I_nA_s : in questo caso, percorrendo la linea in un senso o nell’altro, varia solo l’elemento del III° gruppo. Su una qualunque linea del grafico si trovano i semiconduttori ternari. Ad esempio lungo la prima linea citata si trova il semiconduttore $G_aA_s(x)P_{(1-x)}$, dove la concentrazione dell’elemento del III° gruppo rimane invariata, mentre variano quelle degli elementi del V° gruppo. Se ci spostiamo lungo la linea, partendo dal G_aP (punto in cui $x \rightarrow 0$), la concentrazione (= frazione molare) del P diminuisce a favore di quella dell’ A_s (x incrementa, $0 < x < 1$), fino ad arrivare al punto in cui è rappresentato il G_aA_s (punto in cui $x \rightarrow 1$). Sulla seconda linea citata si trova il semiconduttore $I_n(x)G_a(1-x)A_s$ ($0 < x < 1$), lungo la quale la concentrazione dell’elemento del V° gruppo rimane invariata, mentre variano quelle degli elementi del III° gruppo. La rete complessiva delle linee è suddivisibile in sottogruppi, detti “quadrilateri” o “cicli”. Ciascun ciclo consta di quattro estremi (quattro punti), ciascuno dei quali rappresenta un semiconduttore binario. Ciascuno dei suoi quattro lati rappresenta un semiconduttore ternario, mentre l’interno del ciclo (la sua superficie) rappresenta il semiconduttore quaternario da cui il quadrilatero stesso prende il nome. I due cicli più importanti per la fabbricazione dei dispositivi optoelettronici sono:

- il quadrilatero $G_aP - G_aA_s - I_nA_s - I_nP$, detto “ciclo $I_nG_aA_sP$ ”; al suo interno è presente il semiconduttore quaternario $I_n(1-x)G_a(x)A_s(1-y)P(y)$
- il quadrilatero $A_lA_s - G_aA_s - G_aS_b - A_lS_b$, detto “ciclo $S_bA_lG_aA_s$ ”; al suo interno è presente il semiconduttore quaternario $S_b(1-x)A_l(x)G_a(1-y)A_s(y)$

Un dispositivo optoelettronico ad eterogiunzioni è realizzabile mediante crescite epitassiali particolari di strati di semiconduttori composti su altri. Oltre alle complicazioni di ordine prettamente tecnologico, occorre tenere in considerazione gli eventuali stress meccanici localizzati all’interfaccia fra un semiconduttore composto ed un altro. Questi stress sono tanto maggiori quanto maggiore è la differenza fra le costanti reticolari caratteristiche dei due materiali cresciuti uno sopra l’altro. Indicando con a_s la costante reticolare del semiconduttore che costituisce il substrato, sul quale viene cresciuto uno strato di un altro semiconduttore, la cui costante reticolare la indichiamo con a_L , possiamo definire lo stress meccanico reticolare (“lattice strain”) ε nel modo seguente:

$$\varepsilon = \frac{|a_L - a_s|}{a_L}$$

Ad esempio lo stress meccanico fra un substrato di G_aA_s ed un layer sovrastante di I_nA_s è circa del 7%, cioè grandissimo; lo stress meccanico fra un substrato di Si ed un layer sovrastante di GeSi (il silicio ed il silicato di germanio non sono presenti nel grafico di figura 58) è di circa il 4%. Lo stress ε è abbastanza trascurabile se i due materiali interfaccianti hanno circa la stessa costante reticolare, ovvero se appartengono, sul grafico di figura 58, ad una medesima linea approssimativamente verticale. È il caso dell’interfaccia fra G_aA_s ed A_lA_s , presso la quale lo stress meccanico ε è circa dello 0.13%, cioè di fatto nullo. Quest’ultima osservazione costituisce una sorta di criterio di “compatibilità reticolare” da adottare per l’integrazione delle eterostrutture.

Lo stress meccanico fra un semiconduttore di costante reticolare a_L , cresciuto sopra un substrato costituito da un semiconduttore di costante reticolare a_s , con $a_L > a_s$, produce un adeguamento della dimensione delle celle fondamentali del semiconduttore del “top layer”, sul piano parallelo all’interfaccia, rispetto alla dimensione delle celle fondamentali del semiconduttore del “bottom layer” (= substrato): a_L si restringe, si comprime, fino ad eguagliare a_s , ovvero la superficie quadrata, la base, delle celle fondamentali del top layer, comprimendosi, va a coincidere con la superficie quadrata delle celle fondamentali del bottom layer. La struttura reticolare del top layer, sul piano parallelo all’interfaccia, diventa “coerente” con quella del bottom layer. In tal modo la superficie interfacciante del top layer è caratterizzata dalla stessa densità superficiale di legami covalenti che caratterizza la superficie interfacciante del substrato, evitando che uno dei due layers proponga, all’interfaccia, un maggior numero di legami rispetto all’altro: in tal modo non si ha formazione di legami incompleti all’interfaccia,

facilitando così il transito dei portatori di carica (elettroni e lacune) attraverso l'interfaccia. Infatti eventuali legami rotti, cioè incompleti, localizzati presso l'interfaccia, ovvero regioni locali di orbitali ibridi sp^3 occupate soltanto da un elettrone, causerebbero un decremento della probabilità P con la quale un portatore di carica, inizialmente localizzato su uno stato $|1\rangle$ all'interno di uno dei due layers interfaccianti, si sposta, per esempio a causa di un campo elettrico imposto dall'esterno, su uno stato $|2\rangle$ all'interno dell'altro layer. P è proporzionale al consueto elemento di matrice della quantità di moto \vec{p} del portatore di carica. In formula:

$$P \propto |\langle 2 | \vec{a} \cdot \vec{p} | 1 \rangle|^2 \quad [e72]$$

dove \vec{a} è il versore di polarizzazione del campo elettrico risultante fra il campo elettrico interno (periodico) al cristallo del layer di partenza, dove è localizzato lo stato iniziale $|1\rangle$, ed il campo elettrico imposto dalla polarizzazione esterna. Contemporaneamente al restringimento di a_L lungo il piano parallelo all'interfaccia si verifica un allungamento di a_L perpendicolarmente al suddetto piano, cioè un incremento dell'altezza delle celle fondamentali del semiconduttore del top layer, rispetto all'altezza delle celle fondamentali del semiconduttore del bottom layer: le celle fondamentali del top layer assumono pertanto una forma a parallelepipedo, la cui altezza è tanto maggiore quanto maggiore è il valore di ε fra i due layers, cioè quanto maggiore è il mismatch reticolare fra i due semiconduttori interfaccianti. Pertanto lo stress meccanico ε lungo il piano parallelo all'interfaccia, prodotto dalla compressione della base delle celle fondamentali del top layer in funzione della base delle celle del substrato, comporta proporzionalmente uno stress meccanico ε_{\perp} perpendicolare al piano dell'interfaccia, ed è quest'ultimo a determinare l'incremento dell'altezza delle celle del top layer. In formule:

$$\varepsilon_{\perp} = -\frac{\varepsilon}{\sigma}$$

dove σ , che ovviamente è un numero puro come ε ed ε_{\perp} , è noto come "coefficiente di Poisson" e dipende da molti fattori, per esempio dall'orientamento reticolare del substrato ([100], [010], [111] ecc...), del bottom layer, rispetto a quello del top layer. Il segno negativo indica che se lo stress parallelo all'interfaccia è legato ad una compressione di a_L lungo tale piano, ovvero ad una compressione della base delle celle del top layer, allora il conseguente stress perpendicolare all'interfaccia è legato ad una dilatazione di a_L , ortogonalmente all'interfaccia, cioè ad una dilatazione dell'altezza delle celle del top layer. Viceversa se $a_L < a_S$ abbiamo, nel caso in cui il top layer divenga coerente con il bottom layer, un allargamento della base delle celle del top layer, ovvero un dilatamento di a_L lungo il piano parallelo all'interfaccia, che eguaglia così a_S , e contemporaneamente una compressione di a_L ortogonalmente all'interfaccia, ovvero una compressione dell'altezza delle celle del top layer. Il layer sovrastante tende ad interfacciarsi coerentemente con il substrato, sia con $a_L > a_S$ che con $a_L < a_S$, nel caso in cui lo stress meccanico accumulato non sia troppo grande, cioè nel caso in cui lo spessore d_L del top layer non sia eccessivo, ovvero sia inferiore ad una certa soglia critica, chiamata "critical thickness" – "spessore critico" d_{CR} , che è inversamente proporzionale allo stress meccanico ε . d_{CR} , in funzione di ε , è definito nel modo seguente:

$$d_{CR}(\varepsilon) = \frac{a_S}{2\varepsilon}$$

ad esempio se cresciamo uno strato epitassiale ("epilayer") di $In_{(0.3)}Ga_{(0.7)}As$ su un substrato di $GaAs$ ($a_S \approx 5.65 \text{ \AA}$) abbiamo uno stress meccanico reticolare ε , all'interfaccia, pari a circa 2.2%, pertanto uno spessore critico d_{CR} pari a circa 13 nm; se cresciamo uno strato epitassiale di $Ge_{(0.3)}Si_{(0.7)}$ su un substrato di Si abbiamo uno spessore critico d_{CR} pari a circa 70 nm. Il grafico semilogaritmico riportato in figura 59 mostra la dipendenza, approssimativamente a ramo di iperbole equilatera, dello spessore critico d_{CR} di un top layer (più precisamente un epilayer, cresciuto con la tecnica MBE) di $In_{(x)}Ga_{(1-x)}As$, cresciuto su un substrato (bottom layer) di $GaAs$, dalla frazione molare x ($0 < x < 1$) di indio. Inoltre è riportata la dipendenza, approssimativamente a ramo di iperbole equilatera, dello spessore critico d_{CR} di un epilayer (cresciuto con la tecnica MBE) di $Ge_{(x)}Si_{(1-x)}$, deposto su un substrato di Si , dalla frazione molare x ($0 < x < 1$) di germanio. In entrambi i casi x è proporzionale allo stress meccanico ε che si determina all'interfaccia dei due layers, pertanto possiamo confondere, da un punto di vista qualitativo, lo stress ε con la frazione molare x , e scrivere la seguente funzione approssimativa $d_{CR}(x)$, la quale è evidentemente riscontrabile nel grafico di figura 59:

$$d_{CR}(x) \propto \frac{1}{x}$$

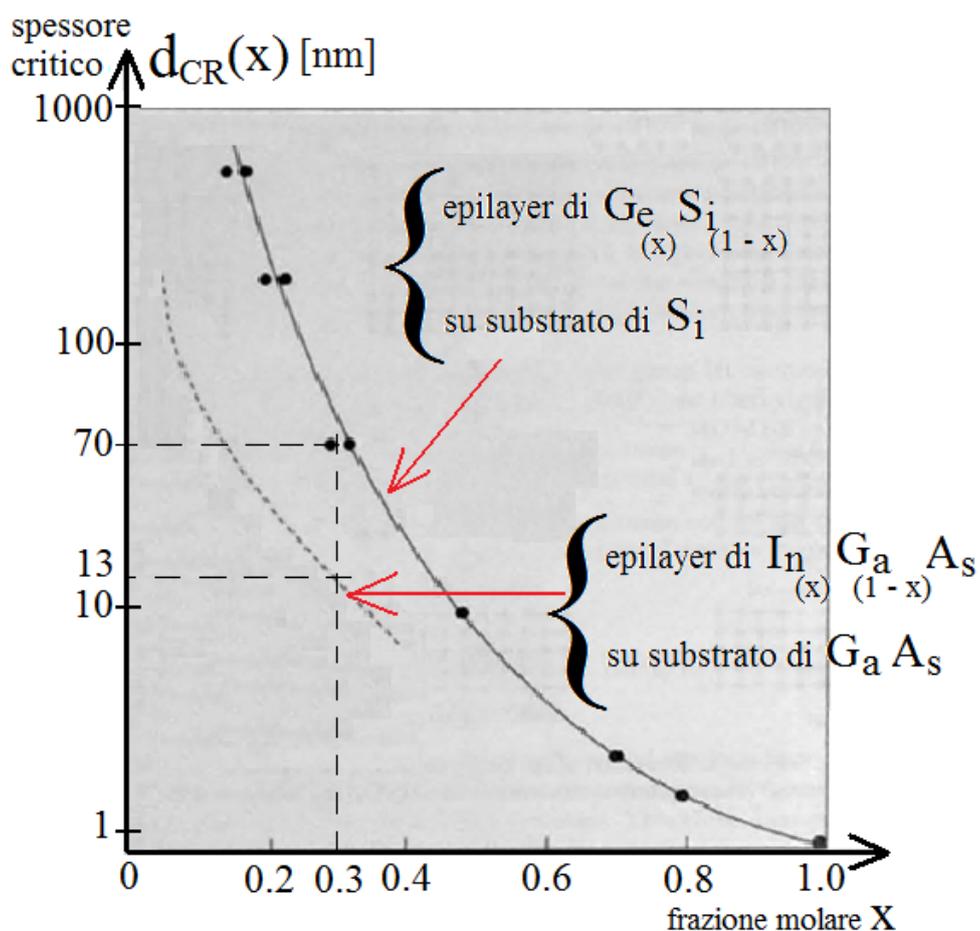


Fig. 59

Grafico semilogaritmico della dipendenza, approssimativamente a ramo di iperbole equilatera, dello spessore critico d_{CR} di un epilayer (cresciuto con la tecnica MBE) di $In_xGa_{1-x}As$, deposto su un substrato di $GaAs$, dalla frazione molare x ($0 < x < 1$) di indio, e della dipendenza dalla frazione molare x di germanio, che è sempre approssimabile con un ramo di iperbole equilatera, dello spessore critico d_{CR} di un epilayer (cresciuto con la tecnica MBE) di Ge_xSi_{1-x} , deposto su un substrato di Si . In entrambi i casi x è proporzionale allo stress meccanico ϵ che si determina all'interfaccia dei due layers.

Riportiamo in figura 60 uno schema qualitativo riassuntivo delle due situazioni possibili ($a_L > a_S$ ed $a_L < a_S$) di mismatch reticolare presso l'interfaccia di una generica eterogiunzione, nel caso in cui $d_L < d_{CR}$, ovvero nel caso in cui il top layer riesca a riordinarsi, da un punto di vista reticolare, in modo coerente rispetto al substrato.

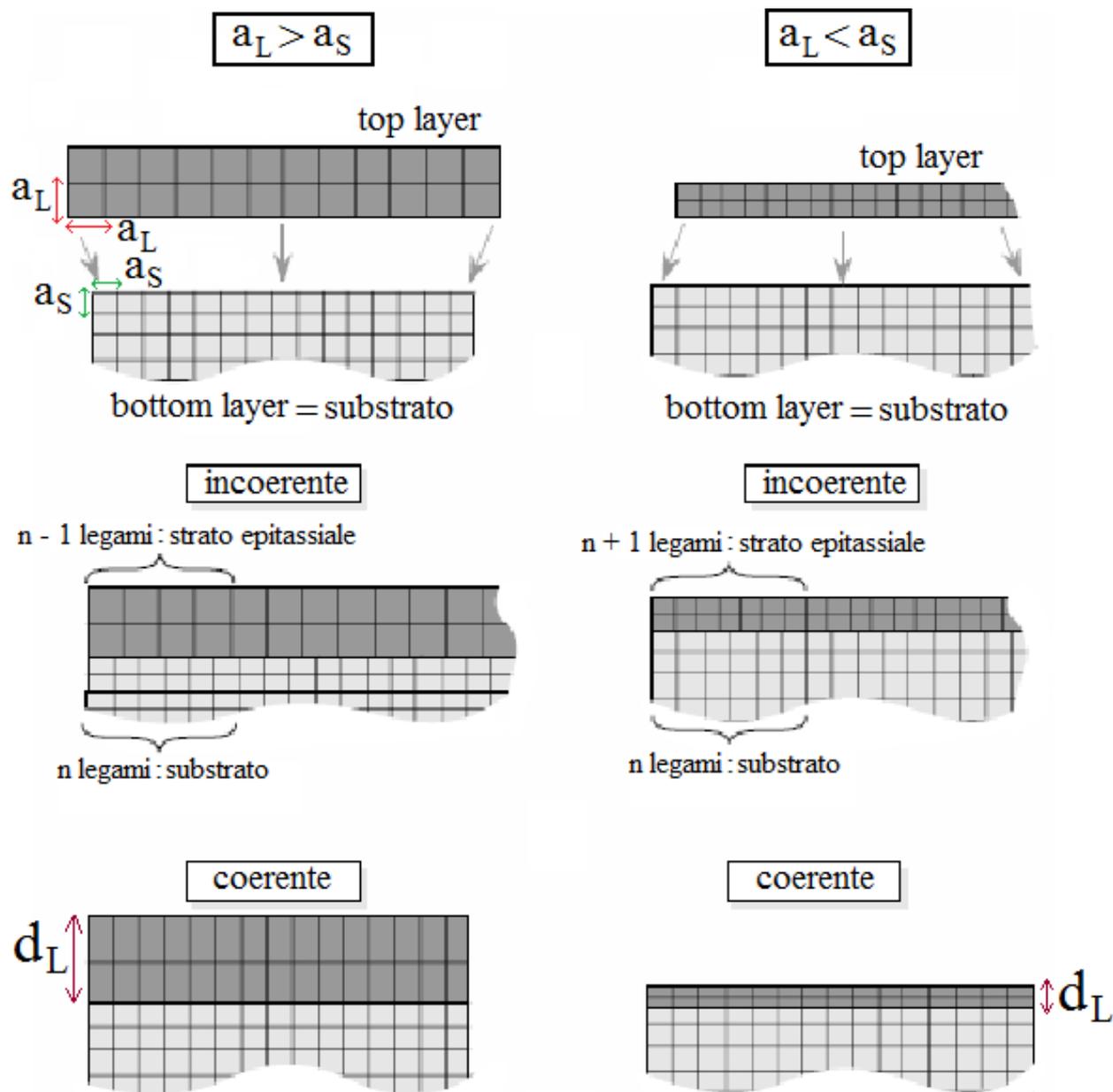


Fig. 60

Schema qualitativo riassuntivo delle due situazioni possibili ($a_L > a_S$ ed $a_L < a_S$) di mismatch reticolare che si possono verificare all'interfaccia di una generica eterogiunzione, nel caso in cui $d_L < d_{CR}$, ovvero nel caso in cui il top layer (epilayer cresciuto sul substrato mediante la tecnica MBE) riesca a riordinarsi, da un punto di vista reticolare, in modo coerente rispetto al substrato.

Lo stress meccanico reticolare ϵ , a cui è sottoposto uno strato di semiconduttore avente costante reticolare a_L , cresciuto per epitassia su un substrato di semiconduttore avente costante reticolare a_S , produce una distorsione della geometria cristallina dell'epilayer che ha ripercussioni anche sulla struttura del diagramma a bande $E - k$ del semiconduttore del top layer. Infatti la forma della BV e della BC del semiconduttore del top layer, nell'ipotesi che d_L non superi d_{CR} , cioè che il cristallo dell'epilayer accumuli energia meccanica di stress, risultano modificate. Ad esempio nel caso in cui l'epilayer depositato sia costituito da un semiconduttore a gap diretto e se la direzione di crescita del top layer (ortogonale all'interfaccia epilayer/substrato) coincide con la direzione cristallografica [001], assunta dal substrato e di conseguenza dal layer cresciuto sul substrato, BV e BC risultano modificate nel modo seguente, come è possibile osservare in figura 61:

- se $a_L > a_S$, cioè nel caso di compressione delle celle fondamentali del top layer lungo il piano parallelo all'interfaccia, al fine di conformarsi a quelle del bottom layer, il gap (diretto) presso il punto Γ si

allarga, la sottobanda di valenza LH splitta rispetto a quella HH (HH rimane sopra LH) di una quantità Δ_{str} pari, come ordine di grandezza, a circa 100 meV (lo splitting fra HH ed LH è notevole), pertanto la doppia degenerazione (o quadrupla, se consideriamo anche lo spin) fra stati elettronici HH e stati elettronici LH è parzialmente rimossa

- se $a_L < a_S$, cioè nel caso di dilatazione delle celle fondamentali del top layer lungo il piano parallelo all'interfaccia, al fine di conformarsi a quelle del bottom layer, il gap (diretto) presso il punto Γ si restringe, la sottobanda di valenza LH splitta rispetto a quella HH (quest'ultima si sposta al di sotto della LH) di una quantità Δ_{str} pari sempre, come ordine di grandezza, a circa 100 meV (splitting notevole), pertanto anche adesso la doppia degenerazione (o quadrupla, se consideriamo lo spin) fra stati elettronici HH e stati elettronici LH è parzialmente rimossa

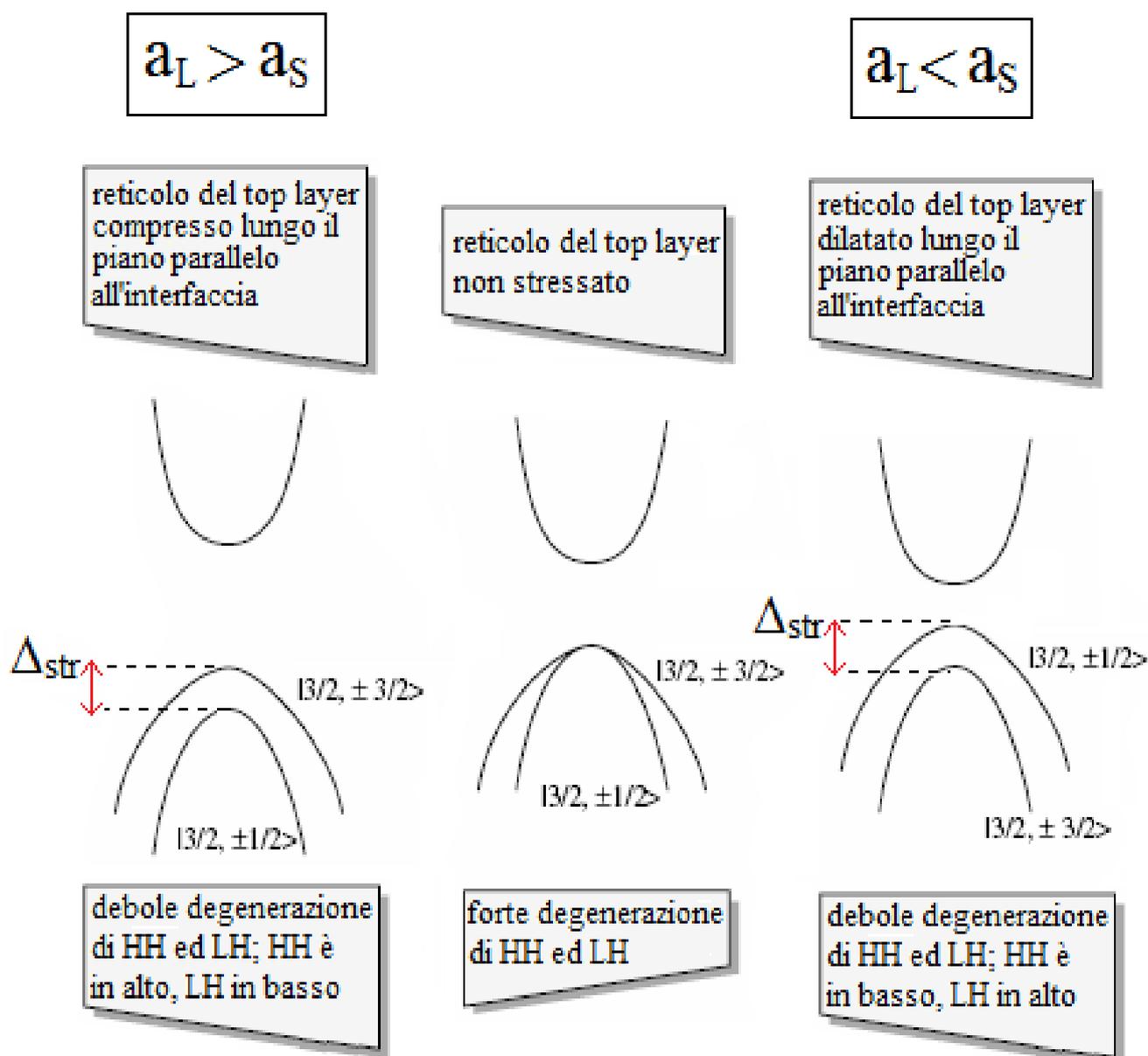


Fig. 61

Rappresentazione qualitativa riassuntiva delle alterazioni energetiche che può subire il diagramma a bande $E - k$ di un semiconduttore a gap diretto, di costante reticolare a_L , cresciuto per epitassia su un substrato fatto da un semiconduttore di costante reticolare a_S , nelle due situazioni possibili ($a_L > a_S$ ed $a_L < a_S$) di mismatch reticolare che si possono verificare all'interfaccia, nel caso in cui $d_L < d_{cr}$.

È evidente come il mismatch reticolare comporti alcune importanti alterazioni delle proprietà ottiche del semiconduttore costituente la “strained active region”, ovvero la zona attiva del fotorivelatore sottoposta a stress meccanici ε ed ε_{\perp} : ad esempio può cambiare la frequenza ottica di massima risposta (elettrica) del fotorivelatore, a causa dell’alterazione del gap, oppure può cambiare l’efficienza quantica QE del fotorivelatore. Quest’ultimo punto può essere compreso, seppur intuitivamente, osservando i diagrammi qualitativi $E - k$ di figura 61 e considerando i sei elementi di matrice di “accoppiamento ottico” riportati nel paragrafo 5.3.3.2, alle pagine 77 e 78. Infatti supponiamo che la regione attiva del fotodiodo riceva una luce polarizzata lungo gli assi convenzionali xy . Supponiamo che nella regione attiva le celle fondamentali abbiano subito una compressione lungo il piano parallelo all’interfaccia con il substrato (il substrato lo consideriamo non attivo otticamente), pertanto siamo nel caso $a_L > a_S$. Se non ci fosse stress reticolare il contributo alla fotocorrente, in uscita dal fotodiodo, dovuto ad HH sarebbe tre volte maggiore di quello dovuto alla sottobanda di valenza LH, ovvero il numero di fotoelettroni forniti dalla sottobanda di valenza HH sarebbe tre volte maggiore di quello fornito da LH. Lo stress meccanico comporta un allontanamento energetico di LH da HH tutt’altro che trascurabile ($\Delta_{str} \approx 100$ meV), cioè gli stati posizionati sulla curva $E(k)$ di LH diventano di fatto ortogonali rispetto agli stati posizionati sulla curva $E(k)$ di HH, pertanto il contributo di LH alla fotoconduzione diminuisce. Tuttavia il contributo di LH è già piccolo in assenza di stress (un terzo di quello di HH), per cui un’ulteriore diminuzione dovuta allo stress comporta una piccola diminuzione della QE del fotorivelatore. La situazione cambia drasticamente nel caso in cui la stessa luce polarizzata lungo xy incide su un fotorivelatore nella cui regione attiva le celle fondamentali hanno subito una dilatazione lungo il piano parallelo all’interfaccia con il substrato (il substrato lo consideriamo sempre non attivo otticamente), pertanto nel caso $a_L < a_S$. Questa volta la sottobanda di valenza HH scende al di sotto di quella LH di un’energia tutt’altro che trascurabile (sempre $\Delta_{str} \approx 100$ meV), cioè gli stati posizionati sulla curva $E(k)$ di HH diventano ortogonali rispetto agli stati posizionati sulla curva $E(k)$ di LH, pertanto il contributo di HH alla fotoconduzione diventa irrisorio. L’unico contributo fotoconduttivo è legato alla sottobanda di valenza LH, che si trova al di sopra di HH, però il contributo di LH è piccolo, come già osservato, pertanto la QE del fotorivelatore diminuisce considerevolmente rispetto al caso precedente. Ancora più netta è la differenza di QE, fra il caso $a_L > a_S$ e quello $a_L < a_S$, legata ad una luce incidente polarizzata lungo l’asse convenzionale z . Se $a_L < a_S$ il contributo fotoconduttivo della sottobanda di valenza HH è decrementato rispetto a quello che si avrebbe in assenza di stress meccanico, dal momento che HH è notevolmente splittata verso il basso rispetto ad LH. Tuttavia per polarizzazione della luce lungo z l’accoppiamento ottico fra la BC e la sottobanda di valenza HH sarebbe comunque nullo, anche in assenza di stress reticolare ε , pertanto lo splitting di HH rispetto ad LH non altera significativamente il contributo fotoconduttivo di HH: si ha un’ulteriore diminuzione dell’accoppiamento ottico fra la BC e la sottobanda di valenza HH, il quale era già approssimativamente nullo. Quindi la QE è legata unicamente all’accoppiamento ottico fra la BC e la sottobanda di valenza LH. Se $a_L > a_S$ il contributo fotoconduttivo della sottobanda di valenza LH, piccolo ma non nullo nel caso di assenza di stress meccanico, è di fatto annullato a causa dello splitting di LH rispetto ad HH, traslazione energetica dovuta al mismatch reticolare fra regione epitassiale otticamente attiva e substrato su cui questa è stata cresciuta. Gli unici stati elettronici di valenza occupati da elettroni atomici che potrebbero, almeno in teoria, scatterare con i fotoni incidenti ed andare in BC, per poter poi essere raccolti da un campo elettrico di polarizzazione esterna e formare la fotocorrente, sono localizzati presso la curva $E(k)$ della sottobanda di valenza HH. Ma questa, come già spiegato, non è accoppiabile otticamente con la BC da una luce incidente polarizzata lungo z (l’elemento matriciale dell’impulso, legato a questa transizione, è nullo). Quindi in quest’ultimo caso l’efficienza quantica QE del fotorivelatore è praticamente nulla, cioè il fotodiodo non risponde (elettricamente) in alcun modo alla luce, polarizzata lungo z , che incide sulla sua regione attiva.

Nel caso in cui lo spessore del layer sovrastante superiori, durante la sua crescita epitassiale sul substrato, il proprio spessore critico, ossia $d_L > d_{CR}$, l’energia meccanica accumulata dal reticolo, nelle regioni limitrofe rispetto all’interfaccia, risulta troppo grande e quindi viene rilasciata, con conseguente ritorno delle celle fondamentali, del semiconduttore costituente il top layer, alla forma cubica originaria, cioè quella tipica del semiconduttore non interfacciato con nessun altro materiale (semiconduttore privo di stress). Pertanto il rilascio della “strain energy” (energia di stress), a causa dell’eccessivo spessore d_L dell’epilayer, comporta l’incoerenza della struttura reticolare del top layer rispetto a quella del substrato. Tale incoerenza può produrre due tipi di irregolarità cristallina all’interno del top layer:

- la formazione di difetti reticolari di linea, come ad esempio le dislocazioni (a vite, a spigolo ecc...), e di punto, come le vacanze, concentrati prevalentemente nelle regioni del top layer vicine all’interfaccia con il substrato. Ciò significa che all’interno dell’epilayer, nel suo bulk, tali difetti sono di fatto assenti, pertanto i portatori di carica transitanti da un layer ad un altro vedono, al più, una barriera di potenziale localizzata unicamente nei dintorni dell’interfaccia e quindi facilmente attraversabile per effetto tunnel;

la probabilità P di attraversamento dell'interfaccia, riportata nella [e72], non risente della barriera in modo significativo. Questa è, fra le due possibili irregolarità cristalline, quella meno deleteria per i dispositivi a fini fotonici.

- La formazione di difetti reticolari di linea, come le dislocazioni, che penetrano all'interno del top layer; in particolare si possono formare colonne ad alta concentrazione di atomi alternate a colonne a bassa concentrazione di atomi o ad assenza di atomi. Queste imperfezioni compromettono significativamente le prestazioni dei dispositivi optoelettronici, decrementando notevolmente la probabilità di transizione P .

In figura 62 è riportato uno schema qualitativo riassuntivo delle due possibili irregolarità reticolari che possono verificarsi all'interno dello strato epitassiale, cresciuto su un substrato, nel caso di rilascio dell'energia meccanica di stress accumulata dal top layer durante la sua crescita, cioè nel caso in cui $d_L > d_{CR}$.

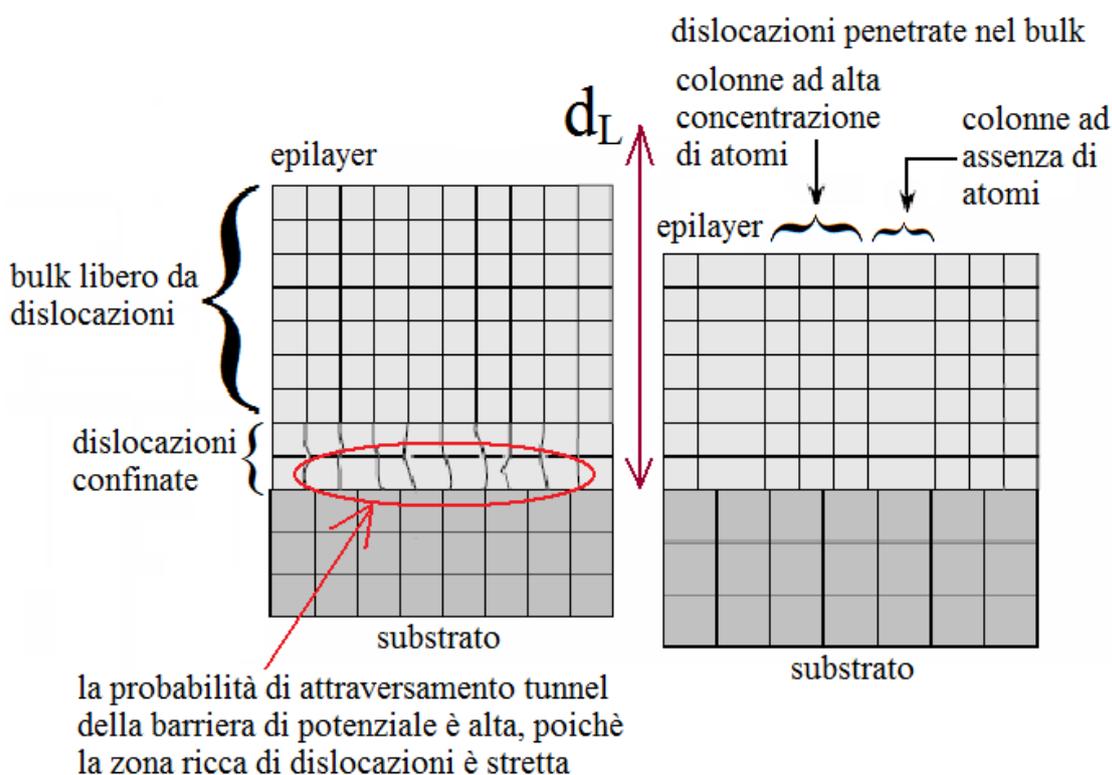


Fig. 62

Schema qualitativo riassuntivo delle due possibili irregolarità reticolari che possono verificarsi all'interno dello strato epitassiale, cresciuto su un substrato di costante reticolare diversa, nel caso di rilascio dell'energia meccanica di stress accumulata dal top layer durante la sua crescita, cioè nel caso in cui $d_L > d_{CR}$.

5.9.2) Il grafico della compatibilità reticolare fra semiconduttori: la legge di Vegard

Riferiamoci sempre al grafico, mostrato in figura 58, della compatibilità reticolare fra semiconduttori III - V e II - VI. Consideriamo adesso due qualsiasi punti, uno rappresentativo di un semiconduttore binario A e l'altro rappresentativo di un semiconduttore binario B; i due punti sono collegati da una linea che rappresenta il semiconduttore ternario $A_{(x)}B_{(1-x)}$. Supponiamo di conoscere sperimentalmente, per mezzo della cristallografia a raggi X (la quale presuppone la riflessione alla Bragg dei raggi X sul reticolo cristallino sotto osservazione), la costante reticolare a_A del semiconduttore binario A e la costante reticolare a_B del semiconduttore binario B. Per conoscere la costante reticolare a_{Alloy} del semiconduttore ternario $A_{(x)}B_{(1-x)}$ ($\forall x \in (0,1)$), cresciuto con la tecnica MBE, si può adottare nuovamente la procedura sperimentale, ovvero calcolare a_{Alloy} con la cristallografia a raggi X. Tuttavia, sotto opportune ipotesi, è stato osservato che il risultato sperimentale di a_{Alloy} è ottenibile, con un

marginale di errore abbastanza trascurabile, anche con l'utilizzo della seguente formula empirica, nota come "legge di Vegard":

$$a_{\text{Alloy}} = a_A (x) + a_B (1 - x) \quad [e73]$$

ad esempio se il semiconduttore binario A è il InAs, il semiconduttore binario B è il GaAs ed il semiconduttore ternario $A_{(x)}B_{(1-x)} = A_{(0.57)}B_{(0.43)}$ è il $\text{In}_{(x)}\text{Ga}_{(1-x)}\text{As} = \text{In}_{(0.57)}\text{Ga}_{(0.43)}\text{As}$ (se $x \rightarrow 0$ abbiamo che $A_{(x)}B_{(1-x)} \rightarrow \text{GaAs}$, mentre se $x \rightarrow 1$ abbiamo che $A_{(x)}B_{(1-x)} \rightarrow \text{InAs}$), possiamo applicare la legge di Vegard, sotto opportune ipotesi, e calcolare la costante reticolare a_{Alloy} del semiconduttore ternario, conoscendo, oltre al valore di x (in questo caso $x = 0.53$), le costanti reticolari a_A e a_B , cioè rispettivamente quella di InAs e quella di GaAs. In formule:

$$a_{\text{In}_{(0.57)}\text{Ga}_{(0.43)}\text{As}} = a_{\text{InAs}} (0.57) + a_{\text{GaAs}} (0.43) \approx [6.04 (0.57)]A^\circ + [5.65 (0.43)]A^\circ \approx 5.86 A^\circ \approx a_{\text{InP}}$$

Pertanto il $\text{In}_{(0.57)}\text{Ga}_{(0.43)}\text{As}$ ed il InP sono due semiconduttori composti che non presentano problemi di mismatch reticolare, ovvero lo stress meccanico ε all'interfaccia dei due materiali è trascurabile e quindi l'eterostruttura $\text{In}_{(0.57)}\text{Ga}_{(0.43)}\text{As}/\text{InP}$ si presta molto bene alla realizzazione di fotorivelatori di II° e III° finestra, come mostrato nel paragrafo 5.9.3.

Affinchè la legge di Vegard fornisca un risultato sufficientemente accurato, cioè un valore della costante reticolare a_{Alloy} del semiconduttore ternario abbastanza prossimo a quello ricavabile dalle evidenze sperimentali, è necessario che siano verificate le due seguenti condizioni:

- il semiconduttore binario A ed il semiconduttore binario B, sublimati e depositi in modo controllato (controllando la frazione molare x) all'interno della camera di lavoro (depressurizzata) in cui avviene la MBE, in modo da crescere il semiconduttore ternario $A_{(x)}B_{(1-x)}$, hanno la stessa struttura cristallina. Pertanto A e B sono entrambi cristalli strutturati a zinoblenda, come nel caso dei semiconduttori binari mostrati in figura 58, oppure entrambi cristalli strutturati a diamante, come nel caso del C, del Si o del Ge, oppure entrambi cristalli strutturati secondo la geometria HCP - "Hexagonal Close Pack", come i nitrucci InN, GaN, AlN e BN (si veda il glossario).
- Il semiconduttore ternario $A_{(x)}B_{(1-x)}$ non deve essere cresciuto in modalità "phase separated alloy", detta anche "clustered alloy", cioè non deve essere un composto cresciuto a fasi separate, a "clusters", a grani; in altre parole la matrice cristallina di $A_{(x)}B_{(1-x)}$ non deve essere formata da un'alternanza, più o meno casuale, di grani di tipo A e grani di tipo B, cioè di zone prevalentemente composte da atomi di A alternate a zone prevalentemente composte da atomi di B, come mostrato nella parte sinistra di figura 63. Il semiconduttore ternario $A_{(x)}B_{(1-x)}$ deve essere cresciuto in modalità "random alloy" ("composto casuale"), ovvero la matrice cristallina di $A_{(x)}B_{(1-x)}$ deve essere formata da un'alternanza casuale di atomi di A e di atomi di B, come mostrato nella parte destra di figura 63.

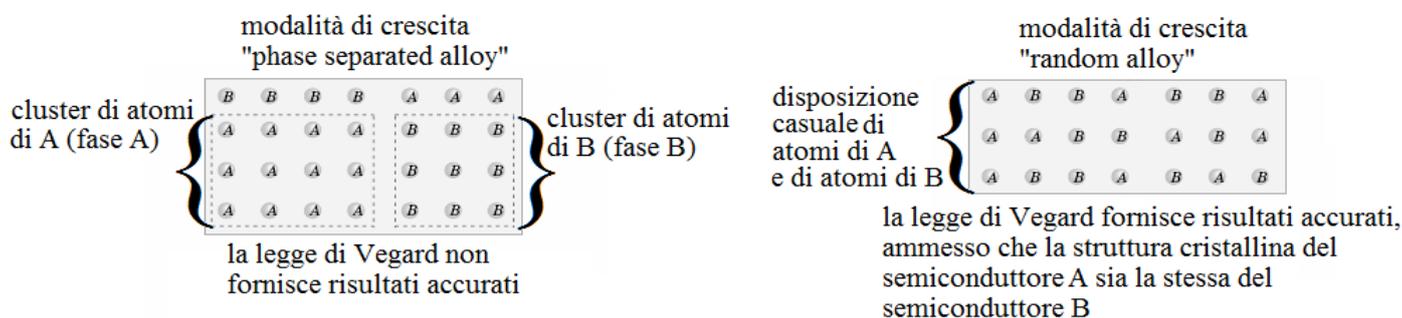


Fig. 63

Schema qualitativo riassuntivo delle due principali modalità di crescita epitassiale di un semiconduttore composto $A_{(x)}B_{(1-x)}$, dove A e B sono due semiconduttori semplici o anch'essi composti che vengono, in modo controllato (controllando la frazione molare x di A), sublimati e depositi su un substrato, all'interno dell'ambiente di lavoro del reattore per MBE. Sulla sinistra è mostrata la matrice cristallina nel caso di crescita "phase separated", nella quale si alternano regioni ricche di atomi di A a regioni ricche di atomi di B. In tal caso la legge di Vegard non fornisce risultati accurati. Sulla destra è mostrata la matrice cristallina nel caso di crescita "random", nella quale si ha la disposizione casuale degli atomi di A e di B. Se il semiconduttore composto $A_{(x)}B_{(1-x)}$ ha una struttura random e la geometria cristallina del semiconduttore A è uguale a quella del semiconduttore B (entrambi strutturati a zinoblenda, come i composti III - V o quelli II - VI, oppure entrambi a diamante, come C, Si e Ge oppure entrambi strutturati secondo la geometria HCP - "Hexagonal Close Pack", come i nitrucci InN, GaN, AlN e BN), la legge di Vegard fornisce risultati accurati.

Utilizzando la legge di Vegard è stato realizzato il grafico mostrato in figura 64: sull'asse delle ordinate è riportata la costante reticolare di alcuni dei più importanti semiconduttori binari III - V, largamente utilizzati in fotonica, sull'asse delle ascisse è riportata la frazione molare x , mentre le linee continue rappresentano la dipendenza da x della costante reticolare dei semiconduttori ternari ottenuti crescendo i binari in modo controllato; infatti ciascuna linea continua congiunge i valori della costante reticolare di due semiconduttori binari diversi.

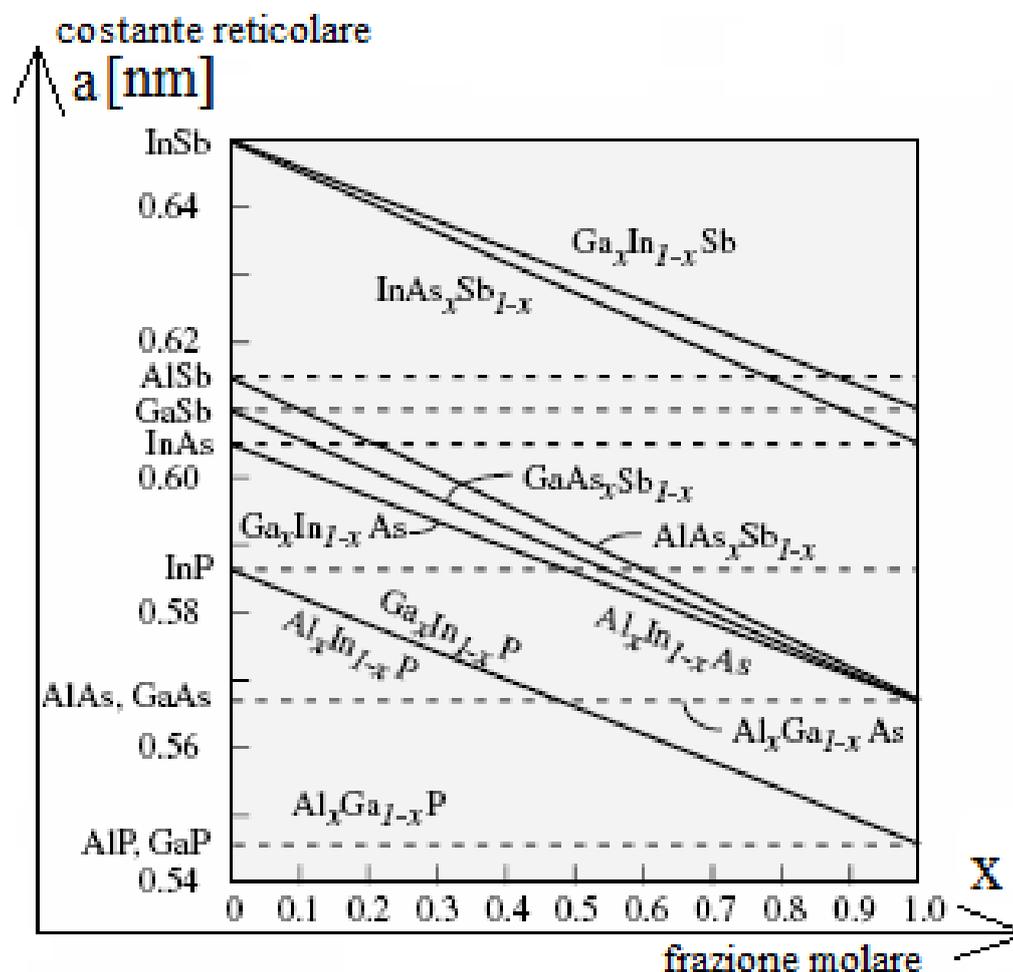


Fig. 64

Grafico realizzato utilizzando la legge di Vegard. Sull'asse delle ordinate è riportata la costante reticolare di alcuni dei più importanti semiconduttori binari III - V, largamente utilizzati in fotonica, sull'asse delle ascisse è riportata la frazione molare x , mentre le linee continue rappresentano la dipendenza da x della costante reticolare di alcuni semiconduttori ternari, ottenuti crescendo quelli binari in modo controllato; ciascuna linea continua congiunge i valori della costante reticolare di due semiconduttori binari diversi.

Discutiamo adesso un modello analitico che consente, partendo dalla legge di Vegard finora esaminata, nota più precisamente come “legge di Vegard per la costante reticolare”, di fittare ed interpolare il diagramma a bande di energia $E(k)$ di qualunque semiconduttore composto $A_{(x)}B_{(1-x)}$ conoscendo unicamente il diagramma a bande $E(k)$ del semiconduttore A e quello del semiconduttore B: la crescita random di $A_{(x)}B_{(1-x)}$ e le due identiche geometrie cristalline dei semiconduttori A e B consentono, per ipotesi, l'applicazione delle legge di Vegard. Inoltre tale modello consente di ricavare delle semplici relazioni “Vegard - like”, con le quali è possibile calcolare alcune grandezze elettriche fondamentali relative al semiconduttore $A_{(x)}B_{(1-x)}$, come le masse efficaci di elettroni e lacune, m_e^* ed m_h^* , definite presso i “bandedges”, per semiconduttori A e B sia a gap diretto che a gap indiretto, oppure come l'energia di gap E_{gap} nel caso in cui i semiconduttori A e B siano a gap diretto.

Consideriamo un generico semiconduttore composto $A_{(x)}B_{(1-x)}$. Per $x = 0$ abbiamo un cristallo costituito, di fatto, unicamente da atomi del semiconduttore B. Supponiamo che in questa matrice cristallina ci sia anche qualche

atomo del semiconduttore A: la probabilità di trovare, intorno ad un fissato atomo A, altri atomi A è trascurabile, cioè proporzionale ad $x = 0$. Quindi l'energia potenziale $U(r)$, prodotta dagli atomi del cristallo, coincide con quella $U_B(r)$ dovuta unicamente agli atomi B, con un periodo spaziale Δr_B ed un'altezza massima delle barriere U_B (= profondità massima delle buche) che dipendono dalla natura del semiconduttore B. In altre parole un qualunque elettrone di valenza, che è "libero" (delocalizzato), cioè è descrivibile mediante un'onda di Bloch, dal momento che la periodicità del cristallo è perfetta, risente con probabilità nulla del potenziale periodico $U_A(r)$ dovuto agli atomi del semiconduttore A e con probabilità unitaria del potenziale periodico $U_B(r)$. Viceversa per $x = 1$ abbiamo un cristallo costituito, di fatto, unicamente da atomi del semiconduttore A. Supponiamo che in questa matrice cristallina ci sia anche qualche atomo del semiconduttore B: la probabilità di trovare, attorno ad un fissato atomo A, altri atomi A è quasi unitaria, cioè proporzionale ad $x = 1$, mentre la probabilità di trovare, attorno ad un fissato atomo A, atomi del semiconduttore B è quasi nulla, cioè proporzionale ad $1 - x = 0$. Quindi l'energia potenziale $U(r)$, prodotta dagli atomi del cristallo, è quella $U_A(r)$ dovuta unicamente agli atomi A, con un periodo spaziale Δr_A ed un'altezza massima delle barriere U_A (= profondità massima delle buche) che dipendono dalla natura del semiconduttore A. In altre parole un qualunque elettrone atomico di valenza, che è "libero" (delocalizzato), cioè è descrivibile mediante un'onda di Bloch, dal momento che la periodicità del cristallo è perfetta, risente con probabilità nulla del potenziale periodico $U_B(r)$ dovuto agli atomi del semiconduttore B e con probabilità unitaria del potenziale periodico $U_A(r)$. Per semplicità espositiva e di calcolo supponiamo che Δr_B sia uguale, almeno approssimativamente, a Δr_A . I disegni mostrati in figura 65 rappresentano qualitativamente e monodimensionalmente le due situazioni appena descritte.

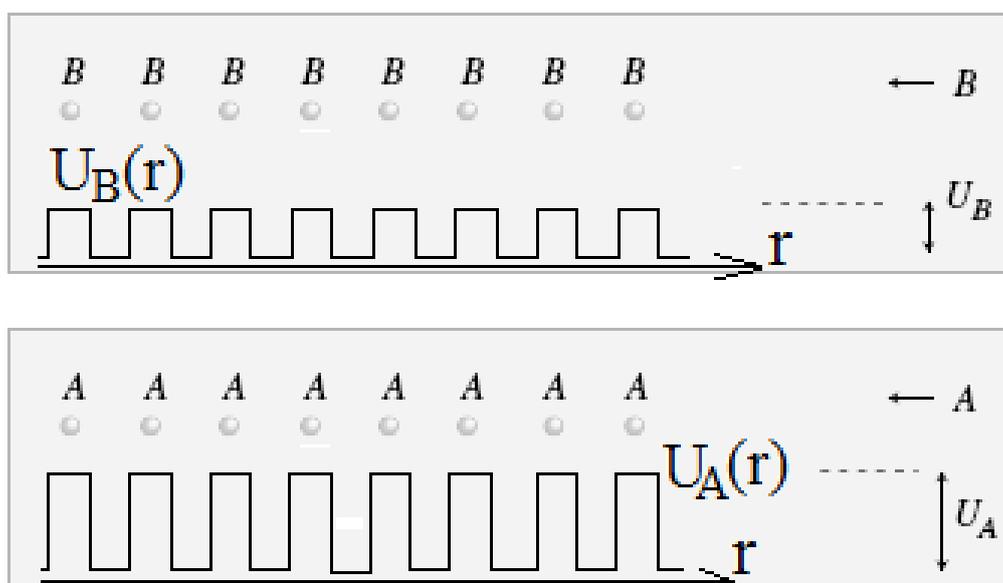


Fig. 65

Rappresentazione qualitativa e monodimensionale del potenziale coulombiano elettrostatico $U(r)$ di cui risente un elettrone "libero" di valenza (descrivibile con un'onda di Bloch) nel caso in cui il semiconduttore composto $A_{(x)}B_{(1-x)}$ sia cresciuto con la frazione molare x di A pari a zero oppure pari ad uno: abbiamo ipotizzato, per semplicità, che il periodo spaziale Δr_B sia uguale, almeno approssimativamente, al periodo spaziale Δr_A . La differenza fra le ampiezze U_B ed U_A dei due potenziali è, invece, non trascurabile.

Ammettiamo adesso di crescere epitassialmente, in modalità "random alloy", un layer di semiconduttore composto $A_{(x)}B_{(1-x)}$, con x compreso fra 0 ed 1. Gli atomi del semiconduttore B e quelli del semiconduttore A si alternano casualmente, pertanto un elettrone atomico di valenza del cristallo di $A_{(x)}B_{(1-x)}$ risente di un potenziale $U_{\text{Alloy}}(r)$ che dipende sia dal potenziale $U_B(r)$ che da quello $U_A(r)$. Tuttavia la casualità della disposizione degli atomi B ed A comporta la casualità spaziale con la quale $U_B(r)$ ed $U_A(r)$ contribuiscono a realizzare il potenziale coulombiano risultante totale $U(r)$, effettivamente visto dagli elettroni atomici di valenza. Questi ultimi risentono di un potenziale risultante totale $U(r)$ non periodico, spazialmente casuale, di forma (spaziale) assolutamente stocastica, pertanto gli elettroni non potrebbero, a rigore, essere trattati come onde di Bloch. Il teorema di Bloch non è valido in $A_{(x)}B_{(1-x)}$ cresciuto in modo random e ciascun elettrone è descritto da un pacchetto ψ_{Alloy} di onde spazialmente localizzato.

$$\psi_{\text{Alloy}}(\mathbf{r}, \mathbf{k}) = \iiint \chi_{\text{Alloy}}(\mathbf{r}, \mathbf{k}) \varphi(\mathbf{k}) e^{i(\mathbf{k} \cdot \mathbf{r})} d^3 \mathbf{k}$$

dove \mathbf{r} è il vettore posizione nello spazio reale di Bravais, $\mathbf{k} = (k_x, k_y, k_z)$ è il vettore d'onda nello spazio di Fourier, $\varphi(\mathbf{k})$ è la "funzione peso" di ciascuna armonica (oscillazione) spaziale propagante $\exp\{i(\mathbf{k} \cdot \mathbf{r})\}$, mentre $\chi_{\text{Alloy}}(\mathbf{r}, \mathbf{k})$ è la funzione involuppo di ciascuna oscillazione spaziale $\exp\{i(\mathbf{k} \cdot \mathbf{r})\}$. Va puntualizzato che il semiconduttore composto $A_{(x)}B_{(1-x)}$, cresciuto in modo random, a rigore possiede una struttura energetica $E(\mathbf{k})$ molto più complessa e disordinata del regolare diagramma a bande $E - \mathbf{k}$ tipico di un cristallo semiconduttore ordinato, perfettamente periodico e privo di difetti, come il semiconduttore B ed il semiconduttore A. Quindi non è formalmente corretto parlare di BV e di BC per il diagramma a bande $E(\mathbf{k})$ di $A_{(x)}B_{(1-x)}$ e associare alla $\chi_{\text{Alloy}}(\mathbf{r}, \mathbf{k})$ una simmetria di tipo p se ψ_{Alloy} descrive un elettrone in BV e allo stesso modo associare alla $\chi_{\text{Alloy}}(\mathbf{r}, \mathbf{k})$ una simmetria di tipo s, o di tipo $s + p$, se ψ_{Alloy} descrive un elettrone in BC in un semiconduttore $A_{(x)}B_{(1-x)}$ a gap diretto oppure a gap indiretto. Nel caso di $A_{(x)}B_{(1-x)}$ random la funzione spaziale $\chi_{\text{Alloy}}(\mathbf{r}, \mathbf{k})$ ha una simmetria più complessa. La situazione di casualità in cui verte l'energia potenziale $U(\mathbf{r})$, della quale risente un qualunque elettrone $\psi_{\text{Alloy}}(\mathbf{r}, \mathbf{k})$, è mostrata qualitativamente nella parte alta di figura 66.

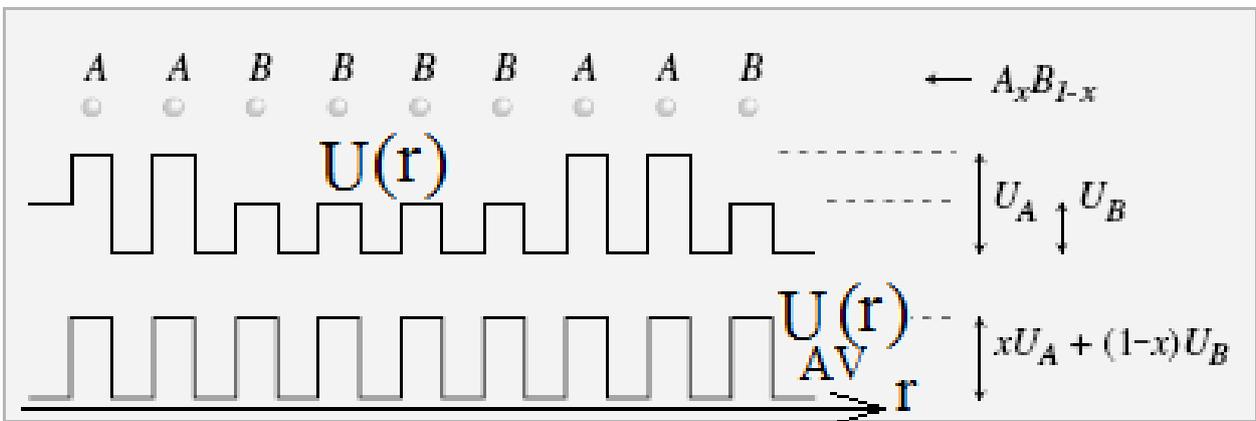


Fig. 66

Nella parte superiore è rappresentato, qualitativamente e monodimensionalmente, il potenziale elettrostatico $U(\mathbf{r})$ di cui risente un qualunque elettrone atomico di valenza, che è descritto da un pacchetto $\psi_{\text{Alloy}}(\mathbf{r}, \mathbf{k})$ di onde localizzate, nel caso in cui il semiconduttore composto $A_{(x)}B_{(1-x)}$ sia cresciuto in modo random, cioè con atomi A casualmente alternati ad atomi B. Abbiamo ipotizzato per semplicità che il periodo spaziale Δr_B sia uguale, almeno approssimativamente, al periodo spaziale Δr_A . La differenza fra le ampiezze U_B ed U_A dei due potenziali è, invece, non trascurabile. Nella parte inferiore è rappresentato, sempre qualitativamente e monodimensionalmente, il potenziale elettrostatico "di cristallo virtuale", perfettamente periodico, $U_{AV}(\mathbf{r})$, di cui risente un qualunque elettrone di valenza, il quale è descrivibile da un'onda di Bloch. $U_{AV}(\mathbf{r})$ è ottenuto da una media, pesata con le frazioni molari x ed $1 - x$, fra $U_A(\mathbf{r})$ e $U_B(\mathbf{r})$.

Riferendoci sempre, a titolo di esempio, alla parte alta di figura 66, abbiamo che all'elettrone atomico di valenza descritto dal pacchetto d'onde $\psi_{\text{Alloy}}(\mathbf{r}, \mathbf{k})$ è associata una densità di probabilità $|\psi_{\text{Alloy}}(\mathbf{r}, \mathbf{k})|^2$ prevalentemente localizzata nella regione del cristallo $A_{(x)}B_{(1-x)}$ individuata dal gruppo dei quattro atomi del semiconduttore B; l'altezza di potenziale U_B , essendo minore dell'altezza di potenziale U_A , determina in tale regione una sorta di buca di potenziale di profondità pari a $U_A - U_B$, nella quale è più probabile trovare l'elettrone di valenza descritto da $\psi_{\text{Alloy}}(\mathbf{r}, \mathbf{k})$. Tuttavia, come mostrato nella parte bassa di figura 66, il suddetto elettrone può essere visto anche come un semplice elettrone "libero", descritto da un'onda di Bloch, che risente non di un potenziale elettrostatico di cristallo random $U(\mathbf{r})$, bensì di un potenziale $U_{AV}(\mathbf{r})$ perfettamente periodico, per il quale è valido il teorema di Bloch, risultante da una media fra il potenziale $U_B(\mathbf{r})$ dovuto agli atomi B ed il potenziale $U_A(\mathbf{r})$ dovuto agli atomi A: la media è pesata dalla frazione molare x e dalla sua complementare $1 - x$, che rappresentano rispettivamente la probabilità di trovare, nel cristallo di $A_{(x)}B_{(1-x)}$, atomi del semiconduttore A ed atomi del semiconduttore B, ovvero di trovare contributi di potenziale dovuti agli atomi A e contributi di potenziale dovuti agli atomi B. In formule:

$$U_{AV}(\mathbf{r}) = xU_A(\mathbf{r}) + (1 - x)U_B(\mathbf{r})$$

$U_{AV}(\mathbf{r})$ è il potenziale, perfettamente periodico, che interesserebbe un ipotetico cristallo (un cristallo virtuale, non realmente esistente) il cui campo elettrico interno, cioè il campo elettrico prodotto dai suoi atomi perfettamente ordinati, produce sugli elettroni atomici un effetto pari alla media pesata (con x ed $1 - x$) fra l'effetto dovuto al campo elettrico prodotto dagli atomi A e l'effetto dovuto al campo elettrico prodotto dagli atomi B. Il modello nella parte bassa di figura 66 prende pertanto il nome di "modello di cristallo virtuale". Grazie a questa semplificazione è consentito descrivere il comportamento di un qualunque elettrone atomico di valenza mediante una semplice onda di Bloch, come previsto nei sistemi cristallini perfetti. Pertanto:

$$\psi_{\text{Alloy}}(\mathbf{r}, \mathbf{k}) \approx u_{AV}(\mathbf{r}, \mathbf{k}) e^{i(\mathbf{k} \cdot \mathbf{r})}$$

dove $u_{AV}(\mathbf{r}, \mathbf{k})$ è una "central cell function" di simmetria p oppure di simmetria s (oppure $s + p$), a seconda che l'elettrone descritto da $\psi_{\text{Alloy}}(\mathbf{r}, \mathbf{k})$ sia posizionato presso il massimo della BV oppure presso il minimo della BC di un semiconduttore a gap diretto (o a gap indiretto). Riprendendo il concetto sintetizzato nella [e10] possiamo risolvere l'equazione di Schrödinger per il semiconduttore A:

$$\hat{H}_{\text{cristalloA}} |\psi_{A\mathbf{k}}\rangle = E_A(\mathbf{k}) |\psi_{A\mathbf{k}}\rangle$$

$$\hat{H}_{\text{cristalloA}} = \hat{H}_{\text{atomico}} + \widehat{\Delta U}_A$$

dove $\widehat{\Delta U}_A$ è un termine potenziale aggiuntivo, discriminante fra problema atomico e problema cristallino, che tiene conto dell'interazione coulombiana cristallina, nel semiconduttore A, responsabile della degenerazione di un livello energetico discreto in una banda. Dall'equazione di Schrödinger di cui sopra ricaviamo, attraverso il calcolo del determinante secolare (si veda il glossario), il diagramma a bande $E_A(\mathbf{k})$ del semiconduttore A. Possiamo altresì risolvere l'equazione di Schrödinger per il semiconduttore B:

$$\hat{H}_{\text{cristalloB}} |\psi_{B\mathbf{k}}\rangle = E_B(\mathbf{k}) |\psi_{B\mathbf{k}}\rangle$$

$$\hat{H}_{\text{cristalloB}} = \hat{H}_{\text{atomico}} + \widehat{\Delta U}_B$$

dove $\widehat{\Delta U}_B$ tiene conto dell'interazione coulombiana cristallina, all'interno del semiconduttore B, responsabile della degenerazione di un livello energetico discreto in una banda. Dall'ultima equazione di Schrödinger possiamo ricavare, sempre attraverso il calcolo del determinante secolare, il diagramma a bande $E_B(\mathbf{k})$ del semiconduttore B. Per calcolare, infine, il diagramma a bande $E_{\text{Alloy}}(\mathbf{k}) = E_{A(x)B(1-x)}(\mathbf{k})$ del semiconduttore composto random $A_{(x)}B_{(1-x)}$, dovremmo risolvere il determinante secolare derivante dalla seguente equazione di Schrödinger:

$$\hat{H}_{\text{Alloy}} |\psi_{\text{Alloy}\mathbf{k}}\rangle = E_{\text{Alloy}}(\mathbf{k}) |\psi_{\text{Alloy}\mathbf{k}}\rangle$$

$$\hat{H}_{\text{cristalloA}_{(x)}B_{(1-x)}} |\psi_{AB\mathbf{k}}\rangle = E_{A_{(x)}B_{(1-x)}}(\mathbf{k}) |\psi_{AB\mathbf{k}}\rangle$$

$$\hat{H}_{\text{cristalloA}_{(x)}B_{(1-x)}} = \hat{H}_{\text{atomico}} + \widehat{\Delta U}_{AV}$$

dove $\widehat{\Delta U}_{AV}$, in conseguenza del modello di cristallo virtuale, è definito nel modo seguente:

$$\widehat{\Delta U}_{AV} = x \widehat{\Delta U}_A + (1 - x) \widehat{\Delta U}_B$$

Da qui si evince che il diagramma a bande $E_{\text{Alloy}}(\mathbf{k}) = E_{A(x)B(1-x)}(\mathbf{k})$ del semiconduttore composto $A_{(x)}B_{(1-x)}$ può essere ottenuto, almeno in una forma approssimata, semplicemente dalla media pesata (sempre con x ed $1 - x$), valutata per ciascun valore del vettore d'onda $\mathbf{k} = (k_x, k_y, k_z)$ del reticolo reciproco, fra il diagramma a bande $E_A(\mathbf{k})$ del semiconduttore A ed il diagramma a bande $E_B(\mathbf{k})$ del semiconduttore B. In formule:

$$E_{\text{Alloy}}(\mathbf{k}) = E_{A_{(x)}B_{(1-x)}}(\mathbf{k}) = x E_A(\mathbf{k}) + (1 - x) E_B(\mathbf{k})$$

Quest'ultima relazione comporta che nel caso in cui A e B siano due semiconduttori a gap diretto vale la seguente espressione, nota come "legge di Vegard per l'energia di gap diretto":

$$E_{\text{gap}\Gamma\text{Alloy}}(k) = E_{\text{gap}\Gamma A(x)B(1-x)}(k) = x E_{\text{gap}\Gamma A}(k) + (1-x) E_{\text{gap}\Gamma B}(k)$$

Per due qualsiasi semiconduttori A e B, sia a gap diretto che a gap indiretto (possono essere entrambi a gap diretto oppure entrambi a gap indiretto oppure uno a gap diretto e l'altro a gap indiretto), vale la seguente espressione, nota come "legge di Vegard per le masse efficaci degli elettroni e delle lacune", dove gli elettroni si assumono localizzati presso il minimo della BC e le lacune presso il massimo della BV:

$$\frac{1}{m_{e\text{Alloy}}^*} = \frac{1}{m_{eA(x)B(1-x)}^*} = x \frac{1}{m_{eA}^*} + (1-x) \frac{1}{m_{eB}^*}$$

$$\frac{1}{m_{h\text{Alloy}}^*} = \frac{1}{m_{hA(x)B(1-x)}^*} = x \frac{1}{m_{hA}^*} + (1-x) \frac{1}{m_{hB}^*}$$

Considerando il semiconduttore a gap diretto B = GaAs ed il semiconduttore a gap indiretto A = AlAs, possiamo applicare il procedimento fin qui descritto e calcolare il diagramma a bande $E_{\text{Alloy}}(k) = E_{\text{Al}(x)\text{Ga}(1-x)\text{As}}(k)$ presso qualunque valore quantizzato di k e per qualunque valore di x compreso fra 0 ed 1, ottenendo il grafico, mostrato in figura 67, dell'energia delle valli X, Γ ed L in funzione della frazione molare x di alluminio. Si noti che per $x < 0.4$ l'energia della valle Γ è inferiore a quella delle altre due valli, pertanto il semiconduttore ternario $\text{Al}(x)\text{Ga}(1-x)\text{As}$ è a gap diretto, mentre per $x > 0.4$ l'energia della valle Γ supera prima l'energia della valle X e poi quella della valle L. Per $x > 0.4$ l'energia della valle X è inferiore a quella delle altre due valli, pertanto il semiconduttore ternario $\text{Al}(x)\text{Ga}(1-x)\text{As}$ è a gap indiretto.

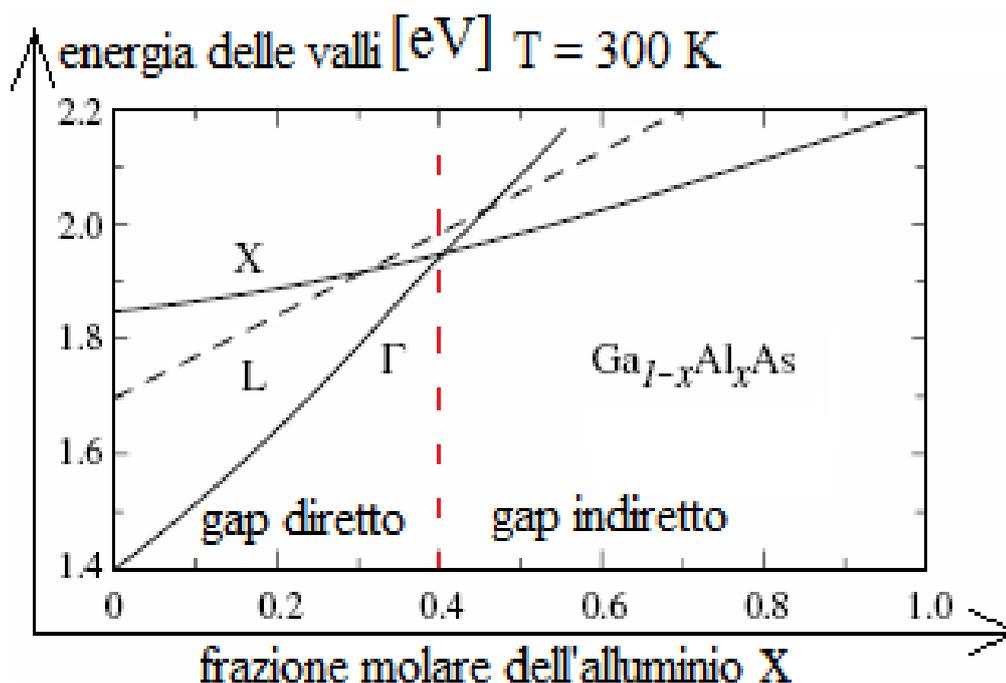


Fig. 67

Grafico dell'energia delle valli X, Γ ed L in funzione della frazione molare x di alluminio, riscontrabili all'interno del diagramma a bande del semiconduttore ternario $\text{Al}(x)\text{Ga}(1-x)\text{As}$. Si noti che per $x < 0.4$ l'energia della valle Γ è inferiore a quella delle altre due valli, pertanto il $\text{Al}(x)\text{Ga}(1-x)\text{As}$ è a gap diretto, mentre per $x > 0.4$ l'energia della valle Γ supera prima l'energia della valle X e poi quella della valle L. Quindi per $x > 0.4$ l'energia della valle X è inferiore a quella delle altre due valli, pertanto il $\text{Al}(x)\text{Ga}(1-x)\text{As}$ è a gap indiretto.

Considerando invece il semiconduttore a gap diretto A = InAs ed il semiconduttore, sempre a gap diretto, B = GaAs, possiamo applicare il solito procedimento e calcolare il diagramma a bande $E_{\text{Alloy}}(k) = E_{\text{In}(x)\text{Ga}(1-x)\text{As}}(k)$ presso qualunque valore quantizzato di k e per qualunque valore di x compreso fra 0 ed 1, ottenendo il seguente grafico dell'energia delle valli X, Γ ed L, nonché della massa efficace dell'elettrone ($m_e^* = m_\sigma^* = m_{ed}^*$, il gap è

diretto $\forall x \in [0, 1]$ normalizzata rispetto alla massa m_0 dell'elettrone libero, in funzione della frazione molare x di indio. Si noti che $\forall x \in [0, 1]$ l'energia della valle Γ è inferiore a quella delle altre due valli, pertanto il semiconduttore ternario $\text{In}_{(x)}\text{Ga}_{(1-x)}\text{As}$ è a gap diretto $\forall x$. Ciò si può evincere anche dal grafico di figura 58, in cui i semiconduttori binari GaAs e InAs sono collegati fra loro da una linea rossa, che indica appunto che il ternario rappresentato è a gap diretto. Si noti altresì l'andamento decrescente di m_e^* all'aumentare di x : la massa efficace dell'elettrone nel InAs è molto piccola ($0.027m_0$), ovvero la concavità della BC del InAs presso il minimo E_C è molto accentuata (grande "warping" della BC, l'elettrone di conduzione è molto leggero). Ciò conferisce al InAs un'elevata mobilità elettronica μ_n : valori tipici possono essere di circa $20000 - 25000 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$.

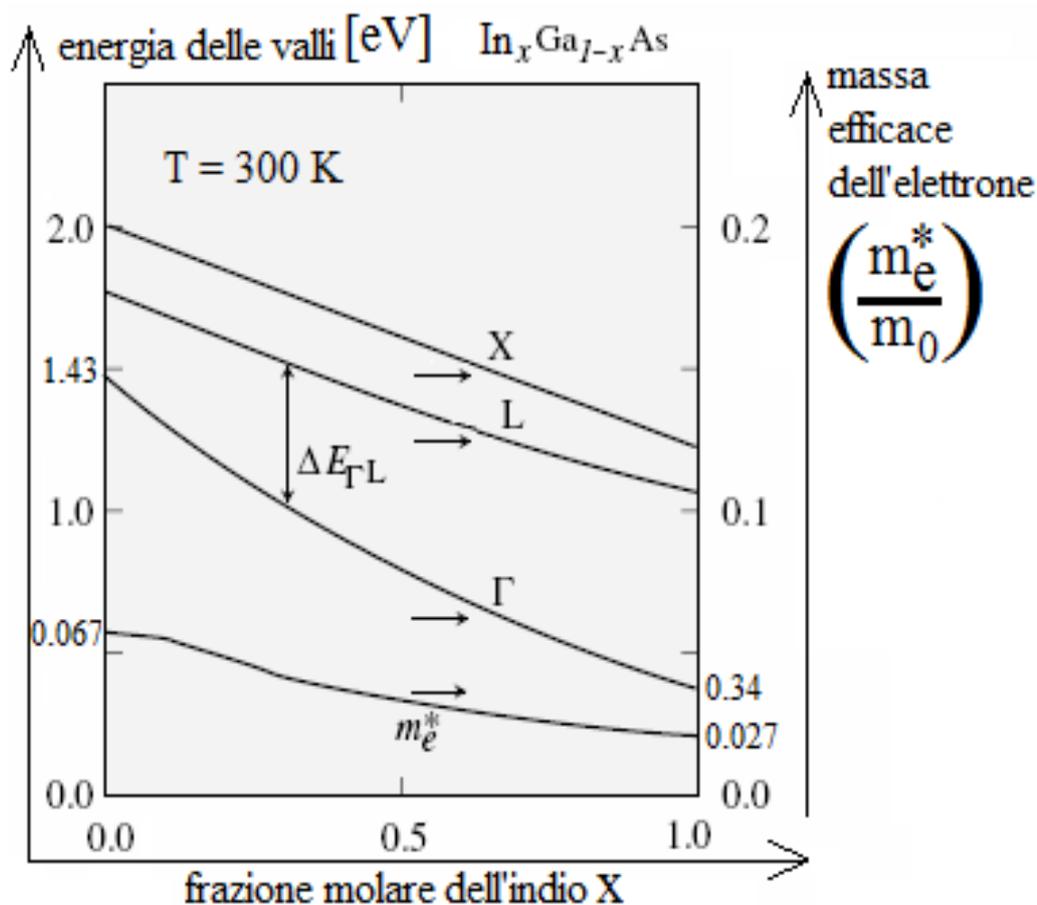


Fig. 68

Grafico dell'energia delle valli X, Γ ed L, nonché della massa efficace dell'elettrone ($m_e^* = m_\sigma^* = m_{ed}^*$, il gap è diretto $\forall x \in [0, 1]$) normalizzata rispetto alla massa dell'elettrone libero, in funzione della frazione molare x di indio del semiconduttore ternario $\text{In}_{(x)}\text{Ga}_{(1-x)}\text{As}$. Si noti che $\forall x \in [0, 1]$ l'energia della valle Γ è inferiore a quella delle altre due valli, pertanto il $\text{In}_{(x)}\text{Ga}_{(1-x)}\text{As}$ è a gap diretto $\forall x$. Si noti altresì l'andamento decrescente di m_e^* all'aumentare di x : la massa efficace dell'elettrone nel InAs è molto piccola ($0.027m_0$), ovvero la concavità della BC del InAs presso il minimo E_C è molto accentuata (grande "warping" della BC, l'elettrone di conduzione è molto leggero). Ciò conferisce al InAs un'elevata mobilità elettronica μ_n : valori tipici possono essere di circa $20000 - 25000 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$.

5.9.3) Il grafico della compatibilità reticolare fra semiconduttori: esempio di progettazione del PIN di terza finestra $\text{InP}/\text{In}_{(0.57)}\text{Ga}_{(0.43)}\text{As}/\text{InP}$ a struttura "MESA"

Per costruire un fotodiodo PIN capace di rivelare, in generale, luce infrarossa approssimativamente nel range $1.3 \div 1.6 \mu\text{m}$, e più precisamente (come nel nostro caso) nella terza finestra di attenuazione (centrata presso $\lambda \approx 1.55 \mu\text{m}$) delle fibre ottiche realizzate in silice (SiO_2), dobbiamo prendere in considerazione il quadrilatero (ciclo) GaP - GaAs - InAs - InP, individuabile nel grafico di figura 58, del quale proponiamo un ingrandimento nella figura 69. Osservando il ciclo ci rendiamo conto che il InP è trasparente nei confronti della radiazione infrarossa incidente di terza finestra, dal momento che il suo λ_{gap} è inferiore alla lunghezza d'onda λ del segnale luminoso; pertanto il suo E_{gap} ($\approx 1.35 \text{ eV}$) è maggiore dell'energia fotonica incidente $h\nu \approx 0.8 \text{ eV}$, per cui nessun fotone

incidente potrà essere assorbito dal InP, se trascuriamo le eventuali risonanze eccitoniche “sub - gap”. Quindi il InP è ideale per realizzare i layers p ed n. Per scegliere il semiconduttore con cui realizzare lo strato intrinseco, cioè il layer di assorbimento ottico, è opportuno seguire la linea verticale (a passo reticolare costante) passante per il InP, la quale passa nelle vicinanze di tutti i materiali “compatibili”, a livello di matching reticolare, con quest’ultimo. La scelta, come si è già osservato nel paragrafo precedente applicando la legge di Vegard per la costante reticolare, può cadere sul semiconduttore ternario $\text{In}_{(0.57)}\text{Ga}_{(0.43)}\text{As}$, cresciuto per epitassia o mediante CVD con l’aggiunta di una piccola concentrazione (frazione molare) di fosforo P, poiché il punto rosso in figura 69, che rappresenta il $\text{In}_{(0.57)}\text{Ga}_{(0.43)}\text{As}$, è posizionato leggermente al di sotto della linea che rappresenta il centro della terza finestra. Il ternario $\text{In}_{(0.57)}\text{Ga}_{(0.43)}\text{As}$, steso opportunamente su un layer di InP, non è interessato da stress meccanici rilevanti ($\varepsilon \approx 0$) presso l’interfaccia con il InP, dato che la costante reticolare del InP è approssimativamente uguale a quella del $\text{In}_{(0.57)}\text{Ga}_{(0.43)}\text{As}$ ($\approx 5.86 \text{ \AA}$). Per giunta il $\text{In}_{(0.57)}\text{Ga}_{(0.43)}\text{As}$ ha una $\lambda_{\text{gap}} > 1.55 \mu\text{m}$, ovvero una $E_{\text{gap}} \approx 0.75 \text{ eV} < h\nu \approx 0.8 \text{ eV}$, per cui tutti i fotoni della luce incidente possono essere assorbiti all’interno di $\text{In}_{(0.57)}\text{Ga}_{(0.43)}\text{As}$; in questo materiale l’assorbimento è inoltre facilitato, per quelle frequenze, da un alto valore di α , anche in virtù dal gap diretto. Le metallizzazioni esterne di contatto, del PIN finora descritto, sono generalmente realizzate in oro: l’oro, oltre ad essere un buon conduttore, aderisce molto bene al InP.

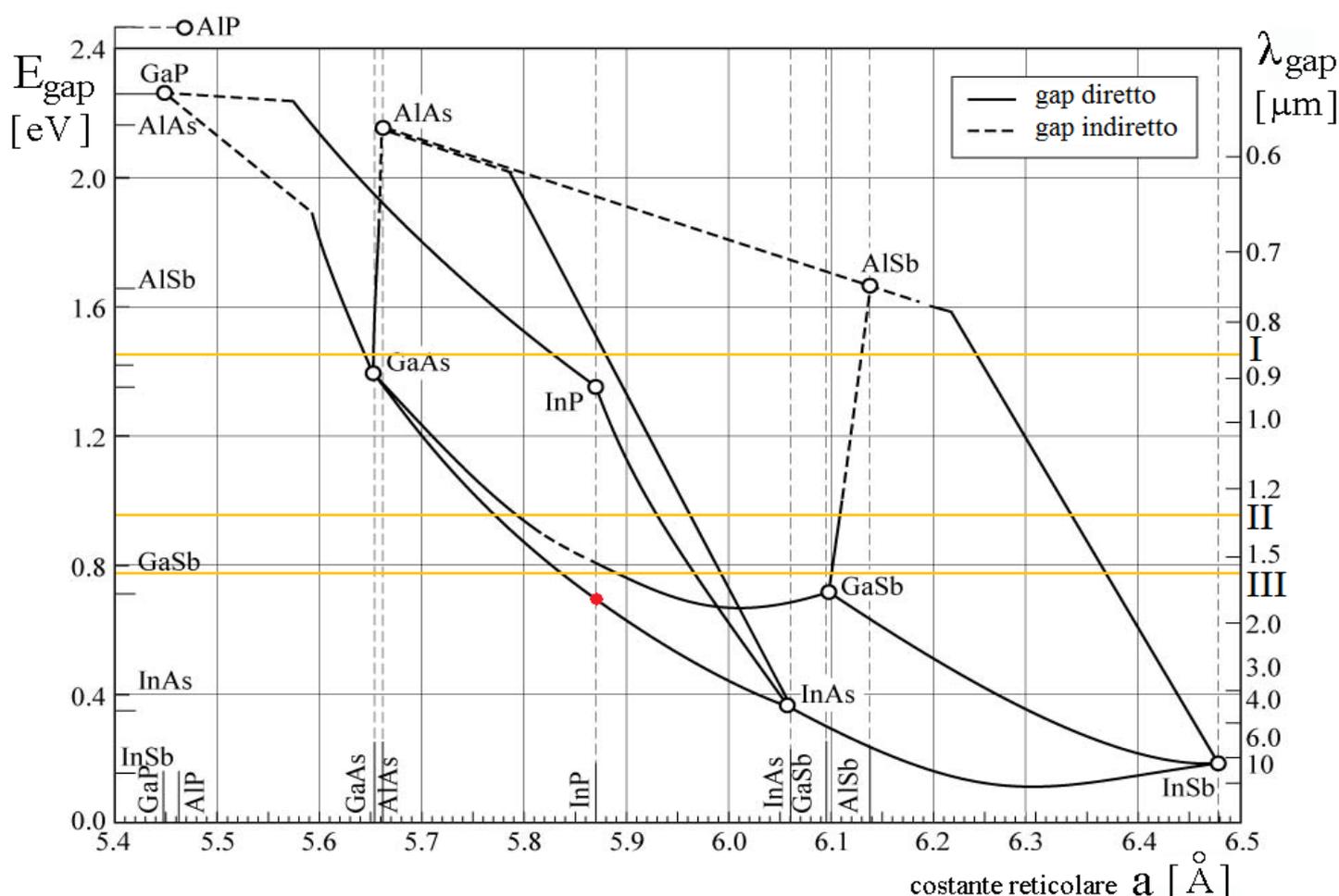


Fig. 69

Ingrandimento di una porzione particolare del grafico della compatibilità reticolare fra semiconduttori, relativa al ciclo InGaAsP. Si notino le lunghezze d’onda centrali delle “tre finestre” (I, II, III) utilizzate nelle trasmissioni su fibra ottica di silice, ovvero le lunghezze d’onda presso le quali l’ossido di silicio (SiO_2), con cui sono usualmente costruite le fibre destinate alle telecomunicazioni, presenta i più bassi valori del coefficiente di assorbimento ottico $\alpha(\lambda)$. Le linee tratteggiate in grigio fra GaAs ed AlAs e fra $\text{In}_{(0.57)}\text{Ga}_{(0.43)}\text{As}$ e InP sono state disegnate perfettamente verticali per semplicità grafica. In realtà queste linee presentano delle leggere curvature, dovute ai lievi mismatches reticolari esistenti fra GaAs ed AlAs e fra $\text{In}_{(0.57)}\text{Ga}_{(0.43)}\text{As}$ e InP. Ad esempio la costante reticolare dell’AlAs è circa pari a 5.660 \AA , mentre quella dell’GaAs è circa pari a 5.653 \AA , per cui il mismatch reticolare è circa dello 0.13%. Tali mismatches potrebbero risultare problematici ai fini della realizzazione dei fotorivelatori e del loro corretto funzionamento: si verificano infatti alcune alterazioni delle proprietà ottiche dei semiconduttori con i quali sono realizzati i fotorivelatori. I mismatches, tuttavia, possono essere sfruttati per realizzare dei buoni dispositivi emettitori di luce: ad esempio opportuni stress meccanici sono, infatti, alla base della coerenza spaziale (direttività) che caratterizza il fascio luminoso emesso dai LASERS MQW (“Multi Quantum Wells”).

In figura 70 riportiamo il grafico semilogaritmico, ottenuto sperimentalmente, del coefficiente di assorbimento ottico specifico $\alpha(\lambda)$ [dB/Km] all'interno di una fibra di silice SiO_2 . Il fotorivelatore realizzato con l'eterostruttura $\text{InP}/\text{In}_{(0.57)}\text{Ga}_{(0.43)}\text{As}/\text{InP}$ è connesso ad un capo della fibra ottica, mentre all'altro è connesso il LED o il LASER che, opportunamente pilotato in modo diretto o meglio ancora in modo indiretto (ad esempio mediante un transistor MESFET all'arseniuro di gallio, oppure con l'ausilio di un SOA – “Semiconductor Optical Amplifier” – “amplificatore ottico a semiconduttore”, oppure ancora per mezzo di un circuito costituito da una coppia differenziale di transistori BJT), emette il segnale di terza finestra modulato (alcune delle modulazioni più importanti sono OOK, ASK, PSK, FSK, PolSK ecc...) che il PIN deve rivelare e convertire in un segnale elettrico da inviare a sua volta in ingresso ad un demodulatore.

Lo spettro risultante di $\alpha(\lambda)$, cioè quello evidenziato nel grafico con il tratto spesso, è dato dalla sovrapposizione di diverse sorgenti di assorbimento di fotoni, ciascuna delle quali contribuisce al depauperamento dell'intensità luminosa che viaggia attraverso la guida ottica. I radicali OH^- e gli ioni Cu^{++} sono residui del processo di fabbricazione delle fibre (di processi ce ne sono diversi, ad esempio la OVD – “Outside Vapor Deposition”, la VAD – “Vapor Axial Deposition”, la MCVD – “Modified Chemical Vapor Deposition”, ecc...). La “dispersione (scattering) Rayleigh” consiste nella dispersione dell'energia ottica associata ad un modo propagativo principale (se la fibra è multimodale, come ad esempio le fibre ottiche “di accesso”, ci sono molti modi propaganti, ed il loro numero è tanto più alto quanto maggiore è la differenza fra l'indice di rifrazione del core e l'indice di rifrazione del cladding) in tanti modi propagativi secondari, ciascuno di piccola energia, in seguito al passaggio del modo propagativo principale, all'interno del core della fibra, da una regione di silice, il cui indice di rifrazione è prossimo ad 1.5, ad un grumo di silicio, avente indice di rifrazione maggiore di 3. Se la lunghezza d'onda della luce è simile al diametro del grumo di silicio (i grumi sono dei precipitati che si formano in seguito al raffreddamento non omogeneo della fibra, alla fine della sua lavorazione), avviene uno sparpagliamento (scattering) energetico isotropico dei modi secondari. Una piccola parte di questi modi secondari forma con l'interfaccia core/cladding un angolo α minore dell'angolo critico α_{CR} , pertanto l'energia a loro associata non si disperde nel cladding, secondo la legge di Snell, e arriva fino al fotorivelatore (a meno di perdite particolari dovute, ad esempio, alle curvature della fibra lungo il percorso), mentre la maggior parte dei modi secondari forma con l'interfaccia core/cladding un angolo α maggiore dell'angolo critico α_{CR} , pertanto l'energia a loro associata si disperde nel cladding lungo un tragitto molto breve e non arriva al fotorivelatore.

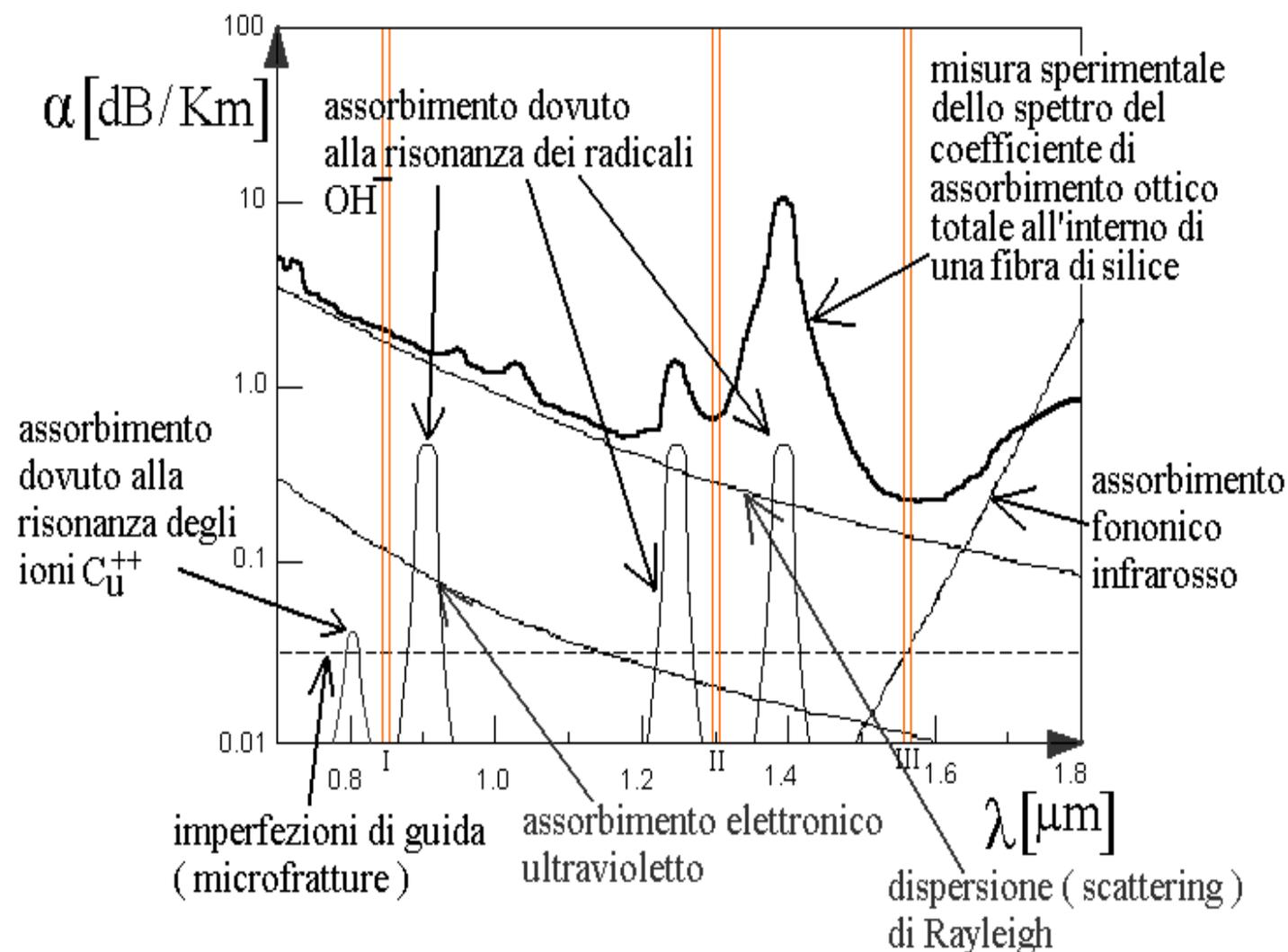


Fig. 70

Grafico semilogaritmico, ottenuto sperimentalmente, del coefficiente di assorbimento ottico specifico all'interno di una fibra di silice SiO₂. Lo spettro risultante, cioè quello evidenziato nel grafico con il tratto spesso, è dato dalla sovrapposizione di diverse sorgenti di assorbimento di fotoni, ciascuna delle quali contribuisce al depauperamento dell'intensità luminosa che viaggia attraverso la guida ottica. I radicali OH⁻ e gli ioni Cu⁺⁺ sono residui del processo di fabbricazione delle fibre (di processi ce ne sono diversi, ad esempio la OVD - "Outside Vapor Deposition", la VAD - "Vapor Axial Deposition", la MCVD - "Modified Chemical Vapor Deposition", ecc...). La "dispersione (scattering) Rayleigh" consiste nella dispersione dell'energia ottica, associata ad un modo propagativo principale, in tanti modi propagativi secondari, ciascuno di piccola energia, in seguito al passaggio del modo propagativo principale, all'interno del core della fibra, da una regione di silice, il cui indice di rifrazione è prossimo ad 1.5, ad un grumo di silicio, avente indice di rifrazione maggiore di 3. Se la lunghezza d'onda della luce è simile al diametro del grumo di silicio (i grumi sono dei precipitati che si formano in seguito al raffreddamento non omogeneo della fibra, alla fine della sua lavorazione), avviene lo sparpagliamento (scattering) energetico isotropico dei modi secondari, con conseguente perdita di una grossa parte dell'energia scatterata sul grumo, legata al modo propagativo principale.

In figura 71 riportiamo la sezione longitudinale della struttura MESA di un fotodiiodo PIN, realizzato mediante l'eterostruttura InP/In_(0.57)Ga_(0.43)As/InP, ottimizzato per la rivelazione di segnali ottici NIR (vicino infrarosso) appartenenti alla terza finestra: la sua eterostruttura è progettata in osservanza ai criteri deducibili dal grafico di figura 69.

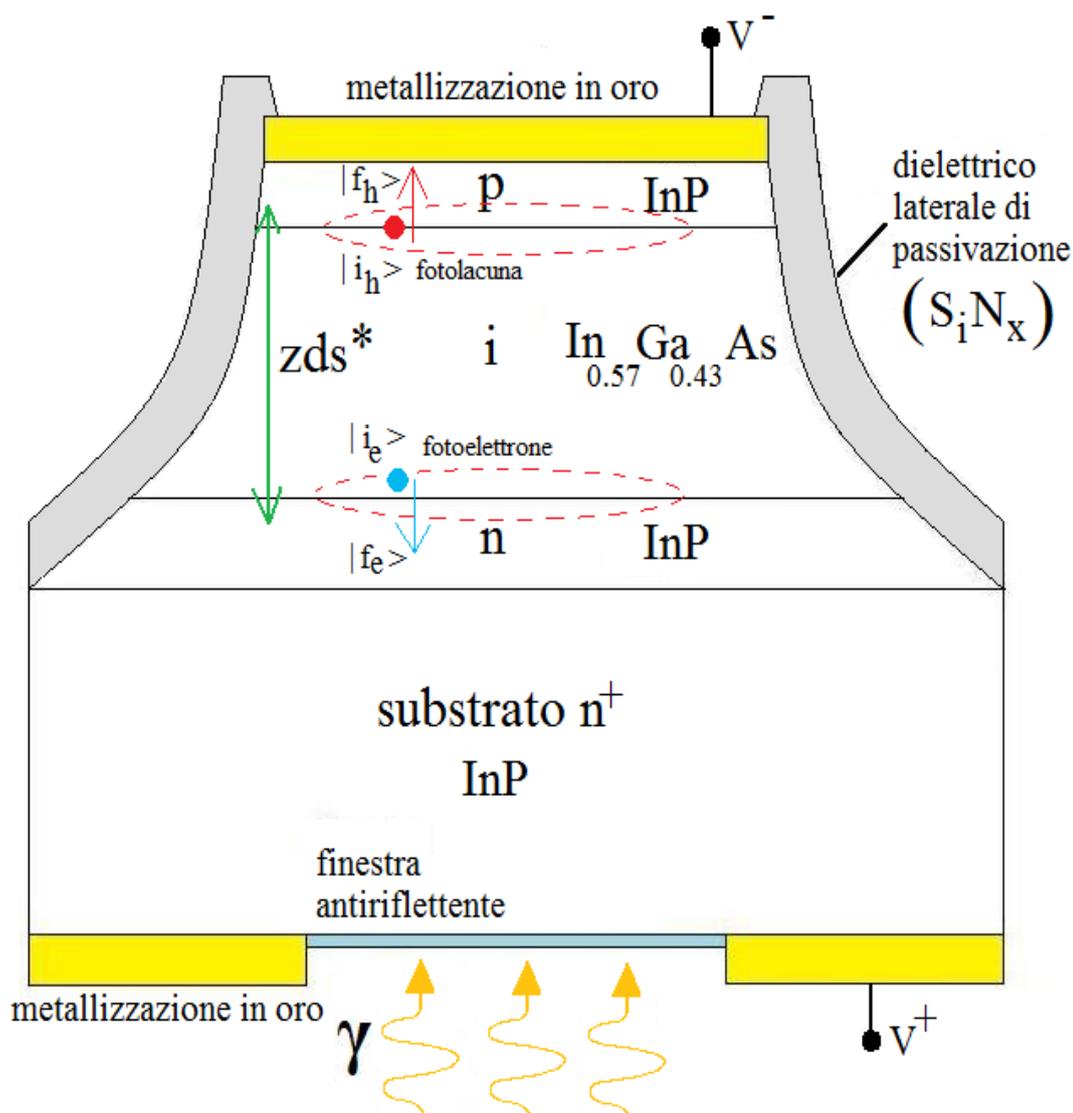


Fig. 71

Sezione longitudinale della struttura MESA di un fotodiiodo PIN, ottimizzato per la rivelazione di luce infrarossa appartenente alla terza finestra di attenuazione delle fibre ottiche in silice ($\lambda \approx 1.55 \mu\text{m}$): la sua eterostruttura InP/In_(0.57)Ga_(0.43)As/InP è progettata in osservanza ai criteri deducibili dal grafico di figura 68.

La disposizione mostrata in figura 71 permette l'aumento del confinamento dell'assorbimento ottico nella zona intrinseca, la cui lunghezza è leggermente inferiore a quella zds^* della zona di svuotamento in presenza di polarizzazione inversa: in tal modo la componente diffusiva della fotocorrente è ridotta praticamente a zero. Ciò comporta un aumento della velocità di risposta del PIN agli impulsi ottici incidenti, ovvero un incremento della banda passante del fotodiiodo. Il coefficiente di assorbimento α per luce a circa $1.5 \mu\text{m}$, in In_(0.57)Ga_(0.43)As, è circa pari a 104 cm^{-1} : sono stati realizzati dispositivi MESA con regione di assorbimento di spessore pari a $3 \mu\text{m}$, aventi un assorbimento ottico pari al 95% della potenza ottica incidente P_0 , e senza una significativa degradazione della velocità di risposta, essendo il tempo di transito abbastanza breve. È importante notare che la lunghezza del layer intrinseco, e conseguentemente la lunghezza zds^* della zona di svuotamento in presenza di polarizzazione

inversa, può anche essere abbastanza grande, dell'ordine di qualche decina di μm ; ciò comporta un'efficienza quantica QE molto buona, senza tuttavia degradare molto il tempo di transito dei portatori di carica fotogenerati attraverso il layer intrinseco. Quest'ultima osservazione è giustificabile considerando che la massa efficace degli elettroni per la mobilità $m_{e\text{Alloy}}^*$, all'interno del semiconduttore ternario $\text{In}_{(0.57)}\text{Ga}_{(0.43)}\text{As}$ con cui è realizzato lo strato intrinseco di assorbimento ottico, è decisamente piccola. $m_{e\text{Alloy}}^*$, come spiegato nel paragrafo 5.9.2, è calcolabile in modo approssimato utilizzando la legge di Vegard per le masse efficaci:

$$\frac{1}{m_{e\text{Alloy}}^*} = \frac{1}{m_{e\text{In}_{(0.57)}\text{Ga}_{(0.43)}\text{As}}^*} = 0.57 \frac{1}{m_{e\text{InAs}}^*} + 0.43 \frac{1}{m_{e\text{GaAs}}^*} \approx 0.57 \frac{1}{0.027m_0} + 0.43 \frac{1}{0.067m_0}$$

da cui si calcola che:

$$m_{e\text{Alloy}}^* = m_{e\text{In}_{(0.57)}\text{Ga}_{(0.43)}\text{As}}^* \approx 0.0363 m_0$$

Questo valore si può ricavare anche dal grafico mostrato in figura 68. Quindi gli elettroni di conduzione fotogenerati all'interno del layer intrinseco sono molto leggeri, molto reattivi allo stimolo elettrostatico esterno legato alla polarizzazione inversa del PIN, ovvero facilmente accelerabili dal campo elettrico imposto dall'esterno ai terminali metallici (Au) del fotodiodo. La concavità della curva $E(k)$ della banda di conduzione del $\text{In}_{(0.57)}\text{Ga}_{(0.43)}\text{As}$ presso il proprio minimo E_C è molto marcata, quindi la mobilità μ_n degli elettroni in BC nel layer intrinseco è alta, pertanto la velocità con la quale i fotoelettroni attraversano la zds^* e raggiungono il bulk n di destinazione è alta e quindi il tempo di transito dei fotoelettroni è piccolo. Dunque la zds^* può anche essere lunga, al fine di massimizzare la QE, pur conservando una banda passante del PIN accettabile (il segnale elettrico in uscita dal PIN "insegue" bene, nel tempo, l'evoluzione del segnale ottico in ingresso). La lunghezza dello strato intrinseco consente di ottenere un buon valore dell'efficienza quantica "interna", ovvero un buon rapporto fra il numero di fotoportatori di carica che giungono presso le due interfacce $\text{In}_{(0.57)}\text{Ga}_{(0.43)}\text{As}/\text{InP}$, evidenziate in figura 71 dalle ellissi tratteggiate, ed il numero di fotoni di terza finestra incidenti sullo strato antiriflettente del PIN. C'è tuttavia la possibilità che i fotoportatori non attraversino le due interfacce, come spiegato nel paragrafo 5.9.1 e nell'espressione [e72], ovvero che il rapporto fra il numero di fotoportatori che effettivamente arrivano ai terminali metallici ed il numero di fotoportatori che raggiungono le due interfacce $\text{In}_{(0.57)}\text{Ga}_{(0.43)}\text{As}/\text{InP}$ non sia unitario, cioè che l'efficienza quantica "esterna" non sia unitaria: ciò degraderebbe l'efficienza quantica "totale", data dal prodotto fra quella interna e quella esterna. Tuttavia il semiconduttore ternario $\text{In}_{(0.57)}\text{Ga}_{(0.43)}\text{As}$ e quello binario InP hanno la stessa struttura cristallina di tipo zincoblenda e lo stesso valore della costante reticolare ($a \approx 5.86 \text{ \AA}$), pertanto non c'è alcuno stress meccanico reticolare alle due interfacce ($\varepsilon \approx 0$), ovvero le celle fondamentali di $\text{In}_{(0.57)}\text{Ga}_{(0.43)}\text{As}$ e di InP , presso le due interfacce, non subiscono alcuna deformazione meccanica parallela ed ortogonale rispetto al piano dell'interfaccia. Tale matching reticolare determina l'assenza di legami ibridi sp^3 rotti, non completi, cioè occupati da un unico elettrone, pertanto elettronicamente molto recettivi, i quali potrebbero catturare un fotoelettrone (di spin opposto rispetto allo spin dell'elettrone già presente sull'orbitale sp^3 , per il principio di esclusione di Pauli), impedendogli di arrivare al bulk n, oppure potrebbero perdere il loro unico elettrone, cioè svuotarsi, e ciò equivale a trattenere una fotolacuna, impedendole di arrivare al bulk p. In figura 72 è mostrata semiquantitativamente la struttura cristallina dell'interfaccia p/i del PIN fin qui trattato: l'assenza di deformazioni reticolari e quindi di stress meccanici, cioè l'assenza di legami covalenti incompleti, favorisce il transito delle lacune fotogenerate nel layer i. La gradualità del passaggio, a livello di composizione atomica della matrice cristallina, dal layer i, costituito dal $\text{In}_{(0.57)}\text{Ga}_{(0.43)}\text{As}$, al layer p, costituito dal InP , è stata notevolmente accorciata nel disegno di figura 72, per ragioni grafiche. Infatti nella realtà non si verifica un passaggio di fatto brusco dalla matrice cristallina del $\text{In}_{(0.57)}\text{Ga}_{(0.43)}\text{As}$ a quella del InP , come mostrato in figura 72: la transizione cristallina è graduale e può interessare uno strato di materiale, la cui matrice cristallina è data da una sorta di "mixing" fra quella del $\text{In}_{(0.57)}\text{Ga}_{(0.43)}\text{As}$ e quella del InP , spesso anche $20 \div 30$ monolayers (un monolayer è spesso circa 3.3 \AA).

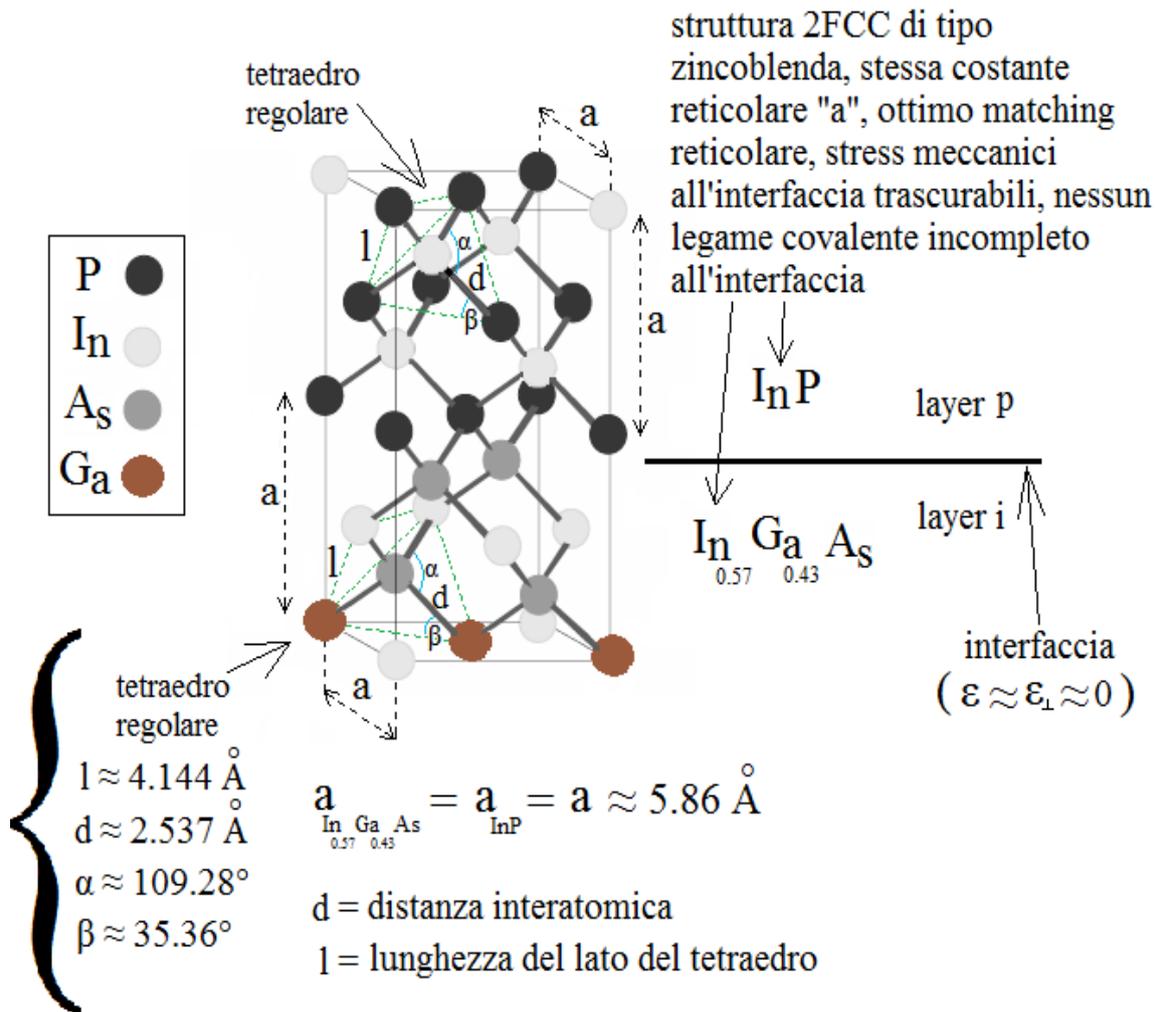


Fig. 72

Disegno semiquantitativo della struttura cristallina dell'interfaccia p/i del PIN MESA di terza finestra fin qui trattato: l'assenza di deformazioni reticolari e quindi di stress meccanici, cioè l'assenza di legami covalenti incompleti, favorisce il transito delle lacune fotogenerate nel layer i. La gradualità del passaggio, a livello di composizione atomica della matrice cristallina, dal layer i, costituito dal $\text{In}_{0.57}\text{Ga}_{0.43}\text{As}$, al layer p, costituito dal InP , è stata notevolmente accorciata nel disegno per ragioni grafiche. Infatti nella realtà non si verifica un passaggio quasi brusco dalla matrice cristallina del $\text{In}_{0.57}\text{Ga}_{0.43}\text{As}$ a quella del InP ; la transizione cristallina è graduale e può interessare uno strato di materiale, la cui matrice cristallina è data da una sorta di "mixing" fra quella del $\text{In}_{0.57}\text{Ga}_{0.43}\text{As}$ e quella del InP , spesso anche 20 ÷ 30 monolayers (un monolayer è spesso circa 3.3 \AA).

L'assenza di legami sp^3 non completi favorisce il passaggio di un fotoelettrone da uno stato elettronico iniziale $|i_e\rangle$, localizzato nel layer i, ad uno stato elettronico finale $|f_e\rangle$, localizzato nel layer n, nonché il passaggio di una fotolacuna da uno stato iniziale $|i_h\rangle$, localizzato nel layer i, ad uno stato finale $|f_h\rangle$, localizzato nel layer p. In formule:

$$P_e \propto |\langle f_e | \vec{a} \cdot \vec{p}_e | i_e \rangle|^2$$

$$P_h \propto |\langle f_h | \vec{a} \cdot \vec{p}_h | i_h \rangle|^2$$

Nei due elementi di matrice della quantità di moto dei fotoelettroni e delle fotolacune riconosciamo \vec{a} , che rappresenta il vettore di polarizzazione del campo elettrico totale (campo interno di "background" del cristallo + campo di polarizzazione esterna), \vec{p}_e , che rappresenta l'impulso del fotoelettrone in procinto di attraversare l'interfaccia layer i/layer n, \vec{p}_h , che rappresenta l'impulso della fotolacuna in procinto di attraversare l'interfaccia layer i/layer p. P_e può essere vista come la probabilità con cui il fotoelettrone attraversa l'interfaccia layer i/layer n; analogamente P_h può essere vista come la probabilità con cui la fotolacuna attraversa l'interfaccia layer i/layer n.

p. Il matching reticolare fra $\text{In}_{(0.57)}\text{Ga}_{(0.43)}\text{As}$ e InP determina valori di P_e e di P_h molto buoni, diciamo abbastanza prossimi all'unità, rendendo quasi unitaria l'efficienza quantica esterna del PIN.

In un PIN simile gioca un ruolo fondamentale non solo la presenza di difetti reticolari alle due interfacce p/i ed i/n, la quale degrada l'efficienza quantica esterna, ma anche la presenza di imperfezioni reticolari all'interno del bulk dei tre layers p, i ed n, quali dislocazioni e vacanze, ed in particolare all'interno del bulk del layer i. Tali difetti individuano livelli energetici all'interno del gap (deep levels), analogamente alle trappole metalliche, i quali possono accettare portatori di carica. È possibile che un fotone, dopo aver superato indenne il InP ed essere stato assorbito dal $\text{In}_{(0.57)}\text{Ga}_{(0.43)}\text{As}$, produca una fotocoppia elettrone/lacuna che non darà alcun contributo alla fotocorrente in uscita, poiché tale coppia si ricombinerà, poco dopo la sua generazione, a causa dei suddetti difetti reticolari. L'effetto deleterio delle imperfezioni reticolari, che causano una diminuzione della QE totale, è più probabile in semiconduttori quaternari (un motivo in più per scegliere, come strato i, il ternario suddetto cresciuto con la minore quantità possibile di fosforo P). Il valor medio della QE per il $\text{In}_{(0.57)}\text{Ga}_{(0.43)}\text{As}$, in corrispondenza di lunghezze d'onda rientranti nella seconda e terza finestra, si aggira intorno a $0.3 \div 0.4$, e tale valore, già abbastanza basso, è destinato a diminuire in proporzione alla concentrazione delle impurezze. Il substrato n^+ di InP può essere costituito da un wafer CZ o FZ, sul quale operiamo le deposizioni epitassiali di n, i e p. Lo strato n funziona da buffer durante la deposizione epitassiale di i, ovvero inibisce la diffusione termica di alcune impurezze, inevitabilmente presenti, dal substrato n^+ al layer intrinseco, garantendo che la QE interna del PIN non si discosti eccessivamente dai suoi valori ideali.

La forma MESA, riportata in figura 71, si ottiene con opportuni attacchi umidi in ambiente acido; quest'ultimo deve essere selettivo nei confronti del InP, ovvero la sua capacità di "etching" deve essere prevalentemente rivolta contro il $\text{In}_{(0.57)}\text{Ga}_{(0.43)}\text{As}$. La sagomatura che ne deriva è causa di una fluttuazione non deterministica (una perdita stocastica) sia della fotocorrente, sia della corrente di buio, chiamata "corrente di leakage" (I_{leak}). Nei pressi della superficie sagomata, infatti, è molto più probabile trovare vacanze, dislocazioni, impurezze sostituzionali o interstiziali che non all'interno del bulk del materiale, per cui, in quelle regioni superficiali, i portatori di carica fotogenerati (così come quelli termici) hanno un tempo di vita medio assai più breve che altrove e possono essere più facilmente catturati dai deep levels legati ai difetti reticolari. Gli stati energetici profondi, cioè interni al gap, localizzati prevalentemente sulla superficie sagomata a forma di MESA, sono legati fortemente anche al riordinamento cristallino a cui un qualunque semiconduttore è soggetto presso le proprie superfici. Infatti in superficie ciascun atomo non risente delle stesse interazioni coulombiane di cui risentono gli atomi posizionati nel bulk del materiale, lontano quindi dalla superficie; le interazioni fra gli atomi posizionati in superficie non consentono agli stessi il raggiungimento della condizione di massima stabilità (ciascun atomo superficiale risente della mancanza di alcuni atomi primi vicini, cioè della mancanza di alcuni legami covalenti), ovvero di minima energia elettrostatica. Il raggiungimento di una condizione simile a quella di massima stabilità è possibile solo se la matrice cristallina del semiconduttore, presso la superficie, assume spontaneamente una configurazione geometrica "di superficie" diversa da quella tipica del materiale, cioè da quella individuabile nel bulk del cristallo. Questa sorta di "autoriordinamento reticolare superficiale" consente, presso la superficie, una certa stabilità elettrostatica. In figura 73 sono riportati i quattro difetti di punto più importanti, le cui distorsioni reticolari interessano regioni del cristallo molto ristrette, localizzate entro qualche cella fondamentale; i difetti di punto possono interessare sia il bulk che la superficie di un semiconduttore (ovviamente la probabilità di trovarli in superficie è maggiore della probabilità di trovarli nel bulk). In figura 74 sono riportati i due più importanti difetti di linea, anch'essi individuabili maggiormente presso la superficie piuttosto che nel bulk; tali difetti si sviluppano in regioni del cristallo molto estese, lungo anche alcune decine di celle fondamentali. In figura 75 sono riportate a sinistra la struttura reticolare del semiconduttore ternario $\text{In}_{(0.57)}\text{Ga}_{(0.43)}\text{As}$ all'interno del bulk (che è la stessa struttura cristallina del InP, con il P al posto di As ed il In al posto di Ga), osservata lungo la direzione cristallografica [001], mentre a destra la struttura reticolare "autoriordinata" del $\text{In}_{(0.57)}\text{Ga}_{(0.43)}\text{As}$ in prossimità della superficie.

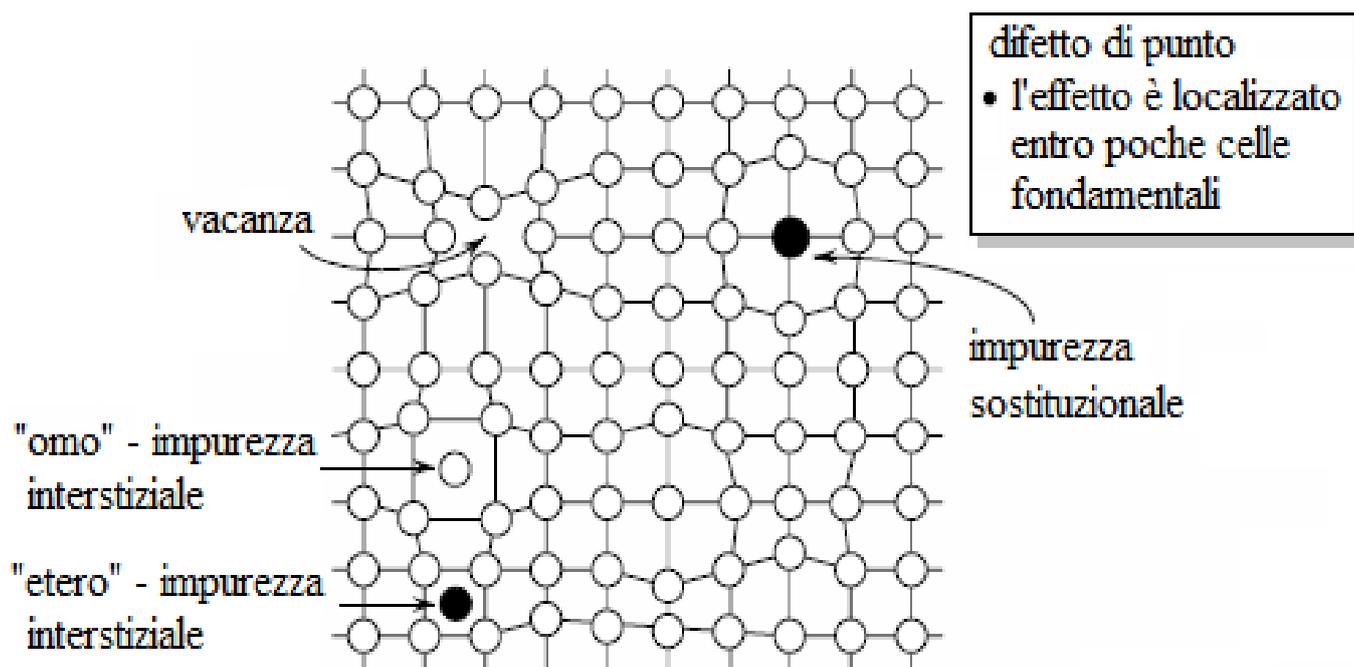


Fig. 73

Rappresentazione qualitativa delle deformazioni reticolari prodotte dai quattro difetti di punto più importanti: tali distorsioni della matrice cristallina interessano regioni del materiale molto ristrette, localizzate lungo poche celle fondamentali. I difetti di punto possono interessare sia il bulk che la superficie di un semiconduttore.

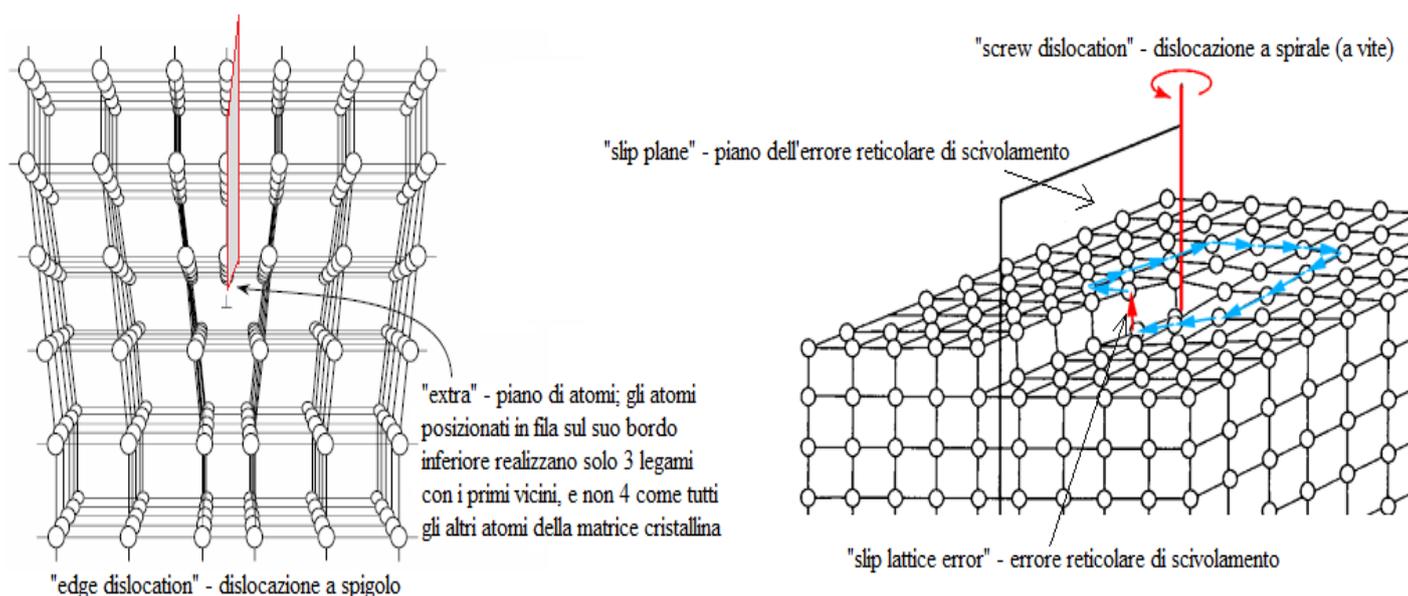


Fig. 74

Rappresentazione qualitativa delle deformazioni reticolari prodotte dai due più importanti difetti di linea ("edge dislocation" - dislocazione a spigolo, "screw dislocation" - dislocazione a spirale), individuabili sia presso la superficie che nel bulk del cristallo semiconduttore. Tali difetti si sviluppano in regioni del materiale molto estese, lungo anche alcune decine di celle fondamentali.

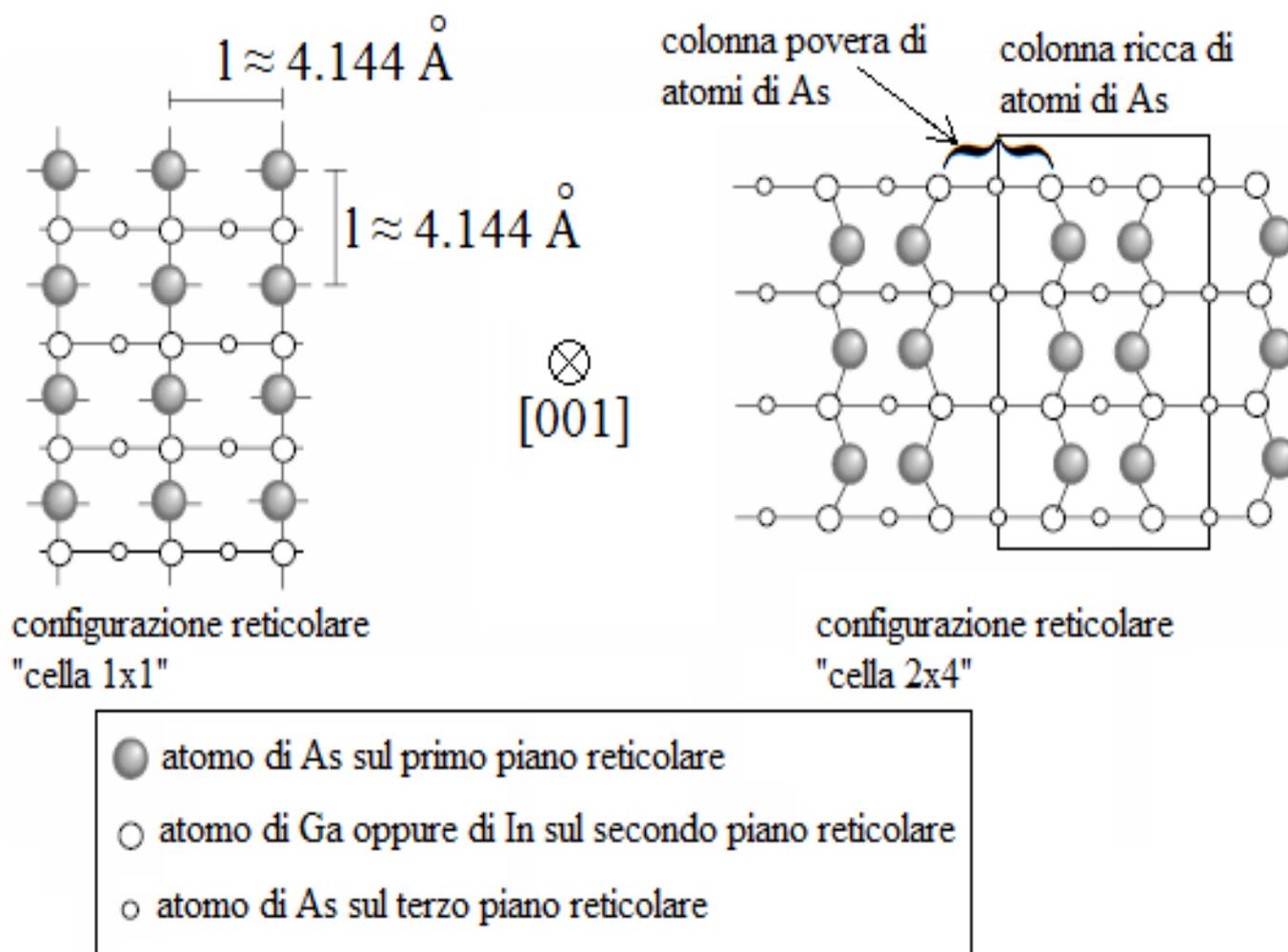


Fig. 75

A sinistra è riportata una rappresentazione qualitativa della struttura reticolare del semiconduttore ternario $\text{In}_{(0.57)}\text{Ga}_{(0.43)}\text{As}$ all'interno del proprio bulk (che è la stessa struttura cristallina del InP , con il P al posto di As ed il In al posto di Ga), osservata lungo la direzione cristallografica [001]. A destra è riportata una rappresentazione qualitativa della struttura reticolare "autoriordinata" del semiconduttore ternario $\text{In}_{(0.57)}\text{Ga}_{(0.43)}\text{As}$ in prossimità della superficie del cristallo: il riordinamento reticolare è funzionale al raggiungimento, da parte degli atomi posizionati sulla superficie, di una condizione di stabilità degradata dall'alta presenza di impurezze e difetti e dalla mancanza di alcuni atomi primi vicini, cioè dalla mancanza di alcuni legami covalenti.

La perdita stocastica I_{leak} di corrente, dovuta alla cattura dei fotoportatori di carica da parte di stati energetici interni al gap, individuati dai difetti reticolari, di superficie e di bulk, e dagli autoriordinamenti reticolari superficiali, è tipica di qualunque struttura (non solo MESA), dato che qualunque dispositivo presenta delle regioni superficiali e dei difetti di crescita. La funzione principale del dielettrico di passivazione è quella di "colmare" gli stati accettori di portatori, dovuti all'interruzione della struttura cristallina, limitando la corrente di leakage che cortocircuita p, i ed n attraverso le regioni superficiali. Un grosso problema tecnologico, relativo alla deposizione dell'isolante di cui sopra, è dovuto al fatto che il $\text{In}_{(0.57)}\text{Ga}_{(0.43)}\text{As}$ è un ternario che non possiede un ossido stabile (per esempio il silicio lo possiede, ed è la silice SiO_2). Si utilizzano quindi nitruri di silicio SiN_x depositi da fase vapore. La deposizione avviene in camere con ammoniaca NH_3 e silano SiH_4 : l'energia per promuovere il legame viene fornita o mediante l'aumento della temperatura (con problemi non trascurabili per i semiconduttori sottostanti) o mediante l'illuminazione con raggi ultravioletti, in modo tale che l'impatto con i fotoni energetici favorisca la formazione del precipitato.

In figura 76 riportiamo un diagramma logaritmico che riassume le regole di progetto per PIN simmetrici, realizzati con l'eterostruttura finora descritta. Possiamo notare la proporzionalità diretta fra la larghezza della zona di svuotamento zds^* , la quale è controllata dalla polarizzazione inversa $|V|$, l'efficienza quantica QE ed il tempo medio impiegato dai portatori fotogenerati per attraversare tutta la zds^* . Questi tre parametri sono inversamente proporzionali alla capacità C della zds^* (quest'ultima può essere vista come un condensatore ad armature piatte e parallele). Consideriamo un PIN InP/In_(0.57)Ga_(0.43)As/InP simmetrico, con $N_A = N_D = 3 \cdot 10^{15} \text{ cm}^{-3}$ e con lunghezza dello strato intrinseco fissata (drogaggi e dimensioni geometriche sono, ovviamente, dei parametri costanti, così come la lunghezza d'onda della radiazione luminosa incidente sul PIN). Si noti altresì la proporzionalità inversa fra la zds^* e la concentrazione dei droganti p ed n, come avviene, del resto, in una qualunque omogiunzione pn brusca. Ipotizziamo di polarizzare inversamente il dispositivo con un voltaggio pari a $|V_P|$, tale per cui il tempo di transito delle fotocariche è pari a 2 ps e la larghezza della zona di svuotamento zds^* è pari a 1 μm (siamo, sul grafico di figura 76, in prossimità del punto di funzionamento P). L'aumento del voltaggio inverso $|V|$ comporta lo spostamento del punto di funzionamento del PIN lungo la retta (colorata di blu, nel grafico) parametrizzata da $N_A = N_D = 3 \cdot 10^{15} \text{ cm}^{-3}$. Tale spostamento rettilineo è collegato all'aumento della larghezza della zds^* ed al conseguente aumento del tempo di transito per le cariche fotogenerate nella zds^* (regione di assorbimento). All'aumentare di $|V|$ aumenta anche la QE, ossia diminuisce la frazione della potenza ottica incidente che viene "persa" all'interno delle zone neutre dei lati p ed n, dato che queste diventano più sottili. C diminuisce, per cui diminuisce la costante di tempo caratteristica dei processi di diffusione dei portatori fotogenerati nelle regioni neutre di p ed n: le correnti di diffusione diventano meno importanti (regimano più velocemente), quindi ciò che risulta limitante per la banda operativa del PIN è il tempo di transito. Possiamo aumentare $|V|$ fino ad arrivare, sul grafico, al punto F ($|V| \rightarrow |V_F|$), presso il quale il fotorivelatore è caratterizzato da degli ottimi valori di zds^* (circa 3.3 μm) e QE (circa 0.99), mentre le fotocariche attraversano la zds^* con un tempo medio ragionevolmente breve (circa 50 ps). Se tuttavia continuiamo ad incrementare il voltaggio inverso, il campo elettrico, all'interno della zds^* , assume un valore (in modulo) talmente alto da comportare il formarsi di correnti di tunneling (a causa dell'eccessivo piegamento delle bande) dovute al breakdown Zener (questo aspetto deleterio, circa il funzionamento di un PIN, lo approfondiremo nel paragrafo 5.10). Graficamente questo fenomeno di tunneling, che causerebbe il totale annegamento del fotosegnale nel rumore, è evidenziato dalla superficie tratteggiata, all'interno della quale entra il punto di funzionamento del PIN, nel caso in cui $|V| > |V_F|$. La regione tratteggiata è una zona "non consentita", ai fini della fotorivelazione, poco prima della quale è consigliabile fissare il voltaggio di polarizzazione inversa. Se consideriamo un PIN simmetrico con i lati p ed n drogati abbastanza pesantemente (quindi un P⁺/I/N⁺), ad esempio $N_A = N_D = 3 \cdot 10^{16} \text{ cm}^{-3}$, si ha che l'elevata pendenza del campo elettrico, all'interno delle regioni svuotate appartenenti ai lati suddetti, determina il raggiungimento della condizione di breakdown ($\epsilon \rightarrow \epsilon_{BD}$) per valori di $|V|$ decisamente più bassi rispetto a quelli di prima (tanto più bassi quanto maggiore è il drogaggio), e ciò è visibile, sul grafico, dal fatto che l'intersezione fra la retta parametrizzata da $N_A = N_D = 3 \cdot 10^{16} \text{ cm}^{-3}$ e l'inizio della regione tratteggiata corrisponde ad un valore di zds^* basso e pertanto non favorevole all'assorbimento dei fotoni (e infatti la QE è abbastanza piccola, circa 40 ÷ 45 %). In questo caso non è possibile aumentare zds^* , e di conseguenza QE, incrementando $|V|$: ciò comporterebbe l'ingresso del punto di funzionamento del PIN all'interno della zona proibita, dove il fotosegnale risulterebbe non intellegibile, cioè indistinguibile dal rumore per un'elettronica di front - end. Per ottenere una regione di svuotamento sufficientemente larga ed una buona efficienza quantica, ovviamente senza entrare nella zona tratteggiata del grafico in figura 76, è necessario avere un livello di drogaggio, per i lati p ed n, minore di $8 \cdot 10^{15} \text{ cm}^{-3}$, ovvero muoversi, al variare di $|V|$, lungo una delle rette che si trovano al di sopra di quella evidenziata in rosso.

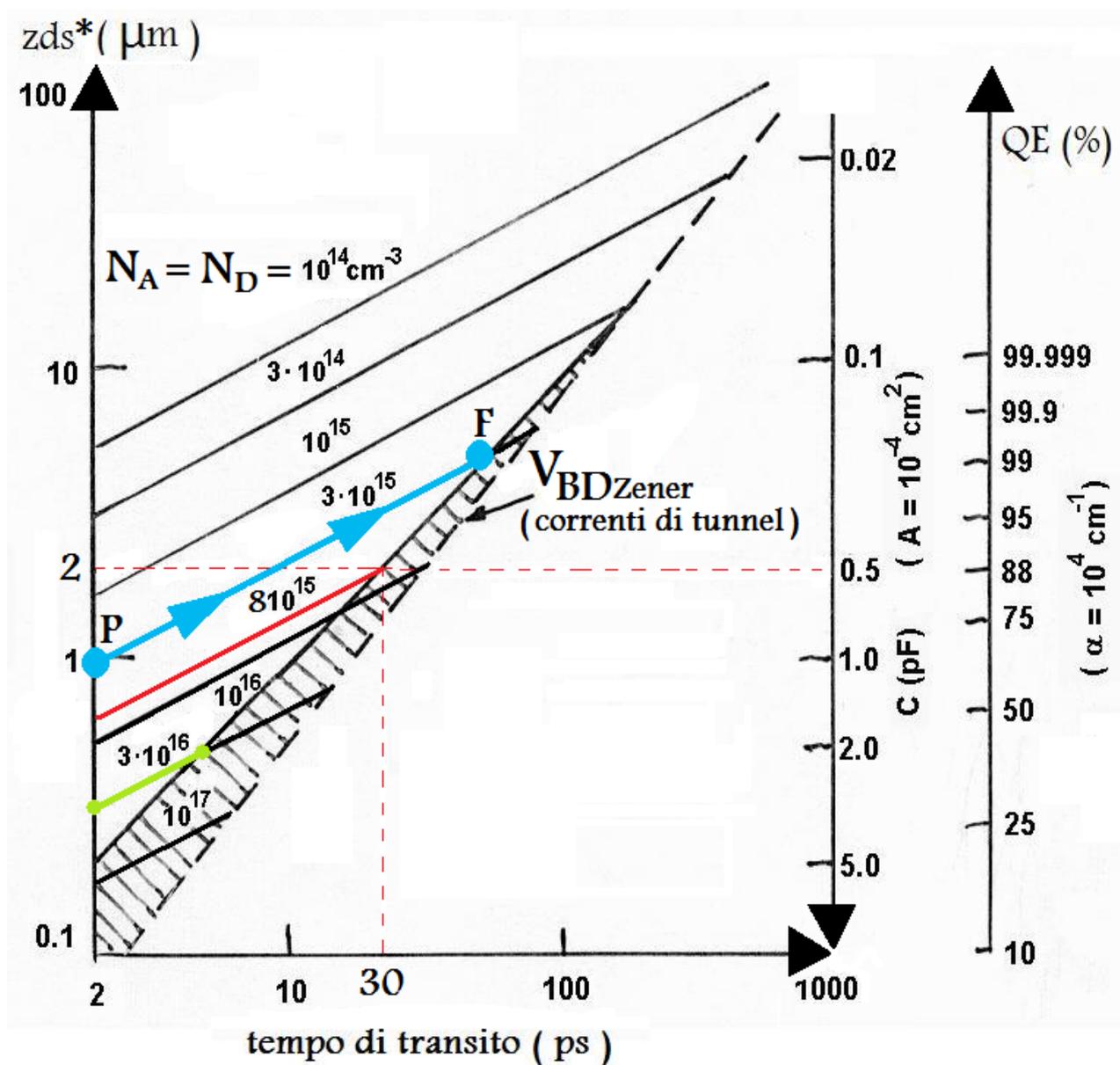


Fig. 76

Diagramma riassuntivo delle regole di progetto per PIN simmetrici realizzati con l'eterostruttura $\text{InP}/\text{In}_{(0.57)}\text{Ga}_{(0.43)}\text{As}/\text{InP}$. Notiamo la proporzionalità diretta fra la larghezza della zona di svuotamento z_{ds}^* ed il relativo tempo medio di transito delle fotocariche, nonché la proporzionalità diretta fra la stessa z_{ds}^* e l'efficienza quantica QE. Si noti come all'aumentare della z_{ds}^* la capacità C (capacità della z_{ds}^* , vista come un condensatore ad armature piatte e parallele) diminuisce: ciò determina la diminuzione della costante di tempo caratteristica dei processi di diffusione dei portatori fotogenerati nelle regioni neutre p ed n.

5.10) L'impossibilità di sfruttare, in un PIN, la moltiplicazione a valanga ai fini della fotorivelazione

Osserviamo il diagramma a bande (molto qualitativo) di un PIN polarizzato fortemente in inversa ed illuminato:

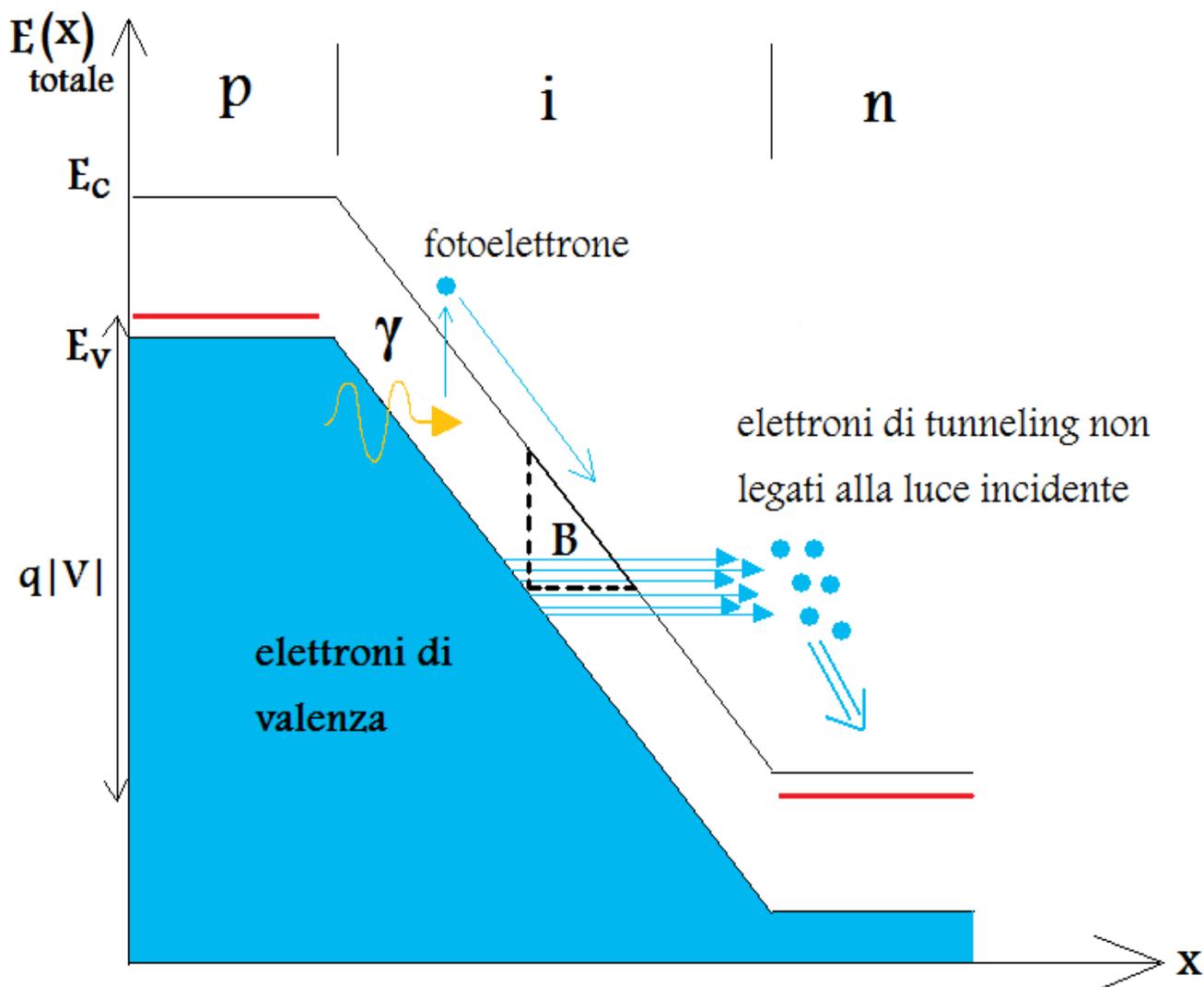


Fig. 77

Diagramma a bande di un PIN polarizzato fortemente in inversa. Si noti come il debole fotosegnale utile è accompagnato da un'intensa corrente di tunneling, causata dall'assottigliamento eccessivo della barriera di potenziale B, vista dagli elettroni di valenza dello strato intrinseco: l'assottigliamento è provocato dall'alto voltaggio inverso $|V|$.

La forte polarizzazione inversa comporta un piegamento delle bande, all'interno del layer intrinseco, molto accentuato. Un fotone assorbito all'interno del layer i può provocare la transizione di un elettrone dalla BV alla BC: il fotoelettrone subisce una cospicua accelerazione verso il lato n, in virtù della grande pendenza di $E_c(x)$, ossia del forte campo, il quale può consentire all'elettrone primario di produrre ionizzazioni da impatto. Tuttavia l'alto voltaggio inverso $|V|$ rende molto sottile la barriera di potenziale B, vista dagli elettroni appartenenti alla BV del layer i, per cui questi elettroni, non legati in alcun modo alla potenza ottica incidente, hanno alta probabilità di attraversare B ed andare nel bulk del lato n (nella sua BC). L'afflusso di portatori (correnti di tunnel) verso i rispettivi bulk, causato dal breakdown Zener, è molto maggiore di qualunque fotosegnale I_L ordinario, seppur moltiplicato (il segnale MI_L "affoga" nel rumore), per cui un PIN non può essere utilizzato come fotomoltiplicatore ad effetto valanga.

5.11) Un esempio di PIN al silicio commerciale: l'elettronica di front - end, le caratteristiche del dispositivo e i passi di processo per la sua integrazione

In figura 78 riportiamo la sezione e la pianta di un PIN al silicio largamente diffuso (disegni non in scala), ottimizzato per la ricezione di segnali ottici NIR appartenenti alla prima finestra di attenuazione delle fibre ottiche realizzate in silice, nel quale i lati p ed n sono pesantemente drogati, mentre lo strato "intrinseco" di fatto è costituito da Si leggermente drogato di tipo n. L'alto livello di drogaggio di p ed n consente di localizzare la zds interamente presso il layer n, anche per alti valori della tensione inversa di polarizzazione, oltre ad attenuare le correnti di diffusione fra quest'ultimo e ciascuna delle due well. Tali correnti sono inessenziali sia perché i fotominoritari, lì generati, ricombinano prima di poter entrare nella zona di campo, sia perché l'intensità luminosa assorbita dalle well è trascurabile, in virtù della piccola profondità a cui si trovano le interfacce p⁺/n⁻ e n⁺/n⁻. L'unica corrente di diffusione apprezzabile è quella delle lacune fotogenerate nella zona neutra di n⁻ (queste infatti, a causa del drogaggio molto basso, hanno una certa probabilità di diffondere e raggiungere la zds). Tuttavia lo spessore considerevole della zona di campo (d) consente di trascurare la frazione di fotoni incidenti che arriva fino al bulk n. Il pesante drogaggio delle well rende spesso la zds ed ohmici i contatti con l'alluminio. La finestra antiriflettente è realizzata con uno strato di silice SiO₂. L'ossido, oltre alla passivazione ed alla diminuzione della riflettività, subita dalla potenza ottica incidente, provvede anche a colmare le vacanze superficiali, attenuando la corrente di leakage ed aumentando la QE del PIN.

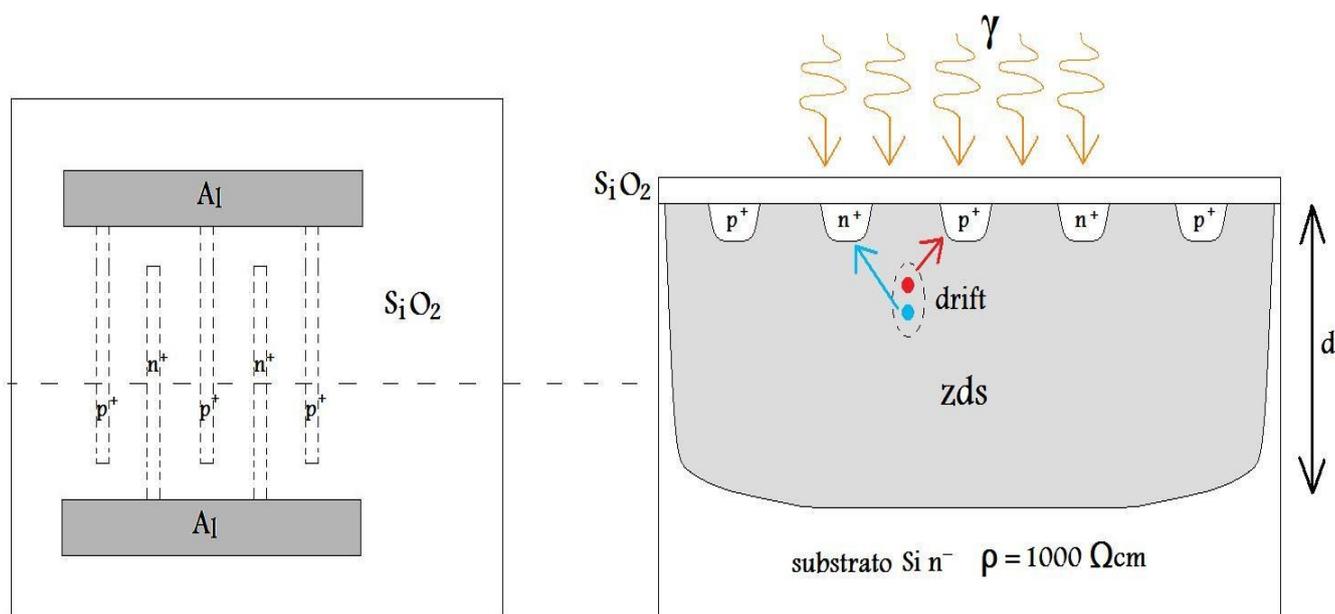


Fig. 78

Pianta e sezione di un fotodiodo PIN al silicio, ottimizzato per la ricezione di fotosegnali infrarossi di lunghezza d'onda $\lambda \approx 0.85 \mu\text{m}$ (ricezione di segnali, ovvero di stringhe di bit, in prima finestra, nel caso in cui il PIN sia connesso ad una fibra ottica in qualità di primo stadio di un ricevitore). A questo scopo sarebbe più indicato un PIN all'GaAs, il quale ha un coefficiente di assorbimento α più alto di quello del Si, per quelle frequenze. Tuttavia il processo realizzativo per il PIN all'GaAs è più complicato e costoso di quello per il PIN al Si.

Riportiamo come esempio alcune grandezze tipiche:

$$I_d = I_0(e^{\frac{V}{\eta V_T}} - 1) - I_L$$

$$\eta \approx 1.1, V_T \approx 26 \text{ mV}$$

Dimensioni zona attiva: $2 \mu\text{m} \times 10 \mu\text{m}$

$d = 25 \mu\text{m}$ (con $V = -5 \text{ V}$), $d = 80 \mu\text{m}$ (con $V = -30 \text{ V}$)

$$QE = (1 - R) \frac{\text{area attiva}}{\text{area totale}} (1 - e^{-\alpha d})$$

$$\alpha^{-1} = 20 \mu\text{m}, \text{ con } \lambda = 850 \text{ nm}$$

$$\text{tempo di diffusione delle lacune} \cong \frac{L_{h,n}^2}{D_{h,n}}$$

$$\text{tempo di transito} \cong \frac{d}{v_{drift}} = \frac{d}{\mu \varepsilon}$$

$$\text{costante di tempo} = R_L C_d$$

$$\text{banda} = \frac{1}{2\pi R_{totale} C_{totale}} \approx 200 \text{ MHz}$$

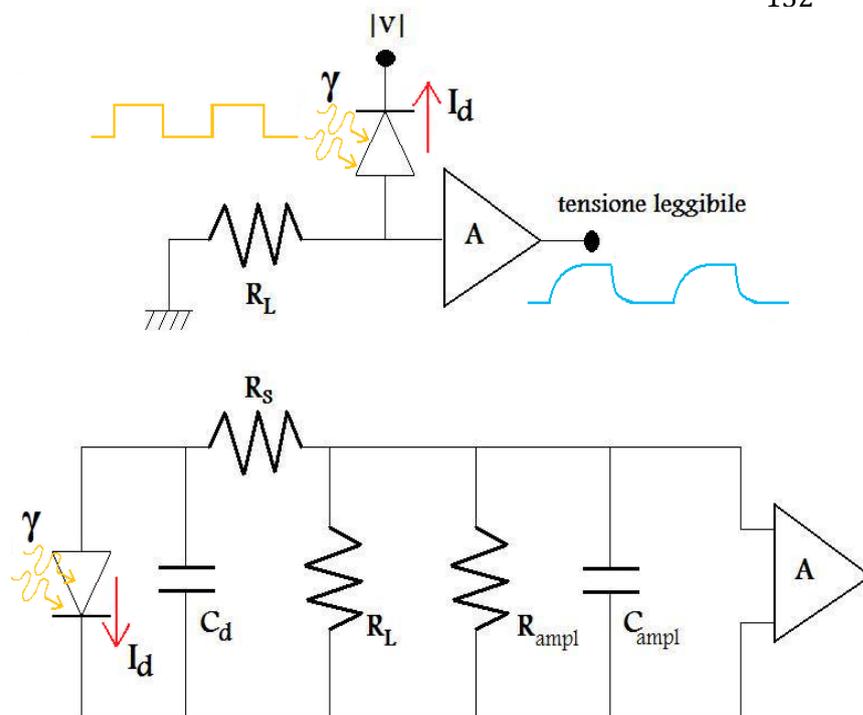


Fig. 79

Equivalentente circuitale, molto semplificato, dell'elettronica di front-end necessaria all'elaborazione del fotosegnale elettrico prodotto dal PIN al Si.

Elenchiamo i passi di processo necessari all'integrazione del dispositivo, partendo da un wafer di Si n-monocristallino (di tipo CZ, spessore tipico di 700 μm , purificato in piú possibile dalle trappole metalliche, mediante gettering intrinseco o estrinseco):

- crescita termica di uno strato di ossido (dry o wet) o deposizione CVD dello stesso
Nel caso si opti per la seconda tecnica riportiamo alcune delle reazioni piú sfruttate (la scelta è determinata dalla temperatura e dal tipo di forno a disposizione)
 - a temperature basse (300 – 450 °C) si sfrutta la reazione del silano con l'ossigeno
 $\text{SiH}_4 + \text{O}_2 \rightarrow \text{SiO}_2 + 2\text{H}_2$
 - a temperature medie (650 – 750 °C) si utilizza la decomposizione, a bassa pressione, del tetraetil ortosilicato (TEOS) $\text{Si}(\text{OC}_2\text{H}_5)_4$
 - a temperature alte (oltre i 900 °C) la LPCVD (pressione nel reattore ridotta, diffusione migliore, velocità di crescita del film maggiore) di diclorosilano ed ossido nitroso
 $\text{SiH}_2\text{Cl}_2 + 2\text{N}_2\text{O} \rightarrow \text{SiO}_2 + 2\text{N}_2 + 2\text{HCl}$
Quest'ultima produce strati di ossido di qualità eccellente, molto simile a quella che caratterizza gli ossidi cresciuti termicamente
- deposizione di un film di fotoresist positivo (steso uniformemente mediante centrifugazione, con velocità di spinnaggio di circa $5 \cdot 10^3 - 10^4$ giri al minuto, ma che dipende dal tipo di resist impiegato)
- precottura del fotoresist ("prebaking")
Questa serve ad evaporare la maggior parte del solvente del resist, così che la matrice polimerica che lo costituisce e il relativo attivatore fotosensibile aderiscano bene all'ossido
- esposizione, alla luce UV, attraverso la maschera 1
La sorgente ultravioletta è costituita, in genere, da una lampada di quarzo contenente vapori di mercurio e xenon. Gli atomi di Hg, dopo essere stati eccitati dalla scarica, emettono (all'atto della diseccitazione) una luce con $\lambda = 436 \text{ nm}$, mentre gli atomi di Xe ne emettono una con $\lambda = 365 \text{ nm}$. La maschera è una lastra di quarzo (il quale è trasparente agli UV) di spessore 3mm. Il cromo (opaco agli UV) vi è steso sopra con un film di spessore dell'ordine nei nm, dove imposto dalle geometrie desiderate. Per ottenere

le dimensioni geometriche previste, è necessario che il sistema di condensatori e proiettori ottici siano calibrati al fine di attenuare il problema della diffrazione della luce.

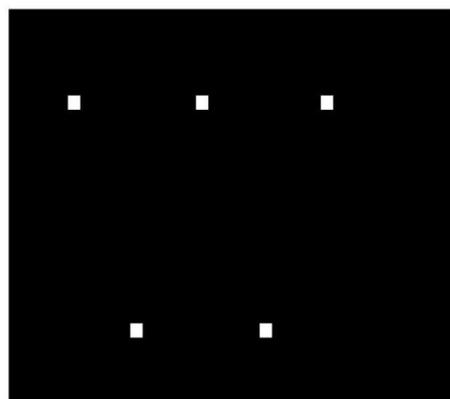
- bagno di sviluppo
Consiste nella rimozione del resist esposto alla luce (il quale è divenuto solubile) immergendolo in opportuna soluzione. In quei punti l'ossido non è più protetto.
- cottura del resist ("postbaking")
Serve ad aumentare la resistenza del resist rimasto. Si effettua a temperature intorno ai 110 – 140 °C
- attacco con acido HF (che è corrosivo per l'ossido ma selettivo verso il Si)
- rimozione del resist
- diffusione termica di boro (oppure impiantazione ionica dello stesso) per convertire localmente il Si in p⁺ (l'ossido presenta un coefficiente di diffusione termica praticamente nullo per la maggior parte degli elementi usati come droganti, boro incluso)
- attacco con acido HF per rimuovere tutto l'ossido rimasto
- crescita di uno strato di ossido termico
- stesura di un film di fotoresist
- precottura
Tempi e temperature di questo passaggio devono tener conto del fatto che adesso ci sono 3 well p⁺, le cui concentrazioni di boro non devono subire grossi cambiamenti (a livello di profilo, profondità di giunzione ecc...), possibili a causa della diffusione termica
- esposizione agli UV attraverso la maschera 2
- bagno di sviluppo
- cottura
- attacco HF
- rimozione del resist
- diffusione termica di fosforo oppure arsenico (oppure impiantazione ionica), al fine di convertire localmente il Si in n⁺
- attacco HF per rimuovere tutto l'ossido rimasto
- crescita di ossido termico o deposizione CVD dello stesso
Le diffusioni termiche precedenti hanno creato well di spessore esiguo (al massimo 4 μm). Ciò è confacente all'esigenza di avere correnti di diffusione, di fotominoritari, quanto più piccole possibili, e quindi di minimizzare la potenza ottica persa. Se utilizziamo l'ossidazione termica, la reazione $\text{Si} + \text{O}_2 \rightarrow \text{SiO}_2$ (dry), oppure $\text{Si} + 2\text{H}_2\text{O}_2 \rightarrow \text{SiO}_2 + 2\text{H}_2$ (wet), provoca la conversione della superficie del Si in ossido, per una profondità pari a metà (circa) dello spessore finale dello strato di SiO_2 . Quindi le well, di per sé molto strette, potrebbero ridursi troppo. È quindi possibile che la CVD, in questo caso, sia da preferire.
- stesura del fotoresist
- precottura
- esposizione agli UV attraverso la maschera 3
- bagno di sviluppo
- cottura
- attacco HF (apertura delle 5 finestre)
- rimozione del resist
- deposizione di un film di alluminio mediante tecniche come l'evaporazione termica, quella ad electron – beam oppure mediante sputtering (oggi le ultime 2 sono le più utilizzate).
L'alluminio ha una resistività ($\approx 2.65 \mu\Omega\text{cm}$) maggiore di quella di conduttori come l'oro ($\approx 2.44 \mu\Omega\text{cm}$) o il rame ($\approx 1.7 \mu\Omega\text{cm}$), ma ha il pregio di aderire molto meglio di questi sia sul Si che sul SiO_2 .
- stesura del fotoresist
- precottura
- esposizione agli UV attraverso la maschera 4 (definizione delle piste di metal)
- bagno di sviluppo
- cottura
- attacco acido (corrosivo per l' Al ma selettivo verso il Si)
- rimozione del resist



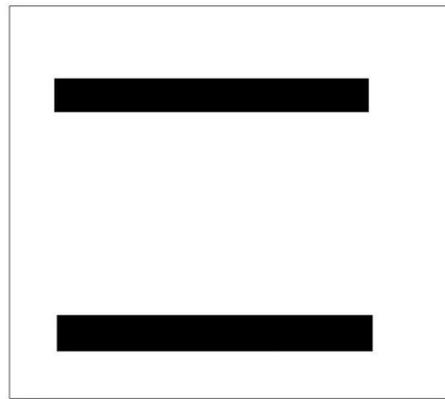
maschera 1



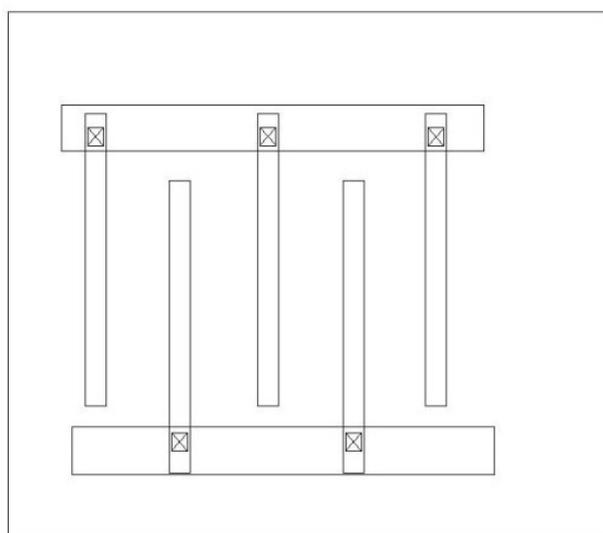
maschera 2



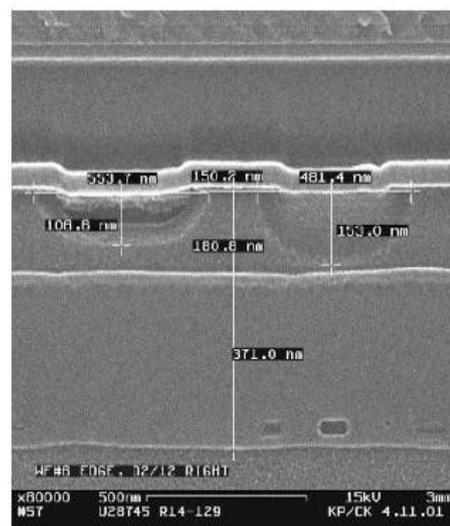
maschera 3



maschera 4



layout



un fotodiodo PIN osservato al microscopio elettronico a scansione

Fig. 80

Riportiamo le quattro maschere utilizzate durante l'integrazione del PIN al Si, il suo layout (leggermente ingrandito rispetto all'effettiva sovrapposizione delle quattro maschere che lo costituiscono) ed un'immagine del dispositivo ottenuta al microscopio elettronico a scansione.

5.12) La proporzionalità inversa tra l'energia di gap E_{gap} e la parte reale $n_r'(\omega)$ dell'indice di rifrazione di un semiconduttore: spiegazione formale

Nel grafico della compatibilità reticolare fra semiconduttori, mostrato in figura 58, possiamo notare che l'energia di gap E_{gap} è inversamente proporzionale alla parte reale $n_r'(\omega) = \text{Re}\{n_r(\omega)\}$ dell'indice di rifrazione $n_r(\omega)$. È interessante spiegarne brevemente il motivo. Ammettiamo, per semplicità espositiva, di avere un sistema elettronico che consta unicamente di due livelli energetici E_c ed E_v , come mostrato nella parte sinistra di figura 80. Abbiamo quindi un materiale, ovviamente solo teorico (non esistente in natura), in cui il complesso sistema energetico a bande $E(k)$, tipico dei cristalli reali, è rappresentato, qualitativamente, da un livello energetico fondamentale E_v , dove gli elettroni sono stabili, cioè il loro tempo di vita medio è infinito, ed un livello energetico metastabile E_c , dove gli elettroni sono stabili solo per breve tempo, cioè il loro tempo di vita medio è finito. I due livelli sono separati da un gap di energia pari a $E_{\text{gap}} = \hbar\omega_0$, analogamente a quanto accade nei semiconduttori, dove il minimo della BC ed il massimo della BV sono appunto separati da un gap E_{gap} caratteristico di ciascun semiconduttore. Supponiamo di inviare, sul materiale energeticamente rappresentato dal semplice sistema a due livelli E_c ed E_v , un fotone di energia $\hbar\omega \neq \hbar\omega_0$. È evidente che non esiste alcuno stato elettronico permesso in cui è localizzato un elettrone atomico in grado di assorbire l'energia ottica $\hbar\omega$ e altresì non esiste alcuno stato elettronico permesso che può accogliere l'elettrone atomico, dopo che questo ha scatterato con il fotone incidente. In altre parole non esiste nessuna coppia di stati elettronici permessi $|i\rangle$ e $|f\rangle$ separati da un'energia pari a $\hbar\omega$, ovvero l'energia fotonica incidente $\hbar\omega$ non può accoppiare nessuno stato iniziale $|i\rangle$, avente energia E , con nessuno stato finale $|f\rangle$, avente energia $E + \hbar\omega$. Quindi il materiale presenta, nei confronti del fotone, una trasparenza passiva che consente a quest'ultimo di attraversare l'intero cristallo senza interagire con gli elettroni atomici. Il materiale, per qualunque pulsazione ottica incidente $\omega \neq \omega_0$, è trasparente, ovvero non assorbente, ed i suoi elettroni non scatterano in alcun modo con i fotoni di energia $\hbar\omega \neq \hbar\omega_0$, come del resto risulta evidente applicando rigorosamente la regola d'oro di Fermi. In altri termini il materiale non si oppone in alcun modo al suo attraversamento da parte della suddetta radiazione luminosa. Se invece irradiamo il materiale con un fotone di energia $\hbar\omega = \hbar\omega_0$ e polarizzato secondo il versore \hat{a} , la probabilità che un elettrone, posizionato sul livello energetico fondamentale (stabile) E_v , avente quantità di moto \vec{p} , scatteri con il fotone e venga promosso sul livello energetico metastabile $E_c = E_v + \hbar\omega_0$, cioè la probabilità che il cristallo assorba il fotone, non è nulla. Tale probabilità dipende dall'overlap fra \hat{a} e \vec{p} , cioè dall'elemento di matrice P della quantità di moto dell'elettrone, la cui espressione è, come noto, la seguente:

$$P = |\langle f | \hat{a} \cdot \vec{p} | i \rangle|^2 = \left| \int \psi_f(r)^* \hat{a} \cdot \vec{p} \psi_i(r) d^3r \right|^2$$

Pertanto lo spettro di assorbimento ottico specifico $\alpha(\omega)$ del materiale è costituito da una delta di Dirac $\delta(\omega - \omega_0)$, posizionata appunto presso la pulsazione ottica ω_0 , mentre risulta nullo per qualunque valore di ω diverso da ω_0 . Ciò significa che per $\omega \neq \omega_0$ il materiale non è assorbente, ovvero non si oppone in alcun modo al flusso fotonico che lo attraversa ($\alpha(\omega \neq \omega_0) = 0$, opposizione alla luce incidente nulla per $\omega \neq \omega_0$), mentre per $\omega = \omega_0$ il materiale è assorbente e pertanto si oppone in modo perfetto alla luce incidente, "distruggendo" appunto il fotone ($\alpha(\omega = \omega_0) = \delta(\omega = \omega_0) \rightarrow \infty$, opposizione alla luce incidente infinita, perfetta, per $\omega = \omega_0$). Lo spettro di assorbimento ottico specifico $\alpha(\omega)$ del materiale, in osservanza alla formula riportata nella [e25], è proporzionale alla parte immaginaria $n_r''(\omega) = \text{Im}\{n_r(\omega)\}$ dell'indice di rifrazione $n_r(\omega)$ del mezzo, e conseguentemente alla parte immaginaria $\varepsilon_2(\omega) = \text{Im}\{\varepsilon(\omega)\}$ della costante dielettrica relativa del mezzo, come si può osservare nella parte destra di figura 81.

Per calcolare lo spettro della parte reale $\varepsilon_1(\omega) = \text{Re}\{\varepsilon(\omega)\}$ della costante dielettrica relativa del materiale, anch'essa riportata nella parte destra di figura 81, dove $\varepsilon_1(\omega)$ è proporzionale alla parte reale $n_r'(\omega) = \text{Re}\{n_r(\omega)\}$ dell'indice di rifrazione $n_r(\omega)$ del cristallo, è necessario invocare l'espressione di Kramers - Kronig che lega, dato un certo materiale, per noi un semiconduttore, lo spettro di $n_r'(\omega)$ con quello dell'assorbimento ottico specifico $\alpha(\omega)$. In formule:

$$n_r'(\omega_0) - 1 = \frac{c}{\pi} P \int_0^\infty \frac{\alpha(\omega)}{\omega^2 - \omega_0^2} d\omega \quad [e74]$$

dove c è la velocità della luce e P è l'elemento di matrice della quantità di moto dell'elettrone in procinto di scatterare con il fotone incidente, il quale possiede energia pari ad $\hbar\omega$. La struttura energetica a due livelli del materiale preso in esame impone, come già spiegato, che lo spettro dell'assorbimento ottico specifico $\alpha(\omega)$

assuma l'andamento a delta di Dirac, colorato in blu nella parte destra di figura 81. Pertanto la [e74] assume la seguente forma:

$$n_r'(\omega_0) - 1 = \frac{c}{\pi} P \int_0^\infty \frac{\delta(\omega - \omega_0)}{\omega^2 - \omega_0^2} d\omega \sim \frac{c}{\omega_0}$$

da cui:

$$\epsilon_1(\omega) = \text{Re}\{\epsilon(\omega)\} \propto n_r'(\omega) = \text{Re}\{n_r(\omega)\} \propto 1 + \frac{c}{\omega_0}$$

Lo spettro di $\epsilon_1(\omega)$ assume pertanto l'andamento approssimato a forma di ramo di iperbole equilatera, colorato in rosso nella parte destra di figura 81.

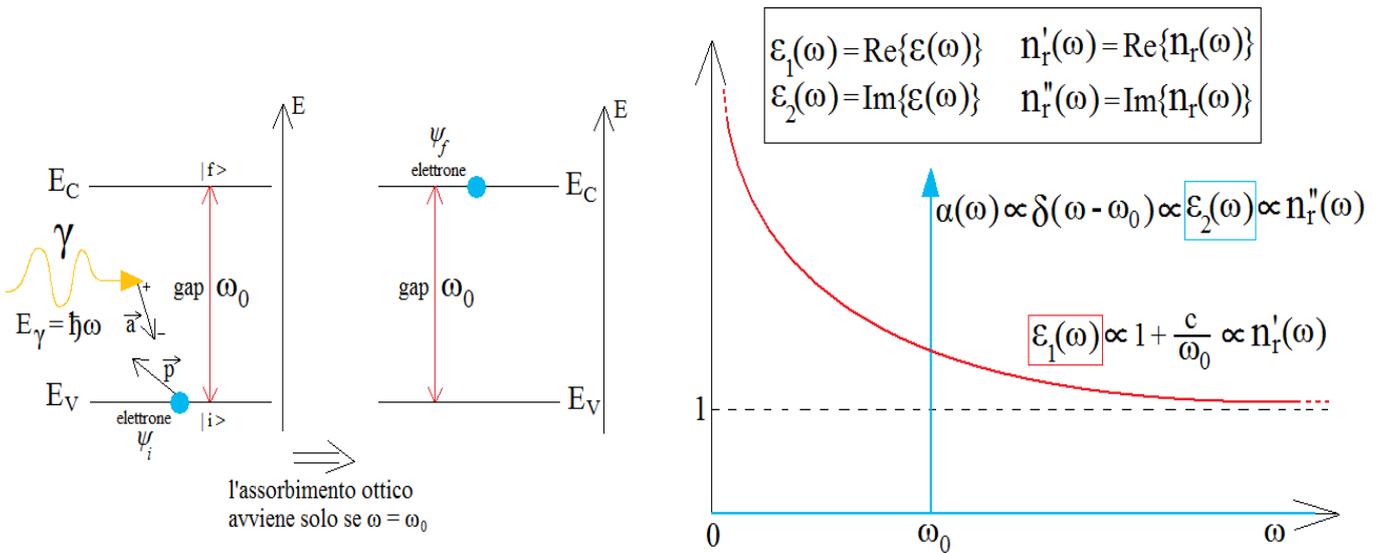


Fig. 81

A sinistra è mostrata la semplice struttura energetica a due livelli (uno inferiore, stabile, ed uno superiore, metastabile) di un ipotetico materiale irradiato da un fotone di energia $\hbar\omega$. Il fotone è polarizzato secondo un vettore di polarizzazione \hat{a} , mentre l'elettrone, in procinto di scatterare con il fotone, possiede un impulso \vec{p} . A destra sono mostrati gli spettri della parte reale $\epsilon_1(\omega)$ e della parte immaginaria $\epsilon_2(\omega)$ della costante dielettrica relativa $\epsilon(\omega)$ dell'ipotetico materiale in esame. Adattando il modello elettronico a due livelli di energia a quello, molto più complesso, del diagramma a bande E - k di un semiconduttore, si evince chiaramente la proporzionalità inversa che sussiste fra $E_{\text{gap}} = \hbar\omega_0$ e la parte reale $n_r'(\omega)$ dell'indice di rifrazione $n_r(\omega)$ di un qualunque semiconduttore.

Il divario energetico $\hbar\omega_0$ fra i due livelli E_C ed E_V , come già accennato, può assumere il significato fisico di gap di energia $E_{\text{gap}} = \hbar\omega_0$ caratteristico del diagramma a bande E - k in un semiconduttore. Pertanto per materiali cristallini aventi un gap $E_{\text{gap}} = \hbar\omega_0$ molto grande, cioè una pulsazione ottica di assorbimento (in generale di transizione ottica) ω_0 molto grande, la parte reale $\epsilon_1(\omega) = \text{Re}\{\epsilon(\omega)\}$ della costante dielettrica relativa del materiale tende ad 1, ovvero al suo valore minimo, pertanto la parte reale $n_r'(\omega) = \text{Re}\{n_r(\omega)\}$ dell'indice di rifrazione del materiale tende ad 1, che rappresenta, come noto, il suo valore minimo. Il valore di $\epsilon_1(\omega)$, e quindi di $n_r'(\omega)$, è esattamente pari ad 1 solo nel caso in cui $E_{\text{gap}} = \hbar\omega_0$ tende all'infinito ($\omega_0 \rightarrow \infty$), e ciò accade in un ambiente privo di stati elettronici permessi, ovvero nel vuoto. Viceversa più ω_0 è piccola, ovvero più l'energia di gap $E_{\text{gap}} = \hbar\omega_0$ del materiale è piccola, e più $\epsilon_1(\omega)$, e quindi $n_r'(\omega)$, è grande. Dunque la proporzionalità inversa fra E_{gap} e $n_r'(\omega) = \text{Re}\{n_r(\omega)\}$, con l'ausilio di un semplice modello elettronico a due livelli di energia, è stata dimostrata.

6) I fotodiodi a valanga (APDs - "Avalanche Photodiodes")

6.1) Le caratteristiche I - V di un fotorivelatore a valanga, sottoposto ad illuminazione; rivelazioni fotovoltaiche e fotoconduttive

In figura 82 è riportato il circuito equivalente (per ora ipersemplificato) di un APD:

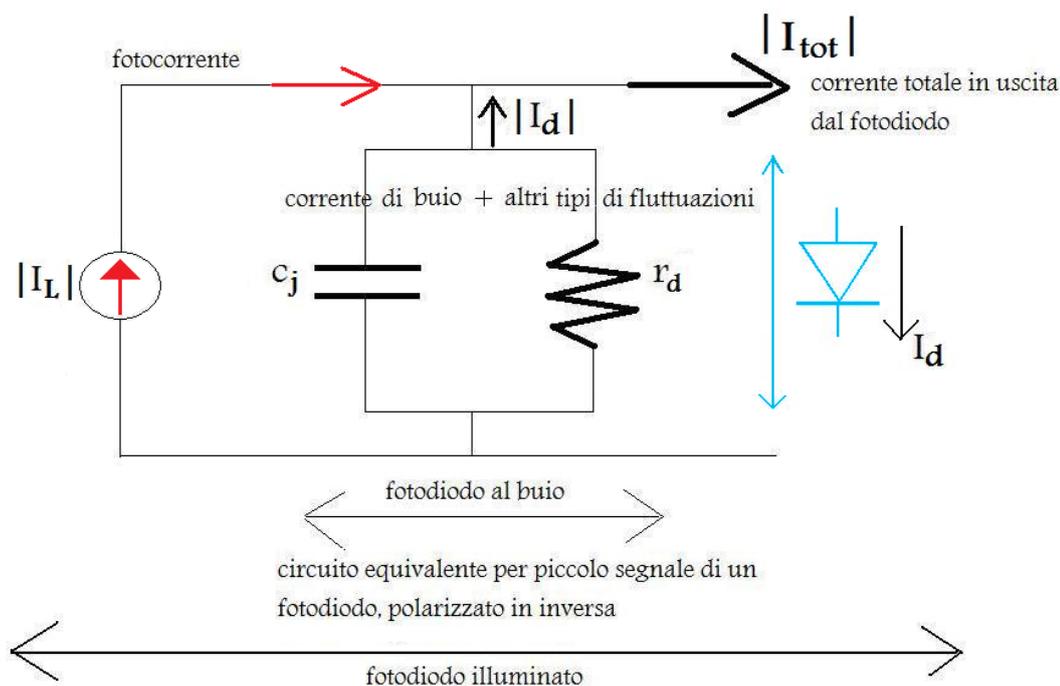


Fig. 82

Circuito equivalente ipersemplificato, per piccolo segnale, di un APD sottoposto ad illuminazione

Il fotodiodo illuminato viene "scomposto" in una sorgente indipendente di fotocorrente, che implementa il segnale utile I_L , e nel parallelo RC, che rende conto della corrente di rumore I_d . Quest'ultima porzione di schema non è altro che l'equivalente per piccolo segnale di un diodo, polarizzato in inversa. r_d è la resistenza differenziale che la giunzione oppone al passaggio della corrente di buio, mentre c_j è la capacità di giunzione (più precisamente è una capacità su unità di superficie attiva, e si misura in F/m²), che nel caso di polarizzazione inversa coincide con la capacità relativa alla zds^* . In formule:

$$I_d = I_0 \left(e^{\frac{V}{\eta V_T}} - 1 \right), \quad \text{con } V_T = \frac{kT}{q} \approx 26\text{mV}$$

Il coefficiente di idealità η lo supponiamo prossimo ad 1 (1.02 ÷ 1.05 è un range tipico), ipotizzando che l'APD sia costituito da un'omogiunzione, nella quale i mismatches reticolari interfacciali sono decisamente trascurabili.

$$r_d = \left(\frac{dI_d}{dV} \right)^{-1} \quad c_j = \frac{\epsilon_S}{zds(V)}$$

La caratteristica I - V dell'APD, in condizione di illuminazione, è modificata dalla fotocorrente I_L , che è legata all'intensità della luce incidente, come mostrato in figura 83:

$$I_{\text{totale}}(V) = I_0 \left(e^{\frac{V}{\eta V_T}} - 1 \right) - I_L$$

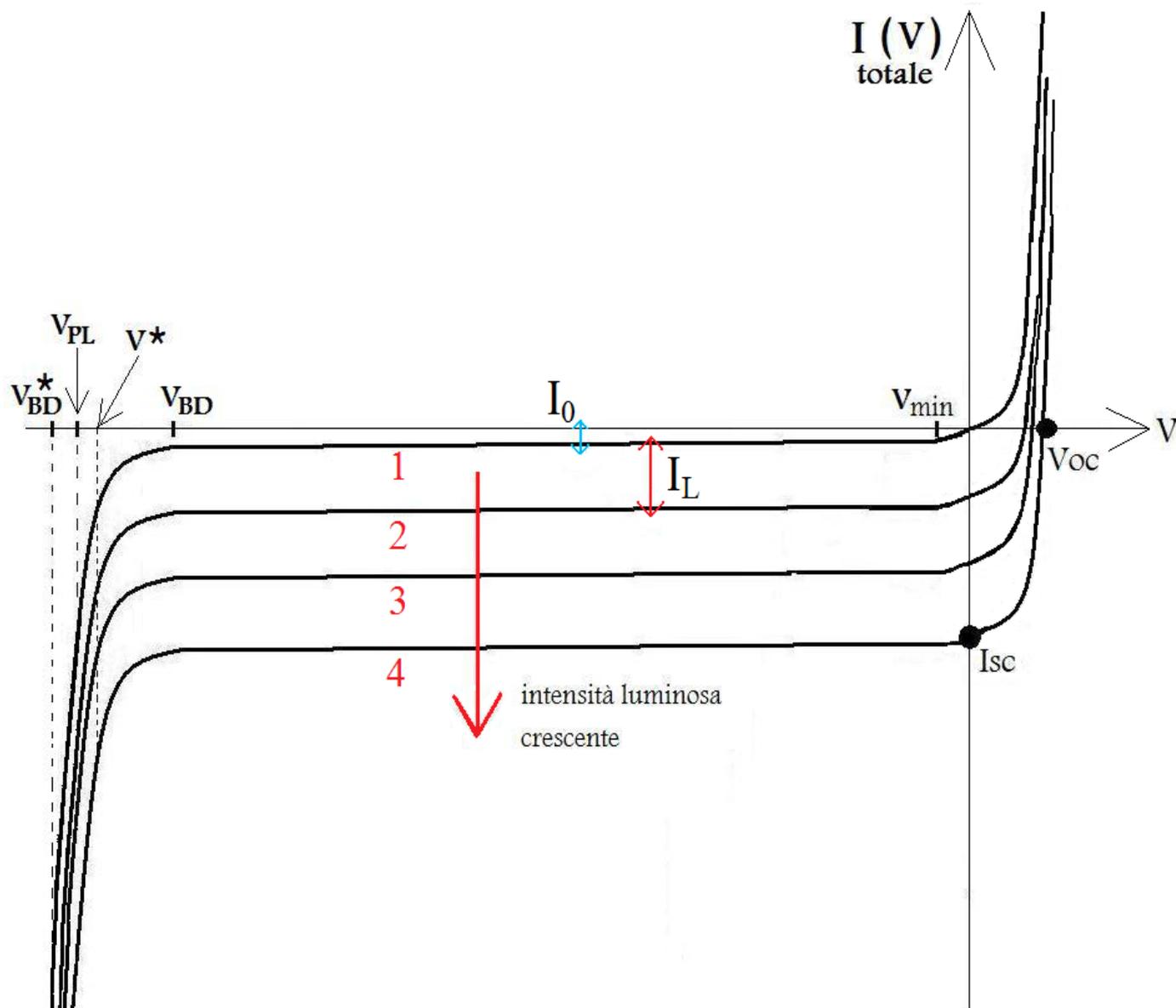


Fig. 83

Curve caratteristiche I - V di un fotodiode a giunzione pn, parametrizzate dall'intensità luminosa incidente sul dispositivo.

Le curve caratteristiche mostrate sono valide anche per un fotodiode PIN, ovviamente per $|V|$ compreso fra 0 e V_{BD} . La caratteristica 1 è quella tipica di un diodo, ovvero in condizione di buio: essendo $I_L = 0$, dai terminali metallici esce $I_{totale} = I_d$, cioè la sola corrente di dark - bulk nel caso di polarizzazione inversa (I_0 nel caso di polarizzazione inversa non forte), oppure la sola corrente di iniezione di minoritari, nel caso di polarizzazione diretta. All'aumentare del numero di fotoni incidenti, con V fissata, aumenta I_L , quindi I_{totale} e pertanto la curva caratteristica è sempre più traslata verso il basso. Si notino le intercette di una qualsiasi delle curve con gli assi, ovvero la corrente di cortocircuito I_{sc} , teoricamente misurabile chiudendo il dispositivo su un amperometro ideale, e la tensione di circuito aperto V_{oc} , misurabile chiudendolo su un voltmetro ideale. Strumenti effettivamente utilizzabili a tale scopo, reperibili in un qualsiasi laboratorio di elettronica, sono il picoamperometro ed il nanovoltmetro.

A seconda di quale tipo di misurazione elettrica eseguiamo ai terminali del fotodiode, possiamo distinguere due tipi di fotorelevazione:

- fotovoltaica
- fotoconduttiva

La prima consiste nel chiudere il fotodiiodo illuminato su un voltmetro di alta qualità, avente idealmente impedenza d'ingresso infinita. Quindi la corrente in uscita dal fotodiiodo è di fatto nulla:

$$I_0 \left(e^{\frac{V_{OC}}{nV_T}} - 1 \right) - I_L = 0$$

da cui si ricava che:

$$V_{OC} = V_T \ln \left(\frac{I_L}{I_0} + 1 \right)$$

Una volta che il voltmetro ha fornito il risultato V_{OC} della misurazione, è immediato, anche se non molto agevole, ricavare il valore della fotocorrente I_L .

La seconda fotorivelazione consiste nel chiudere il fotodiiodo illuminato su un amperometro di alta qualità, avente idealmente impedenza d'ingresso nulla. La tensione ai terminali del fotodiiodo è di fatto nulla, per cui:

$$|I_{totale}(V = 0)| = I_{SC} = I_L$$

Lo strumento di misura, stavolta, fornisce direttamente il valore della grandezza elettrica di nostro interesse.

6.2) Le correnti di rumore di un fotodiiodo a valanga: origine fisica delle fluttuazioni stocastiche e relative densità spettrali di potenza di rumore

Il circuito equivalente di figura 82 può essere reso un po' più realistico mediante il modello riportato in figura 84:

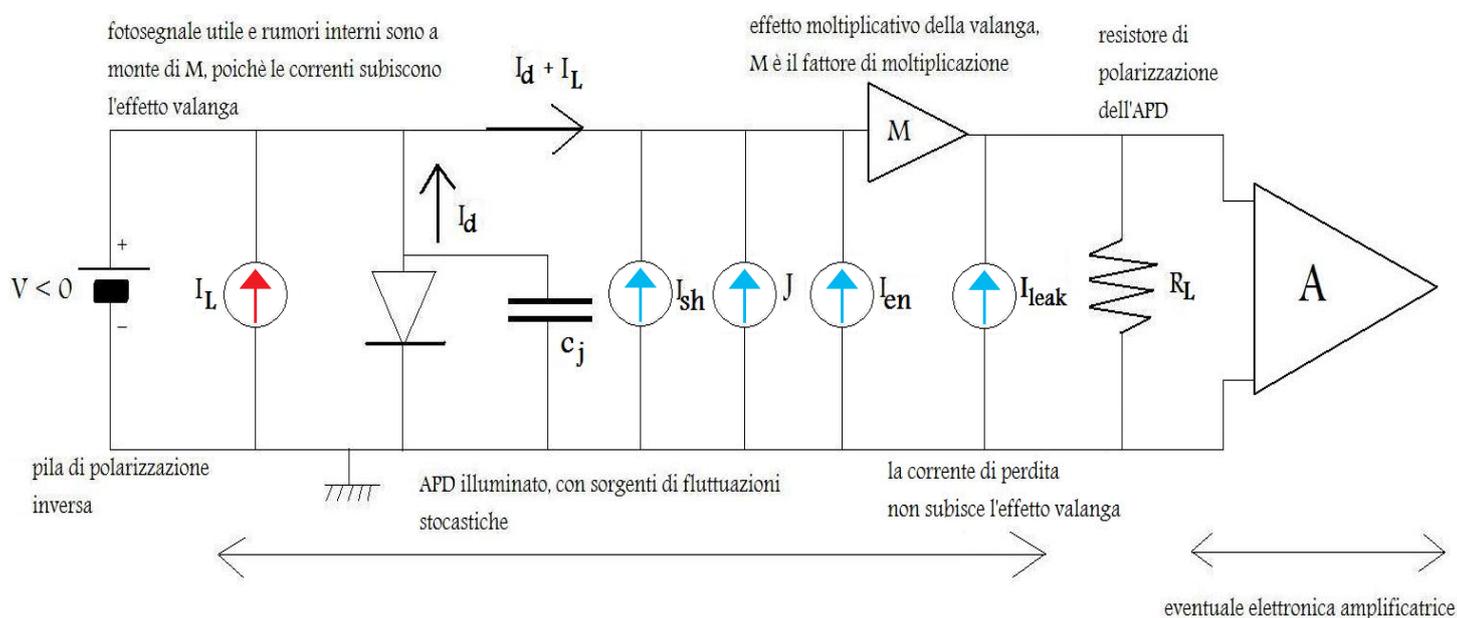


Fig. 84

Circuito equivalente di un APD illuminato (valido anche per un PIN, con $M = 1$) e dell'elettronica di front-end a questo connessa. Il sistema, nel suo complesso, può rappresentare l'equivalente circuitale di un "ricevitore convenzionale" adottato, nell'ambito delle comunicazioni ottiche su fibra, agli inizi degli anni 70, oppure l'equivalente circuitale del primo stadio (stadio di fotorivelazione) di un sistema PET. È possibile notare la pila di polarizzazione inversa, la capacità c_j relativa alla zona di svuotamento della giunzione, il generatore che implementa la fotocorrente I_L (prodotta in risposta all'evento luminoso incidente), alcune delle più importanti sorgenti di rumore del dispositivo, come la corrente di rumore shot I_{sh} , la corrente di rumore termico J , la corrente di rumore dovuta all'effetto dell'elettronica amplificatrice esterna I_{en} , l'amplificatore M che rende conto del fattore di moltiplicazione dovuto al breakdown a valanga, la corrente di leakage I_{leak} , il resistore R_L che converte il segnale di corrente, in uscita dall'APD, in uno di tensione, l'amplificatore esterno A . Quest'ultimo, se approssimato ad un sistema L.C.T.I. (Lineare - Continuo - TempoInvariante), possiede una risposta in frequenza $A(f)$ localizzata in banda base: la banda di A (B_{RX}) è molto minore della banda della risposta in frequenza $H(f)$ dell'APD (B_H).

Abbiamo una giunzione polarizzata in inversa (con in parallelo la solita capacità relativa alla zds^*), dalla quale, in virtù della polarizzazione, proviene una corrente di buio I_d (corrente di dark - bulk). Questa si somma alla corrente di fotoconduzione I_L , pertanto la corrente totale $I_L + I_d$, a causa del campo elettrico elevato, viene moltiplicata per un fattore $M \gg 1$ (generalmente $10 < M < 100$). Abbiamo separato, a livello di simbolismo circuitale, le cause fisiche, quali estrazioni di cariche ed assorbimenti di fotoni, delle correnti che concorrono alla formazione di quella totale, dall'effetto valanga che si ha nella zds^* (eventi di ionizzazione), ponendo tutti i generatori indipendenti, che implementano quelle correnti, a monte dell'amplificatore avente guadagno M . Come sappiamo I_d è data dal contributo di due correnti: una (I_d') dovuta all'estrazione di minoritari, l'altra (I_d'') alla ionizzazione da impatto (termogenerazione di coppie e/h). Tuttavia sia l'una che l'altra sono interessate da fluttuazioni non deterministiche (rumori) intrinseche ai meccanismi fisici che costituiscono le cause di tali correnti. A I_d' è associata una fluttuazione a media nulla i_d' , la quale rende ragione del fatto che, a istanti diversi, il campo estrae quantità diverse di minoritari. I_d' rappresenta il "valor medio" dei minoritari estratti, mentre i_d' costituisce una successione, nel tempo, di correnti impulsive, frutto della casuale e istantanea variazione del numero di minoritari estratti. Si noti che l'impulsività di i_d' è conseguenza della natura granulare della corrente. Questa successione di impulsi reali è rappresentabile, matematicamente, come una successione di delta di Dirac "filtrate" dal semiconduttore, ragion per cui la densità spettrale di potenza (A^2/Hz) del rumore i_d' è bianca (i_d' è un "rumore gaussiano bianco"). Anche I_d'' è interessata da rumore gaussiano bianco i_d'' , poiché il numero di portatori termogenerati per impatto, tutti derivanti da un singolo portatore estratto da un bulk, è stocastico. Nell'istante in cui una carica, eventualmente estratta da un bulk, produce una ionizzazione per impatto, l'elettrone secondario termogenerato, superando la barriera di potenziale rappresentata dal gap di energia, produce di per sé un impulso di corrente (si noti, anche in questo caso, la granularità di cui sopra). Quest'ultimo si crea anche quando una coppia è termogenerata spontaneamente, e quindi anche in assenza di campo. Poniamo $i_d = i_d' + i_d''$: se i_d' e i_d'' sono considerati, anche se non è vero, come processi stocastici indipendenti, la densità spettrale di potenza di i_d ($S_{i_d}(f)$) è la somma delle densità spettrali di potenza dei suoi due contributi. $S_{i_d}(f)$ ha l'andamento di figura 85.

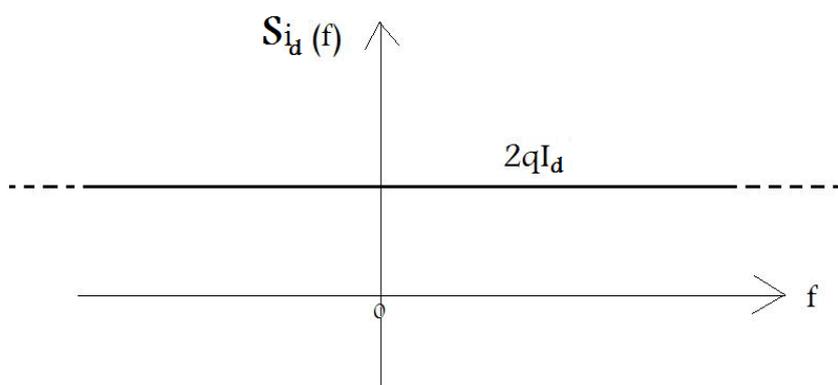


Fig. 85

Densità spettrale di potenza (A^2/Hz) della corrente di dark - bulk

Anche alla fotocorrente I_L , anch'essa granulare, è associato un rumore impulsivo i_L , dato che un elettrone fotogenerato, che supera la barriera dell' E_{gap} , produce un impulso di corrente. i_L è un rumore gaussiano bianco, la cui densità spettrale di potenza $S_{i_L}(f)$ è mostrata in figura 86.

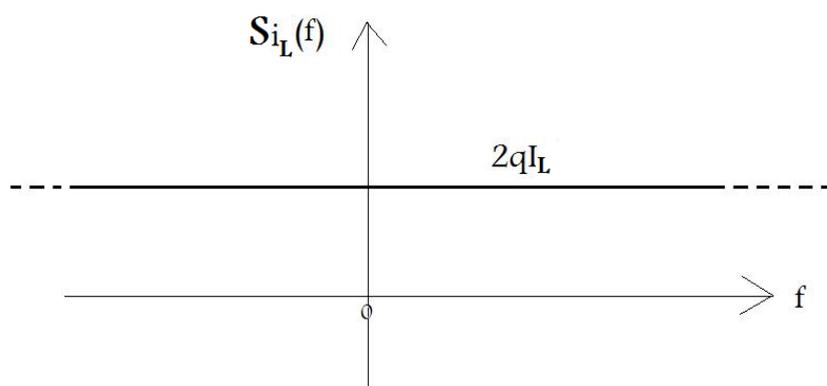


Fig. 86

Densità spettrale di potenza (A^2/Hz) della fotocorrente

L'amplificatore esterno A è costituito da una rete comprendente elementi passivi ed attivi (quest'ultimi sono transistori e diodi); dato che gli elementi attivi lavorano in intorno molto piccoli dei loro punti di riposo, ovvero le ampiezze delle variazioni informative, cioè i piccoli segnali, sono trascurabili rispetto alle ampiezze delle grandezze costanti che polarizzano i dispositivi attivi, possiamo linearizzarne il comportamento. Quindi A è approssimabile con un sistema L.C.TI. (Lineare - Continuo - TempoInvariante), per cui è possibile associargli una risposta in frequenza $A(f)$, localizzata in banda base: la banda di A , che indichiamo con B_{RX} , è molto minore della banda optoelettronica, che indichiamo con B_H , della risposta in frequenza $H(f)$ associata all'APD.

Ipotizzando che l'amplificatore esterno A veda, ai suoi terminali d'ingresso, un'impedenza approssimabile con la resistenza di polarizzazione (ovvero di lettura) R_L , e che l'impedenza di ingresso di A possa essere considerata infinita rispetto ad R_L , allora, in osservanza al teorema di Nyquist, è possibile tener conto del rumore termico relativo ad R_L , che è percorsa dalla corrente totale in uscita dall'APD, mettendo in serie alla stessa un generatore di tensione $e_T(t)$. Questo rappresenta il rumore termico (gaussiano bianco, a valor medio nullo), la cui densità spettrale di potenza $S_T(f)$ è pari a $4kTR_L$ ed il cui andamento è mostrato in figura 87.

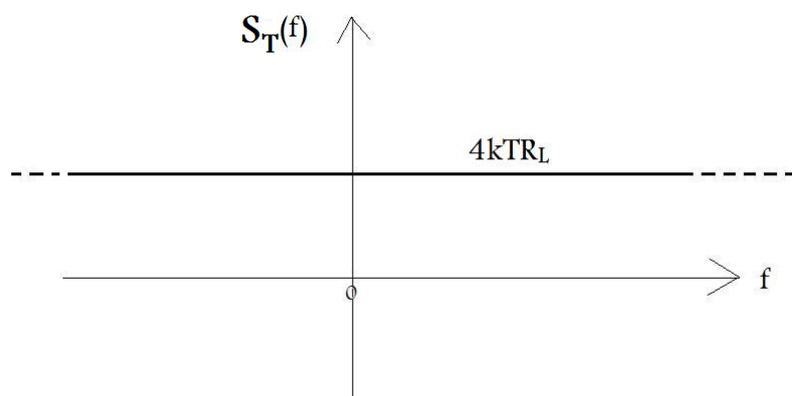


Fig. 87

Densità spettrale di potenza (V^2/Hz) della tensione $e_T(t)$ di rumore termico

Chiamiamo "corrente di rumore shot" dell'APD la quantità $I_{sh} = i_d + i_L$. Supponiamo che i due termini siano processi indipendenti, per cui $S_{I_{sh}}(f) = S_{i_d}(f) + S_{i_L}(f) = 2q(I_L + I_d)$.

Supponiamo che l'APD, avente banda optoelettronica B_H , riceva in ingresso una radiazione luminosa avente potenza ottica utile P (un valore tipico potrebbe essere dell'ordine del μW), ad esempio proveniente da una fibra ottica a cui è connesso il fotorivelatore. Questo segnale luminoso incidente ha una frequenza centrale f_0 (ad

esempio 345 THz, quindi è un segnale trasmesso in prima finestra), compresa all'interno della banda B_{RX} di A ($B_{RX} \ll B_H$), una larghezza di banda Δf (≈ 10 MHz) e, nel caso di condizioni ideali di assorbimento totale, produce una corrente di fotoconduzione pari a:

$$I_L = M \frac{Pq}{hf_0}$$

In realtà I_L sarà minore, per un fattore $QE < 1$ (l'efficienza quantica del dispositivo APD):

$$I_L = QE M \frac{Pq}{hf_0}$$

Un valore possibile potrebbe essere $QE = 0.8$, per cui l'80% della potenza ottica è effettivamente assorbita, e trasdotta in un segnale elettrico di uscita. In altre parole in risposta ad un segnale ottico incidente monocromatico formato da 100 fotoni misuriamo mediamente, ai terminali dell'APD, il contributo di carica relativo a 80 coppie e/h.

Avremo pertanto sia I_L che I_d accompagnate dalla corrente di rumore shot I_{sh} . Questa avrà una potenza N_I (potenza di rumore di corrente) pari a:

$$N_I = \int_{-B_{RX}}^{+B_{RX}} S_{I_{sh}}(f) df = 2q(I_L + I_d)B_{RX} \quad [A^2]$$

J rappresenta, invece, la corrente di "rumore Johnson", cioè di rumore termico. Il rumore termico, a temperatura superiore allo zero assoluto, anche in assenza di campo elettrico, trova origine dal moto casuale dei portatori nel conduttore, i quali hanno energia media distribuita intorno a kT . La sua potenza di rumore di corrente N_E , sulla banda B_{RX} , vale:

$$N_E = \int_{-B_{RX}}^{+B_{RX}} S_J(f) df = \int_{-B_{RX}}^{+B_{RX}} \frac{2kT}{R_L} df = 2S_J(f)B_{RX} = \frac{4kT}{R_L} B_{RX} \quad [A^2]$$

dove l'espressione di $S_J(f)$ è ricavata dal passaggio dal circuito equivalente di Thevenin, nel quale $e_T(t)$ è in serie ad R_L (circuito che implementa il teorema di Nyquist) a quello di Norton, in cui $J(t)$ è in parallelo ad R_L . In termini più rigorosi (benchè equivalenti, in condizioni e per amplificatori standard), al posto di R_L dovremmo mettere la "resistenza efficace", data dal parallelo fra R_L ed r_d , mentre al posto della temperatura assoluta T , a cui si trova R_L , ci vorrebbe quella efficace T_{eff} , legata alla cifra di rumore NF ("Noise Figure") dell'amplificatore esterno A (si veda il glossario), secondo la seguente relazione:

$$T_{eff} = T(10^{\frac{NF}{10}} - 1)$$

Il generatore I_{en} tiene conto del rumore introdotto dall'amplificatore esterno A; il fatto di averlo posto a monte di M è solo un'equivalenza circuitale, ininfluenza per il calcolo degli effetti esterni, ma la sua origine è a valle dell'APD. Tale rumore è determinato, principalmente, dall'effetto di capacità parassite, nel caso in cui A amplifichi una corrente, prodotta dall'APD, ad alte frequenze, dall'effetto di processi Johnson e shot, nel caso di corrente a frequenze intermedie, oppure dal rumore Flicker, nel caso di corrente a basse frequenze. L'importanza dei singoli contributi di I_{en} può essere determinata, qualitativamente, osservando il grafico bilogarithmico della cifra di rumore NF (grafico "a vasca da bagno"), riportato in figura 88, relativo al quadripolo attivo della rete amplificatrice A in uso.

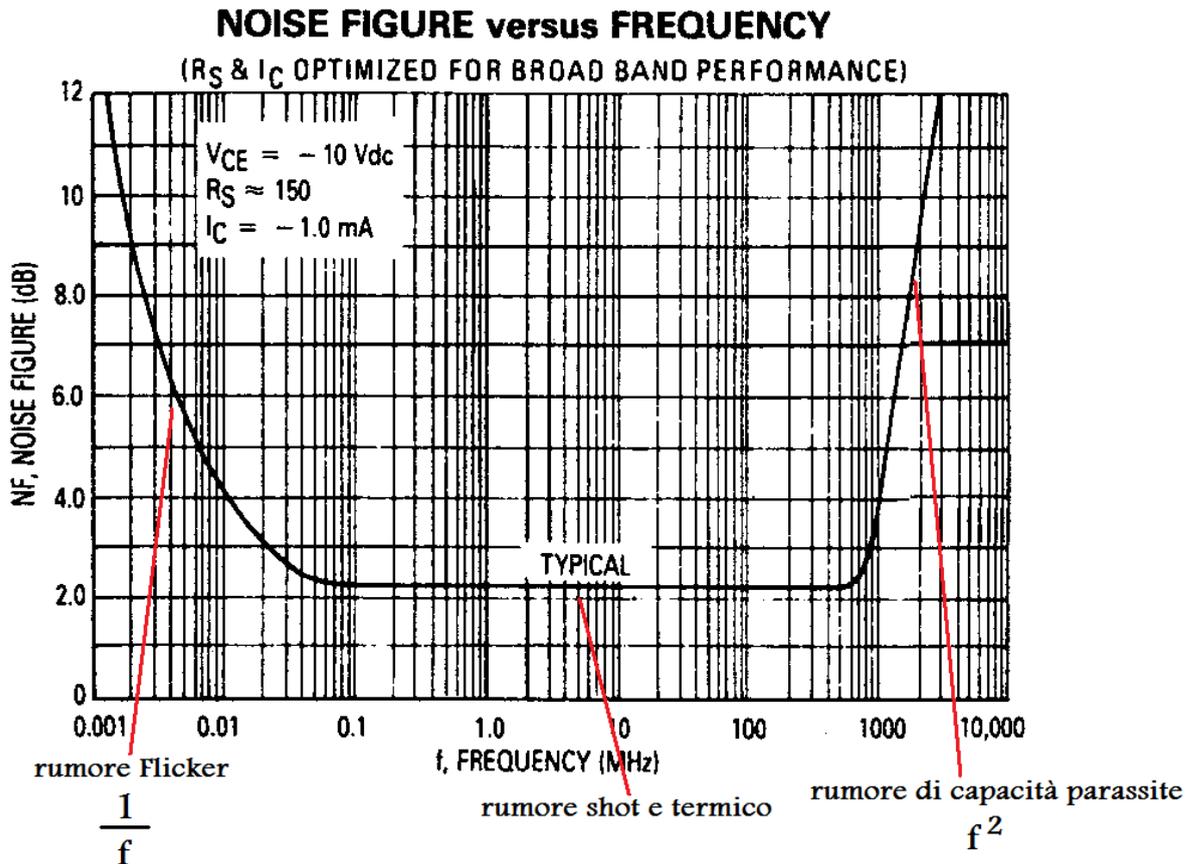


Fig. 88

Grafico della cifra di rumore, in funzione della frequenza, del transistor bipolare pnp 2N4957, della Motorola, teoricamente utilizzabile come quadripolo attivo all'interno di un circuito di amplificazione esterna A. Il grafico del BJT citato è riportato solo a titolo di esempio: si noti, infatti, che una frequenza portante di 345 THz (rientrante nella prima finestra) non è neppure compresa sull'asse delle frequenze, dunque un BJT così obsoleto non potrebbe funzionare per i nostri scopi, dal momento che il segnale utile annegherebbe nel rumore dovuto alle capacità parassite del BJT.

L'amplificatore esterno A, oltre a sporcare il segnale utile I_L , rallenta l'intero sistema, così che le costanti di tempo della rivelazione non sono più quelle tipiche dell'APD, bensì quelle dell'elettronica esterna. Ad I_{en} è legata una potenza di rumore elettronico $N_{en} = S_{en}(f)B_{RX}$ (assumendo che $S_{en}(f)$ sia costante su B_{RX} , e ciò accade nel caso in cui lo sia anche la cifra di rumore di A). Un livello tipico per la densità spettrale di potenza suddetta è:

$$S_{en}(f) = \left(\frac{180 \text{ pA}}{\sqrt{\text{Hz}}} \right)^2$$

Anche la corrente di perdita superficiale possiede un valor medio I_{leak} , a cui si sovrappone una fluttuazione stocastica i_{leak} , dovuta alla granularità e alla casualità con le quali i fotoportatori, primari e secondari, ed i portatori estratti dai bulk (e termogenerati da questi, per impatto) vengono catturati dai difetti reticolari, soprattutto da quelli superficiali. Anche i_{leak} è un rumore gaussiano bianco, a cui è associata una densità spettrale di potenza $S_{leak}(f)$ ed una potenza di rumore N_{leak} :

$$N_{leak} = \int_{-B_{RX}}^{+B_{RX}} S_{leak}(f) df = 2qI_{leak}B_{RX} \quad [A^2]$$

Ricordiamo che per un processo stocastico "stazionario" $x(t)$, al quale sono associabili "forme di stazionarietà" quali il valor medio ed il valor quadratico medio, il valore efficace X^* soddisfa la seguente relazione:

$$(X^*)^2 = \langle x(t)^2 \rangle = \frac{1}{\tau} \int_{-\frac{\tau}{2}}^{+\frac{\tau}{2}} x(t)^2 dt = \text{Potenza di } x(t) = \int_{-B_{RX}}^{+B_{RX}} S_x(f) df$$

Se $x(t) \in \mathcal{N}(0, \sigma_x^2)$, ovvero se $x(t)$ è una variabile aleatoria gaussiana a valor medio nullo, allora il suo valor quadratico medio coincide con la sua varianza:

$$\langle x(t)^2 \rangle = \sigma_x^2$$

σ_x rappresenta la deviazione standard di $x(t)$. τ è l'intervallo temporale di osservazione, che deve essere sufficientemente lungo. A questo punto possiamo ricavare i valori efficaci delle correnti di rumore suddette:

$$I_{sh}^* = \sqrt{2q(I_L + I_d)B_{RX}}$$

$$J^* = \sqrt{\frac{4kTB_{RX}}{R_L}}$$

$$I_{en}^* = \sqrt{S_{en}(f)B_{RX}}$$

$$I_{leak}^* = \sqrt{S_{leak}(f)B_{RX}}$$

Possiamo calcolare, infine, il valore efficace della corrente totale di rumore dell'APD, ricordando che il rumore è un fenomeno statistico, per cui i diversi contributi devono essere sommati quadraticamente, in termini di potenza e non di ampiezza.

$$I_{tot}^* = \sqrt{(I_{sh}^*)^2 + (J^*)^2 + (I_{en}^*)^2 + (I_{leak}^*)^2}$$

6.3.1) Modellizzazione matematica della fotorivelazione intrinseca di un APD

A questo punto possiamo calcolare il rapporto segnale/rumore in uscita dall'amplificatore esterno A e ciò equivale, avendo riportato a monte di M l'effetto di rumore dovuto ad A, a calcolare il SNR direttamente sul carico R_L (resistenza di shunt, generalmente pari a 50Ω). Assumiamo che il contenuto frequenziale Δf del segnale ottico, e quindi della risposta elettrica dell'APD, sia compreso all'interno della banda di A ($\Delta f \in B_{RX}$, $B_{RX} \ll B_H$). R_L serve a polarizzare l'APD, ovvero a convertire il segnale di corrente, prodotto dal rivelatore, in un segnale di tensione amplificabile da A, nell'ipotesi che quest'ultimo abbia impedenza di ingresso infinita. In alternativa ad un amplificatore esterno A ad alta impedenza, è possibile utilizzare un amplificatore A a transimpedenza, al fine di poter minimizzare l'effetto del rumore termico utilizzando una R_L alta, senza tuttavia diminuire troppo la banda del ricevitore.

Supponiamo di illuminare un APD con una radiazione avente una certa potenza ottica P costante nel tempo, una frequenza centrale f_0 e un'occupazione di banda Δf : assumiamo che P sia sufficientemente alta, e quindi che l'intensità incidente I sia sufficientemente alta, per poter ritenere che i rumori shot legati alla fotocorrente I_L siano molto maggiori di quelli termici, elettronici e di perdita (la radiazione luminosa, il cui vettore di Poynting ha modulo proprio pari all'intensità $I_v = \mu c$, è portatrice di molti fotoni). Inoltre consideriamo la fotocorrente alta a tal punto da poter trascurare la corrente di buio $I_d = I_d' + I_d''$ (l'unico rumore shot importante è quello di I_L , cioè $i_L \gg i_d' + i_d''$). Una situazione simile (la migliore possibile), almeno in linea teorica, è riscontrabile nei sistemi di comunicazioni ottiche su fibra noti come "sistemi coerenti", rimasti in uso dalla fine degli anni 70 agli inizi degli anni 90, i quali costituiscono gli equivalenti ottici dei ricevitori eterodina.

In queste condizioni la potenza di rumore, intrinseca nell'APD, è tutta imputabile alla corrente shot di I_L ($i_L(t)$). In altre parole c'è solo $i_L(t)$ a logorare il SNR. Riferendoci sempre alla figura 84, consideriamo dapprima il rivelatore a monte di M, cioè consideriamo l'APD svincolato dal proprio guadagno, pertanto privo dell'effetto valanga: esaminiamo quindi solo la parte "intrinseca" dell'azione di fotorivelazione, così da poter calcolare il SNR "intrinseco" di un APD ($SNR_{APD(i)}$). Qualitativamente è come calcolare il SNR per un rivelatore privo di guadagno interno, ovvero per un PIN ($M = 1$). Come già visto quando un fotone (γ_1) viene assorbito dal semiconduttore può generare una coppia elettrone/lacuna. Nell'istante casuale t_1 , in cui avviene tale assorbimento, l'elettrone primario fotogenerato, superando il gap energetico, produce un impulso di corrente modellizzabile mediante una delta di Dirac posizionata in t_1 , la quale sottende, con l'asse del tempo, un'area pari alla carica elementare q. Questo impulso non è ancora rilevabile sui terminali metallici del fotodiodo: è una corrente non ancora percepita in uscita, una "potenziale" corrente. Tale impulso "stimola" il semiconduttore, che può essere visto come un

sistema avente risposta impulsiva $h(t)$, il quale prende in ingresso la delta all'istante t_1 e produce, di fatto simultaneamente, una risposta data dalla convoluzione (\otimes) fra $h(t)$ e $\delta(t - t_1)$, che è ovviamente la stessa risposta impulsiva traslata di t_1 , ovvero $h(t - t_1)$. In altre parole il semiconduttore può essere visto come un sistema elettronico, un circuito, nel quale scorre una corrente impulsiva $\delta(t - t_1)$: il materiale, il cristallo semiconduttore, non può rispondere fornendo in uscita una corrente altrettanto impulsiva, essendo un sistema caratterizzato da una certa "inerzia elettronica", ossia da una certa banda B_H finita, cosicché la risposta elettrica avrà una durata non nulla. La durata della risposta $h(t - t_1)$ è tanto più breve quanto maggiore è la banda B_H del semiconduttore, ovvero quanto più larga è la sua risposta in frequenza $H(f)$. Evidentemente in un semiconduttore a gap diretto, dove la concavità della banda di conduzione, presso il proprio minimo E_C , è ben marcata (presenza di "warping"), cioè dove la massa efficace degli elettroni è piccola, la durata nel tempo dell'impulso di risposta $h(t - t_1)$ è senz'altro minore della durata di $h(t - t_1)$ fornita in uscita da un fotodiode costruito con un semiconduttore a gap indiretto, dove la concavità della BC è inferiore e quindi la massa efficace degli elettroni è maggiore ("inerzia elettronica" maggiore).

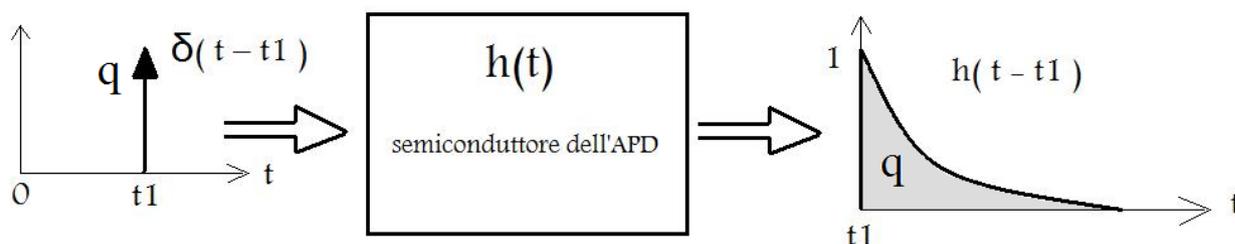


Fig. 89

Rappresentazione concettuale dell'azione di filtraggio di un impulso granulare di corrente, prodotto dal passaggio, per qualsivoglia motivo, di un elettrone dalla banda di valenza a quella di conduzione, da parte del materiale costituente il fotorelevatore.

La superficie sottesa da $h(t)$ con l'asse del tempo è la carica elementare q , la quale, andando in banda di conduzione, ha prodotto l'ingresso a delta. $h(t - t_1)$ è una corrente rilevabile in uscita dal fotodiode ed è il risultato dell'opera di filtraggio che il semiconduttore ha fatto nei confronti della delta di corrente in ingresso. Se sul rivelatore arrivano, come ipotizzato, molti fotoni, allora il semiconduttore avrà come ingresso una successione di delta, tutte posizionate ad istanti casuali t_k , alle quali corrispondono altrettanti fotoni assorbiti, dunque altrettanti fotoelettroni andati in BC. Il sistema, supposto lineare, avrà come uscita una corrente:

$$I(t) = h(t) \otimes [\sum_{k=-\infty}^{+\infty} \delta(t - t_k)] = \sum_{k=-\infty}^{+\infty} h(t - t_k)$$

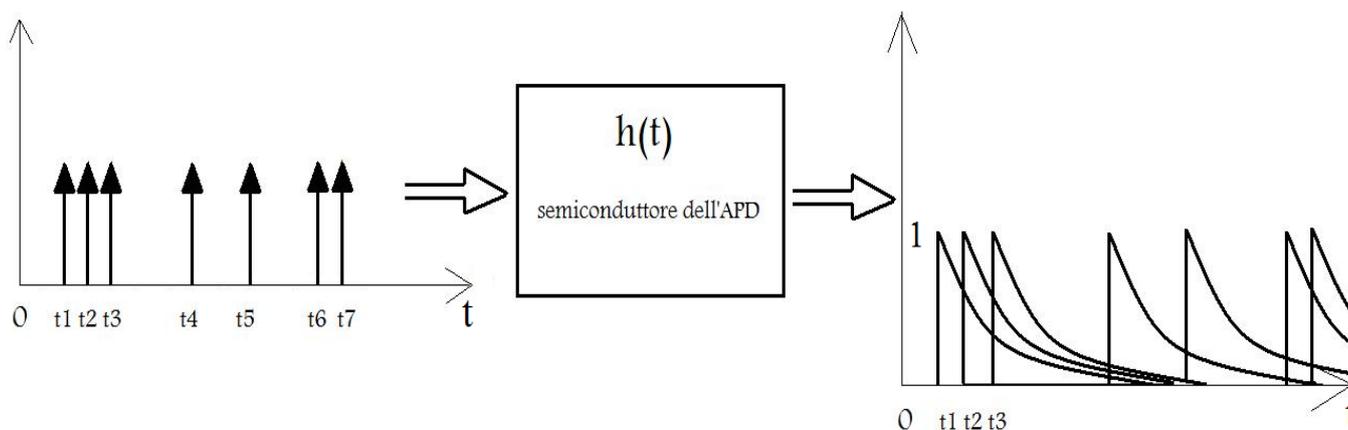


Fig. 90

Rappresentazione dell'azione di filtraggio, da parte del semiconduttore, di un treno di impulsi granulari di corrente, nell'ipotesi in cui il materiale sia modellizzabile mediante un sistema lineare.

$I(t)$ è la fotocorrente intrinseca dell'APD, cioè la fotocorrente non ancora moltiplicata per effetto valanga, oppure la fotocorrente di uscita di un PIN. Se la luce fosse monocromatica sarebbe abbastanza corretto rappresentare le risposte del semiconduttore alle delta mediante impulsi rilevabili $h(t)$ aventi tutti la stessa ampiezza: la radiazione del nostro esempio è a banda frazionale stretta ($\Delta f = 10 \text{ MHz} \ll f_0 = 345 \text{ THz}$) e quindi è approssimabile a monocromatica, anche se a rigore non lo è. Per questo motivo le $h(t)$ in uscita andrebbero disegnate, a rigore, con ampiezze diverse; la forma di una $h(t)$ dipende dalla frequenza del fotone in ingresso e dalle caratteristiche del semiconduttore (caratteristiche del suo diagramma a bande $E - k$), come ad esempio dalla propria mobilità elettronica. $I(t)$ è chiamata "processo d'impulsi di Poisson omogeneo e filtrato": "filtrato" in quanto la fotocorrente $I(t)$ è la sovrapposizione, nel tempo, di impulsi $h(t)$, rilevabili sui terminali di uscita di un PIN, prodotti dal filtraggio di impulsi di corrente, tutti a forma di delta di Dirac, da parte del semiconduttore; "omogeneo" in quanto si è ipotizzato che la potenza ottica P , incidente sul fotorivelatore, sia costante nel tempo, quindi che il numero medio (μ) di fotoni che, nell'unità di tempo, arrivano sul dispositivo sia costante nel tempo (in altre parole il numero medio di impulsi $h(t)$, nell'unità di tempo, rilevabili sui terminali metallici del dispositivo è sempre costante, è omogeneo).

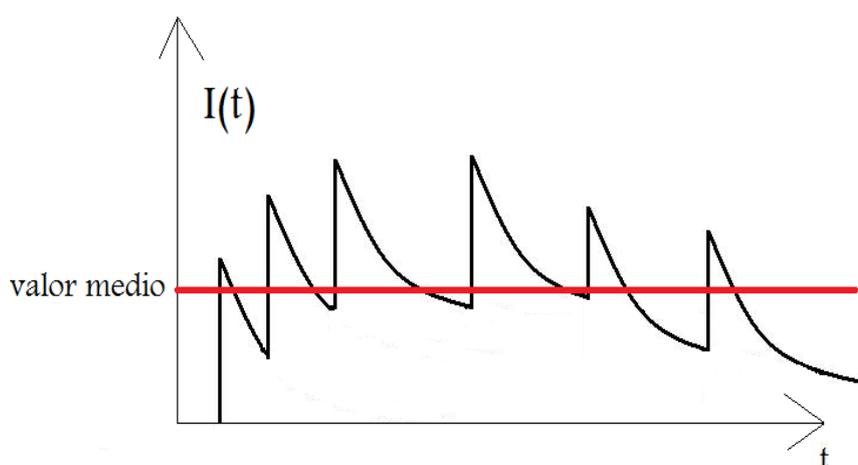


Fig. 91

Grafico qualitativo di una possibile realizzazione della fotocorrente $I(t)$, costituita da un valor medio $I_L = \langle I(t) \rangle$, che rappresenta il fotosegnale utile, e da un residuo alternato $i_L(t)$, che rappresenta il corrispondente rumore shot.

$I(t) = \langle I(t) \rangle + i_L(t) = I_L + i_L$, dove $I_L = \langle I(t) \rangle$ è il valor medio di $I(t)$ e rappresenta il fotosegnale utile (gradito), mentre $i_L(t)$ è il residuo alternato di $I(t)$, a valor medio nullo, ovvero la relativa corrente di rumore shot. $I(t)$ è definita anche come una "variabile aleatoria poissoniana". Una variabile aleatoria poissoniana può essere approssimata con una variabile aleatoria gaussiana nel caso in cui la sua "intensità", ovvero il suo valor medio (valore di aspettazione), sia sufficientemente alto. Nella fattispecie $I(t)$ possiede un valor medio (segnale utile) alto, in quanto si è ipotizzato che P incidente fosse alta, oltre che costante, tale da rendere la corrente di rumore shot i_L dominante su tutte le altre correnti di rumore. Dunque $I(t)$ può essere approssimata con una variabile aleatoria gaussiana avente come valor medio I_L e come valor quadratico medio la somma del quadrato del valor medio di $I(t)$ con la varianza di $I(t)$.

6.3.2) Calcolo del limite quantico per il SNR in uscita da un fotorivelatore

Prima di procedere con il calcolo del SNR, teniamo presente l'asserto del teorema di Campbell, secondo cui la relazione fra il fotosegnale utile, cioè il valor medio I_L della fotocorrente aleatoria poissoniana $I(t)$, ed il numero medio μ di fotoni che, nell'unità di tempo, arrivano sul fotorivelatore è:

$$I_L = \mu \int_{-\infty}^{+\infty} h(t) dt = \mu q$$

È evidente che μ (s^{-1}) è strettamente correlato all'intensità ottica incidente I ($= I_v$).

Dunque:

$$\begin{aligned} \text{SNR}_{\text{APD(i)}} = \text{SNR}_{\text{PIN}} &= \frac{\text{potenza di segnale utile}}{\text{potenza di rumore}} = \frac{\text{valor quadratico medio del fotosegnale utile}}{\text{valor quadratico medio del relativo rumore shot}} = \\ &= \frac{\langle I_L^2 \rangle}{\langle i_L(t)^2 \rangle} = \frac{I_L^2}{\text{potenza di rumore shot}} = \frac{I_L^2}{\int_{-B_{\text{RX}}}^{+B_{\text{RX}}} S_{i_L}(f) df} \end{aligned}$$

dove:

$$S_{i_L}(f) = |H(f)|^2 \mu$$

Il sistema costituito dal semiconduttore ha una risposta impulsiva $h(t)$, e quindi una risposta in frequenza $H(f)$, ottenuta tramite la trasformata bilaterale di Fourier di $h(t)$: questo sistema riceve in ingresso fotoni (vale a dire μ , ovvero un certo numero medio di fotoni al secondo) e produce rumore shot (o rumore di granularità), con relativa densità spettrale di potenza ($S_{i_L}(f)$, A^2/Hz). In figura 92 sono rappresentate $H(f)$ ed $A(f)$ (assunte reali) ed il segnale luminoso incidente sul fotorivelatore, con il range di frequenze Δf occupato da quest'ultimo.

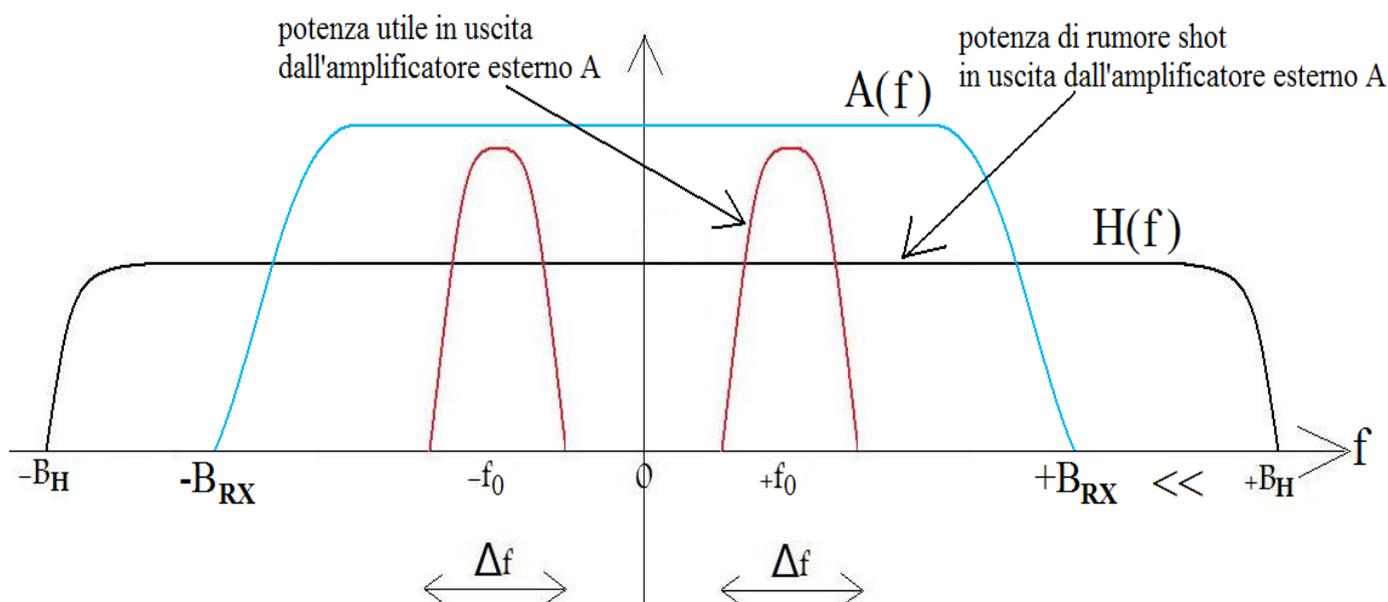


Fig. 92

Nel grafico (qualitativo) sono riportati la risposta in frequenza $H(f)$ del fotorivelatore (PIN o APD), la quale è indicativa dell'estensione spettrale e dell'ampiezza del rumore shot i_L , la risposta in frequenza $A(f)$ dell'amplificatore esterno A, il quale "preleva" sia della potenza utile relativa al segnale ottico informativo sia della potenza di rumore shot (quest'ultima degrada il SNR in uscita da A) ed il modulo del contenuto spettrale del segnale ottico incidente sul dispositivo. Nel caso in cui la frequenza ottica f_0 , prodotta da un LASER connesso al capo trasmittente della fibra, sia modulata in intensità (modulazione IM - "Intensity Modulation", la cui forma più semplice è la ben nota modulazione OOK - "On/Off Keying", nella quale il bit "0" è trasmesso spegnendo l'emettitore di luce, mentre il bit "1" accendendolo), f_0 costituirebbe la portante ottica attorno alla quale sono concentrate tutte le armoniche del segnale ricevuto.

Dal momento che $B_{\text{RX}} \ll B_{\text{H}}$, l'amplificatore esterno in banda base (A) "vede" un rumore shot i_L praticamente bianco, ovvero un $|H(f)|$ di fatto costante e circa pari a $|H(0)|$, oltre al segnale ottico utile centrato in f_0 ($f_0 < B_{\text{RX}}$).

$$H(f) = F\{h(t)\} = \int_{-\infty}^{+\infty} h(t) e^{-j2\pi ft} dt$$

Per cui:

$$H(0) = \int_{-\infty}^{+\infty} h(t) dt = q$$

Allora:

$$S_{i_L}(f) = |H(f)|^2 \mu \cong |H(0)|^2 \mu = q^2 \mu$$

Adesso monolateralizziamo la densità spettrale di potenza del rumore shot $i_L(t)$ legato alla fotocorrente utile I_L :

$$\int_{-B_{RX}}^{+B_{RX}} S_{i_L}(f) df = \int_0^{+B_{RX}} 2S_{i_L}(f) df$$

Dunque:

$$\int_0^{+B_{RX}} 2S_{i_L}(f) df = 2q^2 \mu B_{RX} = 2qI_L B_{RX}$$

$$\text{SNR}_{\text{APD}(i)} = \text{SNR}_{\text{PIN}} = \frac{I_L}{2qB_{RX}} = \frac{\mu}{2B_{RX}} \quad [\text{e75}]$$

La quantità riportata nella [e75], calcolata sotto l'ipotesi di intensa illuminazione incidente (la fotocorrente I_L è molto maggiore delle correnti di dark - bulk I_d , di elettronica esterna I_{en} , termica J e di leakage I_{leak}), è nota con il nome di "limite quantico" del SNR_{APD} (o del SNR_{PIN}) e rappresenta, fissata la potenza ottica incidente P , il limite superiore, ossia il massimo teorico (irraggiungibile), del rapporto segnale/rumore in uscita da un PIN o da un APD di buona qualità (vedremo nel paragrafo 6.5 cosa si intende per APD di "buona qualità"). Maggiore è l'intensità luminosa incidente I_v , maggiore è il valor medio della fotocorrente, e maggiore è il SNR.

Va puntualizzato che la [e75] rappresenta il limite superiore del SNR solo se l'amplificatore esterno A , collegato al fotorivelatore, è in banda base, come da noi ipotizzato, ovvero solo se l'elettronica di front - end del fotorivelatore costituisce un ricevitore convenzionale, quindi non coerente. Nel caso di ricevitore coerente, il quale rappresenta l'equivalente ottico del ricevitore eterodina, largamente utilizzato, prima dell'avvento degli amplificatori a fibra drogata con erbio, per trasmissioni su fibra ottica di tipo WDM ("Wavelength Division Multiplexing" - moltiplicazione a divisione di lunghezza d'onda), è possibile dimostrare che l'effettivo limite quantico per il SNR in uscita da A , dove A è un amplificatore in banda passante, è il doppio di quello appena calcolato. Nel caso di ricevitore omodina si dimostra altresì che l'effettivo limite quantico è il quadruplo di quello riportato in [e75].

6.4) Modellizzazione matematica della fotomoltiplicazione per un APD ed effetto del breakdown a valanga sul SNR_{APD}

Vediamo adesso che effetto produce la moltiplicazione a valanga sul SNR_{APD} , sempre nell'ipotesi di potenza ottica incidente P tale da rendere trascurabili le fluttuazioni dovute all'amplificatore esterno, al rumore termico, alle perdite superficiali e all'estrazione di minoritari (ipotizziamo cioè di essere nella migliore situazione possibile). Un fotone, se assorbito dal semiconduttore all'istante t_k , produce una fotocoppia e/h : l'elettrone va in BC, producendo un impulso di corrente (non rilevabile ai terminali del fotodiodo, cioè non ancora filtrato) uguale a quello descritto nel paragrafo 6.3.1, ossia modellizzabile con una delta $\delta(t - t_k)$ di area pari a q . Il sistema reagisce a questo ingresso, fornendo in uscita una corrente impulsiva filtrata del tipo $M_k h(t - t_k)$, la quale sottende, con l'asse del tempo, una superficie pari a $M_k q$: $M_k h(t - t_k)$ è conseguenza di una valanga di eventi ionizzanti, tutti derivanti dall'elettrone primario. Inoltre un elettrone fotogenerato all'istante t_1 produrrà, in uscita dal sistema, una risposta avente un'ampiezza M_1 ed un'area pari a $M_1 q$ statisticamente diverse dall'ampiezza M_2 e dall'area $M_2 q$ della risposta del sistema ad un elettrone fotogenerato all'istante t_2 : questo a causa del fatto che la moltiplicazione a valanga è un processo statistico, intrinsecamente affetto da aleatorietà (la probabilità P_k , da cui M_k dipende, varia casualmente, nello spazio della zds^* e nel tempo).

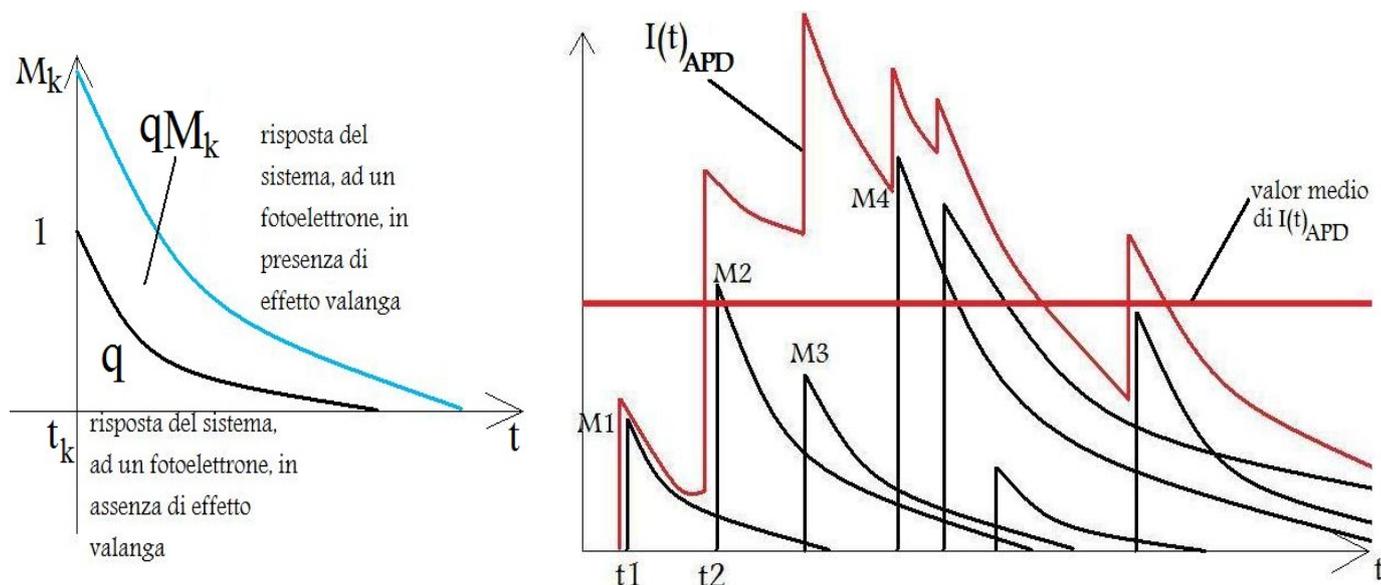


Fig. 93

Nell'immagine a sinistra possiamo notare la differenza, qualitativa, fra una corrente impulsiva $h(t)$, rilevabile ai terminali metallici dell'APD, prodotta dal semiconduttore in assenza di breakdown a valanga, ed una prodotta in presenza di moltiplicazione da impattazione. Nell'immagine a destra è raffigurata una possibile realizzazione della fotocorrente $I(t)_{APD}$ (variabile aleatoria poissoniana), in uscita dall'APD illuminato da una potenza ottica P supposta costante nel tempo. Anche $I(t)_{APD}$ consta di un valor medio (segnale utile) sovrapposto ad un residuo alternato (rumore shot).

$$I(t)_{APD} = \sum_{k=-\infty}^{+\infty} M_k h(t - t_k) = \langle I(t)_{APD} \rangle + i_{L_{APD}}(t)$$

$I(t)_{APD}$ è noto come "processo di impulsi di Poisson marcato", poiché ciascuna risposta impulsiva, rilevabile ai terminali dell'APD, ha un'ampiezza ed un'area "marcate" dall'entità di quella particolare realizzazione del breakdown a valanga che ha prodotto la risposta; le M_k sono quantità statistiche, supposte indipendenti fra loro, a valori interi ($M_k \geq 1 \forall k$). Più precisamente le M_k costituiscono delle variabili aleatorie poissoniane, approssimabili a delle variabili aleatorie gaussiane, come mostrato in figura 94: la distribuzione gaussiana riportata è riferita alla realizzazione di $I(t)_{APD}$ mostrata in figura 93.

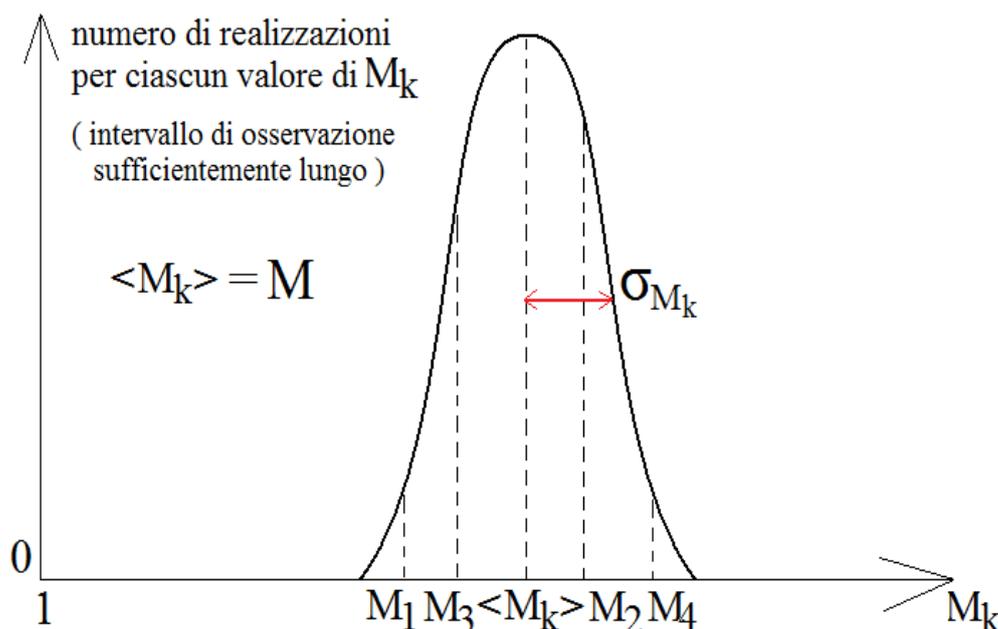


Fig. 94

Rappresentazione qualitativa della distribuzione gaussiana delle variabili stocastiche M_k , le quali, per l'esattezza, costituiscono delle variabili aleatorie poissoniane. La distribuzione mostrata è riferita alla realizzazione di $I(t)_{APD}$ riportata in figura 92.

In particolare possiamo affermare che:

$$M_k \propto \frac{E_{ek}}{E_{ik}}$$

La risposta $M_k h(t - t_k)$ fornita dal sistema, funzionante in modalità valanga, è proporzionale in ampiezza al numero di cariche secondarie generate dal moto accelerato dell'elettrone primario e_k , durante il suo tragitto verso il bulk n: tale numero è proporzionale all'energia cinetica di e_k (E_{ek}), che è legata all'elevato voltaggio inverso $|V|$ applicato ai terminali, e inversamente proporzionale all'energia di ionizzazione (E_{ik}), la quale è proporzionale alla temperatura.

Ammettendo che la risposta impulsiva $h(t)$ sia decisamente breve, ossia che la porzione di risposta in frequenza $H(f)$ "vista" dall'amplificatore A, ovvero sottesa da $A(f)$, "prelevata" da $A(f)$, sia praticamente costante in frequenza (ipotesi valida nel caso in cui $B_{RX} \ll B_H$, riferendoci sempre alla figura 92), possiamo affermare quanto segue:

$$\langle I(t)_{APD} \rangle = \langle \sum_{k=-\infty}^{+\infty} M_k h(t - t_k) \rangle \approx \langle M_k \rangle \langle \sum_{k=-\infty}^{+\infty} h(t - t_k) \rangle = \langle M_k \rangle \langle I(t) \rangle$$

$\langle M_k \rangle$ è l'ampiezza media delle risposte $h(t)$ del sistema ai fotoportatori, ossia il "guadagno medio di fotorivelazione" M . $\langle M_k \rangle$ è altresì il valor medio della distribuzione gaussiana seguita dalle M_k (come mostrato in figura 94). $\langle I(t) \rangle$ è il valor medio I_L della fotocorrente intrinseca $I(t)$.

Quindi:

$$\langle I(t)_{APD} \rangle = M I_L$$

$$SNR_{APD} = \frac{\langle (M I_L)^2 \rangle}{\langle i_{L_{APD}}(t)^2 \rangle} = \frac{(M I_L)^2}{\langle i_{L_{APD}}(t)^2 \rangle} = \frac{M^2 I_L^2}{\int_{-B_{RX}}^{+B_{RX}} S_{i_{L_{APD}}}(f) df}$$

$$S_{i_{L_{APD}}}(f) = \langle M_k^2 \rangle |H(f)|^2 \mu \cong \langle M_k^2 \rangle |H(0)|^2 \mu = \langle M_k^2 \rangle q^2 \mu$$

$$\int_{-B_{RX}}^{+B_{RX}} S_{i_{L_{APD}}}(f) df = \int_0^{+B_{RX}} 2 \langle M_k^2 \rangle q^2 \mu df = 2 \langle M_k^2 \rangle q^2 \mu B_{RX} = 2q I_L B_{RX} \langle M_k^2 \rangle$$

Dunque:

$$SNR_{APD} = \frac{M^2 I_L^2}{2q I_L B_{RX} \langle M_k^2 \rangle} = \frac{I_L^2}{2q I_L B_{RX} \frac{\langle M_k^2 \rangle}{M^2}} = \frac{I_L^2}{2q I_L B_{RX} F(M)} = \frac{SNR_{APD(i)}}{F(M)} = \frac{\text{limite quantico}}{F(M)} \quad [e76]$$

dove:

$$F(M) = \frac{\langle M_k^2 \rangle}{\langle M_k \rangle^2} = \frac{\langle M_k^2 \rangle}{M^2} = \frac{M^2 + \sigma_{M_k}^2}{M^2} = 1 + \frac{\sigma_{M_k}^2}{M^2} \quad [e77]$$

$F(M) (\geq 1)$ è noto come "fattore di rumore in eccesso": tale quantità tiene conto del fatto che il processo di moltiplicazione a valanga, essendo stocastico, introduce di per sè del rumore.

Dalla [e76] si deduce che se l'APD è illuminato da una radiazione avente potenza ottica utile alta (intensità I_v alta, molti fotoni incidenti nell'unità di tempo, ovvero μ alto), la corrente di fotoconduzione è alta e quindi se la corrente di rumore shot i_L , associata alla fotocorrente, domina su tutte le altre sorgenti di rumore, la moltiplicazione a valanga deteriora il SNR intrinseco del rivelatore di un fattore $F(M)$: il peggioramento del SNR è tanto più grave quanto maggiore è il fattore di rumore in eccesso.

6.5) Il fattore di rumore in eccesso di un APD

Per capire il grafico in figura 95 è necessario introdurre il concetto (lo riprenderemo successivamente) su cui si basano i “coefficienti di ionizzazione” α_e e α_h . Durante un processo di valanga si ha una turbolenta successione di eventi ionizzanti, dai quali nascono coppie elettrone/lacuna. Tuttavia è possibile che, all’interno del volume attivo (z_{ds}^*) in cui si verifica la moltiplicazione, solo uno dei due tipi di portatori di carica sia effettivamente capace di innescare una valanga. Ci sono APDs, ad esempio quelli al silicio o all’arseniuro di gallio, progettati per rivelare segnali ottici in prima finestra, nei quali soltanto gli elettroni primari fotogenerati e gli elettroni secondari, generati progressivamente per impatto ionizzante, sono capaci di produrre dei successivi eventi di ionizzazione, per cui la moltiplicazione è affidata soltanto agli elettroni. Chiamando α_e ed α_h , rispettivamente, “coefficiente di ionizzazione per gli elettroni e per le lacune”, possiamo definire il “rapporto di ionizzazione” K_a nel modo seguente:

$$K_a = \frac{\alpha_h}{\alpha_e}$$

Possiamo pertanto affermare che gli APDs al silicio o all’arseniuro di gallio sono caratterizzati da $\alpha_e \gg \alpha_h$, e quindi da $K_a \ll 1$. Ci sono poi APDs, come ad esempio alcune eterostrutture progettate per rivelare segnali ottici in terza finestra, nei quali sono le lacune ad avere la maggiore attitudine a produrre ionizzazioni, ovvero la moltiplicazione a valanga è affidata di fatto unicamente a loro: per questi fotorivelatori $\alpha_e \ll \alpha_h$, per cui $K_a \gg 1$. Mediante un’analisi dei fenomeni connessi alla ionizzazione per impatto si ricava un’espressione del fattore $F(M)$, parametrizzato da $K = \min\left\{K_a, \frac{1}{K_a}\right\}$, il cui andamento è riportato in figura 95.

fattore di rumore in eccesso $F(M)$

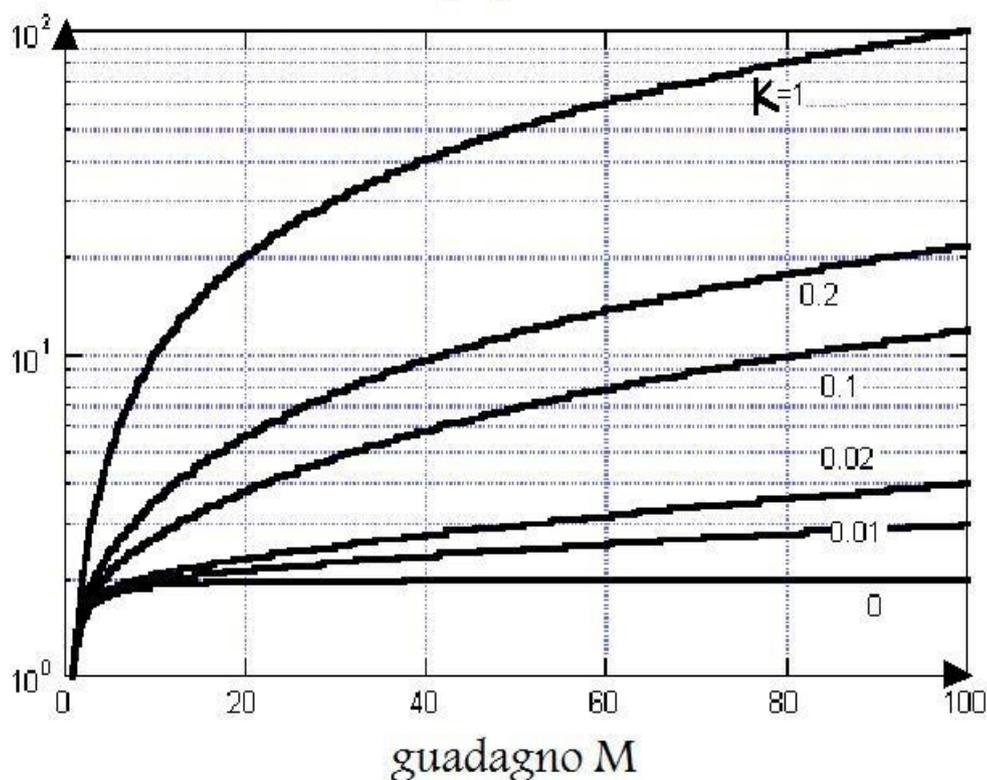


Fig. 95

Grafico del fattore di rumore in eccesso $F(M)$ introdotto da un comune APD, in funzione del guadagno M . Le curve $F(M)$ sono parametrizzate da $K = \min\left\{K_a, \frac{1}{K_a}\right\}$. Il rapporto di ionizzazione K_a , che dipende dal tipo di semiconduttore, dalla temperatura e dal voltaggio inverso applicato, è legato all’attitudine delle due specie di portatori di carica a realizzare ionizzazioni da impatto. L’obiettivo tecnologico dovrebbe essere integrare un fotorivelatore ad alto guadagno medio intrinseco M , ma in cui il breakdown a valanga è sostenuto quasi esclusivamente da una delle due specie di portatori di carica ($K \rightarrow 0$). In tal modo l’APD avrebbe un SNR buono, sia se irradiato da una potenza ottica bassa che da una potenza ottica alta.

$$F(M) = M \left[1 - (1 - K) \frac{(M-1)^2}{M^2} \right] = KM + (1 - K) \left(2 - \frac{1}{M} \right) \cong M^x$$

L'ultima approssimazione è frequentemente utilizzata per ragioni di comodità: x , noto come "esponente di rumore", sempre presente nei datasheet degli APDs, è un numero reale compreso, teoricamente, fra 0 ed 1, ma in realtà compreso fra 0.2 ed 1. Infatti un fattore di rumore in eccesso unitario, ovvero $x \rightarrow 0$, non è ottenibile nella pratica. Un APD "ideale", vale a dire un APD nel quale il processo di moltiplicazione a valanga non è soggetto ad aleatorietà granulare, è un fotorivelatore nel quale tutti gli impulsi filtrati rilevabili sui suoi terminali di uscita hanno ampiezza pari ad M , e quindi area pari ad Mq . In altre parole per ciascun fotoelettrone primario, cioè per ciascuna delta di Dirac di corrente $\delta(t - t_k)$, si innesca un breakdown a valanga in seguito al quale viene raccolto, ai terminali di uscita del fotodiiodo, il medesimo impulso filtrato $Mh(t - t_k)$ ($M_1 = M_2 = M_3 = \dots = M_k = \dots = M$), e quindi il medesimo numero di fotocariche ($M_1q = M_2q = M_3q = \dots = M_kq = \dots = Mq$).

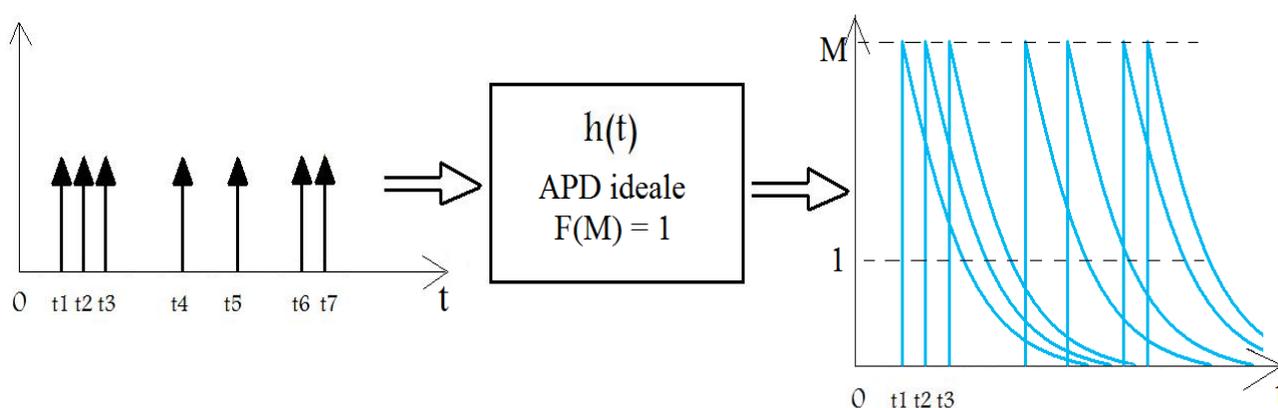


Fig. 96

Rappresentazione dell'azione di filtraggio, da parte di un ipotetico APD ideale ($F(M) = 1$, la moltiplicazione a valanga non introduce rumore), di un treno di impulsi granulari di corrente, nell'ipotesi in cui il semiconduttore sia modellizzabile mediante un sistema lineare. Si noti come gli impulsi filtrati $Mh(t - t_k)$ abbiano tutti la stessa ampiezza (M) ed area (Mq), come se il breakdown a valanga non risentisse di alcuna aleatorietà granulare; in altre parole come se, anziché essere un processo statistico, fosse un fenomeno "esatto", deterministico.

Per il suddetto APD "ideale" la grandezza statistica M_k è distribuita secondo una funzione di distribuzione gaussiana (si faccia riferimento alle figure 94 e 96) estremamente concentrata intorno a $M = \langle M_k \rangle$, dato che il numero di realizzazioni dei valori delle M_k diverse da M è decisamente trascurabile. Quindi abbiamo $\sigma_{M_k}^2 = 0$, per cui, considerando la [e77], concludiamo che $F(M) = 1$.

Si noti l'analogia fra la [e77] e la definizione di cifra di rumore (NF) di un quadripolo (la prima formula del glossario): $\sigma_{M_k}^2$ rappresenta la potenza di rumore (N_Q) introdotta dall'APD, il quale può essere pensato come un quadripolo Q optoelettronico che, nel ricevere il segnale di ingresso, ovvero un treno di delta di Dirac, introduce intrinsecamente del rumore, costituito dalla casualità dell'ampiezza M_k di ciascun impulso filtrato $M_k h(t - t_k)$. Più la polarizzazione inversa $|V|$, a cui sottoponiamo l'APD, è alta, più M (il valor medio delle M_k) è alto, più l'aleatorietà sull'ampiezza di ogni singolo impulso filtrato è alta (ciascun M_k è libero di fluttuare entro un range più ampio), più la funzione di distribuzione gaussiana delle M_k è larga, più $\sigma_{M_k}^2$ è grande, maggiore è $F(M)$, ovvero maggiore è il rumore (N_Q) introdotto dall'APD durante la fotorivelazione. L'analogia appena discussa giustifica che il "fattore di rumore in eccesso" $F(M)$ venga spesso chiamato "cifra di rumore dell'APD".

Per rivelare una radiazione luminosa avente potenza ottica abbastanza elevata potremmo pensare, essendo la moltiplicazione a valanga deleteria ai fini del SNR, e fissata l'attitudine alla ionizzazione di entrambi i portatori di carica (cioè noti α_e ed α_h , ovvero fissata una delle caratteristiche di figura 95), di utilizzare un APD avente guadagno M basso o meglio ancora un PIN. Ciò consentirebbe di attenuare il fattore di rumore in eccesso F e quindi di attenuare il peggioramento del SNR intrinseco, causato dalla valanga. Tuttavia il vantaggio, così acquisito, varrebbe solo per rivelazioni di luce intensa; se volessimo rivelare una radiazione poco intensa, con quello stesso APD o con il PIN, avremmo un peggioramento del SNR (vedremo nel paragrafo 6.6 che M alto è vantaggioso quando l'intensità luminosa incidente I_v è bassa, ovvero quando i rumori dominanti sono quello

termico, quello di perdita e quello dell'elettronica esterna). In questo caso conviene scegliere un APD con M abbastanza alto, così da ottenere una buona sensibilità per I_v bassa, ma realizzato con un materiale semiconduttore nel quale la moltiplicazione a valanga è affidata o solo alle lacune o solo agli elettroni: in altre parole o $K_a \gg 1$ o $K_a \ll 1$. In entrambi i casi K è piccolo (≈ 0), per cui F è piccolo (≈ 1). $F(M)$ è di fatto quasi ideale, cioè quasi unitario, nel caso in cui la capacità moltiplicativa, per impatto ionizzante, è fortemente decrementata per una specie di portatori e fortemente incrementata per l'altra; esiste un settore di ricerca, chiamato "ingegneria delle bande", che studia i metodi tecnologici (MBE soprattutto) per la crescita ed il trattamento di semiconduttori semplici, composti ed eterostrutture che consentano di conseguire questo risultato.

Un modo intuitivo per ottenere un basso fattore di rumore in eccesso $F(M)$ potrebbe essere scegliere un semiconduttore nel quale il diagramma a bande dell'energia totale E in funzione del numero d'onda k ($= 2\pi/\lambda$) presenti una grossa differenza fra la concavità della curva $E(k)$ della BC, presso il proprio minimo E_c , e la concavità della curva $E(k)$ della BV, presso il proprio massimo E_v . In tal modo la massa efficace per la mobilità degli elettroni sarà molto diversa da quella delle lacune, e quindi, a parità di campo ε esterno (accelerante), le due specie di portatori di carica avranno energie cinetiche molto diverse; pertanto α_e sarà molto diverso da α_h .

6.6) Il rapporto segnale/rumore per PIN e APD in condizioni di funzionamento reali (le espressioni complete)

Se adesso riprendiamo la [e75] possiamo ricavare il SNR per un fotodiodo PIN, considerando, stavolta, anche i rumori presenti nel caso realistico di un fotorivelatore non perfetto, sotto il punto di vista della sua realizzazione tecnologica, e nel caso in cui l'elettronica di front - end contribuisca al logoramento della qualità del segnale ottico rivelato: consideriamo, dunque, anche i rumori di dark - bulk, termici, di perdita ed elettronici, presenti in condizioni di funzionamento reali.

$$\begin{aligned} \text{limite quantico} &= \frac{I_L}{2qB_{RX}} = \frac{I_L^2}{2qI_L B_{RX}} \xrightarrow{\text{in generale}} \text{SNR}_{PIN} = \\ &= \frac{I_L^2}{2q(I_L + I_d)B_{RX} + S_J(f)B_{RX} + S_{leak}(f)B_{RX} + S_{en}(f)B_{RX}} = \frac{I_L^2}{\left[2q(I_L + I_d) + \frac{4kT}{R_L} + 2qI_{leak} + S_{en}(f) \right] B_{RX}} \end{aligned}$$

Abbiamo aggiunto alla "formula intrinseca" la densità spettrale di potenza di rumore Johnson, di rumore di perdita e di rumore elettronico esterno. Abbiamo inoltre sommato alla densità spettrale di rumore shot, relativo alla fotocorrente, quella di rumore shot concernente la corrente di buio (anche questa è granulare, pertanto modellizzabile con la teoria di Poisson). Quest'ultima operazione è giustificata dal fatto che, all'interno della zds^* , un elettrone giunto in BC a causa dell'assorbimento di un fotone è indistinguibile da un elettrone giunto in BC a causa dell'estrazione dei minoritari oppure a causa dell'impattazione ionica prodotta da quest'ultimi. Ricaviamo ora il SNR di un APD funzionante in condizioni reali, partendo dalla [e76]:

$$\begin{aligned} \text{SNR}_{APD} \text{ (in condizioni ottimali)} &= \frac{M^2 I_L^2}{2qI_L B_{RX} \langle M_k^2 \rangle} \xrightarrow{\text{in generale}} \\ \text{SNR}_{APD} &= \frac{M^2 I_L^2}{2q(I_L + I_d) \langle M_k^2 \rangle B_{RX} + S_J(f)B_{RX} + S_{leak}(f)B_{RX} + S_{en}(f)B_{RX}} = \\ &= \frac{M^2 I_L^2}{2q(I_L + I_d)F M^2 B_{RX} + \frac{4kT}{R_L} B_{RX} + 2qI_{leak} B_{RX} + S_{en}(f)B_{RX}} \approx \\ &\approx \frac{M^2 I_L^2}{\left[2q(I_L + I_d)M^x M^2 + \frac{4kT}{R_L} + 2qI_{leak} + S_{en}(f) \right] B_{RX}} \quad [e78] \end{aligned}$$

In figura 97 proponiamo una rappresentazione grafica bilogaritmica (qualitativa) del SNR_{APD} in funzione di M (la curva colorata di rosso): sono mostrati il numeratore ed il denominatore dell'espressione [e78], rispettivamente la potenza di segnale utile $P(M)_{segnale}$ e la potenza di rumore totale $P(M)_{rumore}$. Quest'ultima consta di una componente indipendente da M , ovvero quella dovuta ai rumori termici, di perdita e di elettronica esterna, e di

una componente fortemente dipendente da M , legata ai rumori shot di fotocorrente e di corrente di dark - bulk. La differenza fra $P(M)_{\text{segnale}}$ e $P(M)_{\text{rumore}}$, funzioni di M entrambe espresse in decibel, costituisce il $\text{SNR}_{\text{APD}}(M)$. Si noti la presenza di un massimo assoluto, per $\text{SNR}_{\text{APD}}(M)$, in corrispondenza di un particolare valore del guadagno ($M_{\text{ottimoAPD}}$). Per valori di guadagno M maggiori di $M_{\text{ottimoAPD}}$ la pendenza di $P(M)_{\text{rumore}}$ è maggiore di quella di $P(M)_{\text{segnale}}$, a causa del fattore di rumore in eccesso $F(M)$ ($\approx M^x$).

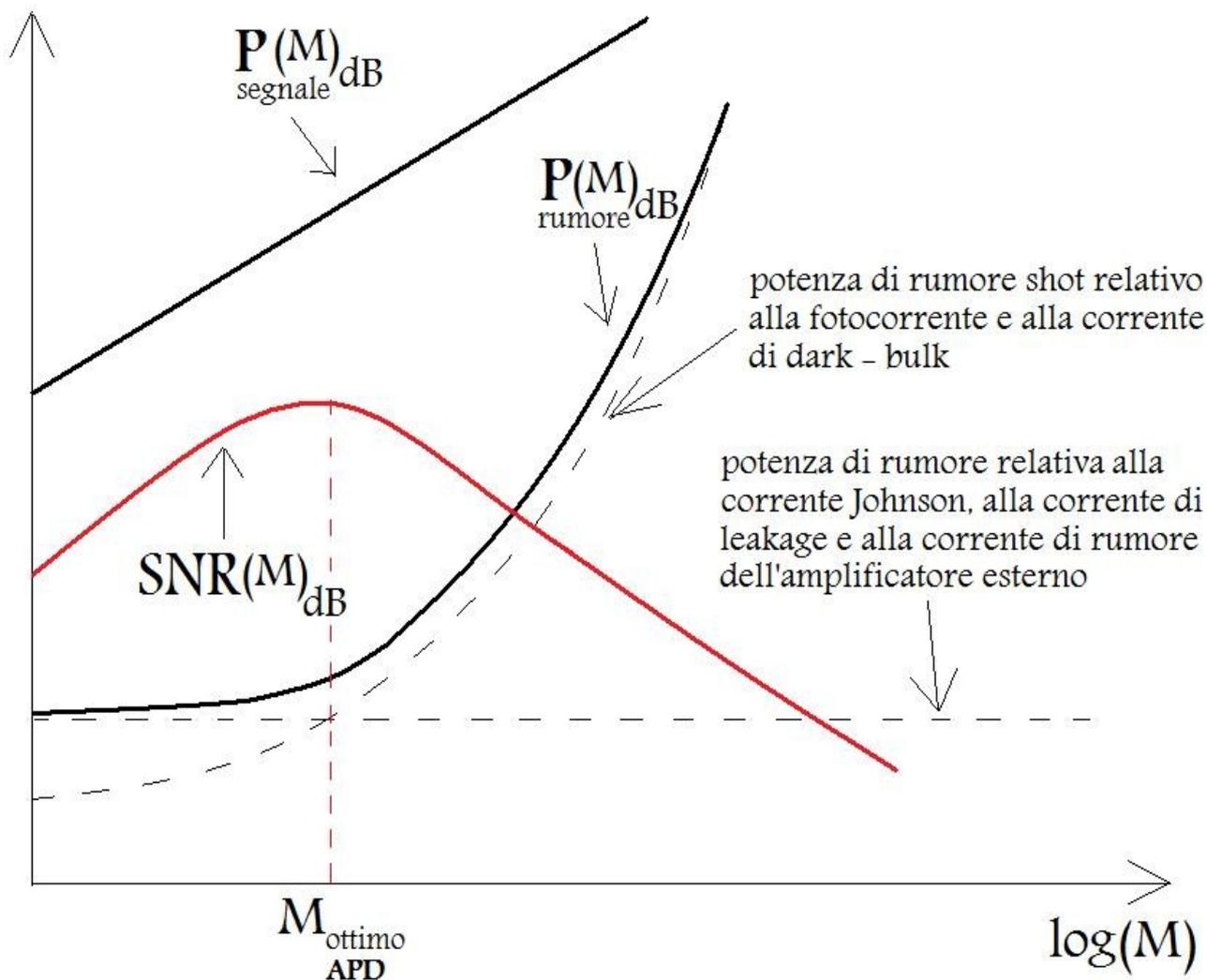


Fig. 97

Grafico bilogaritmico (qualitativo) del rapporto segnale/rumore ($\text{SNR}(M)_{\text{APD}}$) in uscita da un APD, nell'ipotesi di condizioni di funzionamento reali, ovvero ammettendo la presenza di tutte le principali sorgenti di rumore. Si notino i due contributi della potenza di rumore ed il guadagno di moltiplicazione a valanga $M_{\text{ottimoAPD}}$, presso il quale SNR_{APD} assume il suo valore massimo.

È possibile esprimere la [e78] anche in forma normalizzata:

$$\text{SNR}_{\text{APD}} = \frac{I_L^2}{\left[2q(I_L + I_d)M^x + \frac{\frac{4kT}{R_L} + 2qI_{\text{leak}} + \text{Sen}(f)}{M^2} \right] B_{\text{RX}}}$$

Per calcolare il valore ottimo di M , che massimizza il SNR_{APD} , mostrato in figura 97, è sufficiente annullare la derivata, rispetto ad M , del denominatore:

$$\frac{d \left[2q(I_L + I_d)M^x + \frac{\frac{4kT}{R_L} + 2qI_{leak} + S_{en}(f)}{M^2} \right]}{dM} = 0$$

da cui:

$$M_{ottimoAPD} = \sqrt[x+2]{\frac{\frac{4kT}{R_L} + 2qI_{leak} + S_{en}(f)}{q^x(I_L + I_d)}} \approx \frac{1}{1 - \left(\frac{V_{PL}}{V_{BD}^*}\right)^\alpha} \quad [e79]$$

La formula approssimata riportata nella [e79] è l'espressione del guadagno di moltiplicazione a valanga calcolata nel paragrafo 4.2.2; V_{PL} individua il punto di lavoro ottimo, V_{BD}^* è la tensione di polarizzazione inversa presso la quale si verifica il breakdown Zener (si faccia riferimento al grafico I - V riportato in figura 83).

È possibile eseguire una comparazione (qualitativa) fra due APD (APD1 e APD2): supponiamo che i due rivelatori siano identici eccetto che per il fattore M , con M di APD2 (M_2) maggiore di quello di APD1 (M_1). Per piccole potenze ottiche incidenti, ovviamente uguali su entrambi i rivelatori, cioè per fotocorrenti I_L piccole (al massimo qualche decina di mA), i rumori Johnson, di perdita ed elettronico dominano su quelli shot (abbiamo comunque supposto che la corrente di buio sia poco rilevante, rispetto a quella di fotoconduzione). Dunque:

$$SNR_{APD1} \approx \frac{I_L^2}{\left[\frac{\frac{4kT}{R_L} + 2qI_{leak} + S_{en}(f)}{M_1^2} \right]_{BRX}} \quad SNR_{APD2} \approx \frac{I_L^2}{\left[\frac{\frac{4kT}{R_L} + 2qI_{leak} + S_{en}(f)}{M_2^2} \right]_{BRX}}$$

pertanto $SNR_{APD2} > SNR_{APD1}$. Il fotodiodo APD2, avendo guadagno intrinseco maggiore, riesce ad estrarre meglio l'informazione, anche perché i suoi rumori termici, di perdita e di amplificazione esterna vengono attenuati di un fattore maggiore. Questa situazione determina gli andamenti del grafico (qualitativo) a sinistra di figura 98, in cui il valore efficace della corrente totale di rumore I_{rumore}^* (pA o frazioni di pA) è riportato in funzione della fotocorrente I_L . È chiaro che per rivelare basse intensità luminose conviene usare un APD con guadagno intrinseco maggiore.

Al di sopra di una certa soglia di potenza ottica, ovvero per una fotocorrente $I_L > I_L'$, i rumori shot dominano su quelli Johnson, di perdita ed elettronico, quindi:

$$SNR_{APD1} \approx \frac{I_L^2}{[2q(I_L + I_d)M_1^x]_{BRX}} \quad SNR_{APD2} \approx \frac{I_L^2}{[2q(I_L + I_d)M_2^x]_{BRX}}$$

per cui $SNR_{APD2} < SNR_{APD1}$. I rumori shot dell'APD2 sono intensificati maggiormente dall'effetto valanga di quanto non lo siano quelli dell'APD1, proprio perché $M_2 > M_1$. Dunque per rivelare radiazioni intense è preferibile un APD con basso guadagno intrinseco. Il confronto è analogo fra un PIN ($f_0 = 345$ THz, $\Delta f = 10$ MHz, $P = 1$ μ W, $QE(f_0) = 0.8$, corrente di buio $I_d = 2$ nA, $R_L = 50\Omega$, $S_{en} = (180 \text{ pA}/\sqrt{\text{Hz}})^2$) e un APD avente le stesse caratteristiche (con in più $M = 100$ e $F = M^{0.3}$). Dal grafico (quotato) a destra di figura 98 è possibile notare che l'APD garantisce miglior SNR fino a fotocorrenti di circa 35 mA, cioè fino a potenze ottiche incidenti pari a circa 63 mW.

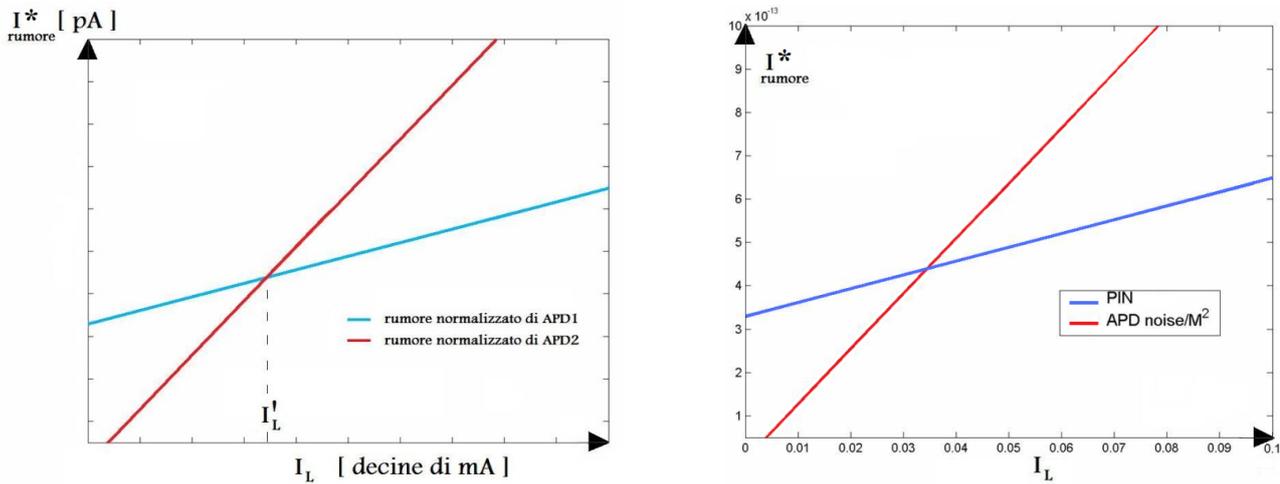


Fig. 98

Confronto fra le prestazioni, in termini di rumore (e conseguentemente di SNR), fra un APD ad alto guadagno (APD2) ed uno a basso guadagno (APD1), nel grafico a sinistra, e fra un APD ed un PIN, nel grafico a destra. Si noti, in entrambi i confronti, la presenza di due regioni di funzionamento dei fotorivelatori: una localizzata presso valori bassi della fotocorrente I_L (ovvero potenza ottica incidente bassa), l'altra presso valori alti della fotocorrente (potenza ottica incidente alta). La prima definisce un regime di funzionamento limitato dai rumori termici, di perdita e di elettronica esterna, in cui è conveniente un guadagno alto, la seconda un regime di funzionamento limitato dal rumore shot, in cui è conveniente un guadagno basso.

6.7.1) La "Responsivity" di un fotorivelatore: descrizione generale

Un parametro fondamentale, tipico di qualunque APD (e di qualunque PIN), è il coefficiente di "sensività" \mathcal{R} , che sta per "responsivity", misurato in A/W, il quale è un indice dell'ampiezza della risposta elettrica dell'APD al variare della lunghezza d'onda della radiazione luminosa incidente, a parità di potenza ottica (1W). \mathcal{R} dipende anche dal tipo di materiale usato per costruire lo strato antiriflesso. Nel grafico a sinistra di figura 99 sono riportati gli andamenti di $\mathcal{R}(\lambda)$ per alcune serie commerciali di APD (le sigle che iniziano per S), con diversi tipi di strato antiriflesso: BQ indica strati antiriflesso costruiti con silice SiO_2 , BR con una particolare resina, BK con vetro borosilicato. Nel grafico a destra abbiamo $\mathcal{R}(\lambda)$ per tre semiconduttori tipici, privi di finestra antiriflesso, con diversi valori dell'efficienza quantica QE.

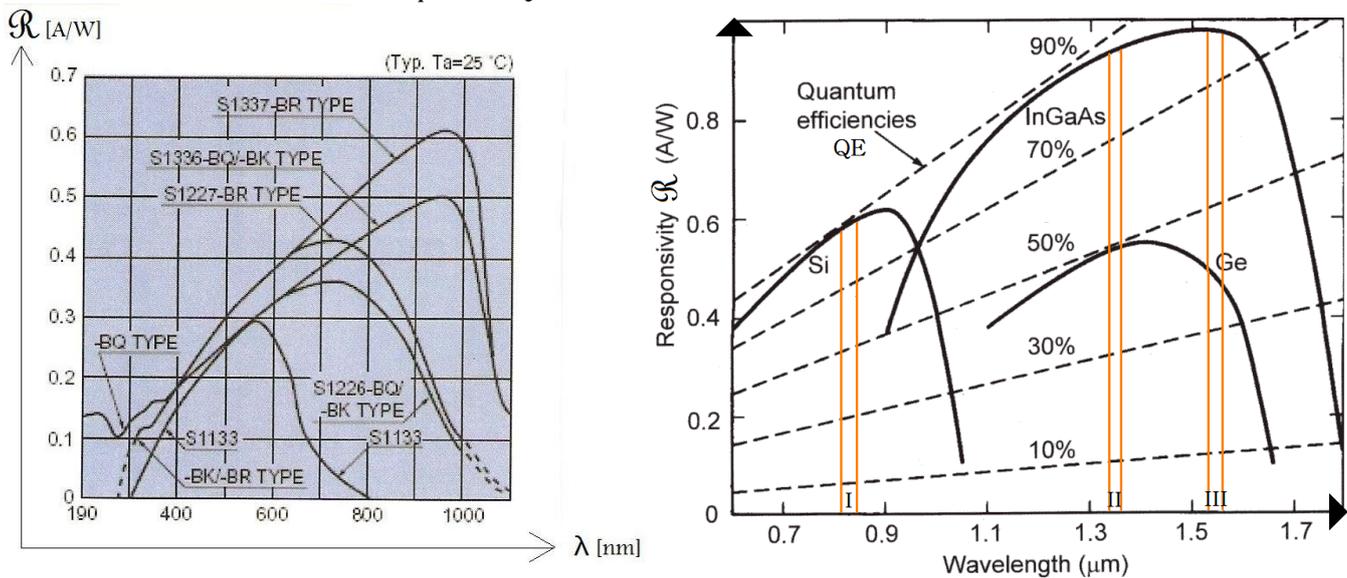


Fig. 99

A sinistra riportiamo il grafico della "responsivity" $\mathcal{R}(\lambda)$ in funzione della lunghezza d'onda λ della radiazione ottica incidente, nel quale le curve sono riferite ad alcune serie commerciali di APD e ad alcuni tipi di finestre antiriflesso. A destra riportiamo il grafico di $\mathcal{R}(\lambda)$ per un APD privo di strato antiriflesso, in riferimento a tre semiconduttori, con i quali è possibile realizzare il fotorivelatore, e all'efficienza quantica QE (si notino le tre finestre utilizzate nelle telecomunicazioni su fibra ottica, costruita con silice SiO_2).

Dal grafico di destra possiamo notare che un segnale ottico appartenente alla prima finestra, trasmesso su fibra di silice SiO_2 , può essere ben convertito in corrente da un APD costruito in Si, alla fine della fibra stessa, dato che la responsivity \mathfrak{r} per il Si assume un valore accettabile presso le lunghezze d'onda intorno a $0.85 \mu\text{m}$ (anche se il semiconduttore migliore, in virtù del gap diretto e dell'alto coefficiente di assorbimento ottico specifico α , è sicuramente il GaAs). Se il segnale ottico da fotorivelare e convertire in corrente elettrica è modulato in seconda o in terza finestra, è chiaro che il Si non può essere adoperato per la realizzazione dell'APD di ricezione: il materiale più adatto, come già spiegato, è il $\text{In}_{(0.57)}\text{Ga}_{(0.43)}\text{As}$.

In condizioni ideali abbiamo già osservato che:

$$I_L = QE M \frac{Pq}{hf_0}$$

Da questa relazione possiamo estrarre l'espressione della sensitività di un APD:

$$\mathfrak{R} = M \frac{QEq}{hf_0} = 0,807 M QE \lambda \quad [\text{e80}]$$

Quindi possiamo affermare che:

$$\mathfrak{R}_{\text{APD}}(\lambda) = M\mathfrak{R}(\lambda)_{\text{APD}[M=1]} = M\mathfrak{R}_{\text{PIN}}(\lambda) \quad [\text{e81}]$$

Nel datasheet di un APD commerciale viene fornito il grafico della responsivity $\mathfrak{r}(\lambda)$ riferita alla condizione di guadagno unitario ($M = 1$), la quale si verifica nel caso in cui la tensione di polarizzazione inversa $|V|$, applicata ai terminali dell'APD, è inferiore a V_{BD} , cioè nel caso in cui la polarizzazione inversa non è sufficientemente alta da consentire l'innescò del breakdown a valanga. Concettualmente è come se nel datasheet venga fornita $\mathfrak{r}(\lambda)$ relativa ad un PIN identico, agli effetti esterni, all'APD polarizzato con $|V| < V_{\text{BD}}$. Sempre nel datasheet di un APD è fornito il grafico di M in funzione di $|V|$, grafico che riporteremo in figura 106, il quale, congiuntamente a $\mathfrak{r}(\lambda)_{(M=1)}$, consente di risalire, note la lunghezza d'onda λ della luce incidente e $|V|$ (e quindi il valore del guadagno fotoconduttivo medio), all'effettiva responsivity $\mathfrak{R}_{\text{APD}}(\lambda)$ dell'APD.

Va puntualizzato che nella [e80] λ è espressa in μm . Si conclude che \mathfrak{r} dipende fortemente dalla lunghezza d'onda λ della radiazione: anche QE è una funzione di λ (della dipendenza della QE di un PIN da λ ci siamo occupati nel paragrafo 5.7), ragion per cui la relazione fra la sensitività e la lunghezza d'onda è tutt'altro che lineare. È possibile ricavare l'espressione fenomenologica tipica di un qualsiasi fotorivelatore, che mette in relazione la variabile ottica d'ingresso del dispositivo (P) con l'uscita elettrica (I_L) da questo fornita:

$$I_L = \mathfrak{R} P$$

Perciò un APD può essere visto come un sistema che riceve in ingresso un segnale luminoso di potenza ottica utile P e produce, in uscita, un segnale elettrico in corrente, secondo una relazione $\mathfrak{r}(\lambda)$, la quale rappresenta una sorta di funzione di trasferimento passa banda, il cui spettro ha la forma generale riportata in figura 100.

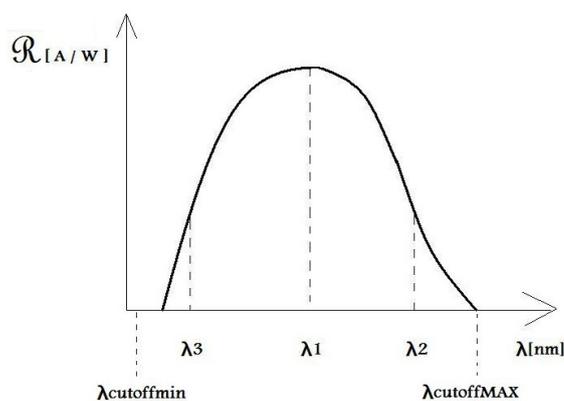


Fig. 100

Grafico qualitativo della "responsivity" $\mathfrak{r}(\lambda)$ per un generico APD

6.7.2) La “Responsivity” di un fotorelevatore: una possibile spiegazione fisica

La responsivity può essere vista come un parametro macroscopico, ovvero determinabile sperimentalmente, strettamente legato al concetto di “spettro di assorbimento” esaminato nel paragrafo 5.3.1. La forma di $\mathfrak{r}(\lambda)$ è legata a quella della Line Shape $g(\nu)$ (o meglio $g(\lambda)$) che caratterizza il semiconduttore utilizzato per realizzare il fotorelevatore. Il fatto che lo spettro della responsivity $\mathfrak{r}(\lambda)$, riportato qualitativamente in figura 100, non sia ben approssimabile con la Line Shape visualizzata in figura 35, ovvero con una funzione di distribuzione gaussiana, è dovuto al fatto che tale approssimazione è accettabile solo per materiali ben riconducibili a strutture energetiche a due livelli, mentre un qualunque semiconduttore è modellizzabile correttamente con strutture energetiche a bande $E - k$, o al più con sistemi a quattro livelli di energia. Dunque è evidente che la forma di $g(\lambda)$, e quindi di $\mathfrak{r}(\lambda)$, si complica.

Per capire meglio il motivo per cui una potenza ottica incidente viene convertita in una fotocorrente la cui ampiezza è diversa a seconda della frequenza della radiazione, è utile affidarsi nuovamente al diagramma dell’energia totale E degli elettroni in funzione del vettore d’onda k .

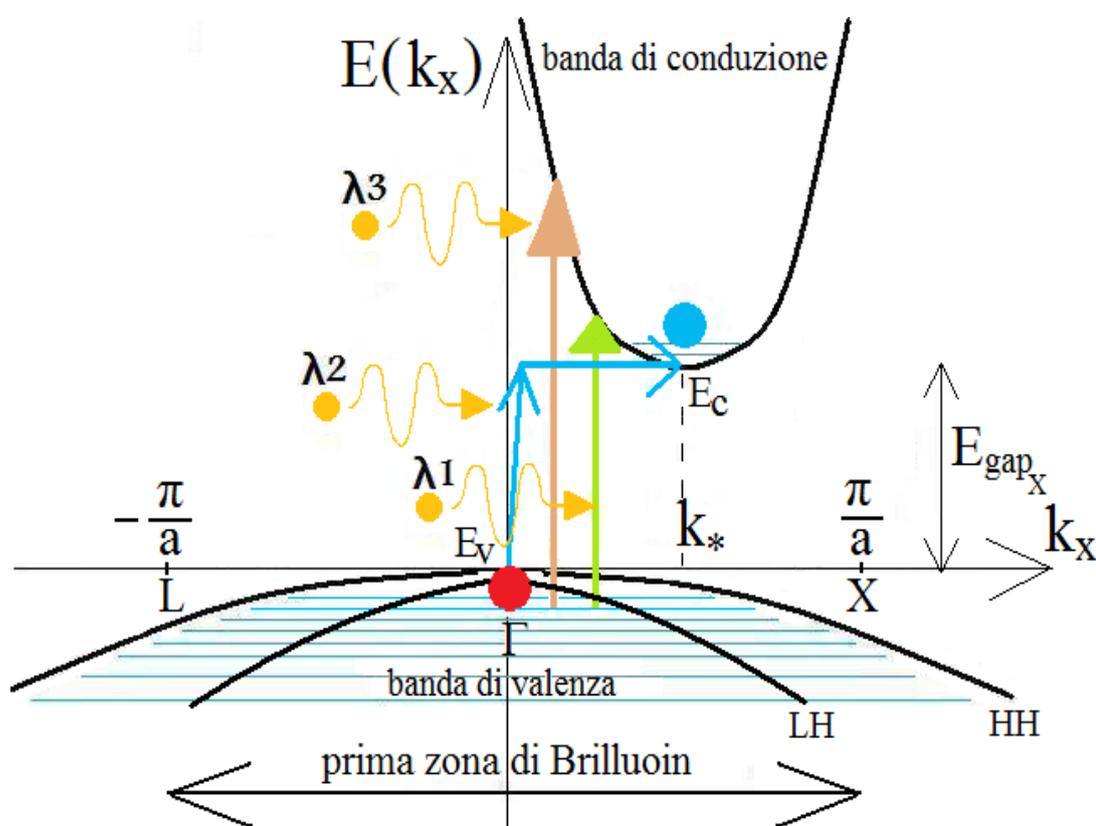


Fig. 101

Diagramma qualitativo $E - k$ di un generico semiconduttore a gap indiretto, con il quale è costruito un fotorelevatore. Mediante questo modello è possibile spiegare, intuitivamente, il motivo per cui a diverse frequenze della radiazione luminosa incidente corrispondono ampiezze diverse della risposta elettrica in uscita dal dispositivo.

Supponiamo di illuminare l’APD con una radiazione monocromatica di potenza pari ad 1 watt, avente lunghezza d’onda λ_2 . Uno qualunque dei fotoni di energia $h\nu_2$, che costituiscono questa radiazione, deve trovare, da un punto di vista energetico, una situazione favorevole per poter cedere la sua energia e stimolare un elettrone a compiere il salto del gap. Innanzi tutto il livello energetico E_v (il massimo della BV) deve essere occupato almeno da un elettrone; infatti se E_v è vuoto la probabilità di assorbimento dei fotoni, cioè la QE, diminuisce, causando quindi una diminuzione di \mathfrak{r} . Per di più il livello energetico E_c (minimo della BC) deve essere vuoto, ad esempio svuotato dall’energia termica kT , o magari da un campo ε applicato dall’esterno, in virtù della polarizzazione inversa. Se infatti E_c è pieno la QE, e conseguentemente la \mathfrak{r} , sono ulteriormente ridotte. In più, come già spiegato, affinché il fotone λ_2 provochi la generazione della coppia è necessario l’intervento del fonone, per cui se il reticolo cristallino del semiconduttore non vibra in modo tale da spostare l’elettrone dallo stato $k = 0$ (Γ) allo stato $k = k_*$, la QE è ancora più bassa. Dunque, a causa di questi tre fattori, la percentuale di fotoni λ_2 che

effettivamente viene assorbita è bassa, per cui $\mathfrak{R}(\lambda_2)$ è bassa. Per λ ancora più alte (cioè per frequenze ancora più basse), per esempio $\lambda = \lambda_{\text{cutoffMAX}}$ (= 1150 nm, per un APD realizzato in silicio), nessun fotone ha energia tale da consentire il superamento del gap ad un elettrone ($h\nu^* < E_{\text{gap}}$, con $\nu^* < \nu_{\text{cutoffMAX}}$) e quindi nessun fotone viene assorbito, $QE = 0$, $\mathfrak{R} = 0$ e l'APD non risponde. Adesso illuminiamo il rivelatore con una radiazione avente $\lambda = \lambda_1$. Statisticamente il fotone λ_1 permette ad un elettrone il salto colorato di verde nel disegno di figura 101 (salto più grande, dato che $\nu_1 > \nu_2$). La probabilità che il livello energetico inferiore, corrispondente al salto verde, sia occupato da un elettrone è alta, come pure la probabilità che il livello a cui arriva tale salto sia vuoto. Quest'ultima probabilità non è unitaria poiché gli elettroni localizzati sul fondo della BC, se soggetti ad energia termica, ovvero pochi kT , saltano su livelli superiori assai poco distanti da E_c . Per giunta qui non c'è alcun bisogno dell'intervento del fonone, dato che $\Delta k = k_f - k_i = 0$. Dunque la percentuale di fotoni assorbiti è maggiore rispetto al caso di $\lambda = \lambda_2$ e pertanto la QE è più alta e la \mathfrak{R} è più alta. Inoltre un elettrone che, saltato il gap, arriva in quel punto, possiede una massa efficace per la mobilità data da (assumendo che la dispersione $E(k)$, nelle vicinanze di E_c , sia approssimabile con una parabola):

$$m_{e1}^* = \frac{\hbar^2}{\left. \frac{d^2 E(k)}{dk^2} \right|_{k_*}}$$

In quel punto m_e^* è quasi minima (comunque abbastanza piccola). Definendo la mobilità degli elettroni come:

$$\mu_n = \frac{qT}{m_e^*}$$

si conclude che la mobilità degli elettroni che compiono il salto verde è molto alta (T è la durata del cammino libero medio). Si ha quindi un motivo in più per ritenere che una fotocorrente di portatori, generati da una radiazione luminosa avente $\lambda = \lambda_1$, sia maggiore che nel caso di λ_2 . Se la luce, che incide sull'APD, possiede una lunghezza d'onda $\lambda_3 < \lambda_1$ (e quindi una frequenza $\nu_3 > \nu_1$), allora il numero di fotoni che costituiscono la potenza ottica incidente (numero pari a 1 joule/ $h\nu_3$) è minore che nel caso precedente, pertanto il numero di portatori che contribuiscono alla fotocorrente è minore. In più ciascun elettrone, dopo aver compiuto il salto colorato di rosa in figura 100, ha una massa efficace per la mobilità maggiore, a causa della forma di $E(k)$.

$$m_{e2}^* > m_{e1}^*$$

Quindi la mobilità degli elettroni è minore e dunque il fotorivelatore risponde con una fotocorrente minore: in altre parole \mathfrak{R} è più bassa. Il limite inferiore della banda, vale a dire $\lambda_{\text{cutoffmin}}$, rappresenta il valore di lunghezza d'onda sotto il quale la radiazione luminosa, incidente sullo strato antiriflesso, viene in gran parte assorbita da quest'ultimo, non arrivando ad interagire col semiconduttore (in quel caso è logico che l'APD non possa rispondere). Se come finestra antiriflettente utilizziamo un layer di silice (SiO_2 , $E_{\text{gap}} \approx 9$ eV), $\lambda_{\text{cutoffmin}}$ è circa 138 nm ($\lambda_{\text{cutoffmin}}$ rientra pertanto nel range della luce ultravioletta).

Occorre sottolineare che la definizione appena utilizzata di massa efficace m_e^* degli elettroni per la mobilità è conseguenza di un'analisi monodimensionale dei dispositivi a semiconduttore, da noi adottata per ragioni di semplificazione analitica. In realtà un elettrone, quando si trova in banda di conduzione, interagisce con un campo elettrico periodico ε_{px} lungo la direzione x , prodotto da un potenziale periodico, lungo x , approssimabile con la forma spaziale semplificata $U_{KPx}(x)$ del modello di Kronig - Penney, con un campo periodico ε_{py} prodotto da un potenziale periodico $U_{KPy}(y)$ e con un campo ε_{pz} , prodotto da un potenziale $U_{KPz}(z)$. I tre campi periodici, così come i corrispondenti potenziali, sono diversi fra loro, poiché su ciascuna dimensione il cristallo semiconduttore presenta, in generale, una periodicità reticolare diversa. Pertanto i tre diagrammi a bande $E_j - k_j$ (con $j = x, y, z$) sono diversi fra loro, per cui la massa che ci consente di trattare l'elettrone come una particella massiva soggetta, lungo l'asse x , esclusivamente ad un campo elettrico esterno, ossia m_{ex}^* , è diversa dalle masse m_{ey}^* ed m_{ez}^* . La massa efficace m_e^* è, come noto, un tensore in funzione di k_x, k_y e k_z (anche se noi la trattiamo in qualità di grandezza scalare), così come lo è l'indice di rifrazione n_r dello stesso cristallo, in funzione di x, y, z .

7) I fotodiodi SAM ("Separate Absorption and Multiplication" APDs)

7.1.1) Il fotodiodo a valanga SAM $n^+/p/\pi/p^+$ al silicio: struttura di base, grafici spaziali delle grandezze elettriche, regioni di assorbimento e di moltiplicazione

Finora abbiamo esaminato fotodiodi nei quali l'assorbimento dell'energia luminosa e l'eventuale moltiplicazione a valanga dei fotoportatori avvengono confinate nella stessa regione. Esistono, tuttavia, dispositivi nei quali l'assorbimento e la moltiplicazione sono spazialmente separate, per cui ciascuno strato è ottimizzato a svolgere una determinata funzione (regione di assorbimento A, regione di moltiplicazione M, ecc....). Cominciamo con il presentare la struttura in silicio $n^+/p/\pi/p^+$. In figura 102 riportiamo la topologia base degli strati e gli andamenti spaziali di densità di carica totale $\rho(x)$, campo elettrico $\mathcal{E}(x)$, potenziale $V(x)$ e diagramma a bande, nell'ipotesi di equilibrio, ovvero in assenza di polarizzazione (π indica un layer p cresciuto mediante epitassia).

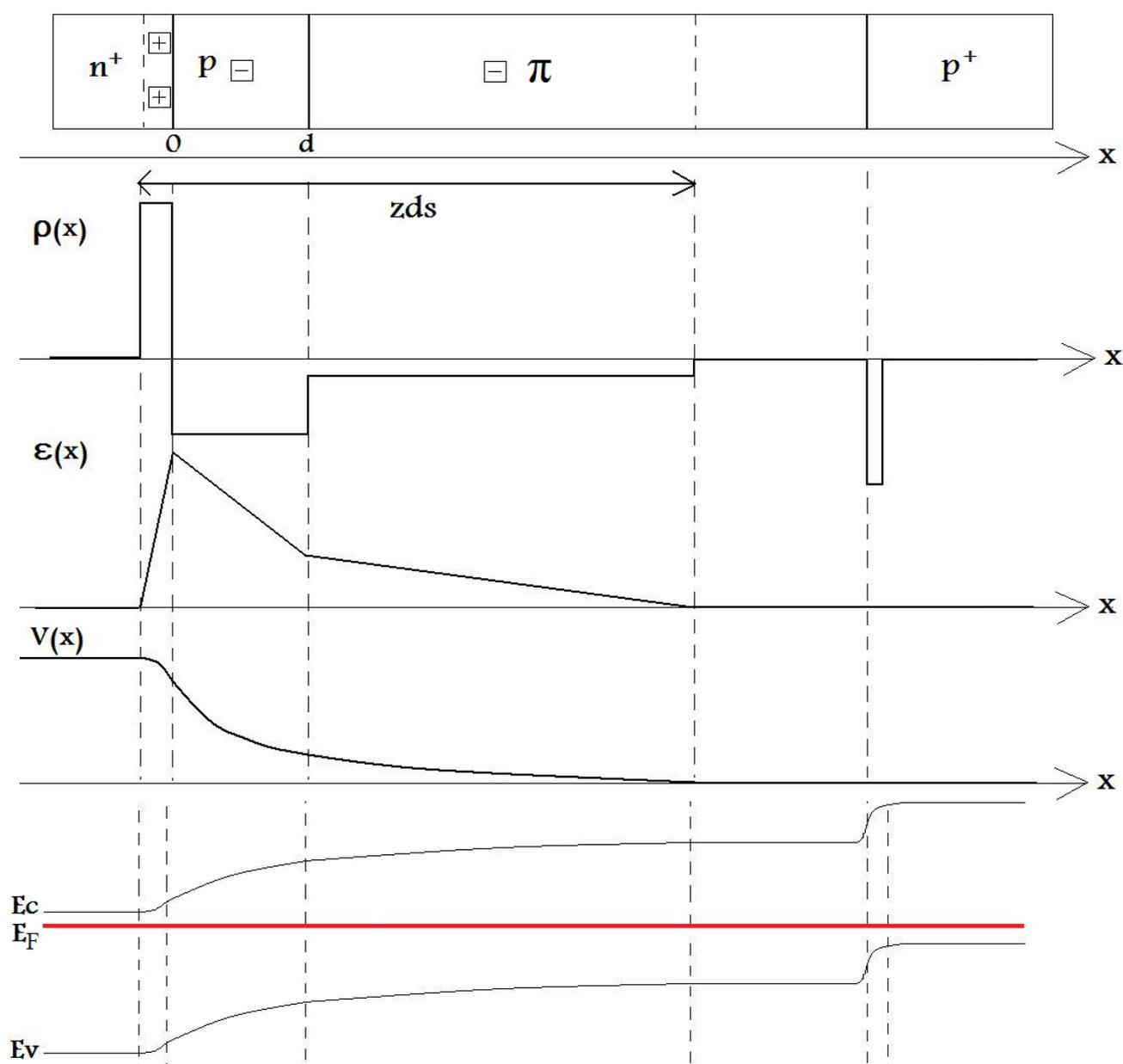


Fig. 102

Fotodiode a valanga di tipo SAM costruito in silicio, realizzato mediante la struttura $n^+/p/\pi/p^+$, in condizione di equilibrio, cioè non polarizzato. Riportiamo gli andamenti spaziali della densità di carica totale $\rho(x)$, del campo elettrico $\varepsilon(x)$, del potenziale $V(x)$ ed il diagramma a bande di energia.

La carica spaziale a destra dell'interfaccia π/p^+ è dovuta allo svuotamento di lacune, mentre il relativo accumulo (a sinistra) non è stato evidenziato, in quanto lo spazio da questo occupato è inessenziale. Ora polarizziamo inversamente la struttura e sottoponiamola ad illuminazione, come mostrato in figura 103.

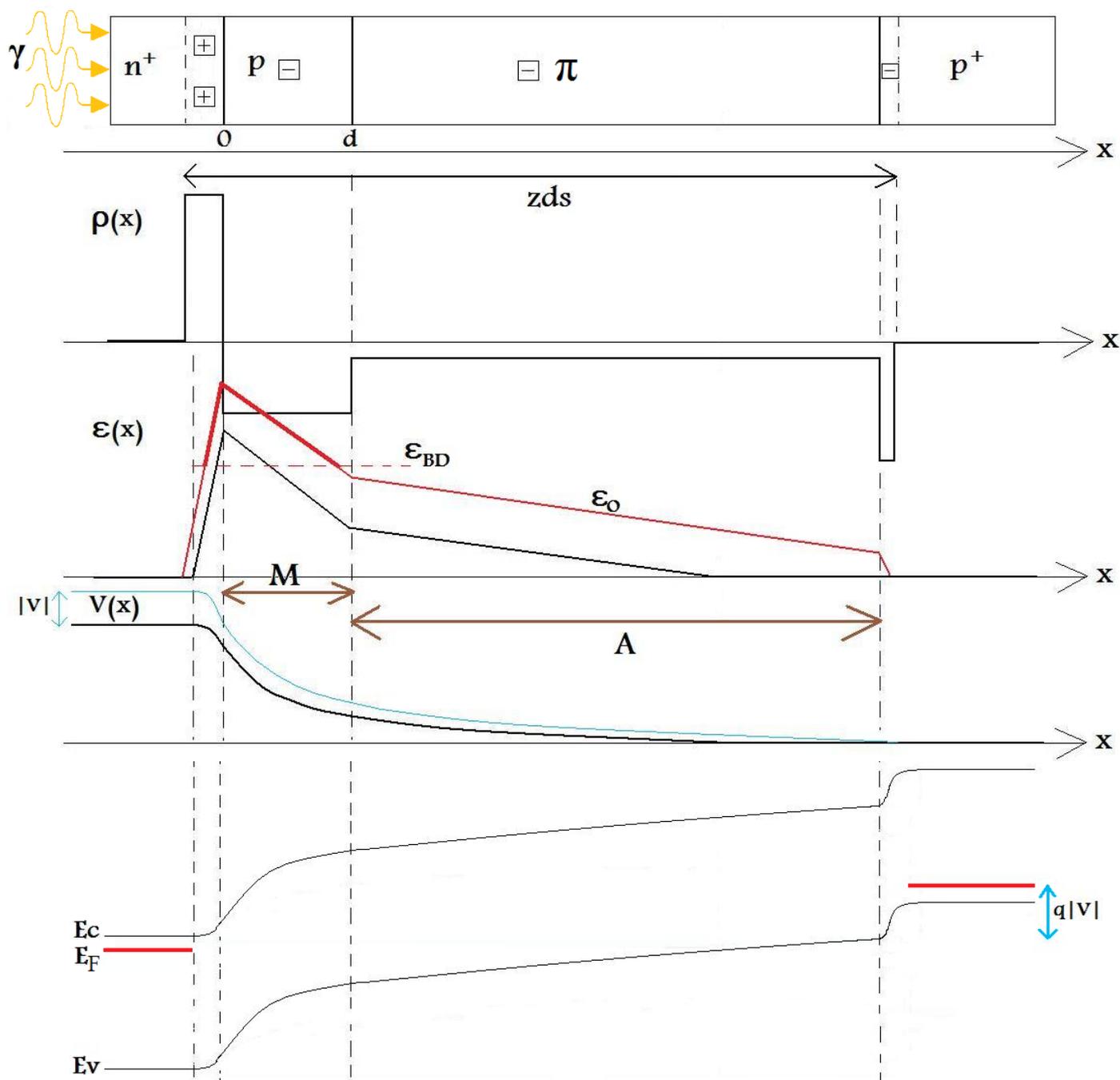


Fig. 103

Fotodiode a valanga di tipo SAM al silicio, realizzato mediante la struttura $n^+/p/\pi/p^+$, in condizione di polarizzazione inversa ($V < 0$) e sottoposto ad illuminazione. Riportiamo gli andamenti spaziali della densità di carica totale $\rho(x)$, del campo elettrico $\varepsilon(x)$, che è aumentato a causa della polarizzazione imposta sui terminali, del potenziale $V(x)$ e del diagramma a bande di energia: in quest'ultimo il livello dell'energia di Fermi è disallineato di una quantità di energia pari a $q|V|$.

La struttura presenta una configurazione del tutto comune. Abbiamo, da destra verso sinistra:

- un substrato p^+ abbastanza sottile, che deve presentare una superficie quanto più perfetta possibile, da un punto di vista reticolare
- una deposizione epitassiale di silicio leggermente drogato di tipo p (π), di spessore 4 – 5 μm , che costituisce la zona di assorbimento A. È questa la ragione per cui la suddetta superficie p^+ deve essere trattata ad hoc, dal momento che l'orientamento cristallografico di π , durante la crescita, è pesantemente influenzato da quello superficiale di p^+ .
- un layer p spesso circa 1 μm , che costituisce la zona di moltiplicazione M, ricavato o con locale diffusione termica di boro o per epitassia
- uno strato n^+ , ricavato per diffusione termica, con spessore dell'ordine di 100 – 200 nm

Quest'ultimo layer deve essere così sottile per favorire la trasmissione dell'intensità luminosa attraverso il materiale, così che questa possa raggiungere lo strato π . Inoltre l'elevato livello di drogaggio (di tipo n) favorisce l'attenuazione, qui quasi totale, della componente diffusiva della corrente, consentendo un'alta velocità di risposta. È possibile, tecnologicamente, ottenere una giunzione n^+/p così poco profonda utilizzando l'arsenico come drogante n , mediante una diffusione termica di arsina (chiamata anche arsina) AsH_3 . Infatti, invocando il modello "a sfere rigide" delle strutture cristalline, si ottiene che il raggio tetraedrico dell' As è lo stesso di quello del Si (1.175 Å) e maggiore di quello del P (depositabile mediante diffusione di fosfina PH_3), per cui l' As diffonde con più difficoltà del P, all'interno della matrice cristallina del Si ($D_{\text{As}} < D_{\text{P}}$ per qualunque temperatura tipica per la diffusione, ovvero l'energia di attivazione di Arrhenius dell'arsenico è maggiore di quella del fosforo). Sopra a n^+ deve essere deposto sia il contatto metallico superiore, possibilmente integrato a forma di corona, al fine di massimizzare la superficie di silicio illuminabile, sia uno strato antiriflettente, ad esempio di ossido. Una sezione ipersemplificata del dispositivo può essere quella mostrata in figura 104.

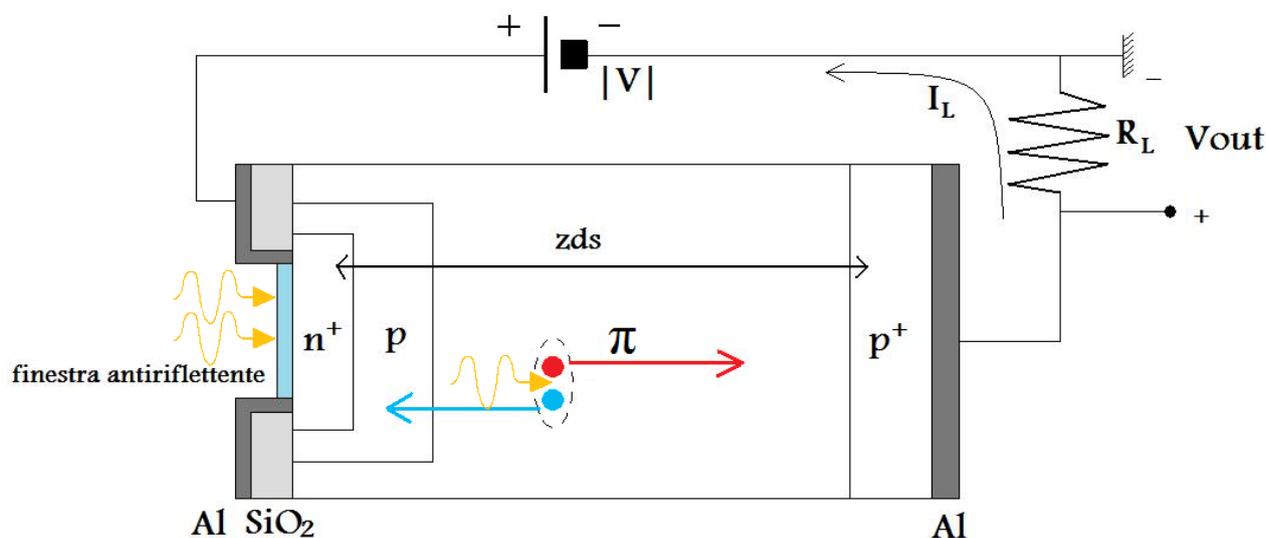


Fig. 104

Rappresentazione semplificata della sezione del SAM – APD al silicio, collegato al circuito esterno di polarizzazione e lettura.

La maggior parte dei fotoni, una volta superato lo strato n^+ , può essere assorbita sia in quello p che in quello π , ma è preferibile che l'assorbimento avvenga in quest'ultimo, per cui il suo spessore è molto maggiore (di quattro o cinque volte) rispetto a quello di p . La motivazione, intuitivamente, può essere compresa tenendo conto che il Si, per radiazioni luminose incidenti appartenenti al vicino infrarosso NIR, al visibile ed al vicino ultravioletto, possiede un rapporto di ionizzazione K_a di circa 0.02, ovvero $\alpha_e \gg \alpha_h$: pertanto la moltiplicazione a valanga è tutta dovuta agli elettroni. Ciò, oltre ad assicurarci un $F(M)$ abbastanza piccolo (bassa rumorosità anche per M grande), ci fa capire che se la maggior parte dei fotoni interagisse col Si presso $x = 0$ (figura 103), la gran parte della profondità della zona a campo elevato sarebbe sfruttata dalla specie di portatori altamente inibita alla moltiplicazione per impatto ionizzante, cioè le lacune, mentre gli elettroni, che dovrebbero generare la

moltiplicazione a valanga, avrebbero a disposizione solo una piccola regione di M (la zona di moltiplicazione) per produrre eventi ionizzanti. Dunque è necessario che i fotoni vengano assorbiti nei pressi di $x = d$, in modo che gli elettroni, muovendosi verso sinistra, abbiano tutta la zona M, in cui $\varepsilon > \varepsilon_{\text{breakdown}}$, per innescare la valanga.

Il fatto che le lacune, muovendosi verso destra, non inneschino valanghe, dal momento che $\varepsilon \approx \varepsilon_0 < \varepsilon_{\text{breakdown}}$, non è penalizzante, dato che queste sono comunque inibite alla moltiplicazione, in quanto α_h è molto piccolo. Per giunta è vantaggioso che ε_0 sia inferiore a ε_{BD} : se così non fosse, cioè se $|V|$ fosse troppo grande, si avrebbe un piegamento delle bande tale da provocare le correnti di tunnel, sostenute da portatori che nulla hanno a che fare con l'evento luminoso. Un simile breakdown (Zener) sovrasterebbe il fotosegnale, rendendone impossibile la lettura. Questo pericolo è tanto più concreto quanto maggiore è la concentrazione di donatori in n^+ (ciò fissa il limite superiore per N_d). Dunque, riferendoci sempre alla figura 103, è chiara la convenienza di assegnare la quasi totalità dell'assorbimento alla zona A (la regione π è completamente svuotata, il più povera possibile di impurezze metalliche e cresciuta con epitassia lenta ed accurata): le fotolacune si dirigono verso il bulk p^+ , i fotoelettroni verso quello n^+ e, passando attraverso la zona M, innescano la valanga che conferisce guadagno al sistema.

7.1.2) Il fotodiode a valanga SAM $n^+/p/\pi/p^+$ al silicio: profilo della densità di corrente di elettroni nella regione di moltiplicazione e valor medio M del guadagno di fotorivelazione

Calcoliamo ora, per questa struttura, il valor medio M del coefficiente di moltiplicazione della densità di corrente. Per svolgere dei calcoli analitici ipotizziamo che la regione di svuotamento nel lato n^+ abbia spessore trascurabile; assunzione più che giustificata, dato che questo spessore risulta pari a circa $\sqrt{2\varepsilon_{\text{Si}}/qN_d}$, con N_d molto alto. Dunque la zona di moltiplicazione è esattamente delimitata fra 0 e d. Assumiamo anche che la zona di assorbimento A di fotoni sia tutta e la sola regione π . Nell'istante dell'illuminazione osserviamo una densità di corrente di fotoelettroni primari, all'altezza di $x = d$, che chiamiamo $J_e(d)$. Questi fotoelettroni devono ancora moltiplicare, e lo fanno durante il loro trascinarsi verso sinistra. Ad una certa altezza x ($0 < x < d$) osserviamo una corrente $J(x) = J_e(x) + J_h(x)$: in quel generico punto contiamo elettroni primari, elettroni secondari e lacune secondarie. Presso $x = 0$ abbiamo una corrente $J_{\text{totale}}(0) = J_e(0) + J_h(0)$. In maniera rigorosa possiamo scrivere:

$$J_h(0) = J_{\text{hdrift}}(0) + J_{\text{hdiff}}(0) + J_{\text{hdarkbulk}}(0)$$

Il primo addendo è trascurabile, avendo ipotizzato che la regione di moltiplicazione sia precisamente compresa fra $x = 0$ e $x = d$, per cui gli ultimi eventi ionizzanti si verificano in $x = 0^+$. Dunque la sezione $x = 0$ non è attraversata da alcuna corrente di trascinamento di lacune. Anche il secondo addendo è trascurabile, poiché le pochissime fotolacune generate in n^+ non riescono a diffondere con successo verso p , a causa dell'elevato drogaggio (ciò concorre a rendere il dispositivo a larga banda). Il terzo addendo è formato dalle lacune termogenerate nel bulk del lato n^+ ed estratte in virtù della polarizzazione inversa (secondo il profilo spaziale espresso nella [e21]). Tuttavia, a causa del drogaggio elevato, la massima concentrazione di "lacune di buio", cioè di lacune di dark - bulk, è pari a n_i^2/N_d , per cui è bassissima, visto che N_d è alto. Pertanto anche quest'ultimo addendo è inessenziale, e concludiamo che $J_h(0)$ è trascurabile. $J_e(0)$ è la corrente di elettroni moltiplicata a valanga, ovviamente solo di drift. In osservanza alla legge di continuità della corrente abbiamo che:

$$J_e(0) = J_e(x) + J_h(x) \quad \forall x \text{ tale che } 0 < x < d \quad [\text{e82}]$$

Fissato un generico punto x e percorrendo un elemento di zona M lungo la direzione dell'asse x , osserviamo un decremento di densità di corrente di elettroni pari a:

$$-dJ_e(x) = [\alpha_e J_e(x) + \alpha_h J_h(x)]dx \quad [\text{e83}]$$

Sostituendo la [e82] nella [e83] otteniamo la seguente equazione differenziale a coefficienti costanti:

$$\frac{dJ_e(x)}{dx} + (\alpha_e - \alpha_h)J_e(x) = -\alpha_h J_e(0)$$

la cui soluzione è:

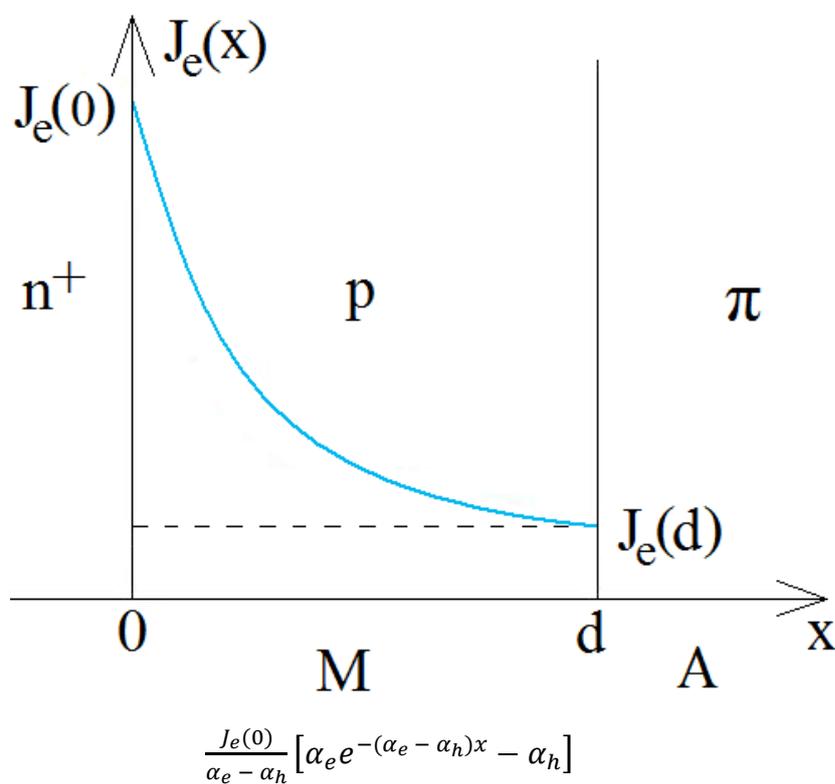
$J_e(x) =$


Fig. 105

Grafico semiquantitativo della densità di corrente di elettroni $J_e(x)$ all'interno della regione di moltiplicazione M (corrispondente allo strato p). All'interfaccia p/ π ($x = d$) si osserva una densità di corrente $J_e(d)$ di elettroni primari fotogenerati nella regione di assorbimento A (corrispondente allo strato π). I fotoelettroni, trascinati dal forte campo elettrico, producono ionizzazioni per impatto, in seguito alle quali si osserva, all'interfaccia n^+ /p, una densità di corrente di elettroni $J_e(0)$ costituita da fotoelettroni primari e soprattutto secondari. Il fattore di moltiplicazione, dovuto al fenomeno del breakdown a valanga, è proprio $M = J_e(0)/J_e(d)$.

È quindi immediato ricavare M:

$$M = \frac{\text{fotocorrente di elettroni moltiplicata}}{\text{fotocorrente di elettroni primari}} = \frac{J_e(0)}{J_e(d)} = \frac{1 - K_a}{e^{-(1 - K_a)\alpha_e d} - K_a} \quad [\text{e84}]$$

Poichè i due coefficienti di ionizzazione α_e ed α_h dipendono, seppur debolmente, dal voltaggio inverso applicato (ovvero dal campo ϵ), possiamo disegnare il seguente grafico qualitativo:

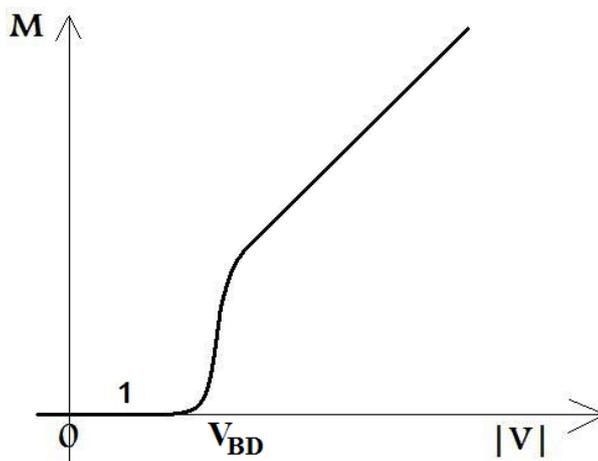


Fig. 106

Grafico qualitativo del fattore di moltiplicazione M in funzione del voltaggio inverso applicato al fotodiodo SAM - APD (di silicio), costruito a struttura $n^+/p/\pi/p^+$. Per bassi valori di voltaggio inverso, ovvero $|V| < V_{BD}$ (per cui all'interno del layer p si ha $\epsilon < \epsilon_{BD}$), non si verifica il breakdown a valanga, quindi $M \approx 1$. Per voltaggi sufficientemente alti, ossia $|V| > V_{BD}$, il dispositivo è in grado di guadagnare, in termini di densità di fotocorrente, secondo un fattore di moltiplicazione M linearmente dipendente da $|V|$.

α_e ed α_h crescono all'aumentare di $|V|$, mentre la loro differenza diminuisce, con conseguente aumento della rumorosità della fotorivelazione. Nel Si, con radiazioni luminose incidenti aventi frequenze rientranti nei range del vicino infrarosso, del visibile e del vicino ultravioletto, α_h sarà sempre inferiore ad α_e , ma per valori di ϵ abbastanza alti il rate di ionizzazione delle lacune "recupera terreno" rispetto a quello degli elettroni, a tal punto che per $\epsilon \approx 5 \cdot 10^5 \text{ Vcm}^{-1}$ si ha che α_h è circa la metà di α_e . Dall'espressione [e84] è evidente che, per $d \rightarrow 0$ (regione di moltiplicazione troppo stretta), $M \rightarrow 1$. La zona di moltiplicazione non deve, inoltre, essere troppo larga, per non correre il rischio di avere un breakdown a valanga distruttivo (valanga non più controllata). Infatti per:

$$d \rightarrow \frac{-\ln(K_a)}{(1 - K_a)\alpha_e}$$

si avrebbe:

$$M \rightarrow \infty$$

Anche per le strutture SAM è valida la formula:

$$I_L = R P = \frac{M Q E q}{h f} P$$

Occorre tenere presente che maggiore è il guadagno M offerto dal dispositivo e minore è la sua banda passante, poiché aumenta il tempo necessario alla generazione dei portatori secondari per effetto valanga e, quindi, il tempo necessario alla loro rimozione dalla zona di moltiplicazione verso il circuito esterno. La relazione fra M ed il tempo di rimozione è quasi una proporzionalità inversa esatta.

7.1.3) Il fotodiodo a valanga SAM $n^+/p/\pi/p^+$ al silicio: la dipendenza spaziale della "probabilità di trigger"

Giustificiamo ora in maniera formale il fatto che è importante che la maggior parte dei fotoni sia assorbita nel layer π , definendo uno dei parametri più importanti per un SAM - APD: la "probabilità di trigger". Fissato un certo voltaggio inverso, la probabilità di trigger di una coppia di fotoportatori appena generata è la probabilità che il fotoelettrone e la fotolacuna producano un evento ionizzante (cioè che inneschino la valanga) durante i rispettivi trascinalenti, prima di essere raccolti al di fuori della zona a campo elevato. Tale probabilità condizionata, essendo legata al comportamento di una coppia, conterà dei contributi di probabilità di entrambe le specie. In formule:

$$P_{tr}(x) = P_e(x) + P_h(x) - P_e(x)P_h(x)$$

Un elettrone che si trova all'interfaccia n^+/p attraverserà una ristrettissima porzione di zona ad alto campo, prima di essere raccolto da n^+ , ed avrà, quindi, una probabilità praticamente nulla di fare da trigger per il fenomeno della valanga: dunque, in base alle ipotesi utilizzate per il calcolo di M , si evince che $P_e(0) \rightarrow 0$, mentre $P_e(d) \rightarrow P_{eMAX}$ (l'elettrone può ionizzare all'interno dell'intera zona di moltiplicazione). Viceversa, per la lacuna, $P_h(d) \rightarrow 0$, mentre $P_h(0) \rightarrow P_{hMAX}$. Evidentemente $P_{tr}(0) \rightarrow P_{hMAX}$ e $P_{tr}(d) \rightarrow P_{eMAX}$ (con $P_{eMAX} > P_{hMAX}$). Studiando la relazione fra la frequenza di ionizzazione delle due specie e la probabilità di trigger è possibile scrivere il seguente sistema 2x2 di equazioni differenziali accoppiate:

$$\begin{cases} \frac{dP_e(x)}{dx} = (1 - P_e(x))\alpha_e(P_e(x) + P_h(x) - P_e(x)P_h(x)) \\ \frac{dP_h(x)}{dx} = (1 - P_h(x))\alpha_h(P_e(x) + P_h(x) - P_e(x)P_h(x)) \end{cases}$$

Le approssimazioni lineari delle due soluzioni, insieme a quella di $P_{tr}(x)$, sono visualizzate nel grafico mostrato in figura 107.

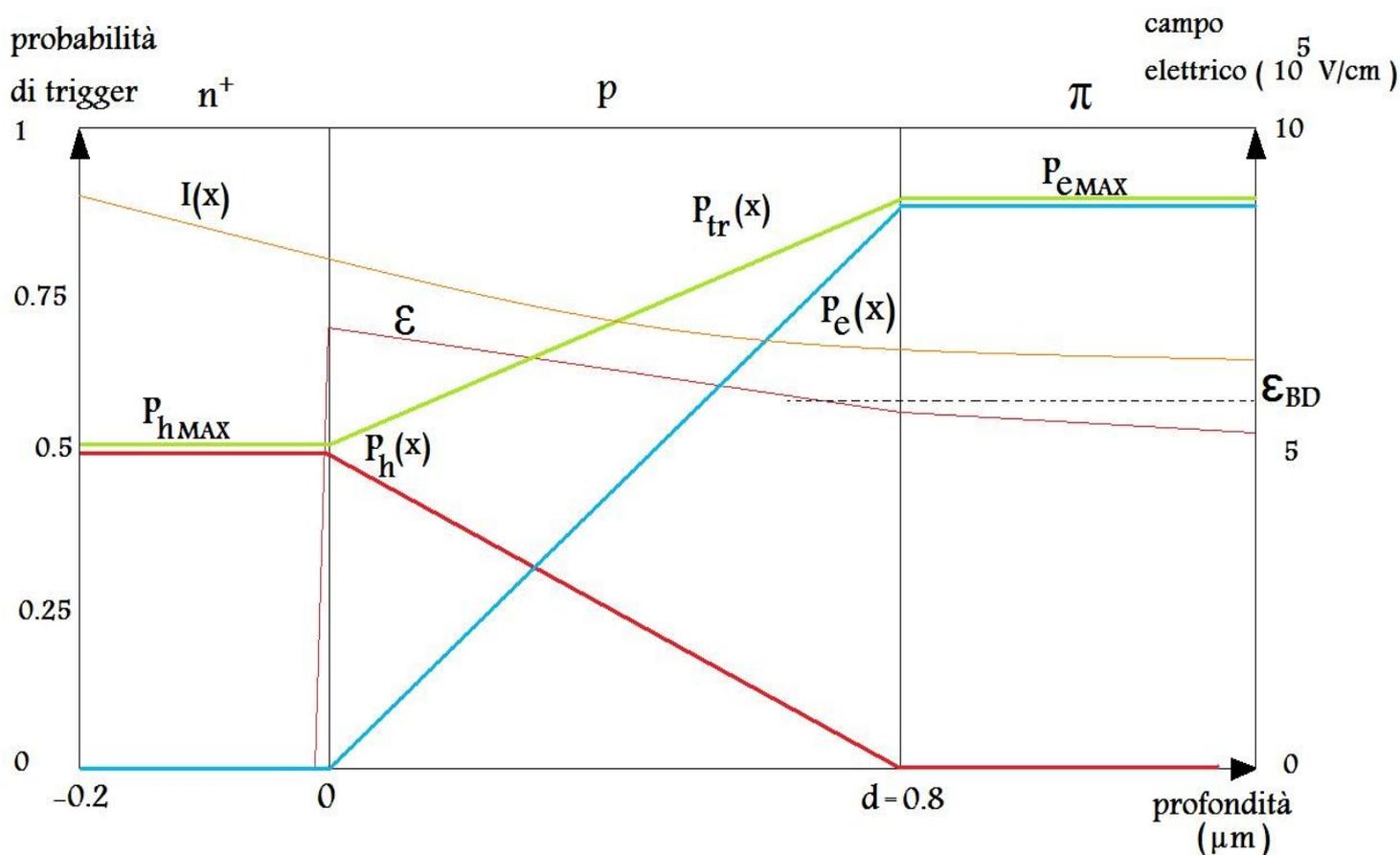


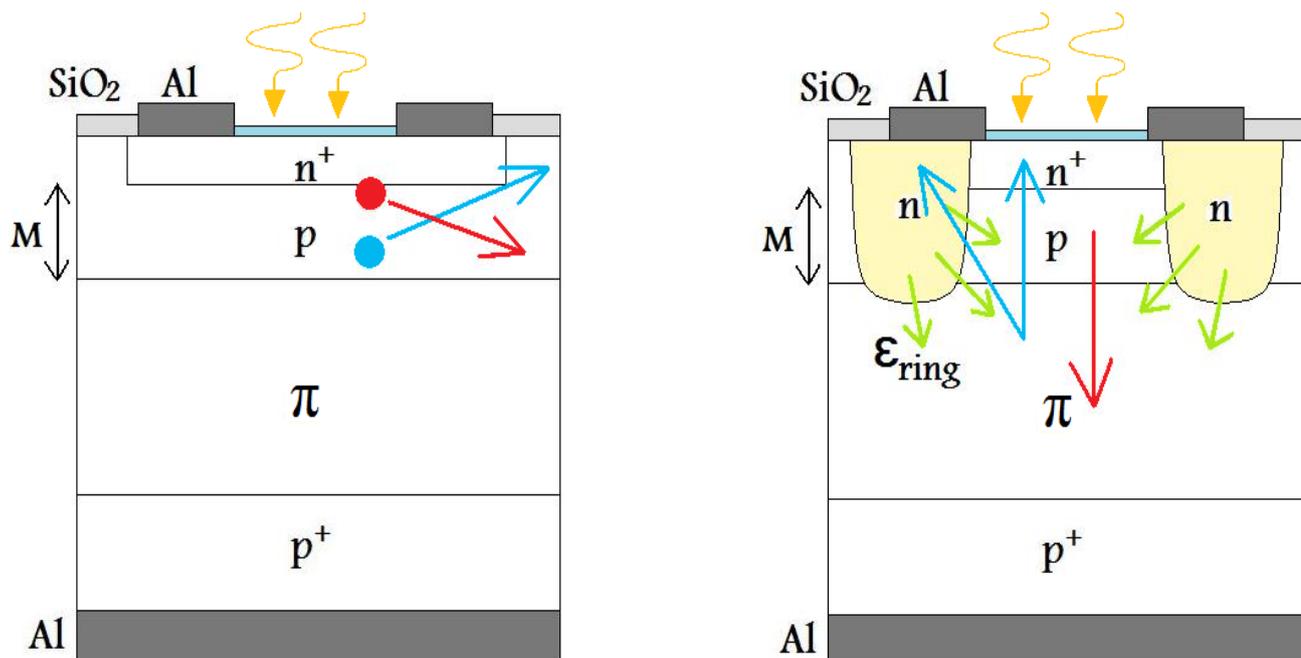
Fig. 107

Riportiamo un ingrandimento della regione $n^+/p/\pi$, con dei valori tipici per le profondità degli strati n^+ e p , relativo alla sezione del fotodiode SAM – APD costruito a struttura $n^+/p/\pi/p^+$. Sono visualizzati gli andamenti spaziali, approssimati linearmente, della probabilità di trigger $P_e(x)$ di un fotoelettrone primario, della probabilità di trigger $P_i(x)$ di una fotolacuna primaria, della probabilità di trigger condizionata $P_{tr}(x)$ della fotocoppia primaria elettrone/lacuna. Sono contestualmente riportati anche gli andamenti del campo elettrico $\varepsilon(x)$ ($< \varepsilon_{BD}$ nella zona di assorbimento della luce, $> \varepsilon_{BD}$ in quella di moltiplicazione del fotosegnale) e dell'intensità luminosa $I(x)$ (ad esponenziale decrescente, data la condizione di piccolo segnale): si noti come la geometria del SAM – APD limiti fortemente l'assorbimento di fotoni nei layers n^+ e p .

È pertanto di primaria importanza progettare dei layers n^+ e p che consentano all'intensità luminosa incidente di raggiungere, in grossa percentuale, lo strato π , laddove le coppie fotogenerate hanno la maggiore probabilità di trigger.

7.1.4) Il fotodiode a valanga SAM $n^+/p/\pi/p^+$ al silicio: controllo e collimazione del breakdown a valanga mediante trench integrata, fattore geometrico, efficienza quantica

Osservando la struttura mostrata in figura 108, a sinistra, ci rendiamo conto di un possibile problema: mentre i fotoelettroni innescano eventi ionizzanti nella zona M, alcuni di loro, sia primari che secondari, potrebbero fuoriuscire dal volume attivo (volume di moltiplicazione a valanga). Ciò comporterebbe una diminuzione dell'entità della valanga, poiché i fotoelettroni dispersi non produrrebbero ionizzazioni per impatto, e quindi dell'ampiezza del fotosegnale elettrico, anche perché quei fotoelettroni potrebbero non essere raccolti dai contatti metallici. Avremmo, pertanto, una diminuzione dell'efficienza quantica QE e della responsività r del



fotodiode. Analogamente alcune fotolacune, fra le tantissime generate nella regione M, potrebbero sfuggire al campo della zona p, disperdendosi: anche questo fenomeno comporterebbe un calo del fotosegnale, poiché le lacune suddette non verrebbero raccolte dal contatto metallico inferiore.

Fig. 108

A sinistra è rappresentata la sezione di un SAM – APD $n^+/p/\pi/p^+$, completa di contatti di alluminio, ossido di silicio di passivazione e finestra dielettrica antiriflettente. Il fotodiode è sottoposto ad illuminazione. È messo in evidenza il problema della fuoriuscita, dal volume attivo, dei fotoportatori all'interno della zona a campo elevato (M). A destra è riportata una possibile soluzione, volta a confinare il breakdown a valanga al di sotto dell'area attiva e a collimare i flussi di carica che ne derivano.

Un modo per risolvere il problema potrebbe essere l'intervento, a livello costruttivo, mostrato sempre nella figura 108, a destra: una "trench" ("ring") di Si di tipo n. Questa, insieme alle regioni p e π, costituisce delle giunzioni presso cui insistono dei campi ε_{ring} , i quali hanno un duplice effetto positivo: raccolgono i fotoelettroni di valanga che, per qualche motivo, non percorrono la zona M lungo la sua totale profondità, immettendoli

comunque nel contatto metallico e garantendone la raccolta ai fini del fotosegnale, e al tempo stesso respingono le fotolacune di valanga che potenzialmente sarebbero uscite dal volume attivo, collimandone il flusso di drift verso il layer π . Dunque la trench, esercitando un'azione di controllo e collimazione del breakdown a valanga, ha un effetto positivo sulla QE e sulla R . Si noti che tale vantaggio è il risultato di un attento equilibrio fra l'isolamento antidispersivo della valanga e la sufficiente larghezza della superficie attiva: infatti se la sezione della trench è eccessiva, la "zona morta" (la regione sottesa dalla corona metallica superiore, dalla trench e da un resistore superficiale di cui parleremo nel paragrafo 8.3) risulta troppo larga rispetto a quella attiva, e conseguentemente la regione M troppo sacrificata; il risultato è una diminuzione dei parametri sopra citati. Dunque sarebbe importante poter progettare un dispositivo equipaggiato con una trench, ma con un "fattore geometrico" ϵ_G , detto anche "fill factor", il più possibile tendente all'unità:

$$\epsilon_G = \frac{\text{area attiva}}{\text{area totale}}$$

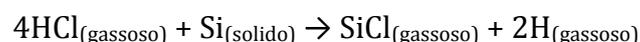
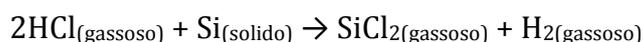
Purtroppo l'attuale stato dell'arte non consente ϵ_G superiori al 44% per il SAM - APD descritto, di area totale pari a $40 \times 40 \mu\text{m}^2$, ed al 50% per la stessa struttura, di area totale pari a $50 \times 50 \mu\text{m}^2$. Il dispositivo in questione ha un'efficienza quantica QE la cui dipendenza spaziale è data dalla seguente relazione:

$$QE(x) = \epsilon_G P_e(x) (1 - R) (1 - e^{-\alpha x}) \quad \forall x \text{ tale che: } 0 < x < d \quad [e85]$$

dove $P_e(x)$ è appunto la probabilità che un fotoelettrone, all'interno della regione di guadagno, dia origine alla valanga presso la generica coordinata x .

7.1.5) Il fotodiodo a valanga SAM $n^+/p/\pi/p^+$ al silicio: crescita epitassiale dello strato di assorbimento π e profilo spaziale della concentrazione di drogante all'interfaccia π/p^+

Torniamo ad esaminare la struttura del SAM - APD $n^+/p/\pi/p^+$ e concentriamoci, adesso, sulla parte posteriore (rispetto alla direzione di ingresso della luce) del dispositivo, ossia l'interfaccia π/p^+ . p^+ è un substrato massivo cresciuto per epitassia oppure ricavato da un wafer CZ mediante taglio diamantato (il più possibile accurato) e successiva lappatura. Sarebbe ancor più vantaggioso utilizzare un wafer FZ oppure un wafer realizzato con la tecnica di Bridgeman, essendo tali wafers molto più poveri, rispetto a quello CZ, di impurezze metalliche. Tuttavia, ad esempio, un wafer FZ costa, come ordine di grandezza, dieci volte di più di un wafer CZ, e per giunta ha una superficie più piccola (circa 20 cm di diametro, rispetto ai circa 30 cm del CZ); pertanto da un wafer FZ si ricava un numero inferiore di substrati utili. È molto importante che il substrato p^+ sia il meno inquinato possibile, affinché durante la successiva crescita epitassiale da fase vapore VPE ("Vapor Phase Epitaxy") di π su p^+ non si verifichi un'elevata diffusione termica delle suddette impurezze (soprattutto di Au, Co, Cd, Fe e Cu) dal substrato al layer in fase di deposizione. Prima di iniziare il processo di accrescimento epitassiale di π è opportuno preparare la superficie del silicio p^+ , eliminando tutte le eventuali impurezze e dissipando tutte le tensioni superficiali, per limitare la formazione di difetti all'interno della regione di assorbimento, i quali degraderebbero la QE del SAM - APD. Tale preparazione si ottiene facendo scorrere, sulla superficie di p^+ , acido HCl ad elevate temperature (1200 °C). Le reazioni chimiche che avvengono alla superficie del substrato sono le seguenti:



In tal modo la superficie imperfetta di p^+ viene erosa ("etching" del silicio) e i prodotti gassosi desorbiti, cosicché la successiva epitassia VPE di π si adatta, cristallograficamente, ad una configurazione reticolare migliore. Occorre far attenzione a non corrodere, con l'acido cloridrico, le pareti di quarzo del reattore e di grafite del

suscettore; per questo tali materiali vengono preventivamente trattati con un processo CVD, il quale li ricopre di un film sottile che li rende inattaccabili dall'acido.

In figura 109 è raffigurato un modello ipersemplificato di reattore utilizzabile per la crescita epitassiale VPE contemporanea di vari layers π su altrettanti substrati massivi p^+ . L'inclinazione del suscettore è finalizzata unicamente a creare una situazione di omogeneità tra i film gassosi stagnanti, ciascuno localizzato appena al di sopra di un layer π in fase di crescita, composti dai reagenti gassosi diffusi nella camera di lavoro. Il fatto che i film gassosi stagnanti siano uguali fra loro comporta una crescita epitassiale di layers π tutti approssimativamente uguali, pertanto alla fine del processo VPE abbiamo un set di omogiunzioni π/p^+ approssimativamente uguali fra loro, pronte per essere ulteriormente trattate al fine di ottenere un set omogeneo di fotodiodi SAM - APD $n^+/p/\pi/p^+$.

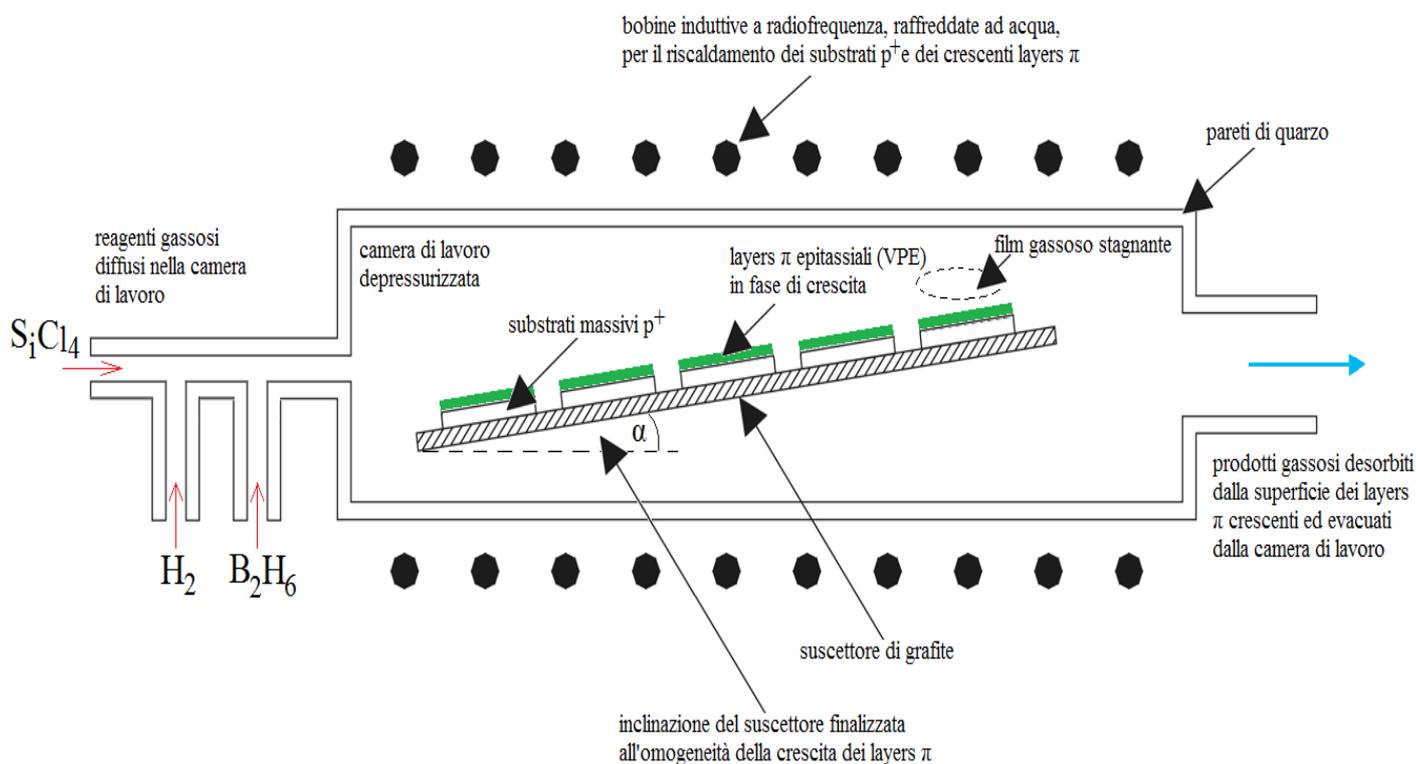


Fig. 109

Rappresentazione di un modello ipersemplificato di reattore utilizzabile per la crescita epitassiale da fase vapore (VPE) contemporanea di vari layers π su altrettanti substrati massivi p^+ , al fine di realizzare un set, più omogeneo possibile, di fotodiodi SAM - APD $n^+/p/\pi/p^+$ al silicio.

Mentre il layer di Si cresce sul substrato p^+ (la reazione della riduzione del tetracloruro di silicio $SiCl_4$ in idrogeno sarà discussa nel paragrafo 8.3.1), il diborano B_2H_6 , che è un precursore del drogante p, viene inizialmente adsorbito dalla superficie di Si epitassiale, ovvero la molecola B_2H_6 instaura (complice l'alta temperatura, dell'ordine del migliaio di $^{\circ}C$) un legame chimico con il Si oppure un'interazione fisica, del tipo dipolo elettrico - dipolo elettrico, all'interfase gassosa - solida. Sempre a causa della temperatura alta, il boro viene estratto dalla molecola "veicolante" B_2H_6 ed è assorbito dal Si: il boro, a questo punto, perde un elettrone e lo ione B^+ inizia a diffondere attraverso il Si epitassiale, cresciuto fino a quel momento. Una volta sistematosi in posizione sostituzionale e terminato il processo di epitassia VPE (ovvero giunti nuovamente a temperature ordinarie), lo ione B^+ acquista prima un elettrone, tornando a rivestire il ruolo di atomo drogante B, dopodichè ne cattura un altro, drogando il Si epitassiale di tipo p.

Dunque l'immissione del diborano nel reattore produce, all'interno dello strato π , un profilo pressochè costante di drogante, indicato con "C" nel grafico di figura 110. Tuttavia, a causa della temperatura elevata, una parte del drogante del substrato p^+ , il cui profilo è indicato con "A" nel grafico di figura 110, diffonde termicamente verso π , dando luogo ad un profilo di concentrazione di boro, intorno all'interfaccia π/p^+ , la cui forma è indicata con "B". B rappresenta il noto "autodrogaggio" dello strato epitassiale π per opera del substrato p^+ (retrodiffusione da p^+ verso π). Ciò può costituire un aiuto nel drogare π , qualora si abbia intenzione di immettere una piccola quantità di B_2H_6 nel reattore, oppure un effetto collaterale di cui tener conto, dato che è necessario che il layer sopra p^+ sia quanto meno drogato possibile (p).

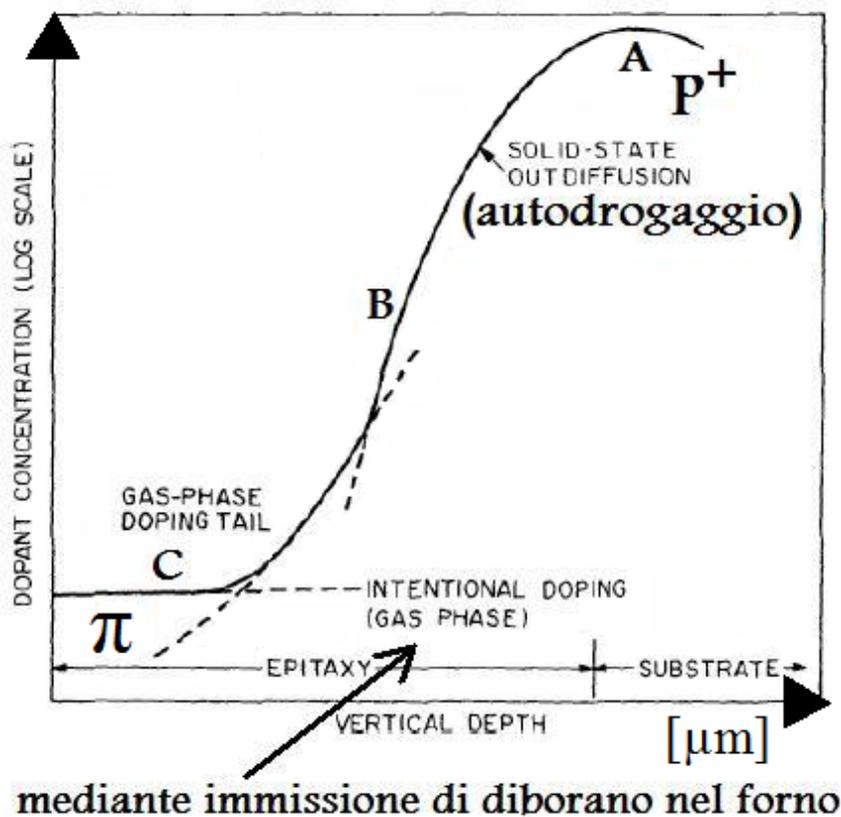


Fig. 110

Profilo della concentrazione di boro nei pressi dell'interfaccia π/p^+ , una volta ultimata la crescita epitassiale di π su p^+ . Si noti la distinzione fra il drogaggio intenzionale, ottenuto mediante l'immissione nel forno del precursore del boro B_2H_6 , e l'autodrogaggio di π dovuto alla retrodiffusione termica di boro da p^+ verso il crescente strato π .

Quindi in π abbiamo due diversi profili di concentrazione di drogante, uno dovuto alla retrodiffusione, l'altro al drogaggio intenzionale ottenuto con il diborano. L'importanza del contributo di ciascuno, nella definizione del profilo risultante, è conseguenza di due fattori: la velocità v della crescita epitassiale del Si e quella v_d della diffusione del drogante B^+ nel Si epitassiale. Se:

$$v < 2\sqrt{v_d}$$

è dominante la retrodiffusione, mentre se accade l'opposto è dominante il drogaggio intenzionale.

Il processo di crescita epitassiale crea uno strato cristallino π con una minore quantità di impurezze rispetto al substrato p^+ , ma con una perfezione reticolare inferiore. In particolare, la struttura cristallina del substrato (quella generalmente usata è la 100, secondo gli indici di Miller) viene ripetuta nel layer π , riproducendo eventuali difetti strutturali. I principali difetti prodotti dalla crescita epitassiale sono qualitativamente mostrati in figura 111.

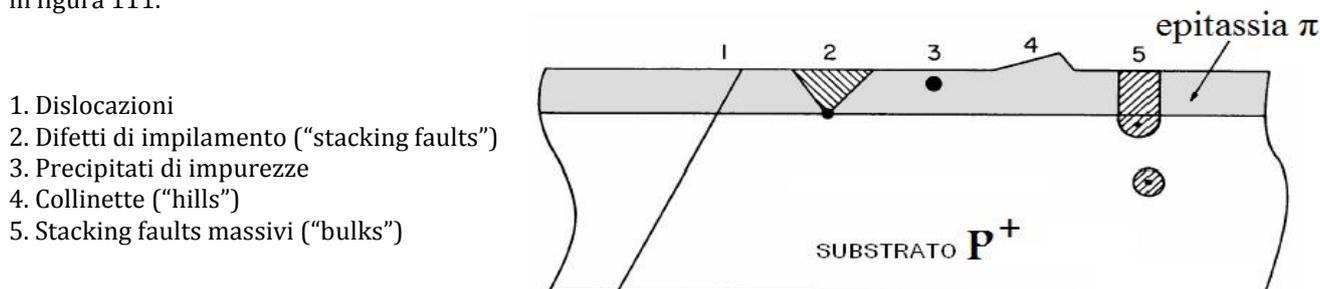


Fig. 111

Rappresentazione dei principali difetti strutturali caratteristici di uno strato cresciuto mediante epitassia

In particolare gli ultimi due si originano da fluttuazioni termiche durante l'accrescimento epitassico.

La well di guadagno p e quella n^+ sono ricavate o per diffusione termica o per impiantazione ionica; al fine di creare uno strato n^+ con spessore dell'ordine dei 100 nm, è opportuno ricorrere alla diffusione termica dell'arsenico, come già spiegato nel paragrafo 7.1.1.

7.1.6) Il fotodiode a valanga SAM $n^+/p/\pi/p^+$ al silicio: il sistema antiriflesso

Nell'espressione [e85] possiamo notare il coefficiente $1 - R$, che indica l'attitudine del sistema, costituito dall'interfaccia aria/finestra dielettrica antiriflesso (che supponiamo costruita con silice SiO_2) e dall'interfaccia contigua finestra antiriflesso/semiconduttore (di silicio), a trasmettere al materiale interno l'intensità luminosa incidente. Nel caso del SAM - APD di silicio, riportato in figura 108, questa "trasmissione" è esprimibile mediante la seguente relazione:

$$T = 1 - R = \frac{n_0 n_{Si}}{|n_0 B + C|^2}$$

dove:

- n_0 = indice di rifrazione dell'aria = $\sqrt{\mu_{rel.aria}(\lambda)\epsilon_{rel.aria}(\lambda)} \approx$ indice di rifrazione del vuoto = 1
- n_{Si} = indice di rifrazione del silicio = $\sqrt{\mu_{rel.Si}(\lambda)\epsilon_{rel.Si}(\lambda)} \approx 3.4 \div 3.5$
- λ = lunghezza d'onda della radiazione luminosa incidente
- B, C = coefficienti che dipendono dai due indici di rifrazione suddetti e da quello dell'ossido SiO_2 che costituisce lo strato antiriflettente: $n_{SiO_2} = \sqrt{\mu_{rel.ossido}(\lambda)\epsilon_{rel.ossido}(\lambda)} \approx 1.4 \div 1.6$

Quindi l'ossido SiO_2 può essere utilizzato sia come strato di passivazione, cioè come "riempitore" delle interruzioni cristallografiche del Si superficiale, così da limitare la corrente di leakage, sia come finestra antiriflettente, volta a ridurre la discontinuità, fra aria e Si, del valore dell'indice di rifrazione n . La discontinuità di n è, infatti, la causa dell'effetto specchio, il quale può causare la riflessione di una consistente frazione della potenza ottica incidente. Pertanto, dopo la definizione della well n^+ , è necessario depositare uno strato antiriflettente, ad esempio di ossido SiO_2 , il cui indice di rifrazione n_{SiO_2} può essere modificato mediante opportuni drogaggi. Se si desidera aumentare n_{SiO_2} , rispetto ai suoi valori tipici (che sono circa $1.4 \div 1.6$, per le lunghezze d'onda λ di vicino infrarosso della luce incidente, utilizzate in comunicazioni ottiche), al fine di avvicinare l'indice di rifrazione della finestra antiriflettente a quello del Si, è necessario immettere nel reattore per CVD, insieme al silano e all'ossigeno ($\text{SiH}_4 + \text{O}_2 \rightarrow \text{SiO}_2 + 2\text{H}_2$, CVD a temperature di $300 - 450 \text{ C}^\circ$), anche una certa percentuale molare di un precursore (AsH_3 , PH_3 , P_2O_5 , vapori di POCl_3) di un drogante del V° gruppo (As, P). In alternativa si può aumentare n_{SiO_2} diffondendo nel reattore CVD una certa percentuale molare di precursori di Ge e Ti, ad esempio il GeO_2 . Se si desidera altresì diminuire n_{SiO_2} , rispetto ai suoi valori tipici, al fine di avvicinare l'indice di rifrazione della finestra antiriflettente a quello dell'aria, è necessario immettere una certa percentuale molare di un precursore di un drogante del III° gruppo, ad esempio il diborano B_2H_6 , che è uno dei più comuni precursori del boro, oppure il B_2O_3 , altro precursore del boro, oppure un precursore del fluoro F, ad esempio il ZrF_4 . Nel grafico semiquantitativo mostrato in figura 112 è riportato il valore dell'indice di rifrazione n_{SiO_2} dell'ossido di silicio SiO_2 in funzione della frazione molare dei principali precursori delle impurezze droganti appena discusse, diffusi all'interno della camera di lavoro durante la deposizione CVD dello strato antiriflettente di ossido. I valori dell'indice di rifrazione mostrati in ordinata si riferiscono a lunghezze d'onda appartenenti allo spettro del vicino infrarosso (NIR), al quale appartengono le tre finestre di attenuazione delle fibre ottiche di silice.

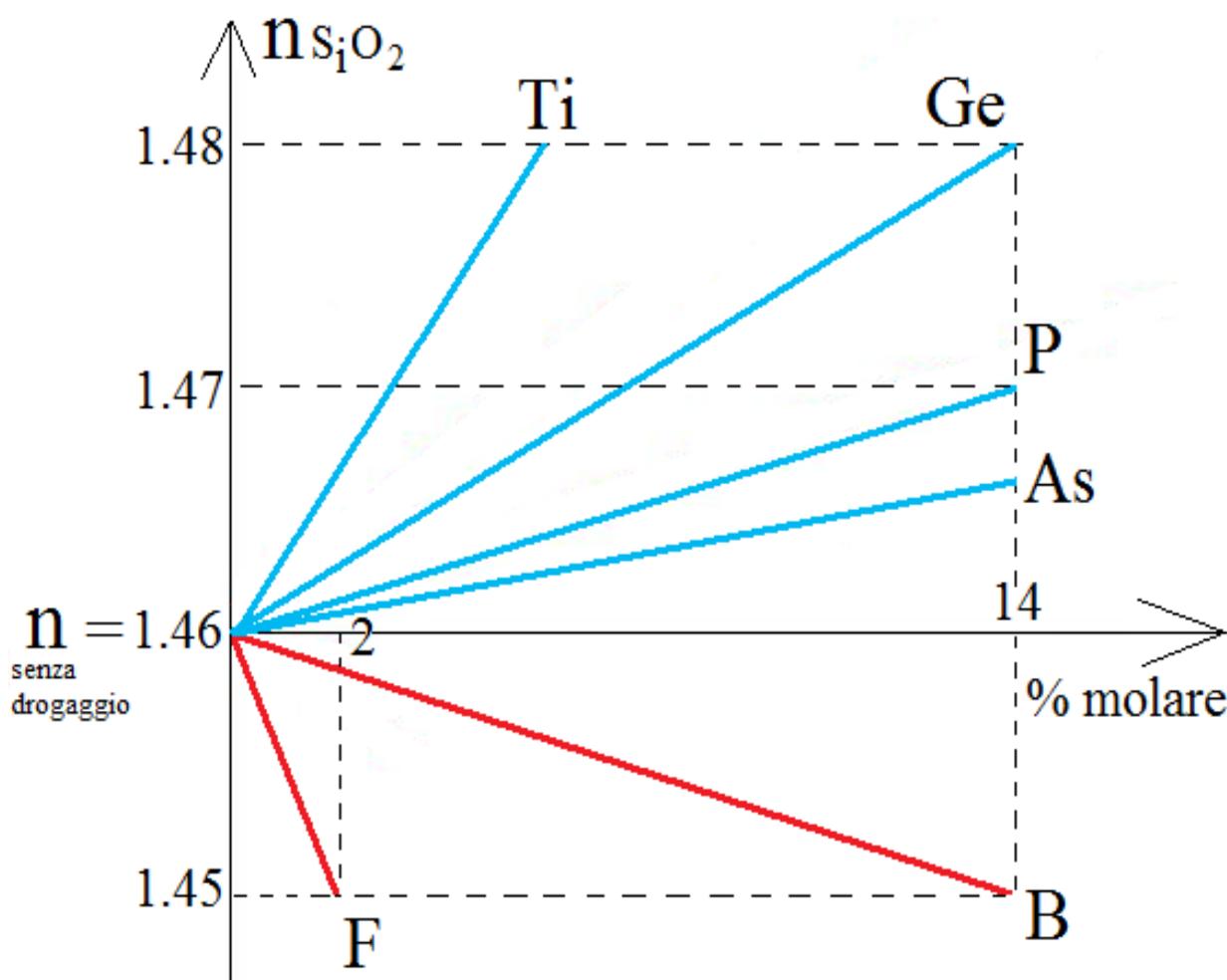


Fig. 112

Grafico semiquantitativo del valore dell'indice di rifrazione dell'ossido di silicio SiO_2 , molto utilizzato per la realizzazione di finestre antiriflesso, in funzione della percentuale molare dei precursori delle impurezze illustrate, nell'ambito della deposizione CVD dello strato antiriflettente di ossido. I valori dell'indice di rifrazione riportati si riferiscono a lunghezze d'onda appartenenti allo spettro del vicino infrarosso (NIR), che è lo spettro ottico utilizzato nel campo delle telecomunicazioni su fibra di silice.

Una finestra antiriflettente di buona qualità è realizzata mediante una struttura vetrosa a multistrato, nella quale ciascun layer è costituito da un film di ossido di silicio SiO_2 (talvolta si utilizza l'ossido di germanio GeO_2) drogato in modo che lo strato abbia un particolare valore dell'indice di rifrazione n_k , compreso tra l'indice di rifrazione n_{k-1} del layer sottostante e l'indice di rifrazione n_{k+1} del layer sovrastante. La finestra multilaminata è pertanto caratterizzata da una gradualità quasi continua dell'indice di rifrazione, attraverso tutto il suo spessore, dall'interfaccia aria/finestra a quella finestra/semiconduttore, e ciò consente, in osservanza alla legge di Snell (si parla infatti di applicazione "quasi continua" della legge di Snell), una buona collimazione dei raggi luminosi incidenti sul fotorivelatore ed una scarsa riflessione della potenza ottica incidente. Questa particolarità realizzativa della finestra antiriflettente partecipa a rendere buona l'efficienza quantica totale del fotodiode. Nella parte a sinistra di figura 113 riportiamo il grafico dell'indice di rifrazione dell'ossido di silicio SiO_2 , privo di drogaggio, in funzione della lunghezza d'onda λ della luce incidente, mentre nella parte a destra mostriamo un esempio molto qualitativo di finestra antiriflettente multilaminata.

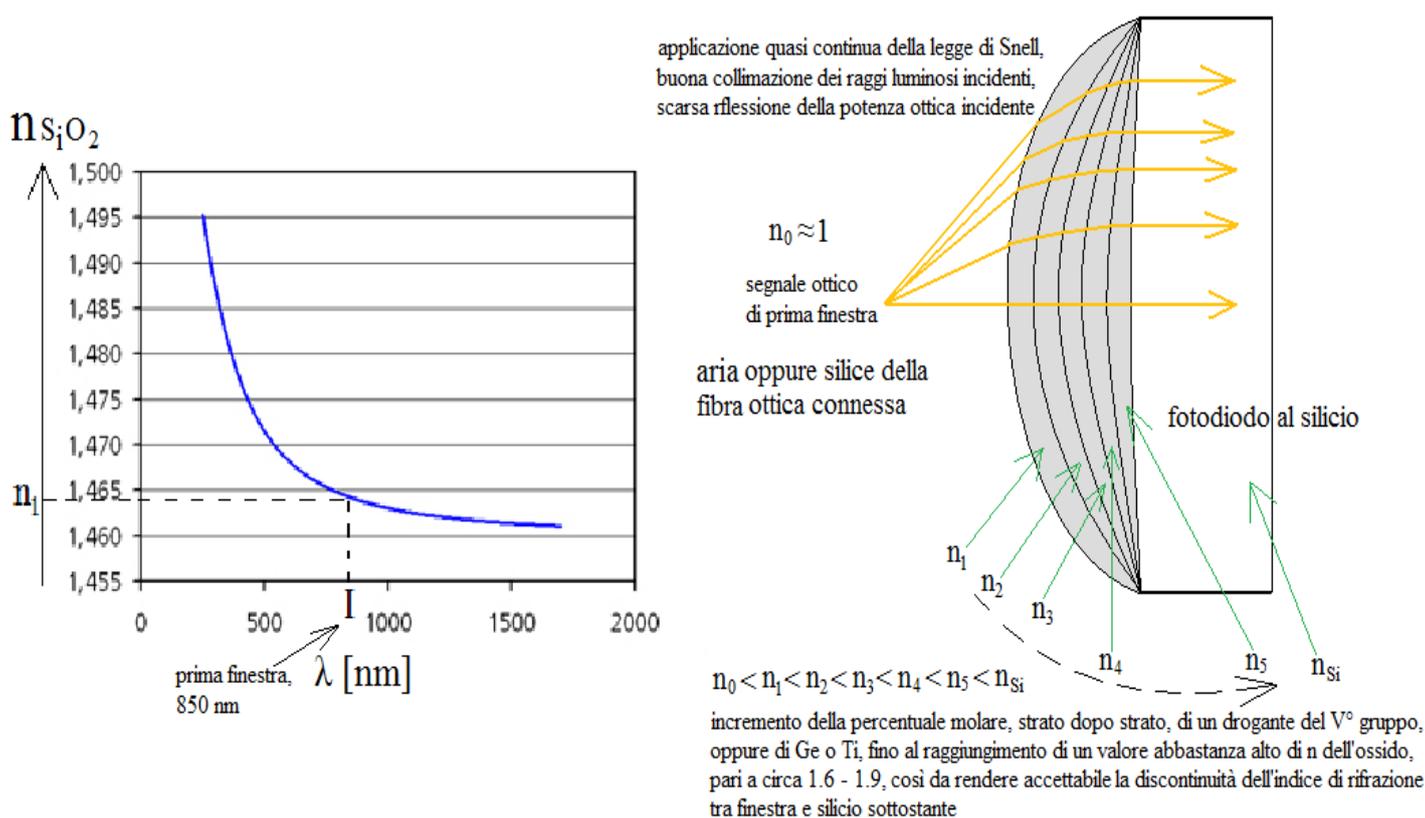


Fig. 113

A sinistra è riportato il grafico dell'indice di rifrazione dell'ossido di silicio SiO_2 , privo di drogaggio, in funzione della lunghezza d'onda λ della luce incidente, mentre a destra è mostrato un esempio molto qualitativo di finestra antiriflettente realizzata mediante una struttura vetrosa multilaminata.

Dunque il coefficiente di trasmittanza T dipende fortemente dalla lunghezza d'onda incidente, dal tipo di ossido e dal suo spessore: variando queste condizioni è stata misurata sperimentalmente una perdita di potenza ottica incidente, a causa della riflessione, oscillante fra il 10 e il 40 %. La realizzazione dello strato antiriflesso deve quindi tenere conto dell'intervallo di frequenza tipico delle applicazioni a cui il dispositivo è destinato. Ad esempio, se utilizziamo strutture complesse a matrice (che esamineremo nel capitolo 8) costituite, nella loro cella fondamentale, dal SAM - APD al Si visto finora, bisogna tener conto che, in ambiente PET, i fotoni a 511 keV (raggi gamma), emessi dal tessuto sottoposto a mapping, non incidono direttamente sul rivelatore, bensì su un cristallo scintillatore LYSO. Questo, mediante meccanismi di cui rimandiamo la spiegazione al capitolo 8, emette luce nella regione del blu (350 - 500 nm), con un picco di massima emissione intorno ai 420 nm. Dunque il tipo di ossido antiriflettente ed il suo spessore devono essere tali da massimizzare T solo per quell'intervallo di lunghezze d'onda. Successivamente, in base al coefficiente di assorbimento α per il Si (che a circa 400 nm vale $5.4 \cdot 10^6 \text{ cm}^{-1}$ ed il cui inverso coincide col cammino libero medio del fotone nel Si), occorre determinare la profondità richiesta per un completo assorbimento (circa $2.3 \mu\text{m}$). In base a quest'ultimo dato appare evidente la necessità di avere n^+ e p più sottili possibile. Con l'utilizzo di scintillatori LYSO e l'accurato progetto dello strato di ossido, si riesce ad ottenere una $QE(\lambda = 420 \text{ nm})$ pari a circa il 95%.

7.2.1.1) Il fotodiodo a valanga SAM $p^+/n/n^-/n^+$ ad eterostruttura: le ragioni del suo utilizzo ed il suo funzionamento

Finora abbiamo discusso il SAM - APD $n^+/p/\pi/p^+$ di silicio, che riprenderemo per descrivere i SiPMs che potrebbero essere usati nella PET. La struttura suddetta è utilizzabile per ricevere radiazioni luminose aventi lunghezza d'onda compresa fra il violetto/blu ed il vicino infrarosso (infatti fino a lunghezze d'onda λ di circa $1 \mu\text{m}$ la responsivity del Si è superiore a 0.4 A/W , per cui l'ampiezza della risposta elettrica è accettabile). Tuttavia esistono numerose applicazioni, alcune delle quali di diagnostica medica, come ad esempio le tecniche di mapping termico di tessuti organici o di oggetti, nelle quali è importante rivelare efficacemente fotoni nel medio o lontano infrarosso; tale esigenza è presente anche nell'ambito delle telecomunicazioni su fibra ottica.

Ad un capo della fibra si trova un emettitore di luce (un LED o meglio ancora un LASER, i quali convertono un segnale elettrico in uno luminoso) che invia stringhe di bit, ciascuna corrispondente ad un particolare livello di ampiezza assunto, nel tempo, dal segnale elettrico che pilota l'emettitore (ovvero il segnale modulante, il quale contiene l'informazione da trasmettere). Il segnale luminoso informativo, così modulato (modulazione di ampiezza "ASK") ed "inserito" nel nucleo della fibra, viaggia lungo la stessa, subendo distorsioni temporali di varia natura, fino ad arrivare all'altro capo della guida ottica, dove è posizionato un fotorivelatore, il quale riconverte il segnale ottico in un segnale elettrico, che una volta demodulato deve essere il più possibile aderente a quello modulante. La frequenza portante ottica, intorno alla quale si concentra il contenuto spettrale del segnale trasmesso, deve appartenere ad una delle tre finestre presso le quali l'ossido di silicio, con cui si realizzano le fibre, presenta un coefficiente di assorbimento $\alpha(\lambda)$ molto basso (misurato, nel settore delle comunicazioni ottiche su fibra, in dB/Km). L'utilizzo della seconda e della terza finestra (centrate, rispettivamente, in 1.33 e 1.55 μm) impone, come emettitori di luce, dei costosi LEDs o degli ancora più onerosi LASERS, entrambi costituiti, fundamentalmente, da delle eterogiunzioni del tipo $\text{InP}/\text{In}_{(1-x)}\text{Ga}_x\text{As}_{(1-y)}\text{P}_y/\text{InP}$, e come fotorivelatori, all'altro capo della fibra, dei dispositivi caratterizzati da una buona risposta elettrica, ovvero da valori di $\alpha(\lambda)$, e quindi di $\kappa(\lambda)$, accettabili, presso quelle lunghezze d'onda del vicino - medio infrarosso.

Fotoni con una lunghezza d'onda di 1.5 μm (terza finestra) hanno un'energia di circa 0.828 eV, inferiore all' E_{gap} del Si, e infatti la sua responsivity presso quelle lunghezze d'onda è nulla. Quindi sembrerebbe necessario costruire un APD, ad omogiunzioni, utilizzando un semiconduttore il cui gap sia compatibile con quelle energie fotoniche: ad esempio il $\text{In}_{(0.57)}\text{Ga}_{(0.43)}\text{As}$, che ha una E_{gap} di circa 0.75 eV, oppure il Ge, la cui E_{gap} è di circa 0.67 eV. Tuttavia questi materiali sono caratterizzati, a quelle lunghezze d'onda, da un $K_a \approx 1$ ($\alpha_h \approx \alpha_e$), per cui, fissato M, dal massimo fattore di rumore in eccesso possibile. Questo inconveniente si risolve realizzando eterogiunzioni fra materiali cristallograficamente compatibili, depositabili mediante epitassie fattibili, secondo le regole e gli standard fissati dalla già citata "ingegneria delle bande".

Osserviamo la struttura $p^+/n/n^-/n^+$ riportata in figura 114:

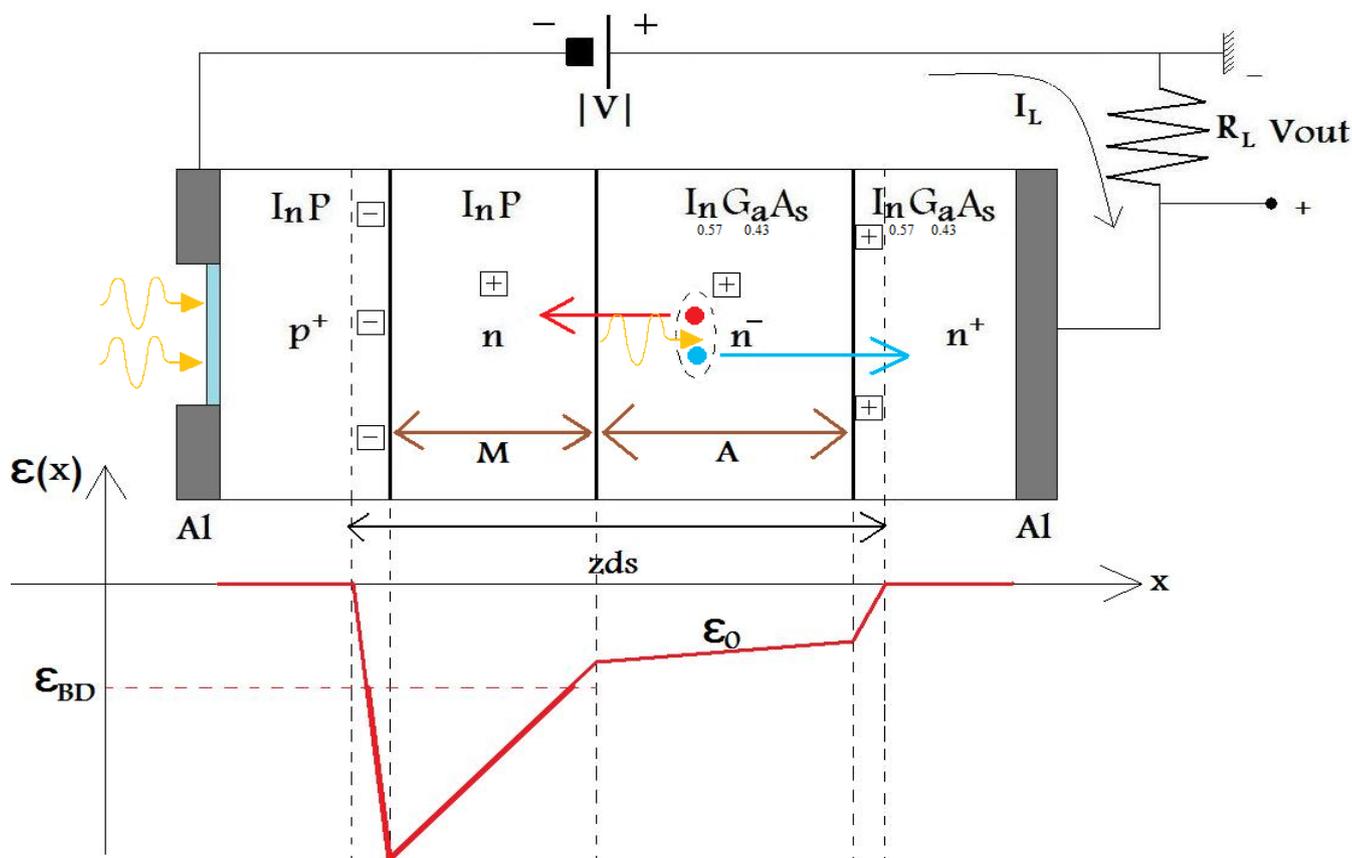


Fig. 114

Rappresentazione schematica di un SAM - APD ad eterogiunzione, costruito a struttura $p^+/n/n^-/n^+$, collegato al circuito esterno di polarizzazione inversa e lettura, e sottoposto ad illuminazione. Riportiamo anche il grafico del campo elettrico $\epsilon(x)$, il quale supera la soglia ϵ_{BD} all'interno della regione di moltiplicazione M. Questa coincide con lo strato n, realizzato in InP, materiale nel quale la moltiplicazione a valanga è prodotta quasi esclusivamente dalle lacune: ciò abbassa la rumorosità del fotodiode.

I fotoni incidono sulla finestra antiriflettente, attraversandola in buon numero, per poi transitare attraverso i layers p^+ ed n : come già visto analizzando i PIN, il InP possiede un gap diretto di circa 1.35 eV, per cui i fotoni di terza finestra non possono essere assorbiti in queste regioni ($E_\gamma < E_{\text{gap}}$). Appena questi raggiungono lo strato di $\text{In}_{(0.57)}\text{Ga}_{(0.43)}\text{As}$ (gap diretto, $E_{\text{gap}} \approx 0.75$ eV), vengono assorbiti con una QE “teorica” (cioè massima) di circa l’80%, ma effettiva, ossia comprensiva di perdite, specialmente superficiali, di circa il 30% ÷ 40%, almeno per dispositivi di qualità commerciale. I fotoelettroni primari vengono trascinati verso n^+ senza innescare alcun evento ionizzante, essendo $\varepsilon_0 < \varepsilon_{\text{BD}}$, e quindi vanno a sostenere la fotocorrente sul circuito di polarizzazione. Le fotolacune primarie vengono trascinate in senso opposto, senza produrre ionizzazioni, finché si trovano nella zona di assorbimento. Appena queste entrano nella regione di guadagno ($\varepsilon > \varepsilon_{\text{BD}}$) iniziano a innescare la valanga, la cui formazione è dovuta unicamente a loro: nel InP, infatti, $\alpha_h \gg \alpha_e$, per cui $K_a \gg 1$, pertanto la rumorosità è bassa anche per guadagni M elevati. Per questa struttura valgono le stesse approssimazioni e gli stessi calcoli (per esempio per M) che abbiamo visto per la struttura duale in Si ($n^+/p/\pi/p^+$). Le fotolacune primarie e secondarie sono raccolte in p^+ , i fotoelettroni secondari sono trascinati fino a n^+ . Per giunta il InP sopporta alti valori di campo, prima che le sue bande si pieghino a tal punto da promuovere, per assottigliamento delle barriere di potenziale, le correnti di tunnel del breakdown Zener.

7.2.1.2) Il fotodiode a valanga SAM $n^+/p/\pi/p^+$ ad eterostruttura GaAs/GaAs/ $\text{Cd}_{(x)}\text{Hg}_{(1-x)}\text{Te}/\text{Si}$ utilizzato per la fotorivelazione di radiazioni ottiche nel medio - lontano infrarosso

Le osservazioni fatte finora consentono di spiegare sinteticamente il funzionamento di un fotorivelatore SAM – APD $n^+/p/\pi/p^+$ ad eterogiunzioni, dove il layer n^+ e quello p di moltiplicazione a valanga sono realizzati in GaAs, il layer π di assorbimento ottico è realizzato in $\text{Cd}_{(x)}\text{Hg}_{(1-x)}\text{Te}$ ed il substrato p^+ in Si. Iniziamo la trattazione spiegando la proprietà più importante del composto ternario con cui è costruito il layer π .

Il tellurio Te è un elemento del VI° gruppo, il mercurio Hg ed il cadmio Cd sono due metalli appartenenti alla regione della tavola periodica nota come “terre rare”. Il composto ternario $\text{Cd}_{(x)}\text{Hg}_{(1-x)}\text{Te}$, in linea generale, può essere definito un semiconduttore a gap diretto; tale cristallo può essere cresciuto mediante MBE controllando la frazione molare x di Cd. È stato provato, sia sperimentalmente che attraverso il metodo di calcolo delle bande di energia TBM, che il diagramma a bande $E(k)$ del $\text{Cd}_{(x)}\text{Hg}_{(1-x)}\text{Te}$ varia abbastanza considerevolmente al variare della frazione molare x di Cd: in particolare è stata osservata una dipendenza approssimativamente lineare dell’energia di gap diretto $E_{\text{gap}}(x)$ in funzione di x . Questa dipendenza è mostrata nel grafico semiquantitativo di figura 115.

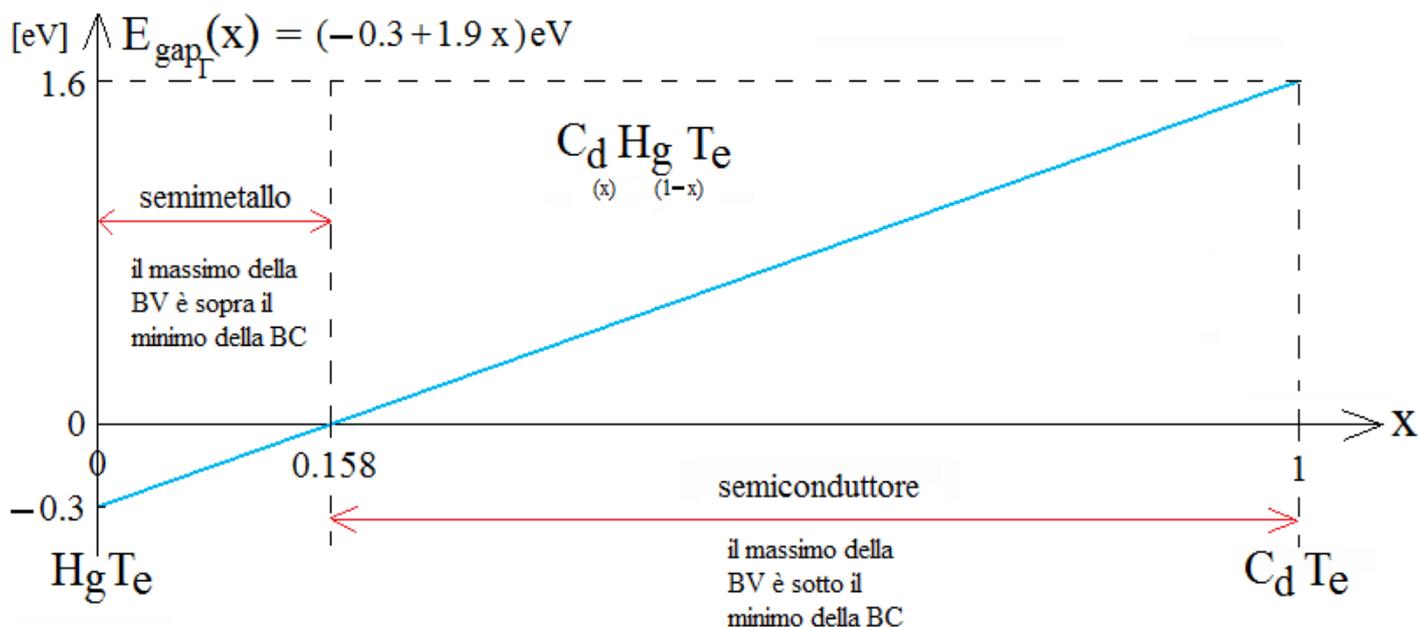


Fig. 115

Grafico semiquantitativo della dipendenza, approssimativamente lineare, dell’energia di gap diretto $E_{\text{gap}}(x)$ del cristallo $\text{Cd}_{(x)}\text{Hg}_{(1-x)}\text{Te}$ in funzione della frazione molare x di cadmio Cd. Per $x < 0.158$ il composto ternario è definito “semimetallo”, poiché la parte superiore della BV è sovrapposta alla parte inferiore della BC (il gap è negativo), pertanto il cristallo può condurre corrente anche a $T = 0$ K, mentre per $x > 0.158$ il composto ternario è definito “semiconduttore”, dato che non c’è alcuna sovrapposizione fra la parte superiore della BV e la parte inferiore della BC (il gap è positivo), per cui il cristallo può condurre corrente solo se $T > 0$ K.

Per x compresa approssimativamente fra 0.158 ed 1 il cristallo $\text{Cd}_{(x)}\text{Hg}_{(1-x)}\text{Te}$ può essere definito a tutti gli effetti un semiconduttore a gap diretto, poiché osservandone il diagramma a bande $E - k$ si può notare che presso il punto di simmetria Γ il massimo della banda di valenza si trova al di sotto del minimo della banda di conduzione. In questo caso il cristallo possiede il gap positivo $E_{\text{gap}\Gamma}$ tipico di tutti i semiconduttori, la cui larghezza dipende da x . Per $x \rightarrow 1$ il cristallo cresciuto mediante MBE è il composto binario CdTe , il cui gap diretto ha una larghezza pari a $E_{\text{gap}\Gamma} \approx 1.6$ eV. Per x inferiore a 0.158 il cristallo $\text{Cd}_{(x)}\text{Hg}_{(1-x)}\text{Te}$ non può essere definito un semiconduttore, poiché osservandone il diagramma a bande $E - k$ si può notare che presso il punto di simmetria Γ il massimo della banda di valenza si trova al di sopra del minimo della banda di conduzione: questa parziale sovrapposizione (“bandedges overlap” – sovrapposizione degli estremi della BC e della BV) della parte superiore della BV con la parte inferiore della BC comporta, come si evince dal grafico di figura 115, un gap negativo ($E_{\text{gap}\Gamma} < 0$) presso il punto Γ , la cui larghezza dipende sempre dalla frazione molare x di Cd. Per $x \rightarrow 0$ il cristallo cresciuto è il composto binario HgTe , il cui gap diretto ha una larghezza pari a $E_{\text{gap}\Gamma} \approx -0.3$ eV, ovvero il massimo della BV si trova sopra al minimo della BC di circa 0.3 eV. In figura 116 riportiamo il diagramma a bande del composto ternario $\text{Cd}_{(x)}\text{Hg}_{(1-x)}\text{Te}$ con $x = 0.316$, cioè del semiconduttore a gap diretto $\text{Cd}_{(0.316)}\text{Hg}_{(0.684)}\text{Te}$, il cui gap diretto ha una larghezza pari a $E_{\text{gap}\Gamma} \approx 0.3$ eV. È inoltre mostrato un ingrandimento della regione del diagramma $E - k$ di maggior interesse fotonico, nonché la stessa regione nel caso di frazione molare di cadmio $x = 0$: in quest’ultimo diagramma $E(k)$ è ben visibile la sovrapposizione dell’estremo superiore della BV (“BV - edge”) con l’estremo inferiore della BC (“BC - edge”).

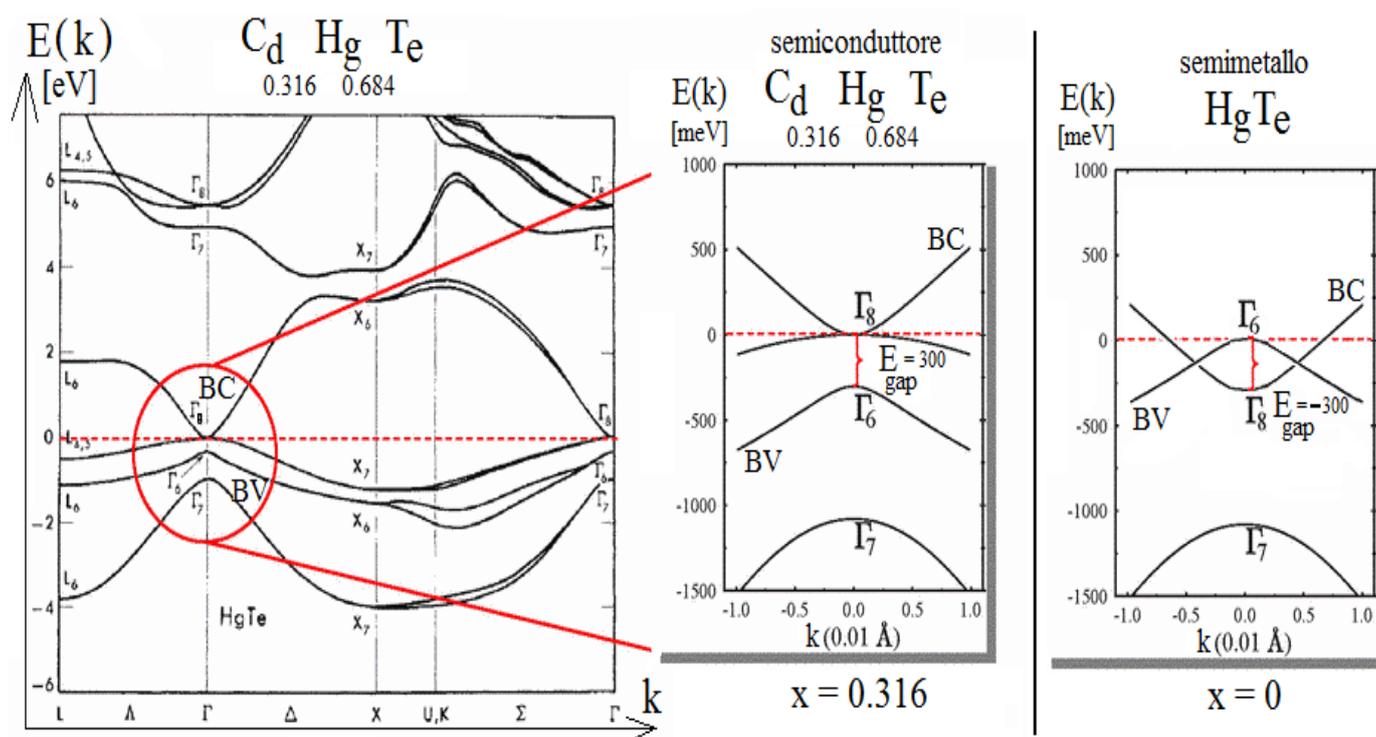


Fig. 116

A sinistra riportiamo il diagramma a bande di energia $E - k$ del composto ternario $\text{Cd}_{(x)}\text{Hg}_{(1-x)}\text{Te}$ con $x = 0.316$, cioè del semiconduttore a gap diretto $\text{Cd}_{(0.316)}\text{Hg}_{(0.684)}\text{Te}$, il cui gap diretto ha una larghezza pari a $E_{\text{gap}\Gamma} \approx 0.3$ eV. Al centro è mostrato un ingrandimento della regione del diagramma $E - k$ di maggior interesse fotonico (si noti il gap diretto positivo, che conferma la definizione di semiconduttore), nonché la stessa regione nel caso di frazione molare di cadmio $x = 0$, nella parte destra della figura. In quest’ultimo caso nel diagramma $E(k)$ è ben visibile la sovrapposizione dell’estremo superiore Γ_6 della BV (“BV - edge”) con l’estremo inferiore Γ_8 della BC (“BC - edge”): si noti il gap diretto negativo pari a $E_{\text{gap}\Gamma} \approx -0.3$ eV.

Il composto $\text{Cd}_{(x)}\text{Hg}_{(1-x)}\text{Te}$ con $x < 0.158$, che a causa della sovrapposizione appena spiegata non può essere definito semiconduttore, è invece chiamato “semimetallo”. La ragione è legata alla possibilità, da parte del $\text{Cd}_{(x)}\text{Hg}_{(1-x)}\text{Te}$, con $x < 0.158$, di condurre corrente elettrica anche a temperatura pari a $T = 0 \text{ K}$, come avviene nei metalli. Si osservi in proposito la figura 117. Supponiamo di applicare una tensione ai terminali di un cristallo di $\text{Cd}_{(x)}\text{Hg}_{(1-x)}\text{Te}$, con $x < 0.158$, termostato ipoteticamente alla zero assoluto, o comunque ad una temperatura prossima a $T = 0 \text{ K}$. Gli stati elettronici localizzati lungo la curva $E(k)$ che costituisce l’arco superiore della BV, indicato in figura 116 ed in figura 117 con Γ_6 , ospitano elettroni atomici di valenza che possono incrementare la loro quantità di moto, il loro impulso $p = \hbar k$, cioè il loro vettore d’onda k , sotto l’azione accelerante del campo elettrico ε dovuto alla polarizzazione imposta dall’esterno. Questi elettroni possono incrementare il proprio k , contribuendo così alla conduzione, perché nei pressi dell’arco superiore Γ_6 della BV ciascuno stato iniziale $|i\rangle$, avente vettore d’onda k_i , in cui può trovarsi un elettrone di valenza ha la stessa energia totale E^* di almeno uno stato finale $|f\rangle$, avente vettore d’onda k_f , localizzato presso l’arco inferiore della BC, che in figura 116 ed in figura 117 è indicato con Γ_8 . Pertanto un elettrone può transitare dallo stato $|i\rangle$ a quello $|f\rangle$ sotto l’azione del campo di polarizzazione e il fatto che $|k_f|$ sia maggiore di $|k_i|$ comporta che l’elettrone, in seguito alla transizione $|i\rangle \rightarrow |f\rangle$, ha subito l’azione di una forza elettrica $F = q\varepsilon$ accelerante, cioè che l’elettrone ha preso parte al processo di conduzione elettrica all’interno del cristallo. La transizione $k_i \rightarrow k_f$ è possibile anche per $T = 0 \text{ K}$, poiché non c’è bisogno di fornire energia termica che consenta agli elettroni di superare un gap energetico, come avviene nei semiconduttori, per avere a disposizione stati $|k\rangle$ permessi grazie ai quali poter incrementare il loro numero d’onda e partecipare così alla conduzione.

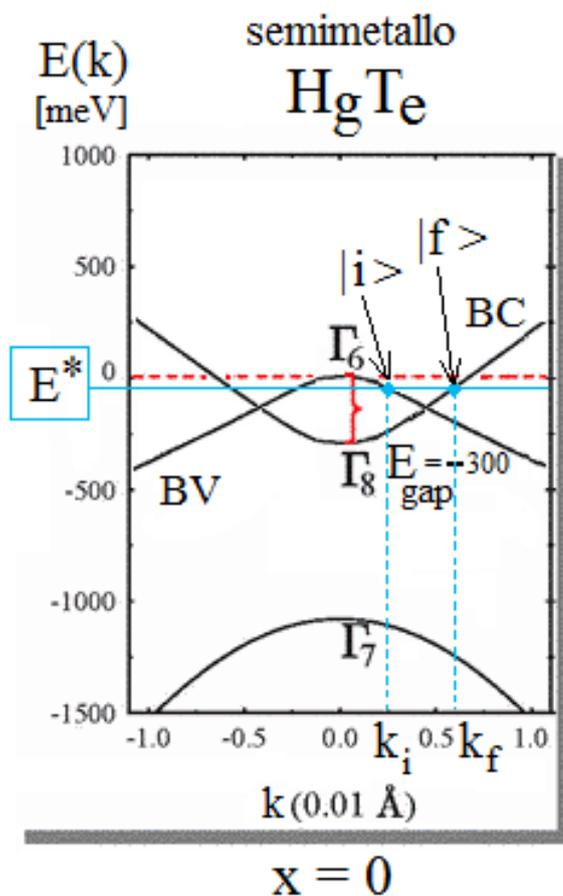


Fig. 117

Spiegazione qualitativa, attraverso la regione di rilevanza fotonica del diagramma $E - k$ del cristallo HgTe, del motivo fisico per cui il composto binario HgTe è convenzionalmente definito “semimetallo” anziché semiconduttore.

Per rivelare radiazioni ottiche FIR, ovvero appartenenti allo spettro del lontano infrarosso, ovvero segnali ottici molto poco energetici (grande lunghezza d'onda λ , bassa frequenza ν), possiamo pertanto utilizzare uno strato realizzato in $\text{Cd}_{(x)}\text{Hg}_{(1-x)}\text{Te}$, con x leggermente maggiore di 0.158, cosicché la larghezza $E_{\text{gap}\Gamma}$ del gap diretto (positivo) del semiconduttore ternario sia piccola a sufficienza da poter assorbire i fotoni infrarossi incidenti. A seconda dell'applicazione specifica alla quale è destinato il fotorivelatore, cioè a seconda della frequenza FIR del segnale ottico da rivelare, ovvero dell'energia ottica $h\nu$ incidente, sceglieremo un opportuno valore dell'energia di gap diretto $E_{\text{gap}\Gamma}$, che deve essere uguale o leggermente minore di $h\nu$; conseguentemente sceglieremo il valore della frazione molare x di cadmio con cui deve essere cresciuto, mediante MBE, il cristallo semiconduttore ternario $\text{Cd}_{(x)}\text{Hg}_{(1-x)}\text{Te}$ con cui realizzare lo strato π di assorbimento ottico del fotodiodo. Tale layer è cresciuto su un substrato massivo p^+ di Si, come si può notare dalla figura 118. È stato scelto il silicio in virtù del fatto che la crescita di un layer di Si mediante VPE oppure la crescita di un lingotto di Si mediante la tecnica CZ, la tecnica FZ oppure ancora quella di Bridgeman, dal quale si taglia un wafer da cui si ricava il layer, è più semplice ed economica dell'analoga crescita di un layer di un semiconduttore composto o della crescita di un lingotto realizzato con quello stesso semiconduttore composto. Inoltre, al di là delle motivazioni puramente economiche, lo stato dell'arte del trattamento del silicio è senz'altro il più maturo e consolidato a livello tecnologico, pertanto è da preferirsi, almeno in questo caso, un substrato massivo p^+ realizzato in silicio piuttosto che in qualunque altro semiconduttore, semplice o composto. All'interfaccia π/p^+ sono presenti forti stress meccanici, causati dal forte mismatch reticolare fra il $\text{Cd}_{(x)}\text{Hg}_{(1-x)}\text{Te}$ ed il Si. Ciò sicuramente comporta una probabilità di transizione delle fotolacune da π a p^+ non unitaria, tuttavia questa non idealità può essere considerata non particolarmente rilevante se la regione di assorbimento ottico $A = \pi$ ed il substrato massivo sono sufficientemente spessi. Inoltre il diagramma a bande di energia $E - k$ del $\text{Cd}_{(x)}\text{Hg}_{(1-x)}\text{Te}$, nei pressi dell'interfaccia π/p^+ , risulta alterato rispetto al diagramma $E(k)$ del semiconduttore ternario all'interno del bulk dello strato π , come spiegato nel paragrafo 5.9.1 e come qualitativamente rappresentato in figura 61. Questa alterazione del diagramma a bande $E(k)$ consente l'assorbimento ottico, confinato presso la regione del layer π limitrofa al substrato massivo p^+ , di segnali luminosi aventi frequenze FIR un po' diverse da quelle assorbite dallo stesso semiconduttore ternario presso il bulk del layer π . Il layer n^+ dal quale entra la luce FIR e lo strato di moltiplicazione a valanga $M = p$ sono realizzati in GaAs. Questo semiconduttore binario possiede un gap diretto, a temperatura ambiente, pari a circa 1.43 eV, pertanto i fotoni infrarossi attraversano i due strati senza essere assorbiti, ovvero il GaAs presenta trasparenza passiva nei confronti delle radiazioni ottiche FIR incidenti. I fotoni, dopo aver attraversato n^+ e p , raggiungono lo strato π , vengono assorbiti in virtù del piccolo gap diretto del $\text{Cd}_{(x)}\text{Hg}_{(1-x)}\text{Te}$ e producono fotocoppie elettrone/lacuna. Le fotolacune vengono trascinate dal campo elettrico, legato alla polarizzazione inversa del fotodiodo, raggiungendo il bulk del substrato p^+ e contribuendo a sostenere la fotocorrente I_L . I fotoelettroni raggiungono l'interfaccia opposta p/π , la attraversano, anche se con una probabilità di transizione non unitaria, a causa degli stress meccanici reticolari esistenti fra GaAs e $\text{Cd}_{(x)}\text{Hg}_{(1-x)}\text{Te}$, e penetrano all'interno del layer di moltiplicazione a valanga p . Qui gli elettroni possiedono una massa efficace per la mobilità molto piccola ($m_e^* \approx 0.067m_0$ nella valle Γ), a causa della marcata concavità della curva $E(k)$ relativa alla banda di conduzione nel GaAs, ovvero i fotoelettroni hanno una grande mobilità μ_n , quindi subiscono una considerevole accelerazione da parte del campo $\varepsilon > \varepsilon_{BD}$, formando pertanto una considerevole fotocorrente I_L . Inoltre nel GaAs la massa efficace degli elettroni di conduzione, in questo caso di fotoconduzione, che è circa pari a $0.067m_0$, è decisamente più piccola della massa efficace delle lacune di fotoconduzione, cioè prevalentemente le fotolacune HH (fotolacune pesanti), che è circa pari a $0.45m_0$, pertanto l'accelerazione che subiscono i fotoelettroni primari e secondari dentro il layer $M = p$ è considerevolmente superiore all'accelerazione che subiscono, nella stessa zona, le fotolacune secondarie. Ciò comporta che nel layer $M = p$ l'attitudine dei fotoelettroni all'impattazione ionica è molto maggiore di quella delle fotolacune, quindi $\alpha_e \gg \alpha_h$, $K_a \ll 1$, $K \rightarrow 0$, pertanto il fattore di rumore in eccesso $F(M)$ del fotodiodo è basso (buon SNR), come spiegato nel paragrafo 6.5.

Una delle applicazioni più interessanti del fotodiodo appena discusso, riportato qualitativamente in figura 118, è la realizzazione di apparecchiature utilizzate per la visione notturna (occhiali ad infrarosso).

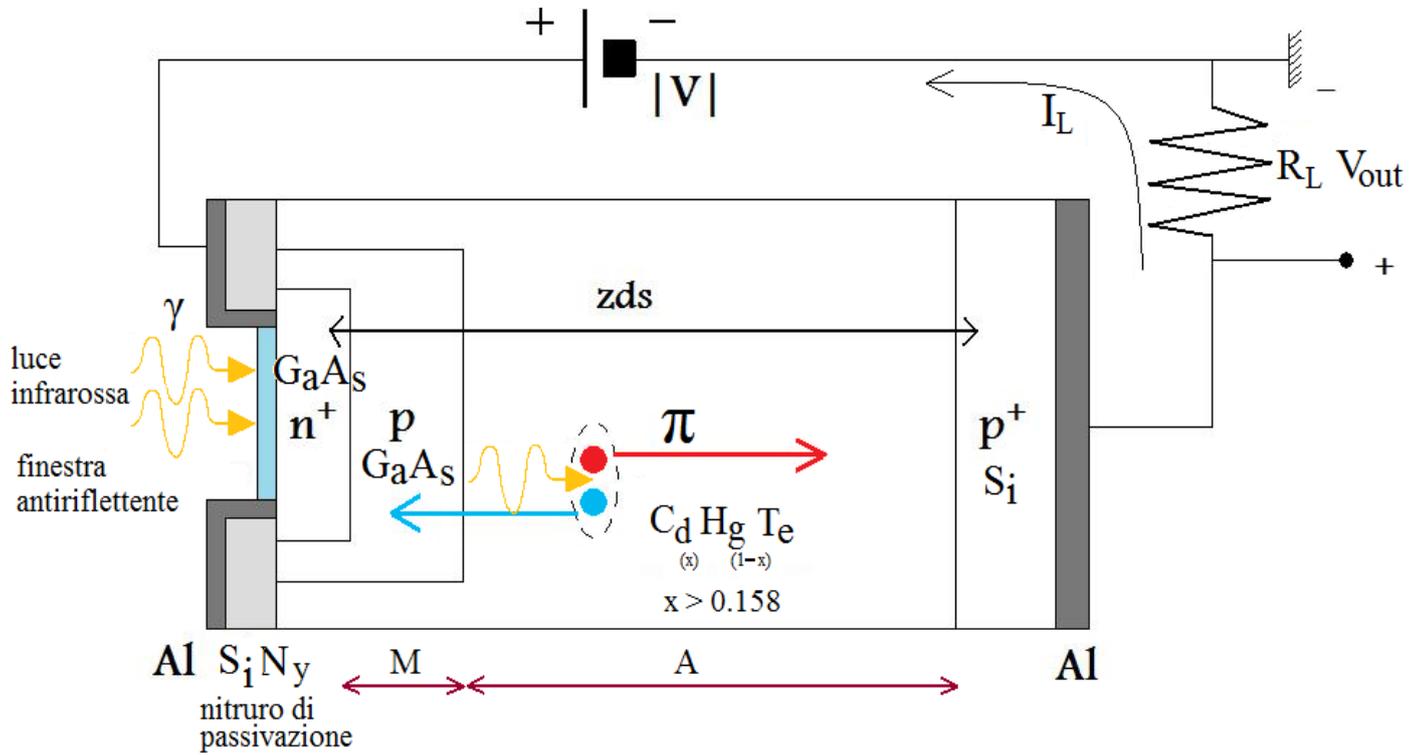


Fig. 118

Rappresentazione schematica di un fotodiode SAM - APD ad eterogiunzioni, costruito a struttura $n^+/p/\pi/p^+$, collegato al circuito esterno di polarizzazione inversa e lettura, e sottoposto ad illuminazione infrarossa. Il layer n^+ e quello di moltiplicazione a valanga $M = p$ sono realizzati in GaAs, il layer di assorbimento ottico $A = \pi$ è realizzato in $Cd_{(x)}Hg_{(1-x)}Te$, con $x > 0.158$, mentre il substrato massivo p^+ in Si. I contatti metallici sono di alluminio, la finestra antiriflettente è costruita con uno strato vetroso multilaminato, mentre il materiale isolante di passivazione è un nitrato di silicio SiN_y , dato che il GaAs, al contrario del Si, non ha ossidi stabili che possano essere cresciuti termicamente.

7.2.2) Il fotodiode a valanga SAM $p^+/n/n^-/n^+$ ad eterostruttura graduata: esempio di "Mesa - Etched" SAGM - APD commerciale

C'è tuttavia un problema all'interfaccia $InP/In_{(0.57)}Ga_{(0.43)}As$, mostrato qualitativamente in figura 119, a sinistra:

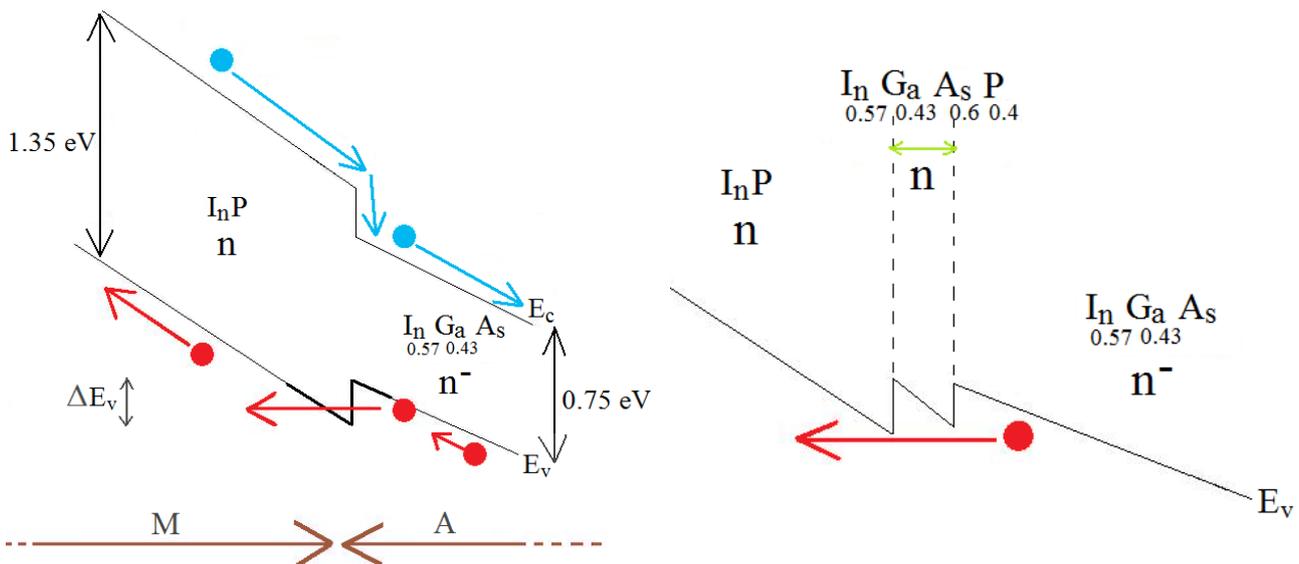


Fig. 119

A sinistra riportiamo la porzione del diagramma a bande qualitativo del SAM - APD $p^+/n/n^-/n^+$ ad eterogiunzione, relativa all'interfaccia $InP/In_{(0.57)}Ga_{(0.43)}As$, in condizioni di polarizzazione inversa ed illuminazione. Si può notare che la barriera di potenziale ΔE_v , dovuta alla grande differenza di E_{gap} fra i due semiconduttori, provoca un accumulo di lacune all'interfaccia regione di moltiplicazione M/regione di assorbimento A, con conseguente aumento dell'inerzia elettrica del dispositivo. A destra riportiamo la soluzione: l'interposizione di un sottile layer $In_{(0.57)}Ga_{(0.43)}As_{(0.6)}P_{(0.4)}$, il cui valore di E_{gap} è compreso fra 1.35 eV e 0.75 eV , permette un "accesso scalinato" della zona M da parte delle fotolacune primarie, facilitando il superamento della barriera.

I fotoelettroni primari, generati nella regione di assorbimento, e quelli secondari, nati per impatto ionico delle lacune nella zona di guadagno, non presentano problemi “di scivolamento” lungo E_c . Lo stesso vale per le lacune secondarie lungo E_v : il problema consiste nella barriera di potenziale ΔE_v , dovuta alla grande differenza di gap fra i due semiconduttori, che le lacune primarie vedono quando sono in procinto di entrare nella regione a campo elevato. Queste hanno certamente una buona probabilità di attraversare la barriera per effetto tunnel, tanto maggiore quanto più è elevata la polarizzazione inversa e dunque minore la larghezza media della barriera. Tuttavia la non unitarietà del coefficiente di trasmissione T provoca un accumulo di lacune all’interfaccia, con conseguente diminuzione della velocità di risposta del fotorivelatore. È opportuno realizzare la struttura schematizzata nel disegno a destra di figura 119, ovvero interporre un sottile layer del quaternario $I_n(0.57)Ga(0.43)As(0.6)P(0.4)$: questo, infatti, avendo un gap intermedio ($0.75 \text{ eV} < E_{\text{gap}} < 1.35 \text{ eV}$), offre alle fotolacune uno scalino energetico intermedio che ne facilita la transizione. Si parla dunque di SAGM – APD (G sta per “Grading” - graduato). In figura 120 riportiamo lo schema di un dispositivo largamente usato per ricevere segnali luminosi, da fibra, in seconda e terza finestra, sagomato a forma MESA: il fotodiodo “Mesa – Etched” SAGM – APD.

- 1 = substrato I_nP tipo p^+
 - 2 = I_nP tipo p^+ ($2 - 3 \mu\text{m}$) strato buffer epitassiale
 - 3 = I_nP tipo n ($2 - 3 \mu\text{m}$) strato di moltiplicazione a valanga
 - 4 = $I_n(0.57)Ga(0.43)As(0.6)P(0.4)$ tipo n ($< 1 \mu\text{m}$) strato a gradino di potenziale (“Graded layer”)
 - 5 = $I_n(0.53)Ga(0.47)As$ tipo n^- ($5 - 10 \mu\text{m}$) strato di assorbimento
- Prodotto guadagno – banda = qualche GHz

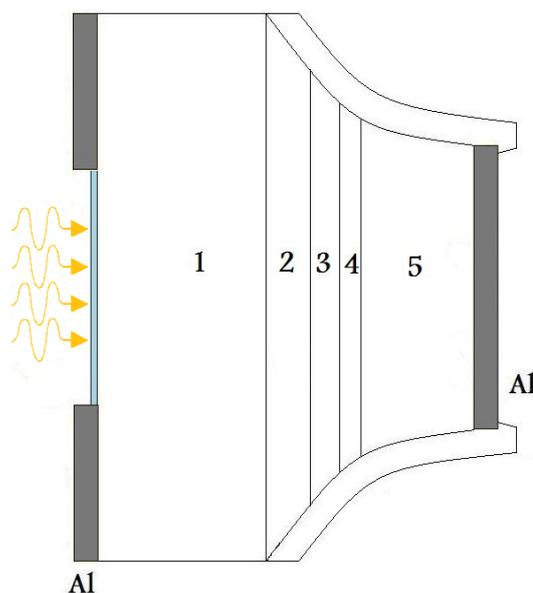


Fig. 120

Schema di un dispositivo largamente usato per ricevere segnali luminosi, da fibra, in seconda e terza finestra, sagomato a forma MESA: il fotodiodo “Mesa – Etched” SAGM – APD. I dati riportati costituiscono delle caratteristiche tipiche.

7.3) Un esempio di eterostruttura graduata a basso fattore di rumore in eccesso

Esaminiamo adesso una struttura che consente di inibire fortemente la capacità delle fotolacune di provocare ionizzazioni per impatto, e di promuovere i fotoelettroni a tale ruolo, abbattendo così il fattore di rumore in eccesso $F(M)$. Il tutto, inoltre, applicando una tensione inversa di polarizzazione piuttosto bassa. Una struttura simile può essere usata come fotorivelatore indipendente, oppure costituire la regione di moltiplicazione di uno più complesso: ad esempio potrebbe essere, nel visibile o nel medio – lontano infrarosso, la zona di guadagno p di un dispositivo $n^+/p/\pi/p^+$. L’idea su cui si basa il funzionamento è l’alternanza fra zone di accelerazione degli elettroni e zone di assorbimento dell’energia cinetica che questi acquistano periodicamente. Riferiamoci alla figura 121.

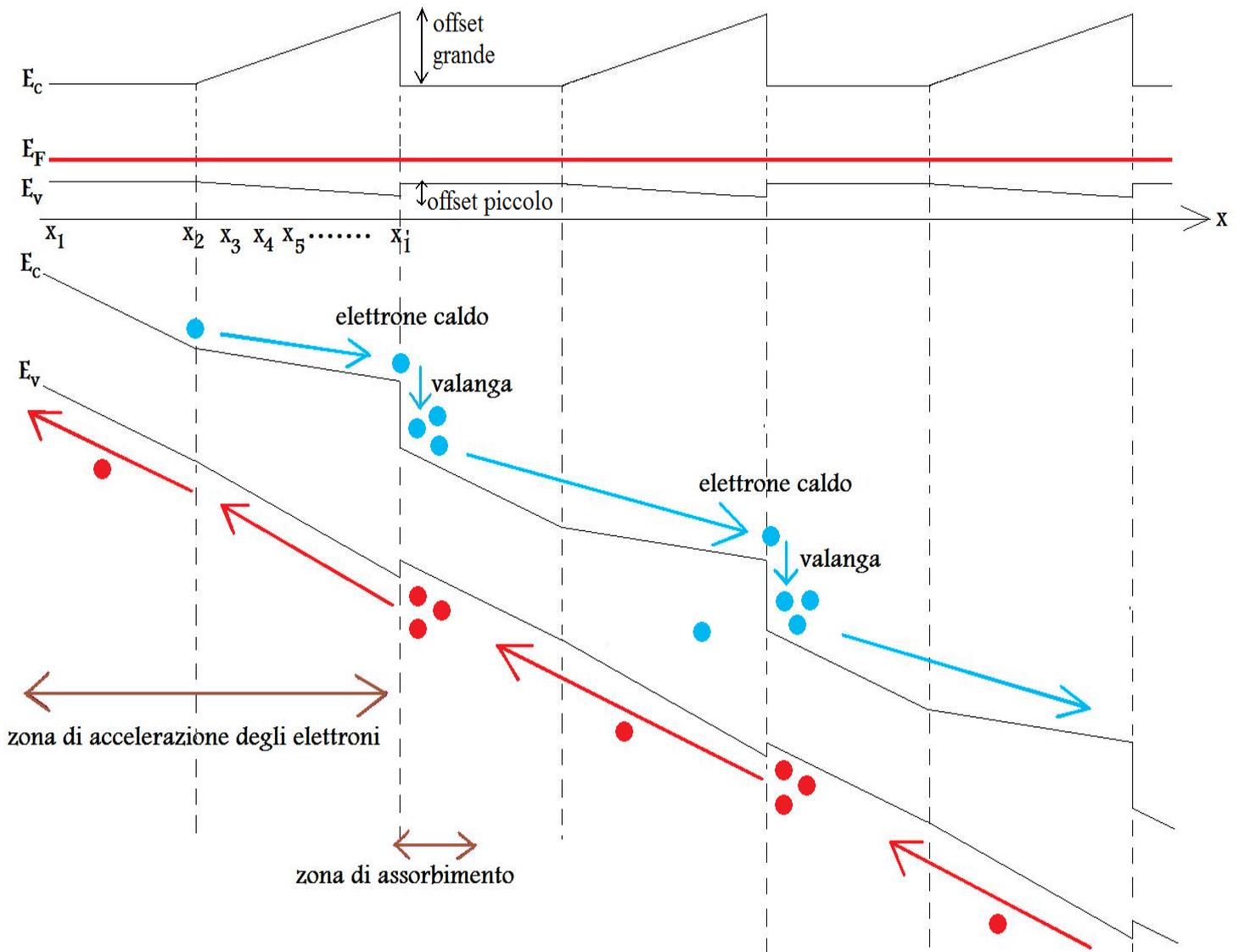


Fig. 121

Diagramma a bande qualitativo di una struttura periodica ad eterogiunzioni, all'equilibrio, in alto, e sottoposta a polarizzazione inversa, in basso, con voltaggio $|V|$ non forte. L'alternanza fra zone di accelerazione degli elettroni, create dalla particolare variazione graduale dell' E_{gap} , e zone di assorbimento dell'energia cinetica, che gli elettroni acquistano periodicamente, è alla base dei processi di breakdown a valanga innescati dai fotoelettroni caldi. Le fotolacune non producono ionizzazioni da impatto a causa della bassa pendenza di E_v : ciò limita la rumorosità del "superreticolo".

Lo strato fra x_1 e x_2 è costituito da $I_{n(0.53)}G_{a(0.47)}A_s$, mentre lo strato fra x_2 e x_1' è formato da un insieme di eterogiunzioni così costituite: fra x_2 e x_3 abbiamo un layer di $I_{n(0.53)}G_{a(0.47)}A_{s(1-x)}P_x$, con x molto piccolo, diciamo tendente a zero, per cui x_2 è quasi un'omogiunzione, fra x_3 e x_4 abbiamo un layer di $I_{n(0.53)}G_{a(0.47)}A_{s(1-x)}P_x$, con una concentrazione di fosforo leggermente maggiore (e quindi di arsenico leggermente minore) di quella che caratterizza il film precedente (x un po' più alto), e così via, fino ad arrivare al film, che termina presso la sezione x_1' , costituito da InP. Durante la crescita dei vari layers da x_2 a x_1' anche le concentrazioni di indio e di gallio diminuiscono, così da garantire una buona compatibilità reticolare attraverso tutto il dispositivo. Questi layers sono cresciuti uno sopra l'altro, a partire dal substrato di $I_{n(0.53)}G_{a(0.47)}A_s$, per mezzo della tecnica CBE, ovvero una particolare MBE, che permette di controllare bene la quantità di materiale crescente. In questo modo è possibile realizzare una transizione graduale da un composto ad un altro in modo da far variare lentamente il gap, evitando la formazione di scalini di energia potenziale per gli elettroni e per le lacune. Dunque si passa lentamente dall' E_{gap} del $I_{n(0.53)}G_{a(0.47)}A_s$ a quello maggiore del InP, con la conseguente formazione di una grande barriera ΔE_c . La distanza fra x_2 e x_1' è di 10 - 20 nm. L'intero dispositivo è una periodizzazione spaziale della configurazione appena descritta. Tutto il semiconduttore della struttura è drogato di tipo p, come si può notare dalla posizione del livello di Fermi nella banda della struttura in assenza di polarizzazione. La piccola barriera ΔE_v è dovuta alla tenue diffusione di lacune fra una generica sezione x_k , compresa fra x_2 e x_1' , e la successiva x_{k+1} , conseguentemente all'allineamento di E_F .

Adesso polarizziamo inversamente questo “superreticolo” con una tensione abbastanza piccola, inferiore a quella con cui abbiamo polarizzato i precedenti SAM – APD, ma tale da rendere negativa la pendenza delle barriere di E_c , ed osserviamo il comportamento dei portatori fotogenerati. Gli elettroni acquistano energia cinetica all’interno di ciascuna delle zone di accelerazione, alla fine delle quali questi risultano molto energetici, “caldi”. La discontinuità di E_c , ovvero il fortissimo campo elettrico interno, localizzato alle interfacce x_1' , conferisce ai fotoelettroni caldi ulteriore energia per innescare una valanga. Questa è formata da fotoelettroni e fotolacune che iniziano a muoversi, come di consuetudine, partendo delle regioni di assorbimento delle energie elettroniche “calde”. La moltiplicazione è unicamente dovuta agli elettroni, incentivata dalla forma di E_c , mentre le lacune primarie e secondarie procedono senza innescare ionizzazioni. Infatti mentre un fotoelettrone trova, presso una qualunque sezione x_1' , una pendenza molto negativa dell’energia potenziale E_c , ovvero un campo elettrico accelerante molto forte, il quale consente agli elettroni di innescare il breakdown a valanga, presso le stesse sezioni x_1' le fotolacune trovano una barriera di potenziale E_v ad alta pendenza, ovvero un campo elettrico decelerante molto forte, il quale rallenta le lacune ed impedisce loro di sfruttare l’energia cinetica, acquistata lungo il layer $In_{(0.53)}Ga_{(0.47)}As$, per innescare ionizzazioni da impatto.

Riportiamo brevemente in appendice A il sistema di funzionamento dell’epitassia da fasci molecolari (MBE - “Molecular Beam Epitaxy”), tecnologia che permette l’integrazione nanometrica del superreticolo appena descritto. Inoltre riportiamo, in appendice B, una possibile tecnica per la rivelazione di radiazioni luminose appartenenti all’intervallo di lunghezze d’onda noto come FIR (“Far Infra Red” - “lontano infrarosso”, luce molto poco energetica), mentre, in appendice C, riportiamo una possibile scelta di semiconduttori III - V per la rivelazione di radiazioni ottiche più energetiche di quelle precedenti, ovvero rientranti nell’intervallo spettrale compreso, approssimativamente, tra la luce visibile e quella ultravioletta.

8) I fotomoltiplicatori al silicio (SiPMs - “Silicon Photon Multipliers”)

8.1) La modalità di funzionamento Geiger e la disposizione matriciale dei SAM - APDs $n^+/p/\pi/p^+$ al silicio

Fin qui abbiamo ipotizzato che gli APD funzionino in “modalità proporzionale”, ovvero che il fotosegnale di corrente I_L , raccolto sulla resistenza esterna R_L , sia direttamente proporzionale al numero di fotoni assorbiti dal dispositivo, in modo tale che l’ampiezza del segnale elettrico sia proporzionale all’intensità ottica incidente I_o . Un funzionamento simile si verifica per tensioni inverse di polarizzazione $|V|$ comprese fra V_{BD} ed un certo valore di soglia V^* (si osservino le figure 83 e 122). In questo intervallo di tensione il guadagno M è abbastanza piccolo e garantisce una stretta correlazione fra entità dell’evento luminoso (sia in intensità che in frequenza) ed entità della risposta elettrica. Tale relazione si perde per tensioni $|V|$ superiori a V^* : sottoposto ad una polarizzazione inversa così elevata, l’APD acquista un guadagno talmente alto che la corrente di uscita non è più riferibile ad un evento luminoso piuttosto che ad un altro. I_L è la stessa in risposta ad un qualunque numero di fotoni assorbiti, e ciò costituisce il principale problema del funzionamento appena descritto, noto come “modalità Geiger non proporzionale”.

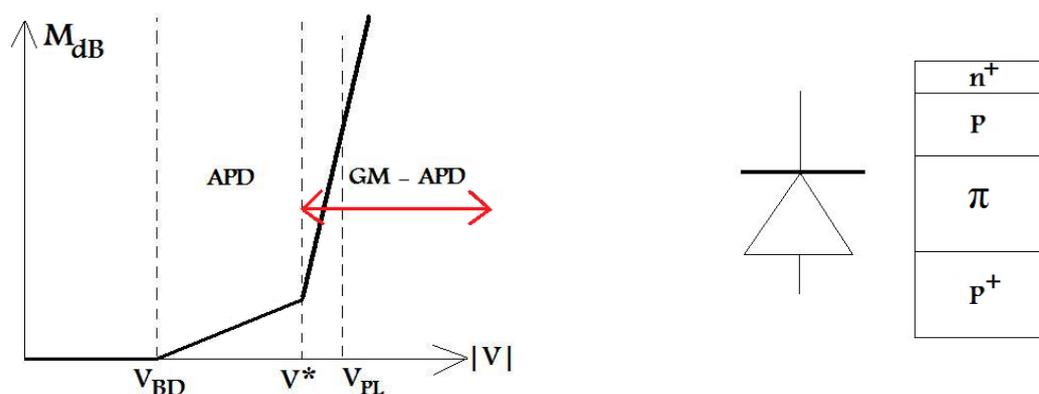


Fig. 122

A sinistra riportiamo il grafico logaritmico del fattore di moltiplicazione della densità di corrente M , in funzione del voltaggio inverso $|V|$ applicato ai terminali del fotodiodo SAM - APD di silicio ($n^+/p/\pi/p^+$), riportato a destra. Si notino, nel grafico di M , la regione di funzionamento proporzionale (APD) e quella di funzionamento non proporzionale (GM-APD).

La soluzione consiste nell'usare non più un singolo SAM - APD (prendiamo la struttura in $\text{Si } n^+/\text{p}/\pi/\text{p}^+$), bensì una matrice costituita da tanti fotorivelatori SAM - APD, chiamati microcelle o pixel, messi in parallelo fra loro e opportunamente connessi. Non appena si verifica un evento luminoso, i suoi fotoni si distribuiscono sulla superficie della matrice di fotodiodi, i quali sono fortemente polarizzati in inversa e tutti funzionanti in modalità Geiger. Ciascuna di queste microcelle, se colpita da uno o più fotoni, produce una valanga, e quindi un segnale elettrico $I_{\text{pix}}(t)$. Tale segnale è più simile ad un impulso digitale che non ad uno analogico (la microcella si comporta come un counter 1 - 0), dal momento che l'informazione portata dall'eventuale $I_{\text{pix}}(t)$ non possiede livelli intermedi. Se questa corrente non scorre, in uscita dalla microcella, allora il pixel non ha ricevuto alcun fotone (nessuna valanga innescata), mentre se c'è corrente (un impulso di corrente) allora il pixel ha rivelato della luce. Se la superficie della matrice è sufficientemente ampia, cioè contiene un elevato numero di microcelle, e l'intensità luminosa incidente sufficientemente bassa (pochi fotoni), allora è molto improbabile che due fotoni colpiscano la medesima cella, per cui è ragionevole associare a ciascun segnale $I_{\text{pix}}(t)$ l'arrivo di un fotone. Conoscendo l'entità di questa corrente e misurando il segnale totale sul circuito esterno, collegato alla matrice, è possibile risalire al numero di fotoni che hanno colpito l'intera matrice di fotodiodi. La struttura bidimensionale, così funzionante, prende il nome di SiPM. Ne esistono vari modelli, contraddistinti da parametri, dimensioni e applicazioni che adesso descriveremo. In figura 123 riportiamo la foto di un SiPM appartenente ad una delle tipologie più diffuse, interessanti anche per un possibile utilizzo in ambito PET.

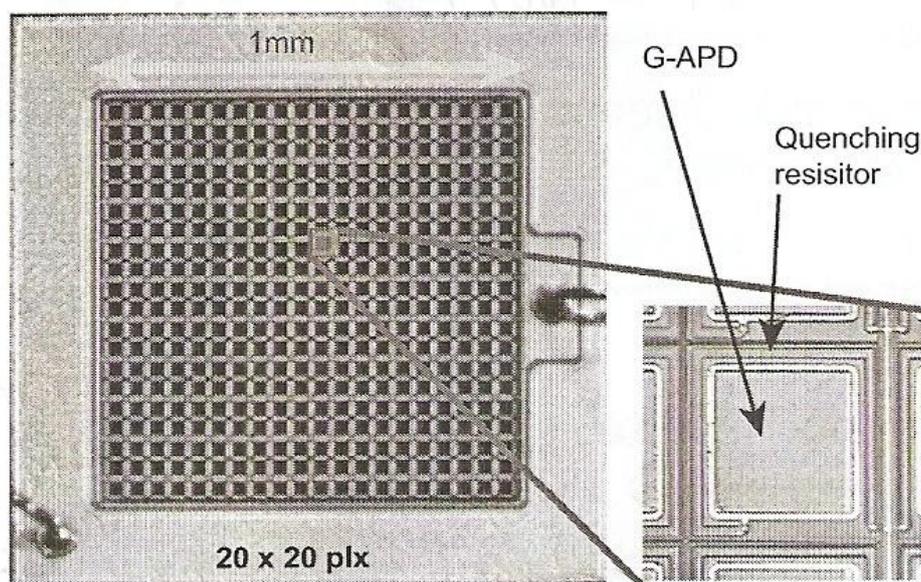


Fig. 123

Foto di un SiPM di superficie totale pari a 1mm^2 , costituito da 400 microcelle (400 GM - APDs, in file di 20), ciascuna di superficie totale pari a $50 \times 50 \mu\text{m}^2$. Nell'ingrandimento del singolo pixel sono ben visibili la finestra dielettrica antiriflettente (SiO_2), la "resistenza di quenching" (R_0) e la griglia metallica per la raccolta del fotosegnale totale.

8.2) Il range dinamico e l'efficienza di rivelazione dei fotoni per un SiPM

La matrice quadrata di figura 123 (1mm^2 di superficie totale, una delle estensioni più diffuse) è costituita da 400 microcelle (400 GM - APDs, in file di 20), ciascuna di superficie totale pari a $50 \times 50 \mu\text{m}^2$. Un altro formato molto comune è la matrice, sempre di 1mm^2 , formata da 625 pixel, in file di 25, ciascuno di superficie totale $40 \times 40 \mu\text{m}^2$. In relazione a tali geometrie possiamo definire uno dei parametri più importanti di un SiPM, noto come "range dinamico": questo è il massimo numero di fotoni incidenti che, teoricamente, la matrice può rivelare in conseguenza di un singolo evento luminoso e corrisponde, almeno idealmente, al numero di pixel che la costituiscono, ovvero al massimo numero di microcelle coinvolgibili dalla suddetta radiazione luminosa. Nel caso del SiPM riportato in figura 123 il range dinamico è quindi pari a 400. Discutiamo adesso la definizione appena data.

Ipotizziamo che sulla matrice arrivi un singolo fotone, la cui frequenza è compatibile con la responsività del Si. Il fotone colpirà una delle 400 microcelle e verrà assorbito dalla regione π del pixel illuminato, innescando una valanga nella sua regione p di moltiplicazione. Dal pixel colpito uscirà una corrente $I_{\text{pix}}(t)$, che verrà raccolta su una resistenza esterna R_L ; la tensione misurata su questa rivelerà l'arrivo di una radiazione luminosa formata da

un solo fotone. Se in un esperimento successivo inviamo sul SiPM una luce costituita da un maggior numero di fotoni incidenti N_ϕ (ma pur sempre molto pochi rispetto al numero totale di pixel: $N_\phi \ll 400 = N_{\text{pixeltotali}}$), questi si sparpagliano sull'intera matrice: statisticamente è ragionevole pensare che ciascuno di questi N_ϕ fotoni colpisca una microcella diversa. Dunque avremo, idealmente, N_ϕ pixel colpiti, ovvero N_ϕ correnti di valanga $I_{\text{pix}}(t)$, ciascuna uscente dal pixel illuminato: queste correnti si sommano, dato che i 400 pixel sono in parallelo fra loro, e la corrente totale $I_L(t)$ porterà, sulla sua ampiezza impulsiva, l'informazione legata alla debole intensità ottica appena arrivata sulla matrice. Dunque finché $N_\phi \ll 400$, corrispondente ad un flusso fotonico $\ll 400$ eventi/mm², abbiamo una relazione pressochè lineare fra il numero di celle colpite ed il numero di fotoni incidenti, per cui l'entità del segnale elettrico d'uscita è correttamente legata, in modo lineare, all'intensità ottica incidente. Tuttavia, se N_ϕ comincia ad essere grande, la probabilità che due fotoni siano assorbiti dallo stesso pixel non è più così bassa, per cui è probabile che il segnale elettrico sottostimi il numero di fotoni arrivati (due o più fotoni che incidono sullo stesso pixel sono interpretati, a causa del funzionamento Geiger, come un singolo fotone). Tale sottostima aumenta all'aumentare di N_ϕ : la funzione che rappresenta il numero di celle colpite $N_{\text{pixelcolpiti}}$, al variare di N_ϕ , è lineare per piccoli valori del numero di fotoni incidenti, e satura per valori di N_ϕ via via crescenti, cioè all'aumentare del numero medio di fotoni per singola microcella. Per intensità luminose molto alte $N_{\text{pixelcolpiti}} \rightarrow 400$, ma è chiaro che l'informazione elettrica non ha più alcun legame con l'effettiva potenza ottica incidente.

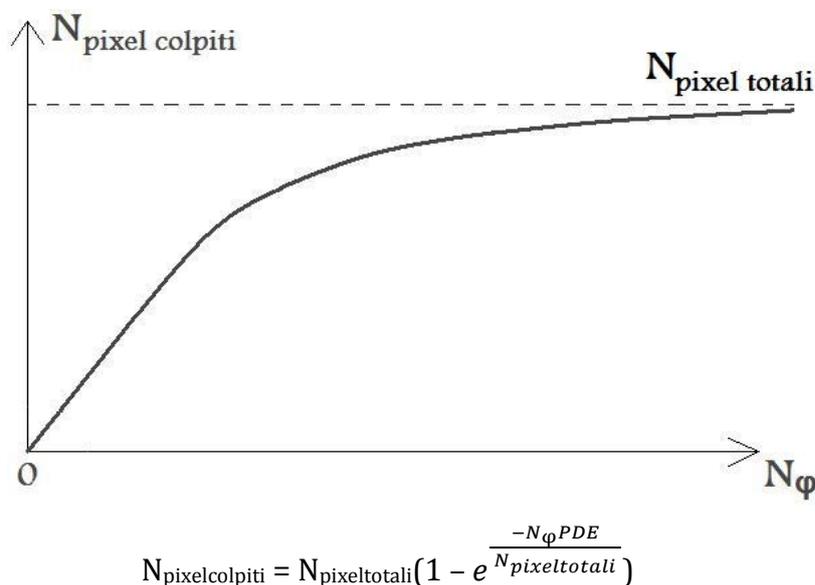


Fig. 124

Grafico del numero di microcelle colpite dalla luce ($N_{\text{pixelcolpiti}}$) in funzione del numero di fotoni incidenti (N_ϕ) su un SiPM, il quale è formato da un numero complessivo di pixel pari a $N_{\text{pixeltotali}}$. L'andamento è praticamente lineare per eventi luminosi costituiti da pochi fotoni, mentre satura per flussi fotonici elevati. Chiaramente solo nel primo caso il segnale elettrico, raccolto sulla griglia, è proporzionale al segnale luminoso incidente.

L'efficienza di rivelazione dei fotoni (PDE - "Photon Detection Efficiency") è un altro parametro fondamentale di un SiPM, ed è così definita:

$$\text{PDE} = \frac{\text{numero dei fotoni rilevati}}{\text{numero dei fotoni incidenti}} = \frac{\text{numero dei pixel attivati}}{N_\phi} = \text{QE} G_f P_t$$

dove QE è la già definita efficienza quantica del singolo SAM - APD che costituisce il pixel, P_t la sua probabilità di trigger, G_f l'efficienza geometrica (fattore geometrico) di tutto il SiPM, da non confondere con quella della singola microcella ϵ_G . Un valore tipico per G_f è $\approx 44\%$; in alcune misure è stato ottenuto un valore per la PDE intorno al 50% per una radiazione incidente verde - gialla, intorno al 30% per una radiazione blu, nonché un incremento lineare rispetto all'overvoltage $\Delta V = |V| - V_{\text{BD}}$.

In figura 125 riportiamo la dipendenza della PDE dalla lunghezza d'onda della radiazione luminosa incidente, per tre particolari SiPMs (i SiPMs sono noti anche come MPPCs - "Multi - Pixel Photon Counters" - contatori di fotoni

multipixelati) prodotti dalla Hamamatsu, azienda giapponese leader mondiale nel settore della componentistica optoelettronica, dei quali riportiamo le caratteristiche geometriche di base:

- modello "S10362 - 11 - 100U", di area totale 1.5x1.5mm², $G_f \approx 78.5$, 100 pixel, ciascuno di area 100x100μm²
- modello "S10362 - 11 - 050U", di area totale 1.5x1.5mm², $G_f \approx 61.5$, 400 pixel, ciascuno di area 50x50μm²
- modello "S10362 - 11 - 025U", di area totale 1.5x1.5mm², $G_f \approx 30.8$, 1600 pixel, ciascuno di area 25x25μm²

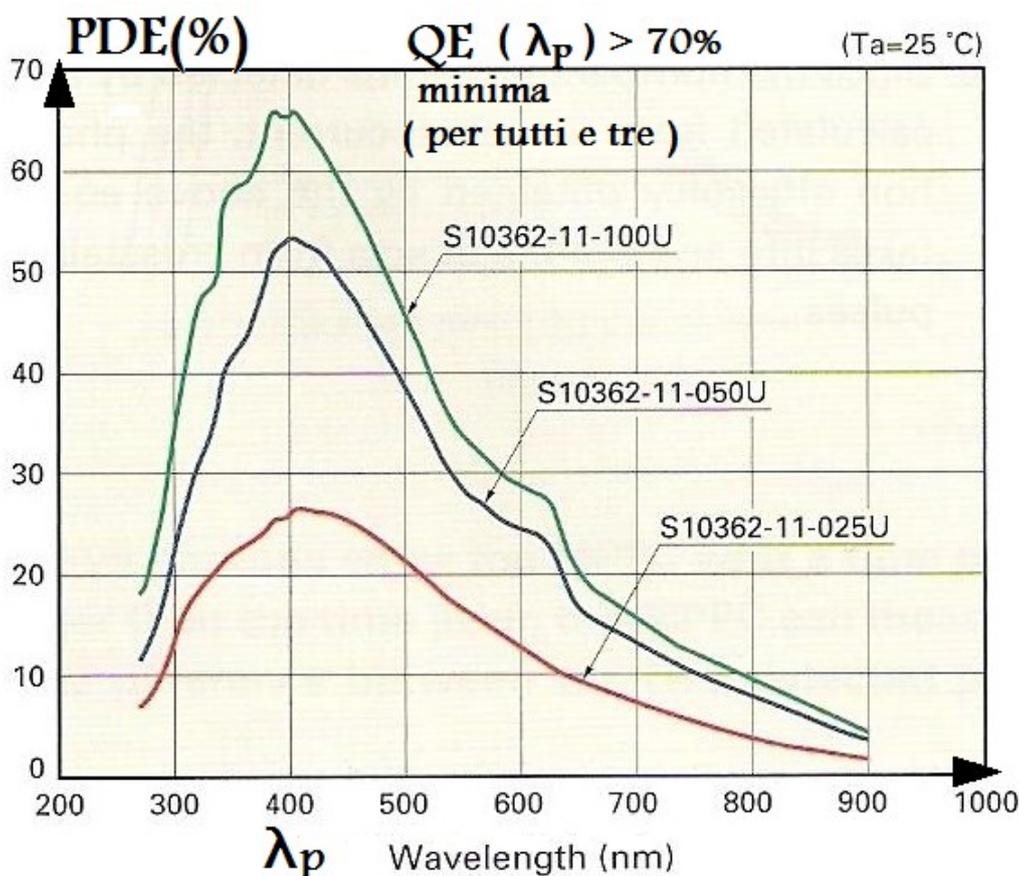


Fig. 125

Grafico, ottenuto da misurazioni in laboratorio, della PDE in funzione della lunghezza d'onda della radiazione luminosa incidente, per tre particolari SiPMs prodotti dalla Hamamatsu.

Si noti come la lunghezza d'onda di picco λ_p , per i tre SiPMs, sia intorno ai 400 nm (luce blu): questa caratteristica suggerisce l'accoppiamento, con ciascuno dei tre fotorivelatori, di un cristallo scintillatore LYSO, di cui parleremo brevemente nel paragrafo 8.9. Si noti, altresì, come all'aumentare della concentrazione di pixel, sulla superficie totale del SiPM, corrisponde una diminuzione di G_f , e conseguentemente una diminuzione della PDE presso qualunque λ , compresa λ_p . La spiegazione della proporzionalità inversa fra concentrazione di pixel (numero totale di pixel, ossia range dinamico) e fattore geometrico la rimandiamo al paragrafo 8.8.

Il fotosegnale totale, in uscita da un SiPM, soddisfa la seguente relazione:

$$I_L(t) \propto q G N_{\text{pixelcolpiti}}$$

dove G è il guadagno di microcella, intuitivamente associabile ad M , ma a cui daremo, a breve, una forma più rigorosa.

8.3.1) Il resistore di quenching: la sua disposizione sull'ossido di microcella e due possibili tecniche di integrazione

In figura 126 riportiamo la sezione di una singola microcella ed il relativo layout:

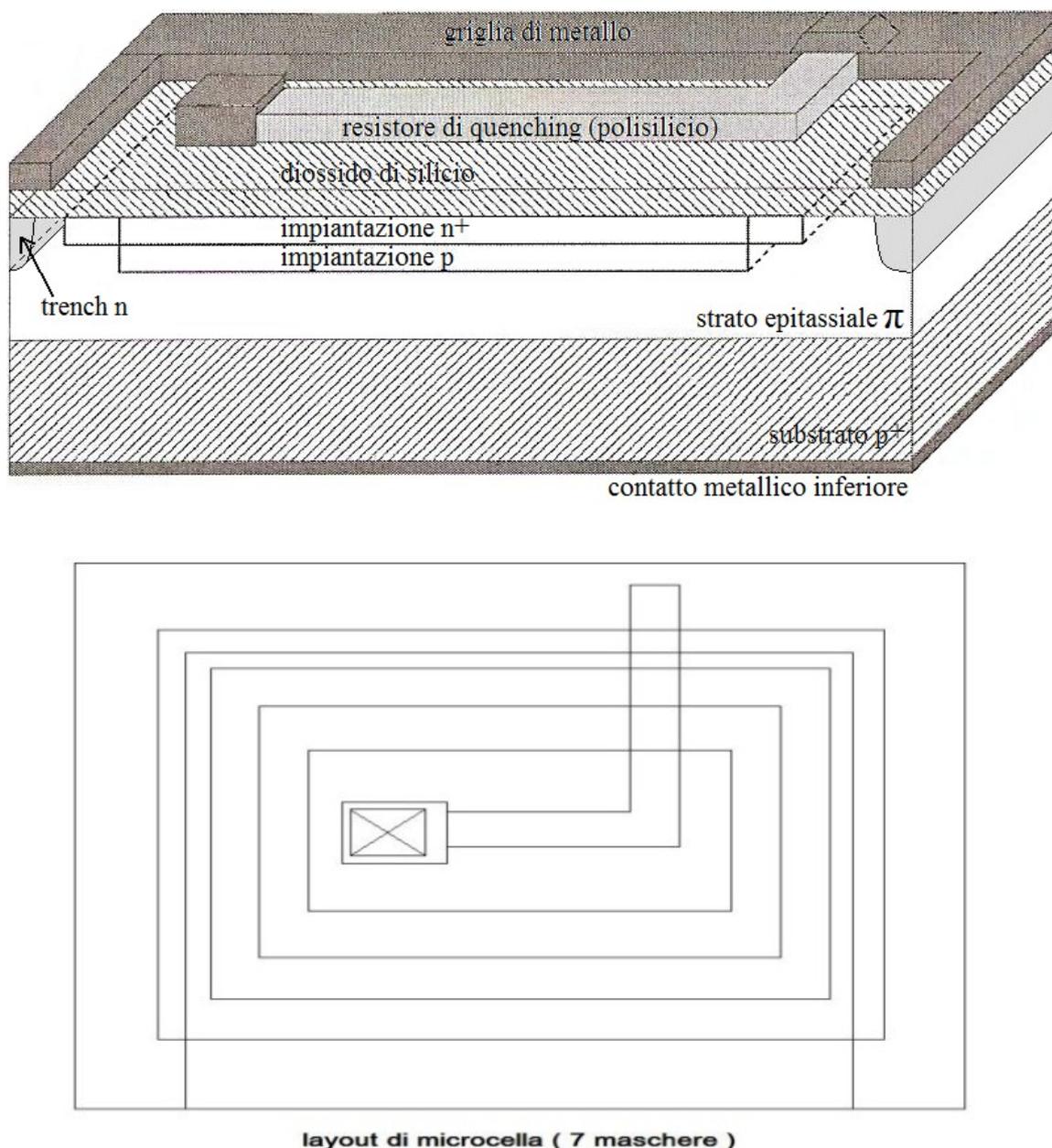


Fig. 126

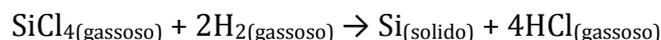
In alto riportiamo la sezione di una microcella, costituita da un fotodiodo al silicio SAM - APD $n^+/p/\pi/p^+$, polarizzato in modalità Geiger non proporzionale. Tale struttura è collegata alla griglia metallica, comune all'intera matrice del SiPM, mediante un "resistore di quenching". Lo strato antiriflesso, che funge anche da layer di passivazione, è costruito con ossido di silicio SiO_2 . In basso è mostrato il layout.

È ben visibile la struttura $n^+/p/\pi/p^+$, dove il substrato p^+ (contattato inferiormente, cioè con il contatto elettrico, di alluminio, nella parte inferiore) ed il layer di assorbimento π sono entrambi depositabili per epitassia da fase vapore VPE. È possibile notare la trench, spesso realizzata con Si di tipo n, la quale migliora l'isolamento locale e la collimazione del fenomeno della valanga. Si noti la finestra aperta nell'ossido per assicurare, alla struttura attiva sottostante, il contatto di A_1 . Quest'ultimo non si sviluppa in una pista terminante sulla griglia metallica: il percorso è mediato da un segmento di materiale altamente resistivo, chiamato in gergo "resistenza di quenching" R_Q ($\approx 300 - 350 \text{ k}\Omega$), ossia la resistenza di "estinzione" della valanga, del cui funzionamento ci occuperemo nel

paragrafo 8.3.2. La lunghezza di questo resistore superficiale è determinata, più che dall'effettiva distanza dalla griglia, dal numero di quadri necessari per far vedere, al fotodiodo sottostante, la resistenza R_Q desiderata.

R_Q è integrabile mediante uno di questi due processi:

- un'epitassia VPE di Si, di cui riportiamo una delle reazioni più utilizzate:



Affinchè il Si, così depositato sull'ossido, cresca in modo quasi amorfo, ovvero costituito da grani aventi ciascuno un proprio orientamento (è tale disorganizzazione cristallografica a conferire alta resistività alla R_Q), è importante che la frazione molare del tetracloruro di silicio, che è il precursore del Si, rispetto all'idrogeno in cui è ridotto, il cui "flow rate", per $t \approx 1270$ °C, è di circa 1 litro/min, sia compresa fra il 2.5 ed il 20%. In questo modo la velocità di crescita del film del resistore è compresa fra 2 e 5 $\mu\text{m}/\text{min}$, sufficientemente elevata per impedire qualunque estesa organizzazione cristallografica del Si, il quale cresce, così, in modalità "poly". Il forno da usare è simile a quello di un normale sistema di deposizione CVD o di crescita VPE. Due possibili reattori utilizzabili a tale scopo sono qualitativamente mostrati in figura 127. Il problema di tale metodo è l'alta temperatura alla quale tenere il substrato (benchè per breve tempo), ossia 1250 - 1275 °C, la quale è sopportabile sia per il Si, che fonde a 1421 °C, sia per l'ossido di silicio, che fonde a 1710 °C, ma non per l'alluminio, la cui temperatura di fusione è 660 °C. Dunque, in caso di epitassia VPE, è necessario usare un siliciuro di metallo refrattario per il contatto fra n^+ e R_Q ; la griglia si fa normalmente di Al , in quanto la sua definizione è frutto di un passo fotolitografico successivo. Altri precursori gassosi del silicio, generalmente immessi nel reattore a fini di crescita epitassiale, sono: il diclorosilano SiH_2Cl_2 , il triclorosilano SiHCl_3 , il silano SiH_4 .

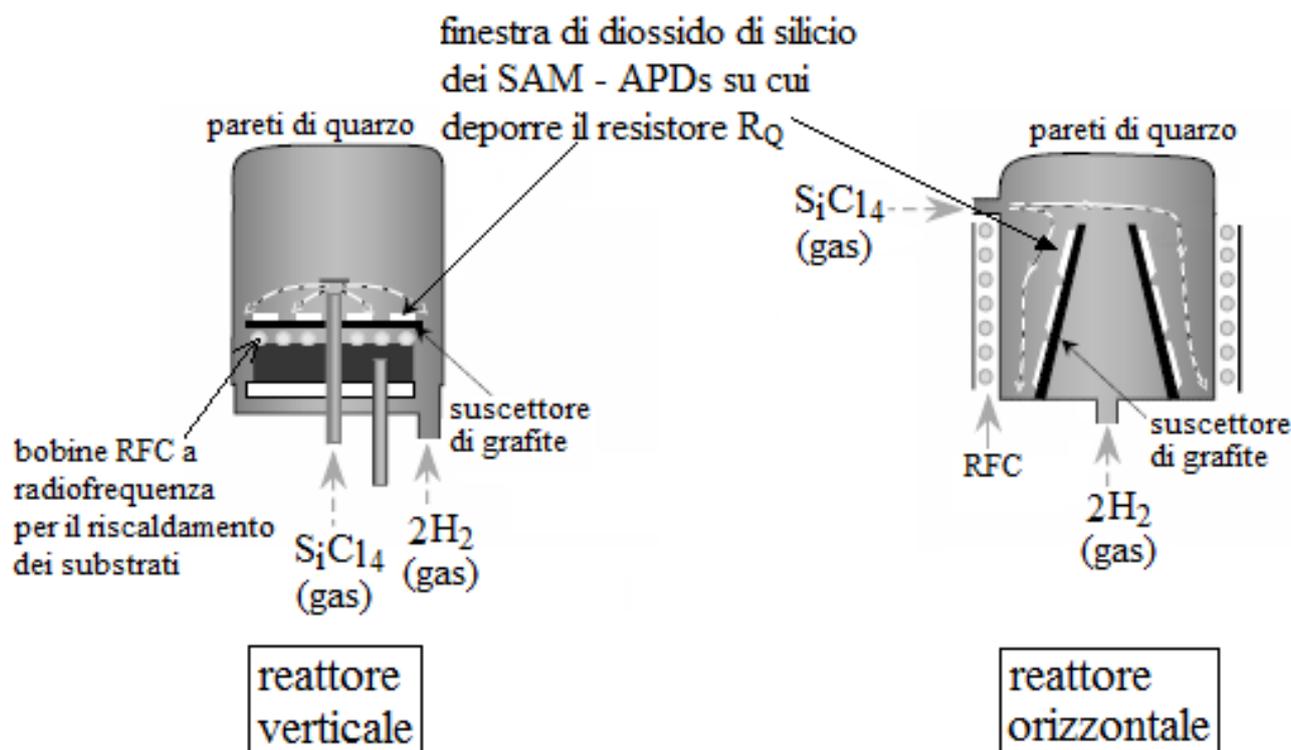


Fig. 127

Rappresentazione qualitativa di due possibili reattori, uno "verticale", l'altro "orizzontale", utilizzabili per la crescita epitassiale da fase vapore VPE di un film di silicio policristallino, il quale può costituire il resistore di quenching R_Q , da integrare sulla finestra antiriflettente e di passivazione, di diossido di silicio, delle microcelle SAM - APDs $n^+/p/\pi/p^+$ di un SiPM. La nomenclatura "verticale" ed "orizzontale" si riferisce alla direzione del flusso del gas di tetracloruro di silicio, diffuso nella camera di lavoro.

- Una deposizione CVD di polisilicio, che in processi LPCVD deve avvenire a temperature comprese fra 575 e 650 °C:



Sotto i 580 °C il film di R_Q cresce totalmente amorfo, cioè particolarmente resistivo, mentre al di sopra di 580 °C R_Q consta di grani aventi orientazioni che dipendono criticamente dalla temperatura.

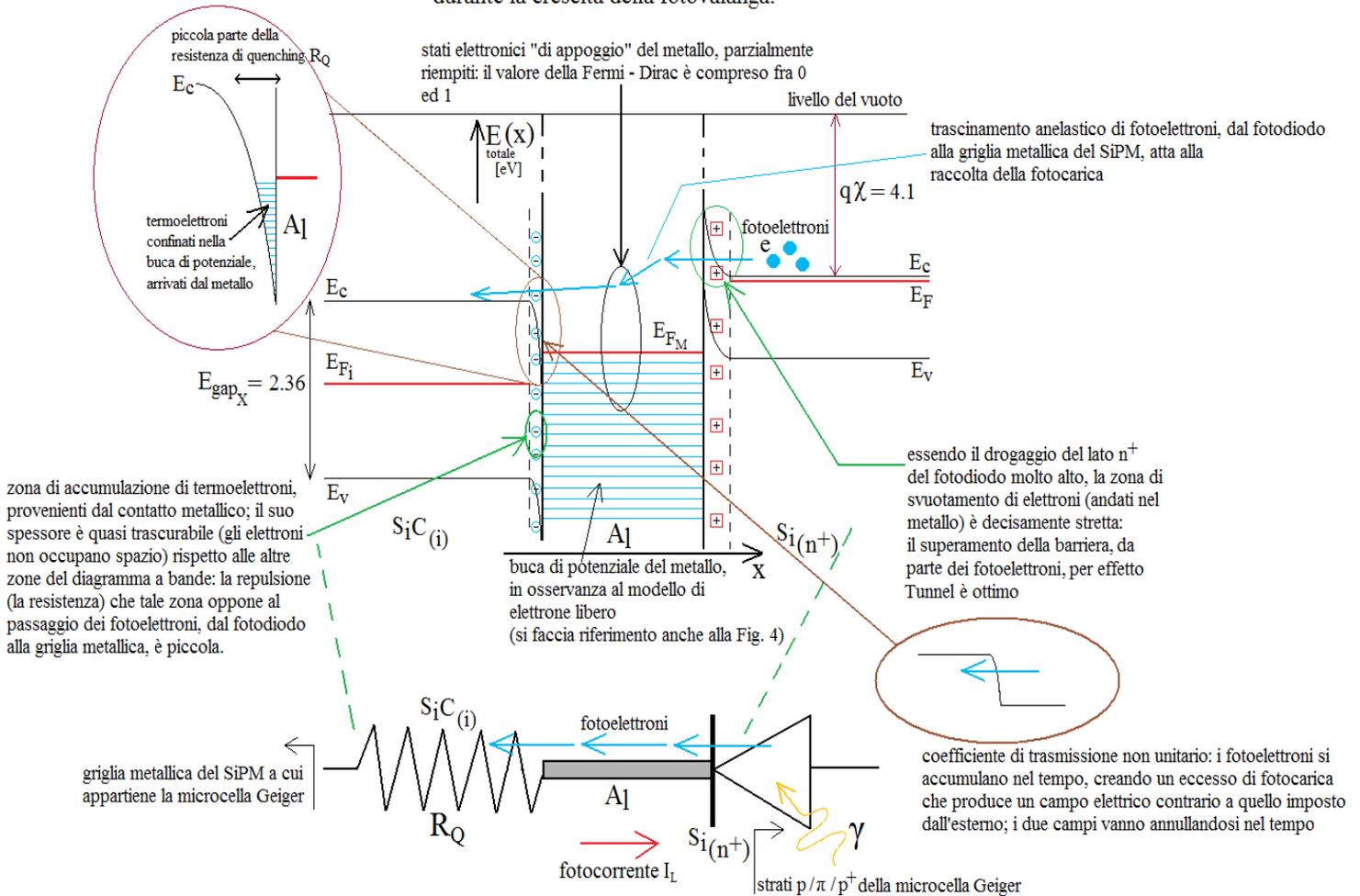
Il substrato p^+ , che è in comune fra tutti i pixel, con il relativo contatto metallico inferiore e la griglia di A_1 hanno il compito di mettere in parallelo tutte le microcelle, al fine di polarizzarle con la medesima pila e di raccogliere, superficialmente, le fotocariche collezionate. Potrebbe essere opportuno collegare, fra griglia e pila, un regolatore di tensione (serie, switching, ecc...), essendo la tensione da questa erogata, in genere, poco stabile, e considerando che il guadagno di pixel è fortemente legato al suo valore. La griglia è contattata col circuito esterno di polarizzazione, che oltre al generatore $|V|$ consta anche del resistore R_L , su cui leggiamo la fototensione V_L . In alcuni dispositivi è possibile trovare R_Q realizzata con una deposizione di carburo di silicio SiC, semiconduttore la cui struttura geometrica è di tipo zincoblenda, avente un gap indiretto, fra BV e BC, pari a circa $3 \div 3.5$ eV (grande gap). Quindi la concentrazione intrinseca n_i di elettroni di conduzione è molto bassa, dell'ordine di $10^6 \div 10^7$ cm⁻³, e ciò comporta che il SiC è un materiale altamente resistivo, dunque adatto alla realizzazione della R_Q (il SiC intrinseco è definito anche un "semisolante").

8.3.2) Il resistore di quenching: il funzionamento

R_Q ha la fondamentale funzione di riportare la microcella Geiger, dopo che questa ha rivelato l'arrivo di un fotone, nella condizione di poterne rivelare successivamente un altro. Infatti l'assorbimento di un fotone innesca un processo di valanga teoricamente inarrestabile nel tempo, paragonabile ad una sorta di feedback positivo (in questo caso, utilizzando il gergo e i modelli tipici della teoria della reazione, si parla di $\beta A > 1$), e ciò contrasta con l'esigenza di utilizzare il pixel per rivelazioni successive. È quindi necessario integrare un elemento che, dopo un certo intervallo di tempo, durante il quale la valanga si innesca, si sviluppa e raggiunge il suo picco, la inibisca e riporti il fotodiodo nella condizione di preassorbimento (una sorta di feedback negativo, ossia una reazione che smorza la "perturbazione" prima autoalimentata, per cui $\beta A < 1$). Ciò avviene perché gli elettroni costituenti la valanga, durante il passaggio attraverso R_Q , vengono rallentati da una densità di stati $D_Q(E)$, localizzati, principalmente, sui bordi dei grani, dove è presente la maggiore concentrazione di vacanze e di difetti reticolari. I grani si comportano da "colla" per gli elettroni diretti verso il circuito esterno, la temporanea accumulazione dei quali provoca una distribuzione di campo elettrico che attenua quella, dovuta ai potenziali di contatto e alla pila, che consente le impattazioni ioniche. In altre parole il campo elettrico prodotto dalla momentanea accumulazione di fotoelettroni attenua e alla fine annulla il campo elettrico legato alla polarizzazione inversa della microcella Geiger. Il campo risultante alla fine del tempo di accrescimento della valanga, di fatto momentaneamente nullo, non incentiva più la prosecuzione della valanga e la corrente $I_{pix}(t)$, dopo un repentino tempo di salita τ_{rise} , inizia a decrescere, con un assai più lungo tempo di discesa τ_{fall} : alla fine il fotosegnale $I_{pix}(t)$ si è estinto completamente e la microcella è pronta per un nuovo evento luminoso.

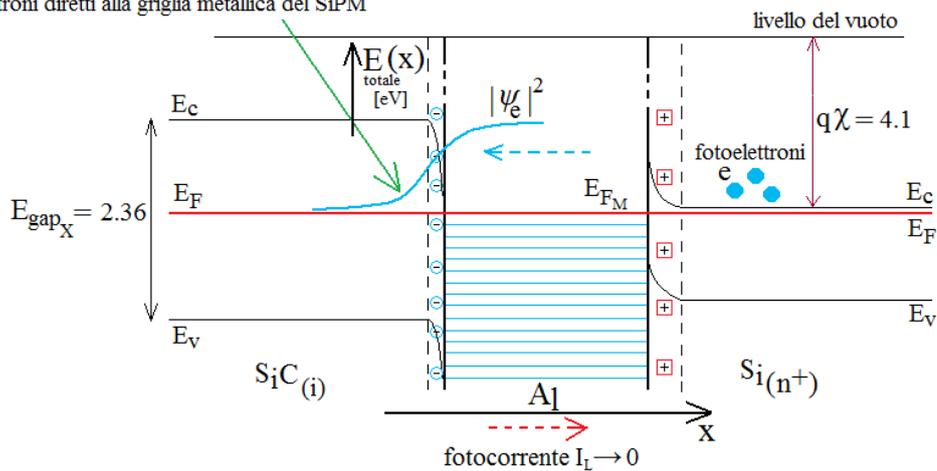
Il momentaneo eccesso di fotoelettroni, che comporta l'attenuazione del campo elettrico accelerante per i fotoelettroni ed innescante per la valanga, può essere interpretato sia come una conseguenza dell'effetto "colla" della densità di stati $D_Q(E)$ di bordo di grani, sia come una conseguenza della discontinuità, fra il silicio drogato n^+ della microcella ed il materiale della resistenza R_Q , dell'energia potenziale E_C degli elettroni in BC. Infatti il gap del silicio monocristallino, con cui è realizzato il pixel SAM – APD, è di circa 1.08 eV, mentre il gap del silicio policristallino, con cui è realizzabile la resistenza R_Q , è di circa 1.5 eV, ed il gap del carburo di silicio SiC cristallizzato secondo il reticolo di tipo zincoblenda (si veda il glossario), che è un altro materiale con cui è possibile realizzare la R_Q , in alternativa al polisilicio, è di circa 2.36 eV. Pertanto i fotoelettroni diretti verso il circuito esterno di lettura, nell'attraversare R_Q , vedono una barriera di potenziale intuitivamente semplificabile con una funzione spaziale di Heavyside, dove il gradino (la barriera) è localizzato lungo il resistore R_Q , come si può notare osservando la figura 128. Questa barriera di potenziale contribuisce, durante la fase di crescita della fotovalanga, a creare l'accumulo di fotoelettroni, lungo R_Q e all'interfaccia R_Q /contatto metallico, a causa della non unitarietà del coefficiente di trasmissione dei fotoelettroni legato al gradino; ciò consente l'abbassamento del modulo del campo elettrico risultante e quindi l'estinzione della fotovalanga stessa.

durante la crescita della fotovalanga:



durante l'estinzione della fotovalanga:

coda evanescente della densità di probabilità dei fotoelettroni diretti alla griglia metallica del SiPM



L'eccesso di fotoelettroni accumulatosi presso l'interfaccia resistore di quenching / metallo produce un campo elettrico che annulla quello legato alla polarizzazione inversa del fotodiode; la momentanea situazione di campo risultante nullo comporta l'allineamento del livello di Fermi e l'incremento della barriera di potenziale vista dai fotoelettroni verso il resistore di quenching. Ciò provoca l'annullamento della fotocorrente I_L e la possibilità per la microcella Geiger di ricevere un nuovo segnale ottico.

Fig. 128

Schema concettuale di una possibile interpretazione del fenomeno dell'estinzione della fotovalanga, in una singola microcella Geiger di un SiPM: l'estinzione è resa possibile dal resistore di quenching R_Q , realizzabile in polisilicio oppure, come in questo caso, in carburo di silicio SiC.

Il fotosegnale $V_L(t)$, prodotto da una microcella, possiede un picco V_{MAX} proporzionale all'overvoltage ΔV , come mostrato in figura 129, a sinistra.

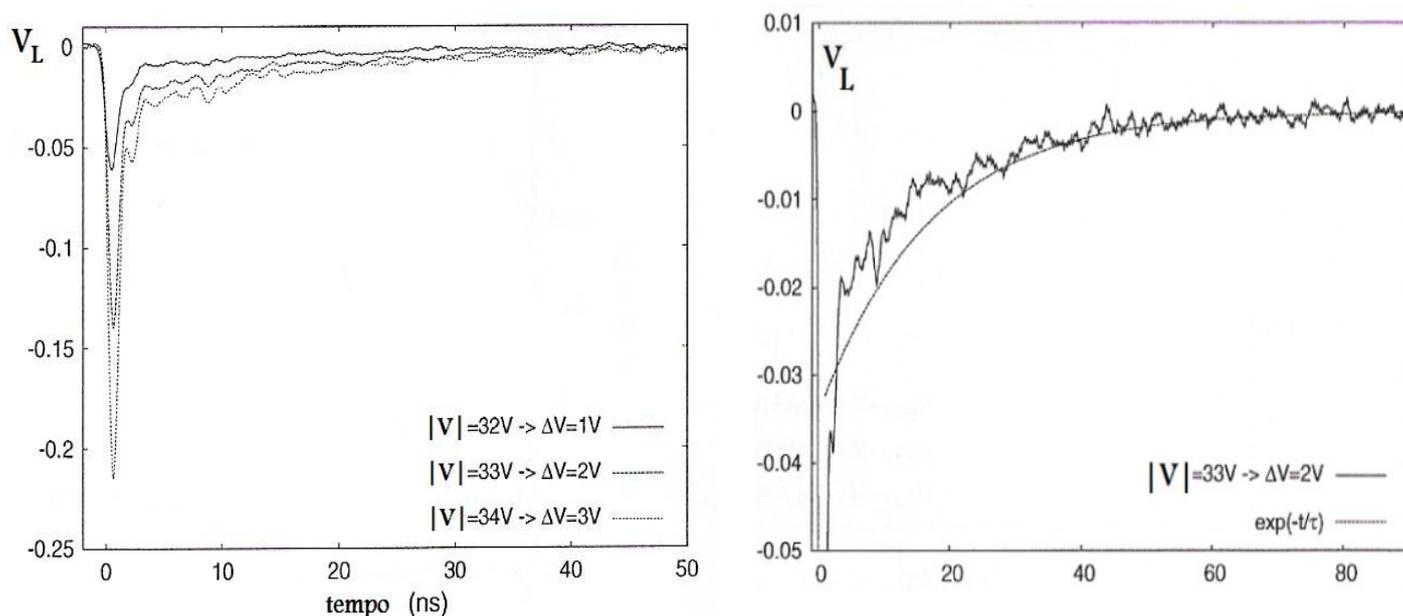


Fig. 129

A sinistra riportiamo il fotosegnale $V_L(t)$, misurato con un oscilloscopio, prodotto da una microcella GM - APD interessata da un evento luminoso. La salita ("leading edge") di $V_L(t)$ corrisponde alla fase di innesco ed autoaccrescimento (feedback positivo) della fotovalanga (fase definita come "carica Geiger"), la discesa ("trailing edge") corrisponde all'estinzione (feedback negativo) della valanga (fase definita come "scarica Geiger"): quest'ultima è resa possibile dal resistore di quenching R_Q . $V_L(t)$ possiede un picco V_{MAX} proporzionale all'overvoltage $\Delta V = |V| - V_{BD}$. Sul display dell'oscilloscopio osserviamo tre fotosegnali, uno per ciascun valore di ΔV . A destra riportiamo un ingrandimento della discesa di uno dei tre fotosegnali: si noti la somiglianza fra l'andamento reale e quello esponenziale, il quale può fungere da modello semplificato.

Qualitativamente è la stessa forma della fototensione prodotta da più microcelle, attivate da N_ϕ fotoni: in tal caso il picco sarà proporzionale a $V_{MAX} N_{pixelcolpiti}$. Ingrandendo uno dei tre andamenti, durante l'estinzione della valanga, ci accorgiamo che il fotosegnale reale è ben approssimabile con una funzione esponenziale. Una semplificazione analoga è applicabile anche alla fase di crescita della valanga, utilizzando la relativa costante di tempo.

8.4) Il circuito equivalente di un SiPM sottoposto ad un impulso luminoso

È possibile modellizzare il comportamento di una o più microcelle del SiPM, durante un evento luminoso, per mezzo del circuito riportato in figura 130. A sinistra abbiamo l'equivalente circuitale, molto semplificato, dell'intero SiPM (nell'esempio un pixel riceve un fotone), mentre a destra abbiamo isolato la parte della matrice coinvolta nell'evento luminoso ed implementato un circuito equivalente per il SAM - APD $n^+/p/\pi/p^+$.

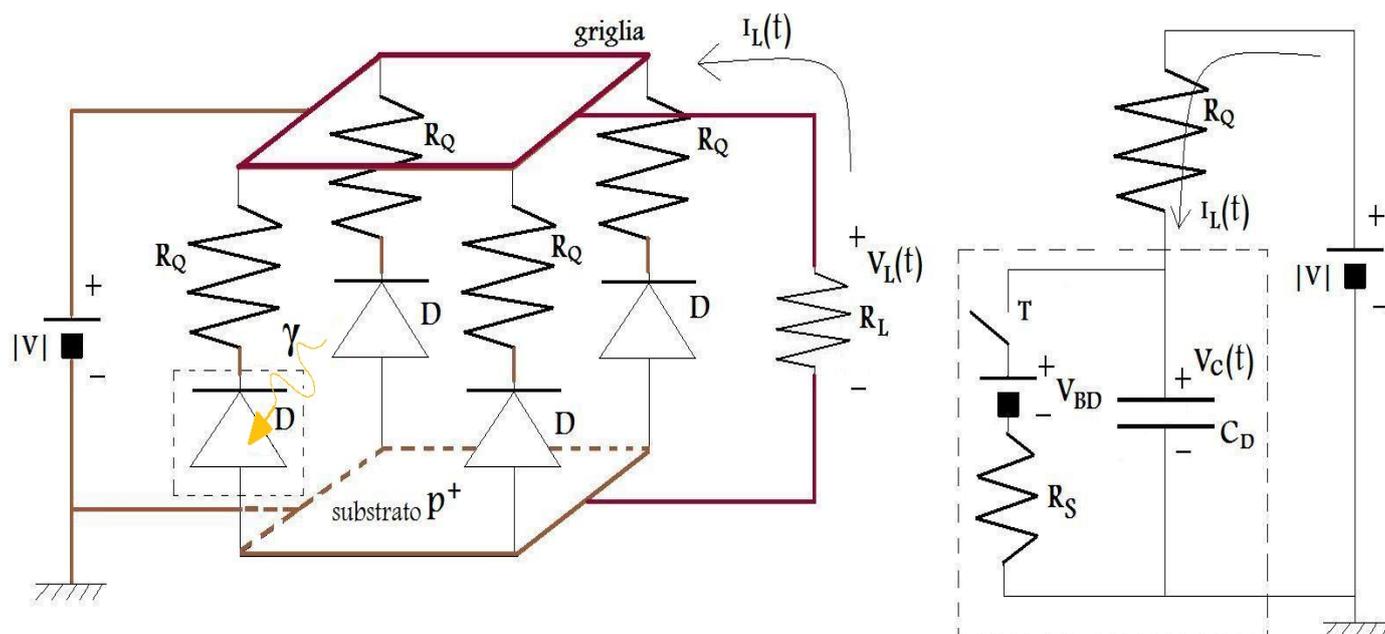


Fig. 130

A sinistra è riportato l'equivalente circuitale di un SiPM (nell'esempio un pixel riceve un fotone): si notino la pila di polarizzazione inversa, la griglia superiore che consente la raccolta delle fotocariche, le microcelle con in serie i resistori di quenching R_Q , il substrato comune p^+ , che insieme alla griglia garantisce la sistemazione in parallelo dei pixel, ed il resistore di lettura R_L . A destra è raffigurata la parte della matrice coinvolta nell'evento luminoso e l'implementazione del circuito equivalente per il singolo SAM - APD.

Supponiamo che prima dell'istante $t = 0$ la matrice non sia illuminata: ai capi di ciascun pixel insiste una tensione di polarizzazione $|V|$, la fotocorrente raccolta dalla griglia è nulla ($I_L(t) = 0$) e quindi la fototensione $V_L(t) = 0$, trascurando qualunque effetto legato al rumore. All'istante $t = 0$ la matrice viene illuminata da un impulso di luce che ipotizziamo costituito da un singolo fotone: il fotone, passato attraverso l'ossido, interagisce con una delle $N_{\text{pixeltotali}}$ microcelle del SiPM, per cui inizia a formarsi una valanga, ovvero una fotocorrente $I_{\text{pix}}(t) = I_L(t)$. Questa cresce assai rapidamente, raggiungendo il 63% ($\approx [1 - e^{-1}] \times 100$) circa del suo valore massimo ($[|V| - V_{BD}]/R_Q$) in un intervallo di tempo brevissimo: un valore tipico potrebbe essere 50 ps. L'andamento della fotocorrente è ricavabile dal circuito di destra, nel quale la situazione del pixel negli istanti di preassorbimento è modellata con l'interruttore T aperto. In questo modo l'unica parte importante del circuito è la squadra $R_Q C_D$, nella quale, a regime, abbiamo corrente I_L nulla e tensione $V_C = |V|$. L'arrivo del fotone è schematizzato, dal punto di vista dell'equivalenza circuitale, dalla chiusura di T a $t = 0$: C_D vede il parallelo fra R_S ed R_Q , ma essendo $R_S \ll R_Q$ ($R_S \approx 1 \text{ k}\Omega$, $R_Q \approx 350 \text{ k}\Omega$) questo è approssimabile con la sola R_S . Dunque C_D si scarica da $|V|$ a V_{BD} , per cui la fotocorrente $I_L(t)$ cresce, e con lei $|V_L(t)|$ (questa fase prende il nome di "carica Geiger"), con la stessa costante di tempo $\tau_{\text{rise}} \approx R_S C_D \approx 50 \text{ ps}$. Non appena I_L raggiunge il picco (il picco della valanga), inizia a farsi sentire la redistribuzione di campo totale causata da R_Q , con la conseguente estinzione del fotosegnale. Questo fenomeno (noto come "scarica Geiger") è schematizzato, circuitualmente, dalla riapertura di T: C_D si carica da V_{BD} a $|V|$, per cui la tensione su R_Q diminuisce ed anche la fotocorrente e la fototensione decrescono, con una costante di tempo $\tau_{\text{fall}} = R_Q C_D \approx 300 \tau_{\text{rise}} \approx 17 \text{ ns}$.

In figura 131 riportiamo i grafici semiquantitativi di $V_C(t)$, $I_L(t)$ e $V_L(t)$ concernenti il modello appena discusso, ed una misura sperimentale di $V_L(t)$.

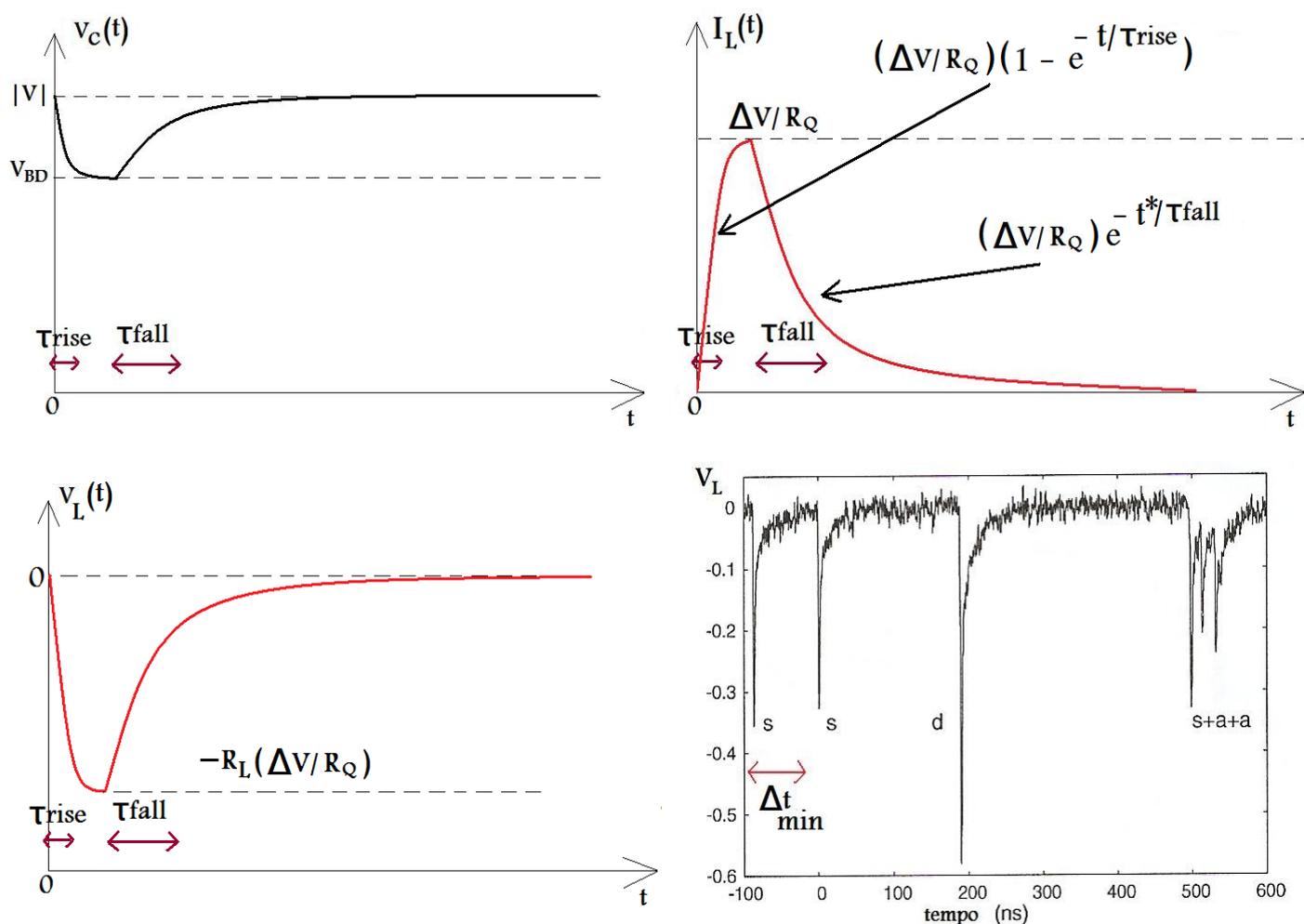


Fig. 131

Grafici semiquantitativi di $V_c(t)$, $I_L(t)$ e $V_L(t)$, concernenti l'equivalente circuitale RC con il quale abbiamo modellizzato un SiPM durante la rivelazione di un evento luminoso, ed una misura sperimentale di $V_L(t)$.

Consistentemente con questa modellizzazione RC, è immediato definire il guadagno G , di singola microcella, come il rapporto fra la carica totale che attraversa la sua giunzione, in conseguenza dell'assorbimento di un fotone, e la carica elementare:

$$G = \frac{Q_{cella}}{q} = C_D \frac{|V| - V_{BD}}{q}$$

$$G_{SiPM} = G N_{pixelcolpiti}$$

All'aumentare della temperatura, con $|V|$ fissata, l'energia cinetica che deve avere una fotocarica, primaria o secondaria, per innescare la valanga aumenta (l'energia di ionizzazione è proporzionale alla temperatura), per cui il guadagno diminuisce, come mostrato dal grafico di figura 132, relativo ad un SiPM di area 1 mm^2 .

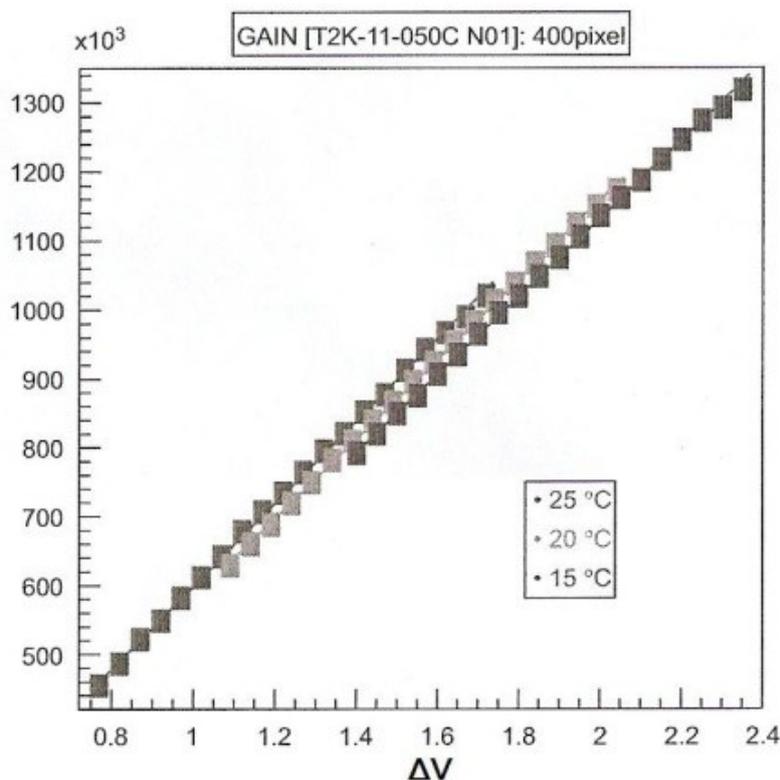


Fig. 132

Grafico del guadagno di un SiPM (G_{SiPM}) costituito da 400 microcelle, in funzione dell'overvoltage ΔV . Le curve visualizzate sono parametrizzate dalla temperatura: si noti come, fissata la polarizzazione inversa, all'aumentare della temperatura diminuisce il guadagno.

8.5) La frequenza di eventi di buio ("dark count rate") di un SiPM

Esaminiamo nuovamente la figura 131: confrontando l'andamento del fotosegnale $V_L(t)$, ottenuto dall'ideale equivalenza circuitale, con quello reale, rilevato sperimentalmente (l'impulso "s", ovvero la risposta del SiPM ad un singolo evento luminoso, possibilmente un singolo pixel), si nota che in quest'ultimo sono presenti importanti fluttuazioni. Abbiamo già esaminato e modellizzato alcune delle principali sorgenti di rumore nei fotorivelatori. Va detto, tuttavia, che l'elevato guadagno di una microcella Geiger consente di estrarre assai efficacemente il fotosegnale da alcuni rumori, quali ad esempio la corrente dovuta all'estrazione di minoritari termici, in virtù della polarizzazione inversa (corrente di dark bulk), oppure la corrente dovuta all'impattazione ionica, di questi ultimi portatori, all'interno della regione di moltiplicazione (corrente di ionizzazione), ecc... Queste correnti di rumore risultano di scarsa importanza, se confrontate con il fotosegnale $V_L(t)$, e ciò costituisce il principale vantaggio della modalità di funzionamento Geiger. Tuttavia il pixel non è esente da disturbi: infatti risulta affetto da fluttuazioni stocastiche, di natura termica, note come "dark count rate" (frequenza di eventi di buio), tanto più importanti quanto maggiori sono la temperatura del Si ed il voltaggio inverso $|V|$ a cui il SiPM è soggetto. All'interno della regione p a campo elevato, della singola microcella, avvengono, nell'unità di volume e di tempo, un certo numero di generazioni e ricombinazioni di coppie di portatori. Come sappiamo un elettrone, per motivi termici, ha una certa probabilità di attraversare la barriera di potenziale costituita dal gap e di andare in BC, lasciando una lacuna in BV. La termogenerazione di questo elettrone non è distinguibile da una fotogenerazione che avviene nello stesso istante, dal momento che entrambi gli elettroni, andando in BC, forniscono al Si una corrente impulsiva a delta di Dirac, la quale viene "filtrata" dal semiconduttore del pixel e proposta, in uscita, sulla griglia di A_1 . Quest'ultimo segnale ha la nota forma $h(t)$. Dunque l'impulso $h(t)$ derivante da un reale evento di fotoni è indistinguibile da quello dovuto ad una generazione termica, ed è tale simultanea sovrapposizione a costituire le fluttuazioni osservabili nella misura sperimentale di $V_L(t)$, riportata in figura 131, soprattutto durante la scarica Geiger. L'ampiezza del disturbo $h(t)$, ovvero D_{kt} è una grandezza statistica e dipende dal k_{esimo} evento termico e moltiplicativo.

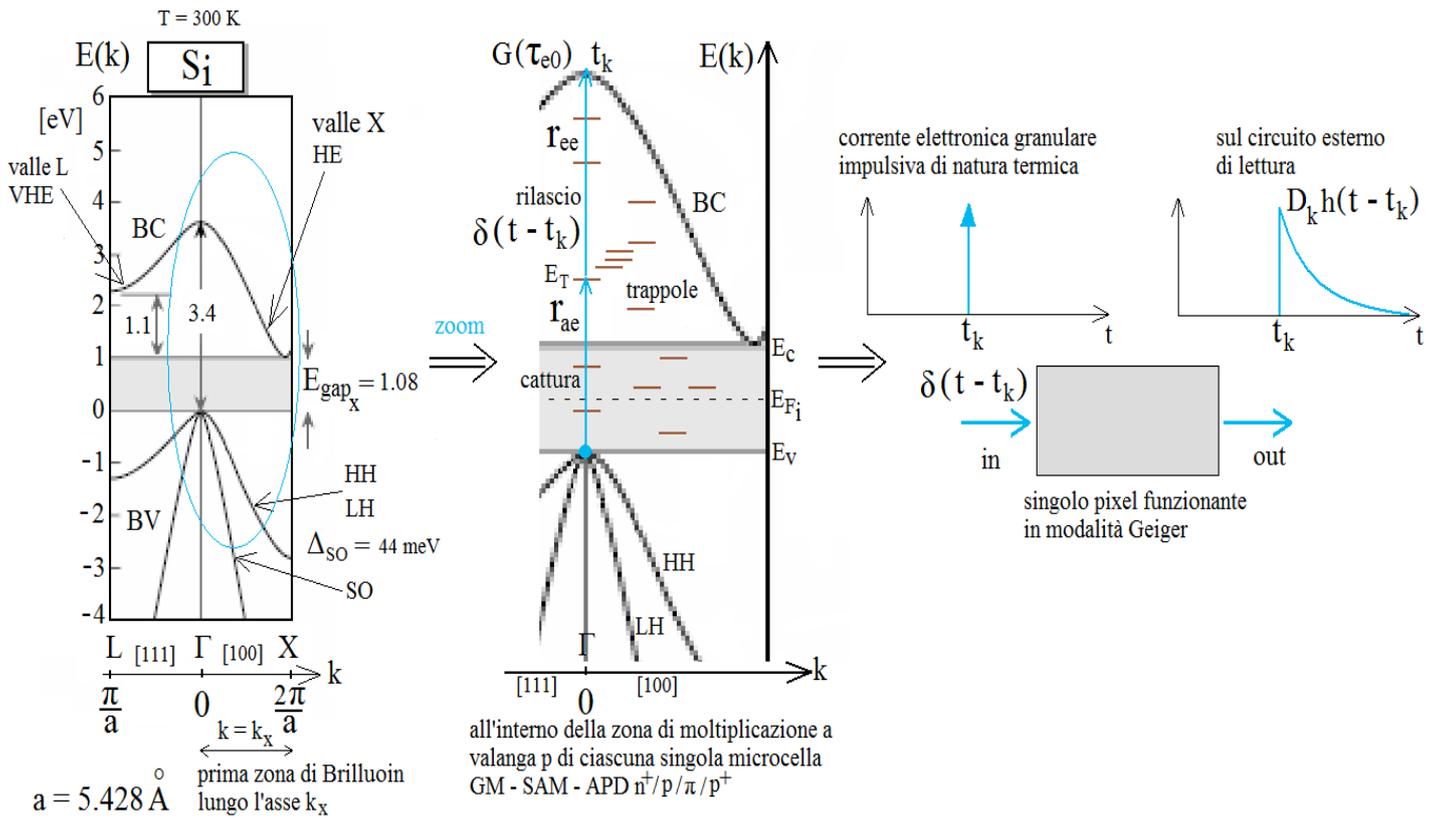


Fig. 133

Rappresentazione del processo stocastico di generazione termica granulare (G) di elettroni in BC (il cui tempo caratteristico è pari a τ_{e0}), all'interno della regione di moltiplicazione (p) di una singola microcella GM - APD. La termogenerazione di portatori, indistinguibile dalla fotogenerazione dovuta alla luce che irradia il SiPM, non è inibita dal fatto che il silicio, con cui è realizzato il SiPM, ha gap indiretto: i deep levels possono provvedere a mediare il processo G, rendendolo più probabile.

Ciascun pixel, al buio, può essere visto come un sistema lineare che riceve in ingresso un certo numero di termogenerazioni elettroniche al secondo, ovvero di funzioni delta, ciascuna in un istante poissoniano t_k , e che emette sulla griglia quello stesso numero di impulsi di buio $D_k h(t - t_k)$ al secondo, i quali si sommano fra loro, producendo una corrente $I_{darkpix}(t)$. La somma di tutte le $I_{darkpix}(t)$ costituisce una $I_{dark}(t)$ che, attraversando R_L , produce la $V_{dark}(t)$ che si sovrappone alla $V_L(t)$ ed il cui effetto è ben visibile sul display dell'oscilloscopio, come mostrato in figura 131. Il fatto che il gap del Si sia indiretto non costituisce un serio motivo di attenuazione del dark count rate, che può arrivare ad alcuni MHz, dal momento che la termogenerazione può essere intermedia, dunque incentivata, dalla presenza dei deep levels, i quali catturano un portatore e, successivamente, lo liberano, facilitandone il superamento del gap.

Con riferimento alla figura 133 possiamo affermare che il rate di generazione granulare termica netta per gli elettroni, indicato con G, è formato da un rate di assorbimento elettronico da parte delle impurezze (r_{ae}) e da un rate di emissione elettronica da parte delle stesse (r_{ee}). r_{ae} indica appunto il numero di elettroni di valenza sull'unità di volume e di tempo, all'interno della regione di moltiplicazione (p) della singola microcella GM - APD, che vengono catturati dai deep levels, supponendo che questi si trovino tutti sullo stesso livello energetico E_T . L'assorbimento di un elettrone da parte di una trappola equivale all'emissione di una lacuna, verso la BV, da parte della stessa trappola, per cui r_{ae} può essere espresso come:

$$r_{ae} = P_h N_T [1 - f(E_T)] \quad [\text{cm}^{-3} \text{ s}^{-1}]$$

P_h rappresenta la probabilità di emissione delle lacune in BV, ossia di cattura degli elettroni di valenza nei centri intrappolatori all'interno del gap, mentre $N_T [1 - f(E_T)]$ corrisponde alla concentrazione di trappole, sul livello E_T , libere da elettroni (f è ovviamente la distribuzione di Fermi - Dirac).

r_{ee} indica il numero di elettroni, in un primo momento catturati dai deep levels, all'interno della regione di moltiplicazione (p) della singola microcella GM - APD, che vengono rilasciati dalle impurezze sull'unità di volume e di tempo, supponendo che questi centri di intrappolamento si trovino tutti sullo stesso livello energetico E_T . r_{ee} può essere espresso come:

$$r_{ee} = P_n N_T f(E_T) \quad [\text{cm}^{-3} \text{s}^{-1}]$$

P_n rappresenta la probabilità di rilascio degli elettroni in BC (può essere interpretata anche come la probabilità con cui la trappola cattura una lacuna proveniente dalla BC), mentre $N_T f(E_T)$ corrisponde alla concentrazione di trappole, sul livello E_T , occupate da elettroni.

È evidente che l'analisi fin qui condotta riferita agli elettroni vale, in modo del tutto analogo, anche per le lacune.

L'espressione del rate netto di generazione termica G , al quale contribuiscono r_{ae} e r_{ee} , è riportata di seguito ed è nota come "equazione di Shockley - Read - Hall":

$$G = v_{th} N_T \sigma \frac{np - n_i^2}{p + n + 2n_i \cosh\left(\frac{E_T - E_{Fi}}{kT}\right)} \quad [\text{cm}^{-3} \text{s}^{-1}]$$

È immediato osservare come questo tasso di generazione sia tanto maggiore quanto più il semiconduttore è fuori dall'equilibrio, ossia quanto più $np \neq n_i^2$, e quindi quanto più la polarizzazione inversa $|V|$ è forte, e quanto più i centri di intrappolamento, supposti tutti sul livello energetico E_T , sono collocati vicino al centro del gap: infatti se $E_T \rightarrow E_{Fi}$ (livello di Fermi intrinseco), il coseno iperbolico tende ad 1, ovvero al suo valor minimo, e quindi G tende al suo valor massimo (a parità di tutte le altre condizioni). v_{th} è la velocità termica degli elettroni e delle lacune e σ ($\approx \sigma_n \approx \sigma_p$) è la sezione d'urto di cattura, caratteristica delle trappole, per gli elettroni e per le lacune.

Se fissiamo $|V|$ e aumentiamo la temperatura, assistiamo ad un aumento del rate relativo ai processi di generazione e ricombinazione termiche: abbiamo, nell'unità di volume e di tempo, un maggior numero di elettroni che vanno in BC e quindi un maggior numero di delta di Dirac in ingresso al pixel, ovvero un maggior numero di impulsi $D_{kh}(t - t_k)$; pertanto $I_{\text{darkpix}}(t)$ è più elevata, e di conseguenza lo è anche la $V_{\text{dark}}(t)$ che si sovrappone al fotosegnale utile $V_L(t)$. Infatti se definiamo τ_{e0} e τ_{h0} come i tempi caratteristici dei processi di cattura e rilascio da parte delle trappole, rispettivamente, degli elettroni e delle lacune, secondo le seguenti espressioni:

$$\tau_{e0} \approx \frac{1}{\sigma_n v_{th} N_T} \quad \tau_{h0} \approx \frac{1}{\sigma_p v_{th} N_T}$$

possiamo notare che all'aumentare della temperatura v_{th} aumenta, e quindi τ_{e0} e τ_{h0} diminuiscono, i processi di cattura e rilascio suddetti diventano più frequenti ed il dark count rate risulta più alto ($V_{\text{dark}}(t)$ è più importante). $n_i = n_i(T)$ è una funzione crescente della temperatura (per T abbastanza alte è, in pratica, una funzione esponenziale), secondo la nota relazione:

$$n_i(T) = \sqrt{N_C N_V} e^{-\frac{E_{gap}}{2kT}} \propto T^{\frac{3}{2}} e^{-\frac{E_{gap}}{2kT}} \approx e^{-\frac{E_{gap}}{2kT}}$$

dove N_C ed N_V sono le ben note densità di stati equivalenti in BC e in BV (citate nel paragrafo 4.1).

Se fissiamo la temperatura T e aumentiamo l'overvoltage ΔV , assistiamo ad un incremento lineare della concentrazione, al secondo, di impulsi di buio, oltre ad un aumento di $\langle D_k \rangle$. Ciò comporta un incremento di $V_{\text{dark}}(t)$. Da qui si comprende l'inopportunità di polarizzare un SiPM con $|V|$ troppo alta: esiste un valore ottimale V_{PL} che consente, anche in modalità Geiger, di massimizzare il SNR di microcella. Le proporzionalità del dark count rate con T e ΔV sono illustrate graficamente in figura 134.

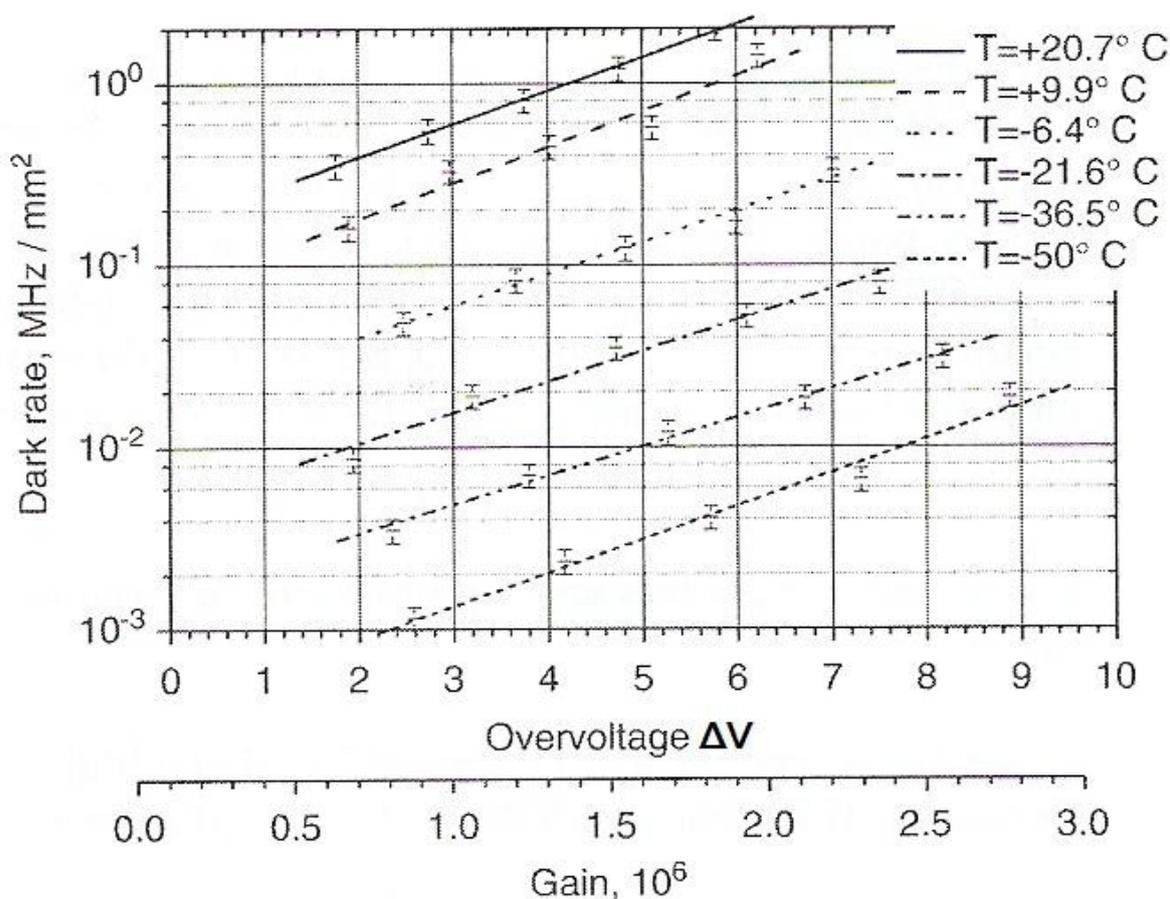


Fig. 134

Grafico semilogaritmico, ottenuto da misure in laboratorio, della dipendenza del dark count rate di un tipico SiPM dall'overvoltage ΔV , alla cui scala quotata è associata quella del guadagno, dovuto alla polarizzazione inversa $|V| = V_{BD} + \Delta V$. Le caratteristiche del dark count rate sono parametrizzate dalla temperatura T e risultano lineari rispetto all'overvoltage.

8.6) Il crosstalk ottico all'interno di un SiPM

Riferiamoci sempre alla misura, tramite oscilloscopio, del fotosegnale $V_L(t)$, riportata in figura 131. Questa mostra due fototensioni "s", lette su R_L , collegabili a due singoli eventi luminosi. Successivamente è presente un segnale "d" che può essere associato all'incidenza, sul SiPM, di un evento di entità doppia rispetto a quelli verificatisi in precedenza (potrebbero essere arrivati due fotoni, nell'ipotesi di $PDE \approx 1$, oppure un ancora maggior numero di fotoni, nel caso più realistico $PDE < 1$). C'è anche la possibilità che l'evento luminoso in questione sia esattamente lo stesso dei due precedenti, solo che ora si è verificata un'interferenza ottica, o "optical crosstalk", che ha comportato l'attivazione di un pixel che nulla aveva a che fare con la radiazione luminosa incidente. Nel momento in cui una microcella assorbe un fotone, si possono verificare uno di questi tre fenomeni:

- Una piccola frazione di fotoportatori (primari o secondari) può uscire dal volume di moltiplicazione del pixel colpito ed andare nella regione di guadagno di una microcella adiacente, innescando lì una valanga. Ad essa corrisponde una corrente $I_{pix}(t)$ che si somma a quella "legittima" uscente dal pixel illuminato. Sul circuito esterno di lettura risulta un fotosegnale tipico di due fotoni, ma in realtà solo una cella ha reagito all'illuminazione: l'altra ne ha solo subito l'interferenza.
- Durante la formazione della valanga, all'interno del pixel colpito, due fotoportatori, magari aiutati da qualche trappola profonda, si ricombinano, producendo (per la conservazione dell'energia) un fotone. Questo lascia il pixel di origine e viaggia verso un limitrofo, dove viene assorbito, dando luogo ad una valanga di crosstalk.
- Sempre durante la moltiplicazione nel pixel illuminato, una fotocarica potrebbe incontrare, lungo il suo tragitto verso il bulk di destinazione, un centro di scattering (un altro portatore di carica, una carica spaziale di drogaggio, una vacanza o un'impurezza metallica trattenenti un portatore di carica, ecc...):

una brusca “frenata” coulombiana comporterebbe l’emissione, da parte della fotocarica in questione, di raggi X. Questi fotoni, di lunghezza d’onda cortissima, viaggerebbero fino ad un pixel limitrofo, dove potrebbero essere assorbiti. Ciò comporterebbe una fotocorrente di crosstalk.

Osserviamo il grafico di figura 135. Se fissiamo una delle caratteristiche, ovvero la distanza fra due pixel della matrice, rispetto ai quali vogliamo valutare la probabilità di interferenza, ed aumentiamo il guadagno, la fotovalanga in una delle due microcelle diventa più consistente e la probabilità che si verifichi quanto sopra spiegato aumenta. Se invece fissiamo il guadagno, è chiaro che all’aumentare della distanza fra due pixel la probabilità che questi interferiscano otticamente fra loro diminuisce.

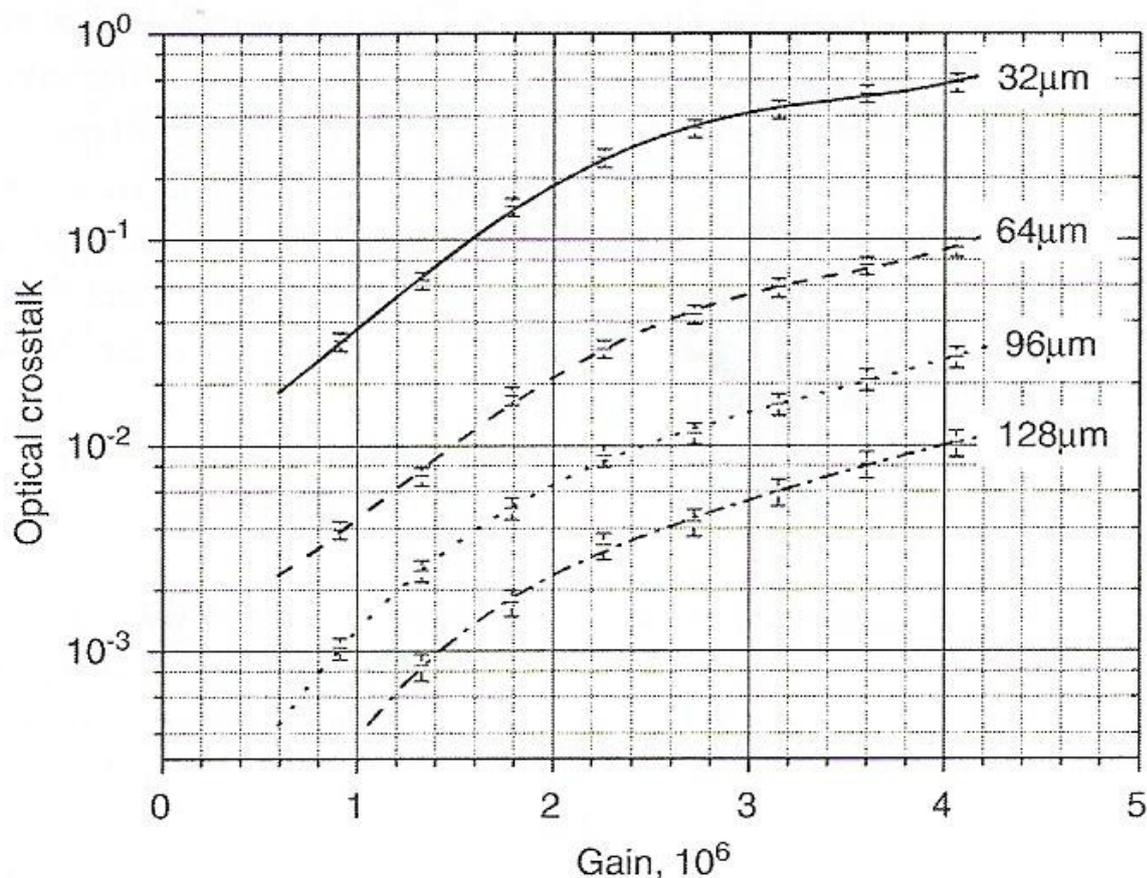


Fig. 135

Grafico semilogaritmico delle curve caratteristiche della probabilità d’interferenza ottica, fra due pixel di un SiPM, in funzione del guadagno del SiPM. Le curve sono parametrizzate dalla distanza (micrometrica) che separa i pixel rispetto ai quali valutiamo la probabilità di crosstalk.

Uno degli effetti più importanti della trench, descritta nel paragrafo 7.1.4 e visualizzata in sezione in figura 108, è proprio limitare tale interferenza, e quindi salvaguardare l’accuratezza dell’informazione portata dal fotosegnale.

8.7) La risoluzione temporale di un SiPM: il problema degli “after pulses” relativo alla scarica Geiger ed i ritardi legati alla carica Geiger

Concentriamoci ora sull’ultima fototensione visualizzata sul grafico, ottenuto dall’oscilloscopio, del segnale $V_L(t)$, visualizzato sempre in figura 131, ovvero quella indicata con “s+a+a”. Quando un fotone viene assorbito da una microcella sappiamo che si crea una valanga di fotoportatori, i quali contribuiscono, magari in misura assai differente ($\alpha_e \neq \alpha_h$), al suo autoaccrescimento, nonché a sostenere la fotocorrente $I_{\text{pix}}(t)$. Tuttavia è probabile che alcuni di questi fotoportatori, durante la carica Geiger, siano catturati dalle trappole presenti nel Si. Dal momento

in cui una fotocarica viene assorbita da uno stato accettore profondo, è chiaro che questa non può contribuire più alla fotoconduzione. Dunque $I_{\text{pix}}(t)$, durante la carica, è attenuata da questi intrappolamenti stocastici. Durante la scarica molte trappole potrebbero rilasciare i portatori precedentemente catturati, per cui le fotocariche appena liberate (elettroni tornati in BC e lacune in BV) vengono trascinate dal campo elevato e messe nella condizione di innescare altre valanghe, note come “after pulses” (nel nostro esempio ce ne sono due), che si sommano a quella estinguente.

Gli after pulses peggiorano la “risoluzione temporale” del SiPM, ovvero la capacità del SiPM di distinguere temporalmente, a monte della lenta elettronica di lettura, due eventi luminosi verificatisi uno dopo l'altro. Minore è l'intervallo temporale fra queste due radiazioni incidenti, che il SiPM è in grado di apprezzare e quindi di separare, a livello di risposta elettrica, maggiore è la sua risoluzione temporale. La probabilità di avere after pulses in un certo istante t , successivo all'inizio della scarica Geiger, è data dalla seguente relazione:

$$P_{\text{afterpulses}}(t) = P_c P_r(t) P_t$$

dove P_c è la probabilità di cattura, proporzionale alla concentrazione di trappole e all'entità della valanga, $P_r(t)$ la probabilità di rilascio nell'istante t , P_t la probabilità di trigger, ovvero la probabilità che i portatori rilasciati provochino ionizzazioni per impatto, dando luogo alle valanghe di ritardo. $P_r(t)$ è data da:

$$P_r(t) = e^{-\frac{t}{\tau}}$$

dove τ è il tempo medio di intrappolamento.

Nell'ipotesi che non si verificano after pulses, l'intervallo minimo di tempo Δt_{min} fra un evento luminoso ed il successivo, affinché il SiPM possa rivelare anche il secondo in modo corretto, è, qualitativamente, dell'ordine di $(R_s + R_Q)C_D$ (= intervallo di tempo di carica Geiger + intervallo di tempo di scarica Geiger), ovvero il tempo necessario alla fotovalanga, associata al primo evento, per nascere, raggiungere il suo picco ed estinguersi. R_s , R_Q e C_D sono gli elementi circuitali del modello analitico del SiPM illuminato, mostrato in figura 130. Uno o più after pulses comportano l'allungamento del tempo di estinzione (intervallo di tempo di scarica Geiger), per cui deve passare più tempo fra una rivelazione e la successiva, con conseguente degradamento della risoluzione temporale.

L'intervallo di tempo di carica Geiger è, a livello di risoluzione temporale, molto meno critico rispetto a quello di scarica, poiché non risente del fenomeno spurio degli after pulses. Tuttavia anche per il tempo di carica possiamo individuare due fattori critici, i quali possono determinare, seppur in modo marginale, un deterioramento della risoluzione temporale del SiPM:

- Il tempo medio di raccolta dei fotoportatori, ovvero il tempo a loro necessario per attraversare tutta la regione di drift che separa il punto di fotogenerazione (o di impattazione ionica) dal bulk di raccolta. Questo tempo dipende, oltre che dalla mobilità, anche dal campo elettrico: se il suo valore è dell'ordine di 10^5 V/cm, una regione di drift lunga $5 \mu\text{m}$ è attraversabile, in media, in circa 10 ps.
- Il tempo impiegato dal processo di valanga per interessare tutta la zona di moltiplicazione, che è il fenomeno con una durata media più lunga, di circa 100 ps, per cui dominante. La valanga, infatti, viene innescata in un punto del volume attivo, sottoposto a campo elevato, e si propaga tanto meno rapidamente quanto maggiore è il guadagno. Pochi istanti dopo l'innescamento si forma un gradiente di carica fra il punto di trigger e le regioni, a questo limitrofe, non ancora investite dal breakdown. Ciò comporta una diffusione laterale della fotocarica di valanga, con conseguente rallentamento della sua propagazione. I tempi caratteristici di quest'ultima sono legati anche alla posizione del punto di innescamento, e risultano leggermente differenti nel caso in cui il fotone interagisca nel cuore del volume di guadagno, piuttosto che in un angolo o lungo un lato.

La risoluzione temporale è, inoltre, notevolmente peggiorata dalla lenta elettronica di front-end: il motivo principale è l'inerzia temporale legata alle capacità parassite, la cui presenza è tanto più significativa quanto maggiore è la discontinuità fra il livello di integrazione del SiPM e quello dell'elettronica di lettura collegata.

8.8) La proporzionalità inversa fra range dinamico e fattore geometrico in un SiPM

Un altro problema tipico dei SiPM è la proporzionalità inversa fra range dinamico e fattore geometrico. Abbiamo visto che il numero finito di microcelle, integrate sulla superficie totale della matrice, è responsabile della saturazione del numero di pixel colpiti e quindi della fotocorrente raccolta dalla griglia, all'aumentare del numero medio di fotoni incidenti per singolo pixel, ragion per cui la risposta elettrica del SiPM è accurata solo per flussi fotonici ridotti. Sembra pertanto auspicabile l'aumento del range dinamico, ovvero l'aumento del numero di pixel per mm². Tuttavia all'aumentare della concentrazione di pixel corrisponde, per i limiti delle attuali tecnologie fotolitografiche, una diminuzione della superficie attiva della matrice, ovvero del fill factor G_f : non si riesce, al di sotto di certe dimensioni, a ridurre la superficie occupata da trench, resistori e piste di metallo dello stesso fattore con cui si riduce l'area totale di microcella ("scaling down" non proporzionale). Ciò comporta la diminuzione della PDE e della QE (come mostrato in figura 125). È necessario, dunque, un compromesso fra le due esigenze.

8.9) Il cristallo di scintillazione LYSO accoppiato ad un SiPM per applicazioni PET

Esaminiamo, a grandi linee, il dispositivo generalmente anteposto a qualunque SiPM, noto come "cristallo scintillatore", atto a ricevere radiazione luminosa molto energetica ed a trasformarla in un pacchetto di fotoni, di opportuna frequenza, assorbibili dalle microcelle della matrice. Come abbiamo già accennato, infatti, in ambito PET il tessuto organico sottoposto a mapping, avendo assorbito l'isotopo tracciante β^+ , emette positroni, i quali annichilano con elettroni atomici del tessuto stesso, producendo fotoni a 511 keV. Questi costituiscono luce appartenente ai raggi gamma, frequenze alle quali il Si presenta responsività nulla. Dunque la radiazione tipica della PET non può essere rilevata direttamente dai SiPM: è necessario uno strumento ottico in grado di convertire l'informazione gamma, legata alla concentrazione dell'isotopo nel tessuto, in una equivalente avente frequenze inferiori, rientranti in un intervallo spettrale per cui la sensibilità del Si sia accettabile. Nella PET sono largamente utilizzati i cristalli scintillatori LYSO, strutture inorganiche costituite da ortosilicato di lutezio - ittrio, drogate con cerio.

Un cristallo scintillatore si presenta come un parallelepipedo avente base simile o identica all'area totale del SiPM con cui deve essere assemblato, in modo da ottenere il miglior allineamento possibile, e lunghezza sufficiente ad assicurare alta probabilità di assorbimento ottico. Per un SiPM di 1mm² è comune l'utilizzo di un LYSO di dimensioni 1mm x 1mm x 10mm, generalmente rivestito di teflon bianco a scopo antiriflessivo; il contatto fra LYSO e SiPM può essere assicurato mediante l'utilizzo di un particolare grasso ottico.

Descriviamo le modalità con le quali il cristallo converte la luce gamma in luce "silicio-compatibile". Quando un pacchetto di fotoni a 511 keV incide sul cristallo e vi penetra (il multistrato di teflon serve ad assicurare un coefficiente di trasmissione accettabile), l'energia cinetica di questo viene parzialmente assorbita dal materiale. L'attitudine del cristallo a tale assorbimento prende il nome di "efficienza di rivelazione": questa è unitaria nel caso (ideale) in cui il pacchetto di fotoni, attraversando il cristallo, deposita tutta la sua energia all'interno del cristallo stesso. L'energia depositata in parte va dissipata, ad esempio per mezzo di vibrazioni fononiche, in parte va ad eccitare, lungo il percorso dei fotoni rifratti nel cristallo, gli "stati" del cristallo stesso. Si può definire "efficienza di scintillazione" Y di un cristallo come la frazione di energia depositata nel cristallo che si converte in fotoni di scintillazione, i quali, generalmente, appartengono allo spettro del visibile. In formule si può scrivere:

$$Y = \frac{\text{numero di fotoni visibili prodotti dal cristallo}}{\text{energia depositata nel cristallo (MeV)}}$$

Ad esempio per un cristallo di ioduro di sodio, drogato con tallio, si ha che:

$$Y_{\text{NaI(Tl)}} = \frac{39000}{\text{MeV}}$$

Si parla di efficienza di scintillazione unitaria nel caso ideale in cui tutta l'energia fotonica deposta nel cristallo venga convertita in energia di scintillazione, ovvero vada ad eccitare i suddetti stati. Va puntualizzato che di solito la frazione di energia depositata nel cristallo che si converte in scintillazione è piuttosto bassa. Gli stati del cristallo sono costituiti sia da elettroni atomici del cristallo sia da elettroni forniti dai droganti usati. L'energia di scintillazione promuove gli elettroni atomici e quelli legati alle impurezze droganti su dei particolari livelli energetici superiori E_s , abbastanza vicini fra loro. Dopo un intervallo di tempo pari a 4 - 5 volte il periodo di decadimento relativo a quei livelli instabili (circa 40ns nel caso del LYSO), la maggior parte degli stati si è già diseccitata, ossia è tornata allo stato fondamentale E_0 (stabile), emettendo dei fotoni di scintillazione, a delle frequenze, tutte intorno a quella centrale $\nu_s - \nu_0$, rispetto alle quali il cristallo è di fatto trasparente. Tale trasparenza passiva è dovuta al fatto che il numero di degenerazione del livello fondamentale è molto maggiore del numero di degenerazione dei livelli (trattabili quasi come un unico livello) superiori ($g_{E_0} \gg g_{E_s}$). Tali fotoni, la maggior parte dei quali, nel caso del LYSO, corrisponde a lunghezze d'onda comprese fra 400 e 450 nm, ovvero luce blu, attraversano praticamente indenni il cristallo, il quale costituisce una sorta di guida ottica, lungo la sua dimensione maggiore, fino ad arrivare all'interfaccia LYSO/SiPM: si definisce "efficienza di raccolta" la frazione di fotoni di scintillazione che riesce a raggiungere il SiPM. Questi vengono rilevati dal SiPM, il quale avrà una finestra antiriflettente progettata in modo da massimizzare il coefficiente di trasmissione per lunghezze d'onda intorno a 420nm, presso le quali il cristallo LYSO presenta l'emissione massima (il massimo della Line Shape).

In figura 136, a sinistra, riportiamo la struttura che potrebbe costituire il primo stadio di un buon apparato PET, ovvero una matrice 4x4 formata da 16 SiPMs, ciascuno costituente un canale di rivelazione, aventi caratteristiche simili fra loro, montati su un substrato p^+ comune. La matrice raffigurata è un prototipo costruito nel centro di ricerca FBK (Fondazione Bruno Kessler) di Trento. Ciascun pad di alluminio è collegato alla griglia di un singolo SiPM, dal quale è possibile raccogliere la fotocarica legata all'evento luminoso che ha interessato quel particolare SiPM. Il segnale luminoso gamma è rilevato e convertito in luce di scintillazione blu da una matrice 4x4 di cristalli LYSO, ciascuno di dimensioni 0.96 mm x 0.96 mm x 10mm, separati l'uno dall'altro da un film di resina bianca di spessore 100 μ m. I 16 cristalli (riportati in figura 136, a destra) si accoppiano perfettamente ai 16 SiPMs.

È possibile effettuare una misura di calibrazione volta a mappare la dispersione spaziale, cioè il crosstalk ottico descritto nel paragrafo 8.6, della risposta elettrica, della matrice dei 16 SiPMs, ad un evento luminoso che, per le leggi dell'ottica lineare, dovrebbe interessare solo un particolare SiPM. La sorgente di luce in questione potrebbe essere un LASER abbastanza coerente in frequenza, che emette luce blu (stiamo considerando la matrice 4x4 priva di scintillatori LYSO, cioè stiamo calibrando solo la matrice di SiPMs), avente un diagramma d'irradiazione $D(\theta, \varphi)$ molto sottile e diretto verso quel particolare SiPM. In realtà quel SiPM, verso cui è prevalentemente diretta la luce del LASER (quest'ultimo, per emettere luce blu, deve possedere un core realizzato con un semiconduttore composto III - V a base di azoto, cioè un semiconduttore nitridico, come il GaN oppure il AlN), non è l'unico, fra i 16 che costituiscono la matrice 4x4, a rispondere all'evento: gli otto SiPMs limitrofi presentano una risposta elettrica tutt'altro che trascurabile, dovuta appunto all'interferenza ottica. L'istogramma tridimensionale noto come "Hitmap", riportato in figura 136, in basso, rappresenta la risposta elettrica, della matrice 4x4, al suddetto evento luminoso. Come si può notare, è presente il crosstalk ottico fra i 16 SiPMs, ma è facilmente distinguibile il canale corrispondente alla posizione, in linea di volo, della sorgente di luce, dato che il canale raccoglie, in questo esempio particolare, più del 90% delle incidenze fotoniche rivelate (più precisamente, più del 90% della fotocarica complessivamente raccolta da tutte le pads laterali).

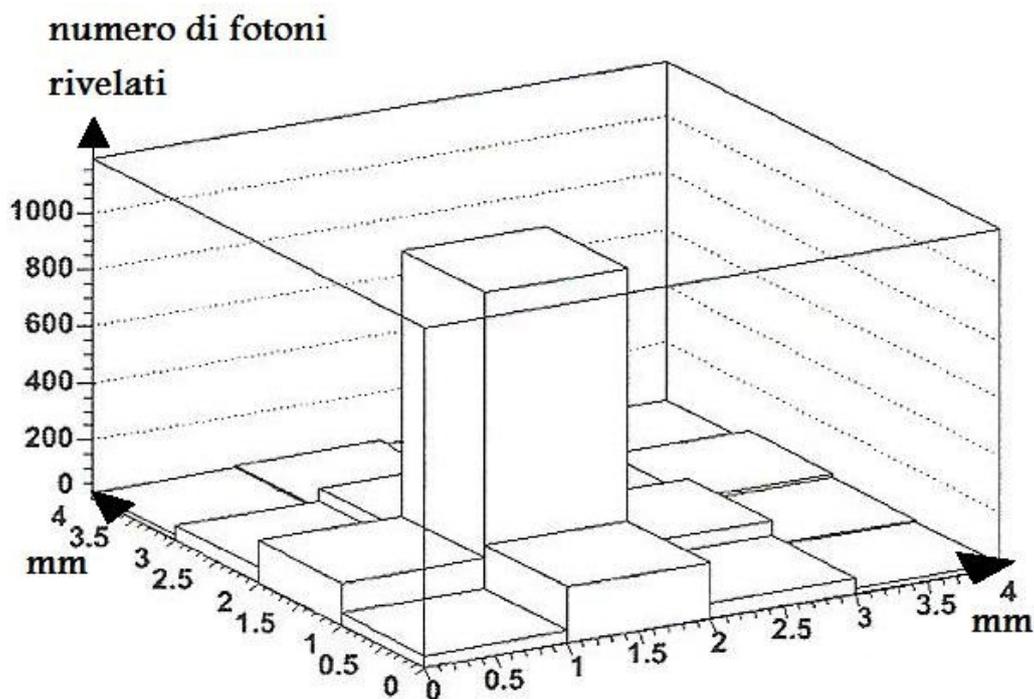
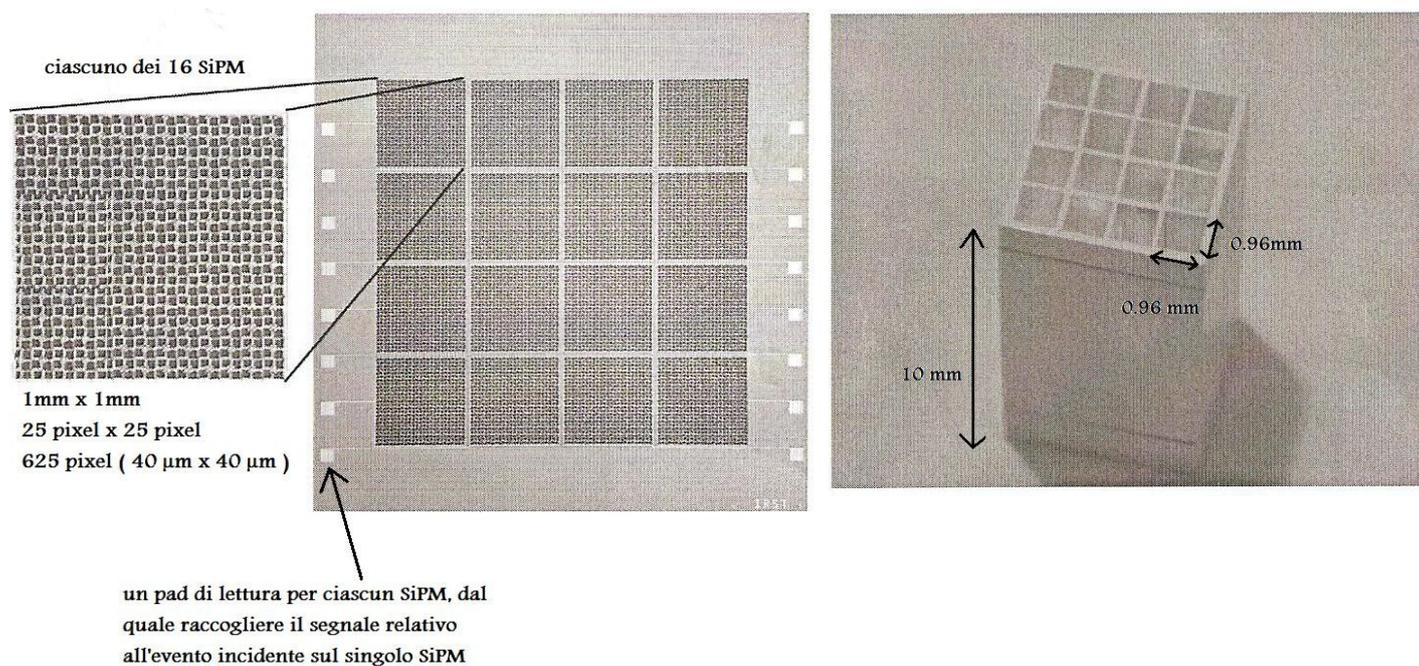


Fig. 136

A sinistra è rappresentata la matrice 4x4 di 16 SiPMs (nell'immagine è riportato il prototipo costruito nel centro di ricerca FBK di Trento), montati a substrato p⁺ comune, con i 16 pad laterali sui quali viene raccolto il fotosegnale; ciascun pad è collegato alla griglia metallica di un singolo SiPM. A destra riportiamo la matrice 4x4 di cristalli scintillatori LYSO da accoppiare alla matrice di SiPMs suddetta. I due sistemi assemblati insieme potrebbero costituire un buono stadio di fotorivelazione per applicazioni PET. In basso riportiamo l'Hitmap relativo ad una misura di calibrazione, effettuata sulla matrice di SiPMs priva dei cristalli scintillatori, al fine di testare l'importanza del crosstalk ottico. Questo risulta presente ma non deleterio ai fini della conservazione dell'informazione circa la posizione, in linea di volo, della sorgente luminosa. Nell'esperimento a cui si riferisce l'Hitmap la sorgente ottica è rappresentata da un LASER di luce blu, realizzato in GaN o in AlN, molto direttivo, cioè avente un diagramma d'irradiazione molto coerente nello spazio, puntato sul SiPM che mostra il maggior numero di eventi luminosi incidenti, nel nostro caso particolare oltre il 90% della fotocarica complessivamente raccolta da tutte le pads laterali della matrice 4x4.

Appendici

A) L'epitassia da fasci molecolari (MBE - "Molecular Beam Epitaxy")

La MBE ("Molecular Beam Epitaxy" - epitassia da fasci molecolari) è una tecnica che permette la crescita di sottili strati di materiali cristallini al fine di formare delle eterostrutture, come ad esempio il superreticolo realizzato in $\text{In}_{(0.53)}\text{Ga}_{(0.47)}\text{As}_{(1-x)}\text{P}_x$ (o più in generale $\text{In}_{(1-y)}\text{Ga}_{(y)}\text{As}_{(1-x)}\text{P}_x$), citato nel paragrafo 7.3 ed il cui diagramma a bande è qualitativamente mostrato in figura 121, su substrati massivi, ad esempio realizzati in $\text{In}_{(0.53)}\text{Ga}_{(0.47)}\text{As}_{(1-x)}$. La caratteristica principale della MBE è di operare in condizioni di ultra - alto vuoto, nel quale non si hanno urti tra le molecole o gli atomi, degli elementi sublimati, che incidono sul substrato. Il vuoto è dell'ordine di 10^{-11} Torr; ciò permette di ottenere materiali di alta purezza. L'utilizzo di un ambiente in ultra - alto vuoto consente, inoltre, di monitorare accuratamente la crescita del materiale in tempo reale, ad esempio tramite la diffrazione di elettroni, procedimento impossibile con le tecniche che operano a pressioni maggiori. Questo monitoraggio in tempo reale, unito alla bassa velocità di crescita che si può ottenere (anche inferiore a 0,1 nm/s), permette di controllare lo spessore dello strato cresciuto con un'accuratezza dell'ordine dello spessore del singolo monostato atomico del materiale in crescita, che è di circa 1 \AA . In figura 137 è mostrato lo schema semplificato di un reattore per MBE.

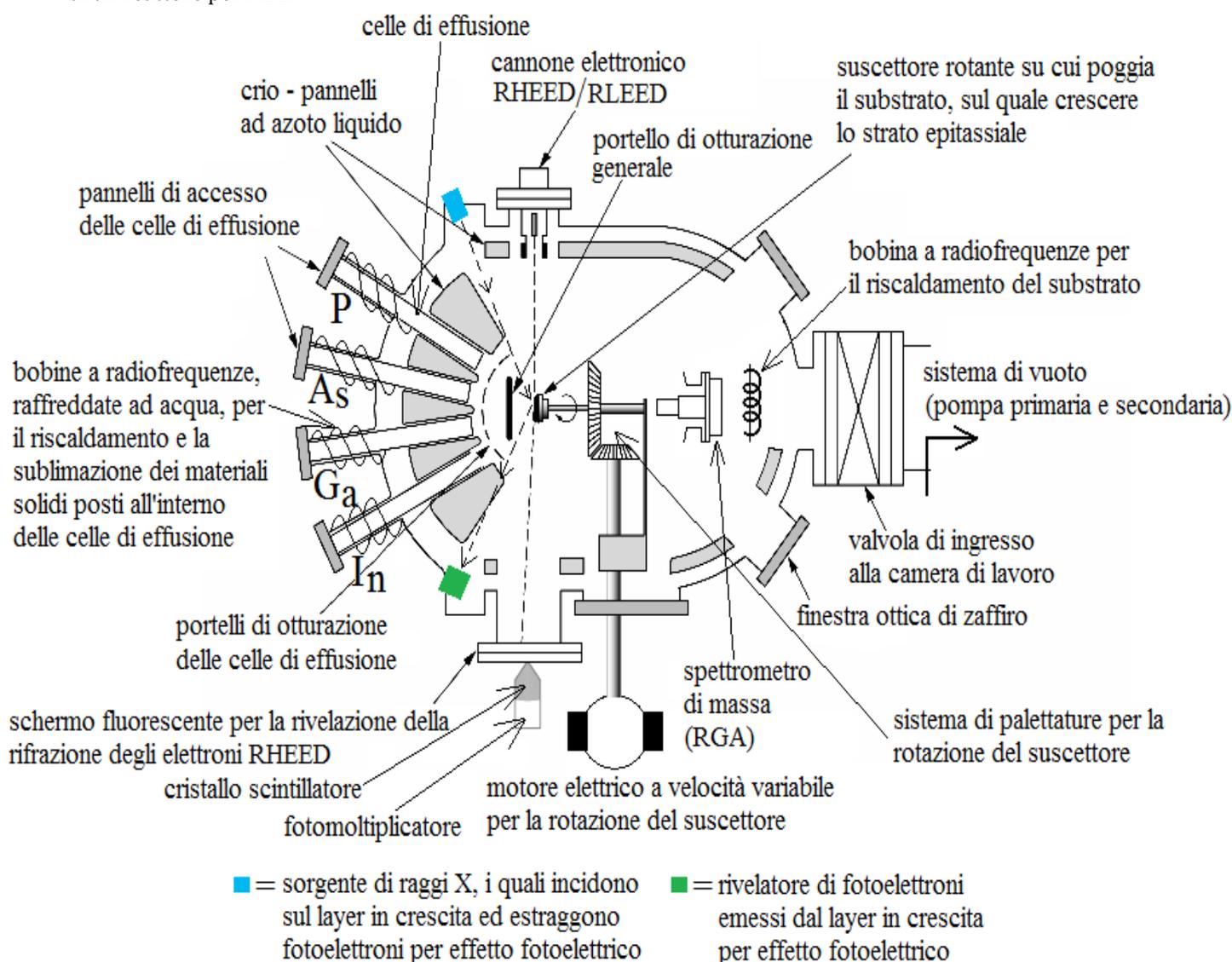


Fig. 137

Schema semplificato di un reattore utilizzabile per la crescita di eterostrutture (il disegno è riferito alla crescita del superreticolo illustrato in figura 120) mediante la tecnica MBE ("Molecular Beam Epitaxy" - epitassia da fasci molecolari). Sono visualizzati la camera di lavoro (camera ad ultra - alto vuoto), la strumentazione essenziale per la crescita degli strati epitassiali ed il controllo di spessore e qualità di quest'ultimi.

In figura 138 è mostrato qualitativamente, solo a titolo di esempio, un comune sistema di vuoto assemblato insieme ad un reattore MBE oppure alla camera di lavoro di un sistema di epitassia da fase vapore VPE; il sistema di vuoto è costituito da una pompa primaria e da una secondaria.

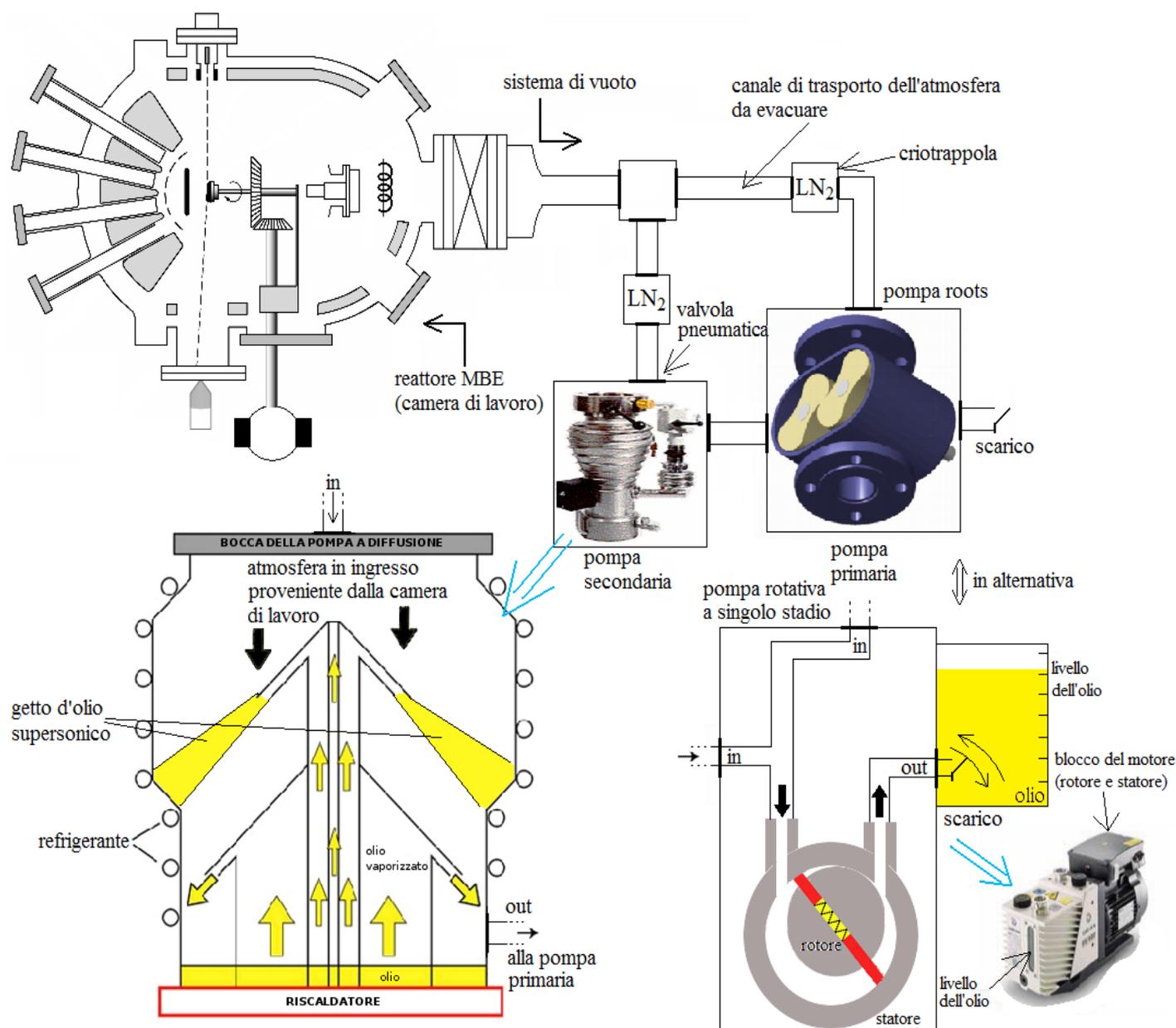


Fig. 138

Schema di principio di un comune sistema di vuoto assemblato insieme ad un reattore MBE oppure ad uno VPE, costituito da una pompa primaria e da una secondaria.

Il costo dell'apparecchiatura MBE è determinato principalmente dal grado di sofisticazione del sistema di vuoto, necessario a favorire la purezza dei layers così cresciuti. Infatti, anziché usare un'economica pompa meccanica rotativa (a singolo stadio, che arriva fino a circa 5mTorr, con una velocità di pompaggio di circa 60m³/h, oppure a doppio stadio, che arriva fino a circa 10⁻⁵ Torr), oppure una pompa "roots" o comunque una qualunque pompa a spostamento di parete, in qualità di pompa primaria, in un sistema MBE avanzato è opportuno utilizzare una ben più costosa criopompa (chiamata anche pompa criogenica), ad esempio ad azoto liquido (- 173 °C), che crea il pre - vuoto (760Torr → 5mTorr) senza l'inconveniente dei reflussi di vapori d'olio, i quali contaminerebbero gli strati epitassiali crescenti (una criopompa non necessita infatti di lubrificazione, al contrario delle pompe meccaniche). Molto usata è la pompa criogenica "ad adsorbimento" mostrata, qualitativamente, in figura 139.

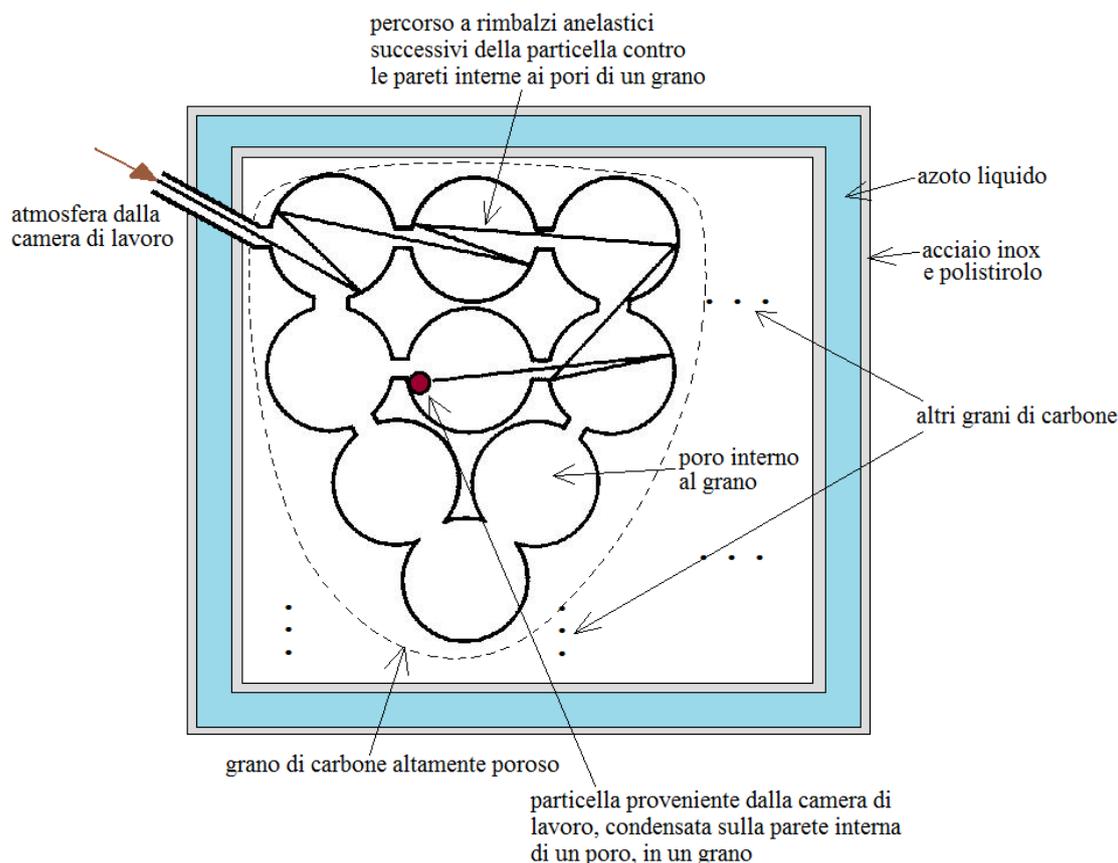


Fig. 139

Schema di principio di una comune pompa criogenica “ad adsorbimento”, utilizzabile come pompa primaria all’interno del sistema di vuoto assemblato ad un reattore MBE, al posto di una più economica pompa a spostamento di parete, tipica di un sistema di vuoto assemblato ad un reattore VPE.

La pompa criogenica “ad adsorbimento” comprende una regione di materiale altamente poroso, ad esempio un volume di granelli di carbone, ciascuno dei quali presenta, al proprio interno, un’elevata porosità, ovvero ciascun grano di carbone è caratterizzato da un elevato rapporto superficie/volume. L’intero volume poroso si trova all’interno di un contenitore a doppia parete, nella cui intercapedine è versato azoto liquido; le pareti sono realizzate con una struttura multilaminare di acciaio inox e polistirolo. I grani di carbone sono raffreddati fino a temperature criogeniche dall’azoto liquido. Le molecole di gas provenienti dalla camera di lavoro, che entrano in un grano, rimbalzano anelasticamente molte volte nei pori interni, perdendo energia finché non restano aderenti alle pareti interne ai pori del grano, grazie alle forze di coesione, le quali garantiscono l’azione di adsorbimento che produce l’evacuazione dell’atmosfera dalla camera di lavoro. Le forze di coesione sono superiori a quelle di agitazione termica, le quali tendono a distaccare le molecole dalle pareti interne dei pori, a causa della bassissima temperatura. Se la temperatura dei grani si dovesse innalzare, l’energia termica diventerebbe maggiore di quella di coesione e le molecole verrebbero rilasciate; tale desorbimento vanificherebbe il vuoto primario ($\sim 5\text{mTorr}$) creato all’interno del reattore MBE. L’inconveniente maggiore di questa pompa, oltre al costo, consiste nella progressiva saturazione della regione porosa.

Come pompa secondaria ($5\text{mTorr} \rightarrow 10^{-11}\text{ Torr}$) non è possibile utilizzare quella, largamente adoperata nella normale epitassia da fase vapore VPE (adatta a crescere omostrutture), a diffusione di vapori d’olio supersonici, la cui velocità di pompaggio è molto buona (circa qualche decina di litri al secondo): si tratta infatti di una pompa a trasferimento di quantità di moto che arriva solo fino a $10^{-6} - 10^{-7}\text{ Torr}$, e per giunta il suo utilizzo comporterebbe un alto tasso di inquinamento dell’ambiente di lavoro; la scelta deve cadere quindi sulle ben più costose pompe turbomolecolari oppure su quelle a sublimazione (ad esempio di titanio).

La pompa turbomolecolare è costituita da un impilamento coassiale di dischi rotanti, ciascuno fornito di palette, alternati a dischi fissi: ciascun disco rotante, detto “rotore”, ruota a diverse migliaia di giri al minuto (20000 – 40000 giri/min, come ordine di grandezza) su un disco immobile sottostante, detto “statore”, il quale è fornito di palette aventi inclinazione opposta rispetto alle palette del disco rotante sovrastante. Le molecole da evacuare

vengono raccolte ed accelerate dalle pale di ciascun disco rotante, collidono sulle pale di quello fisso e da qui rimbalzano verso le pale del disco rotante successivo. Il funzionamento è pertanto analogo a quello delle turbine di un jet: le molecole di gas vengono rimosse meccanicamente grazie all'azione combinata delle due serie di lame alternate lungo l'asse della pompa. In altre parole la combinazione di rotori e statori consente di spingere dall'alto verso il basso le molecole di gas del reattore MBE, nonostante la pressione vada aumentando, dall'alto verso il basso, a causa della diminuzione dell'inclinazione delle pale di rotori e statori. Una pompa turbomolecolare può essere montata in qualunque posizione, sia verticalmente che orizzontalmente, e non solo verticalmente, come accade invece per le pompe a diffusione d'olio. L'avviamento e lo spegnimento della pompa è generalmente molto rapido; inoltre una pompa turbomolecolare è caratterizzata da una altissima velocità di aspirazione (dell'ordine di 200 litri al secondo).

Un inconveniente di queste pompe, oltre all'alta rumorosità e l'alto costo, è l'inquinamento, seppur leggero, dovuto ai vapori d'olio, in quanto esse necessitano di lubrificazione. Quindi, se come pompa secondaria usiamo una turbomolecolare, dobbiamo interporre fra questa e la camera di lavoro, diciamo a metà del canale di trasporto, una criotrappola (LN_2), dove far condensare gli efflussi d'olio. La criotrappola è costituita, almeno in linea generale, da un contenitore di acciaio inox a doppia parete, nella cui intercapedine viene versato azoto liquido: sulla parete interna così raffreddata, che è a contatto con l'atmosfera da evacuare, le impurezze organiche provenienti dalla pompa turbomolecolare condensano, senza arrivare quindi ad inquinare l'eterostruttura in fase di crescita. In figura 140 mostriamo una foto ed un disegno di una generica pompa turbomolecolare; si può notare l'impilamento coassiale alternato fra rotori, muniti di pale aventi una certa inclinazione, e statori, muniti di pale aventi un'inclinazione opposta.

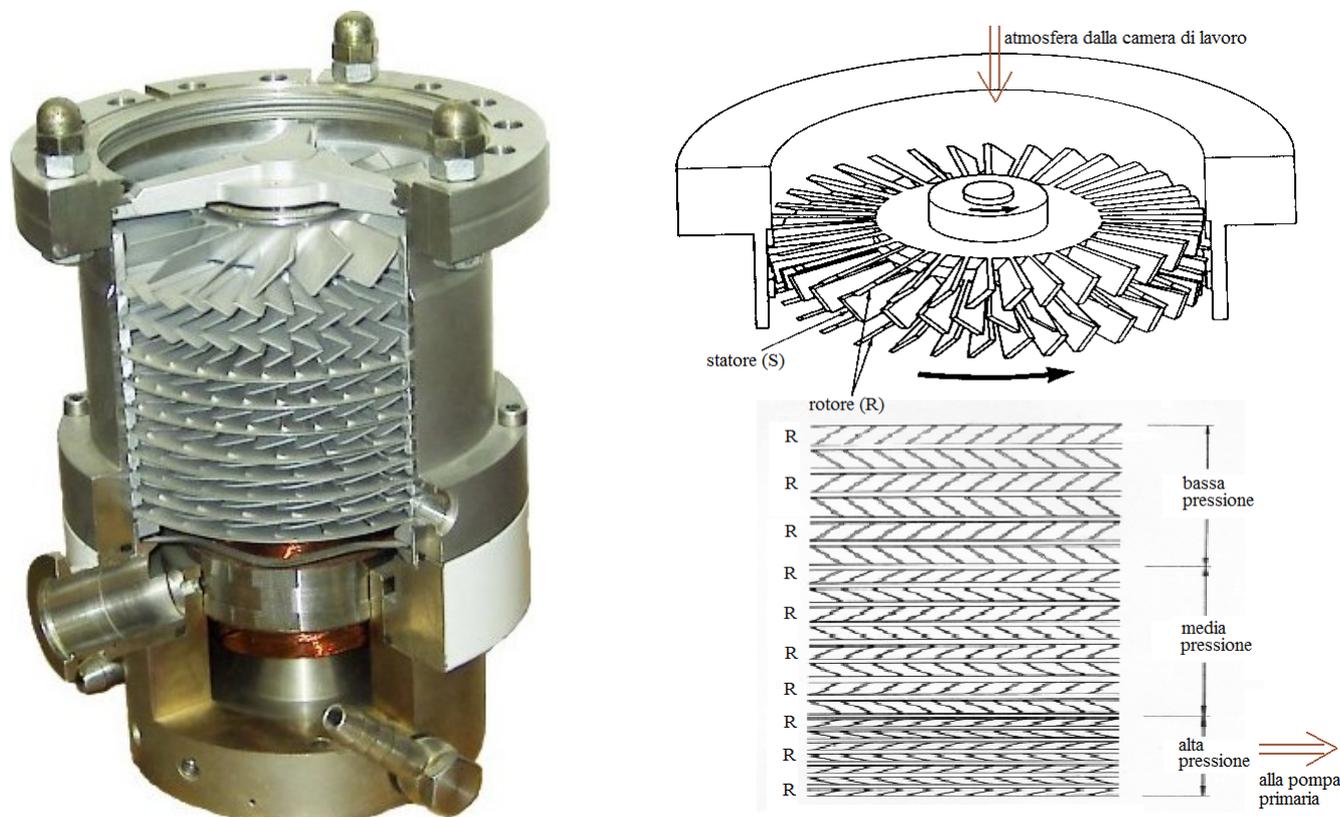


Fig. 140

Foto e schema di una pompa turbomolecolare utilizzabile come pompa secondaria, all'interno del sistema di vuoto assemblato ad un reattore MBE, al posto di una pompa a diffusione di vapori d'olio, tipica di un sistema di vuoto assemblato ad un reattore VPE.

La pompa a sublimazione (o pompa "a getter") consta fondamentalmente di una sorgente di materiale solido altamente reattivo nei confronti dei gas atmosferici (O_2 , H_2 , N_2 , CO_2 , CO , ecc...). Tale sorgente, ad esempio di titanio Ti, viene foggata a forma di filamento, in cui è presente anche una certa quantità di molibdeno Mo, per motivi di lavorazione del filamento stesso (più precisamente il filamento è formato per l'85% da Ti e per il restante 15% da Mo). Gli estremi del filamento sono collegati ad un alimentatore di tensione, il quale produce una corrente elettrica che attraversa il sottile filamento. Questo si surriscalda per effetto Joule fino a raggiungere una temperatura di circa $1500\text{ }^\circ\text{C}$, dopo di ch  una parte del titanio sublima e si condensa sulle pareti interne

della cella cilindrica che contiene il filamento di Ti/Mo. Queste pareti possono essere raffreddate con acqua o con azoto liquido. L'evacuazione dei gas presenti in atmosfera dal reattore MBE è legata proprio al chemi-adsorbimento dei gas (O_2 , H_2 , N_2 , CO_2 , CO , ecc...) sul film di Ti altamente reattivo. I gas atmosferici, quando incidono sul film, reagiscono con questo, formando composti stabili; ad esempio per le molecole di azoto, provenienti dalla camera di lavoro, vale la seguente reazione: $N_2 + 2Ti \rightarrow 2TiN$. La capacità di evacuazione dipende dal coefficiente di adesione di un gas sul film di titanio: H_2 , N_2 , O_2 , CO_2 , e CO sono evacuati con velocità discretamente alte (10 litri/scm², se le pareti sono raffreddate con azoto liquido, da 3 a 10 litri/scm² con raffreddamento ad acqua). I gas nobili e gli idrocarburi non interagiscono con il film di titanio. Le pompe a sublimazione di titanio vengono solitamente impiegate per abbassare ulteriormente la pressione in camere di lavoro già sotto un vuoto medio-alto, raggiungendo valori finali di pressione di circa 10^{-9} - 10^{-10} Torr. In tal modo l'ambiente che ospita l'epitassia MBE viene efficacemente svuotato.

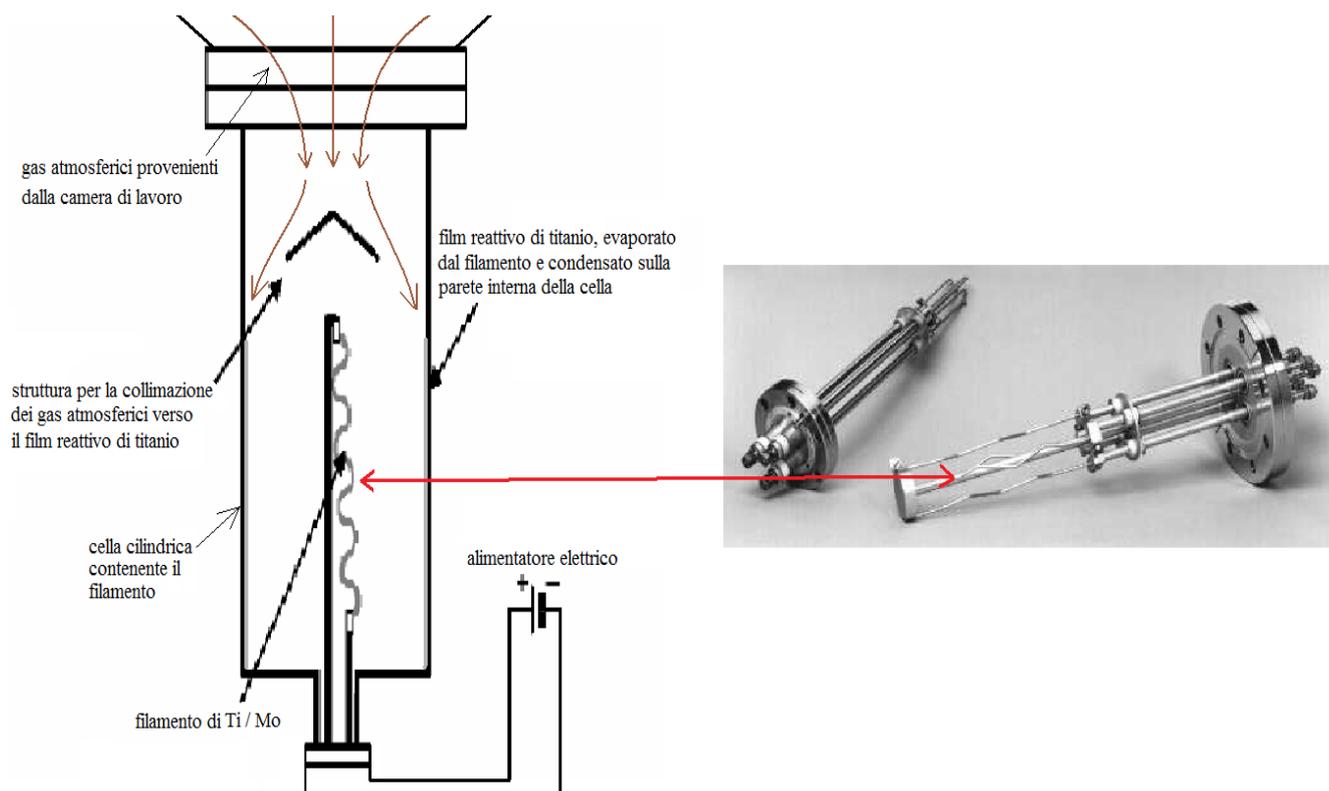


Fig. 141

Schema e foto di una pompa a sublimazione di titanio utilizzabile come pompa secondaria, all'interno del sistema di vuoto assemblato ad un reattore MBE, al posto di una pompa a diffusione di vapori d'olio, tipica di un sistema di vuoto assemblato ad un reattore VPE.

Il trattamento di pulizia della superficie di deposizione consiste, inizialmente, nell'eliminare depositi di ossido superficiale ("l'ossido nativo" ad esempio, che si forma sulla superficie del silicio) o altre impurità assorbite nel substrato mediante un riscaldamento ad elevata temperatura (1000 - 1250 °C), ottenuto tramite una corrente di gas idrogeno. Successivamente si esegue un processo di "sputtering", cioè la rimozione di impurezze superficiali per mezzo di un fascio di particelle energetiche inerti, ad esempio ioni argon Ar^+ , che erodono per collisione la superficie del substrato; il processo di "sputtering" è seguito da una ricottura ("annealing") del materiale superficiale, al fine di riordinarne la struttura reticolare.

Durante il processo di crescita, gli elementi effusivi (gallio, fosforo, arsenico, indio, ecc...) sono presenti in forma solida e vengono fatti sublimare nei forni ad efflusso (celle di effusione), per creare i fasci di molecole che si andranno a depositare (condensare) sulla superficie del substrato, il quale è posizionato sul suscettore rotante all'interno della camera di lavoro. Uno dei modelli di celle più utilizzati, soprattutto per l'evaporazione di Si, Ge, Sb e As, è il tipo Knudsen. Le sorgenti solide degli elementi da sublimare vengono riscaldate per induzione elettromagnetica da delle bobine alimentate a radiofrequenza, raffreddate ad acqua ed avvolte intorno alle celle di effusione. Al variare della radiofrequenza e/o dell'ampiezza della corrente elettrica di alimentazione in una

bobina, è possibile variare la temperatura della sorgente solida posizionata all'interno del forno ad efflusso avvolto dalla bobina stessa; il controllo della sublimazione della sorgente solida è quindi legato al controllo dell'alimentazione elettrica della bobina che avvolge la cella ad efflusso. Pertanto il controllo delle frazioni molari dei vari elementi, che costituiscono ciascun layer di una eterostruttura in fase di crescita, si riduce al controllo dell'entità (frequenza/ampiezza) delle alimentazioni elettriche delle varie bobine avvolte intorno alle celle del reattore. I criopannelli sono le pareti di contenitori riempiti con azoto liquido e servono ad isolare termicamente ciascun forno ad efflusso dagli altri. Infatti senza l'utilizzo dei criopannelli potrebbe accadere che alimentando una bobina con una corrente la cui ampiezza viene fatta crescere nel tempo, al fine di incrementare il tasso di sublimazione dell'elemento contenuto dentro la cella avvolta dalla bobina, cioè di aumentare la presenza dell'elemento nello strato in fase di crescita, aumenta anche la temperatura delle altre celle (una sorta di accoppiamento termico), cioè aumenta in modo spurio, non voluto, il tasso di sublimazione degli altri elementi e conseguentemente anche la presenza degli elementi nello strato epitassiale crescente. In altri termini senza i criopannelli si perderebbe il controllo, mediante il pilotaggio delle correnti nelle bobine, delle frazioni molari degli elementi sublimati all'interno dell'eterostruttura in fase di crescita. Un criopannello posto fra due celle impedisce che il calore crescente, proveniente da una delle due, si propaghi fino a raggiungere l'altra. I portelli di otturazione delle celle di effusione contribuiscono, attraverso l'azione meccanica di apertura/chiusura, al controllo dell'uscita dei fasci molecolari dai forni. All'interno della camera di lavoro di alcuni apparati MBE è presente un'ulteriore sorgente di evaporazione, costituita da un cannone ad elettroni; ad esempio per produrre elevati flussi di evaporazione di Si, oppure per l'evaporazione e la deposizione di film metallici di contatto, è spesso utilizzato il cannone TEMESCAL modello SFIH - 270 - 2 con relativo alimentatore da 10 kW TELEMAR modello TT - 10.

La crescita degli strati è costantemente monitorata da due sistemi di caratterizzazione "in situ" che possono lavorare uno in sostituzione dell'altro oppure contemporaneamente, nel caso in cui si desideri una maggiore accuratezza ed affidabilità dell'azione di monitoraggio. Sono presenti infatti sia uno spettrometro di massa, che rivela la composizione massiva del materiale sublimato che giunge nei suoi pressi, e in base a ciò determina, istante per istante, la composizione molare del layer in crescita, sia un sistema di rivelazione del fenomeno della diffrazione di elettroni, noto come apparato RHEED - "reflection high energy electron diffraction" - tecnica basata sulla diffrazione di elettroni altamente energetici riflessi. Esiste anche una tecnica alternativa a quest'ultima, chiamata tecnica RLEED - "reflection low energy electron diffraction" - tecnica basata sulla diffrazione di elettroni a bassa energia riflessi. Un cannone elettronico RHEED (oppure uno RLEED) emette costantemente un fascio elettronico molto energetico (o poco energetico); gli elettroni energetici, la cui energia cinetica può arrivare fino a 10 keV, nel caso in cui si utilizzi un cannone RHEED CAMECA modello 410, incidono sul layer epitassiale in fase di crescita e vengono riflessi e diffratti verso uno schermo fluorescente. Al variare, nel tempo, delle frazioni molari degli elementi (P, As, Ga, In, ecc...) costituenti il layer superficiale in crescita, colpito dal fascio elettronico, variano anche sia la modalità di riflessione che la figura di diffrazione visualizzata sullo schermo fluorescente. Tale figura fluorescente è rivelata e convertita in una figura di scintillazione da un cristallo scintillatore e successivamente rivelata da un fotomoltiplicatore, a semiconduttore oppure a tubo. In alternativa sia la figura di diffrazione RHEED che quella RLEED possono essere acquisite e digitalizzate tramite una telecamera a CCDs ("charge coupled devices" - dispositivi ad accoppiamento di carica), in bianco e nero, collegata ad una scheda "frame grabber" montata su un computer. Monitorando pertanto l'evoluzione temporale dell'immagine fluorescente è possibile acquisire informazioni in tempo reale circa lo stato di crescita dell'eterostruttura. Lo spettrometro di massa è utilizzabile anche come analizzatore di gas residui, ovvero come analizzatore dei gas "desorbiti" (emessi) dallo strato epitassiale in fase di crescita. Questa terza modalità di caratterizzazione "in situ" del layer in crescita è nota come RGA - "residual gas analyzer": un esempio di sistema RGA, abbastanza utilizzato, è l'analizzatore di gas residui a singolo filtro da 0 a 100 a.m.u. (unità di massa atomica) FISONs modello Quartz 100. Esiste un'ulteriore tecnica di monitoraggio e caratterizzazione del layer epitassiale in fase di crescita, chiamata XPS - "X - ray photoelectron spectroscopy" - spettroscopia di fotoelettroni estratti da raggi X, nella quale una sorgente di raggi X, ad esempio il cui target di collisione elettronica, all'interno del tubo a vuoto, è formato da una piastra anodica realizzata in alluminio $Al_{K\alpha}$ oppure in magnesio $Mg_{K\alpha}$, invia i fotoni γ_x verso il cristallo epitassiale in fase di crescita. Il pedice K_{α} indica che quando il target è irradiato dagli elettroni emessi dal catodo, gli elettroni atomici dei due materiali, con cui è possibile realizzare le placche di collisione all'interno del tubo a vuoto, subiscono una particolare transizione di eccitazione da un livello energetico inferiore E_m ad uno superiore E_n , per poi subire, trascorso un tempo di vita medio τ_n caratteristico di E_n , una successiva diseccitazione verso un livello $E_m^* < E_n$, dove in generale $E_m^* > E_m$. Questa particolare transizione, chiamata "transizione elettronica K_{α} ", si può verificare sia nell'alluminio che nel magnesio. In seguito alla diseccitazione $E_n \rightarrow E_m^*$, la piastra di magnesio emette fotoni γ_x di energia $h\nu_x = E_n - E_m^* \approx 1253$ eV, mentre la piastra di alluminio emette fotoni γ_x di energia $h\nu_x = E_n - E_m^* \approx 1486$ eV. Alcuni modelli di tubi a vuoto per la produzione dei raggi X, adatti alla caratterizzazione di alcuni tipici layers MBE, sono costruiti dalla VG Instruments. Pertanto il layer MBE viene irradiato dai raggi X, alcuni elettroni atomici del cristallo epitassiale vengono estratti, per effetto fotoelettrico, dal layer e raggiungono un rivelatore di elettroni (molto utilizzato è il modello RIBER MAC II). Ipotizziamo di avere un campione di substrato realizzato con un solo

elemento, come ad esempio P, As, Ga, In, As, Si, Ge, Sb ecc..., ovvero con un elemento che comunemente è possibile trovare in un tipico layer MBE. Se irradiamo il campione con raggi X è possibile costruire un grafico come quelli semiquantitativi riportati nella parte centrale di figura 142. Ciascun grafico semiquantitativo riporta in ascissa l'energia totale E , rispetto al livello del vuoto (il livello ad energia nulla), dei gusci indicizzati dal numero quantico principale $n = 1, 2, 3 \dots$, e degli orbitali $s, p, d \dots$, caratteristici delle strutture energetiche degli elementi prima citati, e in ordinata il numero $W_{\gamma_e}(E)$ dei fotoelettroni al secondo, giunti al rivelatore di elettroni, estratti da un substrato realizzato con un certo elemento, a causa di una radiazione X incidente, ad esempio avente energia $h\nu_x \approx 1253 \text{ eV}$ (cioè la sorgente dei raggi X è realizzata in magnesio $Mg_{K\alpha}$), fotoelettroni provenienti da ciascun livello energetico E del materiale a cui appartenevano prima dell'estrazione. I picchi si riferiscono agli orbitali $ns, np, nd \dots$ ai quali appartengono i fotoelettroni γ_e prima della loro estrazione ad opera dei fotoni X incidenti sul materiale. Pertanto conosciamo, data una particolare radiazione X incidente ($h\nu[Mg_{K\alpha}] = 1253 \text{ eV}$) ed un particolare angolo di incidenza dei raggi X sul substrato, lo spettro di emissione fotoelettrica $W_{\gamma_e}(E)_k$ di tutti gli elementi $k = P, As, Ga, In, As, Si, Ge, Sb \dots$ rintracciabili in una comune eterostruttura realizzabile con la tecnica MBE. Quindi istante per istante è possibile valutare l'incidenza degli eventi, rilevati dal rivelatore di elettroni, per ciascun orbitale (... 3s, 3p, 3d, 4s, 4p, 4d, 4f...), cioè la frequenza di incidenza, sul rivelatore, dei fotoelettroni provenienti da ciascun orbitale. Pertanto si può stimare, istante per istante, mediante un sistema computerizzato collegato a valle del rivelatore di elettroni, il grafico $W_{\gamma_e}(E)_{MBE}$ del numero di fotoelettroni al secondo, giunti al rivelatore, estratti dallo strato superficiale dell'eterostruttura MBE, cioè estratti dal layer superficiale in fase di crescita epitassiale, ad opera della radiazione X incidente sul layer; i fotoelettroni estratti provengono dai livelli energetici E riportati in ascissa. Dallo spettro di emissione fotoelettrica $W_{\gamma_e}(E)_{MBE}$ si può stimare, istante per istante, il contributo di ciascun grafico $W_{\gamma_e}(E)_k$ a quello risultante $W_{\gamma_e}(E)_{MBE}$, cioè si può stimare la frazione molare di ciascun elemento ($k = P, As, Ga \dots$) presente nel layer MBE in fase di crescita.

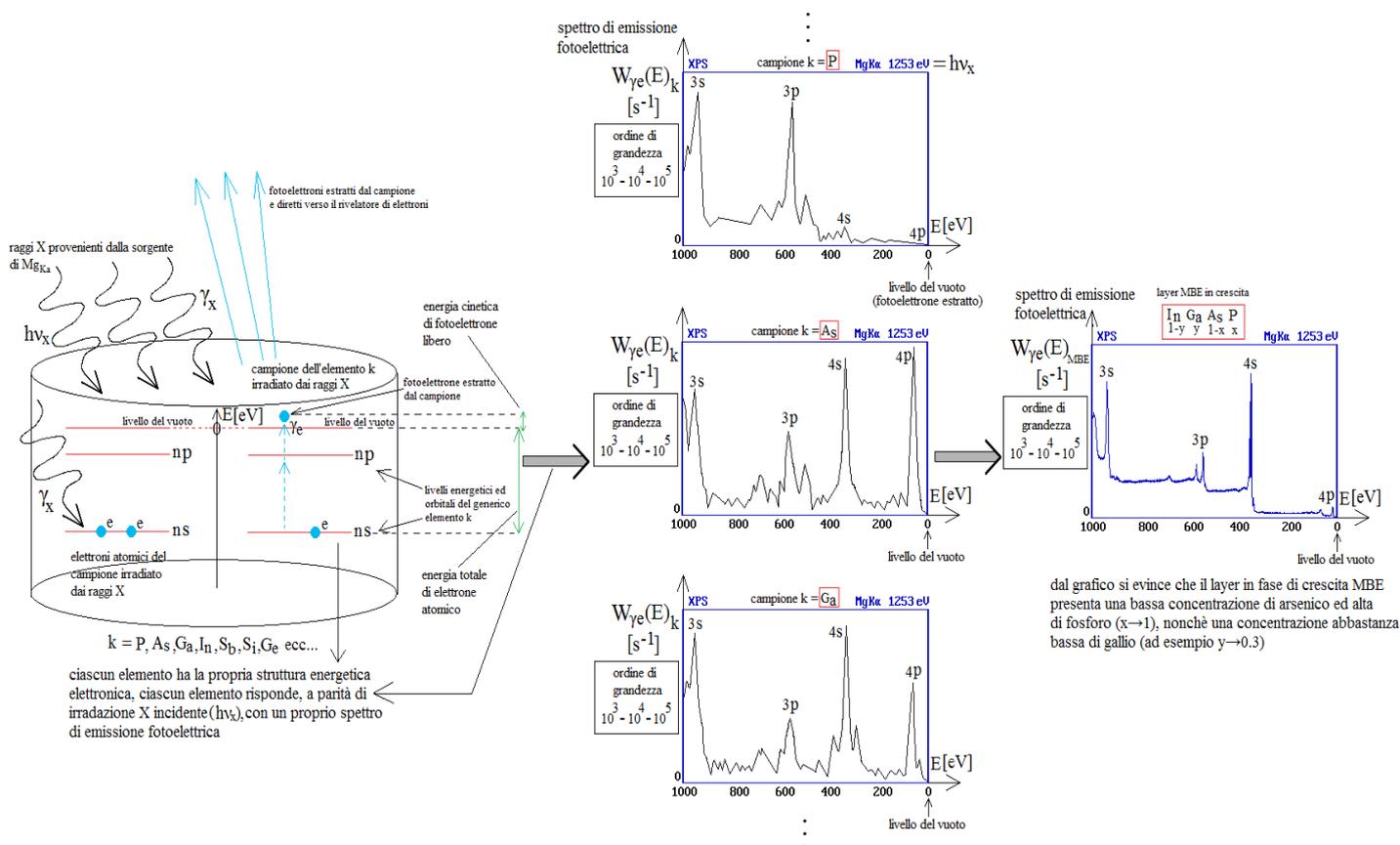


Fig. 142

Schema riassuntivo della tecnica XPS - "X - ray photoelectron spectroscopy" - spettroscopia di fotoelettroni estratti da raggi X, che è utilizzabile per il monitoraggio e la caratterizzazione del layer epitassiale MBE in fase di crescita.

Un motore elettrico a velocità di rotazione variabile, assemblato ad un'asta e ad un sistema di palettature, consente la rotazione del suscettore e quindi del substrato, favorendo la deposizione uniforme dei fasci molecolari.

I fasci molecolari si disporranno seguendo una struttura cristallina simile a quella del substrato, creando strati di purezza molto elevata. È da puntualizzare che sulla superficie su cui incidono i fasci molecolari non si verificano reazioni chimiche, contrariamente a quanto accade in un processo di crescita CVD, ma soltanto una deposizione degli atomi e delle molecole incidenti, provenienti dalle celle ad efflusso. Un vantaggio derivante da questa tecnica di crescita epitassiale è dato dalla bassa temperatura di lavoro (400 – 800 °C), che evita fenomeni di retrodiffusione ed autodrogaggio. È possibile realizzare strati di semiconduttori composti e profili di drogaggio molto complessi, in quanto si dispone di un totale controllo dei fasci molecolari e la crescita epitassiale MBE è molto più lenta rispetto a quella della crescita epitassiale da fase vapore VPE: una velocità tipica della MBE è di circa $4 \div 7$ nm/minuto, che come ordine di grandezza è circa un millesimo della velocità tipica della VPE, che è dell'ordine di $1 \mu\text{m}/\text{minuto}$. I fenomeni fisici che regolano l'epitassia da fasci molecolari (in realtà qualunque tipo di epitassia) sono "l'adsorbimento", il "desorbimento" e la "diffusione". L'adsorbimento è un fenomeno in cui molecole o atomi, ad esempio molecole appartenenti ai fasci uscenti dai forni ad efflusso, oppure queste e le molecole del substrato, instaurano un'interazione del tipo dipolo elettrico – dipolo elettrico (forze di Van der Waals). Il processo inverso, mediante il quale si ha la liberazione di un materiale adsorbito, è il desorbimento, che produce scarti gassosi che devono essere evacuati dalla camera di lavoro. La diffusione è una sorta di "migrazione delle molecole", le quali si spostano sulla superficie del substrato, vi scivolano sopra, finché non trovano una posizione energeticamente favorevole per il loro adsorbimento.

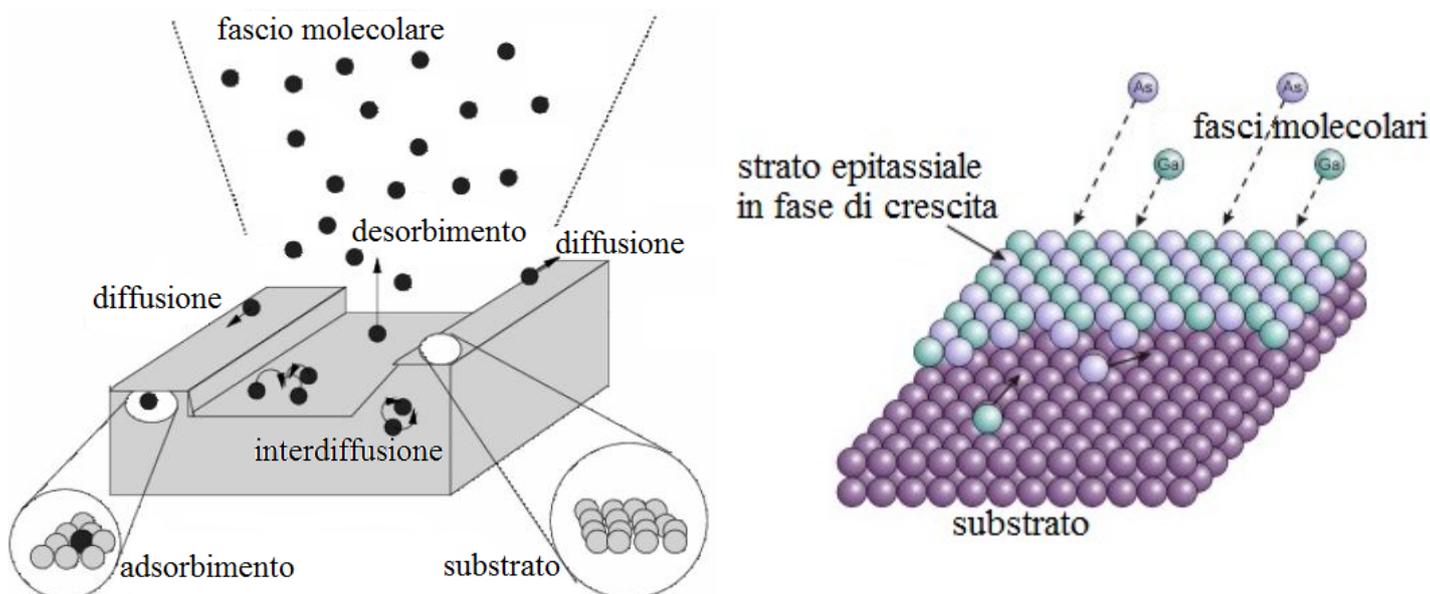


Fig. 143

Rappresentazione qualitativa della deposizione dei fasci molecolari, durante un processo MBE, su un substrato massivo e dei tre fenomeni fisici più importanti che avvengono, nei pressi della superficie del substrato, durante un processo di crescita epitassiale: l'adsorbimento, il desorbimento e la diffusione.

In figura 144 è riportata la foto di una moderna apparecchiatura utilizzabile per la crescita da fasci molecolari di strati epitassiali nanometrici.

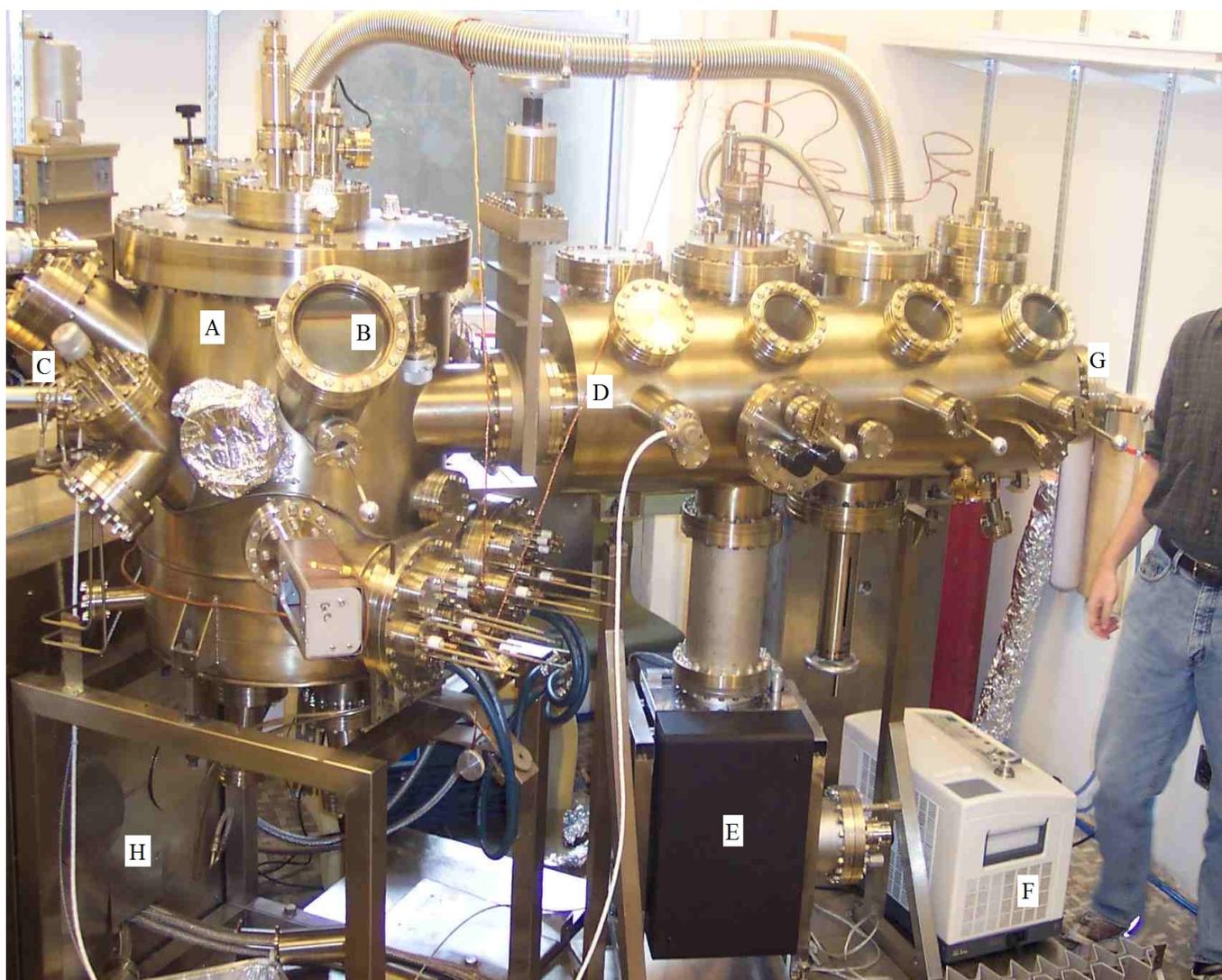


Fig. 144

Foto di un apparato MBE

Si notino:

- A) il cilindro verticale che costituisce la camera di lavoro depressurizzata, cioè il reattore MBE, dove avviene la crescita controllata degli strati epitassiali
- B) la finestra ottica di zaffiro
- C) le celle ad effusione per la sublimazione delle sorgenti solide collocate al loro interno
- D) la struttura cilindrica orizzontale che mette in comunicazione la camera di lavoro con il sistema di vuoto. Tale struttura costituisce il canale di trasporto per l'atmosfera, da evacuare, proveniente dal reattore MBE.
- E) una pompa a sublimazione per la creazione del vuoto secondario
- F) una pompa meccanica rotativa per la creazione del vuoto primario
- G) valvola a chiusura pneumatica
- H) una pompa a sublimazione per la creazione del vuoto secondario

B) Una possibile tecnica per la rivelazione di luce appartenente al lontano infrarosso

Per rivelare radiazioni nel lontano infrarosso, ad esempio nell'ambito di applicazioni termografiche, ovvero per $\lambda > 2.5 \mu\text{m}$, è possibile utilizzare anche omogiunzioni di silicio, sotto opportune condizioni di drogaggio e termostatazione. Consideriamo il lato n di un fotodiiodo di Si: le impurezze droganti individuano, come è noto, degli stati donatori localizzati poco al di sotto di E_C : nel caso del fosforo questa distanza energetica ΔE è pari a circa 0.045 eV, per valori di drogaggio tali da non considerare il Si n degenere, cioè per $N_D < 10^{20} \text{ cm}^{-3}$. Pertanto un eventuale fotone incidente con energia superiore a 0.045 eV, ovvero $\lambda < 27.5 \mu\text{m}$, può cedere l'energia ad un elettrone, reso disponibile da un atomo drogante, permettendogli di superare la barriera di potenziale ΔE e contribuire alla fotoconduzione. Tuttavia il SNR di un dispositivo così fatto sarebbe inaccettabile, dal momento che a temperature ordinarie ($50 \text{ K} < T < 500 \text{ K}$, ovvero nell'intervallo delle temperature di svuotamento) gli stati donatori sono tutti ionizzati ($n \approx N_D$), per cui la fotocorrente sarebbe completamente coperta da una corrente di portatori prodotti termicamente (grande rumore di fondo). Per risolvere questo problema è necessario tenere il fotodiiodo di Si a temperature estremamente basse ($T < 50 \text{ K}$, presso le quali $n = N_D^+ \ll N_D$), mediante l'utilizzo di sistemi di refrigerazione ad elio oppure ad azoto liquidi (con notevole levitazione dei costi). In questo modo la ionizzazione termica degli stati donatori è fortemente inibita (la Fermi - Dirac, sul livello energetico di tali stati, è prossima ad 1), per cui gli elettroni possono essere promossi in BC solo attingendo energia da eventuali fotoni incidenti.

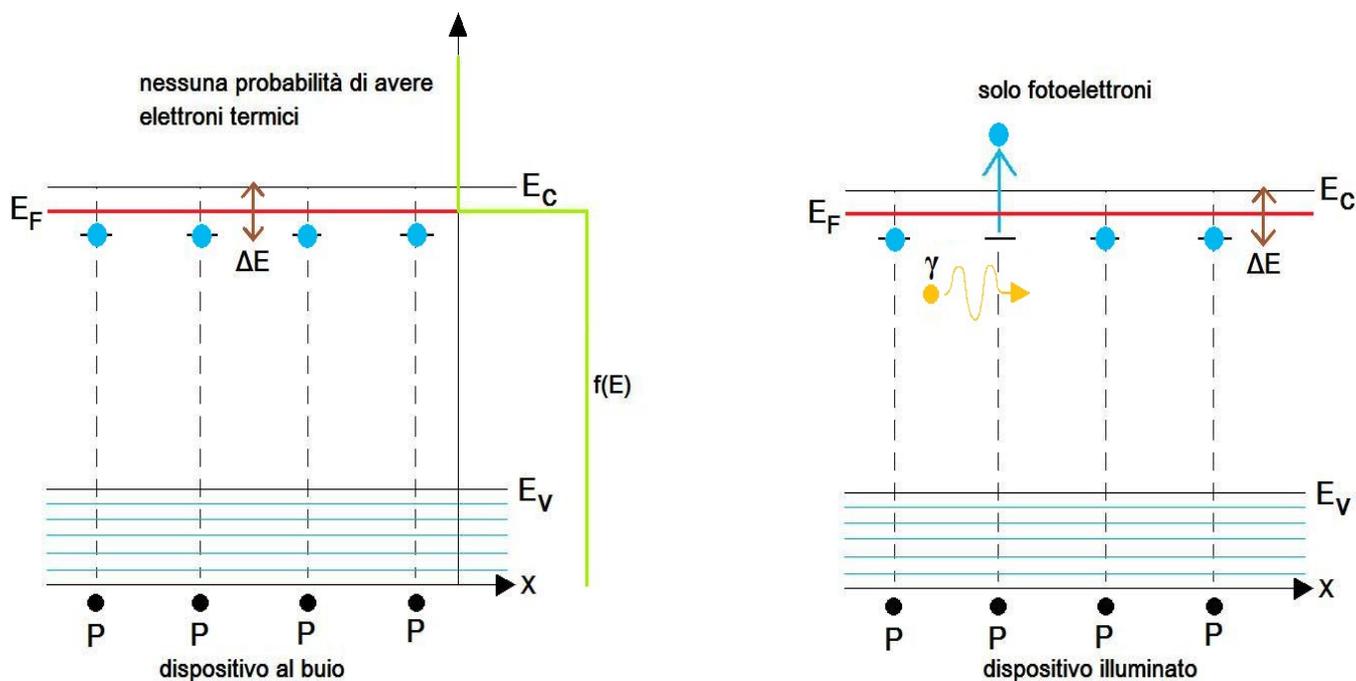


Fig. 145

A sinistra riportiamo il diagramma a bande di energia, in funzione della posizione, del lato n di un fotodiiodo al silicio (l'impurezza drogante del V° gruppo è il fosforo), non irradiato dalla luce. La bassa temperatura a cui è tenuto il fotodiiodo inibisce la ionizzazione termica degli stati donatori localizzati presso i droganti n; ciò consente di limitare fortemente il rumore termico. A destra riportiamo lo stesso diagramma, relativo al fotodiiodo sottoposto ad illuminazione: gli elettroni promossi in BC sono solo quelli dovuti all'assorbimento dei fotoni incidenti.

C) Una possibile scelta di semiconduttori III - V per la rivelazione di frequenze ottiche comprese fra il visibile e l'ultravioletto

Come già accennato, sia nel paragrafo 5.9.2 che nel glossario, esistono semiconduttori binari III - V a base di azoto, a gap diretto, chiamati "semiconduttori nitridici" ("nitrides semiconductors"), che possono essere cresciuti, mediante la tecnologia MBE, sopra substrati di zaffiro oppure di carburo di silicio SiC. Questi semiconduttori nitridici, la cui struttura reticolare segue la geometria HCP - "Hexagonal Close Pack" (si veda il glossario), sono: InN, GaN, AlN, BN. La costante reticolare a di questi semiconduttori, che coincide con il lato dell'esagono regolare che costituisce la base della struttura cristallina HCP, è abbastanza piccola ($2.55 \div 3.5 \text{ \AA}$) rispetto al passo reticolare a dei semiconduttori semplici e composti (III - V e II - VI) strutturati secondo geometrie di tipo diamante e zincoblenda. In figura 146 è mostrato il grafico della compatibilità reticolare fra i quattro semiconduttori nitridici; è riportata anche la costante reticolare del SiC, spesso utilizzato come substrato.

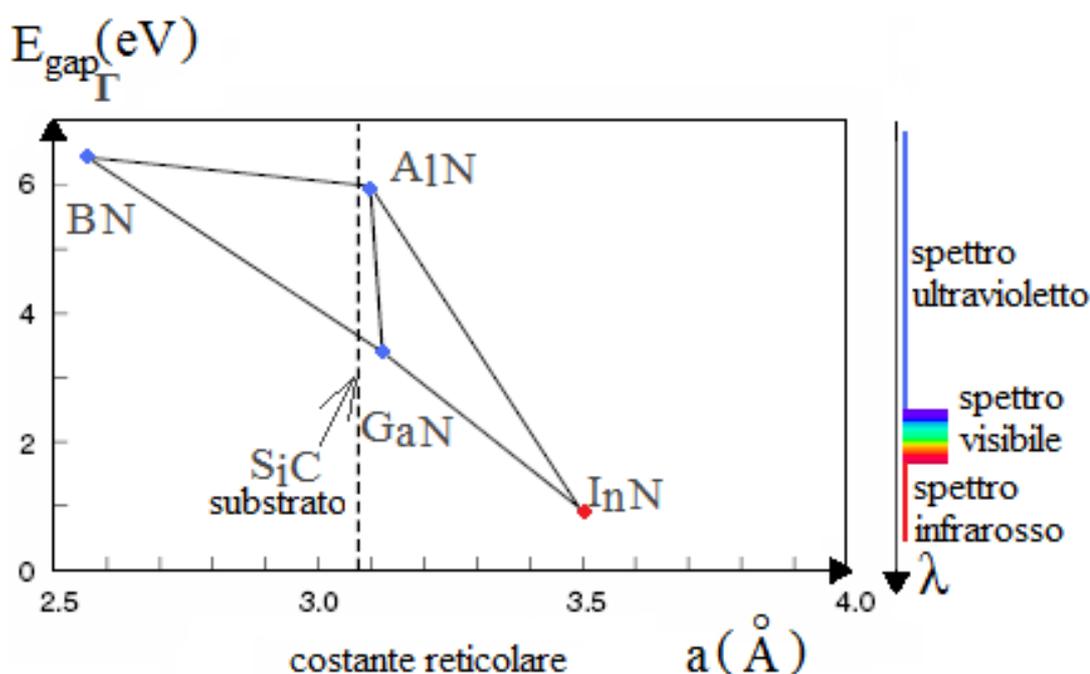


Fig. 146

Grafico della compatibilità reticolare fra i quattro semiconduttori nitridici; è riportata anche la costante reticolare del carburo di silicio SiC, spesso utilizzato come substrato. L'utilizzo di GaN, AlN e BN per la rivelazione di radiazioni ultraviolette risulta evidente dal valore dell'energia di gap diretto $E_{\text{gap}\Gamma}$ dei suddetti semiconduttori.

L'unico elemento di incertezza, nel grafico mostrato in figura 146, è rappresentato dal InN: utilizzando il metodo TBM ("Tight Binding Method") per il calcolo delle bande di energia e la costruzione dei grafici $E - k$, risulta che il InN abbia un gap diretto $E_{\text{gap}\Gamma}$ pari a circa $1.5 \div 1.6 \text{ eV}$; tuttavia misure sperimentali condotte a partire dal 2001 hanno evidenziato un gap diretto pari a circa 0.9 eV . Il grafico di figura 146 riporta, come $E_{\text{gap}\Gamma}$ del InN, proprio quest'ultimo risultato. Tuttavia, nonostante la discrepanza fra dato teorico e dato sperimentale, è abbastanza consolidato il risultato secondo cui il InN è un semiconduttore a gap diretto utilizzabile per la fotorivelazione di luce appartenente all'intervallo delle lunghezze d'onda NIR, cioè del vicino infrarosso. Non si esclude, pertanto, una possibile futura applicazione nel settore delle telecomunicazioni su fibra ottica. Nel caso degli altri tre semiconduttori nitridici, la teoria relativa al TBM concorda con le evidenze sperimentali: il gap del GaN è pari a circa 3.2 eV , il gap del AlN è pari a circa 6 eV , quello del BN a circa 6.4 eV .

Il fatto che i tre semiconduttori nitridici aventi una $\lambda_{\text{gap}\Gamma}$ appartenente allo spettro ultravioletto, ovvero InN, GaN, AlN, possiedono una costante reticolare a decisamente più piccola ($2.55 \div 3.1 \text{ \AA}$) di quella dei comuni semiconduttori strutturati secondo la geometria a diamante o a zincoblenda ($4.43 \div 6.1 \text{ \AA}$), e congiuntamente

una gap diretto abbastanza più grande (maggiore di 3 eV, contro i $1 \div 1.5 \div 2 \div 2.3$ eV dei comuni semiconduttori), può essere compreso anche utilizzando il semplice modello analitico, del potenziale elettrostatico periodico “di background” all’interno di un cristallo, rappresentato dal ben noto modello a buche e barriere rettangolari di Kronig – Penney. Come è noto, applicando il modello nel caso di un generico cristallo ideale unidimensionale, cresciuto ad esempio lungo l’asse z, otteniamo una relazione del tipo:

$$-1 < F(E) = \cos(k_z a) < 1$$

dove E è l’energia totale di un elettrone atomico del cristallo, F è una funzione dell’energia E, a è la costante reticolare, cioè la somma fra la larghezza b di una barriera e la larghezza l della buca adiacente, e k_z il vettore d’onda di un elettrone atomico lungo l’asse z del cristallo, considerato perfettamente periodico e privo di difetti e impurezze (vale pertanto il teorema di Bloch). La relazione suddetta mette, come noto, in evidenza che esistono alcune energie elettroniche totali E, tali per cui F(E) è compresa fra -1 ed 1, che possono essere assunte dagli elettroni del cristallo, ed altre energie elettroniche totali E, tali per cui la relazione non è verificata (cioè F(E) ha modulo maggiore di 1), che non possono essere assunte dagli elettroni. In altre parole esistono bande di energia permesse agli elettroni atomici e gaps di energia non occupabili da questi, che non hanno stati permessi agli elettroni. L’espressione completa di F(E) è una funzione molto complicata di parametri geometrici del cristallo unidimensionale, come la larghezza b della barriera e la larghezza l della buca ($a = b + l$), e di parametri elettrostatici, come l’energia elettronica totale E dell’onda di Bloch associata all’elettrone preso in esame e l’altezza V_0 della barriera.

$$F(E) = \frac{\beta^2 - \alpha^2}{2\alpha\beta} \sinh(\beta b) \sin(\alpha l) + \cosh(\beta b) \cos(\alpha l) = \cos[k_z(l + b)] = \cos(k_z a)$$

con:

$$\alpha = \sqrt{\frac{2m_0 E}{\hbar^2}}$$

$$\beta = \sqrt{\frac{2m_0(V_0 - E)}{\hbar^2}}$$

Possiamo semplificare l’espressione di F(E) se supponiamo che la barriera sia infinitamente stretta, cioè $b \rightarrow 0$, per cui la costante reticolare a coincide, di fatto, con la larghezza l della buca ($a \approx l$), e infinitamente alta, ovvero $V_0 \rightarrow \infty$, e che il prodotto fra la larghezza della barriera e la sua altezza V_0 (bV_0) assuma un valore costante finito. Sotto queste ipotesi semplificative è possibile esprimere la F(E) nel modo seguente:

$$F(\alpha) = \frac{\theta}{\alpha} \sin(\alpha a) + \cos(\alpha a) = \cos(k_z a)$$

con:

$$\alpha \ll \beta$$

$$\beta \approx \sqrt{\frac{2m_0 V_0}{\hbar^2}}$$

$$\theta = \frac{\beta^2 b a}{2}$$

$$\beta b \propto \beta \sqrt{V_0} \rightarrow 0$$

$$\sin(\beta b) \rightarrow \beta b, \cosh(\beta b) \rightarrow 1, l = a - b \approx a$$

m_0 è la massa dell’elettrone libero, α e β si misurano in cm^{-1} , θ è un numero puro, così come F(α). In figura 147 è riportato il grafico della funzione F(α) per $\theta = 3/2$, confrontato con un generico diagramma a bande di energia E

– k , molto qualitativo, relativo al generico cristallo di semiconduttore nitridico (GaN, AlN, BN) unidimensionale, cresciuto lungo l'asse z , analizzato mediante il modello analitico di Kronig – Penney.

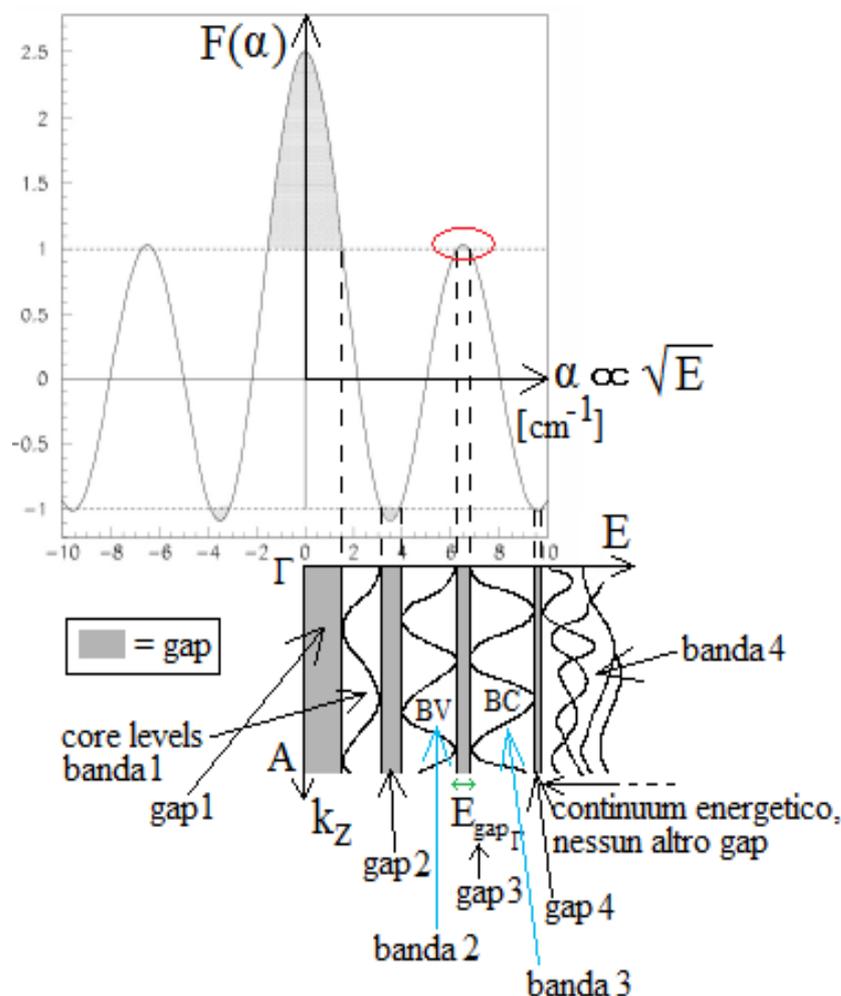


Fig. 147

Grafico della funzione $F(\alpha)$ per $\theta = 3/2$, confrontato con un generico diagramma a bande di energia $E - k$, molto qualitativo, relativo al generico cristallo di semiconduttore nitridico (GaN, AlN, BN) unidimensionale, cresciuto lungo l'asse z , analizzato mediante il modello analitico di Kronig – Penney.

Presso le energie elettroniche totali E , in corrispondenza delle quali il modulo della funzione $F(\alpha)$ è maggiore di 1, il diagramma a bande $E - k$ non presenta stati elettronici permessi, cioè occupabili dagli elettroni del cristallo, ovvero il diagramma $E - k$ presenta gaps di energia; presso le energie E , in corrispondenza delle quali il modulo della funzione $F(\alpha)$ è minore di 1, il diagramma a bande $E - k$ presenta stati elettronici permessi, ovvero presenta bande consentite di energia. I semiconduttori nitridici GaN, AlN e BN sono caratterizzati, come già osservato, da una costante reticolare a inferiore rispetto a quella tipica dei semiconduttori comuni, come quelli semplici Si e Ge e come quelli composti III – V e II – VI, pertanto la frequenza dell'oscillazione, lungo l'asse α ($\alpha \propto \sqrt{E}$), della funzione $F(\alpha)$ di uno dei tre semiconduttori nitridici è minore della frequenza dell'oscillazione, sempre lungo l'asse α , della funzione $F(\alpha)$ dei semiconduttori comuni. Pertanto la funzione $F(\alpha)$ dei semiconduttori nitridici ha un andamento più liscio, lungo α , della funzione $F(\alpha)$ relativa ai semiconduttori comuni: ne consegue che i gaps e le bande del diagramma $E - k$ di uno dei tre semiconduttori nitridici sono più larghi dei gaps e delle bande del diagramma $E - k$ dei semiconduttori comuni. In particolare nei semiconduttori nitridici la curva di $F(\alpha)$ evidenziata dall'ellisse rossa, in figura 147, è più liscia rispetto alla stessa curva di $F(\alpha)$ dei semiconduttori comuni, pertanto la larghezza $E_{\text{gap}\Gamma}$ del gap diretto, avente valenza ottica, indicato nel disegno con "gap 3", è maggiore della larghezza E_{gap} del gap, di valenza ottica, caratteristico dei semiconduttori comuni. Anche le

disequaglianze di $E_{\text{gap}\Gamma}$ fra i quattro semiconduttori nitridici mostrati in figura 146 sono spiegabili, utilizzando questo modello, a partire dalle disequaglianze della costante reticolare a .

In figura 148 mostriamo i diagrammi a bande di energia $E - k$ di InN, GaN e AlN.

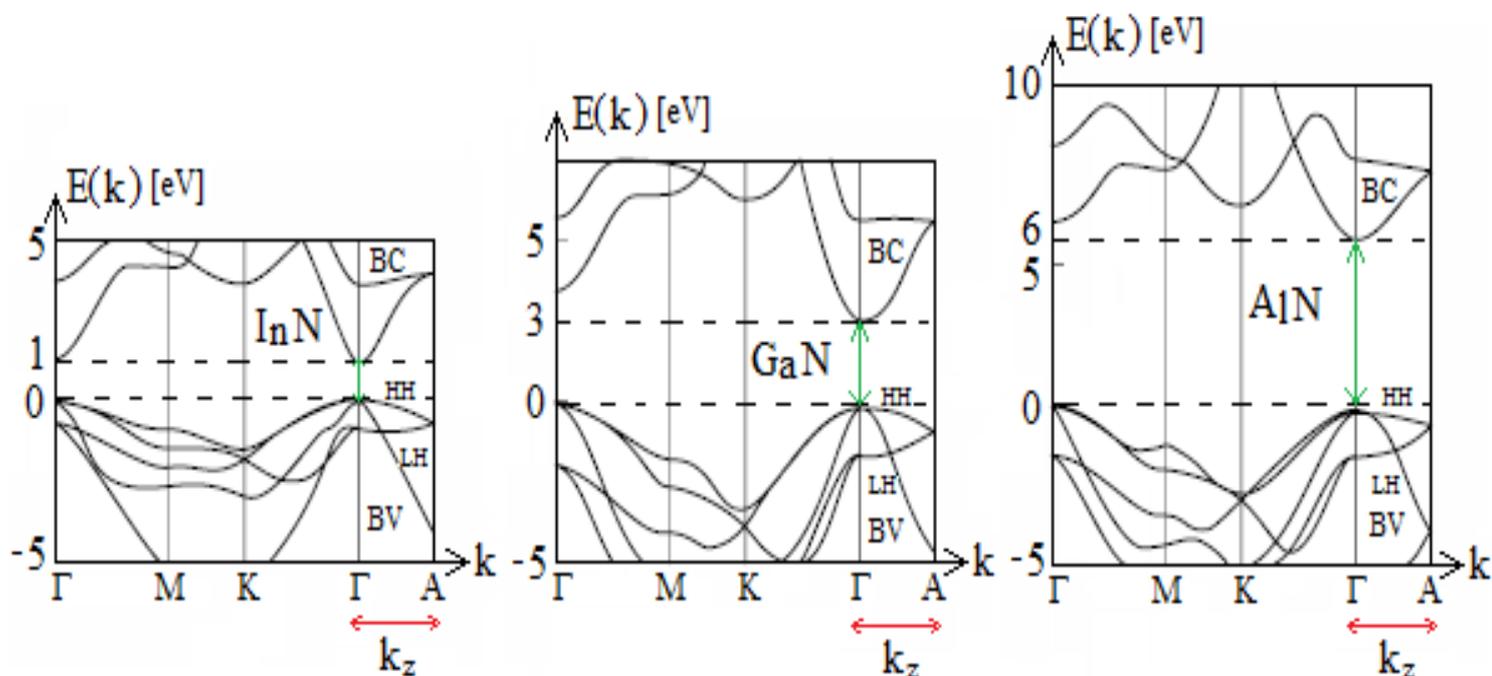


Fig. 148

Diagrammi a bande di energia $E - k$ dei tre semiconduttori nitridici a gap diretto InN, GaN e AlN

Dai diagrammi $E - k$ di figura 148 si evince che i semiconduttori binari GaN ed AlN, così come quello ternario $\text{Al}_{(1-x)}\text{Ga}_{(x)}\text{N}$, il quale può essere facilmente cresciuto epitassialmente in virtù del buon matching reticolare fra i due semiconduttori binari, sono adatti alla rivelazione di luce più energetica rispetto a quella infrarossa e visibile, ossia di luce blu - violetta e ultravioletta, a causa del grande gap diretto $E_{\text{gap}\Gamma}$. Il semiconduttore nitridico ternario $\text{Al}_{(1-x)}\text{Ga}_{(x)}\text{N}$ potrebbe rivestire un ruolo importante nella realizzazione di dispositivi elettronici di lettura di informazione binaria (0,1) altamente concentrata sulla superficie di un supporto di memoria, come ad esempio un compact disk (CD). Uno degli esempi più attuali di una possibile applicazione dei suddetti semiconduttori nitridici è la tecnologia "Blue - Ray" di archiviazione e lettura digitali di informazione, destinata a sostituire quelle basate sul DVD e sul CD - ROM. Il $\text{Al}_{(1-x)}\text{Ga}_{(x)}\text{N}$ potrebbe essere impiegato per la costruzione del core di un LASER (nel core si crea l'inversione positiva di popolazione dei portatori di carica, la ricombinazione dei quali produce l'emissione dei fotoni) che, a causa del pompaggio elettrico a cui è sottoposto, cioè della corrente di iniezione I_{bias} di portatori di carica minoritari dovuta alla polarizzazione diretta della giunzione, emette una fascio molto coerente e direttivo di luce blu - violetta o ultravioletta. Tale luce, di lunghezza d'onda λ_{inc} molto piccola ($\lambda_{\text{inc}} < 400 \text{ nm}$), è indirizzata sulla superficie di un CD, sulla quale l'informazione da leggere è memorizzata attraverso una successione di grumi di materiale, di protuberanze, di "dots", lunghi e corti (un formato rappresenta 1, l'altro 0), che rappresentano il codice binario con cui l'informazione stessa è codificata. La luce incide sulla superficie del CD e viene scatterata indietro, dal momento che la sezione media dei dots superficiali ha dimensione simile alla lunghezza d'onda λ_{inc} della luce incidente. Pertanto, essendo la luce incidente molto energetica, ovvero di lunghezza d'onda λ_{inc} corta, anche la densità di informazione memorizzata sul CD, il quale è chiamato "Blue Disk", può essere più alta di quella memorizzata, mediante dots, sulla superficie di un tradizionale DVD o CD - ROM; quest'ultimi infatti sono dischi sui quali incide luce infrarossa, pertanto la dimensione dei dots deve essere sufficientemente grande (maggiore di 700 nm), affinché la luce di grande lunghezza d'onda possa essere opportunamente scatterata indietro. La luce blu, proveniente dal LASER il cui core è realizzato in $\text{Al}_{(1-x)}\text{Ga}_{(x)}\text{N}$, una volta scatterata sulla superficie del Blue Disk, torna indietro, pertanto può essere rivelata da un fotodiode, polarizzato in inversa, il cui volume attivo è realizzato sempre in $\text{Al}_{(1-x)}\text{Ga}_{(x)}\text{N}$, adatto cioè all'assorbimento della luce blu - violetta o ultravioletta. Il fotodiode converte l'informazione ottica,

portata dalla luce blu scatterata sui dots della superficie del Blue Disk, in un'informazione elettrica interpretabile da un'elettronica di front - end, che in questo caso consiste in un lettore Blue - Ray.

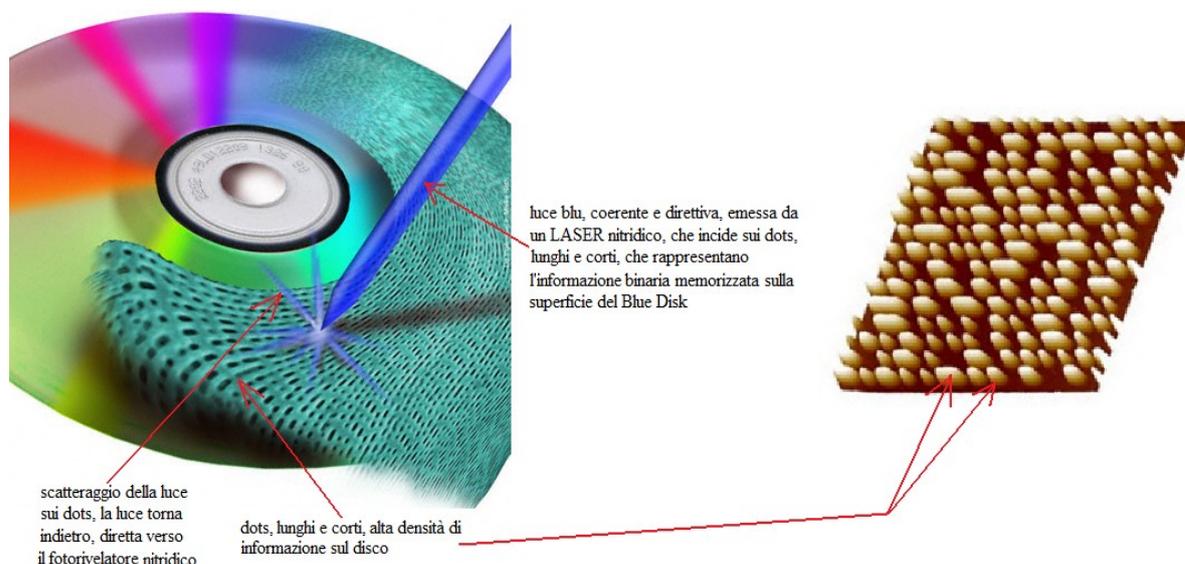


Fig. 149

Possibile applicazione dei semiconduttori nitridici GaN e AlN in un sistema elettronico di lettura dell'informazione binaria memorizzata, mediante "dots" lunghi e corti, sulla superficie di un Blue Disk. I semiconduttori nitridici potrebbero essere impiegati sia per la realizzazione del LASER che proietta il fascio di luce blu sulla superficie del CD, sia per la realizzazione del fotorivelatore, al quale giunge la luce blu scatterata sui dots, la quale è portatrice dell'informazione binaria rappresentata dai dots stessi.

D) Schema di principio dell'elettronica di lettura per la TOF PET

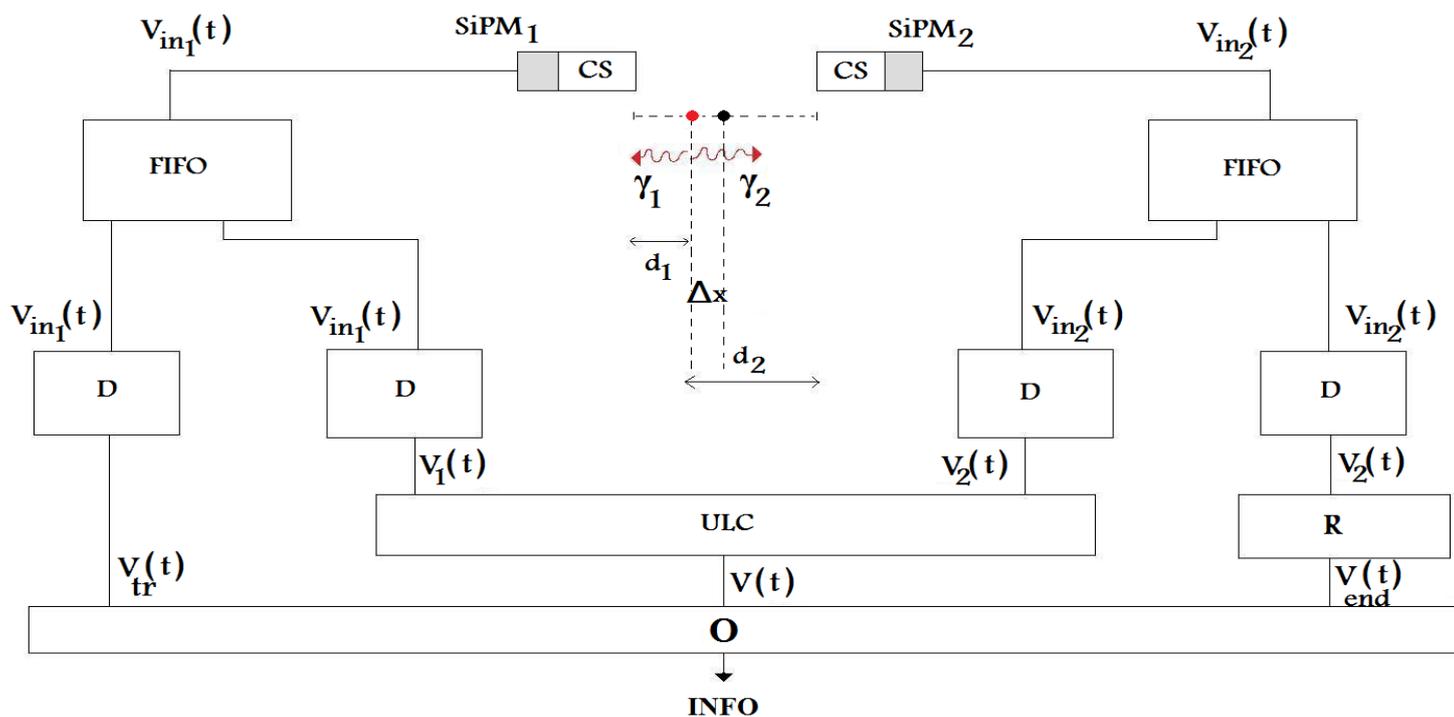


Fig. 150

Schema a blocchi dell'elettronica di front - end usata per la visualizzazione di immagini, del tessuto sottoposto a mapping, in ambito TOF PET.

Riferiamoci allo schema concettuale a blocchi riportato in figura 150. Supponiamo che il punto nero sia esattamente il centro della LOR (definita nel paragrafo 2.2) relativa ai due fotoni γ_1 e γ_2 , quello rosso il punto in cui avviene l'annichilazione elettrone atomico/positrone, ovvero il punto, con buona approssimazione, in cui si è concentrato l'isotopo tracciante β^+ ; di tale punto vogliamo conoscere la posizione, con la minore incertezza possibile, utilizzando la LOR fra i due fotorivelatori come sistema di riferimento. d_1 è la distanza fra il punto di annichilazione e lo scintillatore LYSO assemblato con il SiPM₁, d_2 è la distanza fra il suddetto punto e lo scintillatore LYSO assemblato con il SiPM₂. Chiamiamo Δx l'offset del punto di annichilazione rispetto al centro della LOR ($\Delta x = (d_2 - d_1)/2$), ovvero l'incertezza relativa alla posizione del punto dal quale partono i due fotoni (due raggi gamma). Nella trattazione che segue trascuriamo i ritardi legati ad entrambi i cristalli scintillatori LYSO, che per semplicità consideriamo identici. Supponiamo che all'istante $t = -d_1/c$ avvenga la produzione di γ_1 e γ_2 , e quindi che all'istante $t = 0$ il fotone γ_1 sia arrivato al LYSO (CS) del SiPM₁. Il LYSO produce dei fotoni di scintillazione (secondo le modalità descritte nel paragrafo 8.9), prevalentemente nello spettro del blu, i quali attivano alcune microcelle del SiPM₁. Quest'ultimo segnala l'evento luminoso di scintillazione, e indirettamente l'evento γ_1 , producendo in uscita un segnale di tensione $V_{in1}(t)$; $V_{in1}(t)$ inizia all'istante $t = 0$, coincidente con l'inizio dell'effetto valanga, raggiunge il suo picco in $t \approx R_s C_D \approx 50$ ps, dopodiché la resistenza di quenching R_Q agisce e smorza la valanga, la quale si estingue con un tempo di caduta dell'ordine di $R_Q C_D \approx 17$ ns.

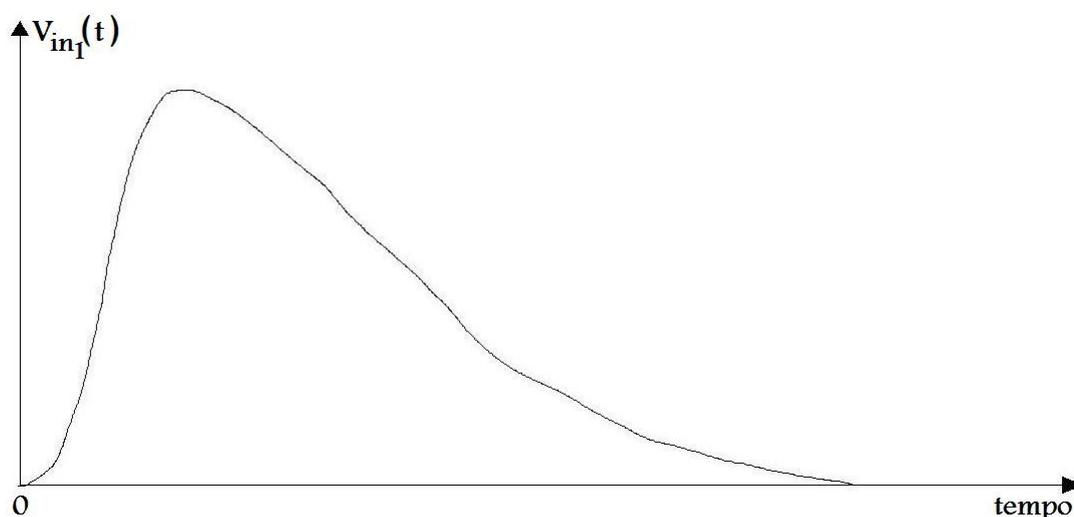


Fig. 151

Grafico temporale, molto qualitativo, del segnale $V_{in1}(t)$ prodotto dal SiPM₁, in risposta all'evento luminoso legato al fotone γ_1 . $t = 0$ è l'istante in cui γ_1 arriva sul cristallo scintillatore (CS) LYSO, e trascurando i ritardi a questo legati, l'istante in cui i fotoni di scintillazione arrivano ad irradiare il SiPM₁. In $t = 0$ inizia a svilupparsi la valanga che costituisce il fotosegnale $V_{in1}(t)$.

Il fotone γ_2 , percorrendo un tragitto più lungo di quello percorso da γ_1 ($d_2 = d_1 + 2\Delta x$), arriva ad incidere sul LYSO del SiPM₂ all'istante $t_2 = 2\Delta x/c$. Il LYSO, supposto identico a quello allineato con il SiPM₁, produce un pacchetto di fotoni di scintillazione uguale al pacchetto che ha irradiato il SiPM₁ in $t = 0$. Quindi, indirettamente, γ_2 innescava una valanga identica a quella innescata dal primo fotone (a meno di fluttuazioni relative alla risposta di scintillazione del cristallo LYSO, alla QE del SiPM, ecc...), essendo γ_1 e γ_2 equienergetici. Dunque la tensione $V_{in2}(t)$ resa disponibile in uscita dal secondo SiPM è qualitativamente uguale a quella precedente, a parte il ritardo t_2 .

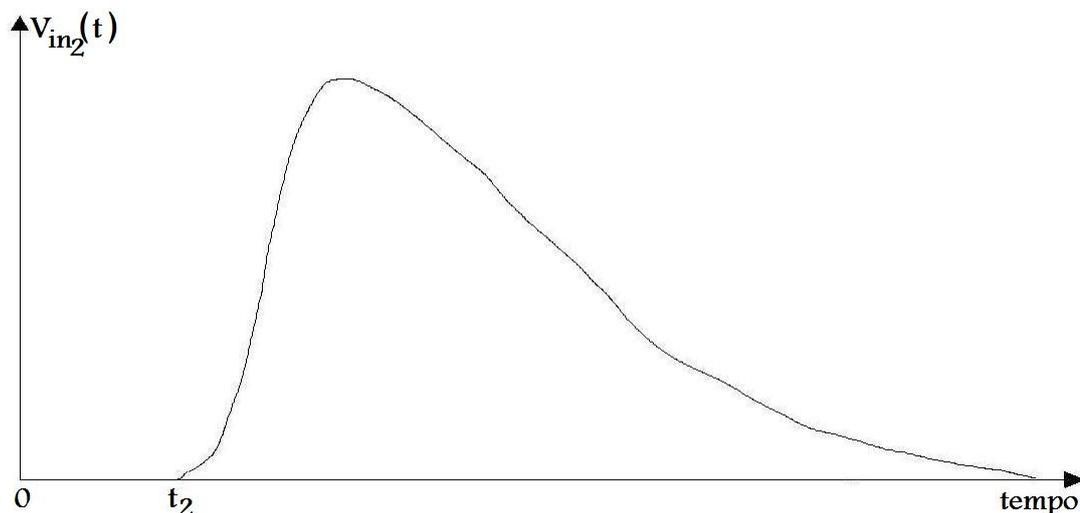


Fig. 152

Grafico temporale, molto qualitativo, del segnale $V_{in2}(t)$ prodotto dal SiPM₂, in risposta all'evento luminoso dovuto al fotone γ_2 . t_2 è l'istante in cui γ_2 arriva sul cristallo scintillatore (CS) LYSO, e trascurando i ritardi a questo legati, l'istante in cui i fotoni di scintillazione arrivano ad irradiare il SiPM₂. In t_2 inizia a svilupparsi la valanga che costituisce il fotosegno $V_{in2}(t)$.

Questi due segnali entrano nei rispettivi blocchi di Fan in - Fan out (FIFO), circuiti che mettono a disposizione in uscita, su più coppie di terminali (due nel nostro caso), la tensione in ingresso, per cui in uscita abbiamo copie multiple di $V_{in1}(t)$ e $V_{in2}(t)$. Entrambe le tensioni vengono inviate sui terminali d'ingresso di un discriminatore di tensione (indicato con D), circuito che per valori di tensione in ingresso V_{in} inferiori ad un valore di soglia V_T ($= V$ di threshold, tipico di ciascun discriminatore), produce in uscita un valore analogico basso interpretabile come uno "zero logico", mentre per valori di V_{in} maggiori di V_T , produce in uscita un valore analogico alto interpretabile come un "uno logico". Il comparatore integrato all'interno di D deve entrare in saturazione per valori di tensione in ingresso tanto più piccoli quanto maggiore è la "sensibilità" energetica che deve caratterizzare la fotorivelazione (si osservi la figura 154). Un modello, ovviamente solo concettuale, di discriminatore di tensione D può essere quello riportato in figura 153.

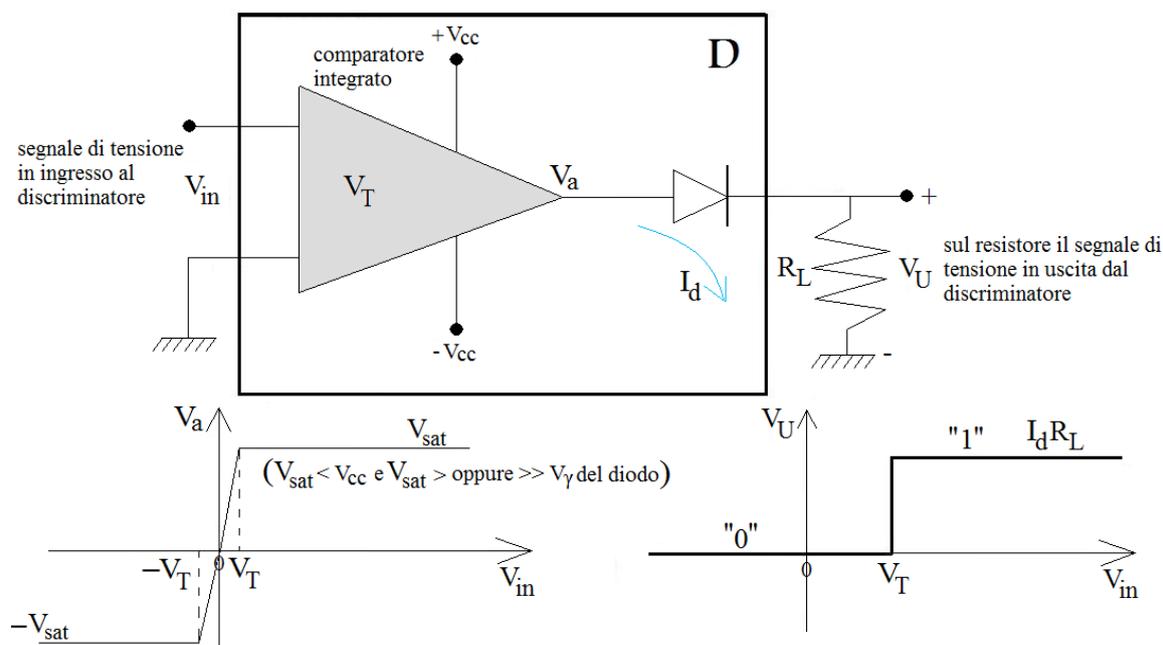


Fig. 153

Rappresentazione concettuale di un possibile modello di discriminatore di tensione (D), utilizzabile all'interno dell'elettronica di front - end per la TOF PET. Si notino la caratteristica ingresso - uscita del comparatore integrato e quella dell'intero discriminatore D.

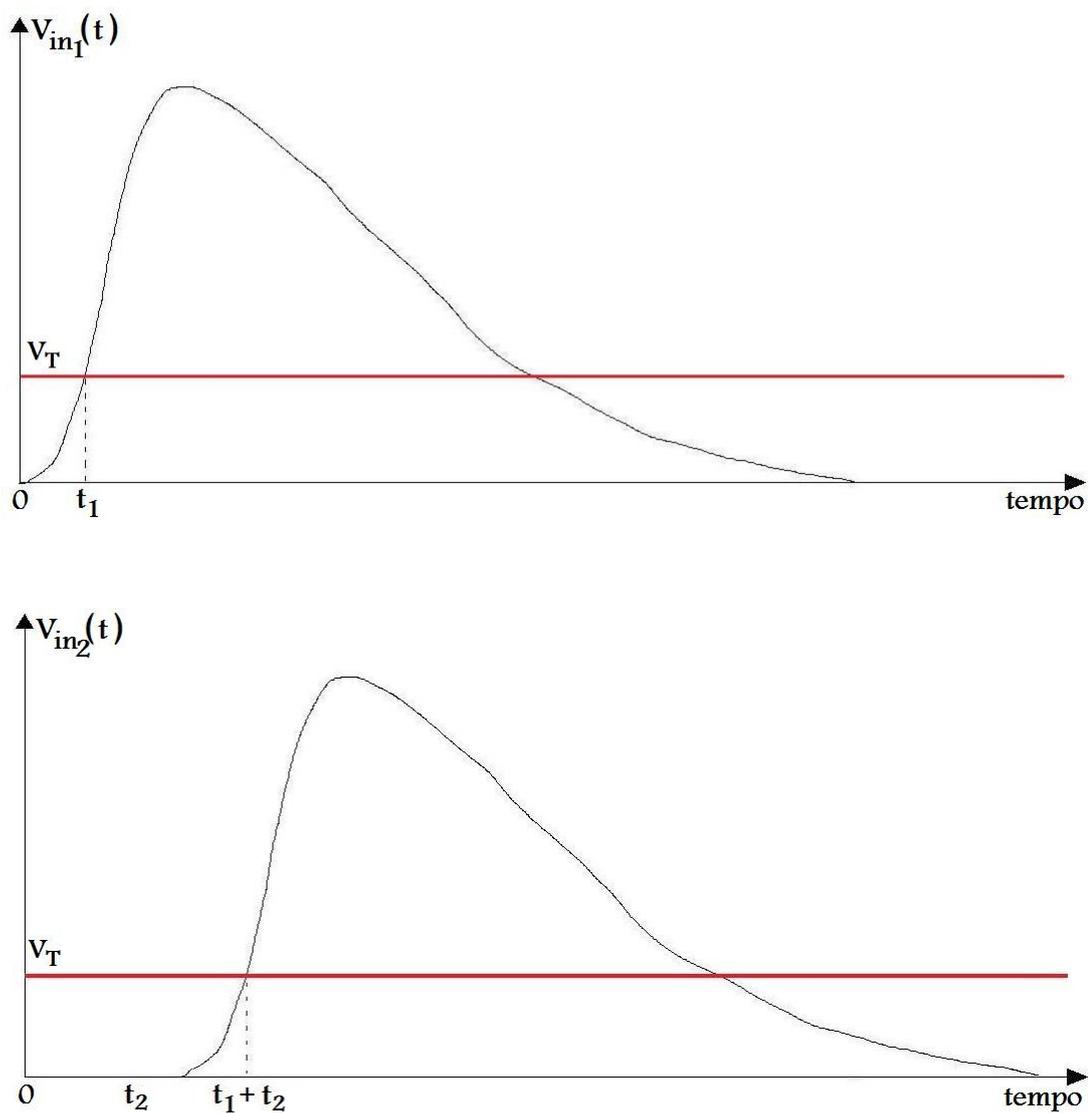


Fig. 154

Grafici temporali, molto qualitativi, dei fotosegnali $V_{in1}(t)$ e $V_{in2}(t)$ prodotti, rispettivamente, dal SiPM₁ e dal SiPM₂, in risposta agli eventi luminosi γ_1 e γ_2 . $V_{in1}(t)$ e $V_{in2}(t)$ entrano ciascuno in un discriminatore di tensione D, mostrato in figura 153, al fine di essere digitalizzati (si osservi la figura 155).

$V_1(t)$ e $V_2(t)$, riportate di seguito, costituiscono la digitalizzazione, fornita dai discriminatori di tensione D, dei fotosegnali $V_{in1}(t)$ e $V_{in2}(t)$.

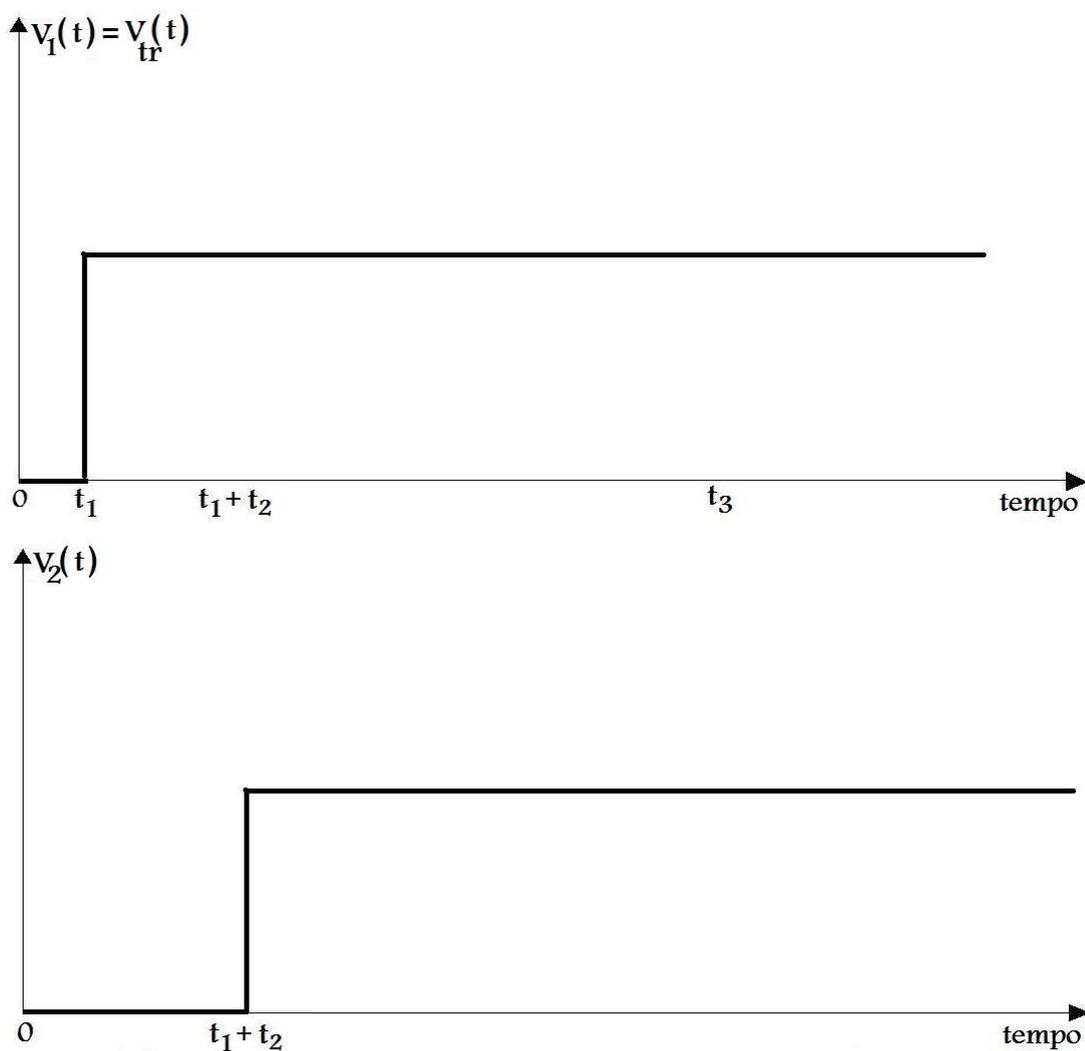


Fig. 155

Grafici temporali delle tensioni $V_1(t)$ e $V_2(t)$, le quali costituiscono la digitalizzazione, fornita dai discriminatori di tensione D, dei fotosegnali $V_{in1}(t)$ e $V_{in2}(t)$.

$V_1(t)$ e $V_2(t)$ entrano nell'unità logica delle coincidenze (ULC): questo circuito, sempre e solo a livello di concetto, può essere visto come una porta XOR, ragion per cui la sua tensione di uscita $V(t)$ sarà quella mostrata in figura 156.

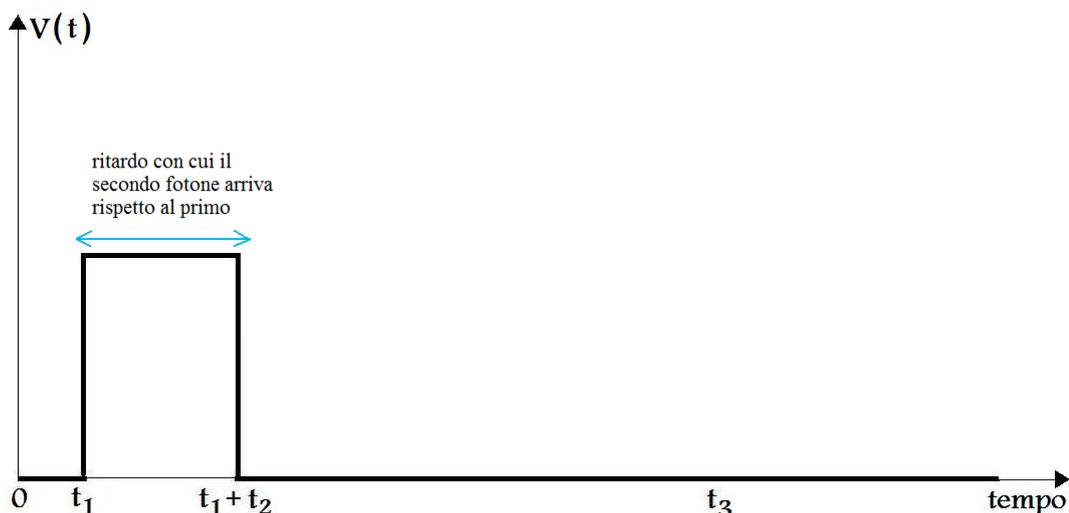


Fig. 156

Grafico temporale della tensione $V(t)$ in uscita dall'unità logica delle coincidenze (ULC): questo blocco può essere visto, a livello concettuale, come un circuito che implementa la funzione XOR.

Il segnale $V_1(t)$ viene anche utilizzato come "trigger" dell'oscilloscopio (O): infatti deve, all'istante t_1 , dare il via allo strumento atto a visualizzare il segnale $V(t)$, mentre la tensione $V_2(t)$ viene portata anche in ingresso ad un ritardatore (R), la cui uscita $V_{\text{end}}(t)$ è la $V_2(t)$ ritardata di un tempo T maggiore del tempo che impiega l'oscilloscopio ad elaborare l'ingresso $V(t)$ prima di poter rendere disponibile in uscita un'informazione INFO stabile. In tal modo $V_2(t)$ funge, sempre per l'oscilloscopio, da segnale di stop. Adesso che t_2 è nota, conosciamo anche $\Delta x (= (c/2)t_2)$, per cui conosciamo, approssimativamente, la posizione del punto di annichilazione lungo la LOR.

Analizziamo ora cosa accade nell'eventualità in cui i due fotoni non abbiano la stessa energia. Supponiamo che γ_1 e γ_2 vengano generati esattamente al centro della LOR, per cui arrivano nello stesso istante ad incidere sui due cristalli scintillatori LYSO: se γ_1 è più energetico di γ_2 , la valanga che interessa il SiPM₁ consta di una maggiore quantità di fotocariche rispetto alla valanga che interessa il SiPM₂, dato che l'energia conferita al SiPM₁ dal LYSO₁ è maggiore dell'energia conferita al SiPM₂ dal LYSO₂. I due segnali elettrici sono simultanei ma $V_{\text{in1}}(t)$ è maggiore di $V_{\text{in2}}(t)$.

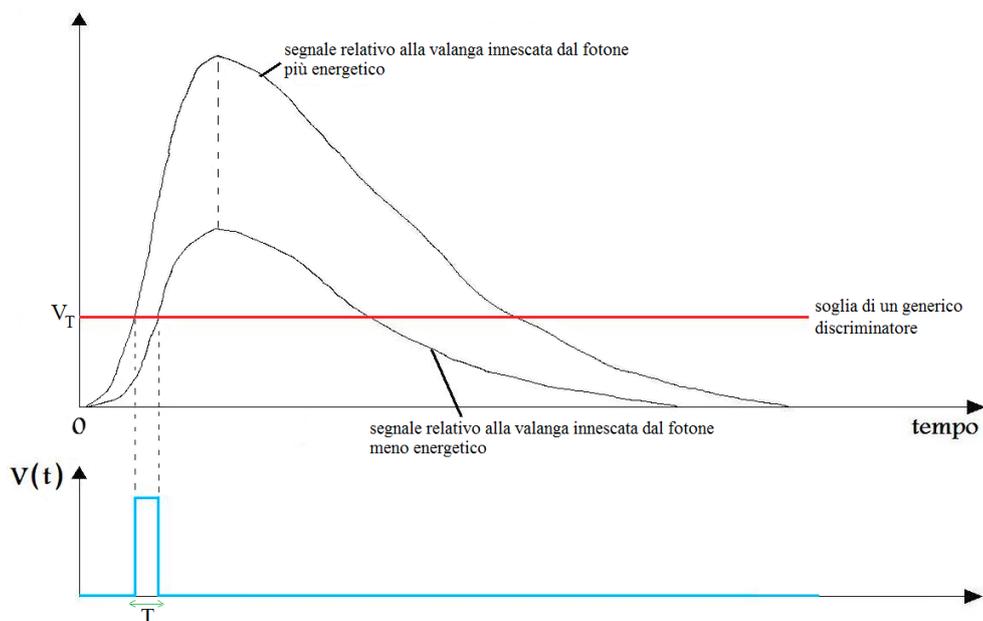


Fig. 157

Rappresentazione dell'errore di walk che può interessare la stima della posizione del punto di annichilazione, nell'eventualità in cui i due fotoni siano equienergetici e nel caso in cui utilizziamo dei discriminatori D comuni, all'interno dell'elettronica di front - end.

Dunque l'elettronica di lettura rileverà un ritardo temporale di γ_2 rispetto a γ_1 , anche se tale ritardo non c'è, e di conseguenza la stima del punto di annichilazione risulterà poco accurata. Questo errore, noto come "effetto walk", è dovuto alla differenza temporale fra gli istanti presi i quali i due segnali attraversano la soglia dei discriminatori: il segnale più alto supera per primo la soglia. Per sopperire a ciò è utile usare dei particolari discriminatori di tensione, detti "discriminatori a frazione costante" ("constant fraction discriminator" – CFD). Il principio su cui si basa il loro funzionamento è semplice (ed è indicativo del loro nome): il segnale di uscita (digitale), che un CFD mette a disposizione dell'elettronica che gli sta a valle, passa da 0 a 1 nell'istante in cui il segnale di ingresso raggiunge una frazione prefissata (esempio il 70%) del suo valore di picco. È da notare che il picco di una valanga è raggiunto in un istante che è quasi indipendente dall'intensità della valanga stessa, e dipende quasi unicamente dai parametri R_s e C_D del circuito equivalente della microcella. Pertanto se usiamo due CFD, uno per $V_{in1}(t)$, l'altro per $V_{in2}(t)$, l'errore di walk è fortemente attenuato, e i due fotoni risultano quasi simultanei, come deve essere.

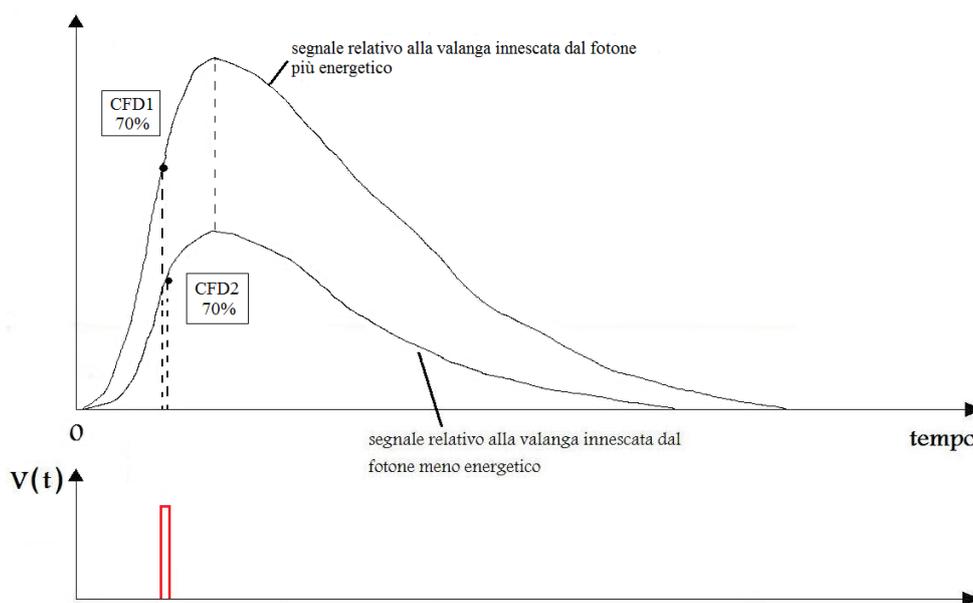


Fig. 158

Rappresentazione qualitativa dell'errore di walk che può interessare la stima della posizione del punto di annichilazione, nell'eventualità in cui i due fotoni siano equienergetici e nel caso in cui utilizziamo dei "discriminatori a frazione costante" ("constant fraction discriminator" – CFD), all'interno dell'elettronica di front – end. Si noti l'attenuazione dell'errore rispetto al caso (figura 157) in cui utilizziamo dei discriminatori D comuni.

Meglio ancora sarebbe usare discriminatori "zero crossing" (ZCD): uno ZCD passa da 0 a 1 nell'istante in cui la derivata del segnale di ingresso assume valore zero. In tal modo l'errore di walk risulta quasi totalmente rigettato e la simultaneità dei fotoni risulta pressoché perfetta. In figura 159 è riportato il ring usualmente adottato per la TOF PET in ambito diagnostico: ciascuna coppia di fotorivelatori, posizionati ai capi di un diametro del ring, può determinare la posizione approssimata di un punto di annichilazione, lungo la LOR costituita proprio dal diametro che congiunge i due SiPMs.

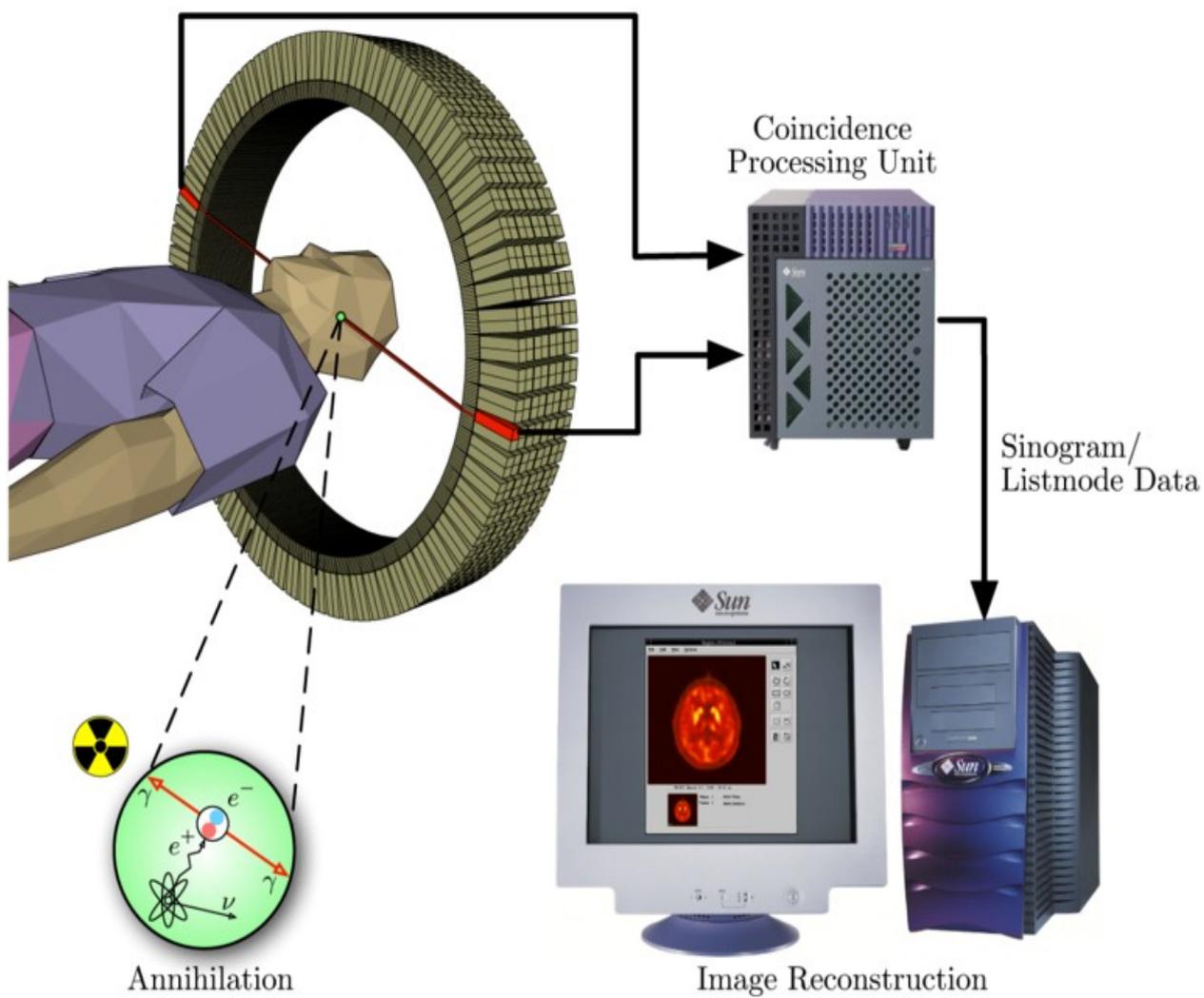


Fig. 159

Rappresentazione del ring usualmente adottato per la TOF PET in ambito diagnostico: ciascuna coppia di fotorivelatori, posizionati ai capi di un diametro del ring, può determinare la posizione approssimata di un punto di annichilazione, lungo la LOR costituita proprio dal diametro che congiunge i due SiPMs.

Bibliografia

Testi:

- 1) B. G. Streetman, S. K. Banerjee: "Solid State Electronic Devices", Prentice – Hall International Editions
- 2) G. Ghione: "Dispositivi per la Microelettronica", McGraw – Hill
- 3) R. S. Muller, T. I. Kamins: "Device Electronics for Integrated Circuits", J. Wiley & Sons
- 4) S. Wolf, R. N. Tauber: "Silicon Processing for the VLSI Era", vol. 1, Lattice Press
- 5) S. M. Sze: "Dispositivi a Semiconduttore", Hoepli
- 6) G. S. May, S. M. Sze: "Fundamentals of semiconductor fabrication", J. Wiley & Sons
- 7) P. S. Kireev: "Semiconductor Physics", Mir Publishers
- 8) J. F. Gibbons: "Semiconductor Electronics", McGraw – Hill
- 9) A. Pirovano, C. Monzio Compagnoni: "Dispositivi optoelettronici integrati", Società Editrice Esculapio
- 10) T. Tamir: "Guided – wave optoelectronics", Springer Verlag
- 11) H. A. Haus: "Waves and fields in optoelectronics", Prentice – Hall International Editions
- 12) D. L. Lee: "Electromagnetic principles of integrated optics", J. Wiley & Sons
- 13) L. D. Hutcheson: "Integrated optical circuits and components", Marcel – Dekker
- 14) H. Nishihara, M. Haruna, T. Suhara: "Optical Integrated Circuits", McGraw – Hill
- 15) C. Kittel: "Introduzione alla Fisica dello Stato Solido", Boringhieri
- 16) B. E. A. Saleh, M. C. Teich: "Fundamental of Photonics", J. Wiley & Sons
- 17) J. Singh: "Optoelectronics: An Introduction to Materials and Devices", McGraw – Hill
- 18) J. Singh: "Semiconductor Devices: An Introduction", McGraw – Hill
- 19) M. Dagenais, R. F. Leheny, J. Crow: "Integrated Optoelectronics", Academic Press
- 20) E. J. Murphy: "Integrated Optical Circuits and Components", Marcel – Dekker
- 21) J. Singh: "Electronic and Optoelectronic Properties of Semiconductor Structures", Cambridge University Press
- 22) K. Konishi, G. Paffuti: "Meccanica Quantistica: nuova introduzione", Edizioni PLUS – Pisa University Press
- 23) C. Hamaguchi: "Basic Semiconductor Physics", Springer

24) B. E. A. Saleh, M. C. Teich: "Fundamentals of photonics", J. Wiley & Sons

25) A. Bogoni, L. Potì: "Elementi di comunicazioni ottiche", Pitagora

26) M. Luise: "Sistemi di trasmissione su fibra ottica", Edizioni ETS, Pisa

27) W. R. Fahrner: "Nanotechnology and Nanoelectronics", Springer

Pubblicazioni:

1) J. Haba: "Status and perspectives of Pixelated Photon Detectors" (2008), Nuclear Instruments and Methods in Physics Research, Section A: Accelerators, Spectrometers, Detectors and Associated Equipment 595 (1), pp. 154 – 160

2) M. Mazzillo, G. Condorelli, D. Sanfilippo, A. Piazza, G. Valvo, B. Carbone, G. Fallica, G. Bonanno, (...): "Silicon photomultipliers for nuclear medical imaging applications" (2008), Proceedings of SPIE – The International Society for Optical Engineering 7003, art. no. 70030I

3) M. Petasecca, B. Alpat, G. Ambrosi, P. Azzarello, R. Battiston, S. Haino, A. Papi, (...): "Thermal and electrical characterization of silicon photomultipliers" (2008), IEEE Transactions on Nuclear Science 55 (3), art. no. 4545198, pp. 1686 – 1690

4) E. Levin, E. Shelegeda, C. Sitarsky, (...): "Advances in the development of solid state photomultipliers for medical imaging" (2008), Progress in Biomedical Optics and Imaging – Proceedings of SPIE 6913, art. no. 69130J

5) G. Laurenti, G. Levi, E. Foschi, C. Guandalini, M. Zuffa, (...): "Time Of Flight Detectors: from phototubes to SiPMs" (2008), Nuclear Instruments and Methods in Physics Research, Section A: Accelerators, Spectrometers, Detectors and Associated Equipment 588 (1-2), pp. 267 – 271

6) G. Llosà, N. Belcari, M. G. Bisogni, A. Del Guerra, G. Collazuol, S. Marcatili, S. Moehrs, C. Piemonte: "Silicon photomultipliers and SiPM matrices as photodetectors in nuclear medicine" (2007), IEEE Nuclear Science Symposium Conference Record 3, art. no. 4436825, pp. 3220 – 3223

7) C. Piemonte: "A new silicon photomultiplier structure for blue light detection" (2005), Nuclear Instruments and Methods in Physics Research, 568 (A), pp. 224 – 232

8) D. Renker: "Photodetectors" (2004), Nuclear Instruments and Methods in Physics Research, 527 (A), pp. 15 – 20

9) V. Golovin, V. Saveliev: "Novel type of avalanche photodetector with Geiger mode operation" (2003), Nuclear Instruments and Methods in Physics Research, 447 (A), pp. 486 – 490

10) R. Battiston, M. G. Bisogni, M. Boscardin, G. Collazuol, F. Corsi, G. F. Dalla Betta, A. Del Guerra, N. Dinu, G. Levi, S. Marcatili, S. Moehrs, C. Marzocca, C. Piemonte, A. Pozza, C. Sbarra, L. Quadrani, G. Llosà: "Novel silicon photomultipliers for PET applications" (2006), IEEE Transactions on Nuclear Science (3), pp. 1875 – 1879

11) G. Collazuol: "Optical Materials" (2007)

12) P. Buzhan, B. Dolgoshein, L. Filatov, A. Ilyin, V. Kaplin, (...): "Silicon photomultiplier and its possible applications" (2003), Nuclear Instruments and Methods in Physics Research, Section A: Accelerators, Spectrometers, Detectors and Associated Equipment 504 (1-3), pp. 48 – 52

- 13) E. A. Georgievskya, S. N. Klemin, L. Filatov, P. Buzhan, B. Dolgoshein, A. Ilyin, (...): "The solid – state silicon photomultiplier for a wide range of applications" (2002), Proceedings of SPIE - The International Society for Optical Engineering 5126, pp. 37 – 42
- 14) G. B. Bondarenko, P. Buzhan, V. Golovin: "Microcell Si avalanche photodiode with limited Geiger mode gain" (2000), IEEE Nuclear Science Symposium and Medical Imaging Conference 1, pp. 7 – 27
- 15) C. Piemonte, R. Battiston, M. Boscardin, G. F. Dalla Betta, A. Del Guerra, N. Dinu, A. Pozza, N. Zorzi: "Characterization of the First Prototypes of Silicon Photomultiplier Fabricated at ITC – irst" (2007), IEEE Transactions on Nuclear Science, vol. 54, art. no. 1
- 16) G. Llosà, N. Belcari, M. G. Bisogni, G. Collazuol, A. Del Guerra, N. Dinu, C. Piemonte, (...): "Evaluation of the First Silicon Photomultiplier Matrices for a Small Animal PET Scanner" (2008), IEEE Transactions on Nuclear Science

Ulteriori fonti:

- 1) A. Diligenti (Prof. Ordinario, Facoltà di Ingegneria, Università di Pisa): "Appunti del corso di Dispositivi e Tecnologie Elettroniche" (a.a. 2006/2007), SEU – Servizio Editoriale Universitario di Pisa
- 2) P. E. Bagnoli (Prof. Ordinario, Facoltà di Ingegneria, Università di Pisa): "Appunti dei corsi di Optoelettronica e Fotonica", Edizioni "il Campano"
- 3) F. Giannetti (Prof. Associato, Facoltà di Ingegneria, Università di Pisa): "Comunicazioni Ottiche", Edizioni "il Campano"
- 4) M. Morganti, F. Cei (Facoltà di Scienze Matematiche Fisiche e Naturali, Università di Pisa): "Appunti di meccanica quantistica", reperibili online sul sito web del Dott. Cei
- 5) S. Bassi: "Tesi di Laurea Specialistica in Fisica Medica sulla Caratterizzazione Spettroscopica di Fotorivelatori al Silicio SiPM" (a.a. 2007/2008)

Figure:

(con adattamenti grafici effettuati dall'autore della tesi, tramite l'editor di immagini "Paint", al fine di conformare il concetto espresso dall'immagine al contesto in cui questa è inserita)

figura 13: tratta dal testo 21)

figura 14: tratta dal testo 21)

figura 16: tratta dal testo 21)

figura 28: tratta dal testo 21)

figura 33: tratta dal testo 21)

figura 40: tratta dal testo 21)

figura 53: tratta dal motore di ricerca Google

figura 58: tratta dal motore di ricerca Google

figura 59: tratta da una presentazione, in formato Power Point, reperita con il motore di ricerca Google

figura 60: tratta dal testo 21)

figura 61: tratta dal testo 21)

figura 62: tratta dal testo 21)

figura 63: tratta dal testo 21)

figura 64: tratta dal testo 21)

figura 65: tratta dal testo 21)

figura 66: tratta dal testo 21)

figura 67: tratta dal testo 21)

figura 68: tratta dal testo 21)

figura 69: tratta dal motore di ricerca Google

figura 70: tratta dal motore di ricerca Google

figura 72: tratta dal testo 21)

figura 73: tratta dal testo 21)

figura 74: tratta dal testo 21)

figura 75: tratta dal testo 21)

figura 76: tratta da appunti, in formato Adobe Reader, reperiti con il motore di ricerca Google

figura 80 (PIN osservato al microscopio elettronico a scansione): tratta dal motore di ricerca Google

figura 88: tratta dal Data Sheet del transistor BJT 2N4957 della Motorola

figura 95: tratta dal motore di ricerca Google

figura 98: tratta dagli appunti dell'esercitazioni del corso di "Optoelettronica" del Prof. Giacomo Langfelder (Politecnico di Milano, Dipartimento di Elettronica ed Informazione)

figura 99: quella a sinistra è tratta dal motore di ricerca Google, quella a destra dagli appunti del corso di "Fotonica Integrata" del Prof. G. Breglio (INFN – sezione di Napoli)

figura 110: tratta dal motore di ricerca Google

figura 111: tratta dal motore di ricerca Google

figura 113, parte sinistra: tratta dal motore di ricerca Google

figura 116: tratta dal sito web del Dipartimento di Fisica dell'Università di Würzburg

figura 122, parte sinistra: tratta dalla pubblicazione 1)

figura 123: tratta dalla pubblicazione 1)

figura 125: tratta dall'opuscolo illustrativo (gennaio 2008) sugli MPPC ("Multi - Pixel Photon Counter") prodotti dalla Hamamatsu Photonics K.K., Solid State Division

figura 126, parte superiore: tratta dalla pubblicazione 15)

figura 127: tratta dal testo 21)

figura 129: tratta dalla pubblicazione 15)

figura 131 (la misurazione mediante oscilloscopio): tratta dalla pubblicazione 15)

figura 132: tratta dalla pubblicazione 1)

figura 134: tratta dalla pubblicazione 1)

figura 135: tratta dalla pubblicazione 1)

figura 136: tratta dalla pubblicazione 16)

figura 137: tratta dal testo 27)

figura 138, le foto: tratte da Google e dall'enciclopedia online Wikipedia

figura 140: la foto di sinistra è tratta da Wikipedia, il disegno di destra è tratto da Google

figura 141: tratta dal motore di ricerca Google

figura 142, il grafico più a destra: tratto da appunti reperiti tramite Google

figura 143: tratta dal motore di ricerca Google

figura 144: tratta dal motore di ricerca Google

figura 146: tratta dal testo 21)

figura 147, il grafico di $F(\alpha)$: tratto da "Appunti di meccanica quantistica", reperibili online sul sito web del Dott. Cei

figura 148: tratta dal testo 21)

figura 149: tratta dal motore di ricerca Google

figura 159: tratta dal motore di ricerca Google

Glossario fondamentale

1) Cifra di Rumore di un generico quadripolo Q (NF, "Noise Figure") = rapporto fra la potenza totale di rumore $N_{tot,out}$, presente ai terminali di uscita del quadripolo "reale" Q (ossia un quadripolo avente, al suo interno, delle sorgenti di rumore intrinseche, cioè un quadripolo che introduce del rumore N_Q), e la potenza di rumore $N_{in,out}$, relativa al segnale informativo in ingresso, presente sempre ai terminali di uscita del quadripolo. $N_{tot,out}$ è la somma della potenza di rumore associata al segnale di ingresso, elaborato dal quadripolo, e la potenza di rumore introdotta dal quadripolo stesso.

$$NF = \frac{N_{tot,out}}{N_{in,out}} = \frac{N_{in,out} + N_Q}{N_{in,out}} = 1 + \frac{N_Q}{N_{in,out}}$$

Un quadripolo "ideale", ovvero non intrinsecamente rumoroso, presenta una cifra di rumore unitaria ($NF = 0dB$), ossia non introduce alcuna potenza di rumore ($N_Q = 0$) alla già presente potenza di rumore $N_{in,out}$ proveniente dall'ingresso: quest'ultima potenza non dipende dal quadripolo, dato che è una caratteristica stocastica del rumore associato al segnale in ingresso. Un quadripolo reale (intrinsecamente rumoroso) presenta una cifra di rumore $NF > 1$, ossia introduce potenza di rumore ($N_Q > 0$). Per un quadripolo Q che funziona da amplificatore possiamo definire la cifra di rumore come il rapporto fra il Rapporto Segnale - Rumore all'ingresso ed il Rapporto Segnale - Rumore all'uscita.

$$NF = \frac{SNR_{in}}{SNR_{out}}$$

Dunque un quadripolo reale peggiora il SNR presente in ingresso ($SNR_{out} < SNR_{in}$), a causa dell'introduzione di N_Q . Questo può comportare un peggioramento dell'intelligibilità dell'informazione, ossia è più difficile, per un'elettronica di lettura posta a valle del quadripolo Q, separare l'informazione (il segnale utile) dal rumore di fondo.

2) Efficienza Quantica (QE, "Quantum Efficiency", $0 < QE < 1$):

- per un fotorivelatore a semiconduttore (PIN, APD, SAM - APD ecc...) = rapporto fra il numero medio di coppie elettrone - lacuna raccolte ai terminali metallici del dispositivo, presso i quali si osserva una corrente sostenuta proprio dai suddetti portatori di carica, ed il numero di fotoni incidenti sul dispositivo.

$$QE = \frac{\text{numero medio di coppie elettrone - lacuna fotogenerate e rilevate dal circuito esterno di polarizzazione/lettura}}{\text{numero di fotoni incidenti}}$$

- per un fotomoltiplicatore a tubo (PMT) = rapporto fra il numero medio di elettroni emessi dal catodo metallico (fotocatodo), illuminato dalla radiazione elettromagnetica, per effetto fotoelettrico, ed il numero di fotoni che incidono sul fotocatodo del dispositivo.

$$QE = \frac{\text{numero medio di elettroni emessi dal catodo metallico per effetto fotoelettrico (emissione termoionica)}}{\text{numero di fotoni incidenti sul catodo metallico}}$$

3) Efficienza di rivelazione dei fotoni per un SiPM (PDE, "Photon Detection Efficiency", $0 < PDE < 1$) = rapporto fra il numero medio di fotoni rivelati da un SiPM, ovvero da una matrice di fotodiodi al silicio SAM - APDs $n^+/p/\pi/p^+$ (pixels, microcelle della matrice), tutti disposti in parallelo e polarizzati fortemente in inversa (funzionanti in modalità Geiger), ed il numero di fotoni effettivamente incidenti sulla superficie della matrice.

$$PDE = \frac{\text{numero medio di fotoni rivelati dal SiPM}}{\text{numero di fotoni incidenti sul SiPM}} = \frac{\text{numero medio di pixel attivati}}{\text{numero di fotoni incidenti sul SiPM}}$$

4) Equazione di Schrödinger = equazione differenziale, alle derivate parziali, che deve essere soddisfatta dalla funzione d'onda $\psi(x,y,z,t)$, affinché questa possa descrivere, in termini probabilistici (secondo "l'interpretazione di Copenhagen, prevalente in meccanica quantistica) il comportamento, nello spazio e nel tempo, di un sistema quantistico reale, ad esempio un elettrone all'interno di un reticolo cristallino. Lo stato $\psi(x,y,z,t)$ costituisce un vettore complesso appartenente allo spazio di Hilbert L^2 , vale a dire è una funzione a quadrato sommabile. La particella massiva, rappresentata dal suddetto stato, possiede energia totale E (= energia potenziale + energia cinetica) e si trova in una regione interessata da un'energia potenziale $U(x,y,z,t)$. Tale equazione deriva, secondo la formulazione originale di Schrödinger, dal concetto secondo il quale su una funzione d'onda $\psi(x,y,z,t) \in L^2$, rappresentativa di uno stato quantistico riscontrabile nella realtà, è possibile applicare l'operatore hamiltoniano \hat{H} , espresso, in forma analitica, nello spazio reale delle coordinate, ottenendo da questa operazione di misura ($\hat{H}\psi$) l'autostato $\psi(x,y,z,t)$ moltiplicato per l'osservabile E ($E\psi$), il quale costituisce l'autovalore dell'energia dello stato. Quest'azione di misura, che non altera lo stato ψ , e che quindi non ne provoca il "collassamento" su uno dei suoi infiniti sottostati, è esprimibile, in forma astratta, dalla "notazione di Dirac", nella quale lo stato ψ può evolvere nel tempo, mentre l'operatore hamiltoniano \hat{H} non lo può fare:

$$\hat{H} |\psi\rangle = E |\psi\rangle \quad [g1]$$

Il simbolo $|\psi\rangle$, contenente l'autofunzione di energia E , è noto come "simbolo di ket". Date due soluzioni normalizzate dell'equazione di Schrödinger $|\psi_m\rangle$ e $|\psi_n\rangle$, abbiamo che il prodotto scalare fra ψ_m e ψ_n risulta pari ad 1 (perfetta sovrapposizione fra i due stati) nel caso in cui $m = n$, ovvero nel caso in cui i due autostati possiedono autovalori dell'energia uguali, ossia $E_m = E_n$, mentre il loro prodotto scalare risulta pari a 0 (sovrapposizione nulla fra i due stati ψ_m e ψ_n) nel caso in cui $m \neq n$, ovvero nel caso in cui i due autostati possiedono autovalori dell'energia diversi, ossia $E_m \neq E_n$. In termini di notazione di Dirac:

$$\begin{aligned} \langle \psi_m | \psi_n \rangle &= \int \psi_m^*(r) \psi_n(r) d^3r = \text{"distribuzione delta di Dirac sull'asse dell'energia totale E"} = \\ &= \delta(E_m - E_n) = \delta_{mn} = \begin{cases} 1, & \text{se } m = n \\ 0, & \text{se } m \neq n \end{cases} \quad [g2] \end{aligned}$$

dove r è il vettore posizione e il simbolo $\langle \psi |$, contenente l'autofunzione ψ_m di autovalore dell'energia E_m , è noto come "simbolo di bra". Quest'ultimo rappresenta, all'interno della definizione di prodotto scalare fra funzioni d'onda a quadrato sommabile (spesso citato con il nome di "overlap integral" - "integrale di sovrapposizione"), il coniugato dello stato ψ_m . Gli stati $|\psi_m\rangle$ e $|\psi_n\rangle$, $\forall mn \in \mathbb{N}$, possono costituire una possibile base ortonormale, detta anche "sistema completo di autofunzioni", sulla quale sviluppare un generico stato $|\psi\rangle$, il quale, nell'ambito della fisica dei semiconduttori, può rappresentare, ad esempio, un elettrone o una lacuna.

Come è noto è possibile definire anche il valor medio di un dato osservabile, ovvero il valor medio del risultato ottenuto dal processo di misura attuato dall'applicazione di un dato operatore hermitiano (tutti gli operatori quantistici sono hermitiani) su un certo stato $|\psi\rangle$. Nel caso riportato nella formula [g1] l'incertezza sul valore della variabile aleatoria E è nullo, proprio in virtù della definizione di autofunzione e autovalore, per cui il valor medio dell'energia coincide con E . Seguendo sempre la notazione di Dirac abbiamo:

$$\langle \hat{H} \rangle = \langle \psi | \hat{H} | \psi \rangle = \int \psi^*(r) \hat{H} \psi(r) d^3r = \int \psi^*(r) \left[\frac{-\hbar^2}{2m_e^*} \nabla^2 + U(r) \right] \psi(r) d^3r = E \quad [g3]$$

Il termine fra parentesi quadre rappresenta l'espressione analitica, nello spazio reale delle coordinate, dell'operatore hamiltoniano: si noti il termine rappresentativo dell'energia cinetica e quello rappresentativo dell'energia potenziale. La [g3] potrebbe descrivere un problema "one body" ("a singolo corpo") relativo, ad esempio, ad un elettrone all'interno di un semiconduttore, avente massa efficace m_e^* , descritto da una particolare funzione d'onda $\psi(r)$, data da un'onda piana monocromatica (onda di De Broglie) moltiplicata per una funzione

avente la stessa periodicità del cristallo, delocalizzato, come previsto dal teorema di Bloch, all'interno del reticolo cristallino, ipotizzato perfetto. $U(r)$, in questo contesto, rappresenta il potenziale periodico prodotto dagli atomi del cristallo ("potenziale di background").

La forma astratta di Dirac dell'equazione di Schrödinger può essere convertita in una forma funzionale al calcolo degli autovalori dell'energia, nota come "notazione di Heisenberg". Sotto questa forma lo stato $|\psi\rangle$ può essere pensato come un vettore, a rigori di dimensione infinita, ma di fatto, in molti casi pratici, di dimensione N pari a 4, 6 oppure 8, sviluppato su una base ortonormale $\{|\psi_k\rangle\}$, scelta tenendo conto della natura del problema; inoltre lo stato $|\psi\rangle$ non evolve nel tempo. Nello studio dei materiali semiconduttori, adatti in campo fotonico, le due basi ortonormali maggiormente utilizzate sono quelle di Clebsch – Gordan e di Luttinger – Kohn (quest'ultima è la più in auge). L'operatore \hat{H} , che può evolvere nel tempo, è esprimibile, in tale contesto, sotto forma di una matrice quadrata $N \times N$. Questa matrice H è hermitiana, dal momento che l'operatore hamiltoniano è hermitiano, per cui la sua trasposta coincide con la sua coniugata ($H^T = H^*$), mentre la stessa matrice H coincide con la sua aggiunta ($H = H^\dagger$). Più formalmente, dati due generici autostati ψ_a e ψ_b dell'operatore \hat{H} , si può affermare che:

$$\langle \psi_a | \hat{H}^\dagger \psi_b \rangle = \langle \hat{H} \psi_a | \psi_b \rangle = \langle \psi_a | \hat{H} \psi_b \rangle^*$$

Assumiamo, in osservanza all'interpretazione di Copenhagen, che un generico stato $|\psi\rangle$ sia una sovrapposizione (= combinazione lineare) di stati $|\psi_i\rangle$, i quali costituiscono un sistema completo di autofunzioni dell'operatore hamiltoniano:

$$|\psi\rangle = \sum_{i=1}^{\infty} \alpha_i |\psi_i\rangle \quad \alpha_i \in \mathbb{C}$$

se sottoponiamo $|\psi\rangle$ ad una singola azione di misura da parte di un certo operatore, ad esempio quello di Hamilton, lo stato $|\psi\rangle$ "collassa" su uno dei suoi infiniti sottostati $|\psi_i\rangle$ (vale a dire che $|\psi\rangle$ viene alterato dall'azione di misura) con una probabilità pari al modulo quadro del relativo coefficiente ($|\alpha_i|^2$). Ciò è dovuto al fatto che il sistema di misura utilizzato per stimare un qualunque osservabile relativo allo stato misurando $|\psi\rangle$ (ad esempio un apparato di fotorivelazione) è un sistema macroscopico caratterizzato da molti gradi di libertà, pertanto un sistema quantistico $|\psi\rangle$, nel momento in cui interagisce con lo strumento di misura, subisce una sorta di "bagno termico" che lo altera irreversibilmente; tale alterazione è descrivibile, in termini quantistici, mediante un collassamento di $|\psi\rangle$ su uno dei suoi sottostati $|\psi_i\rangle$. Utilizzando la notazione di Heisenberg possiamo pensare lo stato $|\psi\rangle$ come un vettore $\in \mathbb{C}^{N \times 1}$ sviluppato nello spazio di Hilbert individuato dalla base ortonormale $\{|\psi_i\rangle\}$, formalmente con $N \rightarrow \infty$, ma nella maggior parte dei casi pratici, come già detto, $N = 4, 6$ oppure 8 . \hat{H} è una matrice $N \times N$ hermitiana, la cui diagonale principale è fatta da elementi reali, per cui i suoi autovalori $E_1, E_2, \dots, E_i, \dots, E_N$ sono tutti reali (un operatore hermitiano è sempre associato ad un osservabile fisico, in altre parole il suo valor medio è sempre reale), dunque $\{E_i\}$ costituiscono gli autovalori dell'energia dei sottostati $|\psi_i\rangle$, mentre la loro somma $\sum_i E_i$ fornisce l'energia totale E dello stato $|\psi\rangle$. Convertendo la notazione di Dirac dell'equazione di Schrödinger nella notazione matriciale di Heisenberg otteniamo:

$$\begin{pmatrix} H_{11} & H_{12} & \dots & H_{1N} \\ H_{21} & H_{22} & & \vdots \\ \vdots & \dots & \dots & H_{NN} \end{pmatrix} \begin{pmatrix} \alpha_1 \psi_1 \\ \vdots \\ \alpha_N \psi_N \end{pmatrix} = \begin{pmatrix} E_1 & 0 & \dots & 0 \\ 0 & E_2 & & \vdots \\ \vdots & 0 & \dots & E_N \end{pmatrix} \begin{pmatrix} \alpha_1 \psi_1 \\ \vdots \\ \alpha_N \psi_N \end{pmatrix} \quad [g4]$$

Compattando la [g4] possiamo scrivere:

$$H \psi = E I \psi \quad [g5]$$

dove I rappresenta la matrice identica. Dalla [g5] è possibile ricavare ed eguagliare a zero il polinomio caratteristico $\gamma(E)$, di grado N , della matrice H :

$$\gamma(E) = \det(H - E I) = \det \begin{bmatrix} H_{11} - E & H_{12} & \dots & H_{1N} \\ H_{21} & H_{22} - E & & \vdots \\ \vdots & \dots & \dots & H_{NN} - E \end{bmatrix} = 0 \quad [g6]$$

Nell'ambito dello studio dei materiali semiconduttori utilizzati in fotonica, la [g6] è nota sotto il nome di "secular equation" (equazione secolare), mentre il determinante nella [g6] è noto come "determinante secolare". Le N soluzioni di $\gamma(E)$ sono gli N valori dell'energia E_1, E_2, \dots, E_N , ciascuno costituente l'autovalore E_i correlato all'autostato ψ_i . Una volta noti gli autovalori E_i della matrice hamiltoniana, oltre ai coefficienti α_i relativi ai sottostati ψ_i , è possibile calcolare il valor medio E dell'energia totale dello stato $|\psi\rangle$. Considerando lo stato ket $|\psi\rangle$ come un vettore colonna $\mathbb{C}^{N \times 1}$ e lo stato bra $\langle\psi|$ come un vettore riga $\mathbb{C}^{1 \times N}$ i cui N elementi sono i complessi coniugati degli N elementi appartenenti allo stato ket ($\langle\psi|$ è lo stato trasposto coniugato di $|\psi\rangle$), possiamo esprimere il calcolo di E nel modo seguente:

$$E = \langle \hat{H} \rangle = \langle \psi | H | \psi \rangle = \int \left\{ (\alpha_1^* \psi_1^*, \dots, \alpha_N^* \psi_N^*) \begin{pmatrix} E_1 & 0 & \dots & 0 \\ 0 & E_2 & & \vdots \\ \vdots & 0 & \dots & E_N \end{pmatrix} \begin{pmatrix} \alpha_1 \psi_1 \\ \vdots \\ \alpha_N \psi_N \end{pmatrix} \right\} d^3 r = \sum_{i=1}^N E_i |\alpha_i|^2$$

Dunque E è la media degli autovalori E_i pesata dai coefficienti α_i . Per le N autofunzioni $\{\psi_i\}$, ricavabili dalla [g5] una volta estratti gli autovalori $\{E_i\}$ dalla [g6], sono valide le stesse proprietà di "overlap" indicate nella [g2]. Va notato, in aggiunta, che se un autovalore E_k ha degenerazione $M (\leq N)$, ovvero se M sottostati $\{\psi_i\}$ di ψ possiedono la stessa energia, questi individuano un sottospazio vettoriale (sottospazio rispetto allo spazio a cui appartiene ψ) i cui stati, dati da una qualunque combinazione lineare degli M stati $\{\psi_i\}$, hanno tutti autovalore E_k .

Se il vettore $|\psi\rangle$ rappresenta la sovrapposizione di N stati elettronici atomici, dove N è il numero degli atomi di un cristallo semiconduttore interagenti fra loro, ad esempio N orbitali di tipo S, e nel caso in cui gli elementi H_{ij} ($\forall i, j \in [1, N]$) della matrice \hat{H} siano dipendenti dal vettore d'onda $k = (k_x, k_y, k_z)$, allora è possibile calcolare la dipendenza da k dei livelli energetici dati dagli N autovalori di \hat{H} , ossia $E_1(k), E_2(k), \dots, E_N(k)$, e dunque il diagramma E - k della banda formata dalla degenerazione di ciascun singolo orbitale (la banda di conduzione, nel caso di degenerazione degli orbitali S in un semiconduttore a gap diretto). Questo metodo di calcolo delle bande, derivante dalla "secular equation", è noto come TBM - "Tight Binding Method".

Partendo dalla notazione di Dirac riportata nella [g1], nel caso generale dipendente dal tempo, è immediato ricavare la forma classica dell'equazione di Schrödinger:

$$\hat{H} |\psi\rangle = E |\psi\rangle = \hat{T} |\psi\rangle \xrightarrow[\substack{\text{l'operatore hamiltoniano} \\ \text{espresso nello spazio reciproco} \\ \text{dei vettori d'onda } k}]{\text{}} \left[\frac{\hat{p}^2}{2m} + \hat{U} \right] \psi(k, t) = E \psi(k, t) = \hat{T} \psi(k, t) \quad [g7]$$

dove l'operatore quantistico \hat{H} è espresso nella sua forma "classica" in base al noto "principio di corrispondenza" fra meccanica quantistica e meccanica classica. \hat{T} rappresenta l'operatore temporale, il quale, applicato ad uno stato, ne fornisce l'evoluzione temporale, fissato un punto dello spazio.

$$[g7] \xrightarrow[\substack{\text{passaggio dell'operatore hermitiano } p^2 \\ \text{dallo spazio reciproco} \\ \text{a quello diretto delle coordinate}}]{\text{}} \left[\frac{-\hbar^2}{2m} \nabla^2 + U(x, y, z, t) \right] \psi(x, y, z, t) = E \psi(x, y, z, t) = i\hbar \frac{\partial \psi(x, y, z, t)}{\partial t}$$

dove $\langle \hat{H} \rangle = E$ è costante nel tempo, nel caso in cui la particella massiva in questione (di massa m) si trovi in un campo elettrico conservativo. La soluzione generica dell'ultima equazione, dipendente dal tempo, è:

$$\psi(r, t) = \psi(r) e^{-i \frac{E}{\hbar} t}$$

dove r è il vettore posizione ed E/\hbar la pulsazione ω con la quale la fase della funzione ruota. Se sostituiamo tale funzione nell'equazione di Schrödinger dipendente dal tempo otteniamo l'equazione di Schrödinger non dipendente dal tempo, in rappresentazione delle coordinate:

$$\nabla^2 \psi(r) + \frac{2m}{\hbar^2} [E - U(r)] \psi(r) = 0 \quad [g8]$$

5) Guadagno (di fotoconduzione):

- per un fotorivelatore a semiconduttore (APD, SAM – APD ecc...) = numero medio di coppie elettrone – lacuna raccolte ai terminali metallici del dispositivo per ciascuna coppia elettrone – lacuna (fotocariche primarie) fotogenerata a causa dell'interazione fra il semiconduttore ed un singolo fotone incidente.

$$G = \frac{\text{numero medio di coppie elettrone - lacuna rivelate dal circuito esterno di polarizzazione/lettura}}{\text{numero di fotocopie primarie elettrone-lacuna}} =$$

$$= \frac{\text{carica totale raccolta ai terminali}}{\text{fotocarica primaria (carica elementare)}} = \frac{Q_{raccolta}}{|e|}$$

G = fattore (medio) di moltiplicazione della densità superficiale della fotocorrente di cariche primarie; tale moltiplicazione è dovuta al fenomeno del breakdown a valanga

- per un fotomoltiplicatore a tubo (PMT) = numero medio di elettroni rilevati sull'anodo del dispositivo per ciascun elettrone emesso dal catodo per effetto fotoelettrico.

$$G = \frac{\text{numero medio di elettroni arrivati sull'anodo e rilevati dal circuito esterno}}{\text{numero di elettroni emessi dal catodo metallico per effetto fotoelettrico}}$$

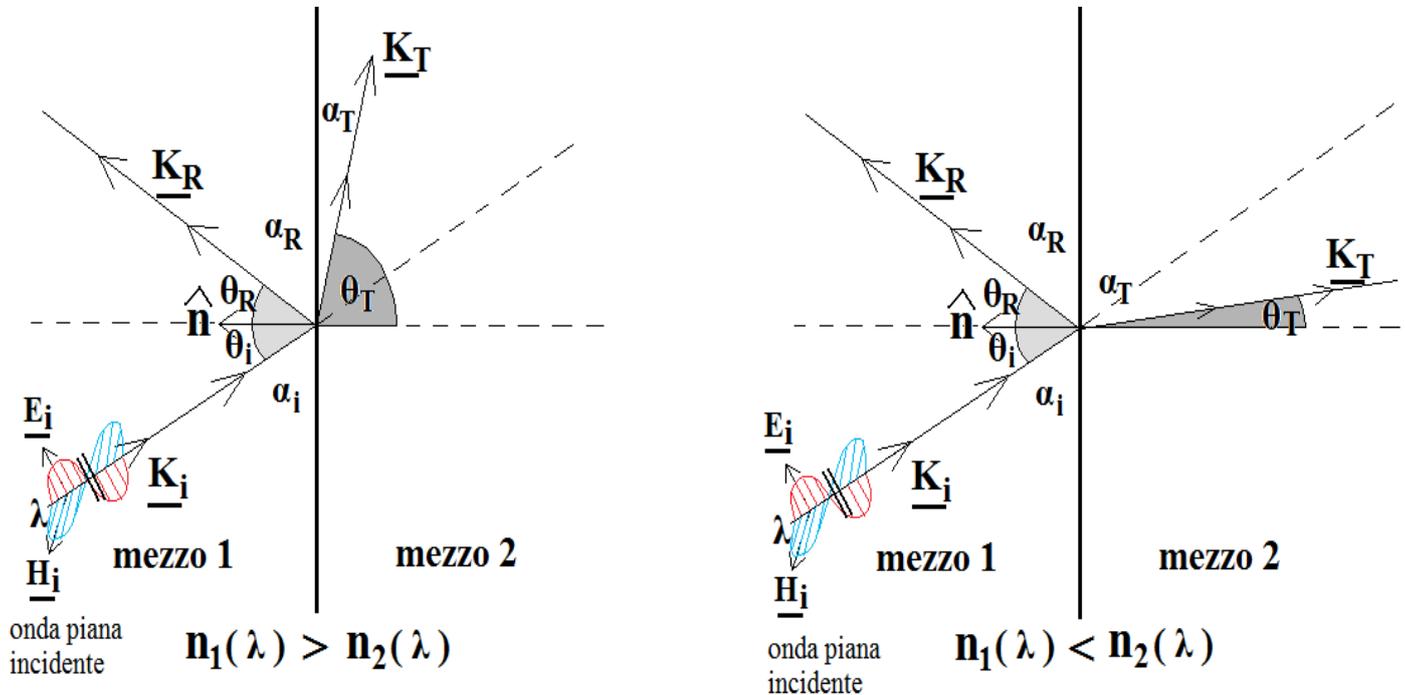
6) Risoluzione temporale di un generico fotorivelatore = capacità di un qualunque fotorivelatore di distinguere, temporalmente, due eventi luminosi verificatisi uno dopo l'altro; minore è l'intervallo temporale fra due radiazioni incidenti consecutive che il fotorivelatore è in grado di apprezzare e quindi di separare a livello di risposta elettrica, maggiore è la sua risoluzione temporale. In seguito alla rivelazione di una radiazione luminosa, e quindi successivamente alla formazione e alla regimazione della risposta elettrica da parte del dispositivo, maggiore è la velocità con la quale il fotorivelatore è in grado di tornare nella condizione di preassorbimento della radiazione, maggiore è la sua banda operativa (banda passante) e maggiore è la sua risoluzione temporale.

7) Legge di Snell per la rifrazione della luce = legge che consente di determinare gli angoli di riflessione θ_R e di rifrazione θ_T per un raggio di luce (onda monocromatica) che incide all'interfaccia fra due mezzi aventi indice di rifrazione diversi, n_1 ed n_2 (uno specchio), noti i due indici di rifrazione alla lunghezza d'onda della radiazione incidente e l'angolo di incidenza θ_i . Gli angoli θ_i , θ_R e θ_T sono formati dai vettori d'onda \underline{k}_i , \underline{k}_R e \underline{k}_T con la direzione individuata dal versore \hat{n} normale all'interfaccia. I complementari angoli α sono quelli formati dagli stessi tre vettori d'onda con il piano costituente l'interfaccia. L'espressione della legge di Snell è:

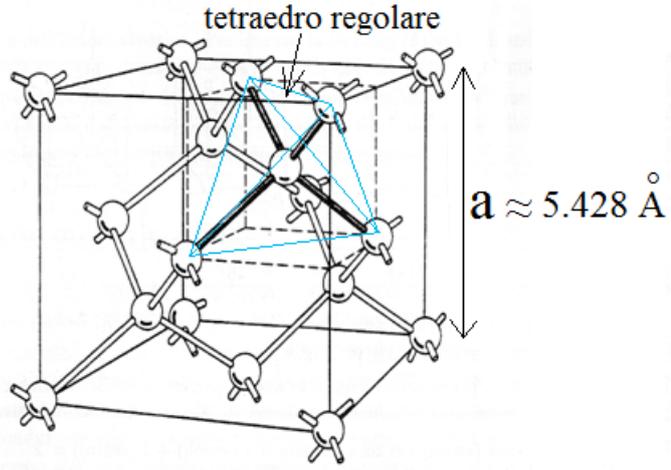
$$n_1(\lambda) \sin(\theta_i) = n_2(\lambda) \sin(\theta_T); \quad \text{con } \theta_i = \theta_R$$

che può essere scritta anche nel modo seguente:

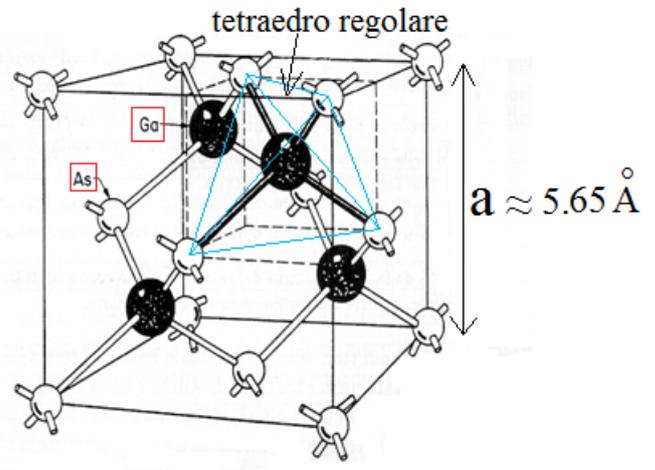
$$n_1(\lambda) \cos(\alpha_i) = n_2(\lambda) \cos(\alpha_T); \quad \text{con } \alpha_i = \alpha_R$$



8) Struttura cristallina, di tipo diamante, del silicio Si (semiconduttore semplice), struttura cristallina, di tipo zincoblenda, dell'arseniuro di gallio GaAs (semiconduttore composto binario III - V) e cella fondamentale di Wigner - Seitz della semplice struttura FCC (cubo a facce centrate) alla base dei reticoli di tipo diamante e zincoblenda:

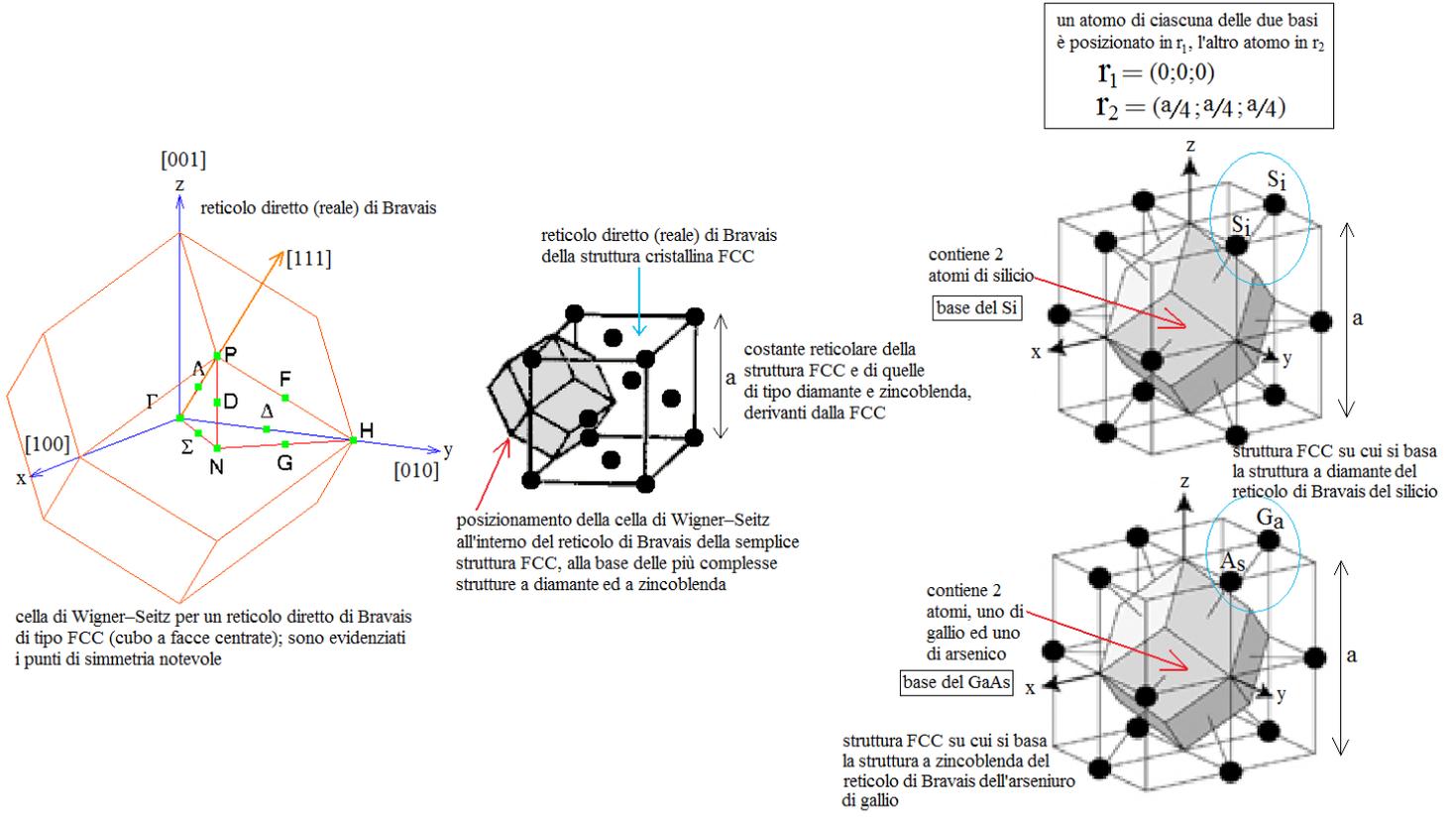


la cella fondamentale del reticolo diretto di Bravais di tipo diamante, cioè la cella di Wigner-Seitz, conta due atomi di Si, che formano la base, uno posizionato in $(0,0,0)$, l'altro in $(a/4, a/4, a/4)$



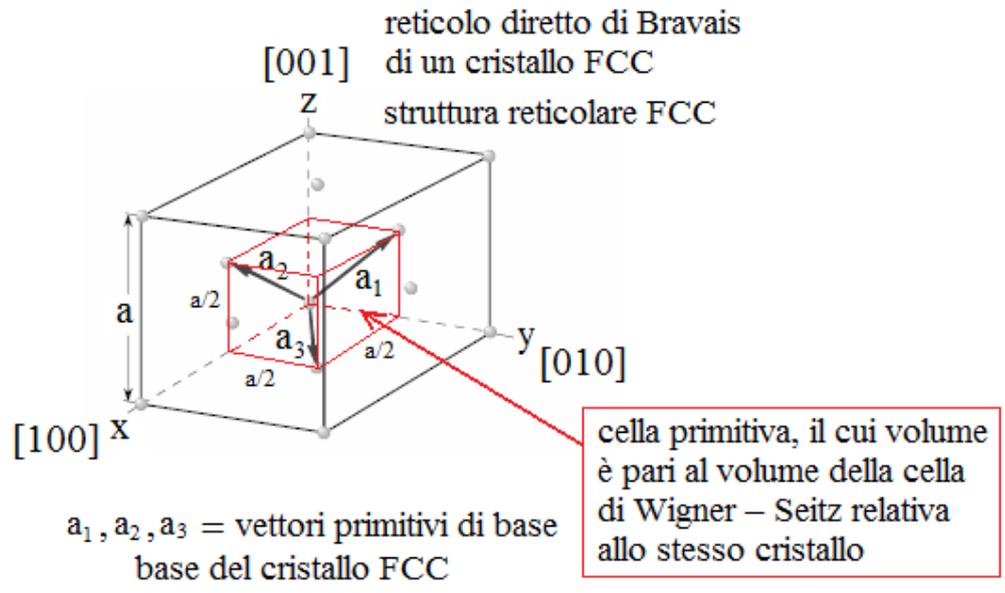
la cella fondamentale del reticolo diretto di Bravais di tipo zincoblenda, cioè la cella di Wigner - Seitz, conta due atomi, uno di As ed uno di Ga, che formano la base, uno posizionato in $(0,0,0)$, l'altro in $(a/4, a/4, a/4)$

La struttura cristallina di altri semiconduttori semplici, come carbonio C e germanio Ge, è identica a quella del silicio Si, ovvero è di tipo diamante, chiaramente con costanti reticolari "a" di valore diverso: quella del carbonio è pari a circa 3.57 \AA , mentre quella del germanio è pari a circa 5.66 \AA . La struttura cristallina di altri semiconduttori composti, sia III - V che II - VI, è identica a quella dell'arseniuro di gallio GaAs, con ciascuno il proprio valore della costante reticolare "a".

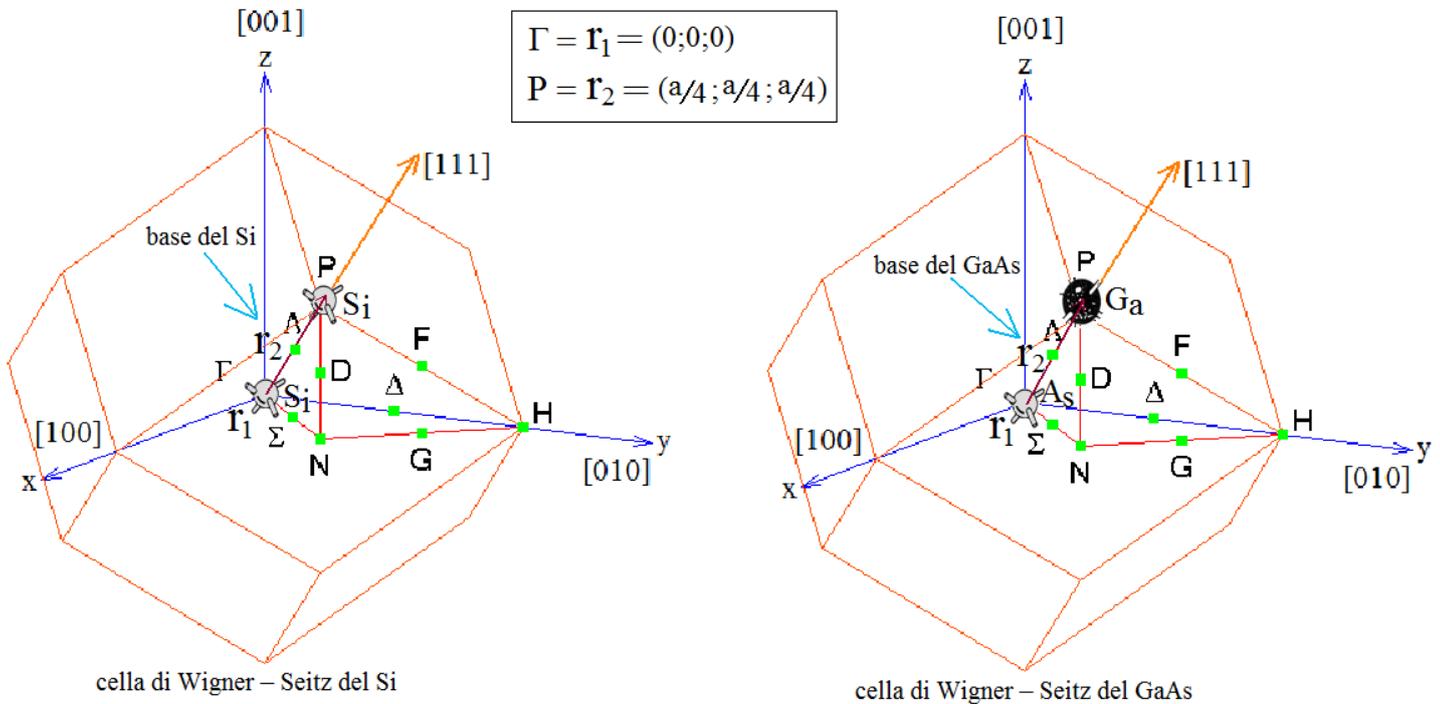


Per costruire la cella fondamentale di Wigner - Seitz del reticolo diretto di Bravais, nel caso di una struttura reticolare FCC, che è alla base delle più complesse strutture cristalline di tipo diamante e di tipo zincoblenda, dobbiamo prima di tutto fissare un generico punto reticolare. Supponiamo che tale punto costituisca l'origine degli assi x, y, z, comunemente indicata con Γ , lungo i quali si sviluppa il reticolo diretto di Bravais. Tracciamo i vettori R ($R_1, R_2, R_3, \dots, R_{12}$) aventi come punto di applicazione comune il punto di origine Γ e come punte tutti i dodici punti reticolari adiacenti a Γ , sui quali sono posizionati, all'interno del cristallo FCC, i dodici atomi primi vicini dell'atomo localizzato su Γ . Costruiamo i dodici piani P perpendicolari a questi vettori R e passanti per il punto mediano di ciascun vettore. Il più piccolo volume formato intorno a Γ e delimitato, "impacchettato", da questi dodici piani, ha una superficie a forma di poliedro, il quale costituisce la cella di Wigner - Seitz.

La cella di Wigner - Seitz contiene esattamente un punto reticolare, cioè un atomo, nel caso di una struttura reticolare FCC (tale atomo rappresenta quindi la base del cristallo FCC); la cella di Wigner - Seitz ha lo stesso volume della cella primitiva originale, mostrata qui di seguito.



La cella di Wigner - Seitz per le strutture cristalline a diamante ed a zincoblenda contiene due punti reticolari, cioè due atomi, i quali formano la base del cristallo; siamo infatti nel caso di due cubi FCC traslati, l'uno rispetto all'altro, lungo la direzione cristallografica [111], cioè lungo la diagonale principale, di un quarto della lunghezza di quest'ultima. Riportiamo qui di seguito la base del reticolo diretto di Bravais del silicio e dell'arseniuro di gallio:

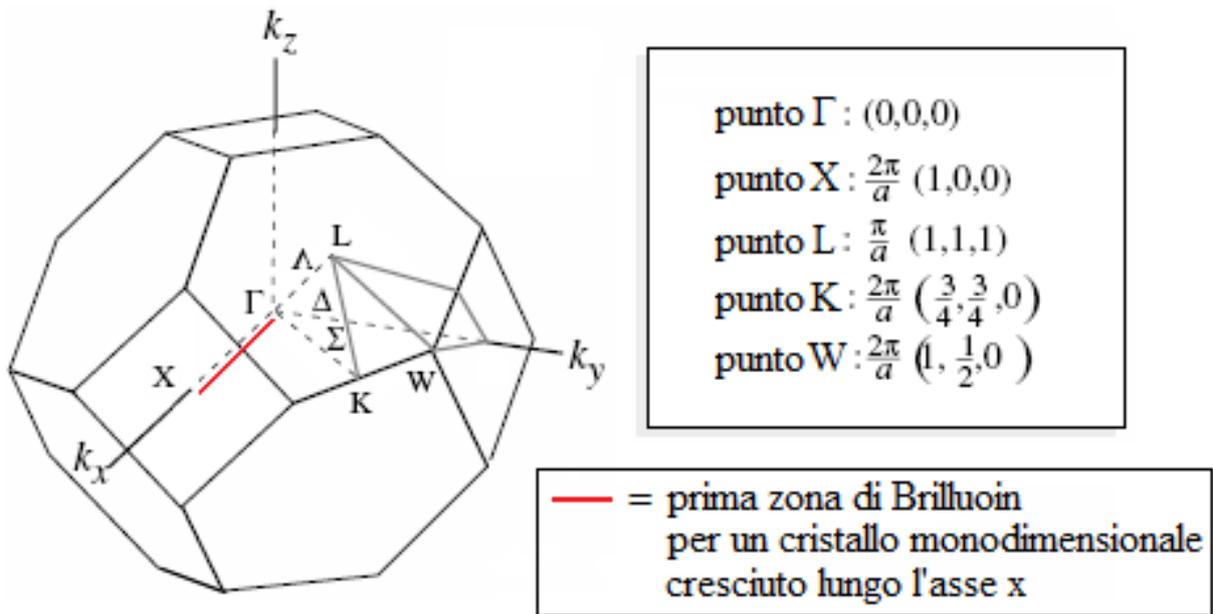


Nella base del reticolo diretto di Bravais del Si un atomo è posizionato nel punto Γ , che è individuato dal vettore posizione $\mathbf{r}_1 = (0;0;0)$, mentre l'altro atomo è posizionato nel punto \mathbf{P} , che è individuato dal vettore posizione $\mathbf{r}_2 = (a/4; a/4; a/4)$, dove $a \approx 5.428 \text{ \AA}$. Nella base del reticolo diretto di Bravais del GaAs l'atomo di As è posizionato nel punto Γ , che è individuato dal vettore $\mathbf{r}_1 = (0;0;0)$, mentre l'atomo di Ga è posizionato nel punto \mathbf{P} , che è individuato dal vettore $\mathbf{r}_2 = (a/4; a/4; a/4)$, dove $a \approx 5.65 \text{ \AA}$.

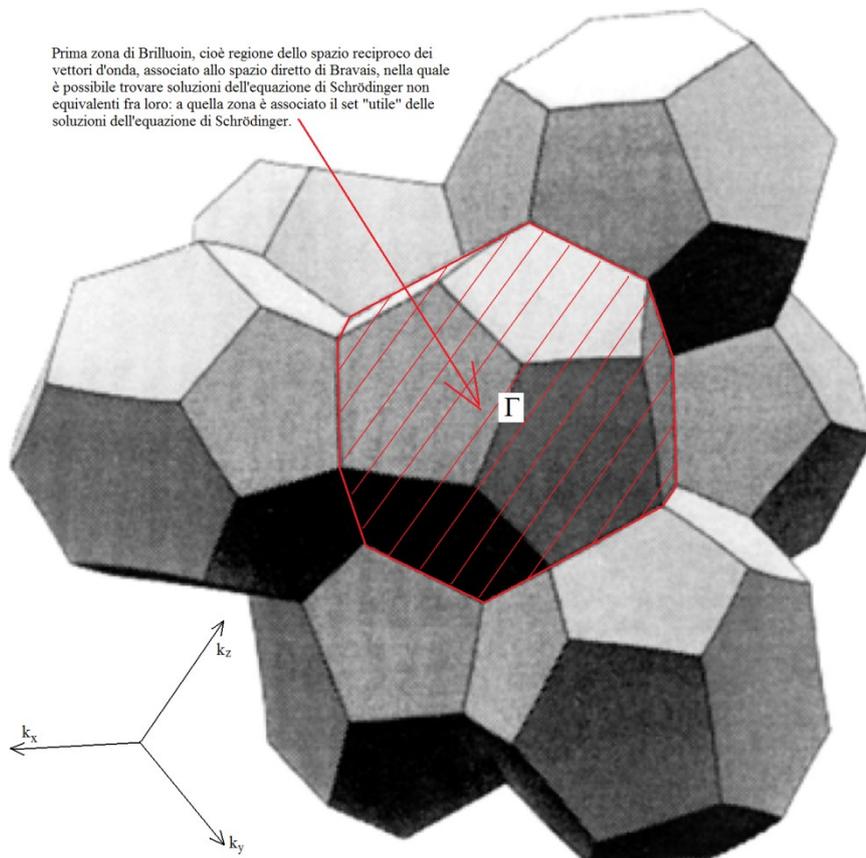
Replicando periodicamente la cella di Wigner - Seitz è possibile riempire tutto lo spazio in cui si sviluppa il cristallo, senza regioni vuote (ciascuna cella di Wigner - Seitz si "incastra" perfettamente con quelle adiacenti), e quindi definirne la geometria attraverso tutto il reticolo diretto di Bravais.

Il volume compreso all'interno della cella di Wigner - Seitz può essere visto come il dominio della trasformata di Fourier, applicazione il cui codominio è la ben nota "prima zona di Brillouin", che è la cella fondamentale di Wigner - Seitz del reticolo reciproco associato al cristallo. Per costruire la prima zona di Brillouin del reticolo reciproco di Fourier, sempre nel caso di una struttura reticolare FCC, si esegue esattamente quanto spiegato in precedenza nel caso del reticolo diretto, solo che la costruzione va eseguita nello spazio dei vettori d'onda \mathbf{k} , i cui assi coordinati k_x, k_y, k_z sono quantizzati secondo i passi $\Delta k_x = 2\pi/L_x, \Delta k_y = 2\pi/L_y, \Delta k_z = 2\pi/L_z$, dove L_x, L_y, L_z sono le dimensioni del cristallo. Al posto dei vettori spaziali \mathbf{R} , che congiungono il punto fissato Γ con i suoi punti $(x;y;z)$ primi vicini (dodici, nel caso di reticolo diretto FCC), avremo i vettori d'onda $\mathbf{G} = \mathbf{p}/\hbar$ ($G_1, G_2, G_3 \dots$) che congiungono l'origine Γ degli assi k_x, k_y, k_z con i punti $(k_x; k_y; k_z)$ primi vicini di Γ . Il più piccolo volume, nello spazio dei \mathbf{k} , formato intorno a Γ e delimitato, "impacchettato", dai piani che attraversano i punti mediani dei vettori \mathbf{G} e che sono ortogonali rispetto agli stessi \mathbf{G} , ha una superficie a forma di poliedro diversa da quella precedente: il nuovo poliedro costituisce la cella di Wigner - Seitz nello spazio reciproco dei vettori d'onda, cioè la prima zona di Brillouin. La forma della prima zona di Brillouin per un reticolo cristallino di tipo FCC è mostrata nella voce 9) del glossario.

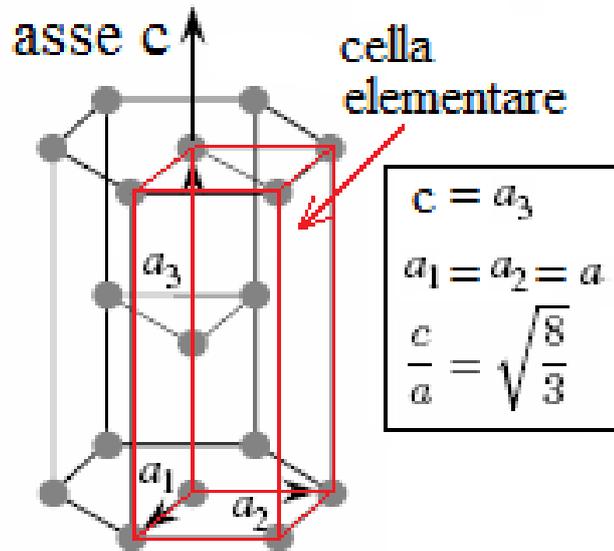
9) Prima zona di Brilluoin per un reticolo cristallino di tipo FCC e per semiconduttori caratterizzati da una struttura cristallina di tipo diamante (C, Si, Ge) oppure da una struttura cristallina di tipo zincoblenda (III - V, II - VI):



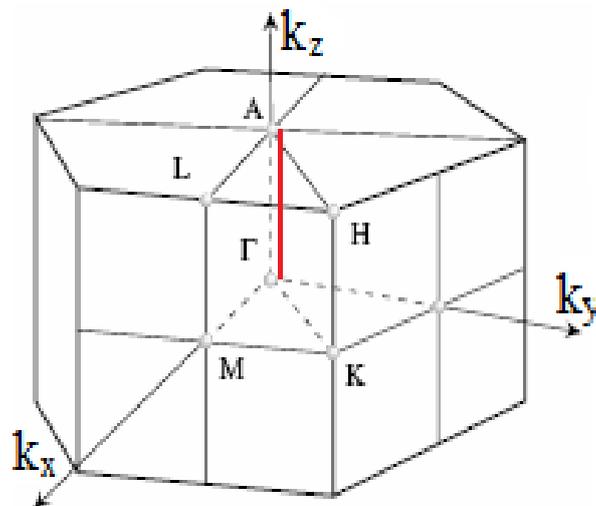
Replicando periodicamente la prima zona di Brilluoin è possibile riempire tutto lo spazio reciproco associato al cristallo, senza regioni vuote (ciascuna zona di Brilluoin si "incastra" perfettamente con quelle adiacenti), pertanto è possibile definire le soluzioni dell'equazione di Schrödinger, e quindi gli autovalori E dell'energia elettronica totale e conseguentemente il diagramma a bande $E - k$ del cristallo.



10) Struttura cristallina, di tipo HCP - "Hexagonal Close Pack", caratteristica dei semiconduttori composti III - V a base di azoto ("nitrides semiconductors") BN, AlN, GaN, InN e dei metalli pesanti di cadmio Cd e cobalto Co:

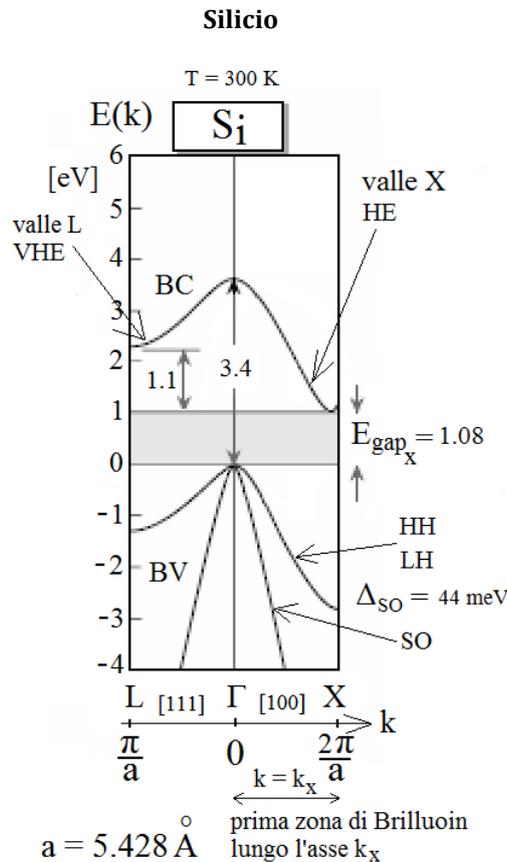


11) Prima zona di Brilluoin per semiconduttori e metalli caratterizzati da una struttura cristallina di tipo HCP:



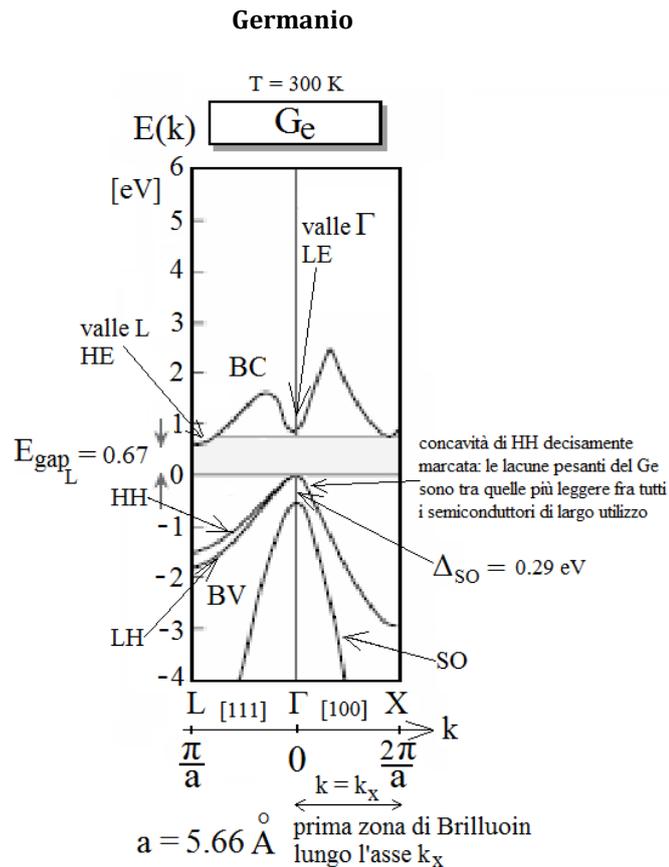
— = prima zona di Brilluoin per un semiconduttore nitridico monodimensionale, cresciuto lungo l'asse z

12) Diagrammi a bande di energia E - k per alcuni dei semiconduttori, sia semplici che composti, più utilizzati in fotonica ed in particolare nella realizzazione di fotorivelatori:



L'applicazione di alte polarizzazioni inverse ad un fotodiiodo di Si (PIN o APD) può comportare due fenomeni spuri da evitare ai fini della fotorivelazione:

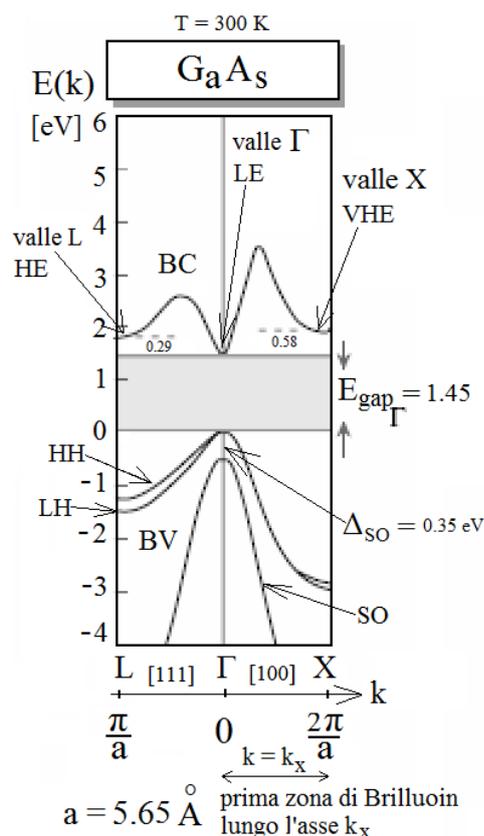
- la formazione, in seguito all'eccessivo piegamento delle bande, di alte correnti di tunneling che si sovrappongono al fotosegnale utile, rendendone impossibile la lettura
- i fotoelettroni di conduzione, altamente energetici a causa del forte campo elettrico imposto dall'esterno, potrebbero raggiungere i punti alti della curva E(k) della valle X, i quali, nello spazio dei vettori d'onda k, distano relativamente poco dai punti della curva E(k) della valle L. I fotoelettroni, durante il loro trascinarsi verso il bulk di destinazione di tipo n, potrebbero subire degli scattering di varia natura (dovuti a fononi, impurezze metalliche, impurezze droganti, difetti reticolari ecc...), a causa dei quali i fotoelettroni possono transitare da uno stato iniziale k_i , appartenente alla valle X, ad uno stato finale k_f , appartenente alla valle L. Più la polarizzazione inversa imposta sui terminali del fotodiiodo è alta, più la distanza, nello spazio di Fourier, fra lo stato k_i e lo stato k_f diminuisce e maggiore è la probabilità che un fotoelettrone transiti, pertanto, dalla valle X alla valle L. Tale transizione comporta un aumento della massa efficace degli elettroni per la mobilità, con conseguente diminuzione della mobilità elettronica e quindi dell'ampiezza della fotocorrente.



L'applicazione di alte polarizzazioni inverse ad un fotodiodo di Ge (PIN o APD) pu  comportare due fenomeni, il primo negativo, ai fini della fotorivelazione, il secondo positivo:

- la formazione, in seguito all'eccessivo piegamento delle bande, di alte correnti di tunneling che si sovrappongono al fotosegnale utile, rendendone impossibile la lettura
- i fotoelettroni di conduzione, altamente energetici a causa del forte campo elettrico imposto dall'esterno, potrebbero raggiungere i punti alti della curva $E(k)$ della valle L, i quali, nello spazio dei vettori d'onda k , distano relativamente poco dai punti della curva $E(k)$ della valle Γ . I fotoelettroni, durante il loro trascinarsi verso il bulk di destinazione di tipo n, potrebbero subire degli scattering di varia natura (dovuti a fononi, impurezze metalliche, impurezze droganti, difetti reticolari ecc...), a causa dei quali i fotoelettroni possono transitare da uno stato iniziale k_i , appartenente alla valle L, ad uno stato finale k_f , appartenente alla valle Γ . Pi  la polarizzazione inversa imposta sui terminali del fotodiodo   alta, pi  la distanza, nello spazio di Fourier, fra lo stato k_i e lo stato k_f diminuisce e maggiore   la probabilit  che un fotoelettrone transiti, pertanto, dalla valle L alla valle Γ . Tale transizione comporta una diminuzione della massa efficace degli elettroni per la mobilit , con conseguente aumento della mobilit  elettronica e quindi dell'ampiezza della fotocorrente.

Arseniuro di Gallio



L'applicazione di alte polarizzazioni inverse ad un fotodiodo di GaAs (PIN o APD) può comportare due fenomeni spuri da evitare ai fini della fotorivelazione:

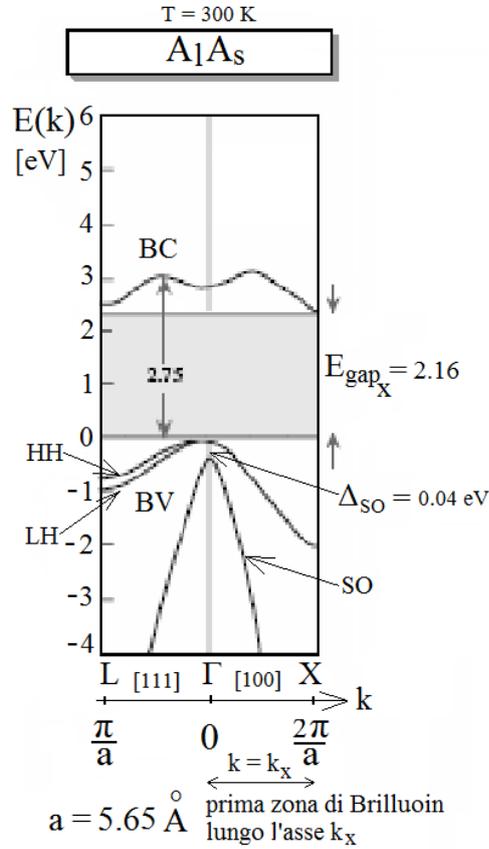
- la formazione, in seguito all'eccessivo piegamento delle bande, di alte correnti di tunneling che si sovrappongono al fotosegnale utile, rendendone impossibile la lettura
- i fotoelettroni di conduzione, altamente energetici a causa del forte campo elettrico imposto dall'esterno, potrebbero raggiungere i punti alti della curva $E(k)$ della valle Γ , i quali, nello spazio dei vettori d'onda k , distano relativamente poco dai punti della curva $E(k)$ della valle L . I fotoelettroni, durante il loro trascinarsi verso il bulk di destinazione di tipo n , potrebbero subire degli scattering di varia natura (dovuti a fononi, impurezze metalliche, impurezze droganti, difetti reticolari ecc...), a causa dei quali i fotoelettroni possono transitare da uno stato iniziale k_i , appartenente alla valle Γ , ad uno stato finale k_f , appartenente alla valle L . Più la polarizzazione inversa imposta sui terminali del fotodiodo è alta, più la distanza, nello spazio di Fourier, fra lo stato k_i e lo stato k_f diminuisce e maggiore è la probabilità che un fotoelettrone transiti, pertanto, dalla valle Γ alla valle L . Tale transizione comporta un aumento della massa efficace degli elettroni per la mobilità, con conseguente diminuzione della mobilità elettronica e quindi dell'ampiezza della fotocorrente. Aumentando ulteriormente la polarizzazione inversa sul fotodiodo la probabilità di transizione dei fotoelettroni dalla valle Γ alla valle X inizia a diventare non trascurabile. Il passaggio dei fotoelettroni da uno stato iniziale k_i , appartenente alla valle Γ , ad uno stato finale k_f , appartenente alla valle X , comporterebbe un ulteriore aumento della massa efficace degli elettroni per la mobilità ed un'ulteriore diminuzione della mobilità elettronica e quindi dell'ampiezza della fotocorrente.

La curva di dispersione $E(k)$ presso il minimo E_c della BC, intorno al punto di simmetria notevole Γ , può essere approssimata, come è ben noto, con una relazione parabolica, ma ancora più accuratamente con la seguente relazione non parabolica:

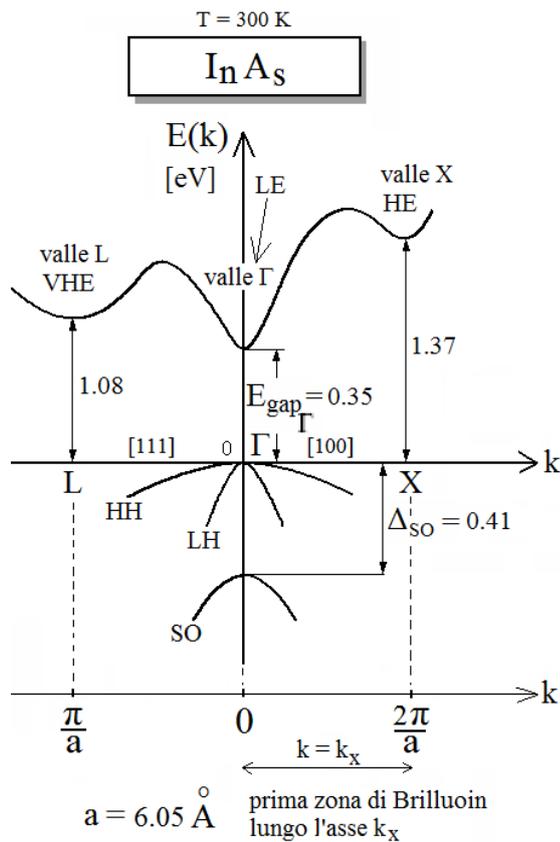
$$E(k)[1 + \alpha E(k)] = \frac{\hbar^2 k^2}{2m_e^*}$$

dove, per il GaAs, $\alpha = 0.67 \text{ eV}^{-1}$.

Arseniuro di Alluminio

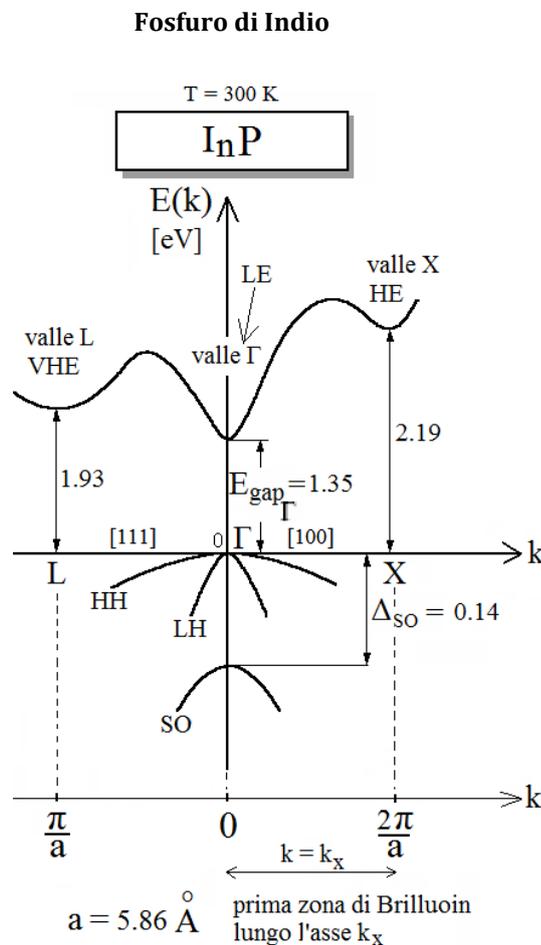


Arseniuro di Indio



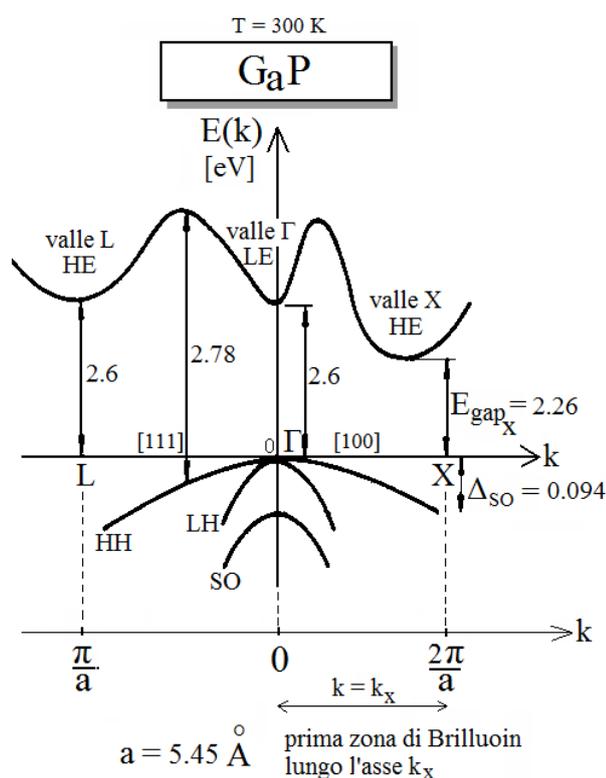
L'applicazione di alte polarizzazioni inverse ad un fotodiiodo di InAs (PIN o APD) può comportare due fenomeni spuri da evitare ai fini della fotorivelazione:

- la formazione, in seguito all'eccessivo piegamento delle bande, di alte correnti di tunneling che si sovrappongono al fotosegnale utile, rendendone impossibile la lettura
- i fotoelettroni di conduzione, altamente energetici a causa del forte campo elettrico imposto dall'esterno, potrebbero raggiungere i punti alti della curva $E(k)$ della valle Γ , i quali, nello spazio dei vettori d'onda k , distano relativamente poco dai punti della curva $E(k)$ della valle L. I fotoelettroni, durante il loro trascinarsi verso il bulk di destinazione di tipo n, potrebbero subire degli scattering di varia natura (dovuti a fononi, impurezze metalliche, impurezze droganti, difetti reticolari ecc...), a causa dei quali i fotoelettroni possono transitare da uno stato iniziale k_i , appartenente alla valle Γ , ad uno stato finale k_f , appartenente alla valle L. Più la polarizzazione inversa imposta sui terminali del fotodiiodo è alta, più la distanza, nello spazio di Fourier, fra lo stato k_i e lo stato k_f diminuisce e maggiore è la probabilità che un fotoelettrone transiti, pertanto, dalla valle Γ alla valle L. Tale transizione comporta un considerevole aumento della massa efficace degli elettroni per la mobilità, e conseguentemente una considerevole diminuzione della mobilità elettronica e quindi dell'ampiezza della fotocorrente utile.



Nel caso di applicazione di alte polarizzazioni inverse ai terminali di un fotodiiodo di InP (PIN o APD) valgono, qualitativamente, le stesse considerazioni fatte nel caso analogo di un fotodiiodo di InAs.

Fosforo di Gallio



L'applicazione di alte polarizzazioni inverse ad un fotodiiodo di GaP (PIN o APD) può comportare due fenomeni, il primo negativo, ai fini della fotorivelazione, il secondo positivo:

- la formazione, in seguito all'eccessivo piegamento delle bande, di alte correnti di tunneling che si sovrappongono al fotosegnale utile, rendendone impossibile la lettura
- i fotoelettroni di conduzione, altamente energetici a causa del forte campo elettrico imposto dall'esterno, potrebbero raggiungere i punti alti della curva $E(k)$ della valle X, i quali, nello spazio dei vettori d'onda k , distano relativamente poco dai punti della curva $E(k)$ della valle Γ . I fotoelettroni, durante il loro trascinarsi verso il bulk di destinazione di tipo n, potrebbero subire degli scattering di varia natura (dovuti a fononi, impurezze metalliche, impurezze droganti, difetti reticolari ecc...), a causa dei quali i fotoelettroni possono transitare da uno stato iniziale k_i , appartenente alla valle X, ad uno stato finale k_f , appartenente alla valle Γ . Più la polarizzazione inversa imposta sui terminali del fotodiiodo è alta, più la distanza, nello spazio di Fourier, fra lo stato k_i e lo stato k_f diminuisce e maggiore è la probabilità che un fotoelettrone transiti, pertanto, dalla valle X alla valle Γ . Tale transizione comporta una diminuzione della massa efficace degli elettroni per la mobilità, con conseguente aumento della mobilità elettronica e quindi dell'ampiezza della fotocorrente.

13) Applicazione del TBM (“Tight Binding Method”) per il calcolo qualitativo della curva di dispersione $E(\mathbf{k})$ della banda di conduzione BC del silicio ed estrapolamento del concetto di massa efficace degli elettroni per la mobilità “longitudinale” m_l^* e “trasversale” m_t^* :

Fissiamo un generico punto all’interno del reticolo diretto di Bravais del silicio: scegliamo l’origine degli assi coordinati, individuata dal vettore posizione $\mathbf{r} = \mathbf{R}_0 = 0$. Questo punto coincide con il punto di simmetria notevole Γ appartenente alla cella di Wigner – Seitz localizzata intorno all’origine degli assi x, y, z , ovvero alla prima cella di Wigner – Seitz del reticolo diretto di Bravais. Attorno alla prima cella di Wigner – Seitz sono posizionate tutte le altre celle, la cui unione costituisce l’intero reticolo diretto di Bravais, ovvero l’intero cristallo di silicio. Ciascuna cella di Wigner – Seitz è individuata da un vettore posizione \mathbf{R} : $\mathbf{r} = \mathbf{R}_0 = 0$ individua la prima cella di Wigner – Seitz, \mathbf{R}_1 individua la seconda cella di Wigner – Seitz, \mathbf{R}_2 la terza, \mathbf{R}_3 la quarta ecc... Supponiamo che nel cristallo di silicio considerato ci siano, oltre alla prima cella di Wigner – Seitz, un numero di celle pari a NC . Una generica cella di Wigner – Seitz, ad esempio individuata dal vettore posizione \mathbf{R}_k , contiene due atomi di silicio, uno individuato dal vettore posizione $\mathbf{R}_k + \mathbf{r}_1 = \mathbf{R}_k$ ($\mathbf{r}_1 = 0$), l’altro individuato dal vettore posizione $\mathbf{R}_k + \mathbf{r}_2$, dove \mathbf{r}_2 giace lungo la direzione cristallografica $[111]$, relativa a ciascuna cella, ovvero $\mathbf{r}_2 = (a/4; a/4; a/4)$, con un sistema di riferimento solidale con la particolare cella presa in considerazione; \mathbf{r}_2 ha modulo pari a $[\sqrt{3/16}]a$, dove a è la costante reticolare del silicio ($a \approx 5.428 \text{ \AA}$). Definiamo lo stato elettronico $|\psi_{\mathbf{k}}(\mathbf{r})\rangle$ presso l’origine Γ degli assi coordinati, cioè in $\mathbf{r} = \mathbf{R}_0 = 0$, avente vettore d’onda \mathbf{k} , nel modo seguente:

$$|\psi_{\mathbf{k}}(\mathbf{r})\rangle = \sum_{\mathbf{R}=\mathbf{R}_0}^{R_{NC}} \sum_{j=1}^2 e^{i(\mathbf{k}\cdot\mathbf{R})} \Phi(\mathbf{r} - \mathbf{R} - \mathbf{r}_j)$$

$|\psi_{\mathbf{k}}(\mathbf{r})\rangle$ è la sovrapposizione dello stato “on – site” $\Phi(\mathbf{r})$, che rappresenta la configurazione elettronica di valenza dell’atomo di Si, appartenente alla prima cella di Wigner – Seitz, posizionato in $\mathbf{r} = \mathbf{R}_0 = 0$ (da cui il termine “on – site”), con lo stato “off – site” $\Phi(\mathbf{r} - \mathbf{r}_2)$, che rappresenta la configurazione elettronica di valenza dell’atomo di Si, sempre appartenente alla prima cella di Wigner – Seitz, posizionato però in \mathbf{r}_2 , e con tutti gli altri stati “off – site” $\exp\{i(\mathbf{k}\cdot\mathbf{R})\}\Phi(\mathbf{r} - \mathbf{R} - \mathbf{r}_j)$ ($\mathbf{R} = \mathbf{R}_1, \mathbf{R}_2, \mathbf{R}_3, \mathbf{R}_4, \dots, \mathbf{R}_k, \dots, \mathbf{R}_{NC}$, $j = 1, 2$), i quali rappresentano le estensioni, gli effetti, presso $\mathbf{r} = \mathbf{R}_0 = 0$, delle configurazioni elettroniche di valenza degli atomi di Si di tutte le altre NC celle di Wigner – Seitz. Come è ben noto il silicio presenta una configurazione elettronica di valenza tipica del quarto gruppo della tavola periodica degli elementi, ovvero del tipo $3s^2 3p^2$. Se i quattro orbitali di valenza del Si non subissero l’ibridazione di tipo sp^3 , ciascuno stato elettronico $\Phi(\mathbf{r} - \mathbf{R} - \mathbf{r}_j)$ assumerebbe la seguente forma:

$$\begin{aligned} \Phi(\mathbf{r} - \mathbf{R} - \mathbf{r}_j) &= \sum_{m=1}^4 c_{mj}(k) \psi_m(\mathbf{r} - \mathbf{R} - \mathbf{r}_j) = \\ &= c_{1j}(k) \psi_{3s}(\mathbf{r} - \mathbf{R} - \mathbf{r}_j) + c_{2j}(k) \psi_{3p_x}(\mathbf{r} - \mathbf{R} - \mathbf{r}_j) + \\ &+ c_{3j}(k) \psi_{3p_y}(\mathbf{r} - \mathbf{R} - \mathbf{r}_j) + c_{4j}(k) \psi_{3p_z}(\mathbf{r} - \mathbf{R} - \mathbf{r}_j) \end{aligned}$$

I quattro vettori $\psi_m(\mathbf{r} - \mathbf{R} - \mathbf{r}_j)$, cioè i quattro orbitali di valenza $3s, 3p_x, 3p_y, 3p_z$, costituiscono una base ortonormale rispetto alla quale sviluppare, sul campo complesso ($c_{mj}(k) \in \mathbb{C}$), lo stato elettronico $\Phi(\mathbf{r} - \mathbf{R} - \mathbf{r}_j)$; per un’approssimazione più accurata di $\Phi(\mathbf{r} - \mathbf{R} - \mathbf{r}_j)$ sarebbe opportuno sviluppare lo stato su una base che tenga conto della degenerazione di spin, ovvero formata dagli otto vettori $3s\uparrow, 3s\downarrow, 3p_x\uparrow, 3p_x\downarrow, 3p_y\uparrow, 3p_y\downarrow, 3p_z\uparrow, 3p_z\downarrow$. Tuttavia i quattro orbitali di valenza dell’atomo di Si subiscono ibridazione sp^3 , pertanto:

$$\begin{aligned} \Phi(\mathbf{r} - \mathbf{R} - \mathbf{r}_j) &= \sum_{n=1}^4 b_{nj}(k) \psi_n(\mathbf{r} - \mathbf{R} - \mathbf{r}_j) = \\ &= b_{1j}(k) \psi_{3[s+p](1)}(\mathbf{r} - \mathbf{R} - \mathbf{r}_j) + b_{2j}(k) \psi_{3[s+p](2)}(\mathbf{r} - \mathbf{R} - \mathbf{r}_j) + \\ &+ b_{3j}(k) \psi_{3[s+p](3)}(\mathbf{r} - \mathbf{R} - \mathbf{r}_j) + b_{4j}(k) \psi_{3[s+p](4)}(\mathbf{r} - \mathbf{R} - \mathbf{r}_j) \end{aligned}$$

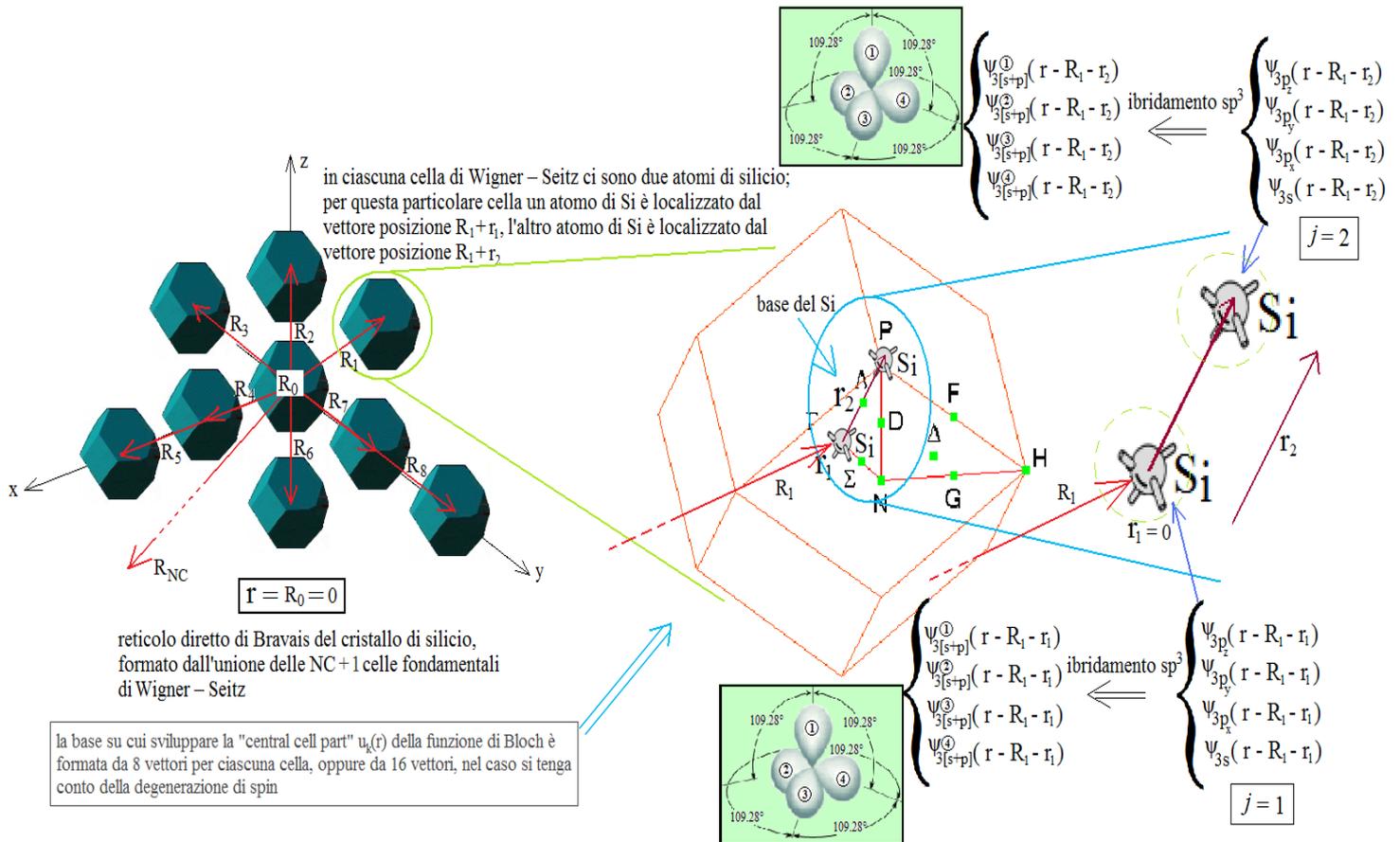
I quattro vettori $\psi_n(\mathbf{r} - \mathbf{R} - \mathbf{r}_j)$ ($n = 1, 2, 3, 4$), cioè i quattro orbitali ibridi di valenza $3[s+p]$, costituiscono la base ortonormale più indicata rispetto alla quale sviluppare, sempre sul campo complesso ($b_{nj}(k) \in \mathbb{C}$), lo stato elettronico $\Phi(\mathbf{r} - \mathbf{R} - \mathbf{r}_j)$; per un’approssimazione più accurata di $\Phi(\mathbf{r} - \mathbf{R} - \mathbf{r}_j)$ sarebbe opportuno sviluppare lo

stato su una base che tenga conto della degenerazione di spin, ovvero formata dagli otto vettori $3[s+p](1)\uparrow$, $3[s+p](1)\downarrow$, $3[s+p](2)\uparrow$, $3[s+p](2)\downarrow$, $3[s+p](3)\uparrow$, $3[s+p](3)\downarrow$, $3[s+p](4)\uparrow$, $3[s+p](4)\downarrow$. Chiaramente il 3 posto davanti al pedice di ibridizzazione $[s+p]$ rappresenta il valore del numero quantico principale che indica il guscio ("shell") energetico più esterno dell'atomo di silicio. Quindi lo stato elettronico $|\psi_k(r)\rangle$ presso l'origine Γ degli assi coordinati x, y, z può essere espresso nel modo seguente:

$$|\psi_k(r)\rangle = \sum_{R=R_0}^{R_{NC}} \sum_{j=1}^2 \sum_{n=1}^4 e^{i(k \cdot R)} b_{nj}(k) \psi_n(r - R - r_j) \quad [g9]$$

Questa espressione di $|\psi_k(r)\rangle$ è nota come "forma di Koster e Slater" e vale per tutti i semiconduttori a struttura reticolare diamantata (C, Si, Ge), e con l'apporto di alcune modifiche anche per i semiconduttori composti a struttura reticolare di tipo zincoblenda (III - V e II - VI). Dalla [g9] risulta evidente che per ciascuna delle $NC + 1$ celle di Wigner - Seitz del cristallo sono necessarie almeno otto funzioni d'onda atomiche $\psi_n(r - R - r_j)$ ($n = 1, 2, 3, 4$ e $j = 1, 2$) per descrivere la "central cell part" dell'onda di Bloch associata ad un generico elettrone del cristallo, cioè la funzione $u_k(r)$, avente periodicità spaziale pari alla costante reticolare "a" del cristallo, che moltiplicata per l'onda piana monocromatica $\exp\{i(k \cdot R)\}$ fornisce l'onda di Bloch dell'elettrone $\psi_{Bloch}(r - R) = u_k(r - R) \exp\{i(k \cdot R)\}$. In formule:

$$u_k(r - R) = \sum_{j=1}^2 \sum_{n=1}^4 b_{nj}(k) \psi_n(r - R - r_j)$$



Più formalmente, tenendo conto della degenerazione di spin, per ciascuna delle $NC + 1$ celle di Wigner - Seitz del cristallo sono necessarie sedici funzioni d'onda atomiche $\psi_n^s(r - R - r_j)$ ($n = 1, 2, 3, 4; j = 1, 2; s = +\frac{1}{2}, -\frac{1}{2}$) per descrivere la "central cell part" dell'onda di Bloch. In formule:

$$u_k(r - R) = \sum_{j=1}^2 \sum_{n=1}^4 \sum_{s=-\frac{1}{2}}^{+\frac{1}{2}} b_{njs}(k) \psi_n^s(r - R - r_j)$$

Scriviamo pertanto l'equazione di Schrödinger indipendente dal tempo:

$$\hat{H} |\psi_k(r)\rangle = E |\psi_k(r)\rangle$$

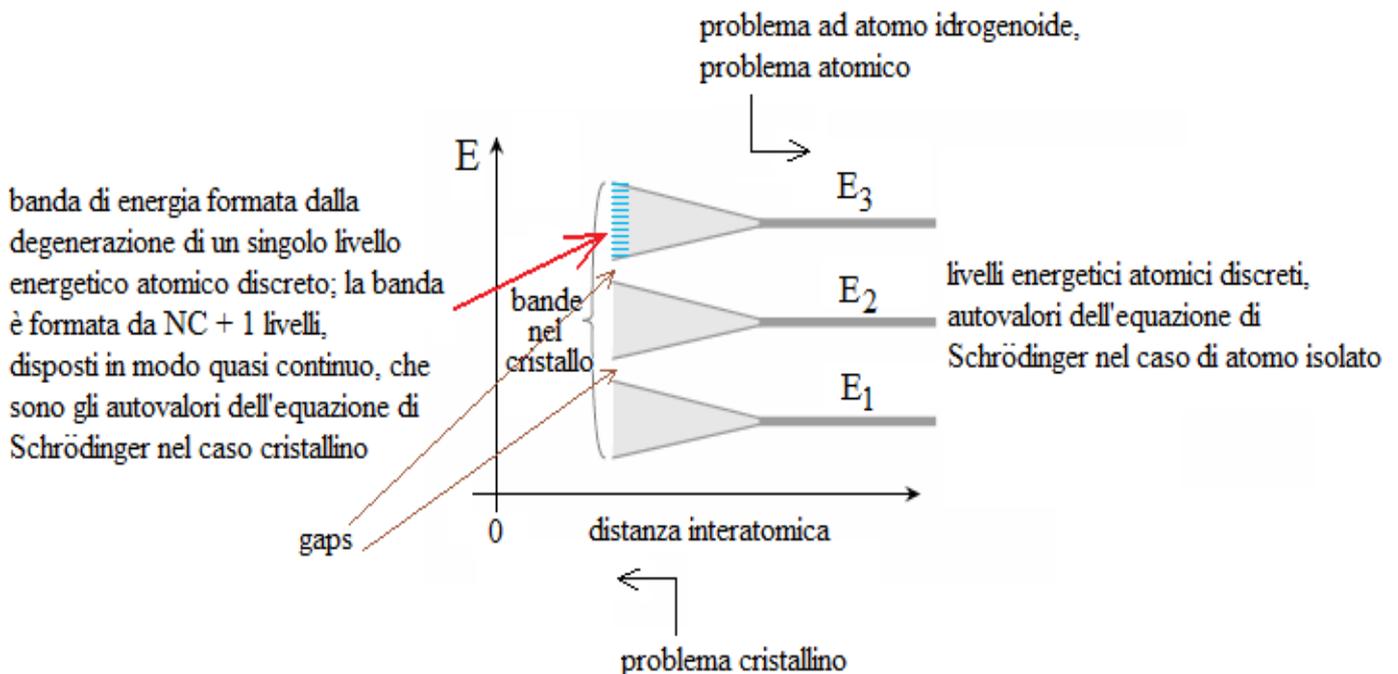
e calcoliamo il determinante secolare che ne deriva, che coincide con il polinomio caratteristico $\gamma(E)$ della matrice hamiltoniana H , imponendolo infine uguale a zero, allo scopo di calcolare gli $NC + 1$ autovalori $E_a(k)$ dell'energia elettronica totale ($a = 1, 2, 3, \dots, NC + 1$) presso il vettore d'onda k . In formule:

$$\gamma(E) = \det[H(k) - E I] = \det \begin{bmatrix} H_{11}(k) - E & H_{12}(k) & \cdots & H_{1(NC+1)}(k) \\ H_{21}(k) & H_{22}(k) - E & & \vdots \\ \vdots & \cdots & \ddots & \\ H_{(NC+1)(NC+1)}(k) - E & & & \end{bmatrix} = 0$$

dove I è la matrice identica. Convenzionalmente il determinante secolare è rappresentato, secondo la notazione di Dirac, nel modo seguente:

$$\| \langle \phi_k^{R \neq R_0}(r - R) | H - E | \psi_k^{R = R_0 = r}(r) \rangle \| = 0 \quad [g10]$$

Il vettore $|\psi_k(r)\rangle$ rappresenta la sovrapposizione di $NC + 1$ stati elettronici di Bloch $u_k(r - R) \exp\{i(k \cdot R)\}$, dove NC è appunto il numero totale di celle "off - site" di Wigner - Seitz del cristallo di silicio, le quali interagiscono elettrostaticamente con la prima cella di Wigner - Seitz, ovvero con la cella "on - site", comportando lo "spreading" (l'allargamento, la degenerazione) di ciascun livello energetico atomico in $NC + 1$ livelli energetici $E_a(k)$, i quali, disposti uno vicino all'altro in modo quasi continuo, costituiscono una banda all'interno del diagramma $E - k$ relativo al silicio. I livelli $E_a(k)$ sono appunto gli $NC + 1$ autovalori dell'energia elettronica totale ($a = 1, 2, 3, \dots, NC + 1$), presso il vettore d'onda k , calcolati risolvendo la [g10].



La risoluzione di [g10], che fornirebbe una forma del diagramma a bande $E(k)$ del silicio abbastanza aderente a quella reale, è molto complicata a livello di calcolo analitico; tuttavia possiamo, solo a titolo di esempio, calcolare una legge di dispersione $E(k)$ di un diagramma a bande del silicio molto qualitativo, relativo alla banda di conduzione BC, assumendo le tre seguenti ipotesi:

- consideriamo un semplice reticolo cristallino di tipo FCC al posto del più complicato reticolo cristallino di tipo diamante, il quale è formato da due cubi FCC, anziché uno solo, con un cubo traslato lungo $[111]$ rispetto all'altro di un quarto della lunghezza della diagonale principale. In altre parole rimuoviamo la

degenerazione di funzioni d'onda per singola cella di Wigner – Seitz, relativa all'indice j ($= 1, 2$), ovvero assumiamo che ciascuna cella possieda un solo atomo di base, anzichè due.

- consideriamo un singolo orbitale $\psi_n(r - R)$ elettronico ibrido di valenza $3[s+p]$ per ciascun atomo di silicio, anzichè quattro orbitali ibridi $3[s+p]$. In altre parole rimuoviamo la degenerazione di funzioni d'onda atomiche per ciascun atomo di Si, relativa all'indice n ($n = 1, 2, 3, 4$): pertanto assumiamo che n possa assumere soltanto un valore.
- consideriamo unitari, presso qualunque vettore d'onda $k = (k_x; k_y; k_z)$, i coefficienti $b_n(k)$.

La risoluzione di [g10] fornisce il seguente risultato:

$$E(k) - E_{3[s+p]} = -[E(k) - E_{3[s+p]}] \sum_{R=R_1}^{R_{NC}} \int \psi_{3[s+p]}^*(r) \psi_{3[s+p]}(r - R) e^{i(k \cdot R)} d^3r + \\ + \int \psi_{3[s+p]}^*(r) \Delta U(r) \psi_{3[s+p]}(r) d^3r + \\ + \sum_{R=R_1}^{R_{NC}} \int \psi_{3[s+p]}^*(r) \Delta U(r) \psi_{3[s+p]}(r - R) e^{i(k \cdot R)} d^3r \quad [g11]$$

Convenzionalmente si utilizzano i seguenti simboli, al fine di presentare l'espressione di $E(k)$, mostrata nella [g11], in modo più compatto:

$$\alpha(R) = \int \psi_{3[s+p]}^*(r) \psi_{3[s+p]}(r - R) d^3r$$

$$\beta_{3[s+p]} = - \int \psi_{3[s+p]}^*(r) \Delta U(r) \psi_{3[s+p]}(r) d^3r$$

$$\gamma(R) = - \int \psi_{3[s+p]}^*(r) \Delta U(r) \psi_{3[s+p]}(r - R) d^3r$$

dove:

$E_{3[s+p]}$ = livello energetico discreto dell'orbitale ibrido $3[s+p]$ nel caso di atomo isolato; da questo singolo livello atomico si forma una banda di dispersione $E(k)$, nel caso in cui l'atomo sia collocato all'interno di un cristallo.

$\Delta U(r)$ = perturbazione dell'energia potenziale coulombiana nel punto del reticolo di Bravais individuato dal vettore posizione $r = R_0 = 0$. Tale perturbazione è dovuta all'interazione elettrostatica fra il potenziale prodotto dall'atomo "on - site", cioè posizionato in $r = R_0 = 0$ (origine Γ degli assi x,y,z), ed i potenziali prodotti dagli atomi "off - site", cioè posizionati nei punti individuati dai vettori posizione $R \neq R_0$ ($R_1, R_2, R_3, \dots, R_k, \dots, R_{NC}$). $\Delta U(r)$ va sommata al potenziale atomico coulombiano $U(r)$ per ottenere l'energia potenziale di cristallo nel punto Γ .

$\alpha(R)$ = "overlap integral", cioè prodotto scalare, fra lo stato elettronico $\psi_{3[s+p]}(r)$ e quello $\psi_{3[s+p]}(r - R)$, con $R \neq R_0$, cioè $R \neq 0$. $\alpha(R) = 0$ in virtù dell'ipotesi analitica che è alla base del TBM, dalla quale discende il nome stesso di questo metodo di calcolo delle bande, ovvero che ciascuno stato elettronico $\psi_{3[s+p]}$ è strettamente legato al nucleo dell'atomo di appartenenza, ossia la densità di probabilità di trovare quello stato, quell'elettrone, presso un qualunque altro sito atomico è trascurabile. Pertanto date $NC + 1$ celle fondamentali del cristallo, le $NC + 1$ funzioni elettroniche atomiche $\psi_{3[s+p]}$, ciascuna strettamente concentrata intorno al nucleo dell'atomo di appartenenza, costituiscono una base ortonormale sulla quale è possibile sviluppare lo stato di un qualunque elettrone del cristallo.

$\beta_{3[s+p]}$ [eV] = "on - site integral" = alterazione dell'energia elettronica totale subita dal livello energetico discreto $E_{3[s+p]}$, nel punto del reticolo di Bravais individuato dal vettore posizione $r = R_0 = 0$. Questa alterazione, determinata da $\Delta U(r)$, è dovuta all'interazione elettrostatica fra l'atomo "on - site", posizionato nell'origine degli assi x,y,z , e gli atomi "off - site", posizionati nei punti individuati dai vettori posizione $R \neq R_0$ ($R_1, R_2, R_3, \dots, R_k, \dots, R_{NC}$).

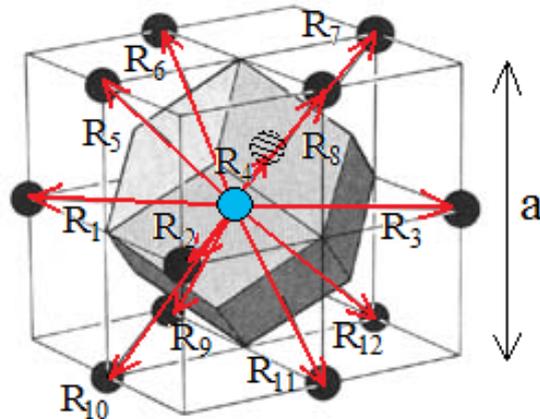
$\gamma(R)$ [eV] = "off - site integral" = energia proporzionale all'intensità dell'interazione elettrostatica fra lo stato elettronico "on - site" $\psi_{3[s+p]}(r)$ e quello "off - site" $\psi_{3[s+p]}(r - R)$, per $R \neq R_0$, cioè $R \neq 0$. Lo stato $\psi_{3[s+p]}(r)$ può

essere visto come lo stato iniziale $|i\rangle$ dal quale può partire un elettrone di conduzione intrinseco, estrinseco oppure un fotoelettrone, mentre lo stato $\psi_{3[s+p]}(r-R)$ come lo stato finale $|f\rangle$ al quale può arrivare un elettrone di conduzione. Pertanto $\gamma(R)$ è strettamente legata alla conducibilità di un cristallo. $\gamma(R) = \gamma(-R)$, ovvero γ presenta simmetria pari nel caso di reticolo FCC. È possibile dimostrare che se $ka \ll 1$ (k = vettore d'onda, a = costante reticolare del Si) e se ciascun atomo del reticolo FCC interagisce unicamente con i suoi dodici primi vicini, mentre l'interazione con gli altri atomi (secondi vicini, terzi vicini ecc...) del cristallo è trascurabile, una possibile definizione di massa efficace dell'elettrone per la mobilità è la seguente:

$$m_e^* = \frac{\hbar^2}{2\gamma a^2} \quad [g12]$$

dove $\gamma = \gamma(R_1) = \gamma(R_2) = \dots = \gamma(R_{12}) = \gamma(R)$, con $R_1 = R_2 = \dots = R_{12}$, è l'intensità dell'interazione fra l'atomo "on - site", in $r = R_0 = 0$, e i dodici atomi "off - site" primi vicini. Questi sono equidistanti dall'atomo "on - site", pertanto i corrispondenti dodici "off - site integrals" sono uguali fra loro e pari a γ . In altre parole l'interazione coulombiana fra l'atomo "on - site" ed il primo vicino individuato da R_1 è uguale all'interazione coulombiana fra l'atomo "on - site" ed il primo vicino individuato da R_2 , che è uguale all'interazione coulombiana fra l'atomo "on - site" ed il primo vicino individuato da R_3 ecc... fino ad R_{12} . Le interazioni fra l'atomo "on - site" e qualunque altro atomo (più lontano dei dodici primi vicini) sono trascurabili ($\gamma \approx 0$). Per cristalli FCC come Au, Ag, Cu, Al e per semiconduttori la cui struttura cristallina è basata su quella FCC, come C, Ge, Si e composti III - V e II - VI, γ è dell'ordine di 1 eV.

struttura reticolare di tipo FCC, che è alla base delle strutture cristalline di semiconduttore di tipo diamante e di tipo zincoblenda



- = atomo individuato dal vettore posizione $r = R_0 = 0$, cioè atomo "on - site"
- = 12 atomi primi vicini individuati dai vettori posizione R_1, R_2, \dots, R_{12} , cioè i 12 atomi "off - site" che interagiscono in modo significativo con l'atomo "on - site"

posizioni dei 12 atomi primi vicini:

$$\frac{a}{2}(\pm 1, \pm 1, 0) ; \frac{a}{2}(\pm 1, 0, \pm 1) ; \frac{a}{2}(0, \pm 1, \pm 1)$$

I parametri energetici $E_{3[s+p]}$, $\beta_{3[s+p]}$, $\gamma(R)$ ed il parametro di overlap (numero puro) $\alpha(R)$ sono quattro parametri di fitting nel metodo TBM: noto il cristallo di cui desideriamo calcolare la dispersione $E(k)$ di una banda del diagramma $E - k$, sono disponibili delle tabelle nelle quali sono riportati i valori dei parametri. Pertanto mediante un software dedicato si assegnano i valori ai quattro parametri, presenti nella [g11], e si esegue il programma, il quale di fatto esegue la [g11]: il programma fornisce le curve $E(k)$ (una per ogni valore del livello energetico di atomo isolato $E_{3[s+p]}$) che formano il diagramma a bande $E - k$ del materiale.

Dalla [g12] si evince che se l'atomo a cui appartiene un elettrone $\psi(r)$ è di fatto isolato, cioè non interagente con nessun altro atomo di un cristallo, atomo del cristallo distante R dall'atomo isolato e avente uno stato elettronico $\psi(r - R)$, allora fra l'atomo isolato ed il resto del cristallo non c'è alcuna interazione significativa, pertanto $\gamma(R) \approx 0$, quindi $m_e^* \rightarrow \infty$ e la mobilità elettronica $\mu_n \approx 0$. La concavità dell'arco inferiore, presso il minimo E_c , della BC è di fatto nulla, ovvero la banda di conduzione presenta una curva di dispersione $E(k)$ praticamente costante rispetto a k , simile ad un singolo livello energetico discreto; questa osservazione concorda con l'ipotesi di atomo isolato. Pertanto non si ha conduzione di corrente elettrica fra lo stato iniziale, di partenza, $\psi(r)$ e quello finale, di arrivo, $\psi(r - R)$. Maggiore è l'intensità γ dell'interazione fra l'atomo, prima isolato, ed il resto del cristallo, e minore è m_e^* e di conseguenza maggiore è la mobilità elettronica μ_n . Al crescere di γ la concavità dell'arco inferiore, presso il minimo E_c , della dispersione $E(k)$ della BC aumenta, cioè la larghezza della BC aumenta, ovvero la BC presenta una curva di dispersione $E(k)$ sempre più concava, sempre più dissimile ad un singolo livello energetico discreto: questa osservazione concorda con il crescente allontanamento dalla condizione di atomo isolato. Pertanto al crescere di γ si ha un incremento della conduzione di corrente elettrica fra lo stato iniziale $\psi(r)$ e quello finale $\psi(r - R)$.

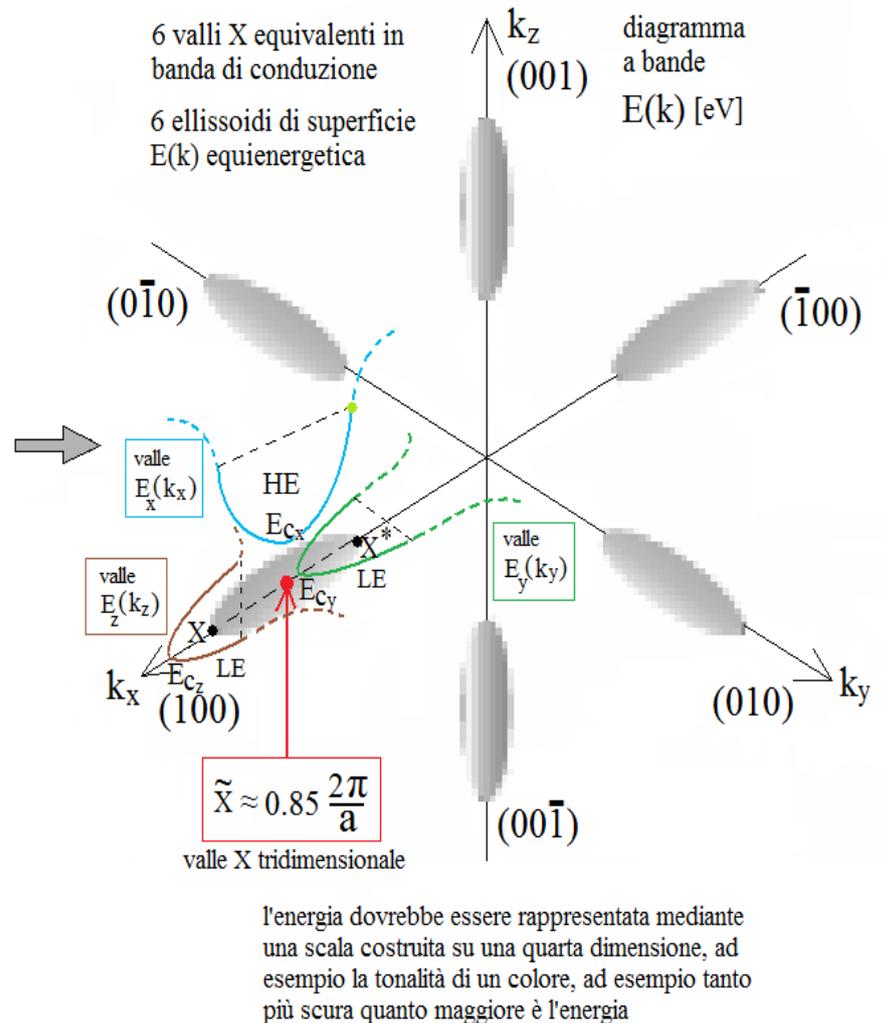
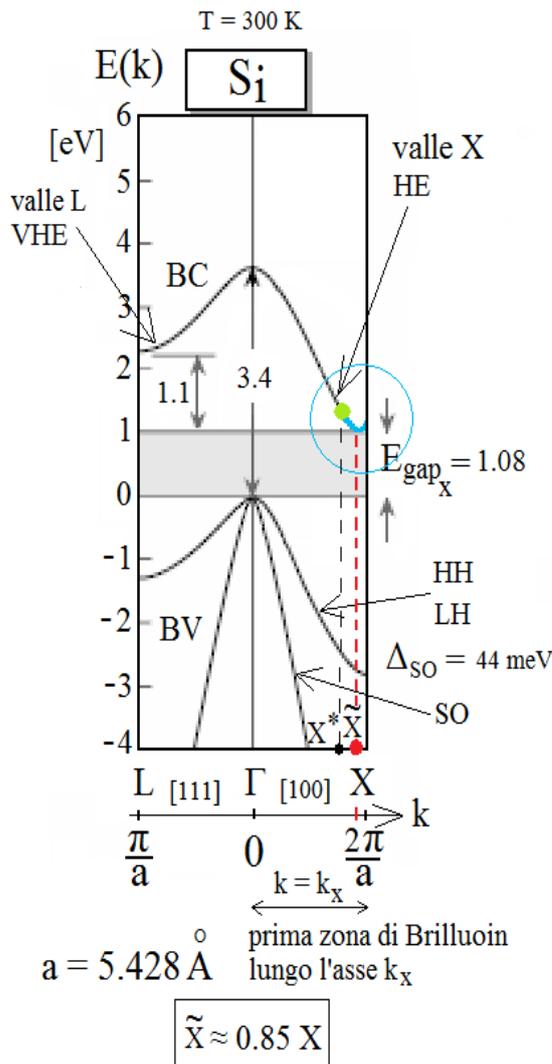
L'espressione di $E(k)$ riportata in [g11] assume la semplice forma mostrata nella [g13], nell'ipotesi in cui $\gamma(R) \neq 0$ solo per le interazioni con i dodici atomi primi vicini di quello posizionato nell'origine degli assi x,y,z .

$$E(k) = E_{3[s+p]} - \beta_{3[s+p]} - \sum_{R=R_1}^{R_{12}} \gamma(R) e^{i(k \cdot R)} \quad [g13]$$

Con semplici calcoli algebrici si arriva alla seguente espressione approssimata della curva di dispersione $E(k)$, relativa alla banda di conduzione di un reticolo cristallino FCC; tale dispersione $E(k)$ si può adattare, seppur molto qualitativamente, a quella relativa alla banda di conduzione di un cristallo semiconduttore come il silicio.

$$E(k) = E_{3[s+p]} - \beta_{3[s+p]} - 4\gamma \left[\cos\left(\frac{k_x a}{2}\right) \cos\left(\frac{k_y a}{2}\right) + \cos\left(\frac{k_y a}{2}\right) \cos\left(\frac{k_z a}{2}\right) + \cos\left(\frac{k_z a}{2}\right) \cos\left(\frac{k_x a}{2}\right) \right]$$

Osservando il diagramma a bande $E - k$ del silicio, ad esempio costruito applicando il TBM in modo meno approssimato, cioè evitando di implementare nei calcoli le tre approssimazioni precedentemente discusse, è possibile notare che la curva di dispersione $E(k)$ della banda di conduzione BC presenta sei valli X equivalenti, i cui punti di minimo E_c sono localizzati, lungo i semiassi positivi e negativi k_x , k_y , k_z , presso il punto $\bar{X} \approx 0.85[2\pi/a]$. Ciascuna delle sei valli X equivalenti, nel dintorno del proprio punto centrale \bar{X} , presenta delle gole, delle vallate, lungo ciascuna delle tre direzioni k_x , k_y , k_z sulle quali si sviluppa il reticolo reciproco di Fourier. In altre parole ciascuna valle X presenta una dispersione $E_x(k_x)$ concava, a forma di valle, lungo l'asse k_x , una dispersione $E_y(k_y)$ concava lungo l'asse k_y ed una $E_z(k_z)$ concava lungo k_z : intorno a ciascuno dei sei punti \bar{X} il diagramma a bande $E - k$ del silicio presenta una dispersione $E(k)$ a forma di valle tridimensionale.



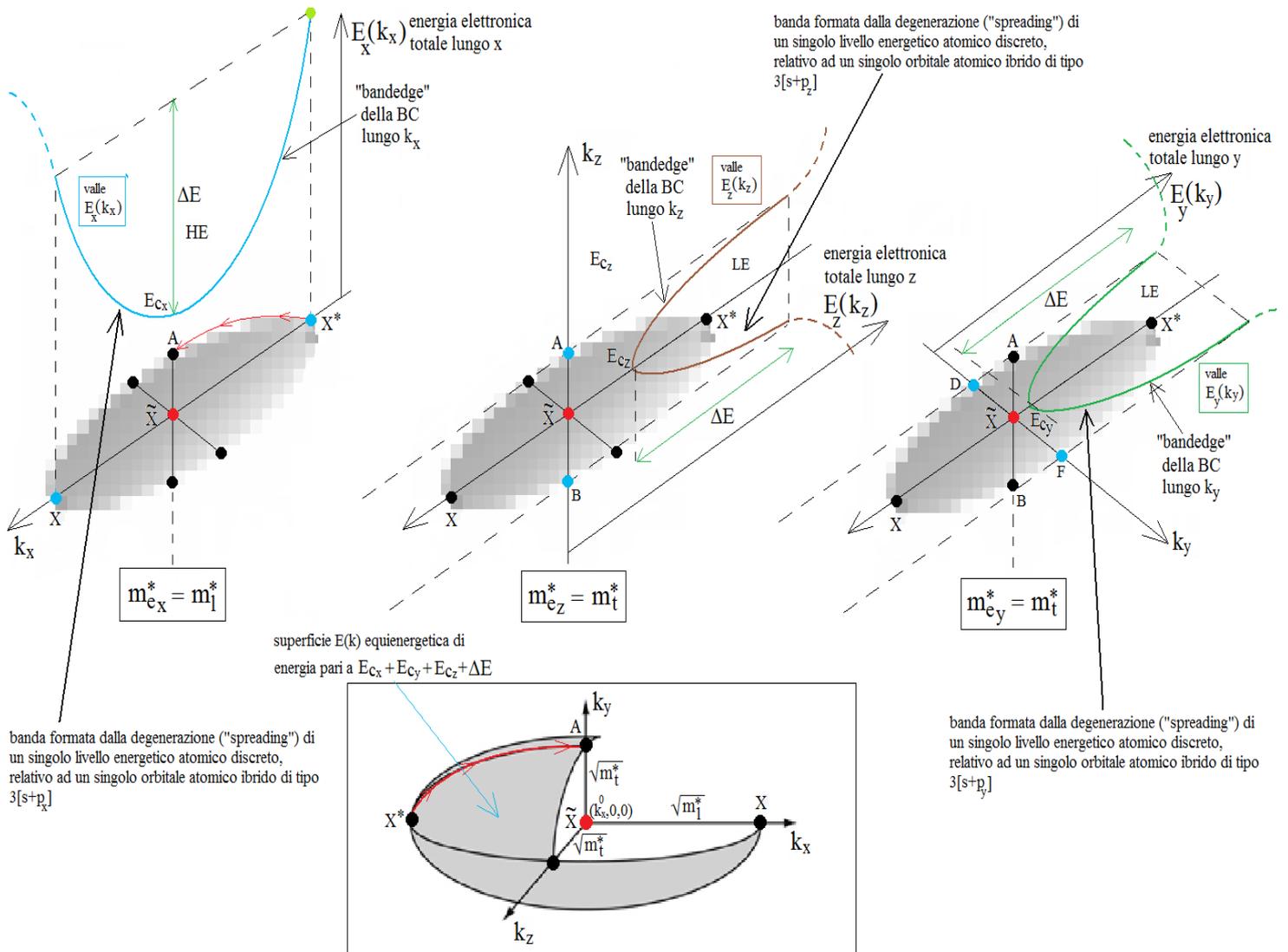
Consideriamo la valle X localizzata intorno al punto \tilde{X} appartenente al semiasse positivo k_x , ovvero la valle X tridimensionale posizionata nella regione del diagramma E - k evidenziata, in figura, dal cerchio azzurro. Posizioniamoci esattamente sull'asse k_x , ovvero assumiamo $k_y = k_z = 0$, e scorriamo l'asse k_x nel suo verso positivo a partire dal punto Γ , cioè da $k_x = 0$, fino ad incontrare il punto $X^* \approx 0.7[2\pi/a]$. Presso il vettore d'onda k_x individuato dal punto X^* la curva di dispersione $E_x(k_x)$, ovvero l'energia elettronica totale lungo l'asse x, cioè in funzione di k_x , assume il valore evidenziato dal punto colorato di verde chiaro. Scorrendo ulteriormente l'asse k_x osserviamo che la dispersione $E_x(k_x)$ assume la forma di una valle, la quale costituisce, presso il punto \tilde{X} , il "bandedge" della BC lungo k_x , il cui minimo è E_{Cx} . Presso il minimo ipotizziamo di poter considerare approssimativamente parabolica la dispersione $E_x(k_x)$, pertanto possiamo definire la massa efficace m_{ex}^* dell'elettrone di conduzione per la mobilità, lungo l'asse x, nel modo seguente:

$$m_{ex}^* = \frac{\hbar^2}{\left. \frac{d^2 E_x(k_x)}{dk_x^2} \right|_{\tilde{X}}} = m_1^* \approx 0.98 m_0$$

dove m_1^* è la massa efficace dell'elettrone di conduzione, per la mobilità, "longitudinale", che nel silicio, presso il punto \tilde{X} , è abbastanza alta, quasi pari alla massa dell'elettrone libero. In altre parole l'elettrone lungo l'asse x, presso il punto \tilde{X} , possiede una grande inerzia, cioè risulta abbastanza difficile da accelerare per mezzo di un campo elettrico di polarizzazione esterna ϵ_x , ovvero l'elettrone ha una grande massa efficace $m_{ex}^* = m_1^*$: infatti la concavità della dispersione $E_x(k_x)$, presso \tilde{X} , è abbastanza bassa. Percorrendo l'asse k_x da X^* a X possiamo osservare che la dispersione parziale lungo k_x , cioè $E_x(k_x)$, passa da un valore alto $E_x(X^*)$ al suo valore minimo $E_x(\tilde{X}) = E_{Cx}$, per poi tornare nuovamente ad un valore alto $E_x(X) = E_x(X^*)$, dove $E_x(X) - E_{Cx} = \Delta E$. ΔE , che è

proporzionale all'integrale di off - site $\gamma(R)$, è la larghezza del "bandedge" della BC lungo k_x . Pertanto presso i punti X^* ed X , lungo l'asse k_x , l'energia elettronica totale $E_x(k_x)$ lungo x è uguale e abbastanza alta; presso i punti X^* ed X l'energia elettronica cinetica lungo x è uguale e pari a ΔE . Percorrendo l'asse k_x da X^* a \tilde{X} l'energia $E_x(k_x)$ diminuisce. Tuttavia se durante il percorso da X^* a \tilde{X} incrementiamo anche il vettore d'onda k_z , ovvero scorriamo contemporaneamente l'asse k_z nel suo verso positivo, partendo da $k_z = 0$, osserviamo che mentre $E_x(k_x)$ diminuisce l'energia elettronica totale $E_z(k_z)$ lungo l'asse z aumenta, a partire dal suo valore minimo $E_z(k_z = 0) = E_{Cz}$. Se partiamo dal punto X^* e ci muoviamo verso il punto A lungo il piano $(k_x, 0, k_z)$, in modo da formare un arco di ellisse $\overline{X^*A}$, osserviamo il decremento di $E_x(k_x)$ e contemporaneamente l'incremento di $E_z(k_z)$, che compensa la diminuzione progressiva di $E_x(k_x)$. Arrivati nel punto A , notiamo che $E_x(k_x)$ è decrementata rispetto al proprio valore presso il punto X^* di un'energia pari a ΔE , mentre $E_z(k_z)$ è incrementata della stessa quantità ΔE rispetto al proprio valore minimo E_{Cz} . Pertanto lungo l'arco di ellisse $\overline{X^*A}$ l'energia totale elettronica è costante e pari a $E_{Cx} + \Delta E + E_{Cz} + E_{Cy}$, dove E_{Cy} è il minimo valore dell'energia elettronica totale lungo y presso la valle X , ovvero il minimo valore assunto dalla dispersione $E_y(k_y)$. Con ragionamenti del tutto analoghi a quelli svolti finora è possibile dimostrare che tutta la superficie dell'ellissoide, e non solo l'arco $\overline{X^*A}$, presso una qualunque delle sei valli X equivalenti, costituisce una superficie $E(k) = E(k_x, k_y, k_z)$ equienergetica così definita:

$$E(k) = E(k_x, k_y, k_z) = \frac{\hbar^2 k_x^2}{2m_{ex}^*} + \frac{\hbar^2 k_y^2}{2m_{ey}^*} + \frac{\hbar^2 k_z^2}{2m_{ez}^*} = E_{Cx} + E_{Cy} + E_{Cz} + \Delta E = \text{costante}$$



Osservando la figura si può notare, come già spiegato, che percorrendo l'asse k_x dal punto X^* a quello X , la curva di dispersione $E_x(k_x)$ passa da un valore alto $E_{C_x} + \Delta E$ al valore minimo E_{C_x} , per poi tornare a quello alto $E_{C_x} + \Delta E$. Questa escursione energetica di $E_x(k_x)$ avviene fra X^* e X , ovvero in un intervallo, entro un segmento, abbastanza lungo dell'asse k_x (tale segmento è lungo circa $0.3[2\pi/a]$). Quindi la concavità della dispersione totale $E(k)$ lungo l'asse k_x , cioè la concavità della dispersione parziale $E_x(k_x)$, è abbastanza bassa, comportando una massa efficace longitudinale m_l^* dell'elettrone in BC abbastanza grande ($m_l^* \approx 0.98 m_0$). Percorrendo l'asse k_z dal punto B a quello A la curva di dispersione parziale $E_z(k_z)$ passa da un valore alto $E_{C_z} + \Delta E$ al valore minimo E_{C_z} , per poi tornare a quello alto $E_{C_z} + \Delta E$. Questa escursione energetica di $E_z(k_z)$, sempre di ΔE , avviene fra B ed A, ovvero lungo un segmento dell'asse k_z decisamente più corto del segmento, lungo l'asse k_x , compreso fra X^* e X , a causa della forma dell'ellissoide, che è allungato nella direzione k_x . La concavità della dispersione $E_z(k_z)$, presso \tilde{X} , è quindi decisamente marcata. Possiamo definire la massa efficace m_{ez}^* dell'elettrone di conduzione per la mobilità, lungo l'asse z , nel modo seguente:

$$m_{ez}^* = \frac{\hbar^2}{\left. \frac{d^2 E_z(k_z)}{dk_z^2} \right|_{\tilde{X}}} = m_t^* \approx 0.19 m_0$$

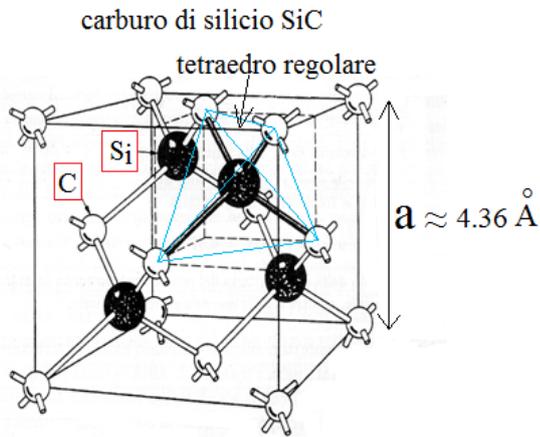
dove m_t^* è la massa efficace dell'elettrone di conduzione, per la mobilità, "trasversale" lungo z , che nel silicio, presso il punto \tilde{X} , è considerevolmente piccola. In altre parole l'elettrone lungo l'asse z , presso il punto \tilde{X} , possiede una piccola inerzia, cioè risulta abbastanza facile da accelerare per mezzo di un campo elettrico di polarizzazione esterna ϵ_z , ossia è molto reattivo nei confronti degli stimoli elettrostatici imposti dall'esterno.

Ragionamenti del tutto analoghi possono essere svolti nel caso della dispersione parziale $E_y(k_y)$ e della sua concavità. Possiamo pertanto definire anche la massa efficace m_{ey}^* dell'elettrone di conduzione per la mobilità, lungo l'asse y :

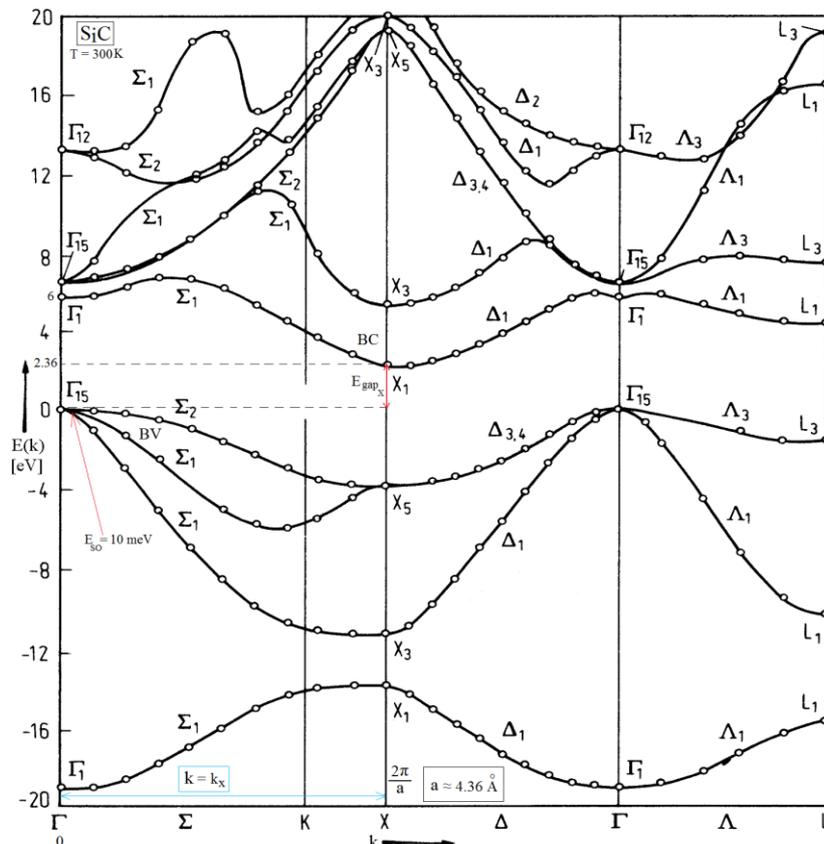
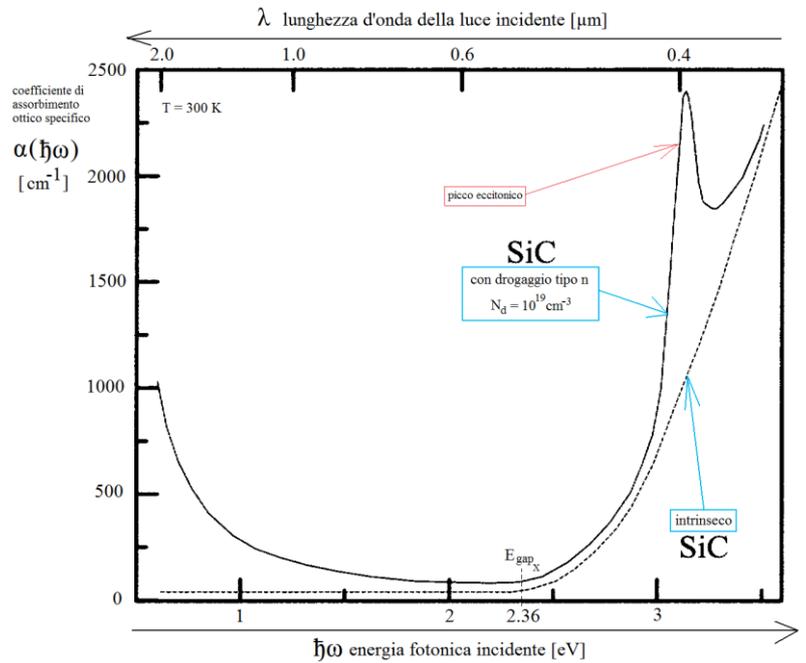
$$m_{ey}^* = \frac{\hbar^2}{\left. \frac{d^2 E_y(k_y)}{dk_y^2} \right|_{\tilde{X}}} = m_t^* \approx 0.19 m_0$$

dove m_t^* è la massa efficace dell'elettrone di conduzione, per la mobilità, "trasversale" lungo y , che nel silicio, presso il punto \tilde{X} , è all'incirca uguale a quella trasversale lungo z . Quindi per il silicio quando si parla di massa efficace dell'elettrone di conduzione, per la mobilità, "trasversale" si intende indistintamente, con buona approssimazione, sia quella valutata lungo k_z che quella valutata lungo k_y .

14) Struttura cristallina, spettro $\alpha(\hbar\omega)$ del coefficiente di assorbimento ottico specifico e diagramma a bande di energia $E - k$ del carburo di silicio SiC (utilizzato ad esempio per l'integrazione micrometrica della resistenza di quenching R_0 nelle microcelle Geiger dei SiPMs), cristallizzato secondo la geometria zincoblenda:



la cella fondamentale del reticolo diretto di Bravais di tipo zincoblenda, cioè la cella di Wigner - Seitz, conta due atomi, uno di C ed uno di Si, che formano la base, uno posizionato in $(0,0,0)$, l'altro in $(a/4, a/4, a/4)$



Il carburo di silicio SiC può cristallizzare in oltre 200 modi diversi, seguendo geometrie reticolari policristalline di varia natura, ad esempio di tipo zincoblenda, HCP, romboedrico ecc...