UNIVERSITÀ DEGLI STUDI DI NAPOLI FEDERICO II

FACOLTÀ DI INGEGNERIA

DOTTORATO DI RICERCA IN INGEGNERIA CHIMICA, DEI MATERIALI E DELLA PRODUZIONE
XVIII CICLO

ANALISI DI MATERIALI SEMICONDUTTORI IN SILICIO-GERMANIO E CARATTERIZZAZIONE DI TRANSISTORI BIPOLARI IN TALE TECNOLOGIA

Coordinatore del corso di Dottorato:
Prof. Nino Grizzuti

Candidato:
Ing. Michele Portico Ambrosio

Tutor:
Prof. Giuseppe Mensitieri

Co-tutor:
Prof. Niccolò Rinaldi

ANNO ACCADEMICO 2005-2006

Sommario

Alla mia famiglia

CAPITOLO 1 – Introduzione

L'ambito delle tecnologie elettroniche per i dischi rigidi (comunemente indicati in inglese come *Hard Disk Drive* o HDD) è estremamente competitivo ed in continuo mutamento. I parametri fondamentali degli HDD – quali la densità di registrazione, il numero totale di *byte* per piatto, il tempo di accesso, nonché il costo per *Megabyte* – hanno seguito una vertiginosa evoluzione. Ciò è stato reso possibile grazie ad una continua ricerca sui materiali magnetici, sulle testine di lettura e scrittura, ed infine grazie al fondamentale apporto delle metodologie più avanzate di processamento di segnale, implementate in circuiti integrati sempre più complessi e veloci. In figura 1.1 (fonte IBM) è illustrato l'incremento della densità di registrazione (*areal density*) negli ultimi quattro decenni.

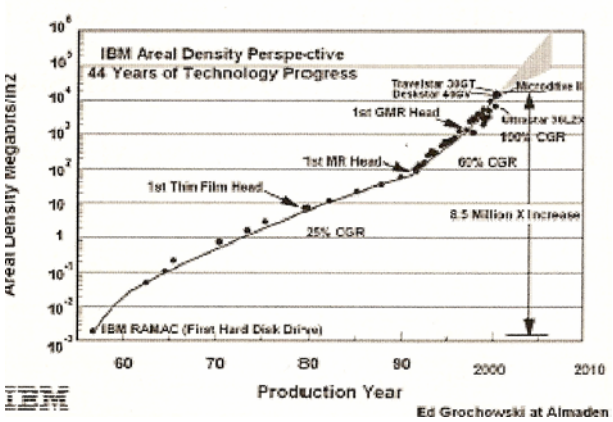

Figura 1.1

Come si nota, la densità di registrazione è passata da circa 2kbit per pollice quadrato – offerti dal primo disco rigido, che fu inventato da IBM verso la metà degli anni cinquanta – ai 100Gbit per pollice quadrato dell'ultima generazione.

Parallelamente, la velocità di trasferimento dei dati da/verso il disco (*Data Rate*) è cresciuta notevolmente negli anni, come riportato in figura 1.2.

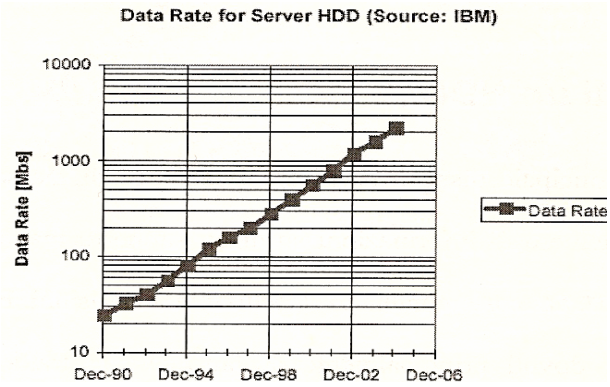

Figura 1.2

Sia l'aumento dell'*areal density* sia l'aumento della velocità di lettura e scrittura implicano numerosi problemi di fattibilità, per tutti i dispositivi elettronici che costituiscono il sistema. Nella figura 1.3 è descritto lo schema a blocchi della componentistica elettronica di un *hard disk drive*.

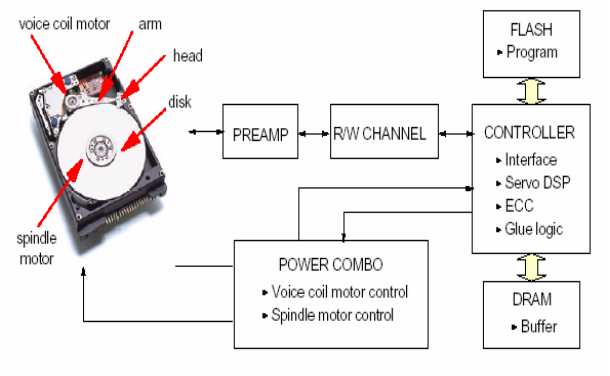

Figura 1.3

L'oggetto di questa attività di ricerca, in particolare, è il primo circuito che si trova ad interfacciare la testina, ovvero il preamplificatore.

1.1 Struttura di un HDD

L'Hard Disk è il principale dispositivo di archiviazione di "massa" che permette di registrare un gran numero di informazioni su di un supporto magnetico riutilizzabile, relativamente stabile e dal costo contenuto rispetto alle dimensioni. Esso presenta alcuni problemi intrinseci, dovuti principalmente al fatto che si tratta di un dispositivo meccanico con parti in movimento soggette ad usura e molto delicate, e quindi la possibilità di guasti è maggiore rispetto ad un dispositivo a stato solido.

Risale al 1957 la prima implementazione, da parte dell'IBM, di un dispositivo utilizzante dischi di alluminio rivestiti di materiale magnetizzabile; era formato da 50 dischi del diametro di 24 pollici che andavano a formare una capacità di 5Mbyte.

Al giorno d'oggi si è arrivati ad avere dischi del diametro di 1.8, 2.5 e 3.5 pollici e capacità dell'ordine dei Gbyte.

All'interno della struttura Hard Disk sono sistemati uno o più dischi denominati "piatti", generalmente realizzati in leghe di alluminio, impilati uno sopra l'altro ad una certa distanza tra loro e costantemente in rotazione attorno ad un asse. Le testine di lettura/scrittura sono disposte su di un braccetto e possono muoversi lungo la superficie utile, come si può vedere nella figura 1.4.

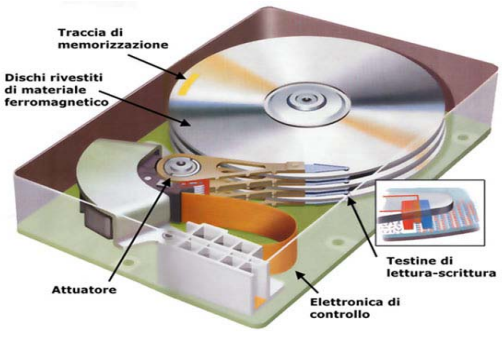

Figura 1.4

I dischi ruotano in senso antiorario ad un regime di rotazione che va da 5400 a 15000 giri al minuto (RPM); in questo modo i dati in esso registrati sono immediatamente disponibili, letti e scritti alla maggiore velocità possibile. Ogni disco ha due lati utilizzabili e su ciascuno si muove la testina. Tutte le testine sono rigidamente connesse insieme e allineate le une con le altre in modo da formare una specie di pettine che effettua le operazioni di lettura/scrittura contemporaneamente su tutte le superfici utilizzate dal dispositivo (figura 1.5).

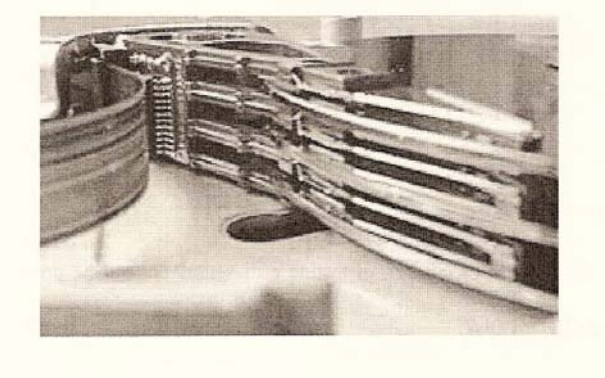

Figura 1.5

Il posizionamento delle testine sui dischi è effettuato da un servomotore particolarmente preciso comandato dalla logica di gestione del disco che ne verifica costantemente posizione, velocità ed accelerazione.

I piatti sono ricoperti da un sottile strato di materiale magnetizzabile; la rotazione dei dischi fa si che si formi un sottilissimo cuscino d'aria sotto la testina, che in effetti "plana" sulla superficie del disco senza toccarlo, permettendo una lunghissima durata del rivestimento magnetico del piatto.

Per scrivere i dati sul piatto, un sensore induttivo presente sulla testina viene percorso da una corrente elettrica e produce una variazione della magnetizzazione sulla superficie del disco.

Per rileggere informazioni viene utilizzato un sensore magneto-resistivo che rileva le variazioni della magnetizzazione e le riconverte in una tensione elettrica che viene amplificata dal preamplificatore e convertita in un segnale digitale che verrà elaborato dalla successiva circuiteria per ricostruire i dati.

I dischi rigidi archiviano i dati in "tracce" concentriche; ogni traccia è ulteriormente divisa in "settori" e il gruppo di tracce su ogni lato di ogni piatto corrispondente alle posizioni allineate del gruppo testine è denominato "cilindro", come si può vedere in figura 1.6.

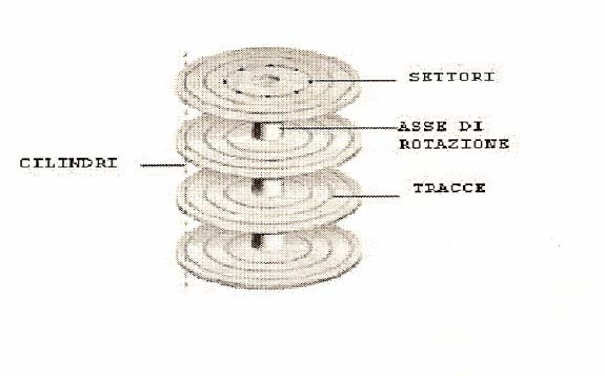

Figura 1.6

L'insieme del numero di cilindri, di testine e di settori per traccia prende il nome di geometria dell'Hard Disk.

Per una corretta prestazione di un disco rigido devono essere rispettati alcuni parametri; quello più importante è "il regime di rotazione del disco"; più grande è, più veloce risulta essere il disco ma causa degli svantaggi perché si ha un maggior rumore ed emissione di calore per attrito e una minore affidabilità nel tempo. Un altro parametro importante è il "tempo di accesso medio" (*Access Time*) che è composto dal "tempo di ricerca " (*Seek Time*), che corrisponde al tempo che il dispositivo impiega a passare da una traccia

all'altra, e dal "tempo di latenza", che è il tempo che la testina deve aspettare, una volta che si è posizionata sulla traccia, per poter leggere il settore giusto.

Infine, occorre considerare la "densità di memorizzazione" che è data dal rapporto tra la capacità totale del disco rigido e il numero di piatti di cui è composto, e il "Data Rate" che corrisponde alla velocità con cui vengono trasmessi i dati, sia in lettura sia in scrittura, nell'unità di tempo.

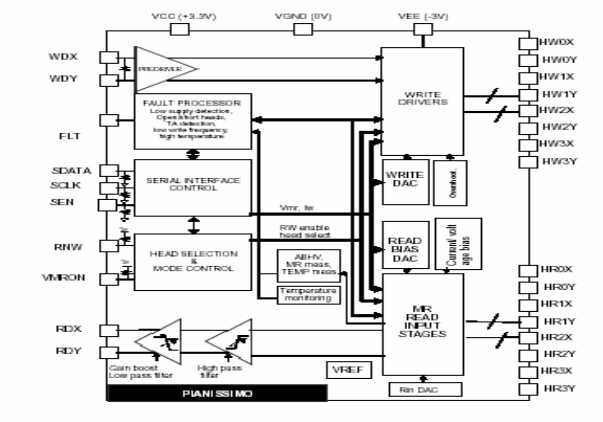

Figura 1.7

1.2 Preamplificatori per HDD – Stato dell'arte

Un preamplificatore per HDD può essere suddiviso, dal punto di vista funzionale, in alcuni sottosistemi, come illustrato in figura 1.7.

1.2.1 Descrizione dei blocchi

Il *low noise amplifier* (LNA) è attivo durante la fase di lettura. Esso riceve in modo differenziale, attraverso una delle due coppie di pad connesse alle testine di lettura (nel caso di *hard disk* a singolo piatto, due superfici utili, quindi due coppie di testine) un debole segnale in tensione, dell'ordine di poche centinaia

di μV e lo amplifica di un fattore che può variare da 100 a 200V/V, a seconda dell'applicazione. Per quanto concerne il LNA, vi sono due parametri fondamentali da ottimizzare in fase di progetto:

- ✓ Banda a -3dB
- ✓ Figura di rumore

Per quanto riguarda la banda, generalmente essa deve estendersi da un corner minimo di non più di DR/4000 ad un massimo di almeno 0.55*DR, dove DR corrisponde al data rate a cui deve lavorare il preamplificatore. I preamplificatori realizzati in tecnologia convenzionale BiCMOS standard lavorano con data rate pari a 800Mbit/s, pertanto la banda del LNA deve estendersi da 200kHz a 440MHz.

La testina, in fase di lettura, è sostanzialmente un sensore magneto-resistivo, la cui resistenza ohmica varia in funzione del flusso magnetico che la attraversa. Tipicamente il valore R_{MR} del trasduttore (a riposo) può variare da 100 a 500Ω; le variazioni indotte dal flusso sono funzione della tecnologia adottata per realizzare il trasduttore (MR=*Magneto Resistance*, GMR=*Giant Magneto Resistance*, TMR=*Tunnel Magneto Resistance*), e possono variare di conseguenza da 2-3% a circa il 20% della resistenza a riposo R_{MR}.

La figura di rumore è definita come segue:

$$NF[dB] = 10 \cdot \log\left[\frac{(V_{nout}/A_v)^2}{4kTR_{MR}}\right]$$

dove V_{nout} è la tensione di rumore efficace complessiva all'uscita del LNA, mentre A_v rappresenta il guadagno in tensione. Secondo tale definizione, la figura di rumore non dovrebbe eccedere 3.5dB (assumendo R_{MR}=500Ω).

Il *write driver* è attivo in scrittura. Esso provvede a pilotare la testina di scrittura selezionata – che è equivalente ad un induttore di 2-4nH – attraverso una linea di trasmissione con impedenza caratteristica pari a 50Ω. Il *write driver* trasforma ogni bit in una transizione di corrente nella testina di scrittura e, di conseguenza, in una variazione di flusso magnetico. I valori di corrente in gioco sono dell'ordine di 50-150mA, al massimo. Ciò che deve essere estremamente controllato sono i fronti di salita e di discesa (che non devono superare 0.3/DR), nonché l'eventuale sovraelongazione programmabile (da 0 a 150% della corrente di scrittura in *steady state*). Inoltre, poiché il pilotaggio avviene attraverso una linea, è necessario adattarla opportunamente per evitare la generazione di riflessioni.

Infine, oltre ai due sottosistemi di lettura e di scrittura, è necessario implementare alcuni blocchi di servizio, i quali per esempio sono responsabili della polarizzazione dell'elemento magneto-resistivo in fase di lettura, della programmazione del preamplificatore attraverso una interfaccia digitale seriale, e di altre eventuali funzioni specifiche richieste dal cliente, in funzione dell'applicazione finale. Il cosiddetto *fault processor*, per esempio, controlla che non si siano verificate situazioni anomale per cui risulti non opportuno proseguire nelle operazioni di scrittura. In caso avvengano eventi di questo tipo, il *fault processor* abilita un segnale che inibisce le operazioni a rischio.

1.2.2 Tecnologia

La tecnologia attualmente impiegata nella realizzazione dei preamplificatori è la BiCMOS standard, che impiega cioè transitori bipolari ad omogiunzione (BJT).

Le attuali tecniche di realizzazione dei transistori bipolari forniscono la possibilità di isolare i dispositivi con la stessa tecnica di ossidazione locale usata per i CMOS. Questo procedimento ha il vantaggio di ridurre notevolmente la capacità parassita collettore-substrato del transistore bipolare, dato che alle regioni fortemente drogate prossime alla superficie, che generano elevate capacità parassite, si sostituisce l'isolamento dell'ossido a bassa capacità parassita. I dispositivi possono anche essere più densamente integrati sul *die*. Inoltre, le tecnologie di fabbricazione dei CMOS e dei bipolari cominciano ad essere piuttosto simili, per cui diventa possibile combinare (a costo di aggiungere qualche ulteriore passo di processo) in una tecnologia BiCMOS i transistori bipolari veloci, sottili, a impiantazione ionica con i dispositivi CMOS. Si possono, in tal modo, migliorare le prestazioni nelle applicazioni digitali dato che l'elevata capacità di portare corrente dei transistori bipolari facilita notevolmente il pilotaggio di grossi carichi capacitivi (*drive capability*). Anche per le applicazioni analogiche questi processi sono molto interessanti, permettendo al progettista di trarre vantaggio dalle caratteristiche peculiari di entrambi i tipi di dispositivi.

Nella figura 1.8 è mostrata una vista semplificata in sezione di un tipico processo BiCMOS per dispositivi ad alta frequenza.

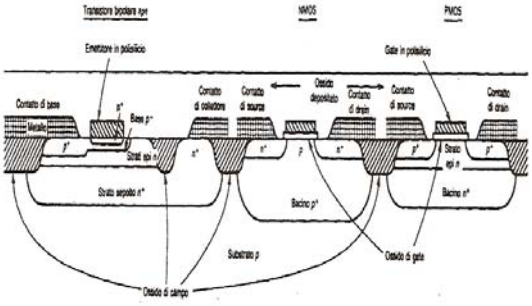

Figura 1.8

Questo processo inizia con passi di mascheratura e di impiantazione di strati sepolti n^+ di antimonio in un substrato di tipo p ovunque si debba formare un transistore bipolare npn o un dispositivo PMOS. Una seconda impiantazione di tipo p di impurezze di boro forma un bacino di tipo p^+ ovunque si debba collocare un dispositivo NMOS. A questo punto si fa seguire la crescita di circa 1μm di strato epitassiale di tipo n, che forma i collettori dei transistori npn e la zona di canale dei dispositivi PMOS. Durante questo e i successivi cicli termici gli atomi di boro più mobili retrodiffondono e il bacino p si estende fino alla superficie, mentre gli strati sepolti di antimonio restano sostanzialmente fermi.

Una mascheratura individua le regioni in cui crescere l'ossido spesso di campo. Si cresce l'ossido cui segue una spianatura laddove l'ossido di campo era cresciuto al di sopra del piano superficiale, riportandolo alla stessa quota delle altre regioni. Infine, con una serie di mascherature e di impiantazioni di tipo p e n si formano le regioni bipolari di emettitore e di base, il contatto bipolare di collettore a bassa resistenza e le zone di *source* e di *drain* dei MOSFET. Si realizzano poi, in sequenza, la crescita dell'ossido e la formazione dei gate in polisilicio e degli emettitori. Si depositano da ultimo i contatti metallici sopra

le zone volute e, successivamente, il *die* viene ricoperto con una deposizione stratificata di SiO_2. Un secondo strato di interconnessioni metalliche viene depositato sopra l'ossido, realizzando tramite *vias* le connessioni al primo strato di metallo, dove necessario. Successivamente, è possibile realizzare altri strati di interconnessioni metalliche, a seconda della complessità del circuito.

1.3 Identificazione della tecnologia ottimale

Le sempre più stringenti richieste del mercato degli HDD in termini di *areal density* e di velocità di lettura/scrittura hanno portato alla necessità di realizzare un'elettronica di controllo e di processamento dei dati in grado di soddisfare queste specifiche.

L'obiettivo della presente attività di ricerca consiste nella identificazione della tecnologia ottimale per la realizzazione di un preamplificatore per il processamento dei dati fino a 3Gbit/s.

La fase introduttiva di ogni analisi di fattibilità comporta lo studio delle tecnologie disponibili e la scelta della più adatta, secondo criteri legati alle *performance* richieste nonché ai vincoli di costo, area e dissipazione.

In particolare, il progetto di preamplificatori a frequenze più elevate richiede l'impiego di BJT più performanti in termini di velocità (tipicamente con frequenze di guadagno unitario $f_T > 45\text{GHz}$). Con la tecnologia BiCMOS standard sopra decritta, questo risultato potrebbe essere ottenuto impiegando una struttura del dispositivo più "aggressiva" e più complessa in termini di processo che, di conseguenza, si traduce in un notevole aumento dei costi.

Recentemente, si sta affermando l'uso di processi BiCMOS al Silicio-Germanio, capaci di fornire transitori bipolari ad eterogiunzione di tipo *npn* estremamente veloci ($f_T > 45\text{GHz}$).

Per eterogiunzione si intende una giunzione *pn* realizzata con due differenti materiali. Nei BJT considerati finora, le giunzioni considerate sono omogiunzioni essendo impiegato lo stesso materiale (silicio) per formare entrambe le regioni di tipo *n* e di tipo *p* . Al contrario, una giunzione tra una regione di tipo *n* di silicio e di tipo *p* di germanio (o un composto di silicio e germanio) forma un'eterogiunzione.

Nei transistori bipolari ad omogiunzione, il drogaggio di emettitore è molto più grande del drogaggio di base per ottenere un'efficienza di iniezione $\gamma \approx 1$. Di conseguenza, la base risulta debolmente drogata mentre l'emettitore è fortemente drogato. E' noto che la f_T dei dispositivi bipolari è limitata fondamentalmente dalla τ_f ($f_T \approx 1/\tau_f$), che è il tempo di attraversamento della base dei portatori minoritari. La massimizzazione di f_T è importante in alcune applicazioni come quelle a radio-frequenza e può essere ottenuta riducendo la larghezza della base. Tuttavia, se il drogaggio di base è fissato per mantenere un'efficienza di iniezione costante, questo approccio determina un incremento della resistenza di base r_b. A sua volta, questa resistenza di base limita la velocità perché forma una costante di tempo con la capacità attaccata al nodo di base. Come risultato, nella tecnologia bipolare standard esiste un *tradeoff* tra un'elevata f_T da un lato e una bassa r_b dall'altro, e di fatto in entrambi i casi estremi la velocità delle cariche è fortemente limitata.

Un modo di superare questo *tradeoff* è quello di aggiungere germanio alla base dei transistori bipolari per formare transistori ad eterogiunzione (HBT – *heterojunction bipolar transistor*). L'idea chiave è che materiali differenti ai due lati della giunzione presentano differenti bandgap. In particolare, il bandgap del silicio è più grande di quello del germanio, e la formazione di SiGe nella base riduce il bandgap proprio in base. Il bandgap relativamente più grande nell'emettitore determina un aumento della barriera di potenziale per le lacune che possono essere iniettate dalla base verso l'emettitore stesso. Perciò, in questa struttura non è necessario avere un drogaggio nell'emettitore molto più grande del drogaggio in base al fine di ottenere la massima efficienza d'iniezione. Di conseguenza, rispetto al caso del BJT, è possibile in un HBT diminuire il drogaggio di emettitore ed aumentare il drogaggio di base, consentendo così di mantenere costante la r_b anche quando si riduce la larghezza di base al fine di incrementare la f_T.

La regione di base degli HBT può essere realizzata facendo crescere un sottile strato epitassiale di silicio-germanio mediante l'impiego della UHV/CVD (*ultra high vacuum/chemical vapour deposition*). Nella figura 1.9 è riportata una sezione schematica del reattore UHV/CVD.

Poiché il processo epitassiale consente la crescita di film con la stessa struttura del substrato, è possibile ottenere uno strato sottile monocristallino di silicio-germanio su un substrato di silicio monocristallino.

E' importante sottolineare che un limite di questo processo è il basso valore della concentrazione di germanio che è possibile impiegare nella realizzazione del composto semiconduttore a causa del differente passo reticolare dei cristalli di germanio e di silicio (il passo reticolare del Ge è più grande).

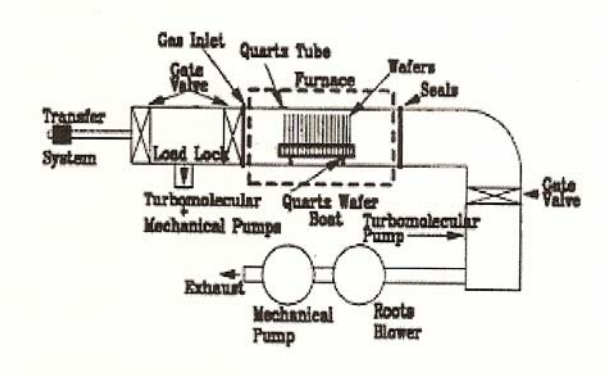

Figura 1.9

Ad esempio, per uno spessore del film di SiGe di circa 100nm, è possibile ottenere un film praticamente privo di difetti nella struttura cristallina con una concentrazione massima di Ge del 15%.

Una possibile soluzione per incrementare ulteriormente le prestazioni del dispositivo è legata alla possibilità di creare in base un profilo di concentrazione di Ge non costante. Infatti, il gradiente di concentrazione del germanio in base determina un campo elettrico che forza gli elettroni a spostarsi attraverso la base, riducendo ulteriormente τ_f e, dunque, incrementando f_T (figura 1.10).

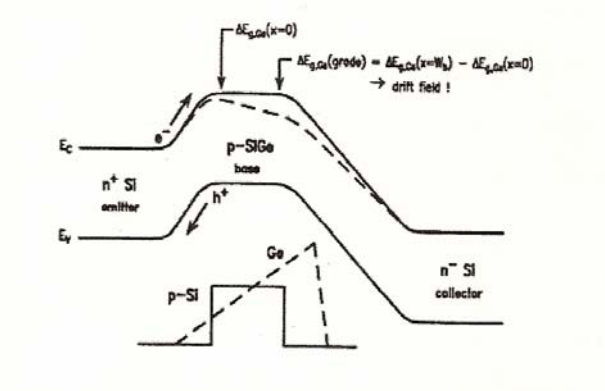

Figura 1.10

Tale soluzione sarà oggetto di studio nella presenta attività di ricerca. In particolare, si procederà allo studio dei dispositivi con differenti profili di

germanio nella regione di base al fine di individuare il profilo ottimo per la specifica applicazione in oggetto.

CAPITOLO 2 - Ingegnerizzazione dei Materiali

Il silicio puó essere, a pieno titolo, considerato un ottimo materiale dal punto di vista tecnologico, tuttavia, dal punto di vista del progettista elettronico non rappresenta il semiconduttore ideale. Infatti, la mobilitá dei portatori per le lacune e gli elettroni è relativamente piccola e la massima velocitá che questi portatori possono raggiungere sotto elevati campi elettrici è di circa 10^7 cm/sec in condizioni normali. Poiché la velocitá di un dispositivo dipende alla fine da quanto velocemente i portatori possono essere trasportati nel dispositivo stesso alle tensioni di lavoro, il Si puó essere riguardato come un semiconduttore "lento". Inoltre, essendo un semiconduttore a bandgap indiretto, presenta un'efficienza quantica bassa, rendendo impraticabile la realizzazione di dispositivi ottici attivi come i diodi laser. D'altra parte, molti dei semiconduttori composti III – V (GaAs o InP) presentano mobilita' e velocita' di saturazione piu' elevate e, in virtu' del loro bandgap diretto, e' possibile realizzare dispositivi ottici più efficienti. In piu', in virtu' del loro processo di realizzazione, i dispositivi III – V possono essere modificati nella loro composizione per esigenze o applicazioni specifiche (ad esempio, per modulare la lunghezza d'onda della luce di un diodo laser). La "customizzazione" a livello atomico di un semiconduttore e' chiamata *bandgap engineering* e fornisce grossi vantaggi in termini di prestazioni per la tecnologia III – V rispetto alla tecnologia in silicio. Sfortunatamente, a fronte di questi benefici, la tecnologia III – V presenta problemi pratici associati alla realizzazione di circuiti integrati a basso costo e con elevati livelli di integrazione. Ad esempio,

per i materiali GaAs o InP non e' possibile allo stato far crescere termicamente ossidi robusti, i wafers sono piu' piccoli e con una maggiore densita' di difetti, si rompono piu' facilmente, etc. Queste deficienze si traducono generalmente in piu' bassi livelli di integrazione, maggiori difficolta' di fabbricazione, produzione minore e costi piu' elevati.

La questione fondamentale e', dunque, capire se e' possibile migliorare le prestazioni dei transistori di Si in modo da renderli competitivi con i dispositivi III – V per le applicazioni RF e a microonde, preservando i vantaggi associati alla tecnologia di realizzazione dei dispositivi in silicio. La risposta e' chiaramente affermativa e questo lavoro di tesi ha lo scopo di analizzare quella tecnologia, basata sulle leghe SiGe, che permette di raggiungere l'obiettivo di cui sopra e che porta alla realizzazione dei SiGe HBT.

Mentre l'idea di base di impiegare leghe di SiGe nella realizzazione di dispositivi in silicio risale agli anni '50, la sintesi di film di SiGe privi di difetti si dimostro' sorprendentemente difficile e solo verso la meta' degli anni '80 furono prodotti i primi film adatti all'impiego nelle tecnologie elettroniche. La difficolta' di realizzare questi film sta nel fatto che, mentre Si e Ge possono essere combinati al fine di ottenere una lega chimicamente stabile, le loro costanti reticolari differiscono del 4.2% e, dunque, i composti in SiGe cresciuti su substrati di Si risultano essere deformati in compressione (*strained*). Si parla in questo caso di crescita pseudomorfica di film di SiGe deformati su substrati di Si, con il film SiGe che tende ad assumere la costante reticolare del Si sottostante. Questi strati *strained* sono soggetti ad un criterio fondamentale di stabilita' che ne limita lo spessore per una data concentrazione di Ge. I film di SiGe depositati che giacciono sotto la curva di stabilita' sono

termodinamicamente stabili e possono essere processati usando forni convenzionali o *rapid-thermal annealing*, oppure l'impiantazione ionica senza generare difetti. Tuttavia, i film di SiGe depositati che giacciono sopra la curva di stabilita' sono "metastabili" e si "rilasseranno" verso la loro naturale costante reticolare (> Si) se esposti a temperature superiori alla temperatura di crescita originaria, generando difetti nel processo che li rendono non utilizzabili per la realizzazione di dispositivi elettronici. Affinche' si possa mettere in piedi una tecnologia SiGe industrializzabile, e' necessario naturalmente che i film di SiGe restino stabili.

2.1 SiGe HBT

L'impiego del Ge nel Si presenta una serie di conseguenze.

Innanzitutto, poiche' il Ge possiede una costante reticolare maggiore di quella del Si, il bandgap di energia del Ge risulta piu' piccolo di quello del Si (0.66eV per il Ge, 1.12eV per il Si), dunque il SiGe avra' un bandgap minore di quello del Si, rendendolo un ottimo candidato per il *bandgap engineering*. La deformazione per compressione associata ai composti in SiGe produce una variazione addizionale del bandgap che ne porta ad un'ulteriore riduzione di circa 75meV per ogni 10% di Ge introdotto. Questo "offset di banda" si presenta fondamentalmente nella banda di valenza, la qual cosa lo rende favorevole per l'impiego nella "customizzazione" dei transistori bipolari *npn*.

Poiche' un film di SiGe deve essere molto sottile affinche' possa restare stabile e dunque privo di difetti, e' certamente un naturale candidato per l'utilizzo nella regione di base di un transistore bipolare (che per definizione deve essere

sottile per il funzionamento ad alta frequenza). Il dispositivo risultante e' costituito da una eterogiunzione emettitore-base *n*-Si/*p*-SiGe e un'eterogiunzione base-collettore *p*-SiGe/*n*-Si, e dunque questo dispositivo e' piu' propriamente chiamato "transistore bipolare a doppia eterogiunzione SiGe", benche' per semplicita' si continuera' ad utilizzare il nome standard di "transistore bipolare ad eterogiunzione SiGe" (SiGe HBT). Il SiGe HBT rappresenta il primo pratico transistore basato sul *bandgap engineering* nel sistema materiale silicio.

Un'altra conseguenza e' legata alla possibilita' che i SiGe HBT possono essere molto facilmente integrati con i transistori CMOS in Si per formare la tecnologia monolitica SiGe HBT BiCMOS. E' questo aspetto che rende competitiva, nel lungo periodo, la tecnologia SiGe, dato che i SiGe HBT possono essere realizzati senza eccessive penalizzazioni in termini di costi rispetto agli standard IC in Si. L'integrazione dei SiGe HBT con la tecnologia Si CMOS e' inoltre il fondamentale punto di partenza per confrontare la tecnologia SiGe e le tecnologie III – V. Affinche' la tecnologia SiGe possa essere competitiva nel lungo periodo, essa deve fornire i vantaggi prestazionali dei SiGe HBT, il livello d'integrazione e la densita' di memoria dei Si CMOS in un singolo ed economico IC che consenta l'integrazione su SoC (ovvero la tecnologia SiGe HBT BiCMOS). I processi tecnologici tipici in SiGe HBT BiCMOS presentano un incremento del 20% nel numero delle maschere rispetto alla tecnologia digitale CMOS e questo e' riguardato come un compromesso accettabile tra prestazioni e costi, a seconda dell'applicazione. In effetti, la tecnologia SiGe HBT BiCMOS rappresenta certamente il futuro dei SiGe HBT, poiche' essa consente l'implementazione di soluzioni *system-on-*

chip su un'ampia base di mercato per applicazioni *wired* e *wireless* ad un costo decisamente accettabile. Questo e' il percorso che la maggior parte delle industrie operanti nel settore della microelettronica sta seguendo oggi.

2.1.1 Prestazioni

Da un punto di vista delle applicazioni a radio frequenza, allo stato dell'arte i SiGe HBT offrono una risposta in frequenza, una figura di rumore e una linearità confrontabili con gli attuali dispositivi III-V; inoltre, tali parametri risultano essere certamente migliori dei corrispondenti nei Si BJT e Si CMOS. Inoltre, i SiGe HBT offrono prestazioni migliori in termini di rumore *1/f* e di rumore di fase rispetto agli altri dispositivi. Essendo transistori bipolari, la transconduttanza per unità di area del SiGe HBT è molto più elevata dei FET in Si e III-V e, per profili graduati di Ge nella regione di base, la conduttanza di uscita del SiGe HBT è decisamente superiore. I SiGe HBT hanno anche la peculiarità di avere un rumore in banda larga minimizzato per densità di corrente molto basse, rendendoli molto interessanti dal punto di vista del consumo di potenza per applicazioni portatili. I dispositivi III-V continueranno a fornire prestazioni di rumore molto buone e, date le più elevate tensioni di breakdown, continueranno a fornire dispositivi di potenza più performanti ma ad un costo decisamente più elevato. Il vantaggio a lungo termine dei SiGe HBT è legato all'integrazione e al costo a livello di sistema, ovvero la capacità di integrarsi facilmente con i CMOS convenzionali li distingue dalle tecnologie III-V. In tal senso, la tecnologia SiGe è essenzialmente equivalente alla tecnologia in Si e presenta tutti i vantaggi associati alle economie di scala nella

realizzazione dei circuiti integrati in Si, compresi i costi del *die* e di produzione.

Per queste ragioni l'interesse nel mondo della microelettronica per la tecnologia SiGe come tecnologia commerciale dei circuiti integrati sta rapidamente crescendo e ormai tutte le principali compagnie di semiconduttori del mondo stanno sviluppando la tecnologia in SiGe.

2.2 Leghe di SiGe

Silicio e germanio sono entrambi semiconduttori del gruppo IV della tavola degli elementi e cristallizzano nella struttura reticolare del diamante (figura 2.1).

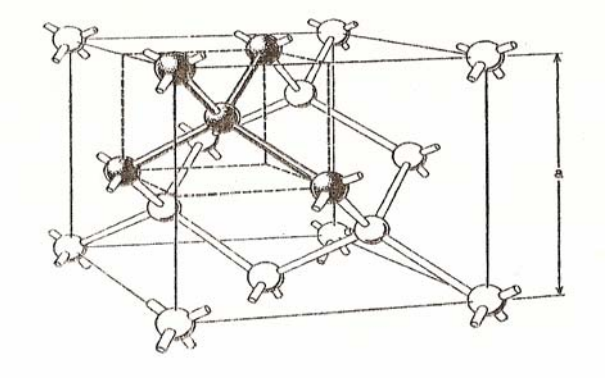

Figura 2.1

I due elementi sono inoltre completamente miscibili e possono dare vita a leghe chimicamente stabili di SiGe che conservano la struttura cristallina del diamante e che presentano una costante reticolare linearmente interpolata, data, al primo ordine, dalla formula di Vegard:

$$a\left(Si_{1-x}Ge_x\right) = a_{Si} + x \cdot \left(a_{Ge} - a_{Si}\right)$$

dove a è la costante reticolare e x è la frazione di Ge. Misure di diffrazione eseguite su film di SiGe mostrano un minore scostamento di questa dipendenza lineare e possono essere rappresentati da una relazione parabolica del tipo:

$$a\left(Si_{1-x}Ge_x\right) = 0.002733x^2 + 0.01992x + 0.5431 \quad (nm) \tag{2.1}$$

come riportato in figura 2.2.

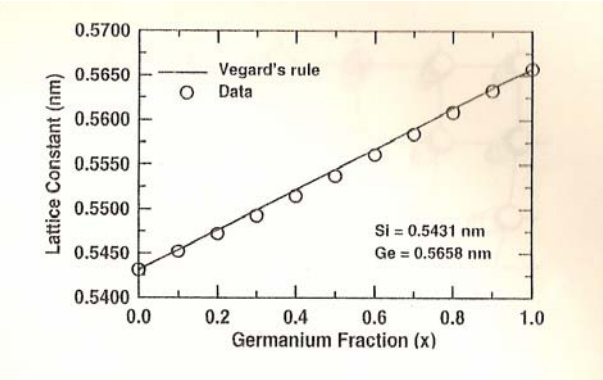

Figura 2.2

2.2.1 Crescita pseudomorfica e rilassamento del film

Il *mismatch* reticolare tra il Si puro ($a = 5.431 A^{\circ}$) e il Ge puro ($a = 5.658 A^{\circ}$) è pari al 4.17% a 300K e aumenta leggermente all'aumentare della temperatura. Quando l'epitassia di SiGe è depositata su uno spesso substrato di Si ospite, il *mismatch* reticolare intrinseco tra il film di SiGe e il substrato di Si sottostante può essere "risolto" in soli due modi.

Primo, il reticolo della lega di SiGe depositata si distorce in modo da acquisire la costante reticolare del Si sottostante. In sostanza, il film di SiGe è forzato ad adottare la costante reticolare più piccola del cristallo ospite. Questo scenario è noto come "crescita pseudomorfica" ed è il risultato desiderato per la maggior parte delle applicazioni. Sotto condizioni di processo che favoriscono la

crescita pseudomorfica, il film di SiGe è forzato in una compressione biassiale (nel piano). In questo caso, la costante reticolare del SiGe nel piano di crescita è determinata dal substrato di Si e il risultato è una distorsione tetragonale (estensione) del cristallo di SiGe, normalmente di forma cubica, nella direzione ortogonale. La lega di SiGe è così sotto compressione, dando vita all'epitassia dello strato *strained* di SiGe. A causa dell'energia di compressione addizionale contenuta nel film di SiGe durante la crescita pseudomorfica, la lega contiene uno stato di energia più elevato rispetto al film *non strained* e, dunque, la natura non favorisce questa condizione di crescita eccetto che sotto un range molto limitato di condizioni.

Secondo e in alternativa, il film di SiGe può "rilassare" durante la crescita verso la costante del reticolo naturale determinata dalla frazione di Si e Ge, come risulta dalla (2.1). Il film di SiGe rilassa mediante la formazione di dislocazioni, originando la rottura della regolarità del reticolo cristallino attraverso la superficie di crescita e, dunque, dando vita ad un film difettato e non adatto alla produzione di massa di dispositivi elettronici. Il rilassamento durante la crescita del SiGe si verifica quando l'energia di compressione è sufficientemente grande che le dislocazioni di spostamento nucleano e, dunque, si muovono.

In sostanza, quando l'energia di compressione nel film supera l'energia di attivazione richiesta per la formazione e il movimento delle dislocazioni, il film stesso rilasserà, rilasciando così l'energia di compressione immagazzinata. Questo meccanismo di rilassamento è alquanto complesso e differenti gradi di compressione residua post-crescita possono risiedere nei film di SiGe. Le dislocazioni per disadattamento, formate durante il processo di rilassamento,

possono essere confinate nel piano d'interfaccia di crescita originale o, in alternativa, possono "infilarsi" attraverso lo strato epitassiale di SiGe. In entrambi i casi, questi difetti possono comportarsi come centri di generazione/ricombinazione per le cariche e come canali per i droganti e, pertanto, rappresentano una situazione da evitare dal punto di vista del progetto del dispositivo. Questi due scenari di crescita sono rappresentati schematicamente nelle figure 2.3 e 2.4.

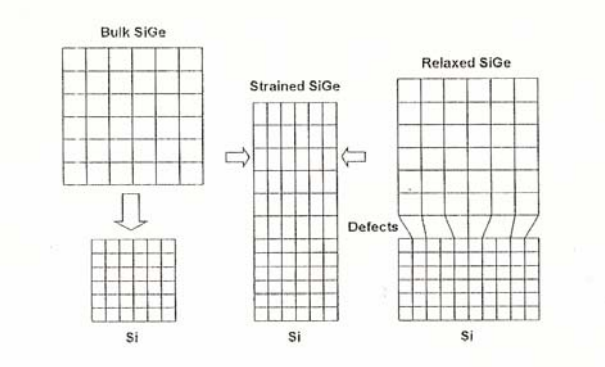

Figura 2.3

In pratica, se immaginiamo un processo di crescita del SiGe senza restrizioni utilizzando una frazione arbitraria di Ge, si dovrebbe procedere come segue. Poiché il substrato di Si è molto spesso (circa 600μm per un wafer di 200mm) e molto rigido, esso rimane sostanzialmente invariato durante la crescita epitassiale. Assumendo un'interfaccia di crescita iniziale pura, la crescita del film di SiGe comincerà in modo pseudomorfo, adottando la costante reticolare del Si sottostante, ma quando viene raggiunto un dato "spessore critico" l'energia di compressione diventa troppo grande per mantenere l'equilibrio locale e il film di SiGe rilasserà verso la sua naturale costante reticolare, con l'energia di compressione in eccesso che genera la formazione di dislocazioni. Questa è la ragione per cui il film di SiGe non può superare un certo spessore.

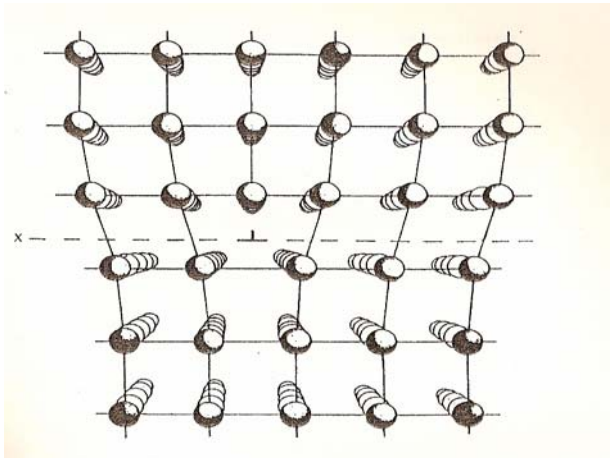

Figura 2.4

Nella pratica, può anche capitare che il film rimanga pseudomorfo fino alla fine del ciclo di crescita, anche se viene superato lo spessore critico. Questo strato di SiGe è detto *metastabile*. I film metastabili rilasseranno durante i successivi *step* di processo che aggiungono energia al sistema e, dunque, non possono essere impiegati nelle tecnologie SiGe.

Al di là delle tecniche di crescita utilizzate, o della struttura e degli schemi di auto-allineamento utilizzati nel transistore, tutti i film di SiGe compressi impiegati nei moderni HBT di SiGe hanno una forma simile. Come riportato in figura 2.5, il film di SiGe depositato effettivamente consiste di una struttura composta da tre strati:

- uno strato buffer di Si, sottile e non drogato;
- lo strato attivo di SiGe drogato con boro;
- uno strato di copertura (*cap layer*) di Si, sottile e non drogato.

Lo strato buffer di Si è usato per far partire il processo di crescita e serve a due scopi. Primo, assicura che sia preservata un'interfaccia di crescita epitassiale pura di SiGe tra il substrato originale di Si, che è stato cresciuto attraverso un processo di epitassia di Si ad alta temperatura, e il successivo strato *strained* di

SiGe che sarà cresciuto con un processo di epitassia a bassa temperatura più complesso.

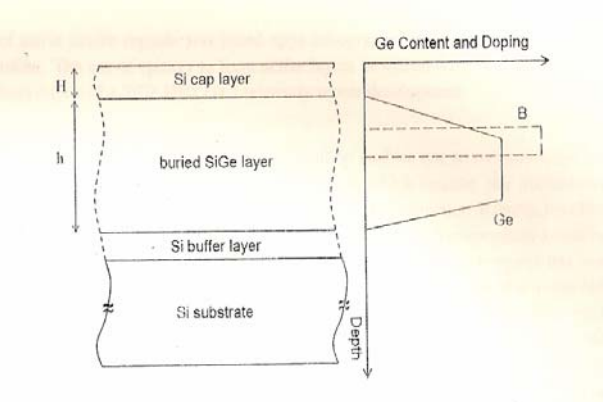

Figura 2.5

Mantenere l'interfaccia di crescita libera dai contaminanti con una perfetta cristallinità è essenziale per ottenere film di SiGe da impiegare nella realizzazione di dispositivi. Secondo, questo strato buffer di Si spesso gioca anche un ruolo importante nel progetto del dispositivo poiché permette l'incorporazione degli strati intrinseci da inglobare facilmente nella giunzione collettore-base e possono essere usati per diminuire il campo di giunzione e fornire un ausilio nel progetto della tensione di breakdown.

Lo strato attivo di SiGe, di spessore h, presenta una composizione di Ge che varia con la posizione, e un picco di drogaggio di boro al suo interno, tipicamente un profilo "a scatola". Lo strato di SiGe costituisce la regione attiva del dispositivo e la forma, lo spessore e la collocazione del profilo di Ge rispetto al profilo di boro in base determinerà, in larga misura, le prestazioni risultanti del transistore.

Infine, il cap layer di Si, di spessore H, serve a 4 scopi. Primo, fornisce una terminazione di Si al composto SiGe. Questo è particolarmente importante poiché la maggior parte degli approcci di fabbricazione degli HBT in SiGe

prevedono qualche forma di *step* di ossidazione per formare lo *spacer* tra emettitore e base usato per l'auto-allineamento, e il SiGe non si ossida bene. Secondo, il cap di Si fornisce uno spazio aggiuntivo per consentire la diffusione verso l'esterno del profilo di boro in base durante il processo, mentre allo stesso tempo fornisce spazio per la diffusione verso l'esterno dell'emitter. Terzo, come per lo strato buffer di Si, un cap layer di Si può essere usato per introdurre un i-layer attivo nella giunzione emitter-base per abbassare il campo elettrico della giunzione e, dunque, ridurre la corrente di tunneling EB parassita, che tipicamente limita l'idealità della corrente di base per basse iniezioni e, dunque, degrada l'affidabilità del dispositivo. Infine, il cap layer di Si contribuisce a migliorare la stabilità complessiva del film, aumentando lo spessore e la frazione di Ge dello strato consentiti.

Un'appropriata pulizia della superficie e una crescita attenta di questo film a tre strati di SiGe può risultare in films di ottima qualità per la realizzazione di dispositivi. In questo caso, si ottiene una struttura cristallina perfetta poichè il film di SiGe è diventato parte, a tutti gli effetti, del cristallo ospite di Si, acquisendone la medesima regolarità strutturale.

2.2.2 La sfida dell'epitassia di Si

Nei primi processi di crescita dell'epitassia di Si i film erano cresciuti per fornire regioni ben controllate di materiale uniformemente drogato su un wafer di silicio meno uniforme che fungeva da *template* per la crescita. La successiva definizione e realizzazione delle regioni attive del dispositivo erano basate sul *patterning* fotolitografico e l'impiantazione ionica. L'uso dell'epitassia per

realizzare strati attivi nei dispositivi in silicio (cioè la regione di base di un HBT in SiGe) è uno sviluppo relativamente recente.

L'impiego dell'epitassia in Si nella fabbricazione dei dispositivi consente di superare i limiti fondamentali dell'impiantazione ionica, che sono sostanzialmente rappresentati dalla distribuzione gaussiana dell'impiantazione, funzione della profondità, la canalizzazione ionica delle specie dopanti impiantate e la necessità di un *annealing* ad alta temperatura per rimuovere i danni dell'impiantazione e per attivare i droganti. Benché siano stati fatti molti progressi nel ridurre la larghezza di base dei BJT in Si, ottenuta per impiantazione ionica, tuttavia tale larghezza, nel regime sub-100nm, risulta essere difficile da controllare nella pratica, limitando così le prestazioni del BJT in Si a valori di picco di f_T sotto i 50GHz. Attraverso l'epitassia del Si, invece, è possibile progettare con precisione il profilo dei droganti nello strato epitassiale e, inoltre, è possibile creare dei composti semiconduttori come il SiGe. Tutto ciò in teoria, perché nella pratica il budget termico dell'epitassia convenzionale in Si, combinato con quello di altri *step* successivi, rende il raggiungimento di questi risultati teorici molto difficile.

L'elevato budget termico associato con l'epitassia di Si convenzionale può essere compreso nel modo seguente. La corretta realizzazione dell'epitassia di Si richiede che si prepari prima una superficie di Si atomicamente "pulita" che servirà da *template* per la crescita epitassiale. L'approccio storico usato nella crescita dell'epitassia di Si è stato quello di riscaldare il wafer in un'atmosfera di idrogeno ad alte temperature (generalmente più di 1000°C) al punto di far evaporare gli ossidi di superficie, rimuovere il carbonio e la contaminazione dei dopanti in superficie. La crescita del film comincia, dunque, ad alte

temperature, assicurando così che i contaminanti residui del sistema di crescita come l'ossigeno o il carbonio non siano incorporati negli strati epitassiali in crescita. Siccome sia la temperatura di pulizia sia quella di crescita per l'epitassia convenzionale di Si sono nel *range* di 1000°C, esse risultano essere incompatibili con le specifiche richieste del processo epitassiale allo scopo di realizzare film da impiegare nei dispositivi. A tali temperature, ogni vantaggio derivante dalla formazione di uno strato preciso del dispositivo per epitassia viene perduto nella successiva diffusione di dopanti. Nel caso estremo di impiego di uno strato *strained* di SiGe, un ulteriore rischio è il rilassamento potenziale alle alte temperature di tali strati che può determinare, come già anticipato prima, la formazione di difetti. Pertanto, la chiave per il successo dell'impiego di uno strato epitassiale di Si (o SiGe) nella realizzazione di dispositivi elettronici ad elevate prestazione consiste nell'ottenere la crescita del film a temperature molto basse (< 600°C). La difficoltà intrinseca di questo passo ha ritardato lo sviluppo degli HBT in SiGe fino alla metà degli anni '80.

2.2.3 Crescita del SiGe

Negli ultimi 25 anni sono state sviluppate diverse tecniche di crescita in grado di produrre film di SiGe da utilizzare nella realizzazione di dispositivi elettronici, ovvero aventi una "qualità da dispositivo". In questo contesto, per "qualità da dispositivo" s'intendono film di SiGe privi di difetti visibili che possono essere usati per produrre singoli HBT in SiGe ideali o quasi ideali. Naturalmente, non tutte le tecniche di crescita di tali film sono adatte per una produzione di scala della tecnologia HBT in SiGe.

Attualmente, le tecniche più impiegate nella produzione di scala dei suddetti dispositivi sono la *ultra-high vacuum/chemical vapor deposition* (UHV/CVD) e la *atmospheric pressure chemical vapor deposition* (APCVD).

2.2.4 Tecniche di crescita

La tecnica UHV/CVD non necessita di un elevato budget termico che, invece, risulta essere tipico della epitassia convenzionale ottenuta con mezzi chimici. La tecnica di crescita prevede una passivazione preliminare della superficie di Si con idrogeno, utilizzando una semplice procedura chimica in soluzione: un attacco di 10-15 secondi in una soluzione diluita 10:1 di H_2O/HF. Lo strato di idrogeno creato durante questo attacco in soluzione riduce la reattività dell'interfaccia di crescita approssimativamente di 13 ordini di grandezza rispetto a quella di una superficie scoperta di Si, con riferimento alla velocità di ossidazione in aria. Questo comportamento passivo si estende allo stesso modo all'adsorbimento di specie droganti. Inoltre, queste superfici di Si passivate con idrogeno si asciugano completamente una volta estratte dal bagno di HF, sicché il maneggiamento dei wafer cresciuti diventa più semplice. I wafer sono asciugati appena fuori dall'attacco e possono, dunque, essere direttamente caricati nella camera di crescita. I wafers "patternati" (ovvero, quelli già trattati secondo le tecnologie elettroniche per la realizzazione di circuiti integrati) che hanno differenti materiali esposti in superficie, semiconduttori o dielettrici, possono, inoltre, essere preparati in questo modo. Lo svantaggio di lavorare con wafers "patternati" sta nel fatto che essi non possono essere completamente asciugati in presenza di altri materiali oltre al Si, la qual cosa

richiede un'asciugatura di queste superfici prima di inserirle nella camera di crescita del film. Tuttavia, nella maggior parte dei processi SiGe HBT, sono impiegate semplici strutture in cui la superficie più esterna esposta è dappertutto silicio, rendendo così il wafer pienamente idrofobico e, dunque, consentendo l'asciugatura completa dei wafers "patternati". La risultante superficie di Si, ricoperta con idrogeno, che viene utilizzata nella UHV/CVD, risulta essere robusta da molti punti di vista.

Una difficoltà presente nella deposizione di regioni attive del dispostivo deriva dalla presenza di impurità elettricamente attive all'interfaccia iniziale di crescita, anche nelle condizioni di vuoto spinto (UHV) impiegate in alcune tecniche di crescita. Nel corso degli anni, sono stati sviluppati diversi metodi per ridurre l'ampiezza e l'impatto di questa contaminazione, tra questi il metodo più comune è quello di depositare uno strato *buffer* di materiale per "seppellire" la contaminazione ben sotto la regione attiva del dispositivo. Tuttavia, se si depositano degli strati "patternati" su un substrato di crescita, quello descritto non è più un approccio perseguibile. In particolare, quando la base epitassiale sta per essere depositata su un wafer che contiene regioni di collettore "patternate", in effetti si sta facendo crescere la giunzione base-collettore e una piccola quantità di dopanti spuri è certamente tollerabile. Nel caso della UHV/CVD, la dose di boro residua all'interfaccia di crescita è nel range di $10^9 - 10^{10}$ cm^{-2} e dunque non provoca nessuna conseguenza nel dispositivo risultante. E' interessante notare che la contaminazione di boro in considerazione proviene dall'ambiente.

Una volta analizzati gli strumenti utilizzati per preparare la superficie iniziale di Si per l'epitassia, adesso è possibile analizzare la crescita epitassiale stessa.

L'impiego di elevate temperature per l'epitassia convenzionale di Si è stata giustificata con la necessità di fornire una certa mobilità atomica, in modo che risultasse uno strato epitassiale di alta qualità. Inoltre, si sapeva che la crescita ad alte temperature sopprimesse l'inclusione di specie dopanti non desiderate nei film da depositare. La finezza della *chemical vapor deposition* sta nel fatto che è possibile crescere uno strato di Si epitassiale a temperatura ambiente se si considera il caso più comune, in cui un monostrato di una sorgente di gas di Si viene adsorbito su un sito reticolare di Si. Questo è, infatti, il caso della crescita per UHV/CVD dove viene utilizzato il silano allo stato gassoso (SiH_4) come sorgente di Si. Per depositare un ulteriore strato è richiesta altra energia termica al fine di estrarre l'idrogeno residuo dalla superficie, consentendo così l'adsorbimento di un monostrato aggiuntivo di film. In questo caso, la mobilità atomica non rappresenta un elemento chiave nel determinare il limite inferiore della temperatura a cui è possibile depositare il Si con tecnica epitassiale tramite la UHV/CVD.

Per ottenere un'adeguata purezza del film durante l'epitassia a bassa temperatura, sono stati impiegati diversi approcci. I più noti sono le tecniche UHV associate alla MBE (*molecular beam epitaxy*), dove vengono raggiunti livelli di vuoto nell'ordine di 10^{-11} torr. Questo livello di vuoto degrada in modo significativo durante la crescita del film, tuttavia, in tal modo, si ottengono film con elevata purezza e perfezione dopo molte ore di crescita.

Per ridurre la complessità e il costo di questo apparato, la tecnica UHV/CVD utilizza una forma chimicamente selettiva della UHV. Una volta riconosciuta la necessità di eliminare solo quelle specie che sono chimicamente attive col Si, viene utilizzata una metodologia UHV semplificata che impiega tubi di

reazione al quarzo. Benchè si riesca ad ottenere livelli di UHV relativamente bassi, tipicamente dell'ordine di 10^{-9} torr, il gas residuo è fondamentalmente idrogeno, la qual cosa non rappresenta un problema quando questo elemento è presente in quantità così basse. I livelli di ossigeno e di acqua sono ridotti ad un ordine di pressione parziale di 10^{-11} torr. Questo approccio chimico selettivo viene ormai comunemente impiegato nelle tecniche di crescita epitassiali.

I film vengono depositati per UHV/CVD a temperature nel range di 400-500°C, che corrispondono alla crescita del Ge puro e del Si puro, rispettivamente. I wafers sono passivati con *HF* come descritto sopra e successivamente caricati nel *load lock* dell'apparato UHV/CVD. E' importante sottolineare che il sistema UHV/CVD è un *batch tool* e i film di SiGe possono essere depositati su più wafer allo stesso tempo, aumentando enormemente la produttività. Dopo aver portato il sistema sotto i 10^{-6} torr, i wafers sono trasferiti sotto idrogeno fluente nella sezione UHV dell'apparato e, a questo punto, ha inizio la crescita epitassiale. Le sorgenti di gas impiegate sono silano (SiH_4), germano (GeH_4), diborano (B_2H_6) e fosfina (PH_3). La velocità di crescita del film può variare da 0.1-100 $\mathring{A}$/minuto come funzione della temperatura e del contenuto di Ge nel film, con velocità tipiche di 4-40 $\mathring{A}$/minuto. Questi limiti di velocità di crescita sono usati per assicurare un controllo dimensionale preciso sulla sezione trasversale del wafer, dell'ordine di 1-2 strati atomici in questo caso. E' richiesto questo livello di precisione se si vuole effettivamente competere con le precisioni di controllo nell'impiantazione ionica, che rappresenta il punto di riferimento per il controllo del drogaggio nei processi che coinvolgono il Si. Tramite UHV/CVD,

è possibile depositare profili di Ge graduati caratterizzati da picchi del 10-30% su dimensioni di 50-150nm con un eccellente controllo sulla sezione trasversale del wafer.

Recentemente, le tecniche APCVD per la deposizione del SiGe sono venute alla luce come tecniche proficue per la crescita di film. La tecnica APCVD deposita Si e SiGe a pressione atmosferica usando sorgenti di gas di SiH_2Cl_2 e GeH_4. La deposizione di film di SiGe è tipicamente ottenuta in un reattore convenzionale di Si epi riscaldato e successivamente raffreddato ad aria. A differenza della UHV/CVD, viene realizzato uno stadio di pre-pulizia *in situ* sui wafers di partenza, seguito da un breve pre-riscaldamento (1070°C per 10 minuti), seguito ancora da un attacco di gas *HCl* per un breve tempo aggiuntivo. In luogo della UHV, vengono usati purificatori di gas e un sistema di *load lock* per controllare la contaminazione di ossigeno e carbonio. I sistemi APCVD sono *tools* a singolo wafer, con i wafers piazzati orizzontalmente su un *holder* di quarzo. In letteratura non sono disponibili molti dati su questa tecnica, tuttavia il fatto che questi reattori per l'APCVD siano stati utilizzati efficacemente per produrre dispositivi HBT in SiGe rappresenta certamente un buon punto di partenza per lo sviluppo di questa tecnica anche a livello industriale.

2.2.5 Struttura a bande

La struttura a bande di energia in una lega di SiGe è chiaramente l'elemento chiave per la sua utilizzazione nella progettazione dei transistori.

Le caratteristiche delle leghe di SiGe adatte alla realizzazione di HBT in SiGe devono essere:

- ✓ un bandgap più piccolo rispetto al Si;
- ✓ un offset di banda che, fondamentalmente, si trova nella banda di valenza;
- ✓ i parametri di trasporto (mobilità, tempi di vita medi , etc.) devono essere migliorati, o almeno devono restare uguali, rispetto al Si.

Come si vedrà, il SiGe *strained* soddisfa tutte queste condizioni.

Sia il Si sia il Ge sono semiconduttori a gap di energia indiretto. Il bandgap principale del Si è 1.12eV a 300K, mentre il Ge ha un bandgap principale di 0.66eV a 300K.

Poiché il Ge presenta un bandgap significativamente più piccolo rispetto al Si (dovuto principalmente alla sua costante reticolare più grande), la conseguenza è che il bandgap del SiGe risulta essere più piccolo di quello del Si. Tuttavia, anche la deformazione in una lega di SiGe pseudomorfico gioca un ruolo importante nel definire le mobilità dei portatori e la struttura a bande.

Come predetto dai primi calcoli teorici, in una lega di SiGe pseudomorfico con una bassa frazione di Ge, la deformazione del film produce una distorsione nella curvatura degli estremi delle bande che perturba le masse efficaci dei portatori e, dunque, la densità degli stati nelle bande di conduzione e di valenza; allo stesso modo, vengono alterati i valori delle mobilità e del tempo di vita dei portatori.

Con riferimento alla densità degli stati, è noto che il prodotto delle densità degli stati $N_C N_V$ nelle bande di valenza e di conduzione si riduce fortemente a causa della distorsione, indotta dalla deformazione, degli estremi delle bande di

conduzione e di valenza. Una conseguenza di questo fenomeno è la riduzione delle masse efficaci degli elettroni e delle lacune. Un metodo relativamente semplice per ricavare le variazioni del prodotto della densità degli stati (che è proporzionale alle masse efficaci), al variare del contenuto di Ge, consiste nel fare misure di corrente di collettore sul transistore, benché questa tecnica non riesca a discriminare tra le variazioni delle masse efficaci delle lacune e degli elettroni. La figura 2.15 mostra i risultati rappresentativi del prodotto $N_C N_V$ in funzione del contenuto di Ge.

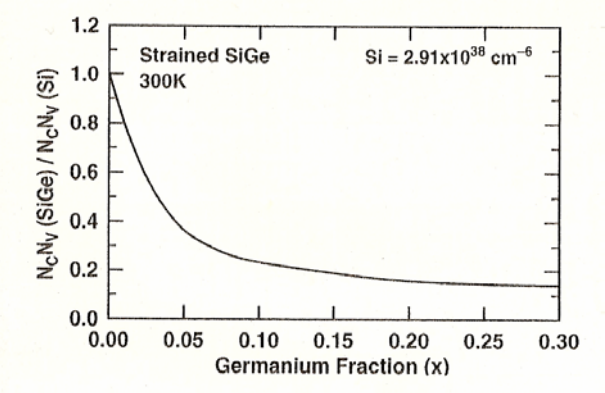

Figura 2.15

La sostanziale riduzione di tale prodotto all'aumentare della frazione di Ge può essere considerata un effetto indesiderabile poiché si traduce direttamente in una riduzione della corrente di collettore nel SiGe HBT. Questo, a sua volta, comporta una riduzione del guadagno di corrente. Tuttavia, la stessa riduzione nelle masse efficaci, che determina il decremento di $N_C N_V$, aumenta anche le mobilità dei portatori, la qual cosa parzialmente attenua l'impatto sulla corrente di collettore.

Gli allineamenti finali degli estremi delle bande nel SiGe *strained* sono mostrati in figura 2.16: ΔE_C e ΔE_V sono positivi e, dunque, i limiti delle bande di conduzione e di valenza sono compresi nei estremi originali del Si.

Nel caso di film di SiGe, l'offset della banda di valenza è di gran lunga il più significativo, come desiderato.

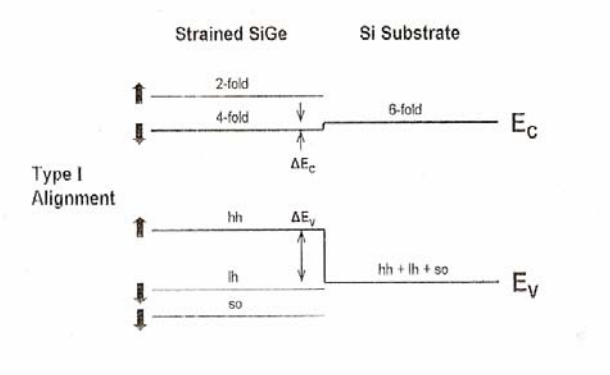

Figura 2.16

Dal punto di vista del progetto del dispositivo, una conoscenza sperimentale accurata degli offset di banda nel SiGe è di fondamentale importanza. Tuttavia, la determinazione precisa degli offset è, in generale, molto difficile da effettuare a causa delle complessità sperimentali. Per esempio, il calcolo di ΔE_g mediante misure sul transistore (tipicamente misure di temperatura della I_C) richiede delle assunzioni sulla dipendenza dall'ampiezza e dalla temperatura dei vari parametri, come ad esempio la mobilità. Inoltre, la regione di base dei SiGe HBT è fortemente drogata (tipicamente sopra i 10^{18} cm^{-3}) e presenta una dipendenza dalla posizione in prossimità delle giunzioni EB e CB. Questi fenomeni aggiungono un contributo al restringimento del bandgap dipendente dal drogaggio che è del tutto indistinguibile elettricamente da quello dovuto alla presenza di Ge e si va ad aggiungere ad esso.

Un metodo più semplice per l'interpretazione dei dati relativi all'offset di banda consiste nelle misure capacità-tensione su condensatori MOS in SiGe leggermente drogati di tipo *p*. La figura 2.17 mostra un esempio di dati di offset di banda di valenza determinati nel SiGe *strained* e mostra chiaramente

il trend ampiamente citato di un offset di banda di valenza quasi lineare in funzione della frazione di Ge per leghe di SiGe caratterizzate da una bassa percentuale di Ge.

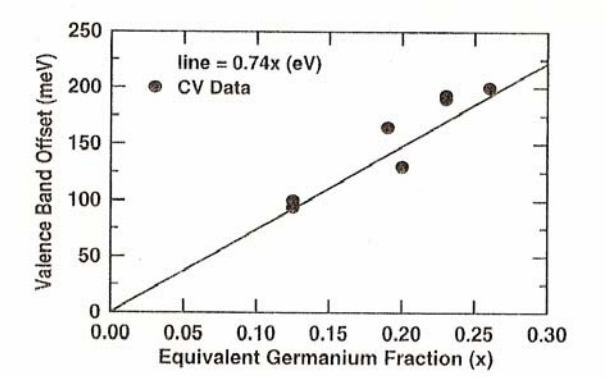

Figura 2.17

Queste misure di offset di banda sono realizzate su film con un contenuto costante di Ge e, dunque, la loro estrapolazione per profili di Ge dipendenti dalla posizione nei SiGe HBT richiede determinate assunzioni aggiuntive.

2.2.6 Parametri di trasporto

Finora si è visto che la presenza del Ge e delle deformazioni reticolari modifica l'energia, la degenerazione, la curvatura locale delle bande di valenza e di conduzione nel Si. Pertanto, ci si aspetta che nel SiGe *strained* le masse efficaci di entrambi i portatori siano significativamente alterate rispetto ai valori originali nel Si. Poiché i parametri di trasporto (le mobilità dei portatori μ_n e μ_p e i loro tempi di vita medi τ_n e τ_p) dipendono dalla struttura a bande e dalle masse efficaci risultanti dei portatori (m_n^* e m_p^*), ci si aspetta che i parametri di trasporto dei portatori possano cambiare mediante l'aggiunta di Ge al Si.

Queste variazioni dei parametri di trasporto nel SiGe sono importanti per alcune ragioni:

- ✓ la corrente di collettore in un SiGe HBT è proporzionale alla mobilità degli elettroni minoritari nella base (μ_{nb});
- ✓ la corrente di base è proporzionale alla mobilità delle lacune minoritarie nell'emettitore (μ_{pe});
- ✓ il tempo di transito in base è inversamente proporzionale a μ_{nb} ;
- ✓ la resistenza di base, che è un parametro importante sia per lo *switching* dinamico sia per le prestazioni di rumore del dispositivo, è proporzionale alla mobilità delle lacune in base (μ_{pb});
- ✓ la probabilità di ricombinazione, le correnti parassite di *leakage* che influenzano il guadagno di corrente e la conduttanza di uscita dipendono inversamente da τ_n e τ_p .

Una conoscenza di come questi parametri di trasporto dipendano dal drogaggio, dalla temperatura e dalla frazione di Ge è particolarmente importante nelle simulazioni delle prestazioni dei SiGe HBT.

Benché sia ovvia l'importanza di determinare l'influenza precisa del Ge sui vari parametri di trasporto, è certamente molto difficoltoso arrivare a dei valori quantitativi per essi. Per esempio, i film di SiGe adatti per le misure dei parametri di trasporto sono generalmente incompatibili con quelli usati negli HBT in SiGe, che richiedono regioni molto sottili e fortemente drogate, spesso con profili graduati di Ge. Inoltre, misurare accuratamente i parametri di trasporto dei portatori minoritari nel SiGe fortemente drogato rappresenta un

problema particolarmente difficile anche nel Si, in quanto richiede un'accurata conoscenza della variazione del bandgap e dati sul tempo di vita.

Nella modellizzazione dei Si BJT viene spesso utilizzato il cosiddetto "modello di mobilità unificato di Klaassen" sia per gli elettroni sia per le lacune. Questo modello, messo a punto dai ricercatori della Philips, utilizza un set di parametri dipendenti dalla temperatura per determinare il bandgap e distingue tra le mobilità dei portatori minoritari e dei portatori maggioritari. Le figure 2.18 e 2.19 mostrano le rispettive mobilità delle lacune e degli elettroni (minoritari e maggioritari) a 300 K su un *range* di drogaggio di interesse pratico nei SiGe HBT.

Si può osservare che in corrispondenza ai valori di drogaggio della base utilizzati correntemente negli HBT in SiGe (10^{18} – 10^{19} cm^{-3}) la mobilità dei portatori minoritari diventa maggiore di quella dei portatori maggioritari, la qual cosa rappresenta un vantaggio dal punto di vista delle prestazioni dinamiche del dispositivo.

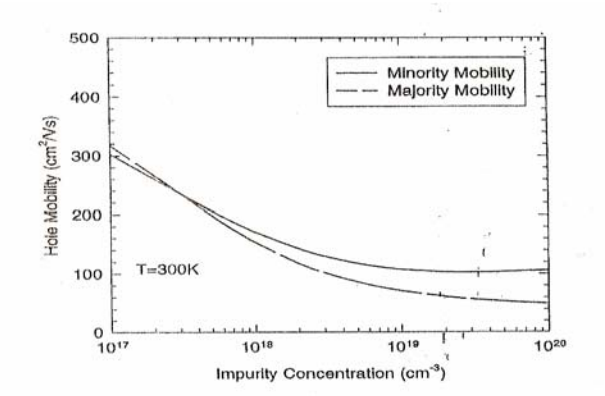

Figura 2.18

Nei film di SiGe *strained*, le mobilità dei portatori saranno alterate rispetto ai loro valori di Si a causa della distorsione locale degli estremi delle bande, a loro volta causata dagli effetti della deformazione del reticolo.

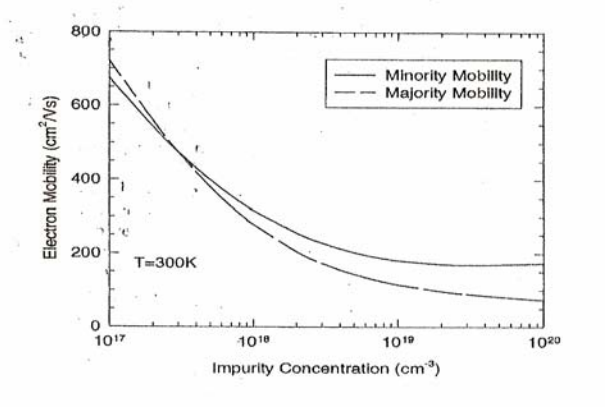

Figura 2.19

Inoltre, a causa della natura non isotropica della deformazione reticolare nel SiGe pseudomorfico cresciuto su un substrato di Si, è attendibile che le variazioni nei valori di mobilità dipenderanno dalla direzione del trasporto (se parallela o perpendicolare all'interfaccia originale di crescita del SiGe).

2.2.7 Mobilità delle lacune

Dall'analisi dei materiali, è noto che l'aggiunta di Ge nel Si aumenta la mobilità delle lacune. La figura 2.20 riporta dei risultati rappresentativi con riferimento alle direzioni di trasporto, parallela all'interfaccia di crescita (*in-plane*) e perpendicolare all'interfaccia di crescita (*out-of-plane*).

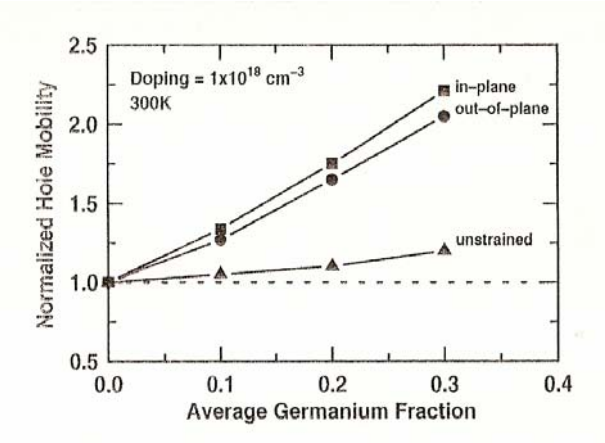

Figura 2.20

Nel caso dei SiGe HBT, la mobilità delle lacune *in-plane* è di gran lunga la più importante poiché essa determina la resistenza di base e, dunque, la risposta in frequenza (attraverso la $f_{\max}$), la velocità di commutazione e le prestazioni di rumore a larga banda. In un tipico SiGe HBT, con il contatto di emettitore in polySi, la regione di emettitore è fortemente drogata (> 10^{20} cm^{-3}) e il tempo di vita medio delle lacune *in-plane* risulta essere fondamentale nella determinazione della corrente di base, molto più della mobilità delle lacune *out-of-plane*. Dalla figura 2.20 si evince che, per la mobilità delle lacune *in-plane,* c'è un incremento di circa il 35% nella μ_p per un contenuto di Ge del 10% a 10^{18} cm^{-3} e che la dipendenza dalla frazione di Ge risulta essere quasi lineare. Ovviamente, anche la deformazione reticolare indotta dalla presenza del Ge gioca un ruolo importante nell'incremento della mobilità delle lacune.

2.2.8 Mobilità degli elettroni

In questo caso, è molto difficile ottenere risultati sperimentali poiché, in una base di tipo *p*, la mobilità degli elettroni minoritari *out-of-plane* assume grande importanza. La figura 2.21 mostra risultati rappresentativi della mobilità degli elettroni per entrambe le direzioni di trasporto (*in-plane* e *out-of-plane*). Si osserva che, per la mobilità degli elettroni *out-of-plane*, si verifica una degradazione del 15% nella μ_n per un contenuto di Ge del 10% a 10^{18} cm^{-3} e che la dipendenza dalla frazione di Ge è ragionevolmente lineare per bassi valori della frazione medesima.

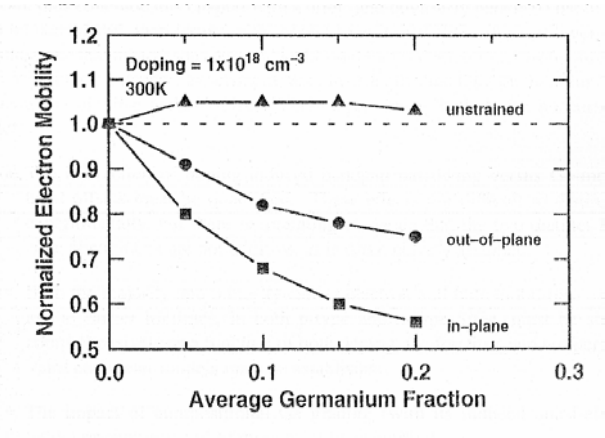

Figura 2.21

2.2.9 Scelta del modello parametrico nel SiGe

La scelta del modello parametrico da impiegare per lo studio e la simulazione dei parametri nel SiGe parte dalla considerazione che i dispositivi HBT in SiGe utilizzano comunque una bassa percentuale di Ge (tipicamente meno del 30% come valore di picco e meno del 15% come valore medio) e, dunque, sono per la maggior parte in Si. Pertanto, i modelli parametrici del Si devono sempre essere utilizzati come punto di partenza nelle simulazioni dei SiGe HBT. E' possibile seguire, a tale proposito, le seguenti regole per il modello da simulare:

- ✓ usare il modello di mobilità unificato di Klaassen per il Si come punto di partenza per le simulazioni sui SiGe HBT;
- ✓ se necessario, aggiungere un fattore di scala lineare (K_p) per tenere in conto l'incremento della mobilità delle lacune all'aumentare della frazione di Ge, secondo la formula:

$$\mu_p(SiGe)(x) = (1 + K_p x)\mu_p(Si)$$

- ✓ se necessario, aggiungere un fattore di scala lineare (K_n) per tenere in conto la degradazione della mobilità degli elettroni all'aumentare della frazione di Ge, secondo la formula:

  $$\mu_n(SiGe)(x) = (1 + K_n x)\mu_n(Si)$$

- ✓ usare i modelli standard per il Si del tempo di vita medio;
- ✓ usare il modello standard per il Si di saturazione della velocità;
- ✓ usare il modello standard del bandgap per il Si e tenere in conto l'offset di banda dovuto al Ge attraverso la seguente espressione:

  $$\Delta E_g \cong \Delta E_v = 0.74x$$

  e assicurandosi che l'offset sia collocato nella banda di valenza;
- ✓ assumere che la riduzione del bandgap dovuta al drogaggio e gli offset di banda dovuti al Ge siano additivi.

2.3 Tecnologia BiCMOS SiGe HBT

La figura 2.22 mostra il trend storico per quanto riguarda i valori di f_T a partire dalla dimostrazione del primo dispositivo *self-aligned* fino ad ora.

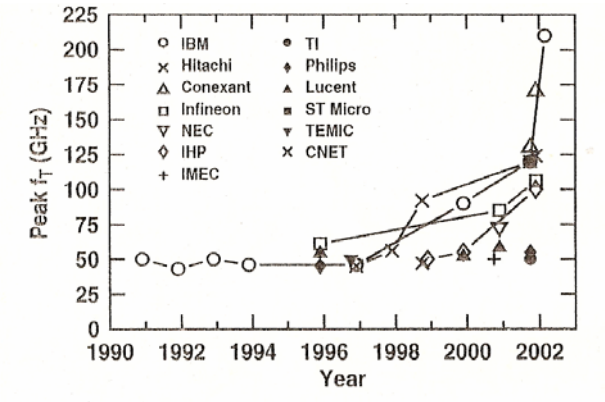

Figura 2.22

A tal proposito, occorre distinguere tra differenti generazioni della tecnologia SiGe, definite dalle prestazioni AC degli HBT in SiGe (cioè, la f_T di picco, che è una funzione del profilo verticale e, dunque, riflette fortemente il grado di sofisticazione nel progetto della struttura, il ciclo termico, la crescita dell'epi, etc.). Dunque, solitamente si etichetta la tecnologia SiGe HBT avente una f_T di picco di 45-55 GHz come la prima generazione, quella con una f_T di picco di 100-120 GHz come la seconda generazione e quella con una f_T di picco $\geq$ 200 GHz come la terza generazione.

Senza considerare l'approccio di integrazione e gli step di processo impiegati, ci sono numerosi elementi e moduli di fabbricazione comuni che esistono tra le varie tecnologie SiGe HBT e includono, per un tipico SiGe HBT della prima generazione:

- ✓ un *subcollector* di partenza n^+ (5-10 $\Omega/\square$) su un substrato di tipo p^- (10-15 Ω-cm) (utilizzando un *subcollector* patternato per consentire l'integrazione con la tecnologia CMOS);
- ✓ uno strato epi di collettore leggermente drogato, ad alta temperatura (0.4-0.5μm di spessore a $5 \cdot 10^{15}$ cm^{-3});
- ✓ *deep trenches* riempiti con polySi per l'isolamento di *subcollectors* di dispositivi adiacenti (0.8-1.2μm di larghezza e 7-10μm di profondità);
- ✓ *shallow trenches* riempiti con ossido (o LOCOS) per l'isolamento locale del dispositivo;
- ✓ un collettore *sinker* impiantato sul *subcollector* (10-20 $\Omega\mu m^2$);

- ✓ uno strato epitassiale di SiGe composto da un buffer di Si (10-20nm di spessore), SiGe drogato con boro, un *layer* attivo (70-100 nm di spessore) e uno strato di protezione di Si (*cap layer*) di 10-30 nm di spessore;
- ✓ una varietà di schemi *self-alignment* emettitore-base "presi a prestito" dalla tecnologia BJT in Si, da usare a seconda della struttura del dispositivo e dell'approccio di deposizione del SiGe (*single-poly*, *double-poly*, etc.). Tutti gli schemi utilizzano qualche tipo di *spacer* emettitore-base (0.1-0.3um di larghezza);
- ✓ un'impiantazione di collettore locale usata per migliorare le prestazioni di J_C e rendere possibile il *tuning* della tensione di breakdown;
- ✓ i contatti della base estrinseca (di solito lo strato epitassiale di SiGe depositato sullo *shallow trench*) con impianti aggiuntivi della base estrinseca per ridurre la *sheet resistance* totale;
- ✓ una base estrinseca silicidata (5-10 $\Omega/\square$);
- ✓ un emettitore di polySi fortemente drogato ($> 5 \cdot 10^{20}$ cm^{-3}), impiantato o drogato *in-situ* (150-200nm di spessore);
- ✓ metallizzazioni multilivello tipo *back end of the line* (BEOL), con Al o Cu. Questi passi sono tipicamente presi dai processi CMOS e possono prevedere da 3 a 6 livelli. Usualmente consistono di piccoli "perni" di tungsteno tra i layers di metallo (*vias*), che prevedono l'impiego di *interlayers* di ossido.

Questi elementi di tecnologia possono essere visualizzati nella *cross section* dello schema di un SiGe HBT di prima generazione mostrato in figura 2.23.

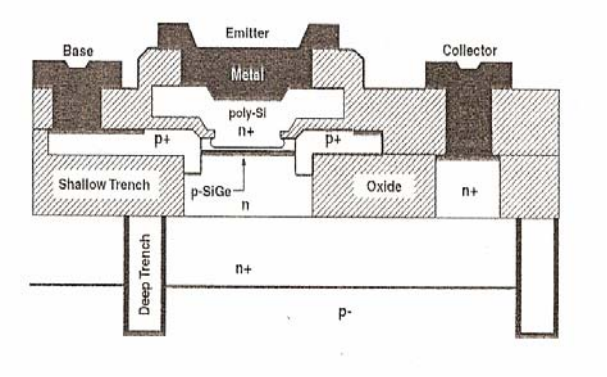

Figura 2.23

Come regola generale, con riferimento alle varie generazioni della tecnologia SiGe, le tecnologie in SiGe della prima generazione sono attualmente impiegate per supportare necessità circuitali per i mercati RF dei cellulari a 900 MHz e a 2.4GHz (GSM e CDMA), applicazioni Ethernet (1-2.5Gbit/sec), Bluetooh, 4-6GHz WLAN, GPS e applicazioni SONET (*synchronous optical networks*) a 10 Gbit/sec. Le tecnologie SiGe della seconda generazione sono impiegate nelle reti a 40 Gbit/sec e nei sistemi a microonde (10 GHz), mentre le tecnologie SiGe della terza generazione vengono impiegate per reti a 80 Gbit/sec e sistemi di comunicazione ISM-band (60 GHz).

2.4 Integrazione dei SiGe HBT nella tecnologia CMOS

Uno dei vantaggi più importanti della tecnologia SiGe rispetto alle tecnologie III-V è la sua capacità di integrarsi con la tecnologia convenzionale CMOS. E' chiaro che, dal punto di vista della compatibilità di processo, della produzione e, infine, dei costi, la capacità di realizzare la tecnologia SiGe negli impianti di fabbricazione della tradizionale tecnologia CMOS rappresenta un enorme vantaggio.

Ovviamente, l'obiettivo non è quello di far competere le tecnologie SiGe con quella CMOS, ma farle coesistere e sfruttare gli schemi di processo della tecnologia CMOS anche per la tecnologia SiGe, la qual cosa si traduce in una notevole riduzione dei costi.

La già citata compatibilità tra gli HBT in SiGe e la CMOS in Si richiede ovviamente un attento progetto del flusso strutturale e di processo al fine di produrre una robusta tecnologia BiCMOS con gli HBT in SiGe. L'aspetto fondamentale in questo contesto è di realizzare l'integrazione suddetta senza degradare le prestazioni dei SiGe HBT e senza perturbare le caratteristiche dei dispositivi CMOS. In particolare, quest'ultimo aspetto è di fondamentale importanza in quanto consente di preservare i tools di modellizzazione e le librerie di design CMOS. Storicamente, sono stati usati due differenti schemi d'integrazione per produrre le BiCMOS HBT in SiGe, ciascuno dei quali con i suoi vantggi e svantaggi: l'integrazione "*base-during-gate*" (BDG) e l'integrazione "*base-after-gate*" (BAG). Lo schema BDG (o i suoi derivati) è stato ampiamente utilizzato per produrre la prima generazione della tecnologia BiCMOS SiGe HBT, mentre lo schema BAG è più facilmente applicabile alle tecnologie della seconda e terza generazione.

Nello schema BDG (anche noto in letteratura come "base=gate") i SiGe HBT condividono i *layers* e i cicli termici CMOS al fine di ridurre la complessità strutturale (figura 2.24).

In questo caso, la base epitassiale in SiGe (e il drogaggio di boro in essa contenuto) è sottoposta all'intero ciclo termico CMOS, la qual cosa porta ad un allargamento del profilo di base.

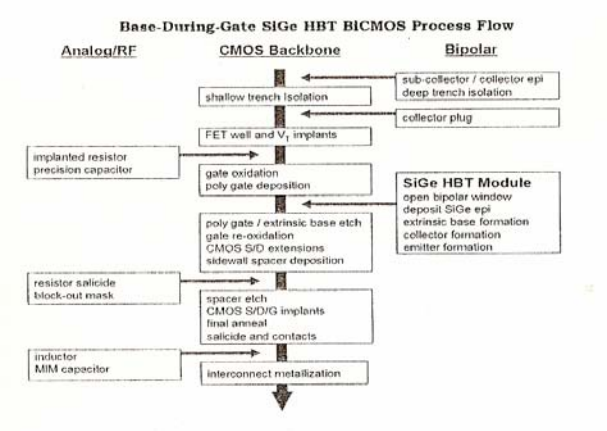

Figura 2.24

Per un ulteriore *scaling* della tecnologia (alla seconda e terza generazione), l'approccio BDG (figura 2.25) diventa più problematico a causa dei cicli termici intrinsecamente grandi associati al processo CMOS.

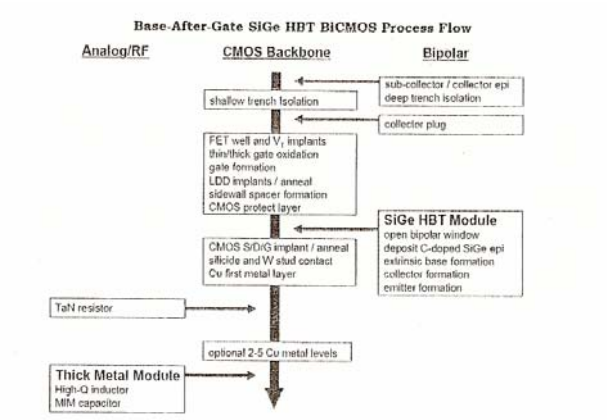

Figura 2.25

Nello schema d'integrazione BAG, i dispositivi CMOS sono completati prima che la base epitassiale in SiGe sia depositata, disaccoppiando in tal modo la realizzazione dei due tipi di dispositivo. Questo approccio rende più semplice una "copia" *step by step* dei passi di fabbricazione CMOS sottostanti a partire dalla preesistente tecnologia CMOS. L'approccio BAG, inoltre, consente più facilmente l'incorporazione dei derivati dei dispositivi CMOS (cioè circuiti CMOS a tensione più alte per I/O e/o circuiti analogici RF). L'unico ciclo termico condiviso tra HBT e CMOS è l'*annealing* finale dell'emettitore e

questo rappresenta un aspetto positivo per l'ottimizzazione del profilo del SiGe HBT. Naturalmente, gli svantaggi dell'approccio BAG sono legati alla maggiore complessità, soprattutto perché gli strati dei bipolari sono depositati sopra la topografia CMOS e successivamente devono essere rimossi.

Anche con la sostanziale riduzione dei cicli termici fornita dallo schema d'integrazione BAG della seconda generazione, l'evoluzione dei livelli di prestazione dei SiGe HBT dalla seconda alla terza generazione ($f_T \geq 200\text{GHz}$) richiede variazioni strutturali che eliminano ogni passo di impiantazione della base estrinseca nello strato epitassiale di SiGe depositato durante il modulo HBT. Queste impiantazioni introducono interstizi nella regione attiva del dispositivo, producendo una maggiore diffusione del boro e rendendo molto difficile disaccoppiare la f_T (cioè il profilo di boro in base) dal progetto della base estrinseca. A tale riguardo, la cosiddetta struttura *raised-extrinsic-base* (figura 2.26) sembra fornire diversi vantaggi ed è stata impiegata nella realizzazione di SiGe HBT con livelli di prestazione molto interessanti.

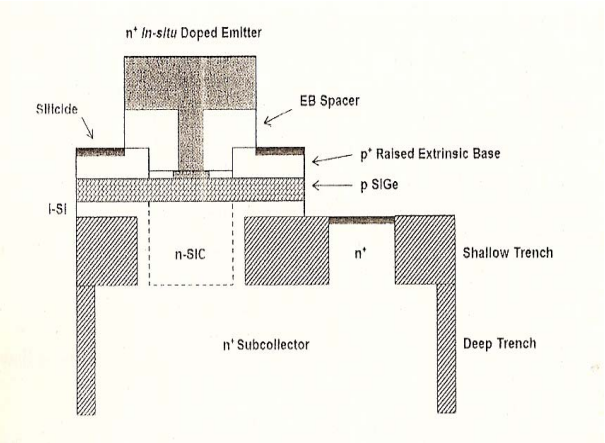

Figura 2.26

2.5 Affidabilità e resa

E' di fondamentale importanza per la tecnologia SiGe dimostrare chiaramente che la sua affidabilità e la sua resa siano confrontabili o migliori di quelle della tecnologia in Si. Ovvero, ogni perdita di affidabilità e di resa dovuta all'incorporazione dei film di SiGe *strained* può potenzialmente bloccare lo sviluppo e la diffusione di tale tecnologia. Non ci sono dati pubblicati sulle tecnologie commerciali SiGe secondo i quali l'uso di film di SiGe termodinamicamente stabili pongono rischi di affidabilità.

Le prove di affidabilità dei transistori bipolari storicamente si muovono su due differenti percorsi: 1) test con la giunzione emettitore-base (EB) inversamente polarizzata, che vengono usati per la *hot electrons injection* nell'ossido, introducendo in tal modo centri di generazione/ricombinazione (G/R) che portano ad incrementare la corrente di base e, dunque, causano la degradazione del guadagno di corrente e un aumento del *low frequency noise*; 2) test del dispositivo utilizzando un'elevata corrente diretta, che ugualmente porta ad un degrado del guadagno di corrente attribuito all'elettromigrazione indotta sul contatto di emettitore, che, a sua volta, porta ad una riduzione della corrente di collettore all'aumentare del tempo di *testing*. I test sulla durata (tempo di vita) dei SiGe HBT, eseguiti con la giunzione EB inversamente polarizzata, sono generalmente condotti sotto elevate polarizzazioni inverse (cioè 3.0V), a basse temperature (-40°C), dove le velocità dei portatori sono più elevate per la riduzione dello *scattering*, mentre le prove con elevate correnti dirette sono condotte con elevati valori di J_C vicino al picco della f_T ($1.0-2.0\,\mathrm{mA}/\mu\mathrm{m}^2$), ad elevate temperature (140°C), dove l'elettromigrazione è più severa.

Tipici dati di test con la giunzione EB inversamente polarizzata, fatti su HBT in SiGe della prima generazione, mostrano variazioni inferiori al 6% nel guadagno di corrente dopo 500 ore di test, a -40°C e con una tensione inversa EB di 2.7V.

Se si confrontano i risultati dei test, nella configurazione con la giunzione EB inversamente polarizzata, sui SiGe HBT con differenti profili di Ge e quelli dei Si BJT con base epitassiale (figura 2.27), si evince chiaramente che non esiste nessun incremento di rischio per quanto concerne affidabilità della tecnologia in SiGe.

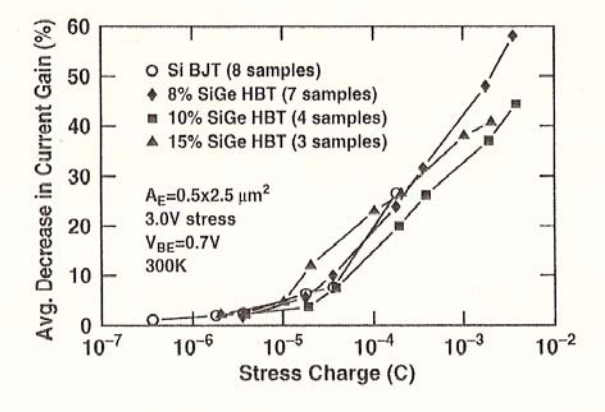

Figura 2.27

Inoltre, le prestazioni del SiGe HBT, con la giunzione EB inversamente polarizzata, risultano essere sostanzialmente migliori di quelle di un equivalente Si BJT. Come risulta evidente dalla figura 2.28, gli impianti di base superficiali a bassa energia, necessari per realizzare BJT in Si impiantati ad elevate prestazioni, presentano il picco del drogaggio di base alla giunzione metallurgica EB e, dunque, aumentano il campo elettrico alla medesima giunzione. Al contrario, per un dispositivo con base epitassiale (Si o SiGe), il morsetto B di base può essere collocato dentro la regione di base come una scatola B e, mentre il ciclo termico allunga B durante il processo, un B

"retrogrado" viene prodotto naturalmente alla giunzione EB, abbassando in tal modo il campo elettrico della giunzione. Poiché l'iniezione degli *hot electrons,* nelle condizioni di polarizzazione inversa della giunzione EB, dipende esponenzialmente dal campo elettrico alla giunzione EB, un transistore a base epitassiale avrà un vantaggio fondamentale su un dispositivo a base impiantata in termini di affidabilità.

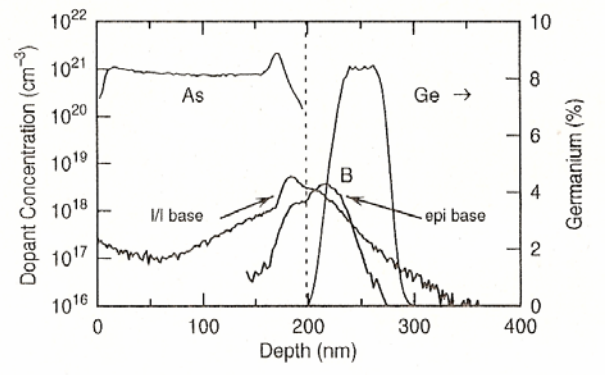

Figura 2.28

Dati tipici di test fatti sui SiGe HBT della prima generazione, caratterizzati da un'elevata corrente diretta, mostrano una variazione inferiore al 5% nel guadagno di corrente dopo 500 ore, a 140°C e $13\,\mathrm{mA}/\mu\mathrm{m}^2$. Usando fattori di accelerazione determinati empiricamente, questo risultato è teoricamente equivalente ad un più che accettabile degrado del guadagno del 10% dopo 100.000 "*power-on-hours*" (POH) sotto condizioni di "uso normale" ($1.25\,\mathrm{mA}/\mu\mathrm{m}^2$ a 100°C). Dato che lo *scaling* della tecnologia naturalmente porta a densità di corrente più elevate nei dispositivi bipolari, sarà importante quantificare queste variazioni con ogni generazione successiva della tecnologia.

La produzione dei SiGe HBT è tipicamente quantificata usando grandi catene di piccoli transistori collegati in parallelo. Un "fallimento" nella catena è

definito come l'intersezione delle seguenti condizioni: corto tra emettitore e collettore, elevato *leakage* nella giunzione EB o un elevato *leakage* nella giunzione CB. Entrambe le metodologie di integrazione BDG e BAG mostrano risultati simili. E' interessante osservare che il meccanismo di fallimento primario nei CMOS e nei SiGe HBT è lo stesso. Assumendo una distribuzione ideale di Poisson legata alla densità dei difetti e all'area di emettitore, si può dedurre la densità netta di difetti associata ad una data tecnologia BiCMOS SiGe HBT, in questo caso producendo numeri nel range 100-500 difetti/cm^2. Per fissare le idee, una densità di difetti di 426 difetti/cm^2 produrrebbe idealmente una produzione del 60% su un circuito integrato contenente 100.000 transistori SiGe HBT da $0.5x2.5\mu m^2$, la qual cosa soddisfa quasi ogni tipo di applicazione.

2.6 Caratteristiche statiche

La presenza delle eterogiunzioni Si-SiGe nelle giunzioni emettitore-base e collettore-base comporta un differente comportamento nella fisica del dispositivo tra i SiGe HBT e i Si BJT.

Le differenze operative essenziali possono essere illustrate in maniera più esaustiva considerando un diagramma a bande di energia. Per semplicità, si consideri un SiGe HBT a base graduata ideale con un drogaggio costante nelle regioni di emettitore, base e collettore. In questo caso, il contenuto di Ge è linearmente graduato dallo 0%, in prossimità della giunzione metallurgica emettitore-base (EB), al massimo valore in prossimità della giunzione metallurgica collettore-base (CB) dopodichè va giù rapidamente fino allo 0%

di Ge. I diagrammi a bande di energia risultanti per i SiGe HBT e Si BJT, in polarizzazione diretta, sono mostrati in figura 2.29. Nella figura è possibile osservare che la riduzione, indotta dalla presenza del Ge, nel bandgap di base si verifica al limite EB ($\Delta E_{g,Ge}(x=0)$) e al limite CB della regione quasi neutra di base $\Delta E_{g,Ge}(x=W_b)$.

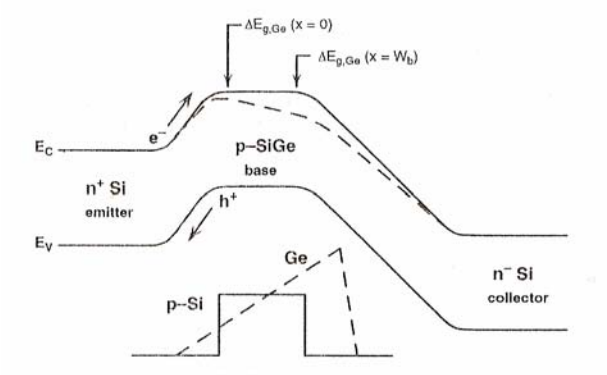

Figura 2.29

Il gradiente della concentrazione di Ge, attraverso la regione di base neutra, induce un campo di *drift* ($\left(\Delta E_{g,Ge}(x=W_b)-\Delta E_{g,Ge}(x=0)\right)/W_b$) nella base medesima che impatterà sul trasporto dei portatori minoritari.

Una prima questione logica riguarda la ragione per cui l'offset di banda di valenza, causato dall'introduzione del Ge nel Si, si trasferisce nella banda di conduzione del SiGe HBT *npn*. Per comprendere questo aspetto, è istruttivo considerare l'introduzione del *layer* di Ge graduato nella base di tipo *p* come un processo a due passi, come riportato nella figura 2.30. Per un drogaggio costante di tipo *p* nella base del Si BJT, sappiamo che sia il livello di Fermi sia la differenza di energia tra il livello di Fermi e il limite superiore della banda di valenza sono fissati. Quando il Ge è introdotto e graduato attraverso la regione neutra di base, viene indotto un offset nella banda di valenza, come riportato nello *Step 1* della figura 2.30. Sappiamo, tuttavia, che il livello di Fermi deve

riallinearsi in modo che sia fissato in energia al suo precedente valore (Si) e, inoltre, che debba essere costante (piatto) se il sistema è in equilibrio. Dunque, rispetto al caso del Si, il livello di Fermi deve diminuire in energia e appiattirsi mediante il trasporto di carica. Dato che il *bandgap* totale è fissato per un dato contenuto di Ge ad ogni posizione *x*, la conseguenza, come riportato nello *Step 2* della figura 2.30, è che il limite inferiore della banda di conduzione nella regione neutra di base è forzato ad andar giù in energia. Dunque, l'offset intrinseco della banda di valenza associato al profilo di Ge (funzione della posizione) si "trasferisce" effettivamente nella banda di conduzione del dispositivo. Questa traslazione dalla banda di valenza alla banda di conduzione è assolutamente propizia, dato che il campo di *drift* indotto, associato al limite inferiore della banda di conduzione, in queste condizioni dipende dalla posizione ed influenzerà positivamente il trasporto degli elettroni minoritari attraverso la base, come desiderato. Si noti che, appena ci si sposta dalla base neutra nella regione di carica spaziale della giunzione CB, si ritorna ad avere l'offset atteso della banda di valenza.

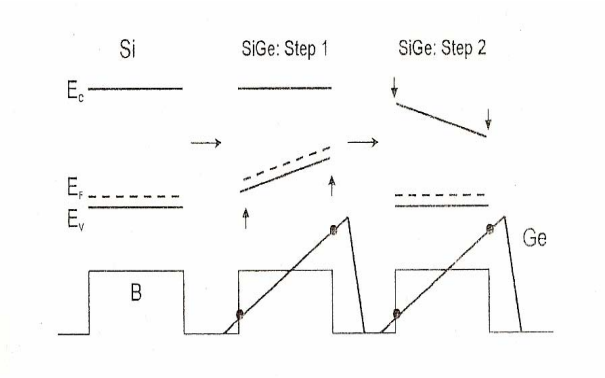

Figura 2.30

Per comprendere come le variazioni dei limiti di banda condizionano il comportamento DC del SiGe HBT, consideriamo prima il comportamento del

Si BJT. Quando viene applicata una tensione diretta V_{BE} alla giunzione EB, gli elettroni sono iniettati dall'emettitore, ricco di elettroni, nella base attraverso la barriera di potenziale EB (si veda la figura 2.29). Gli elettroni iniettati diffondono attraverso la base e sono trasportati nel campo elettrico della giunzione CB, dando vita ad una utile corrente di collettore. Allo stesso tempo, la polarizzazione diretta, applicata alla giunzione EB, produce una iniezione delle lacune dalla base nell'emettitore. Se la regione di emettitore è fortemente drogata rispetto alla base, tuttavia, la densità delle lacune iniettate sarà piccola in confronto alla densità degli elettroni iniettati nella base e, dunque, viene prodotto un guadagno di corrente finito $\beta \propto n/p$.

Come è possibile vedere in figura 2.29, l'introduzione del Ge nella regione di base ha due conseguenze tangibili dal punto di vista DC: 1) la barriera di potenziale alla giunzione EB, vista dagli elettroni che dall'emettitore vengono iniettati nella base, si riduce. Intuitivamente, si capisce come questo possa produrre un incremento esponenziale degli elettroni iniettati a parità di V_{BE}, portando così ad una più elevata corrente di collettore e, dunque, ad un più elevato guadagno, poiché la corrente di base resta invariata. Dato che gli effetti dei limiti di banda generalmente di associano fortemente alle proprietà dei transistori, ci aspettiamo una forte dipendenza di J_C dal contenuto di Ge. Come conseguenza pratica, l'introduzione del Ge effettivamente disaccoppia il drogaggio di base dal guadagno di corrente, dotando, in tal modo, i progettisti del dispositivo di una maggiore flessibilità rispetto al progetto del Si BJT. Se, per esempio, l'applicazione circuitale in considerazione non richiede alti guadagni di corrente (ad esempio, $\beta = 100$), possiamo effettivamente "scambiare" il guadagno di corrente più alto, indotto dall'offset di banda e

dovuto al Ge, per un livello di drogaggio più alto in base, ottenendo un abbassamento netto della resistenza di base e, dunque, migliori caratteristiche dinamiche di *switching* e di rumore. 2) La presenza di un contenuto finito di Ge nella giunzione CB influenzerà positivamente la conduttanza di uscita del transistore, producendo una tensione di Early più elevata. In sostanza, il bandgap di base più piccolo in prossimità della giunzione CB determina una dimensione minore dello svuotamento nella regione neutra di base all'aumentare della tensione V_{CB} (effetto Early) rispetto al Si BJT. Questo fenomeno si traduce in una più elevata tensione di Early rispetto al Si BJT.

2.6.1 Densità di corrente di collettore e guadagno di corrente

Per comprendere meglio le potenzialità dei SiGe HBT occorre prima relazionare formalmente le variazioni nella densità della corrente di collettore e, dunque, il guadagno di corrente, alle variabili fisiche del problema. E', inoltre, istruttivo confrontare attentamente le differenze tra il SiGe HBT e il Si BJT nelle stesse condizioni operative. Nella presente analisi, i due dispositivi presentano una geometria identica, i profili di drogaggio nelle regioni di emettitore, base e collettore sono identici, tranne che per la presenza del Ge nella base del SiGe HBT. Per semplicità di analisi, si assume un profilo di Ge linearmente graduato dalla giunzione EB alla giunzione CB, come descritto nella figura 2.31. Le espressioni risultanti possono essere applicate ad un'ampia varietà di profili di SiGe, dal profilo di Ge costante (*box profile*) a quello triangolare (linearmente graduato) e considerando come caso intermedio il profilo trapezoidale di Ge (una combinazione dei profili costante e

linearmente graduato). L'analisi che segue prevede, inoltre, che i dispositivi operino in condizioni di bassa iniezione, con una ricombinazione trascurabile di *bulk* e di superficie; inoltre, questa analisi viene applicata ai dispositivi SiGe HBT *npn*.

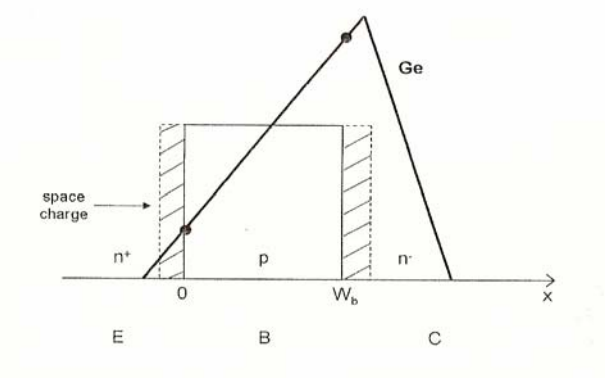

Figura 2.31

Le conseguenze teoriche delle variazioni di bandgap indotte dal Ge sulla J_C possono essere derivate in forma chiusa per un profilo di drogaggio costante in base ($p_b(x) = N_{ab}^-(x) = N_{ab}^- = const$), considerando la relazione generalizzata di Moll-Ross per la densità di corrente di collettore, che vale per basse iniezioni in presenza di un drogaggio di base non uniforme e di un bandgap in base non uniforme ad una data V_{BE} e ad una data temperatura T:

$$J_C = \frac{q\left(e^{qV_{BE}/kT} - 1\right)}{\int_0^{W_b} \frac{p_b(x)dx}{D_{nb}(x)n_{ib}^2(x)}} \tag{2.2}$$

Dove $x=0$ e $x=W_b$ sono i valori dei limiti della base neutra rispettivamente dalla parte della giunzione EB e della giunzione CB. In questo caso, il drogaggio di base è costante ma sia n_{ib} sia D_{nb} sono funzioni della posizione; il primo attraverso l'offset di banda indotto dal Ge, l'ultimo dovuto all'influenza del profilo di Ge (funzione della posizione) sulla mobilità degli elettroni

($D_{nb} = kT/q\mu_{nb} = f(Ge)$). Si noti che J_C dipende solo dalle variazioni nel *bandgap* di base indotte dal Ge. In generale, la densità intrinseca dei portatori nel SiGe HBT può essere scritta come:

$$n_{ib}^2(x) = (N_C N_V)_{SiGe}(x) e^{-E_{gb}(x)/kT} \tag{2.3}$$

Dove il termine $(N_C N_V)_{SiGe}$ rappresenta le variazioni indotte dal Ge (funzioni della posizione) associate alle densità degli stati efficaci nelle bande di valenza e conduzione. Nell'equazione (2.3) l'espressione del *bandgap* di base può essere divisa nei suoi vari contributi, come riportato in figura 2.32.

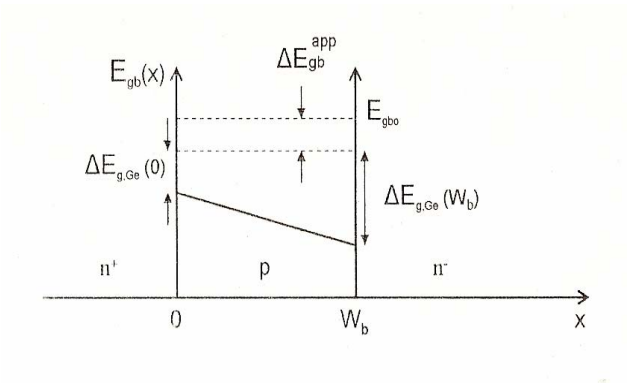

Figura 2.32

In figura, E_{gb0} è il bandgap del Si per bassi drogaggi (1.12eV a 300K), ΔE_{gb}^{app} rappresenta la riduzione apparente del bandgap nella regione di base dovuta all'elevato drogaggio, $\Delta E_{g,Ge}(0)$ è l'offset di banda dovuto al Ge all'ascissa $x=0$ e $\Delta E_{g,Ge}(W_b)$ rappresenta l'offset di banda dovuto al Ge all'ascissa $x=W_b$. Dunque, si può scrivere $E_{gb}(x)$ come:

$$E_{gb}(x) = E_{gb0} - \Delta E_{gb}^{app} + \left[\Delta E_{g,Ge}(0) - \Delta E_{g,Ge}(W_b)\right] \cdot \frac{x}{W_b} - \Delta E_{g,Ge}(0) \tag{2.4}$$

Sostituendo la (2.4) nella (2.3) si ottiene:

$$n_{ib}^2(x) = \gamma n_{i0}^2 e^{\Delta E_{gb}^{app}/kT} e^{\left[\Delta E_{g,Ge}(W_b) - \Delta E_{g,Ge}(0)\right]x/(W_b kT)} e^{\Delta E_{g,Ge}(0)/kT} \tag{2.5}$$

dove si è fatto uso del fatto che per il Si possiamo definire una densità intrinseca dei portatori per bassi drogaggi come:

$$n_{i0}^2 = N_C N_V e^{-E_{g0}/kT} \tag{2.6}$$

e dove abbiamo definito un rapporto efficace di densità degli stati tra SiGe e Si come:

$$\gamma = \frac{\left(N_C N_V\right)_{SiGe}}{\left(N_C N_V\right)_{Si}} < 1 \tag{2.7}$$

L'equazione (2.5) può essere inserita nella relazione generalizzata di Moll-Ross per ottenere:

$$J_C = \frac{q\,\widetilde{D}_{nb}\left(e^{qV_{BE}/kT} - 1\right)\widetilde{\gamma} n_{i0}^2 e^{\Delta E_{gb}^{app}/kT} e^{\Delta E_{g,Ge}(0)/kT}}{N_{ab}^- \cdot \int e^{-\left[\Delta E_{g,Ge}(W_b) - \Delta E_{g,Ge}(0)\right](x/W_b kT)} dx} \tag{2.8}$$

dove si sono definite $\widetilde{D}_{nb}$ e $\widetilde{\gamma}$ come quantità medie rispetto alla posizione attraverso la base, secondo la relazione :

$$\widetilde{D}_{nb} = \frac{\int_0^{W_b} \frac{dx}{n_{ib}^2(x)}}{\int_0^{W_b} \frac{dx}{D_{nb}(x) n_{ib}^2(x)}} \tag{2.9}$$

Usando le tecniche di integrazione standard e definendo

$$\Delta E_{g,Ge}(grade) = \Delta E_{g,Ge}(W_b) - \Delta E_{g,Ge}(0) \tag{2.10}$$

si ottiene:

$$J_{C,SiGe} = \frac{q\widetilde{D}_{nb}}{N_{ab}^-} \cdot \frac{\left(e^{qV_{BE}/kT} - 1\right)\widetilde{\gamma} n_{i0}^2 e^{\Delta E_{gb}^{app}/kT} e^{\Delta E_{g,Ge}(0)/kT}}{\frac{W_b kT}{\Delta E_{g,Ge}(grade)}\left\{1 - e^{-\Delta E_{g,Ge}(grade)/kT}\right\}} \tag{2.11}$$

Infine, definendo un rapporto tra le diffusività degli elettroni minoritari del SiGe e del Si come:

$$\tilde{\eta} = \frac{\left(\tilde{D}_{nb}\right)_{SiGe}}{\left(D_{nb}\right)_{Si}} \tag{2.12}$$

si ottiene l'espressione finale per $J_{C,SiGe}$:

$$J_{C,SiGe} = \frac{qD_{nb}}{N_{ab}^{-}W_b} \cdot \left(e^{qV_{BE}/kT} - 1\right) n_{i0}^2 e^{\Delta E_{gb}^{app}/kT} \cdot \left\{ \frac{\tilde{\gamma}\tilde{\eta}\Delta E_{g,Ge}(grade)/kTe^{\Delta E_{g,Ge}(0)/kT}}{1 - e^{-\Delta E_{g,Ge}(grade)/kT}} \right\} \tag{2.13}$$

Nei limiti delle assunzioni fatte precedentemente, questo può essere considerato un risultato esatto. Dai risultati ottenuti è possibile osservare che J_C, in un SiGe HBT, dipende esponenzialmente dal valore del limite EB dell'offset di banda indotto dal Ge ed è proporzionale al gradiente del *bandgap* indotto dal Ge. Data la natura della dipendenza esponenziale, è ovvio che è possibile ottenere un forte aumento di J_C, per una data V_{BE}, per piccole quantità di Ge introdotto e che è possibile progettare le caratteristiche del dispositivo per ottenere il guadagno di corrente desiderato. Si noti che il termine rappresentativo dell'energia termica *kT* compare al denominatore degli offset di banda indotti dal Ge. Anche questo risultato era prevedibile a partire dalla considerazione che gli effetti dei limiti di banda generalmente si accoppiano alle equazioni di trasporto del dispositivo.

Se si considera un SiGe HBT e un Si BJT con identica tecnologia di contatto degli emettitori e, inoltre, si assume che il profilo di Ge, dalla parte della giunzione EB della regione neutra di base, non si estenda troppo nell'emettitore a tal punto da cambiare la densità della corrente di base, le aspettative dei risultati sperimentali sono che, per un SiGe HBT e un Si BJT costruiti in modo da confrontarli, la J_B dovrebbe essere comparabile, mentre J_C, per una data

V_{BE}, dovrebbe aumentare nel caso di SiGe HBT. La figura 2.33 conferma sperimentalmente queste aspettative.

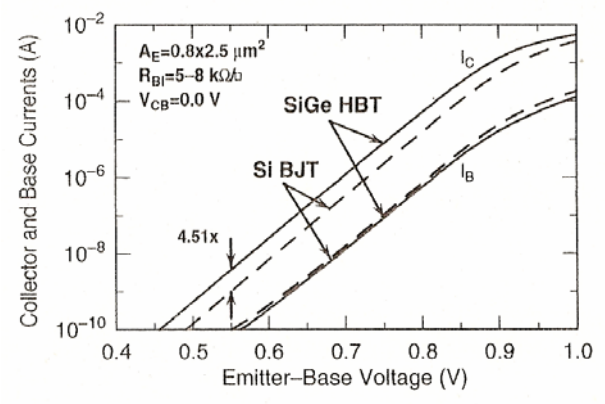

Figura 2.33

In questo caso, si nota che il rapporto del guadagno di corrente tra i SiGe HBT e i Si BJT identicamente costruiti può essere scritto come:

$$\frac{\beta_{SiGe}}{\beta_{Si}} \cong \frac{J_{C,SiGe}}{J_{C,Si}} \tag{2.14}$$

e dunque possiamo definire un fattore incrementale del guadagno di corrente come:

$$\left.\frac{\beta_{SiGe}}{\beta_{Si}}\right|_{V_{BE}} \equiv \Xi \equiv \left\{ \frac{\tilde{\gamma}\tilde{\eta}\,\Delta E_{g,Ge}(grade)/kTe^{\Delta E_{g,Ge}(0)/kT}}{1-e^{-\Delta E_{g,Ge}(grade)/kT}} \right\} \tag{2.15}$$

Risultati sperimentali tipici per Ξ sono mostrati in figura 2.34 e sono confrontati con i risultati teorici calcolati facendo uso della formula (2.15).

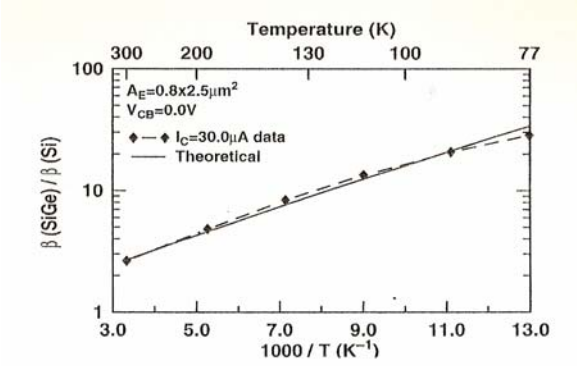

Figura 2.34

A questo punto è possibile fare alcune approssimazioni. Primo, è possibile assumere che $\Delta E_{g,Ge}(grade) >> kT$. Questa approssimazione può essere considerata come lo scenario del "forte gradiente di Ge". In questo caso la (2.13) si riduce a:

$$J_{C,SiGe} \cong \frac{qD_{nb}}{N_{ab}^{-}W_b}\left(e^{qV_{BE}/kT}-1\right)n_{i0}^2 e^{\Delta E_{gb}^{app}/kT}\cdot\left\{\tilde{\gamma}\tilde{\eta}\,\frac{\Delta E_{g,Ge}(grade)}{kT}e^{\Delta E_{g,Ge}(0)/kT}\right\} \qquad (2.16)$$

Si noti, tuttavia, che occorre utilizzare questa espressione con prudenza. Per controllare la sua validità per un profilo realistico, si assuma di avere un profilo di Ge triangolare dallo 0% al 15% in un SiGe HBT che opera a 300 K. Considerando un offset di banda di circa 75meV per una concentrazione di Ge del 10%, troviamo che $\Delta E_{g,Ge}(grade)/kT = 4.3$ e, dunque, l'approssimazione iniziale risulta essere non del tutto rispettata. Chiaramente, però, non appena la temperatura diminuisce, la validità di questa approssimazione migliora rapidamente non appena il termine *kT* diminuisce. Per esempio, un profilo triangolare di Ge dallo 0% al 15% produce $\Delta E_{g,Ge}(grade)/kT = 17$ a 77 K.

In aggiunta al profilo fortemente graduato, è possibile definire allo stesso modo un'approssimazione di "debole gradiente di Ge", che dovrebbe essere valida, per esempio, nel caso di profilo costante (*box profile*) di Ge. In questo caso si ha $\Delta E_{g,Ge}(grade) << kT$. Espandendo l'esponenziale del gradiente di Ge al denominatore della (2.13) in serie di Taylor e semplificando si ottiene:

$$J_{C,SiGe} \cong \frac{qD_{nb}}{N_{ab}^{-}W_b}\left(e^{qV_{BE}/kT}-1\right)n_{i0}^2 e^{\Delta E_{gb}^{app}/kT}\left\{\tilde{\gamma}\tilde{\eta}e^{\Delta E_{g,Ge}(0)/kT}\right\} \qquad (2.17)$$

2.6.2 Drogaggio di base non costante

In generale, non esiste una soluzione in forma chiusa per $J_{C,SiGe}$ se sia il drogaggio di base sia il profilo di Ge dipendono dalla posizione. Nel caso di drogaggio di base non costante è possibile, tuttavia, definire una "riduzione di bandgap efficace indotta dal Ge" ($\Delta E_{g,Ge}(eff)$) secondo la relazione:

$$n_{ib}^2(x) = \tilde{\gamma} n_{i0}^2 e^{\Delta \tilde{E}_{gb}^{app}/kT} e^{\Delta \tilde{E}_{g,Ge}(eff)/kT} \tag{2.18}$$

dove la tilde si riferisce, ancora una volta, ad una quantità mediata sulla posizione. Fisicamente, $\Delta E_{g,Ge}(eff)$ può essere pensato come un offset di banda medio di Ge attraverso la base neutra. Dunque, si può definire la *sheet resistance* della base intrinseca come:

$$R_{bi} = \left\{ \int_0^{W_b} q \mu_{pb}(x) N_{ab}^-(x) dx \right\}^{-1} \tag{2.19}$$

Osserviamo che R_{bi} è l'integrale della carica di base neutra ed è un importante parametro perché è direttamente misurabile attraverso una struttura di test indipendente *on-wafer*. Dunque, si vede che il fattore incrementale efficace del guadagno di corrente per un SiGe HBT con un arbitrario profilo di drogaggio in base è dato dall'espressione:

$$\Xi_{eff} = \frac{J_{C,SiGe}(eff)}{J_{C,Si}} = \frac{\tilde{\mu}_{nb,SiGe} \tilde{\mu}_{pb,SiGe}}{\tilde{\mu}_{nb,Si} \tilde{\mu}_{pb,Si}} \frac{R_{bi,SiGe}}{R_{bi,Si}} \cdot \left\{ e^{\left(\Delta \tilde{E}_{gb,SiGe}^{app} - \Delta \tilde{E}_{gb,Si}^{app}\right)/kT} e^{\Delta \tilde{E}_{g,Ge}(eff)/kT} \right\} \tag{2.20}$$

dove è stata considerata la possibilità di una differenza nel profilo di drogaggio di base tra i SiGe HBT e i Si BJT. L'equazione (2.20) è utile perché consente di confrontare i dati elettrici relativi ai due tipi di dispositivi per dedurre informazioni sul profilo di Ge. Come esempio, se supponiamo $R_{bi,SiGe} \cong R_{bi,Si}$, allora la (2.20) diventa:

$$\Xi_{eff}(T) \cong e^{\Delta\tilde{E}_{g,Ge}(eff)/kT} \tag{2.21}$$

Dunque, se si misurano $J_{C,SiGe}$ e $J_{C,Si}$ per una data V_{BE} come funzioni della temperatura per due identiche geometrie di emettitore, il grafico di $\log\Xi_{eff}(T)$ in funzione di *1000/T* sarà lineare e, dunque, permetterà una determinazione sperimentale di $\Delta E_{g,Ge}(eff)$. Confrontando Ξ_{eff} con Ξ per uno specifico profilo di Ge, si possono dedurre, elettricamente, informazioni sulla forma del profilo di Ge a partire dai dati della corrente di collettore. Per esempio, per un profilo di Ge triangolare troviamo:

$$\Delta\tilde{E}_{g,Ge}(eff) \cong \Delta\tilde{E}_{g,Ge}(0) + kT\ln\left\{\Delta\tilde{E}_{g,Ge}(grade)/kT\right\} \tag{2.22}$$

2.6.3 Altri profili SiGe

L'analisi condotta finora vale per un *range* di profili di Ge che va dal triangolare a quello costante. Esiste, tuttavia, una classe di profili di Ge tecnologicamente importanti che possono essere considerati combinazioni ibride dei profili costante e triangolare, detti trapezoidali (figura 2.35).

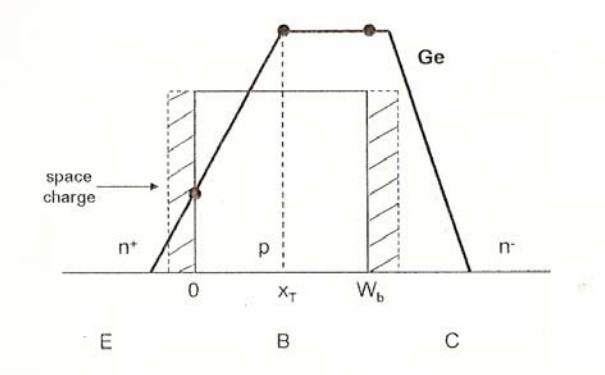

Figura 2.35

In questo caso, si prende un profilo lineare e lo si tronca ad una certa posizione intermedia x_T all'interno della base neutra, mantenendo così costante il

contenuto di Ge da x_T a W_b e, successivamente, portandolo rapidamente a zero, come al solito. Questo approccio di profilo trapezoidale permette di indurre un gradiente di Ge più alto in corrispondenza del limite della regione di base relativo alla giunzione EB maggiormente drogata, mantenendo così una buona risposta dinamica con un picco di Ge più basso. La regione di profilo costante di Ge nella base neutra, almeno in principio, non degrada le prestazioni AC poiché il lato CB della base neutra, tipicamente, avrà un campo di *drift* indotto dal gradiente del drogaggio, in aggiunta al campo di *drift* indotto dal gradiente di Ge, che contribuirà al trasporto degli elettroni. Per un profilo trapezoidale di Ge, è possibile derivare le espressioni per la densità di corrente di collettore in presenza di un drogaggio di base costante. In questo caso, gli offset di banda indotti dal Ge possono essere scritti come:

$$\Delta E_{g,Ge}(x)=\begin{cases}\Delta E_{g,Ge}(0)+\Delta E_{g,Ge}(grade)\left(\dfrac{x}{x_T}\right) & 0\le x\le x_T\\ \Delta E_{g,Ge}(W_b) & x_T\le x\le W_b\end{cases} \tag{2.23}$$

e la densità dei portatori intrinseci è, dunque:

$$n_{ib}^2(x)=\begin{cases}\gamma n_{i0}^2 e^{\Delta E_{gb}^{app}/kT} e^{\Delta E_{g,Ge}(0)/kT} e^{\left[\Delta E_{g,Ge}(grade)x/x_T\right]/kT} & 0\le x\le x_T\\ \gamma n_{i0}^2 e^{\Delta E_{gb}^{app}/kT} e^{\Delta E_{g,Ge}(W_b)/kT} & x_T\le x\le W_b\end{cases} \tag{2.24}$$

Se si divide l'integrale di Kummel nell'equazione di Moll-Ross in due parti, si integra e si confronta il risultato con quello ottenuto per il Si BJT, alla fine si ottiene:

$$\frac{J_{C,SiGe}}{J_{C,Si}}=\frac{\tilde{\gamma}\tilde{\eta}e^{\Delta E_{g,Ge}(0)/kT}}{\dfrac{\xi kT}{\Delta E_{g,Ge}(grade)}+\left\{1-\xi\left(1+\dfrac{kT}{\Delta E_{g,Ge}(grade)}\right)\right\}e^{-\Delta E_{g,Ge}(grade)/kT}} \tag{2.25}$$

dove è stato definito $\xi = x_T / W_b < 1$ come il punto intermedio trapezoidale normalizzato. In questo caso, $\xi = 0$ corrisponde al profilo costante puro del Ge (*box profile*) e $\xi = 1$ al profilo triangolare puro del Ge. Inoltre, assumendo che il rapporto del guadagno di corrente tra un SiGe HBT e un Si BJT sia semplicemente uguale al rapporto delle J_C tra i dispositivi, è possibile tracciare il grafico del guadagno di corrente come una funzione del reciproco della temperatura al variare di ξ, come mostrato nella figura 2.36 (si considera un contenuto di Ge del 10% all'ascissa $x=W_b$). Dalla figura si evince che il risultato per il profilo trapezoidale di Ge giace tra quello corrispondente ai profili costante e triangolare.

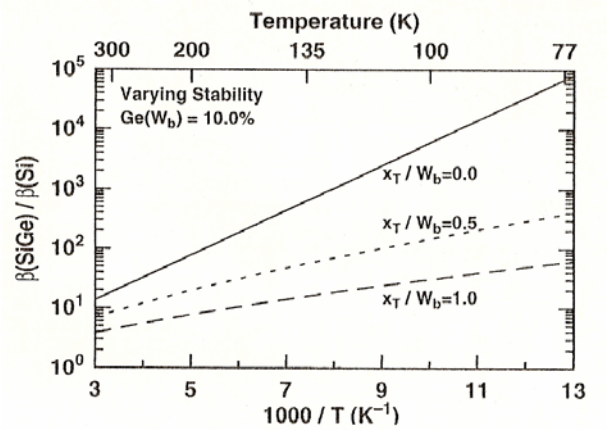

Figura 2.36

Si noti che, facendo semplicemente variare ξ, il contenuto di Ge integrato per i profili cambia allo stesso modo e, dunque, la stabilità del film di Ge diminuirà al decrescere di ξ. Alternativamente, si può fissare il contenuto di Ge a *x=0* e dunque far variare il contenuto di Ge all'ascissa $x=W_b$ in modo che il contenuto totale integrato di Ge rimanga costante. Il fattore incrementale di guadagno in questo caso (per un contenuto di Ge del 2% a *x=0*) è mostrato in figura 2.37.

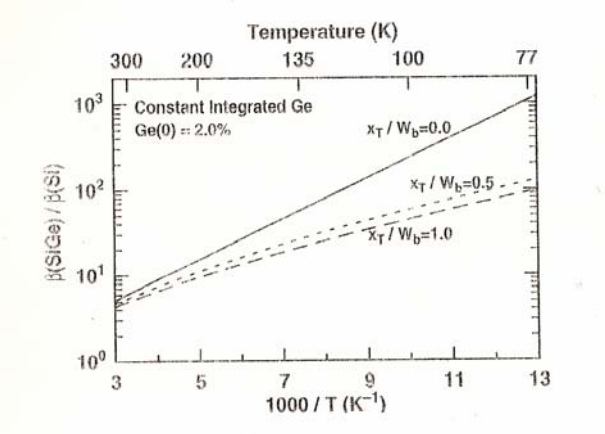

Figura 2.37

2.6.4 Ottimizzazione di β

A partire dall'analisi precedente, è possibile ora fare diverse osservazioni riguardanti gli effetti del Ge sulla corrente di collettore in un SiGe HBT.

- ✓ La presenza del Ge, in qualsiasi forma del profilo di concentrazione, nella base di un transistore bipolare incrementa J_C per una data V_{BE} rispetto ad un equivalente Si BJT.
- ✓ L'incremento in J_C dipende esponenzialmente dal valore al limite EB dell'offset di banda indotto dal Ge e linearmente dal gradiente di Ge attraverso la base. Questa dipendenza gioca un ruolo importante nella determinazione del profilo ottimo di Ge.
- ✓ Alla luce di ciò, il profilo di Ge costante (*box profile*) risulta migliore per l'incremento del guadagno di corrente rispetto al profilo triangolare, mantenendo inalterati tutti gli altri parametri.
- ✓ L'incremento di J_C indotto dal Ge è termicamente attivato (dipende esponenzialmente dal reciproco della temperatura) e, dunque, una

riduzione della temperatura produrrà una forte amplificazione di tale incremento.

2.6.5 Conduttanza di uscita

La conduttanza di uscita dinamica ($\partial I_C / \partial V_{CE}$ ad una data V_{BE}) di un transistore è un parametro di progetto critico per molti circuiti analogici. Intuitivamente, riguardo le caratteristiche di uscita del transistore, si vorrebbe che la corrente di uscita fosse indipendente dalla tensione di uscita e, dunque, avesse idealmente una conduttanza di uscita nulla (ovvero una resistenza di uscita infinita). Nella pratica, naturalmente, questo non accade. Aumentando la V_{CB} si svuota la base neutra nella regione di base, spostando così il valore di confine della base neutra ($x=W_b$) verso l'interno. Poiché W_b determina la densità di portatori minoritari dalla parte CB della base neutra, la pendenza del profilo di concentrazione degli elettroni minoritari e, dunque, la corrente di collettore, necessariamente crescono. Quindi, per un drogaggio di base finito, I_C deve aumentare all'aumentare di V_{CB}, fornendo così una conduttanza di uscita finita. Questo meccanismo è noto come "effetto Early" e, per convenienza sperimentale, definiamo una tensione di Early come:

$$V_A = J_C(0)\left\{\left.\frac{\partial J_C}{\partial V_{CB}}\right|_{V_{BE}}\right\}^{-1} - V_{BE} \cong J_C(0)\left\{\left.\frac{\partial I_C}{\partial W_b}\right|_{V_{BE}} \frac{\partial W_b}{\partial V_{CB}}\right\}^{-1} \tag{2.26}$$

dove $J_C(0) = J_C$ ($V_{CB} = 0V$). La tensione di Early è una misura semplice e conveniente della variazione della conduttanza di uscita al variare della V_{CB}. Una rappresentazione schematica dell'effetto Early, e la definizione di V_A, è mostrata nella figure 2.38 e 2.39.

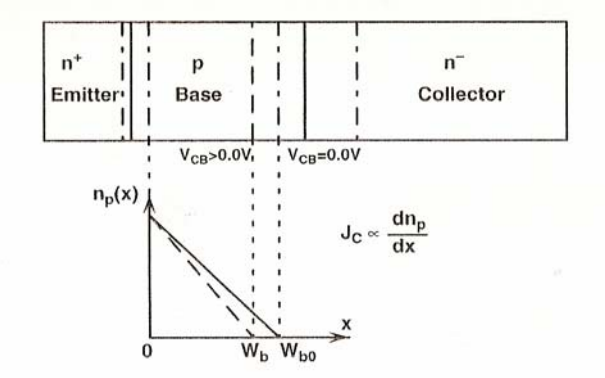

Figura 2.38

Come si vedrà successivamente, è particolarmente difficile mantenere un elevato guadagno di corrente, un'elevata risposta in frequenza ed un'elevata V_A in un Si BJT.

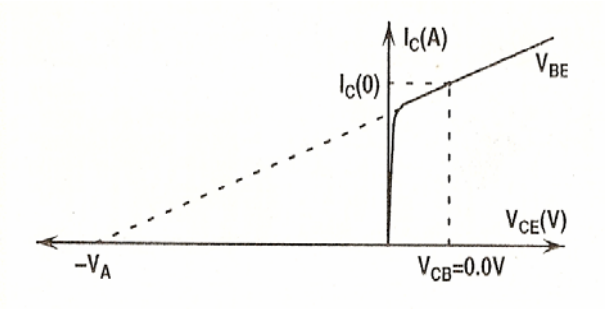

Figura 2.39

Per un Si BJT, dalle equazioni precedenti è possibile ottenere:

$$V_{A,Si} = \frac{\int_0^{W_b} p_b(x)dx}{p_b(W_b)\left\{\frac{\partial W_b}{\partial V_{CB}}\right\}} = \frac{Q_b(0)}{C_{cb}} \tag{2.27}$$

dove $Q_b(0)$ è la carica di base totale a $V_{CB}=0V$, C_{cb} è la capacità di svuotamento collettore-base; inoltre, si assume che V_{BE} sia trascurabile rispetto a V_{CB}. Si noti che C_{cb} è indipendente dal drogaggio ionizzato di collettore (N_{dc}^{+}) e dal drogaggio ionizzato di base (N_{ab}^{-}). Per stimare la sensitività di V_A

su N_{dc}^{+} e N_{ab}^{-}, si può considerare un Si BJT con profili di drogaggio di base e di collettore costanti. In questo caso, è possibile scrivere:

$$V_{A,si} = -W_b(0)\left\{\left.\frac{\partial W_b}{\partial V_{CB}}\right|_{V_{BE}}\right\}^{-1} \tag{2.28}$$

dove $W_b(0)$ è la larghezza della base neutra per $V_{CB}=0V$. La dipendenza di W_b dalla tensione e dal drogaggio può essere ottenuta dalla seguente equazione:

$$W_b \cong W_m - \sqrt{\left(\frac{2\varepsilon}{q}\right)(\phi_{bi} + V_{CB})\left\{\frac{N_{dc}^{+}}{N_{ab}^{-}\left(N_{ab}^{-} + N_{dc}^{+}\right)}\right\}} \tag{2.29}$$

dove W_m è la larghezza della base metallurgica, ϕ_{bi} è la tensione intrinseca della giunzione CB. Usando le (2.28) e (2.29) si può calcolare V_A come funzione del drogaggio, come mostrato in figura 2.40 (W_m=100nm, ΔV_{CB}=1V).

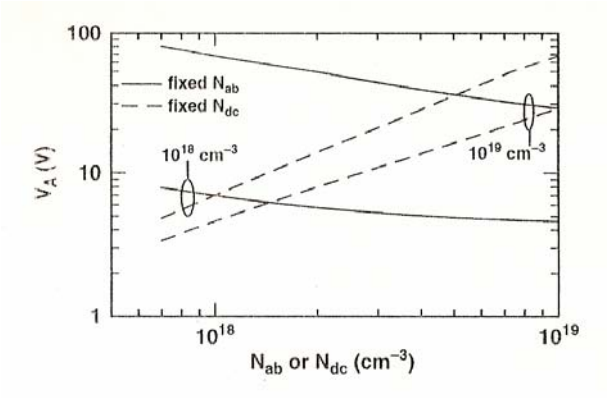

Figura 2.40

Se si fissa N_{ab}^{-}, si osserva una diminuzione di V_A all'aumentare di N_{dc}^{+}, questo perché la quantità di svuotamento di base per unità di polarizzazione è aumentata per un drogaggio di collettore più alto. Se, invece, si fissa N_{dc}^{+}, si osserva un aumento rapido di V_A all'aumentare di N_{ab}^{-}, la qual cosa ha senso perché la base si svuota molto più difficilmente all'aumentare del drogaggio di base, mantenendo inalterato tutto il resto. Nel progetto dei Si BJT, un dato

dispositivo ha generalmente una specifica tensione di *breakdown* collettore-emettitore (BV_{CEO}) determinata dalle richieste del circuito. In prima approssimazione, questa BV_{CEO} determina il livello di drogaggio del collettore. Mentre questo fatto può apparire favorevole nell'ottenere elevate V_A, si deve, tuttavia, sottolineare che il guadagno di corrente è reciprocamente correlato alla carica di base integrata (vedi (2.2)). Dunque, l'aumento di N_{ab}^{-} per migliorare V_A risulta in una forte riduzione di β. Inoltre, per un Si BJT e per una data larghezza di base, un aumento di N_{ab}^{-} degraderà la frequenza di *cutoff* del transistore (a causa della riduzione della mobilità degli elettroni minoritari). Si può immaginare di poter aumentare N_{dc}^{+} per recuperare le prestazioni AC perdute, ma, come si può vedere in figura 2.41, questo a sua volta degrada la V_A. Questo aspetto rappresenta un problema fondamentale nel progetto di un Si BJT: è intrinsecamente difficile ottenere simultaneamente un elevata V_A, un elevato β e un'elevata f_T. In pratica, occorre trovare un compromesso nel design per V_A, β e f_T. Intuitivamente, i vincoli di progetto di un Si BJT si verificano perché β e V_A sono entrambe legate al profilo di drogaggio in base.

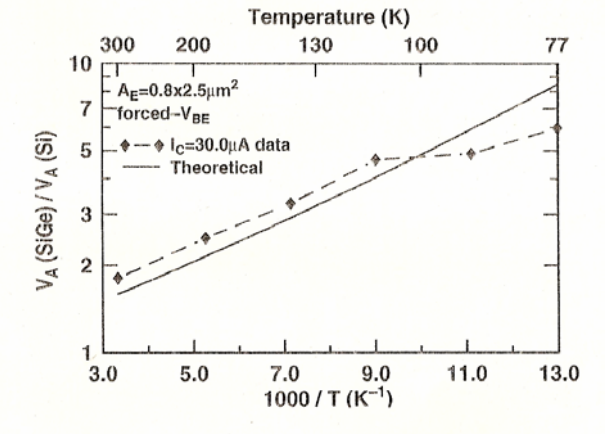

Figura 2.41

L'introduzione di Ge nella regione di base di un Si BJT può alterare questo vincolo in modo favorevole disaccoppiando di fatto β e V_A dal profilo di drogaggio in base.

Per ricavare formalmente V_A in un SiGe HBT, si possono combinare la (2.2) e la (2.26) per ottenere:

$$V_{A,SiGe} = \frac{-\int_0^{W_b} \frac{p_b(x)dx}{D_{nb}(x)n_{ib}^2(x)}}{\frac{\partial}{\partial V_{CB}}\left\{\int_0^{W_b} \frac{p_b(x)dx}{D_{nb}(x)n_{ib}^2(x)}\right\}} \tag{2.30}$$

da cui si può scrivere:

$$V_{A,SiGe} = \left\{\frac{-D_{nb}(W_b)n_{ib}^2(W_b)}{p_b(W_b)} \cdot \int_0^{W_b} \frac{p_b(x)dx}{D_{nb}(x)n_{ib}^2(x)}\right\}\left[\frac{\partial W_b}{\partial V_{CB}}\right]^{-1} \tag{2.31}$$

Confrontando la (2.25) e la (2.31), si può vedere che la differenza fondamentale tra la V_A in un SiGe HBT e in un Si BJT nasce dalla variazione di n_{ib}^2 come funzione della posizione (la variazione di W_b con V_{CB} è, al primo ordine, per entrambi i dispositivi in SiGe e in Si). Si osservi che se n_{ib} è indipendente dalla posizione (nel caso di profilo costante), allora la (2.31) si riduce alla (2.27) e non si manifesta alcun aumento di V_A dovuto al Ge (benché si verifichi comunque un forte incremento di β). D'altra parte, se n_{ib} è funzione della posizione (nel caso di profilo di Ge linearmente graduato), V_A dipenderà esponenzialmente dalla differenza nel bandgap tra $x=W_b$ e quella regione all'interno della base dove n_{ib} è più piccola. Ovvero, il profilo di base è effettivamente "pesato" dal contenuto di Ge crescente dal lato della base neutra che affaccia sul collettore, rendendo più difficile svuotare la base

neutrale per una data V_{CB} applicata, aumentando così effettivamente la tensione di Early del transistore.

Per un profilo di Ge linearmente graduato, si può usare la (2.5) e la (2.31) per ottenere il rapporto di V_A tra il SiGe HBT e il Si BJT:

$$\left.\frac{V_{A,SiGe}}{V_{A,Si}}\right|_{V_{BE}} \equiv \Theta \cong e^{\Delta E_{g,Ge}(grade)/kT}\left[\frac{1-e^{-\Delta E_{g,Ge}(grade)/kT}}{\Delta E_{g,Ge}(grade)/kT}\right] \tag{2.32}$$

Il risultato importante è che il rapporto delle V_A tra un SiGe HBT e un Si BJT è una funzione esponenziale del gradiente del *bandgap* indotto dal Ge attraverso la base neutrale. Tipici risultati sperimentali per Θ sono mostrati in figura 2.42 per un SiGe HBT e un Si BJT costruiti in modo comparabile.

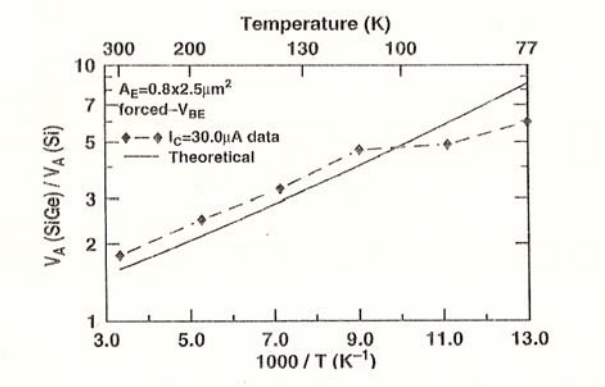

Figura 2.42

Calcoli teorici ottenuti usando la (2.32) come funzione del profilo di Ge sono mostrati in figura 2.43 a 300 K e a 77 K.

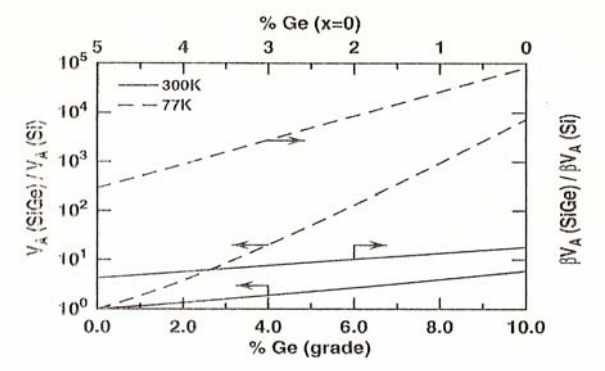

Figura 2.43

In una maniera simile a come è stato fatto per J_C, è possibile a questo punto fare due rilevanti approssimazioni. Primo, è possibile fare l'assunzione che $\Delta E_{g,Ge}(grade) >> kT$. Questa approssimazione può essere chiamata lo scenario del "forte gradiente di Ge". In questo caso la (2.32) si riduce a:

$$\left.\frac{V_{A,SiGe}}{V_{A,Si}}\right|_{V_{BE}} \cong \frac{e^{\Delta E_{g,Ge}(grade)/kT}}{\Delta E_{g,Ge}(grade)/kT} \tag{2.33}$$

Come discusso sopra, bisogna prestare attenzione nell'applicare questa approssimazione. Per controllare la sua validità per un profilo realistico, si assuma di avere un profilo di Ge triangolare dallo 0% al 15% in un SiGe HBT operante a 300 K. Prendendo un offset di banda di approssimativamente 75mV per il 10% di Ge, si trova che $\Delta E_{g,Ge}(grade)/kT = 4.3$ rispetto all'unità, un'approssimazione ragionevole ma non del tutto valida. Tuttavia, appena la temperatura diminuisce, la validità di questa approssimazione migliora rapidamente al decrescere di *kT*. Ad esempio, nel caso appena citato, il profilo di Ge triangolare dallo 0% al 15% produce $\Delta E_{g,Ge}(grade)/kT = 17$ a 77 K.

In aggiunta al profilo fortemente drogato, si può anche definire un approssimazione di "debole gradiente di Ge", che dovrebbe essere valida, per esempio, nel caso di profilo costante di Ge. In questo caso, $\Delta E_{g,Ge}(grade) << kT$. Espandendo gli esponenziali del gradiente di Ge nella (2.32) in serie di Taylor e semplificando, si vede che Θ = 1 e, dunque, non si ottiene nessun incremento di V_A in un Si BJT, pur essendo presente Ge nella regione di base.

2.6.6 Prodotto guadagno-tensione di Early

Alla luce della precedente discussione riguardante le difficoltà intrinseche nell'ottenere elevate V_A assieme ad alti valori di β, è possibile definire convenzionalmente una figura di merito per il progetto di circuiti analogici: il cosiddetto prodotto βV_A. In un Si BJT convenzionale, un confronto tra la (2.2) e la (2.27) mostra che tale prodotto è, al primo ordine, indipendente dal profilo di base e, dunque, non subisce nessun miglioramento dallo *scaling* della tecnologia, così come, invece, accadrebbe per la risposta in frequenza del transistore. Per un SiGe HBT, tuttavia, sia β sia V_A sono disaccoppiati dal profilo di base e possono essere indipendentemente regolati variando la forma del profilo di Ge. Combinando la (2.15) e la (2.32) si trova che il rapporto tra i βV_A di un SiGe HBT e un Si BJT possono essere scritti come:

$$\frac{\beta V_{A,SiGe}}{\beta V_{A,Si}} = \tilde{\gamma}\tilde{\eta} e^{\Delta E_{g,Ge}(0)/kT} e^{\Delta E_{g,Ge}(grade)/kT} \tag{2.34}$$

I risultati di tipici esperimenti per il rapporto dei βV_A sono mostrati in figura 2.44.

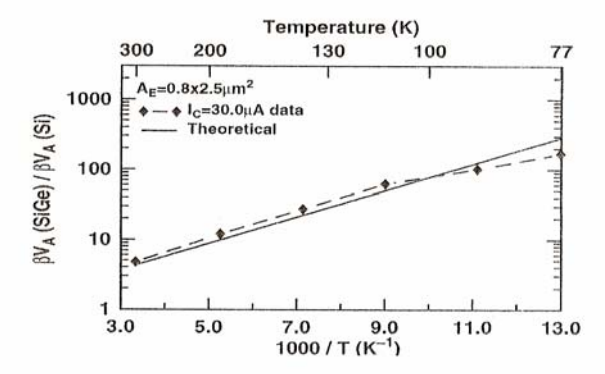

Figura 2.44

Si osservi che βV_A è una funzione termicamente attivata sia dell'offset di banda indotto dal Ge alla giunzione EB sia dal gradiente indotto dal Ge attraverso la

base neutra. Come si può vedere nella figura 2.44, βV_A in un SiGe HBT è significativamente migliorato rispetto ad un Si BJT, per qualsiasi profilo di Ge, benché il profilo triangolare resta la scelta migliore per l'ottimizzazione di V_A e βV_A . Essendo funzioni della temperatura, inoltre, V_A e βV_A sono fortemente incrementati con il raffreddamento, fornendo valori enormi ($\beta V_A > 10^4$) a 77 K per un profilo triangolare di Ge del 10%.

2.6.7 Altri profili nel SiGe

Per il profilo trapezoidale è possibile derivare le espressioni per V_A in presenza di un drogaggio di base costante. La tensione di Early dipende dal rapporto delle densità della corrente di collettore e dalla pendenza della densità della corrente di collettore rispetto a V_{CB} (equazione 2.26) che può essere espressa generalmente come:

$$\frac{\partial J_{C,SiGe}}{\partial V_{CB}} = \left\{ \Xi \left(\frac{\partial J_{C,SiGe}}{\partial W_b} \right) + J_{C,Si} \left(\frac{\partial \Xi}{\partial W_b} \right) \right\} \left(\frac{\partial W_b}{\partial V_{CB}} \right) \tag{2.35}$$

con $\xi = x_T / W_b < 1$, mentre Ξ è il rapporto del guadagno di corrente. Dalla (2.25) è possibile esprimere la variazione del rapporto dei guadagni di corrente da V_{CB} come:

$$\frac{1}{\Xi} \frac{\partial \Xi}{\partial W_b} = \left[\frac{e^{\Delta E_{g,Ge}(grade)/kT} - 1 - \Delta E_{g,Ge}(grade)/kT}{\xi\left(e^{\Delta E_{g,Ge}(grade)/kT} - 1\right) + (1-\xi)\Delta E_{g,Ge}(grade)/kT} \right] \cdot \left(\frac{\xi^2}{x_T} \right) \tag{2.36}$$

e infine si ottiene l'espressione del rapporto delle V_A per un generico profilo trapezoidale SiGe:

$$\left. \frac{V_{A,SiGe}}{V_{A,Si}} \right|_{V_{BE}} = 1 - \xi + \frac{\xi\left(e^{\Delta E_{g,Ge}(grade)/kT} - 1\right)}{\Delta E_{g,Ge}(grade)/kT} \tag{2.37}$$

Dall'equazione (2.37) si evince che, applicando i limiti per i profili triangolare e costante, è possibile ritrovare le espressioni standard per V_A (vedi equazione (2.32)). Si noti che un espressione simile può essere ottenuta per βV_A combinando le equazioni (2.37) e (2.25). Come per il caso precedente, si può fissare il contenuto di Ge a $x=0$ e, dunque, far variare il contenuto di Ge all'ascissa $x=W_b$, in modo che il contenuto di Ge integrato totale rimanga costante. In figura 2.45 sono mostrati i fattori di incremento di V_A e βV_A (per un contenuto di Ge pari al 2% a $x=0$). Come atteso, i risultati per il profilo trapezoidale giacciono tra quelli relativi al puro profilo costante e al puro profilo triangolare.

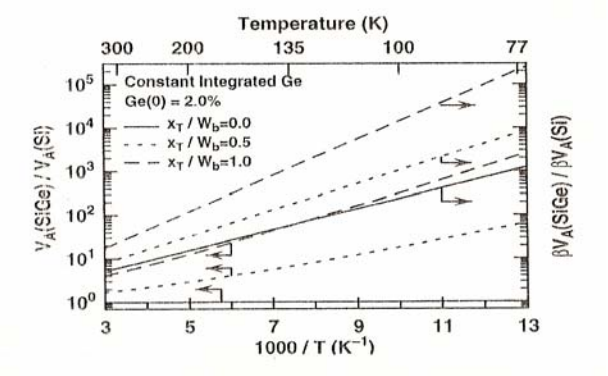

Figura 2.45

2.6.8 Ottimizzazione di V_A e βV_A

A partire dalle analisi fatte in precedenza, è possibile fare diverse osservazioni riguardanti gli effetti del Ge sulla tensione di Early e sul prodotto guadagno – tensione di Early nei SiGe HBT:

- ✓ A differenza di J_C, la sola presenza di un più grande contenuto di Ge al lato CB della base neutra rispetto al lato EB (gradiente di Ge finito) aumenterà V_A per una fissata V_{BE} rispetto al Si BJT.

- ✓ Questo incremento di V_A dipende esponenzialmente dal gradiente di Ge attraverso la base. Questa dipendenza osservata giocherà un ruolo nella comprensione del miglior approccio per l'ottimizzazione del profilo, generalmente preferendo profili con forti gradienti (cioè quelli triangolari).
- ✓ Alla luce di ciò, per due profili di Ge con stabilità costante, un profilo triangolare di Ge è migliore per un aumento della tensione di Early rispetto al profilo di Ge costante, lasciando tutto il resto uguale.
- ✓ L'aumento di V_A indotto dal Ge è termicamente attivato (esponenzialmente dipendente dal reciproco della temperatura) e, dunque, abbassare la temperatura significa ottenere una forte amplificazione dell'incremento.
- ✓ Dato che β e V_A hanno una dipendenza esattamente opposta dal gradiente di Ge e dall'offset di banda EB, il prodotto βV_A in un SiGe HBT sperimenta uno scenario vincente.
- ✓ Un ragionevole compromesso per il design del profilo di Ge che bilanci le necessità di ottimizzazione DC di β, V_A e βV_A dovrebbe essere il profilo trapezoidale, con un piccolo contenuto di Ge alla giunzione EB (3-4%) e un più grande contenuto di Ge alla giunzione CB (10-15%). Ovviamente, si sta considerando il caso di gradiente finito di Ge.

2.7 Modelli dei circuiti equivalenti

Dal punto di vista storico, il primo e più importante modello di circuito equivalente per un transistore bipolare è il modello di Ebers-Moll mostrato in

figura 2.46. Un'assunzione fondamentale in questo modello è che il comportamento complessivo del transistore possa essere visto come una sovrapposizione dei modi operativi diretto e inverso. Dunque, I_F rappresenta la corrente di emettitore totale per il modo operativo diretto e $\alpha_F I_F$ rappresenta la componente della corrente di elettroni I_F o corrente di collettore diretta. Il parametro α_F rappresenta il guadano di corrente a base comune in diretta. Similmente, I_R è la corrente di emettitore totale per il modo operativo inverso, e $\alpha_R I_R$ rappresenta la componente di corrente elettronica di I_R. Il parametro α_R è il guadagno di corrente inverso a base comune. Sia I_F sia I_R hanno una forma funzionale *I-V* esponenziale:

$$I_F = I_{F0}\left(e^{qV_{BE}/kT} - 1\right)$$

$$I_R = I_{R0}\left(e^{qV_{BC}/kT} - 1\right) \qquad (2.38)$$

Dove I_{F0} e I_{R0} rappresentano le correnti di saturazione delle correnti di emettitore diretta ed inversa.

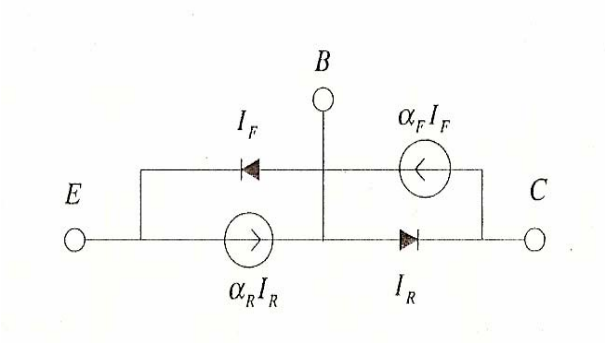

Figura 2.47

Un altro modello di circuito equivalente per il transistore bipolare, che meglio descrive il trasporto dei portatori, è mostrato in figura 2.48a, che è anche noto come la "versione di trasporto" del modello di Ebers-Moll. La corrente di collettore viene scelta, in luogo della corrente di emettitore, come corrente di

riferimento. Per un modo operativo diretto, la corrente di collettore è trasportata dall'emettitore verso il collettore, mentre la corrente di base è iniettata nell'emettitore:

$$I_{CF} = I_S\left(e^{qV_{BE}/kT} - 1\right)$$

$$I_{BF} = \frac{I_{CC}}{\beta_F} \tag{2.39}$$

dove β_F è il guadagno di corrente nella regione di lavoro diretta. Allo stesso modo, nella regione inversa, si ottiene:

$$I_{CR} = I_S\left(e^{qV_{BC}/kT} - 1\right)$$

$$I_{BF} = \frac{I_{CR}}{\beta_R} \tag{2.40}$$

dove β_R è il guadagno di corrente nella regione di lavoro inversa. Si noti che la corrente di saturazione I_S è identica per le correnti di collettore sia nel modo diretto che nel modo inverso. La figura 2.48a può essere ridisegnata come in figura 2.48b, che è più adatta per l'analisi dei circuiti ad emettitore comune.

Si può facilmente vedere che I_S è relazionata a I_{F0} e I_{R0} da:

$$I_S = \alpha_F I_{F0} = \alpha_R I_{R0} \tag{2.41}$$

e questa relazione è anche nota come la "proprietà di reciprocità del transistore bipolare". La reciprocità può essere facilmente compresa come il risultato del trasporto dei portatori minoritari. La corrente di portatori minoritari (elettroni) nella base è determinata dalle proprietà della regione di base, che è condivisa dai modi operativi diretto ed inverso del transistore. In generale, la regola di reciprocità vale solo approssimativamente per i SiGe HBT.

Il modello equivalente per ampi segnali mostrato in figura 2.48b può essere linearizzato per un dato punto operativo DC.

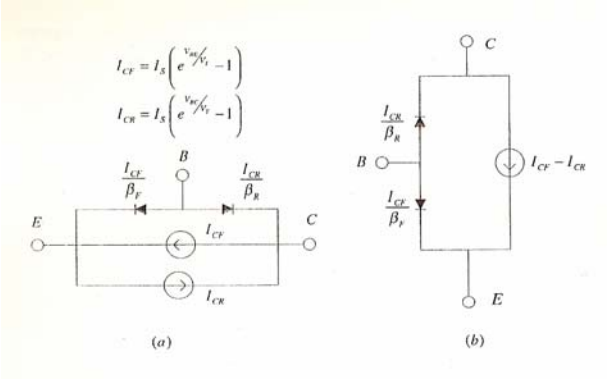

Figura 2.48

Il circuito equivalente è chiamato "circuito equivalente per piccoli segnali". Sotto le condizioni di modo operativo diretto ($V_{BE} > 0$ e $V_{CB} > 0$), questo modello si riduce al ben noto modello lineare per piccoli segnali ibrido a π mostrato in figura 2.49a. Il generatore della corrente di trasporto I_{CF} diventa il generatore della corrente di transconduttanza $g_m v_{be}$ con $g_m = qI_C/kT$. Il diodo EB diretto diventa la conduttanza $g_{be} = g_m/\beta$ e $r_\pi = 1/g_{be}$. Il diodo CB inversamente polarizzato diventa un circuito aperto e, ad elevate correnti, le cadute di tensione sulle resistenze parassite non possono essere più trascurate. I loro effetti possono essere inclusi aggiungendo resistenze appropriate. L'aumento di I_C all'aumentare di V_{CB} dovuto all'effetto Early può essere tenuto in conto aggiungendo r_0 in parallelo al generatore di corrente $g_m v_{b'e'}$. Il circuito equivalente finale è mostrato in figura 2.49b.

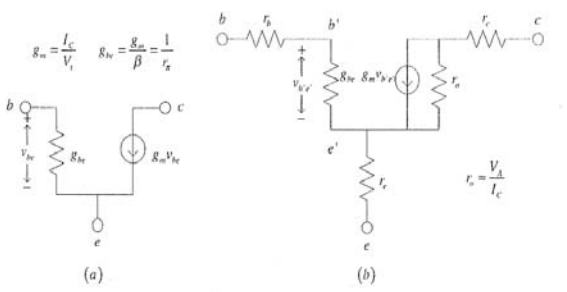

Figura 2.49

Per sfruttare l'elevato potenziale legato alla f_T offerto dall'impiego di una regione di base più piccola e da tempi di transito nell'emettitore più piccoli nei SiGe HBT, la densità della corrente di collettore J_C deve essere sufficientemente alta in modo che il tempo di carica, associato alle capacità parassite e di svuotamento, sia più piccolo rispetto alla somma dei tempi di transito nella base e nell'emettitore ($\tau_b + \tau_e$). Dunque, più piccolo è il tempo di transito, più elevata deve essere la J_C. Per sopprimere l'effetto Kirk ad elevate J_C, i SiGe HBT sono tipicamente progettati con collettori fortemente drogati. Un elevato drogaggio di collettore porta ad un elevato campo elettrico nella giunzione CB inversamente polarizzata (nel modo di funzionamento attivo del transistore) e, dunque, un elevato rate di ionizzazione da impatto. Per circuiti con SiGe HBT di pratica applicazione, operanti o ad elevate densità di corrente di collettore o ad elevate polarizzazioni collettore-base, la moltiplicazione a valanga è un effetto importante che deve essere accuratamente misurato e modellizzato. Nelle applicazioni digitali, il fattore di moltiplicazione a valanga (*M-1*) determina la tensione di breakdown così come la tensione d'inversione della corrente di base, che a sua volta determina il massimo valore utile di V_{CE} per un funzionamento stabile del dispositivo.

In polarizzazione inversa, il campo elettrico nella regione di carica spaziale della giunzione CB è grande. Gli elettroni iniettati dall'emettitore si muovono verso il collettore attraverso la regione di carica spaziale CB. Per ottenere un campo elettrico sufficientemente alto, gli elettroni possono acquisire energia sufficiente dal campo elettrico per creare una coppia elettrone-lacuna attraverso l'impatto con il reticolo cristallino (una semplice analisi mostra che l'energia di

soglia minima per la ionizzazione da impatto è *1.5 x* E_g). Questo processo di generazione dei portatori è chiamato "ionizzazione da impatto". Gli elettroni e le lacune generate per ionizzazione da impatto possono successivamente acquisire energia dal forte campo elettrico e creare altre coppie elettrone-lacuna per ulteriore ionizzazione da impatto. Questo processo moltiplicativo di ionizzazione da impatto è noto come "moltiplicazione a valanga". L'effetto netto è che la corrente degli elettroni che lasciano la regione di carica spaziale CB (la I_C osservata al collettore) è maggiore di quella che entra nella regione di carica spaziale CB (la I_C che sarebbe osservata senza la moltiplicazione a valanga). Il rapporto delle due correnti è noto come "fattore di moltiplicazione a valanga *M*":

$$M = \frac{I_{n,out}}{I_{n,in}} \tag{2.42}$$

Dove $I_{n,in}$ e $I_{n,out}$ sono le correnti elettroniche verso e dalla regione di carica spaziale CB. In pratica, viene usato il termine *M-1* in luogo di *M* perché descrive meglio l'efficienza dell'incremento risultante della corrente di collettore. L'incremento netto della corrente di elettroni dovuto alla ionizzazione da impatto è semplicemente $(M-1)I_{n,in}$. Poiché gli elettroni e le lacune sono sempre generate in coppia, una uguale quantità di corrente di lacune viene generata in questo processo e fluisce nella base di tipo *p*. La corrente di base netta osservata al terminale di base è dunque ridotta di $(M-1)I_{n,in}$:

$$I_B = I_{p,e} - (M-1)I_{n,in} \tag{2.43}$$

dove $I_{p,e}$ è la componente della corrente di base dovuta all'iniezione delle lacune nell'emettitore. Si è trascurata la componente I_B dovuta alla ricombinazione nella base neutra, che è molto piccola rispetto a quella dovuta alla iniezione delle lacune nell'emettitore, nei moderni SiGe HBT. Anche la corrente di *leakage* inversa della giunzione CB I_{CB0} è stata trascurata, in quanto è molto più piccola della $I_{p,e}$ nel normale modo di funzionamento. Si noti che I_{CB0} non può essere trascurato nell'analisi della tensione di *breakdown* in *open base* (BV_{CEO}). La relazione tra le varie componenti delle correnti di lacune e di elettroni in presenza della moltiplicazione a valanga in un SiGe HBT sono illustrate in figura 2.50 per il normale modo di funzionamento del dispositivo.

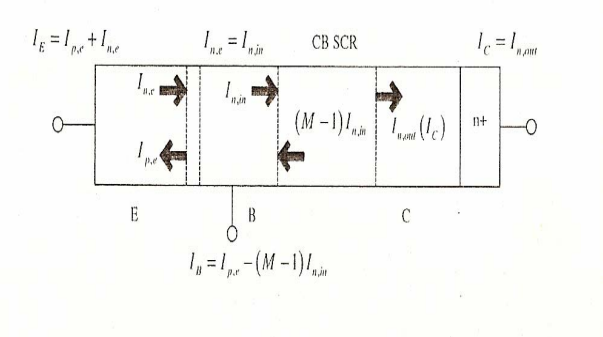

Figura 2.50

2.8 Tensioni di breakdown

Le tensioni di *breakdown* sono spesso caratterizzate attraverso l'applicazione di una polarizzazione inversa tra due dei tre terminali, lasciando il terzo terminale aperto. Per esempio, BV_{CBO} tipicamente è la tensione di *breakdown* collettore-base con l'emettitore aperto. Allo stesso modo, BV_{CEO} fa riferimento

alla tensione di *breakdown* collettore-emettitore con la base aperta. Entrambe le tensioni sono spesso indicate nelle specifiche della tecnologia IC.

Durante il funzionamento del circuito, la massima V_{CE} che può essere supportata dal SiGe HBT è molto più elevata della BV_{CEO} poiché la terminazione DC alla base non è mai, in realtà, elettricamente aperta. Questo vuol dire che la rete di polarizzazione DC presenta sempre un'impedenza finita tra la base e la massa. Come conseguenza, la tensione di *breakdown* collettore-emettitore è notevolmente più alta rispetto alla BV_{CEO}. Per i segnali RF, l'impedenza tra la base e la massa è anche più piccola, rendendo quindi la tensione di *breakdown* effettiva a sua volta più grande. Un'altra questione che deve essere considerata è la dipendenza dalla corrente del fattore di moltiplicazione a valanga *M-1*. Per elevati valori di J_C dove la f_T è massimizzata, *M-1* è molto più piccola rispetto al valore per bassi valori di J_C , dove BV_{CEO} e BV_{CBO} sono tipicamente misurati.

Un modo semplice di definire la BV_{CBO} e di adattare i valori misurati di *M-1* in funzione di V_{CB} attraverso l'equazione di Miller:

$$M = \frac{1}{1-\left(V_{CB}/BV_{CBO}\right)^m} \tag{2.44}$$

dove BV_{CBO} e m sono semplicemente definiti come parametri di *fitting*. BV_{CBO} può essere visto semplicemente come una rappresentazione di *M-1*. In questo caso, un valore più piccolo di BV_{CBO} corrisponde ad un più elevato valore di *M-1* nel dispositivo. La figura 2.51 mostra un grafico di *M-1* per $V_{BE} = 0.6V$. Si osservi che $BV_{CBO} = 6.1265V$ può essere ottenuto dai valori di *M-1* in funzione dei dati di V_{CB} (m *=5.252*). Assieme al valore del parametro m, il valore

estratto di BV_{CBO} consente un'agevole valutazione di *M-1,* poiché la definizione di BV_{CBO} è direttamente correlata a *M-1*.

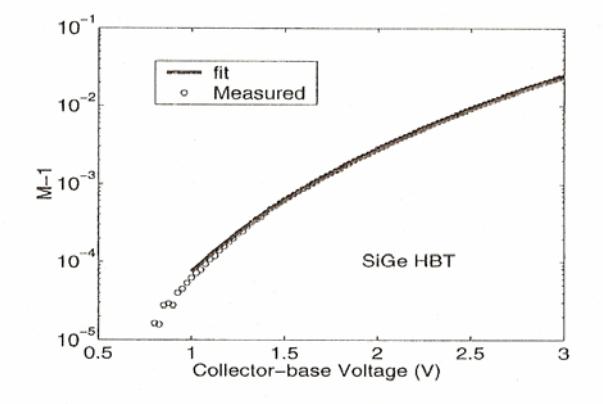

Figura 2.51

L'equazione (2.44), tuttavia, non è molto utilizzata per la modellizzazione della linearità RF poiché la moltiplicazione a valanga dipende strettamente dalla densità di corrente J_C.

In generale, BV_{CEO} è molto più piccola di BV_{CBO}. Questo può essere fisicamente compreso esaminando il processo di moltiplicazione a valanga nella condizione di base aperta (figura 2.52).

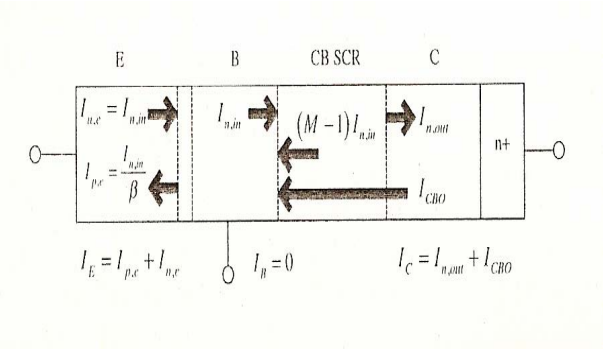

Figura 2.52

La corrente di *leakage* della giunzione CB, I_{CBO}, che è stata precedentemente trascurata in figura 2.50, ora deve essere considerata a causa del terminale di base. Adesso, I_{CBO} appare come una corrente di lacune nella base e può fluire solamente nell'emettitore poiché la base è aperta. Questa corrente è amplificata

da β durante questo processo, producendo un incremento nella corrente di elettroni che fluisce nella giunzione CB. Questa corrente elettronica, a sua volta, crea una corrente di lacune attraverso la moltiplicazione a valanga, che ancora fluisce nella giunzione EB come corrente di base. Viene così raggiunta una condizione di stato stazionario e la risultante corrente di diodo I_{CEO} è relazionata al processo di amplificazione della corrente e, dunque, a β. A causa del *feedback* positivo tramite l'amplificazione β, si prende $M-1=1/\beta$ per ottenere BV_{CEO} . Come confronto, per ottenere BV_{CBO} è necessario un valore infinito di *M-1*, pertanto non c'è nessun processo di *feedback*.

La corrente totale di collettore consiste di I_{CBO} e $I_{n,out}=M\cdot I_{n,in}$. La corrente totale di emettitore consiste di $I_{p,e}=I_{n,in}/\beta$ e $I_{n,in}$, dove I_B= *0A* a causa della condizione di base aperta. Perciò:

$$I_E = I_C$$

$$I_{n,in}\left(1+1/\beta\right)=MI_{n,in}+I_{CBO} \tag{2.45}$$

e $I_{n,in}$ è, dunque, dato da :

$$I_{n,in}=\frac{I_{CBO}}{1/\beta-\left(M-1\right)} \tag{2.46}$$

Per bassi valori di V_{CE}, *M-1 = 0*. Perciò, $I_{CEO}=\left(1+\beta\right)I_{CBO}$. Dunque, la corrente di *leakage* a base aperta, I_{CEO} , è β volte più grande della corrente di *leakage* nella giunzione CB, I_{CBO} . All'aumentare di V_{CE}, *M* aumenta e dunque il *breakdown* si verifica quando *M-1* si avvicina a $1/\beta$. Nell'analisi fatta si è trascurato l'effetto Early, benché esso introduca degli errori trascurabili. E' importante sottolineare che β presenta questa dipendenza unica dalla corrente

di *bias* nei SiGe HBT. Il valore di β per livelli molto bassi di iniezione (ad esempio, per $I_B = I_{CB0}$) può essere molto più piccolo del valore di β in condizione di basse iniezioni. In questo caso, quando il *breakdown* si verifica, la corrente aumenta rapidamente e il β aumenta rapidamente fino al suo valore corrispondente alla bassa iniezione. Allo stesso modo, per medie iniezioni, β può essere molto più piccolo che a basse iniezioni a causa degli effetti del gradiente di Ge. Per questo, β deve essere trattato come una funzione della polarizzazione per un accurata modellizzazione di BV_{CEO} nei SiGe HBT.

L'incremento del drogaggio di collettore nei SiGe HBT è dato dalla necessità di ottenere elevate densità di corrente, una condizione necessaria per sfruttare a pieno il potenziale legato ai valori di f_T offerto dai più bassi tempi di transito delle cariche nei SiGe HBT. Fondamentalmente, la costante di tempo relazionata alla carica della capacità di svuotamento è inversamente proporzionale a J_C e può essere ridotta solamente aumentando J_C. All'aumentare del drogaggio di collettore, tuttavia, aumenta il campo elettrico nella giunzione CB e, dunque, *M-1*, riducendo così le tensioni di *breakdown* del transistore. Tipicamente, un transistore con un elevato picco della f_T presenta un più elevato drogaggio di collettore e, dunque, più bassi valori di BV_{CBO} e BV_{CEO}. Il prodotto $f_T x\ BV_{CEO}$ è spesso denominato *limite di Johnson*. Un valore tipico di questo limite è 200 GHz·V. Recentemente, sono stati realizzati dei prototipi di SiGe HBT con $f_T x\ BV_{CEO} > 400$ GHz·V. Inoltre, molte tecnologie SiGe commerciali offrono degli HBT con differenti valori di BV_{CEO} mediante l'uso di un'impiantazione di collettore selettiva. Pertanto, i progettisti possono avere un'ulteriore grado di libertà nel progetto dei circuiti

elettronici e possono scegliere il dispositivo desiderato con valori opportuni di f_T e BV_{CEO} e, dunque, possono avvantaggiarsi di un più basso valore della capacità CB e un più basso valore di *M-1*, rispetto ai dispositivi con f_T più elevate.

2.9 Caratteristiche dinamiche

La presenza delle eterogiunzioni Si-SiGe nelle giunzioni EB e CB determina un sostanziale cambiamento nella fisica del dispositivo SiGe HBT rispetto al convenzionale Si BJT.

Le differenze fondamentali da un punto di vista AC possono essere meglio illustrate considerando il diagramma schematico a bande di energia. Per semplicità, si consideri un SiGe HBT ideale, con una base graduata e un drogaggio costante nelle regioni di emettitore, base e collettore. In tale struttura, il contenuto di Ge è linearmente graduato, dallo 0% alla giunzione metallurgica EB fino al massimo valore di Ge alla giunzione metallurgica CB, dopodichè rampa rapidamente allo 0% di Ge. I diagrammi a bande di energia risultanti per i due dispositivi, identicamente polarizzati nella regione attiva, sono mostrati in figura 2.53. Si osservi che il gradiente del Ge attraverso la base neutra induce un campo di *drift* intrinseco nella base neutra che impatterà positivamente sul trasporto dei portatori minoritari.

Per comprendere intuitivamente come le variazioni dei limiti di banda condizionano il comportamento AC del SiGe HBT, si consideri dapprima il comportamento dinamico del Si BJT.

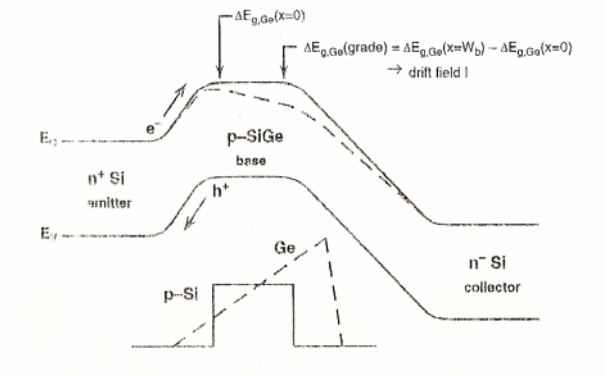

Figura 2.53

Gli elettroni iniettati dall'emettitore nella regione di base devono diffondere attraverso la base (per un drogaggio costante) e sono successivamente trascinati nel campo elettrico della giunzione CB, producendo una corrente di collettore utile che risulta essere funzione del tempo. Il tempo richiesto dagli elettroni per attraversare la base (tempo di transito in base) è significativo e tipicamente determina le prestazioni complessive AC del transistore (cioè, la f_T di picco). Allo stesso tempo, la polarizzazione diretta applicata alla giunzione EB produce dinamicamente una retro-iniezione di lacune dalla base nell'emettitore. Per una fissata corrente di collettore di polarizzazione, l'immagazzinamento dinamico delle lacune nell'emettitore (il tempo di ritardo di immagazzinamento di carica nell'emettitore) è reciprocamente legato al guadagno di corrente AC del transistore (β_{ac}).

Come si può vedere nella figura 2.53, l'introduzione del Ge nella regione di base ha una conseguenza importante poiché il campo di *drift* indotto dal gradiente di Ge attraverso la base neutra è allineato in una direzione (dal collettore all'emettitore) tale che accelererà gli elettroni minoritari iniettati attraverso la base. Si è, quindi, in grado di aggiungere una ampia componente di campo di *drift* al trasporto degli elettroni, accelerando efficacemente il

trasporto diffusivo dei portatori minoritari e, dunque, diminuendo il tempo di transito in base. Anche se gli offset di banda nei SiGe HBT sono tipicamente piccoli per gli standard della tecnologia III-V, il gradiente di Ge sulla breve distanza della base neutra può generare grossi campi elettrici. Per esempio, un profilo di Ge linearmente graduato con un contenuto modesto del picco di Ge del 10%, su una larghezza di base neutra di 50nm, produce un campo elettrico di circa 15kV/cm, sufficiente per accelerare gli elettroni, portandoli in prossimità della velocità di saturazione ($v_s \cong 1x10^7\ cm/\sec$). Poiché il tempo di transito in base tipicamente limita la risposta in frequenza del Si BJT, ci si aspetta che la risposta in frequenza sia significativamente migliorata introducendo questo campo di *drift* indotto dal Ge. Inoltre, è noto che l'offset di banda indotto dal Ge alla giunzione EB aumenterà esponenzialmente la densità di corrente di collettore (e, dunque, il β) del SiGe HBT rispetto al Si BJT. Poiché il tempo di ritardo dovuto all'immagazzinamento della carica nell'emettitore è inversamente legato a β, ci si aspetta che la risposta in frequenza del SiGe HBT benefici del vantaggio legato alla diminuzione del tempo di ritardo dovuto all'immagazzinamento della carica nell'emettitore.

Per basse iniezioni, la frequenza di taglio per guadagno unitario (f_T) in un transistore bipolare può essere scritta generalmente come:

$$f_T = \frac{1}{2\pi\tau_{ec}} = \frac{1}{2\pi}\left[\frac{kT}{qI_C}\left(C_{te}+C_{tc}\right)+\tau_b+\tau_e+\frac{W_{CB}}{2v_{sat}}+r_cC_{tc}\right]^{-1} \qquad (2.47)$$

dove $g_m = kT/qI_C$ è la transconduttanza intrinseca per basse iniezioni, C_{te} e C_{tc} sono le capacità di svuotamento delle giunzioni EB e CB, τ_b è il tempo di transito in base, τ_e il tempo di ritardo dovuto all'immagazzinamento della

carica nell'emettitore, W_{CB} è la larghezza della regione di carica spaziale CB, v_{sat} è la velocità di saturazione e r_c è la resistenza di collettore dinamica. Nella (2.47), τ_{ec} rappresenta il tempo totale di ritardo emettitore-collettore e determina il limite ultimo della velocità di commutazione del transistore bipolare. Dunque, per una data corrente di polarizzazione, le riduzioni di τ_b e τ_e dovuti alla presenza del SiGe si traducono direttamente in un aumento di f_T e $f_{\max}$ del transistore, per una data corrente di polarizzazione.

2.9.1 Effetti di modulazione di carica

L'azione di un transistore, sia esso un bipolare o un MOSFET, è fisicamente realizzata attraverso la modulazione in tensione delle cariche nel transistore che, a sua volta, porta ad una modulazione della corrente di uscita. La modulazione di tensione delle cariche risulta in una corrente capacitiva che aumenta con la frequenza. La larghezza di banda del transistore è, dunque, limitata da vari effetti di immagazzinamento della carica nella struttura estrinseca ed intrinseca del dispositivo. Un'analisi esatta degli effetti di immagazzinamento della carica richiede la risoluzione delle equazioni di trasporto dei semiconduttori nel dominio della frequenza. In pratica, gli effetti di immagazzinamento della carica spesso sono tenuti in conto assumendo che le distribuzioni di carica istantaneamente seguono le variazioni delle tensioni ai terminali in condizioni dinamiche (assunzione "quasi-statica").

Il primo effetto di modulazione di carica in un SiGe HBT è la modulazione delle cariche spaziali associate alle giunzioni EB e CB. Le variazioni di

tensione attraverso le giunzioni EB e CB portano a variazioni degli spessori dello strato di carica spaziale (svuotamento) e, dunque, della carica spaziale totale. Il comportamento capacitivo è simile a quello di un condensatore a facce piane parallele perché le variazioni nella carica si verificano alle facce opposte dello strato di carica spaziale (che è svuotato dei portatori in condizioni di polarizzazione inversa), ai confini delle transizioni della regione neutra. Le capacità risultanti sono definite "capacità di svuotamento EB e CB". In condizioni di elevate iniezioni, la modulazione delle cariche dentro lo strato di carica spaziale diventa significativa. La capacità risultante è definita come la capacità di transizione ed è importante per la giunzione EB poiché questa è direttamente polarizzata. In condizioni di basse iniezioni, la capacità CB è simile a quella di una giunzione *pn* inversamente polarizzata ed è funzione della tensione di polarizzazione CB. Tuttavia, ad elevate iniezioni, anche nella regione attiva la capacità CB risulta essere funzione della corrente di collettore a causa della compensazione di carica dovuta ai portatori mobili.

Il secondo effetto di modulazione della carica è dovuto ai portatori minoritari iniettati nelle regioni della base neutra e di emettitore. Per mantenere la neutralità di carica, una eguale quantità di portatori maggioritari in eccesso sono indotti dai portatori minoritari iniettati. Entrambi i portatori maggioritari e minoritari rispondono alle variazioni di tensione EB, producendo effettivamente una capacità EB. Questa capacità è storicamente definita come "capacità di diffusione", poiché è associata alla diffusione dei portatori minoritari in un transistore bipolare ideale con un drogaggio uniforme di base.

Ciò che è essenziale al fine di ottenere l'azione del transistore è la modulazione della corrente di uscita con una tensione di ingresso. La modulazione di carica

è solo uno strumento per modulare la corrente e deve essere minimizzata al fine di mantenere ideale l'azione del transistore alle alte frequenze. Per esempio, una grande capacità di diffusione EB causa una corrente d'ingresso ampia che aumenta con la frequenza, diminuendo così il guadagno di corrente alle frequenze più elevate. Al livello fondamentale, per una data modulazione della corrente di uscita, è auspicabile una quantità decrescente di modulazione di carica al fine di ottenere frequenze operative più alte. Una naturale figura di merito per l'efficienza del transistore è il rapporto tra la modulazione della corrente di uscita e quella della carica totale:

$$\tau_{ec} = \frac{\partial I_C}{\partial Q_n} \tag{2.48}$$

che ha le dimensioni del tempo ed è dunque chiamato "tempo di transito". Qui, Q_n rappresenta la carica elettronica integrale attraverso l'intero dispositivo e può essere divisa in varie componenti. La derivata parziale nella (2.48) indica che c'è modulazione sia della carica sia della corrente. Ancora, l'equazione (2.48) può essere riscritta usando la tensione d'ingresso come una variabile intermedia:

$$\tau_{ec} = \frac{\partial I_C / \partial V_{BE}}{\partial Q_n / \partial V_{BE}} = \frac{g_m}{C_i} \tag{2.49}$$

dove C_i è la capacità d'ingresso totale e g_m è la transconduttanza. C_i può essere suddivisa in due componenti $C_{be} = \partial Q_n / \partial V_{BE}$ e $C_{bc} = \partial Q_n / \partial V_{BC}$. I tempi di transito collegati alla modulazione di carica della base neutra e dell'emettitore neutro sono i tempi di transito in base e di transito nell'emettitore, rispettivamente. La modulazione della carica di base, richiesta per produrre una data modulazione di corrente in uscita, può essere diminuita

introducendo un campo di *drift* attraverso il gradiente di Ge, riducendo in tal modo il tempo di transito in base ed estendendo la funzionalità del transistore a frequenze molto più alte. Questa riduzione indotta dal gradiente di Ge nella modulazione di carica è la ragione fondamentale per cui i SiGe HBT presentano una migliore risposta in frequenza rispetto ai Si BJT. Il gradiente di Ge è semplicemente un modo conveniente attraverso il quale si riduce la modulazione della carica.

2.10 Fattori di prestazione RF

La figura 2.54 mostra un circuito equivalente ad alta frequenza per piccoli segnali per un transistore bipolare che si prenderà in considerazione allo scopo di analizzare le prestazioni RF del transistore. Per semplicità, si è trascurata la resistenza di emettitore, la resistenza di collettore e la resistenza di uscita dovuta all'effetto Early. Qui, la capacità EB C_{be} è la somma della capacità di diffusione EB $g_m\tau_f$ e della capacità di svuotamento EB C_{te}, mentre $g_m = qI_C/kT$ e $g_{be} = g_m/\beta$, relazioni che valgono per un transistore ideale.

2.10.1 Guadagno di corrente e frequenza di taglio

L'amplificazione di corrente ad alta frequenza del SiGe HBT è tipicamente misurata attraverso il guadagno di corrente per piccolo segnale con l'uscita corto-circuitata (cioè h_{21}). Si immagini di pilotare il terminale di base con una corrente di piccolo segnale $i_b = i_0 e^{j\omega t}$ e di cortocircuitare l'uscita (il collettore), come mostrato in figura 2.55. La tensione di nodo v_b, dunque, è:

$$v_b = \frac{1}{g_{be} + j\omega(C_{be} + C_{bc})} i_b \tag{2.50}$$

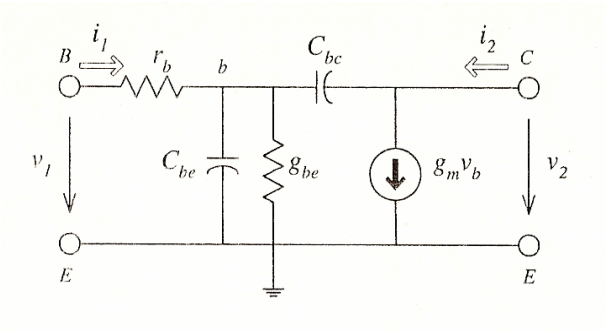

Figura 2.54

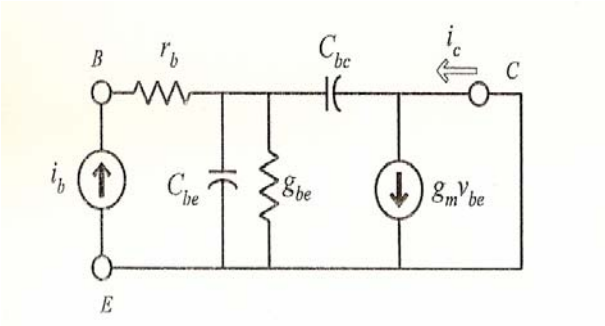

Figura 2.55

Il carico capacitivo effettivo per l'input dovuto alla capacità di Miller C_{be} è ancora C_{bc} a causa del guadagno di tensione nullo risultante dall'output corto-circuitato. Essendo la capacità di giunzione CB, inversamente polarizzata, molto più piccola della capacità di giunzione EB, direttamente polarizzata, è possibile trascurare il suo contributo alla corrente di uscita i_c:

$$i_c \approx g_m v_b = \frac{g_m}{g_{be} + j\omega(C_{be} + C_{bc})} i_b \tag{2.51}$$

Perciò, si ottiene:

$$h_{21} = \left.\frac{i_c}{i_b}\right|_{v_c=0} = \frac{g_m}{g_{be} + j\omega(C_{be} + C_{bc})} = \frac{\beta}{1 + j\omega(C_{be} + C_{bc})/g_{be}} \tag{2.52}$$

Si noti che h_{21} è costante a basse frequenze e, dunque, decresce ad alte frequenze. Ovviamente, la parte immaginaria aumenta con ω e domina alle alte frequenze. In queste condizioni, l'equazione precedente diventa:

$$h_{21} = \frac{g_m}{j\omega(C_{be} + C_{bc})} \tag{2.53}$$

che è equivalente a:

$$h_{21}xf = \frac{f_T}{j} \tag{2.54}$$

$$f_T = \frac{g_m}{2\pi(C_{be} + C_{bc})} \tag{2.55}$$

Il prodotto $|h_{21}xf|$ è costante sul *range* di frequenze in cui questa assunzione vale. Questa costante è denominata f_T, la frequenza di transizione o, più comunemente, la frequenza di taglio. In pratica, f_T è estratta estrapolando il valore misurato di $|h_{21}|$ in funzione dei dati di frequenza in un *range* dove la pendenza vale -20 dB/decade. La frequenza a cui il valore estrapolato di $|h_{21}|$ si riduce all'unità è definita come f_T (ovvero, la frequenza di taglio per guadagno unitario). In pratica, l'estrapolazione è necessaria qui perché si è usualmente non interessati a transistori operanti alla frequenza corrispondente al guadagno di corrente unitario, che può essere differente dalla f_T estrapolata, a seconda dei parassiti e di altri fattori. Invece, si è interessati al guadagno disponibile a frequenze molto più basse, dove il guadagno di corrente è molto più alto dell'unità. Nel *range* di frequenze dove $|h_{21}|$ presenta una pendenza di -20 dB/decade, $|h_{21}|$ può essere facilmente stimato come f_T / f.

I SiGe HBT allo stato dell'arte esibiscono valori di f_T sopra i 200GHz, che sono molto più alti delle frequenze operative dei *bulk* dei sistemi *wireless*

esistenti, che sono tipicamente sotto 10GHz. In questo caso, la (2.52) può essere riscritta come (usando la (2.55)):

$$h_{21} = \frac{\beta}{1 + j\, f / f_\beta} \tag{2.56}$$

$$f_\beta = \frac{f_T}{\beta}$$

Qui $|h_{21}|$ è uguale a β alle basse frequenze, si riduce di 3dB a $f = f_\beta = f_T / \beta$ e, successivamente, decade all'aumentare di f ad una pendenza teorica di -20 dB/decade. Dunque, per un SiGe HBT con f_T = 100GHz e β=100, la frequenza a 3dB è $f_\beta = 1\text{GHz}$. Per una frequenza operativa di 2GHz, che è vicina a f_β, occorre usare la (2.56) per una stima di $|h_{21}|$ in luogo di f_T / f. La figura 2.56 mostra un esempio di h_{21} misurato in funzione della frequenza da 2 a 110GHz per un SiGe HBT. La f_T estrapolata è 117GHz. Una deviazione apprezzabile dalla linea retta a 20dB/decade è osservata sotto 7GHz, la qual cosa indica che occorre impiegare la (2.56) per la stima di h_{21}. Allo stesso modo, si può osservare una deviazione dalla pendenza di 20dB/decade sopra i 40GHz.

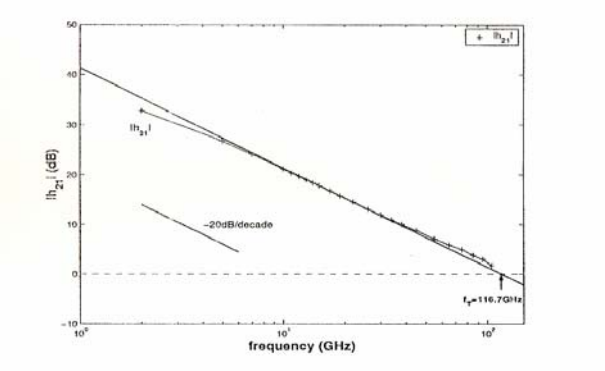

Figura 2.56

Il semplice modello presentato fin qui consente di fornire l'ampiezza di h_{21} ad alte frequenze con una discreta accuratezza, ma non consente di descrivere in

maniera altrettanto accurata la fase di h_{21}, a causa degli effetti quasi-statici. La fase di h_{21} è importante, per esempio, per il progetto di amplificatori retroazionati e di oscillatori. Un semplice ma efficace trucco è quello di sostituire la transconduttanza g_m in figura 2.54 con una transconduttanza complessa y_m:

$$y_m = \frac{g_m}{1 + j\, f / f_N} \tag{2.57}$$

Dove $f_N \approx 3/2\pi\tau_f$, con τ_f tempo di transito diretto (la somma dei tempi di transito in base, emettitore e collettore: $\tau_b + \tau_e + \tau_c$). Nel caso dei dati esposti sopra, la correzione di fase complessiva per h_{21} è approssimativamente 18° a $f=f_T$.

2.10.2 Densità di corrente in funzione della velocità

La natura fondamentale dei SiGe HBT richiede l'uso di elevate densità di corrente al fine di ottenere elevate velocità operative. La dipendenza della densità di corrente operativa da f_T è illustrata meglio esaminando l'inverso di f_T attraverso la (2.55):

$$\frac{1}{2\pi f_T} = \frac{C_{be} + C_{bc}}{g_m} \tag{2.58}$$

Poiché $C_{be} = g_m \tau_f + C_{te}$, $C_{bc} = C_{tc}$ e $g_m = qI_C / kT$, la (2.58) può essere riscritta come:

$$\frac{1}{2\pi f_T} = \tau_f + \frac{kT}{qI_C} C_t \tag{2.59}$$

dove $C_t = C_{te} + C_{tc}$. Poiché C_{te} e C_{tc} sono proporzionali all'area di emettitore, la (2.59) può essere riscritta in termini della densità di corrente di polarizzazione J_C come:

$$\frac{1}{2\pi f_T} = \tau_f + \frac{kT}{qJ_C} C_t' \tag{2.60}$$

dove $C_t' = C_t / A_E$ rappresenta le capacità totali di svuotamento EB e CB per unità di area di emettitore e $J_C = I_C / A_E$ è la densità di corrente di collettore operativa. Dunque, la frequenza di taglio f_T è fondamentalmente determinata dalla densità di corrente di polarizzazione J_C, indipendentemente dalla lunghezza di emettitore del transistore. Per valori molto bassi di J_C il secondo termine è molto grande e f_T è molto piccola, indipendentemente dal tempo di transito diretto τ_f. All'aumentare di J_C il secondo termine decresce e può diventare più piccolo di τ_f. Per elevati valori di J_C, invece, si verifica il *push-out* di base (effetto Kirk) e lo stesso τ_f aumenta con J_C portando al *roll-off* di f_T. Una tipica caratteristica f_T - J_C è mostrata in figura 2.57 per un SiGe HBT di prima generazione.

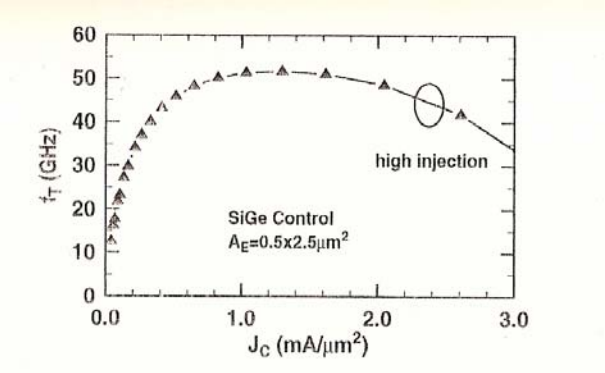

Figura 2.57

I valori di τ_f e $C_t^{'}$ possono essere facilmente estratti dal diagramma di $1/2\pi f_T$ in funzione di $1/J_C$, come mostrato in figura 2.58. In prossimità del picco di f_T, l'andamento risulta essere approssimativamente lineare, la qual cosa indica che $C_t^{'}$ è quasi costante per questo *range* di polarizzazione ad elevati valori di f_T. Dunque, $C_t^{'}$ può essere ottenuto dalla pendenza, mentre τ_f può essere determinato dall'intercetta sull'asse y per corrente infinita ($1/J_C = 0$).

Per incrementare la f_T in un SiGe HBT, il tempo di transito τ_f deve essere ridotto usando una combinazione di uno *scaling* verticale del profilo e di un gradiente di Ge attraverso la base. Allo stesso tempo, la densità di corrente operativa J_C deve essere aumentata in proporzione al fine di rendere il secondo termine nella (2.60) trascurabile rispetto al prima termine (τ_f).Questo è un criterio fondamentale per il progetto di SiGe HBT ad alta velocità. Più alto è il valore di picco di f_T, più elevato è il valore richiesto di J_C.

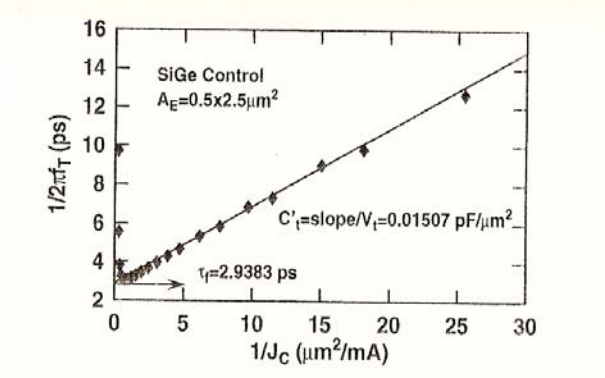

Figura 2.58

Per esempio, la minima densità di corrente richiesta è stata incrementata da $1\,\mathrm{mA}/\mu\mathrm{m}^2$ per i SiGe HBT della prima generazione con un picco di f_T di 50GHz a $10\,\mathrm{mA}/\mu\mathrm{m}^2$ per un picco di f_T di 200GHz per i SiGe HBT della

terza generazione. Naturalmente, densità di corrente più elevate portano a più severi effetti di *self-heating*, che devono essere accuratamente previsti e trattati nella modellizzazione e nel progetto dei circuiti. Inoltre, i problemi di elettromigrazione e altri vincoli di affidabilità legati ad elevate J_C hanno prodotto una necessità sempre più impellente di schemi di metallizzazione in rame.

Al fine di mantenere le prestazioni del transistore entro limiti accettabili anche per elevati valori di J_C , il drogaggio di collettore deve essere incrementato al fine di ritardare gli effetti di elevate iniezioni di carica. Questo aumento di drogaggio, ovviamente, riduce la tensione di *breakdown* del dispositivo. Pertanto, un *trade-off* tra tensione di *breakdown* e velocità del dispositivo è, dunque, inevitabile per tutti i transistori bipolari. Poiché il drogaggio di collettore nei SiGe HBT è realizzato tipicamente per impiantazione auto-allineata, è possibile ottenere dispositivi con diverse tensioni di *breakdown* (e dunque diverse f_T) nella stessa sequenza di fabbricazione, fornendo così ai progettisti una flessibilità maggiore.

Un'altra conseguenza della (2.60) è che la minima J_C richiesta per sfruttare in pieno il potenziale dei transistori, caratterizzati da una piccola τ_f , dipende da $C_t^{'}$. Dunque, $C_{te}^{'}$ e $C_{tc}^{'}$ devono essere minimizzate nel dispositivo. Inoltre, la riduzione di $C_{tc}^{'}$ è importante anche per ottenere un incremento del guadagno di potenza (ovvero, la massima frequenza di oscillazione $f_{\max}$).

2.10.3 Tempi di transito nella base e nell'emettitore

Per comprendere la risposta dinamica del SiGe HBT e il ruolo che il Ge gioca nella risposta in frequenza del transistore, occorre dapprima relazionare formalmente le variazioni del tempo di transito in base e nell'emettitore alle variabili fisiche del problema. E' anche istruttivo confrontare attentamente le differenze tra un SiGe HBT e un Si BJT. Nella presente analisi, si considerano il SiGe HBT e il Si BJT della stessa geometria e si assume che i profili di drogaggio di base, emettitore e collettore dei due dispositivi siano identici, a parte la presenza del Ge nel SiGe HBT. Per semplicità, si considera un profilo di Ge che è linearmente graduato dalla giunzione EB alla giunzione CB, come descritto nella figura 2.59.

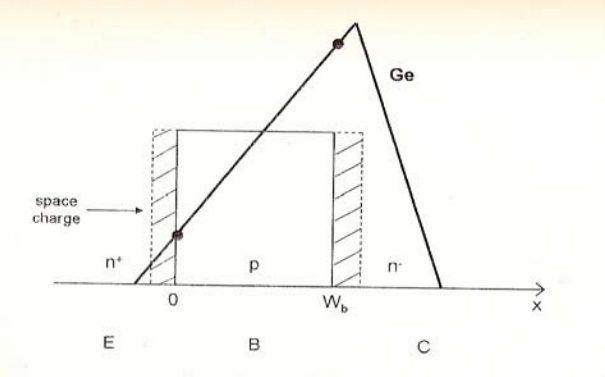

Figura 2.59

Le conseguenze teoriche delle variazioni di *bandgap* indotte dal Ge sulla τ_b possono essere derivate in forma chiusa per un profilo di drogaggio di base costante ($p_b(x) = N_{ab}^{-}(x) = N_{ab}^{-} = const$), considerando la relazione sul tempo di transito generalizzata di Moll-Ross, che vale per basse iniezioni in presenza di un drogaggio di base non uniforme e di un *bandgap* di base non uniforme per valori fissati di V_{BE} e T:

$$\tau_b = \int_0^{W_b} \frac{n_{ib}^2(x)}{p_b(x)} \left[\int_x^{W_b} \frac{p_b(y)dy}{D_{nb}(y) n_{ib}^2(y)} \right] dx \tag{2.61}$$

A questo punto, è possibile inserire la (2.4) nella (2.3) per ottenere la (2.5) e sostituire la (2.5) nella (2.61) per ottenere:

$$\tau_{b,SiGe} = \int_0^{W_b} \frac{n_{ib}^2(x)}{N_{ab}^-} \left\{ \int_\varsigma^{W_b} \frac{N_b^-}{D_{nb}} \left[\frac{1}{\gamma n_{io}^2} e^{-\Delta E_{gb}^{app}/kT} e^{-\Delta E_{g,Ge}(0)/kT} e^{-\Delta E_{g,Ge}(grade) y/kTW_b} dy \right] \right\} dx \tag{2.62}$$

Eseguendo l'integrazione si ottiene:

$$\tau_{b,SiGe} = \frac{W_b^2}{\widetilde{D}_{nb}} \frac{kT}{\Delta E_{g,Ge}(grade)} \cdot \left\{ 1 - \frac{kT}{\Delta E_{g,Ge}(grade)} \left[1 - e^{-\Delta E_{g,Ge}(grade)/kT} \right] \right\} \tag{2.63}$$

Come atteso, si vede che il tempo di transito in base in un SiGe HBT dipende reciprocamente dalla quantità di gradiente di *bandgap* indotto dal Ge attraverso la base neutra. E' istruttivo confrontare τ_b in un SiGe HBT con quello di un Si BJT. Nel caso di un Si BJT, è noto che:

$$\tau_{b,Si} = \frac{W_b^2}{2D_{nb}} \tag{2.64}$$

e dunque si può scrivere:

$$\frac{\tau_{b,SiGe}}{\tau_{b,Si}} = \frac{2}{\widetilde{\eta}} \frac{kT}{\Delta E_{g,Ge}(grade)} \left\{ 1 - \frac{kT}{\Delta E_{g,Ge}(grade)} \left[1 - e^{-\Delta E_{g,Ge}(grade)/kT} \right] \right\} \tag{2.65}$$

dove si è usato il rapporto delle diffusività elettroniche tra SiGe e Si come espresso nella (2.12). Nei limiti delle assunzioni fatte sopra, questo può essere considerato un risultato esatto. Come aspettato dalla discussione intuitiva sul diagramma a banda, si osservi che τ_b e, di conseguenza, f_T in un SiGe HBT dipendono reciprocamente dal fattore di gradiente di *bandgap* indotto dal Ge e, dunque, per un gradiente di Ge finito attraverso la base neutra, τ_b vale meno dell'unità; pertanto, ci si aspetta un aumento della f_T per un SiGe HBT rispetto

ad un Si BJT. La figura 2.60 conferma sperimentalmente questo risultato. Come si può vedere in figura 2.60, poiché f_T è aumentato su un ampio *range* di corrente di collettore utile, è possibile risparmiare potenza per una data frequenza operativa, rispetto al Si BJT. Questo *trade-off* potenza-prestazione può, nella pratica, essere anche più importante del semplice incremento della risposta in frequenza, in particolare per applicazioni portabili. In questo caso, se si decidesse di far lavorare il transistore ad una frequenza di 30GHz, si potrebbe ridurre la corrente di un fattore 5x. Si noti che, così come per l'espressione della densità di corrente di collettore (2.16), anche l'energia termica *kT* gioca un ruolo chiave nella (2.63), dove giace al numeratore e avrà, dunque, importanti implicazioni per la risposta in frequenza del SiGe HBT a temperature criogeniche.

In figura 2.61 sono mostrati calcoli teorici in funzione della forma del profilo di Ge a 300K e a 70K. Il contenuto integrale di Ge è mantenuto costante e il profilo di Ge varia da quello triangolare (10%) a quello costante (5%).

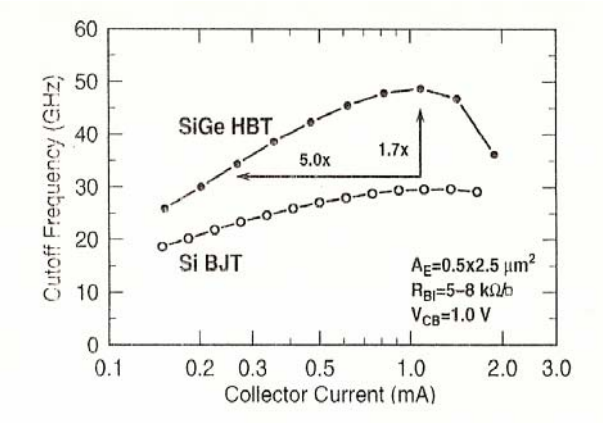

Figura 2.61

2.11 Rilevanti approssimazioni

In maniera simile a quanto fatto per la densità della corrente di collettore, è possibile fare due approssimazioni fisicamente rilevanti per ottenere ulteriori informazioni. Primo, si può assumere che $\Delta E_{g,Ge}(grade) >> kT$. Questa approssimazione può essere definita come lo scenario con un "forte gradiente di Ge". In questo caso la (2.63) si riduce a:

$$\tau_{b,SiGe} \cong \frac{W_b^2}{2\tilde{D}_{nb}} \frac{kT}{\Delta E_{g,Ge}(grade)} \tag{2.66}$$

Si noti, tuttavia, che occorre fare attenzione nell'applicazione di questa approssimazione. Per controllare la sua validità per un profilo realistico, si assuma di avere un profilo di Ge triangolare dallo 0% al 15% in un SiGe HBT operante a 300K. Considerando un offset di banda di approssimativamente 75meV per 10% di Ge, si trova che $\Delta E_{g,Ge}(grade)/kT = 4.3$, il quale valore non rispetta esattamente l'approssimazione precedente. Tuttavia, quando la temperatura diminuisce, il campo di validità di questa approssimazione aumenta rapidamente poiché il fattore *kT* decresce. Per esempio, nel caso esposto sopra, un profilo di Ge triangolare da 0% al 15% produce $\Delta E_{g,Ge}(grade)/kT = 17$, chiaramente >>1.

In aggiunta al caso di profilo fortemente graduato, è possibile, inoltre, definire un'approssimazione di "debole gradiente di Ge" che dovrebbe essere valida, per esempio, nel caso di un profilo costante (*box profile*) di Ge. In questo caso, $\Delta E_{g,Ge}(grade) << kT$. Espandendo in serie di Taylor l'esponenziale del fattore di gradiente di Ge e semplificando, si ottiene:

$$\tau_{b,SiGe} \cong \frac{W_b^2}{2\widetilde{D}_{nb}} \tag{2.67}$$

che è proprio il risultato ottenuto per un Si BJT (vedi la (2.64)). Come ci si aspetta, si vede che il gradiente di Ge è un fattore chiave per ottenere l'incremento desiderato nella risposta in frequenza rispetto ad un Si BJT poiché il campo di *drift*, che contribuisce al trasporto degli elettroni attraverso la base, è indotto proprio dal gradiente di Ge.

CAPITOLO 3 – Tecnologie elettroniche

Per il progettista di circuiti integrati risulta essere di fondamentale importanza la conoscenza dei dettagli del processo di fabbricazione, per due ragioni. Primo, la tecnologia dei circuiti integrati si è diffusa in maniera intensiva perché fornisce il vantaggio economico del processo planare per la realizzazione di circuiti complessi a basso costo attraverso la lavorazione per lotti (*batch processing*). Dunque, la conoscenza dei fattori che influenzano il costo della fabbricazione dei circuiti integrati è essenziale sia per la selezione di un approccio circuitale per risolvere un dato problema di progetto da parte dei designer, sia per la selezione di un particolare circuito per la "customizzazione" da parte dell'utente.

Secondo, la tecnologia dei circuiti integrati presenta un set completamente differente di vincoli di costi per il progettista rispetto alla tecnologia dei componenti discreti. La scelta di un dato approccio circuitale per realizzare una specifica funzione richiede una comprensione dei gradi di libertà disponibili con la tecnologia e della natura dei dispositivi che sono più facilmente realizzati su un chip integrato.

Attualmente, i circuiti integrati analogici sono progettati e realizzati in tecnologia bipolare, in tecnologia MOS e in tecnologie che combinano entrambi i tipi di dispositivi in un unico processo. La necessità di combinare funzioni digitali complesse sullo stesso circuito integrato con funzioni analogiche ha comportato un incremento nell'uso della tecnologia MOS, in particolare quelle funzioni come la conversione analogico-digitale richiesta per

le interfacce tra i segnali analogici e i sistemi digitali. Tuttavia, la tecnologia bipolare è attualmente usata e continuerà ad essere usata in un ampio *range* di applicazioni che richiedono un'elevata capacità di pilotaggio di correnti elevate e livelli più alti di prestazioni analogiche di precisione.

3.1 Fotolitografia

Quando un campione di silicio cristallino è posto in un ambiente ossidante si forma uno strato di diossido di silicio sulla sua superficie. Questo strato agisce da barriera per la diffusione delle impurità. In tal modo, le impurità separate dalla superficie del silicio con lo strato di ossido non diffondono nel silicio stesso durante i successivi passi di processo realizzati ad alte temperature. Quindi, si può formare una giunzione *pn* in una posizione selezionata sul campione prima coprendo il campione medesimo con uno strato di ossido (*step* di ossidazione), successivamente rimuovendo l'ossido nella regione selezionata e, dunque, realizzando uno *step* di pre-deposizione e di diffusione. La rimozione selettiva dell'ossido nelle aree desiderate viene realizzata attraverso la fotolitografia. Questo processo è illustrato nell'esempio concettuale di figura 3.1. Ancora una volta si assume che il materiale di partenza sia un campione di silicio di tipo *n*. Prima si realizza uno *step* di ossidazione in cui si fa crescere termicamente uno strato di diossido di silicio (SiO_2) sulla superficie, il cui spessore usualmente varia tra 0.2μm e 1μm. Il *wafer* che segue questo *step* è mostrato in fig. 3.1a. Successivamente, il campione è rivestito con uno strato sottile di materiale fotosensibile chiamato *fotoresist*. Quando questo materiale viene esposto ad una particolare lunghezza d'onda, subisce una variazione

chimica e, nel caso di *fotoresist* "positivo", diventa solubile in determinate soluzioni chimiche in cui il *fotoresist* non esposto è insolubile. Questo passo è illustrato in fig. 3.1b. Per definire le aree di diffusione desiderate sul campione di silicio, viene collocata una fotomaschera sulla superficie del campione stesso; questa fotomaschera è opaca dappertutto tranne che in alcune aree in corrispondenza delle quali avviene la diffusione. Una radiazione di una determinata lunghezza d'onda è diretta sul campione, come mostrato in figura 3.1c e cade sul *fotoresist* solo in corrispondenza delle aree non opache della maschera. Queste aree del *fotoresist* sono, dunque, chimicamente dissolte nello *step* di sviluppo, come mostrato in figura 3.1d. Le aree del *fotoresist* sono inaccessibili all'operatore.

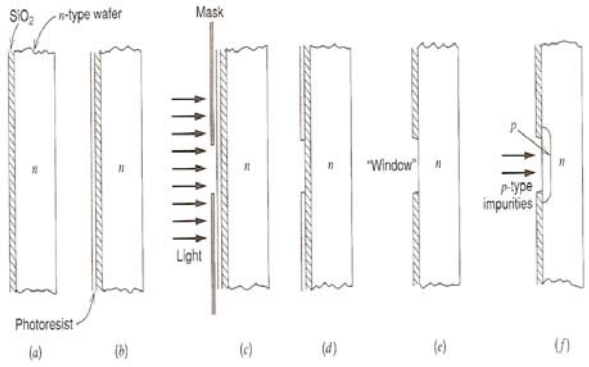

Figura 3.1

Poiché l'obiettivo è la formazione di una regione priva di SiO_2, lo *step* successivo è l'*etching* dell'ossido. Questo *step* può essere realizzato immergendo il campione in una soluzione di *etching* (acido idrofluoridico) oppure esponendolo ad un plasma prodotto elettricamente in un *plasma etcher*. In ogni caso, il risultato è che nelle regioni in cui il *fotoresist* è stato rimosso, l'ossido viene portato via, lasciando scoperta la superficie di silicio.

Il *fotoresist* restante è successivamente rimosso con una operazione chimica, lasciando il campione con delle finestre nell'ossido in corrispondenza delle posizioni desiderate, come mostrato in figura 3.1e. Il campione, adesso, è sottoposto ad uno *step* di pre-deposizione e di diffusione, il che risulta nella formazione di regioni di tipo *p* dove l'ossido è stato rimosso, come mostrato in figura 3.1f. In alcuni casi, l'impurità da aggiungere localmente alla superficie di silicio è depositata attraverso l'impiantazione ionica. Questo metodo di inserzione può spesso essere realizzato attraverso il diossido di silicio in modo che lo *step* di *etching* dell'ossido possa essere evitato.

La dimensione minima della regione diffusa che può essere ordinariamente creata con questa tecnica nella produzione dei dispositivi si è ridotta nel tempo e, ad oggi, è circa 0.2μm x 0.2μm. Il numero di tali regioni che possono essere fabbricate simultaneamente può essere calcolato notando che il campione di silicio usato nella produzione dei circuiti integrati è una fetta circolare, tipicamente da 4 pollici a 12 pollici in diametro e 250μm di spessore. Dunque, il numero di giunzioni *pn* elettricamente indipendenti di dimensioni 0.2μm x 0.2μm che possono essere formate su tale *wafer* è dell'ordine di 10^{11}. Nei moderni circuiti integrati viene utilizzato un certo numero di *step* di diffusione e di mascheratura al fine di formare strutture più complesse come i transistori, ma i punti chiave sono che la fotolitografia è in grado di definire un ampio numero di dispositivi sulla superficie del campione e che tutti questi dispositivi sono fabbricati in lotti allo stesso tempo. Dunque, il costo degli *step* di fotomascheratura e diffusione applicati al *wafer* durante il processo è diviso tra i dispositivi o circuiti sul *wafer* medesimo. Questa capacità di realizzare

migliaia di dispositivi allo stesso tempo è la chiave per il vantaggio economico della tecnologia dei circuiti integrati.

3.2 Crescita epitassiale

I primi transistori planari e i primi circuiti integrati usavano solo gli *step* di fotomascheratura e diffusione nel processo di fabbricazione. Tuttavia, i circuiti integrati diffusi presentavano forti limitazioni rispetto ai circuiti a componenti discreti. In un transistore bipolare a tripla diffusione, come illustrato in fig. 3.2, la regione di collettore è formata con una diffusione di tipo *n* nel *wafer* di tipo *p*. Gli svantaggi di questa struttura sono che la resistenza di collettore serie è elevata e la tensione di *breakdown* collettore-emettitore è bassa. La prima limitazione si verifica perché la concentrazione di impurità nella porzione della diffusione di collettore sotto la giunzione collettore-base è bassa, dando alla regione un'alta resistività. La seconda si verifica perché la concentrazione di impurità in prossimità della superficie del collettore è relativamente alta, risultando in una bassa tensione di *breakdown* tra le diffusioni di collettore e base alla superficie. Per superare questi problemi, la concentrazione di impurità dovrebbe essere bassa alla giunzione collettore-base al fine di ottenere elevate tensioni di *breakdown* ed elevata sotto la giunzione per ottenere basse resistenze di collettore. Tale profilo di concentrazione non può essere realizzato utilizzando solamente le diffusioni e per questo è stata sviluppata la tecnica di crescita epitassiale.

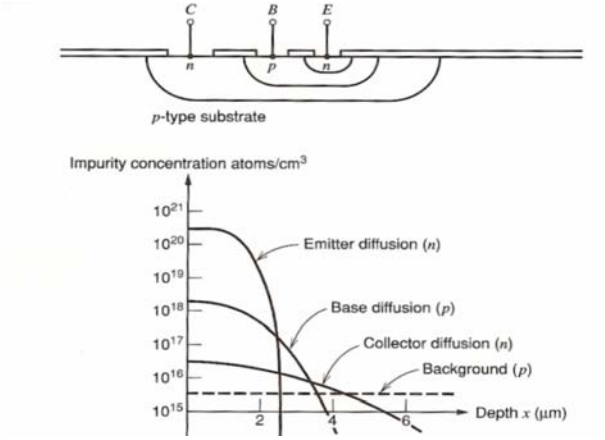

Figura 3.2

La crescita epitassiale (*epi*) consiste nella formazione di uno strato di silicio monocristallino sulla superficie del campione di silicio in modo che la struttura del cristallo di silicio sia continua attraverso l'interfaccia. La concentrazione d'impurità nello strato epitassiale può essere controllata indipendentemente e può essere più grande o più piccola che nel materiale di substrato. Inoltre, lo strato epitassiale è spesso di tipo opposto di impurità rispetto al substrato su cui è cresciuto. Lo spessore degli strati epitassiali usati nella fabbricazione dei circuiti integrati varia da 1μm a 20μm e la crescita degli strati è ottenuta collocando il *wafer* in un'atmosfera ambiente contenente tetracloruro di silicio ($SiCl_4$) o silano (SiH_4) ad elevate temperature. La reazione chimica che ne consegue comporta che il silicio elementare si depositi sulla superficie del *wafer* e lo strato risultante di silicio presenti una struttura cristallina con pochi difetti se le condizioni di crescita sono attentamente controllate. Tale strato è adatto come materiale di partenza per la fabbricazione di transistori bipolari. L'epitassia è inoltre anche utilizzata nella maggior parte delle tecnologie BiCMOS.

3.3 Impiantazione ionica

L'impiantazione ionica è una tecnica per l'inserzione diretta degli atomi di impurità nel *wafer* di silicio. Il *wafer* è collocato in una camera a vuoto e gli ioni della specie di impurità desiderata sono diretti verso il campione ad elevata velocità. Questi ioni penetrano la superficie del *wafer* di silicio fino ad una profondità media che varia tra meno di 0.1μm a circa 0.6μm, a seconda della velocità con cui colpiscono il campione. Il *wafer* è poi mantenuto a temperatura moderata per un periodo di tempo (per esempio, 800°C per 10 minuti) al fine di consentire agli ioni di diventare mobili e di inserirsi all'interno del reticolo cristallino. Questa fase viene chiamata *anneal step* ed è essenziale per consentire la riparazione di ogni danno al cristallo causato dall'impiantazione. I vantaggi principali dell'impiantazione ionica rispetto alla diffusione convenzionale sono:

1. le piccole quantità di impurità possono essere depositate con riproducibilità;
2. la quantità di impurità depositata per unità di area può essere controllata con precisione.

Inoltre, la deposizione può essere fatta con un elevato livello di uniformità attraverso il *wafer*. Un'altra utile proprietà degli strati impiantati è che il picco del profilo di concentrazione di impurità può essere creato in modo che si presenti sotto la superficie del silicio, a differenza degli strati diffusi. Questo consente la fabbricazione di strutture bipolari impiantate con proprietà che sono significativamente migliori di quelle dei dispositivi diffusi.

3.4 Ossidazione locale

Nelle tecnologie MOS e bipolari nasce spesso la necessità di realizzare regioni della superficie di silicio che sono coperte con uno strato di diossido di silicio relativamente sottile, accanto ad aree coperte da strati di ossido relativamente spesso. Tipicamente, le prime regioni costituiscono le aree dei dispositivi attivi, mentre le altre costituiscono le regioni che isolano elettricamente i dispositivi tra di loro. Una seconda richiesta è che la transizione dalle regioni spesse a quelle sottili debba essere realizzata senza introdurre uno *step* verticale grosso nella geometria della superficie del silicio, in modo che sia la metallizzazione sia gli altri *pattern* che vengono depositati successivamente possano giacere su una superficie relativamente planare. L'ossidazione locale è usata per raggiungere questo risultato. Il processo di ossidazione locale comincia con un campione che già presenta un ossido sottile cresciuto su di esso, come mostrato in figura 3.3a. Uno strato di nitruro di silicio (*SiN*) viene prima depositato sul campione e successivamente rimosso con uno *step* di mascheratura da tutte le aree dove l'ossido spesso deve essere cresciuto, come mostrato in figura 3.3b. Il nitruro di silicio si comporta come una barriera per gli atomi di ossigeno che potrebbero, altrimenti, raggiungere l'interfaccia Si-SiO_2 e determinare un'ulteriore ossidazione. Dunque, quando viene realizzato un ulteriore *step* di ossidazione di lunga durata e ad alta temperatura, si forma l'ossido spesso nelle regioni in cui non c'è il nitruro, mentre nessuna ossidazione si ottiene sotto il nitruro. La geometria risultante dopo la rimozione del nitruro è mostrata in figura 3.3c. Si noti che la parte superiore della superficie del diossido di silicio presenta una transizione "dolce" dalle aree spesse a quelle sottili e che l'altezza di questa transizione è inferiore alla differenza di spessore dell'ossido perché

l'ossidazione nelle regioni di ossido spesso tende a "consumare" parte del silicio sottostante.

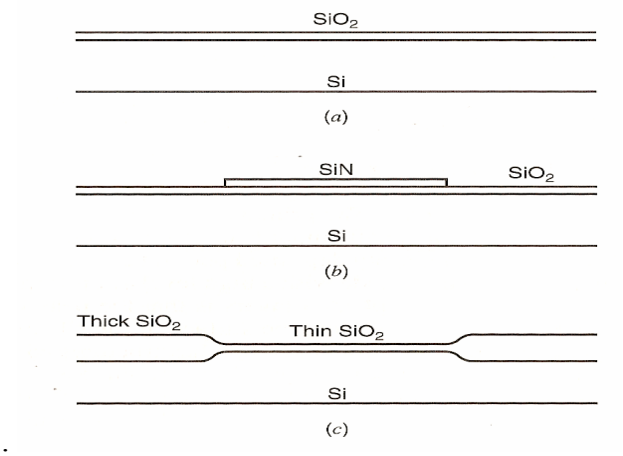

Figure 3.3

3.5 Deposizione del polisilicio

Molte tecnologie di processo utilizzano strati di silicio policristallino depositati durante la fabbricazione. Dopo la deposizione dello strato di silicio policristallino sul *wafer*, le caratteristiche volute sono definite usando uno *step* di mascheratura e possono servire per realizzare le *gate* di silicio dei transistori MOS, gli emettitori dei transistori bipolari, i piatti dei condensatori, resistori e strati di interconnessione. La *sheet resistance* di tali strati può essere controllata attraverso il controllo delle impurità aggiunte e varia tra 20Ω/□ fino a valori molto elevati. Il processo impiegato per depositare lo strato è simile a quello dell'epitassia. Tuttavia, poiché la deposizione è usualmente sopra uno strato di diossido di silicio, il *layer* non si forma come un'estensione a singolo cristallo del silicio sottostante ma si crea sottoforma di film granulare (polisilicio).

3.6 Fabbricazione di transistori bipolari ad alte tensioni

Le tecniche di fabbricazione dei circuiti integrati sono sensibilmente cambiate a partire dall'invenzione del processo planare. Questo cambiamento è stato determinato dagli sviluppi nella fotolitografia, nelle tecniche di processamento e dalla necessità di ridurre le alimentazioni in molti sistemi. Gli sviluppi nella fotolitografia hanno ridotto la minima *feature size* dalle decine di micron a livelli submicrometrici. Il controllo preciso permesso dalla impiantazione ionica ha reso questa tecnica lo strumento più usato per la pre-deposizione degli atomi di impurità. Infine, molti circuiti oggi operano con alimentazioni di 3-5V invece dei 15V usati in passato per ottenere elevati *range* dinamici in circuiti integrati come gli amplificatori operazionali. La riduzione delle tensioni operative consente di ridurre gli spazi tra i dispositivi nei circuiti integrati. Questo permette inoltre di ottenere strutture meno profonde con maggiori capacità in frequenza. Questi effetti derivano dal fatto che lo spessore degli strati di svuotamento della giunzione si riduce riducendo le tensioni. Dunque, i processi per ottenere circuiti integrati a frequenze operative più alte sono progettati per operare con alimentazioni al massimo di 5V e non sono generalmente impiegabili per alimentazioni maggiori. Infatti, esiste un *trade-off* fondamentale tra maggiori capacità in frequenza di un processo e la sua tensione di *breakdown*.

La realizzazione di un circuito integrato bipolare a giunzione isolata prevede una sequenza di passi di mascheratura e di diffusione. Il materiale di partenza è un *wafer* di silicio di tipo *p*, usualmente con uno spessore di 250μm e con una concentrazione di impurità di circa 10^{16} atomi/cm^3. Per la realizzazione di un transistore bipolare *npn*, il primo passo, illustrato in figura 3.4, forma uno

strato di tipo *n* poco resistivo che eventualmente può diventare un percorso a bassa resistenza per la corrente di collettore del transistore. Questo *step* è chiamato *buried-layer diffusion* e lo strato stesso è chiamato *buried layer*. La *sheet resistance* dello strato è nel *range* 20÷50Ω/□ e l'impurità usata è solitamente arsenico o antimonio perché queste impurità diffondono lentamente e, dunque, non si ridistribuiscono durante i successivi *step* di processo.

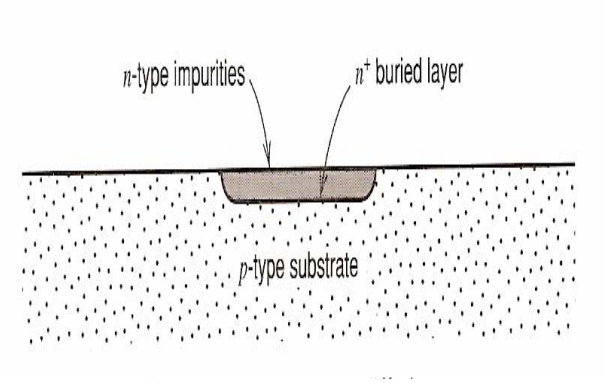

Figura 3.4

Dopo il passo del *buried layer*, il *wafer* è privato dell'ossido e su di esso viene cresciuto uno strato epitassiale, come mostrato in figura 3.5. Lo spessore dello strato e la sua concentrazione di impurità di tipo *n* determinano la tensione di *breakdown* collettore-base dei transistori nel circuito poiché questo materiale forma la regione di collettore del transistore medesimo.

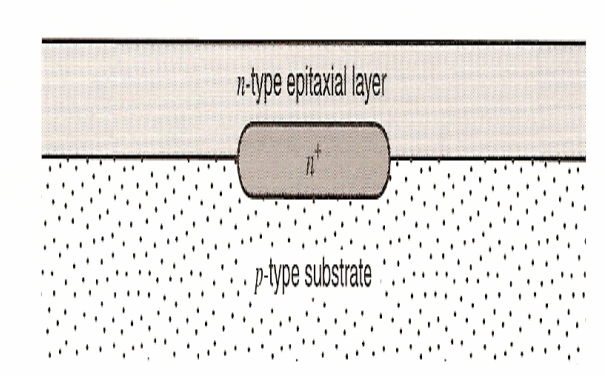

Figura 3.5

Dopo la crescita epitassiale, uno strato di ossido viene fatto crescere sulla superficie dello strato epitassiale. Successivamente, vengono realizzati uno *step* di mascheratura e una diffusione e pre-deposizione di boro (tipo *p*), dando vita alla struttura mostrata in figura 3.6. La funzione di questo passo di diffusione è di isolare i collettori dei transistori tra di loro con giunzioni *pn* polarizzate inversamente ed è chiamato *diffusione di isolamento*.

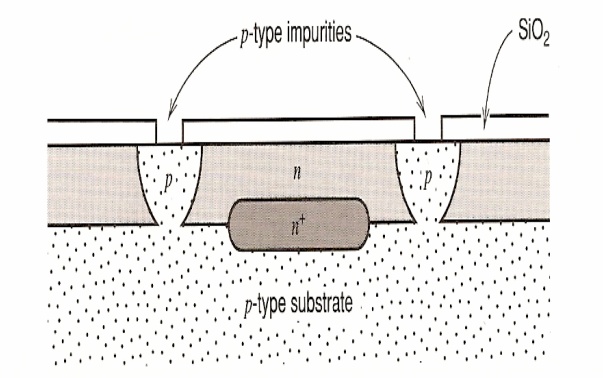

Figura 3.6

I passi successivi sono la mascheratura, la pre-deposizione e la diffusione della base, come mostrato in figura 3.7. L'ultimo *step* è usualmente una diffusione di boro e lo strato risultante presenta una *sheet resistance* di 100÷300Ω/□ e una profondità di 1÷3μm alla fine del processo. Questa diffusione forma non solo le basi dei transistori, ma anche molti dei resistori nel circuito, pertanto il controllo della *sheet resistance* diventa molto importante.

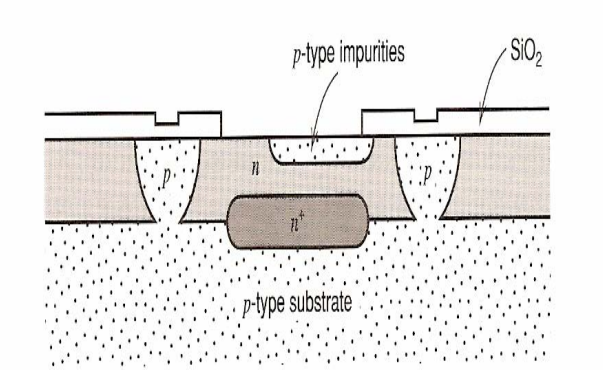

Figura 3.7

Successivamente, vengono formati gli emettitori dei transistori attraverso uno *step* di mascheratura, una pre-deposizione di tipo *n* e una diffusione, come mostrato in figura 3.8.

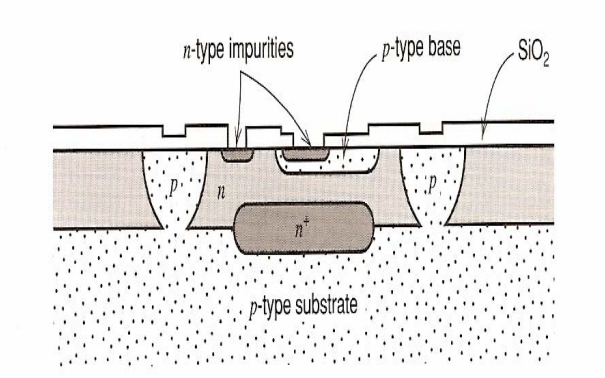

Figura 3.8

La *sheet resistance* è compresa tra 2Ω/□ e 10Ω/□ e la profondità è 0.5÷2.5μm dopo la diffusione. Questo *step* di diffusione è anche usato per formare una regione a bassa resistenza, la quale serve come contatto della regione di collettore. Questo è necessario perché il contatto ohmico è difficile da realizzare direttamente tra la metallizzazione in alluminio e il materiale epitassiale ad alta resistività. Il successivo *step* di mascheratura, la maschera di contatto, è usato per aprire delle buche nell'ossido sopra l'emettitore, la base e il collettore dei transistori in modo da realizzare i contatti elettrici con queste regioni. Le finestre di contatto sono anche aperte per i componenti passivi sul chip. L'intero *wafer* è, dunque, coperto con un sottile strato di alluminio (circa 1μm) che collegherà elettricamente gli elementi del circuito. L'effettivo *pattern* di interconnessione è definito dall'ultimo *step* di mascheratura, in cui l'alluminio è portato via nelle aree dove il *fotoresist* è rimosso. La struttura finale è mostrata in figura 3.9.

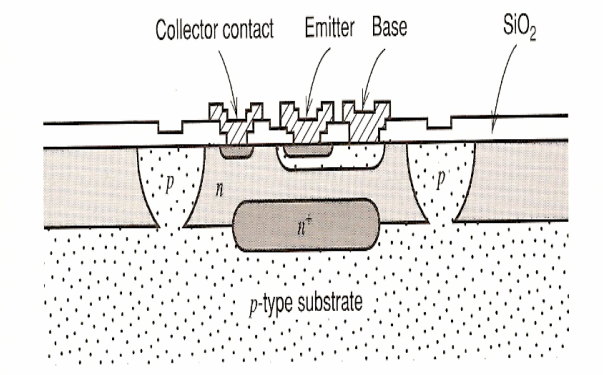

Figura 3.9

Una fotografia al microscopio di un'effettiva struttura dello stesso tipo è mostrata in figura 3.10.

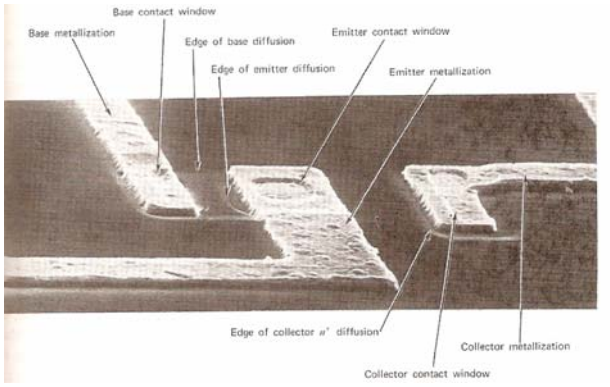

Figura 3.10

L'effetto "terrapieno" sulla superficie del dispositivo risulta dal fatto che viene fatto crescere dell'ossido aggiuntivo durante ogni ciclo di diffusione, in modo che esso sia più spesso sulla regione epitassiale, dove non c'è rimozione di ossido, sia meno spesso sopra le regioni di base e di isolamento che sono entrambe aperte durante lo *step* di mascheratura della base e sia più sottile sopra la diffusione di emettitore. Un tipico profilo di diffusione è mostrato in figura 3.11.

Questa sequenza permette la fabbricazione simultanea di un ampio numero (spesso migliaia) di circuiti complessi su un singolo *wafer*. Il *wafer* è poi collocato in un *tester* automatico, che controlla le caratteristiche elettriche di

ogni circuito sul *wafer* e mette un segno sui circuiti che non soddisfano le specifiche. Il *wafer* viene, quindi, spezzato e suddiviso nei vari circuiti individuali. I chip di silicio risultanti sono chiamati *dice* e il chip è il *die*. Ogni *die* buono è dunque montato in un *package*, pronto per il *testing* finale.

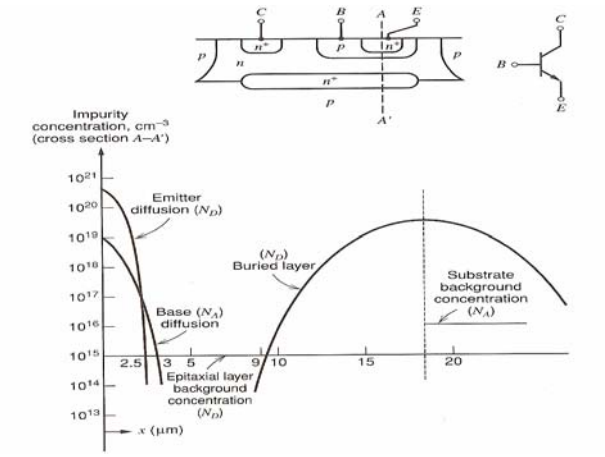

Figura 3.11

3.7 Processi avanzati di fabbricazione di bipolari

Un elevato numero di circuiti integrati analogici bipolari attualmente realizzati usa la tecnologia di base descritta nella precedente sezione. La sequenza di fabbricazione è relativamente semplice e a basso costo. Tuttavia, molte delle applicazioni di importanza commerciale richiedono una più elevata capacità in termini di risposta in frequenza, il che si traduce direttamente nella necessità di più elevate risposte in frequenza dei transistori. La richiesta di velocità più elevate impone una struttura del dispositivo con una larghezza di base più piccola per ridurre il tempo di transito in base e dimensioni complessive più piccole per ridurre le capacità parassite. Le dimensioni più piccole del dispositivo richiedono che la larghezza degli strati di svuotamento della giunzione nella struttura siano ridotti in proporzione, che a sua volta richiede

l'uso di tensioni operative del circuito più basse e concentrazioni di impurità più elevate nella struttura del dispositivo. Per soddisfare queste necessità, è stata sviluppata una classe di tecnologie di fabbricazione dei bipolari che, confrontate con quella del processo per alte tensioni descritto precedentemente, usano strati epitassiali drogati più pesantemente e più sottili, regioni ossidate in modo selettivo per l'isolamento al posto delle giunzioni diffuse e uno strato di polisilicio come sorgente di droganti per l'emettitore.

Il punto di partenza del processo è simile a quello per il processo convenzionale, con uno *step* di mascheratura e di impianto che risulta nella formazione di un *buried layer* fortemente drogato di tipo n^+ in un substrato di tipo *p*. Dopo questo, viene fatto crescere uno strato epitassiale sottile di tipo *n*, di circa 1μm di spessore. Il risultato, dopo questi passi, è mostrato in figura 3.12.

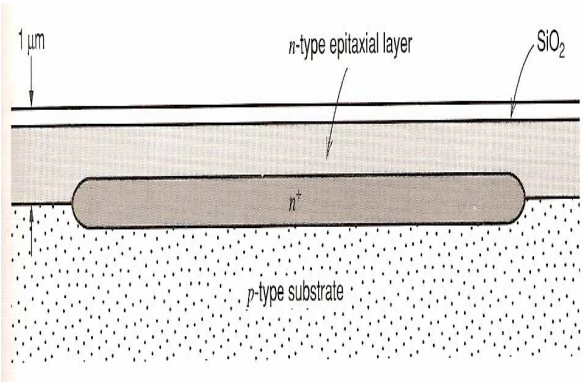

Figura 3.12

Successivamente, è realizzato uno *step* di ossidazione selettivo per formare le regioni che isoleranno il transistore e la regione di collettore dal resto del transistore. Viene realizzato uno *step* di ossidazione, tuttavia, prima della crescita effettiva dello strato spesso di SiO_2, viene realizzato uno *step* di *etching* per rimuovere il materiale di silicio dalle regioni dove l'ossido sarà

cresciuto. Se questo non viene fatto, l'ossido presenta elevate gibbosità nelle regioni in cui viene cresciuto. Gli *step* intorno a queste gibbosità causano delle difficoltà nella copertura dei successivi strati di metallo e polisilicio che saranno depositati. La rimozione di materiale di silicio prima della crescita di ossido risulta in una superficie praticamente planare dopo la crescita dell'ossido e rimuove i problemi degli *step* di copertura nelle fasi successive. La struttura risultante è mostrata in figura 3.13.

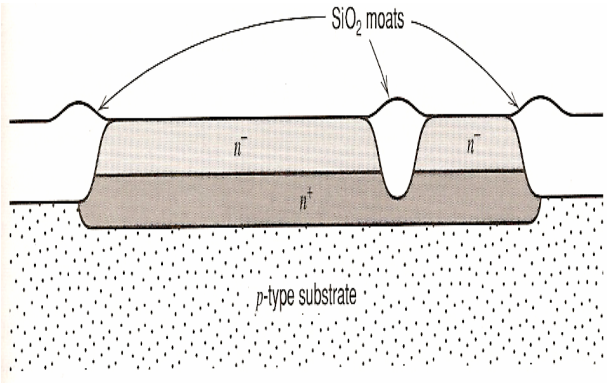

Figura 3.13

Si noti che le regioni di SiO_2 si estendono fino al substrato di tipo *p*, isolando elettricamente le regioni epitassiali di tipo *n*. Queste regioni sono spesso chiamate "fossati". Poiché la crescita degli strati di ossido più spessi di un micron richiede tempi molto lunghi, questo metodo di isolamento è usato nella pratica solo per strutture di transistori molto sottili.

Successivamente, sono realizzati due *step* di mascheratura e di impianto. Un impianto n^+ è poi realizzato nella regione di contatto di collettore e diffuso giù fino al *buried layer*, producendo un percorso a bassa resistività verso il collettore. Una seconda maschera viene impiegata per definire la regione di base e successivamente viene realizzato un impianto di tipo *p* per la base sottile. La struttura risultante è mostrata in figura 3.14.

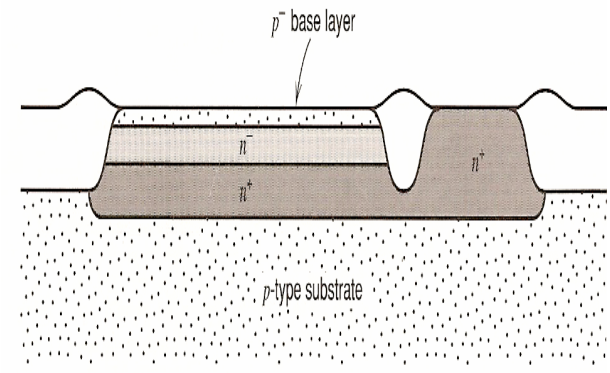

Figura 3.14

La sfida principale nella realizzazione di questo tipo di dispositivi è la formazione di strutture di emettitore e di base molto sottili e, dunque, la realizzazione di contatti ohmici a bassa resistenza per queste regioni. Questo è più spesso ottenuto usando polisilicio come sorgente drogante. Uno strato drogato n^+ di polisilicio è depositato e mascherato per lasciare il polisilicio solo nella regione immediatamente sopra l'emettitore. Durante gli *step* successivi ad alta temperatura, il drogante (usualmente arsenico) diffonde all'esterno del polisilicio e nel silicio cristallino, formando una regione di emettitore fortemente drogata e molto sottile. Dopo la deposizione del *poly*, viene realizzato un forte impianto di tipo *p*, che risulta in uno strato di tipo *p* fortemente drogato in tutti i punti della regione di base tranne che direttamente sotto il polisilicio, dove il *poly* stesso agisce come una maschera per impedire agli atomi di boro di raggiungere questa parte della regione di base. La struttura risultante è mostrata in figura 3.15.

Questo metodo di realizzazione delle regioni a bassa resistenza per contattare la base è chiamato struttura auto-allineata perché l'allineamento della regione di base con l'emettitore si verifica automaticamente e non dipende dall'allineamento delle maschere.

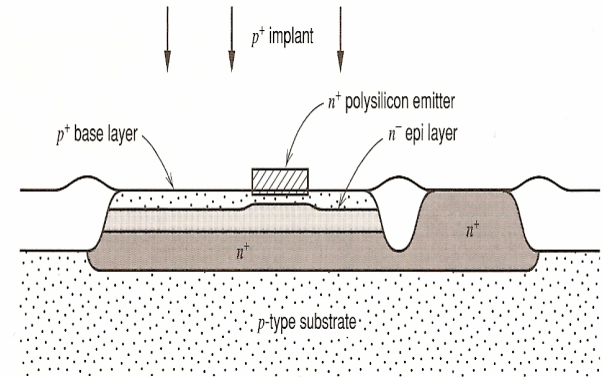

Figura 3.15

La struttura finale dopo la metallizzazione è mostrata in figura 3.16.

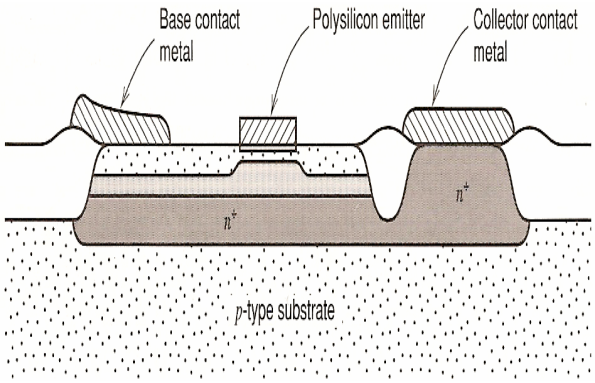

Figura 3.16

Poiché i fossati sono realizzati in SiO_2, le finestre per i contatti possono sovrapporsi in esse, un fatto che riduce drammaticamente le minime dimensioni ottenibili delle regioni di emettitore e di base. Tutto il silicio e il polisilicio esposto è coperto con *silicide* ad alta conduttività (è un composto di silicio e di un metallo refrattario come il tungsteno) per ridurre le resistenze serie e di contatto. Per i transistori a dimensione minima, il contatto di emettitore è fatto estendendo il polisilicio alla regione all'esterno dell'area attiva del dispositivo e formando in quella zona un contatto di metallo col polisilicio stesso. Una fotografia di tale dispositivo è mostrata in figura 3.17 e un tipico profilo di impurità è mostrato in figura 3.18. L'uso del contatto di emettitore remoto con la connessione di polisilicio aggiunge della resistenza

serie di emettitore, per cui, per geometrie del dispositivo maggiori o in quei casi in cui la resistenza di emettitore è critica, viene usato un emettitore più grande e il contatto è collocato direttamente sulla superficie dell'emettitore di polisilicio stesso. I processi di produzione dei circuiti integrati basati su tecnologie simili a quella descritta forniscono transistori bipolari aventi f_T di circa 10GHz, mentre quelli basati sui processi per alte tensioni a diffusione profonda raggiungono valori di f_T di circa 500MHz.

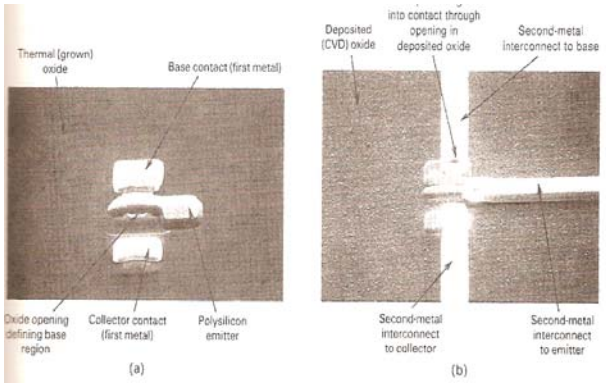

Figura 3.17

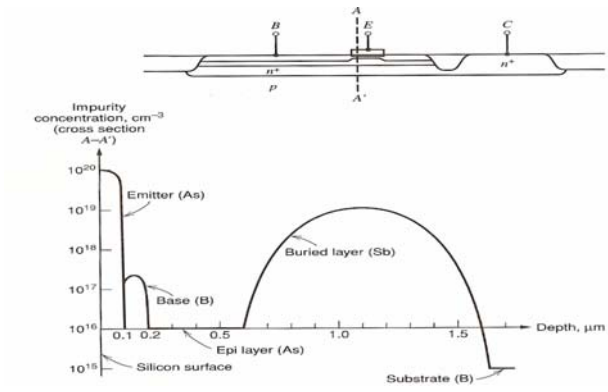

Figura 3.18

3.8 Crescita epitassiale – Aspetti sperimentali

L'epitassia da raggio molecolare (*molecular beam epitaxy* – MBE) è solo una delle diverse tecniche sperimentali per preparare le eterostrutture epitassiali con una precisione atomica. Le alternative sono principalmente la *Chemical Vapour Deposition* (CVD) e le sue principali varietà. Il vantaggio principale della MBE è la sua maggiore flessibilità rispetto alla tecnica CVD con riferimento alla scelta della velocità di crescita, della temperatura e della composizione dello strato. Dunque, la crescita tramite MBE può essere realizzata lontano dall'equilibrio e possono essere ottenute etero-interfacce con risoluzione atomica.

La preparazione completa e adeguata delle superfici del campione è assolutamente fondamentale per la successiva crescita tramite MBE. La qualità del cristallo degli strati epitassiali dipende di gran lunga dalla pulizia della superficie del substrato. Piccole contaminazioni sulla superficie possono deteriorare seriamente le interfacce e degradare le prestazioni dei dispositivi. Pertanto, i substrati sono soggetti tipicamente ad un'accurata pulizia in soluzione chimica prima di introdurli nella camera MBE. L'ossido naturale è sottoposto ad una procedura di desorbimento *in-situ* a temperature superiori a 900°C prima della deposizione degli strati. Tipicamente, il substrato viene sottoposto ad una profonda pulizia in una soluzione ossidante H_2SO_4(95%)/H_2O_2(30%) (5:1) per 5 minuti al fine di ossidare e rimuovere le particelle organiche restanti e, dunque, viene risciacquato in acqua deionizzata. Successivamente, il substrato è messo in un bagno a temperatura di 80°C che consiste di una soluzione di NH_4OH(25%):H_2O_2(30%):*DI*-H_2O(1:1:5). Viene

poi risciacquato in acqua deionizzata per altri 15 minuti, dopodichè viene immerso in un altro bagno, anch'esso a temperatura di 80°C e per 15 minuti, con *HCl(*30%):H_2O_2(30%):*DI-*H_2O (1:1:5). Un ultimo bagno in acqua deionizzata per 15 minuti conclude la pulizia in bagno chimico. Il primo bagno rimuove i contaminanti organici insolubili, mentre il secondo bagno elimina i contaminanti atomici dei metalli pesanti e ionici dalla superficie. Uno strato sottile di ossido viene lasciato sulla superficie attraverso questa procedura. Successivamente, il substrato viene asciugato con azoto e montato su un adattatore in Si per la crescita epitassiale.

Una superficie di silicio puro che è esposta ad aria ambiente inizia immediatamente ad ossidarsi e forma uno strato amorfo sottile di ossido naturale avente uno spessore di circa $20 \div 40\,\text{Å}$. Questo riduce la densità degli stati superficiali. Tuttavia, per una crescita epitassiale, è necessaria la rimozione di questo ossido prima della deposizione del materiale.

Per l'epitassia da fascio molecolare di silicio, si vede che l'approccio più semplice e più efficace è rappresentato da uno *step* di desorbimento *in-situ* dell'ossido. Durante il processo di desorbimento, il silicio diffonde nel diossido di silicio e lo converte in ossido di silicio secondo la reazione seguente:

$$SiO_2 + Si \rightarrow 2SiO$$

Le molecole di ossido di silicio sono volatili a temperature superiori a circa 870°C. Il processo di desorbimento può essere osservato usando uno spettrografo di massa come mostrato in fig. 3.19.

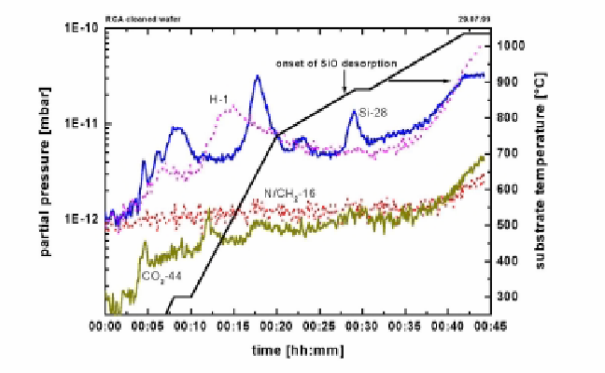

Figura 3.19

3.8.1 Sistema MBE

Il sistema MBE usato comprende una camera con blocco del carico per l'introduzione e il trasferimento dei substrati e una camera di crescita dove avviene la crescita (figura 3.20). La crescita è monitorata e controllata attraverso un software.

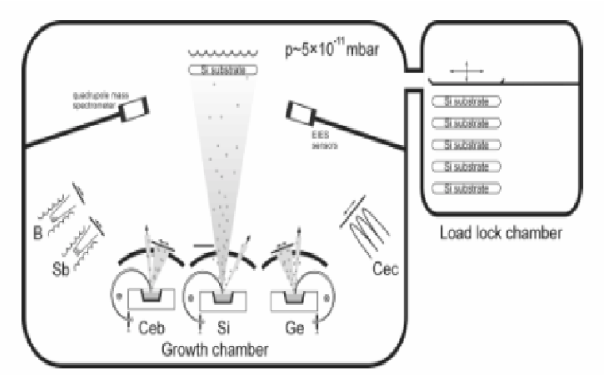

Figura 3.20

3.8.2 Vuoto

Tranne che per i substrati puliti e per il materiale di evaporazione, le condizioni di vuoto spinto sono assolutamente necessarie per ottenere strati epitassiali di elevata qualità con una minima quantità di difetti e impurità. Il flusso incidente

di atomi e di molecole Φ del gas di *background* su una superficie può essere stimato dalla relazione seguente:

$$\Phi = \frac{\beta p}{\sqrt{2\pi m k_B T}}$$

Dove p è la pressione di background, β un fattore di proporzionalità, T la temperatura di evaporazione delle rispettive molecole e m il loro peso molecolare. Perciò, una riduzione della pressione di *background* durante la crescita è altamente desiderabile. La camera di crescita viene pompata con una turbopompa e una pompa ad assorbimento ionico. Un sublimatore in titanio raffreddato ad azoto liquido è disponibile nella camera di crescita per assorbire quanto più gas residuo possibile. Gli evaporatori e tutte le superfici che si riscaldano durante la crescita sono raffreddate ad acqua.

A queste basse pressioni, la lunghezza media di cammino libero λ delle molecole è molti ordini di grandezza più grande delle dimensioni della camera; dunque, il loro moto è balistico. La lunghezza media di cammino libero di una molecola tra due collisioni con altre molecole è data dalla relazione:

$$\lambda = \frac{k_B T}{\sqrt{2}\pi\sigma^2 p}$$

dove σ è la sezione trasversale dell'interazione ($\approx 2 \div 5$ Å).

3.8.3 Misura della velocità di crescita

Il sistema MBE è costituito da sei evaporatori: evaporatori a fascio elettronico per il silicio, il germanio e il carbonio, una cella di sublimazione per il carbonio, una cella ad alta temperatura per il drogaggio di tipo *p* con boro e una

cella di espansione dell'antimonio per il drogaggio di tipo *n*. Per la crescita epitassiale di strati di SiGe vengono utilizzate solamente le camere di evaporazione del Si e del Ge.

Il principio costitutivo di un evaporatore a fascio elettronico è descritto nella figura 3.21. Un filamento caldo riscaldato con resistenze emette elettroni termoionici e un'elevata tensione di 10kV li accelera verso l'anodo perforato.

Il fascio elettronico è, dunque, deflesso con un campo magnetico seguendo un arco di 270°. Il fascio presenta una densità di energia sufficiente per riscaldare localmente ed evaporare il materiale. Con una potenza di ≈1kW è possibile ottenere velocità di crescita di $\approx 0.15\,\mathring{A}/s$ per il Ge e $\approx 0.4\,\mathring{A}/s$ per il Si.

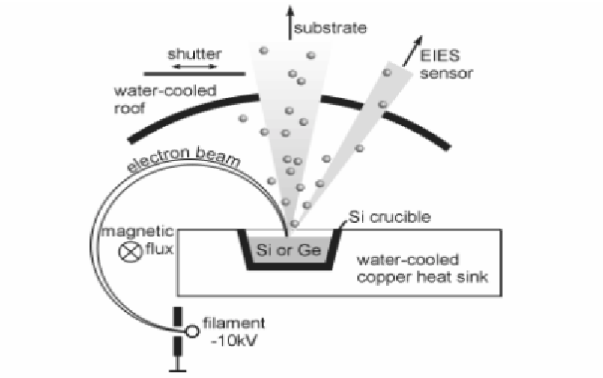

Figura 3.21

Il materiale nell'evaporatore è collocato, in entrambi i casi, nei crogiuoli realizzati in silicio puro. I crogiuoli di silicio stessi sono montati in vasche di rame raffreddate ad acqua. Una copertura raffreddata ad acqua confina il fascio molecolare. Una montatura in Si (non mostrata in figura) è collocata sopra i crogiuoli per impedire al materiale staccato dalla copertura di cadere nel crogiuolo.

Le velocità di crescita sono monitorate mediante sensori per la "spettroscopia ad emissione d'impatto elettronico" (EIES). Questa è una tecnica ottica in cui

viene misurata l'intensità di emissione delle transizioni dello specifico elemento dagli atomi che sono stati eccitati dall'impatto elettronico. Il segnale dipende dalla densità dei rispettivi atomi ed è, dunque, una funzione della velocità di evaporazione. Usando questa tecnica, possono essere monitorare velocità di crescita tra $0.05\,\text{Å}/s$ e $2\,\text{Å}/s$.

La velocità del silicio è calibrata misurando lo spessore dello strato omoepitassiale. Viene impiegata la tecnica della diffrazione a raggi X per ottenere il contenuto di Ge degli strati eteroepitassiali e, dunque, per calibrare la velocità del Ge rispetto a quella del Si.

3.8.4 Misura della temperatura

Un meandro di grafite riscaldato da resistenze è impiegato come sorgente di radiazione termica. I substrati sono mantenuti in corrispondenza della sorgente e sono riscaldati assorbendo la radiazione dal meandro dalla parte opposta. Un anello periferico di silicio montato in mezzo per ridurre le strisce sul cristallo causate dalle disomogeneità della temperatura fornisce una più omogenea distribuzione della temperatura sui substrati.

Una termocoppia disposta nel mezzo, vicina alla parte posteriore del *wafer*, viene utilizzata per le misure di temperatura. La termocoppia è stata precedentemente calibrata rispetto ad un'altra montata direttamente sul *wafer*. In più, è stato installato un pirometro per fornire una misura di temperatura indipendente opzionale a temperature comprese tra 550°C e 1050°C. Il pirometro, a sua volta, è stato calibrato rispetto alla termocoppia. Lo spot di misura sul substrato è di circa 1cm^2.

La distribuzione radiale di temperatura è stata determinata usando il pirometro; si è visto che risulta omogenea in un *range* di pochi gradi intorno al valore centrale.

3.9 Caratterizzazione strutturale

Per la valutazione della morfologia delle superfici, le tecniche di *scanning probe* sono gli strumenti ideali. Sono metodi di *imaging* locale e combinano un'elevata risoluzione con un ampio *range* di misura dinamico. La morfologia della superficie può essere "mappata" a partire da una scala di 10μm fino al livello atomico. Tuttavia, con queste tecniche è possibile eseguire il *probing* solamente delle superfici.

La *transmission electron microscopy* offre la possibilità di rilevare la morfologia degli strati epitassiali sepolti, la forma vera delle caratteristiche scoscese della superficie. Inoltre, produce informazioni sulla qualità del cristallo delle interfacce e delle densità dei difetti.

3.9.1 Atomic force microscopy (AFM)

Nella tecnica AFM una punta con un diametro di circa 100 Å viene fatta scorrere sulla superficie del campione. Essa è collocata alla terminazione libera di una sospensione la cui lunghezza è di circa 200μm. Le forze esercitate dal campione sulla punta determinano l'incurvamento della sospensione. La deflessione della sospensione è misurata con uno spot di un laser riflesso al di là della parte posteriore della sospensione stessa su un "*detector* a fotodiodo sensibile alla posizione" (PSPD). Lo scanner a tubo piezoelettrico su cui è

montato il campione è pilotato in modo che lo spot del raggio laser riflesso sia centrato tra i due fotodiodi. La morfologia della superficie è direttamente correlata con il segnale di pilotaggio dello scanner. Le costanti piezoelettriche dei cristalli usati negli scanner sono estremamente piccole, vale a dire solo una debole funzione della tensione applicata. Dunque, è possibile raggiungere un'ottima risoluzione dei profili superficiali. L'AFM lavora su campioni conduttivi e non conduttivi. Il suo principio di funzionamento è riportato nella figura 3.22.

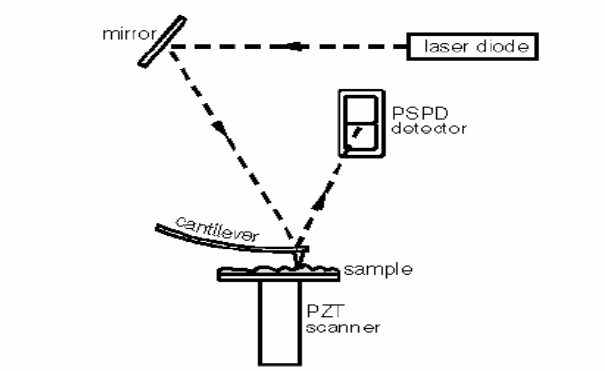

Figura 3.22

La forza complessiva tra la punta e la superficie del campione ha diversi contributi. In assenza di campi elettrostatici o magnetostatici, la forza di Van der Waals è generalmente responsabile di una forza attrattiva per medie e grandi distanze tra punta e campione. Nel regime di piccole distanze, il principio di Pauli è responsabile per la parte repulsiva.

A seconda della distanza tra punta e campione, possono essere distinti tre modi operativi nelle tecniche AFM:

1) *contact mode*, in cui la punta è in contatto fisico con la superficie;

2) *non contact mode*, che opera nel regime di forze attrattive e, in questo caso, la punta oscilla alla sua frequenza di risonanza. La forza attrattiva del

campione perturba l'oscillazione naturale determinando una variazione nell'ampiezza delle oscillazioni, che possono essere rilevate;

3) *tapping mode*, che è un modo di funzionamento compreso tra i due precedentemente descritti. La punta è alternativamente nel regime repulsivo e attrattivo ed è come se "bussasse" sulla superficie (figura 3.23).

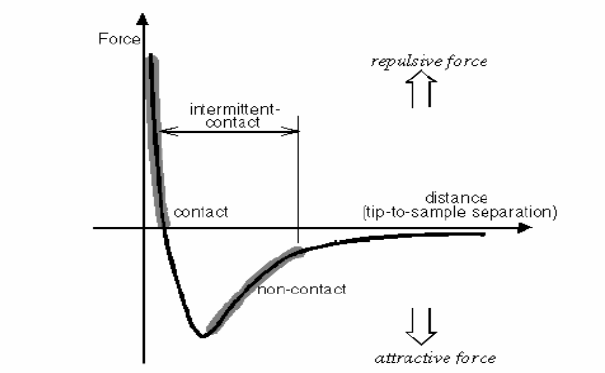

Figura 3.23

In ogni caso, qualunque sia il modo di funzionamento, è fondamentale che la punta di scansione sia più sottile delle variazioni della superficie da analizzare.

3.9.2 Scanning tunneling microscopy (STM)

La tecnica STM è oggi diventata uno strumento molto potente per esaminare i fenomeni di crescita dinamica come l'evoluzione delle morfologie o la crescita delle superfici su scala atomica. Le investigazioni tramite STM sulla crescita possono essere realizzate sia dopo il processo stesso sia durante il processo, a basse temperature.

Sono richieste condizioni di vuoto spinto per evitare l'adsorbimento e l'ossidazione del campione e della punta. Per misure STM ad alta risoluzione, sono richieste punte estremamente sottili e campioni conduttivi, ad esempio metalli o semiconduttori. Usualmente, le punte sono realizzate in tungsteno.

Tipicamente, viene applicata una tensione di circa 0.1÷2V tra la punta e il campione. Successivamente, la punta viene portata con cautela in prossimità della superficie del campione fino a che la distanza punta-campione W sia sufficientemente piccola da consentire il flusso di una corrente quanto-meccanica di *tunneling*. Per un *setup* tipico, la corrente di *tunneling* è tra 0.1 e 30nA per una distanza punta-campione di pochi angstrom. Tale corrente dipende esponenzialmente da W ed è, perciò, molto sensibile alle ondulazioni della superficie. La teoria della perturbazione dà:

$$I_{tunnel} \propto \int_0^{eU} \rho_{LDOS}^{S}\left(E_F - eU + \varepsilon\right)\cdot \rho_{LDOS}^{T}\left(E_F + \varepsilon\right)\cdot |M|^2 \, d\varepsilon \approx U\rho_{LDOS}^{S}\left(0, E_F\right)\cdot e^{-2kW}$$

dove E_F è l'energia di Fermi, $\rho_{LDOS}^{S,T}$ è la densità locale degli stati del campione (S) e della punta (T), $|M|$ è l'elemento della matrice di trasferimento, $k = \sqrt{2m\Phi_B}/\hbar$ è la componente di smorzamento del vettore d'onda, Φ_B è il potenziale di barriera dell'effetto tunnel.

Poiché la corrente di *tunneling* è molto sensibile a W, la regione della punta dove fluisce il 90% della corrente è estremamente stretta è idealmente ristretta ad un singolo atomo (fig. 3.24). Questo fornisce la risoluzione laterale a scala atomica per cui la tecnica STM è divenuta famosa nel tempo.

La tecnica STM prevede due modalità operative di funzionamento: la modalità ad altezza costante e la modalità a corrente costante. In particolare, nella modalità a corrente costante, la distanza tra punta e campione è mantenuta costante mediante un loop di *feedback*.

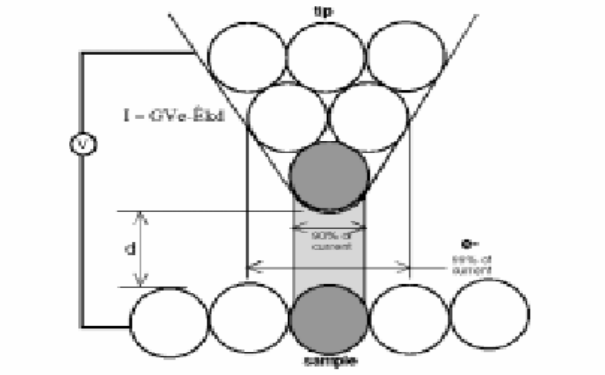

Figura 3.24

Inoltre, la tensione che deve essere applicata allo scanner piezoelettrico per mantenere costante la corrente può essere traslata in un'immagine topografica della superficie (fig. 3.25).

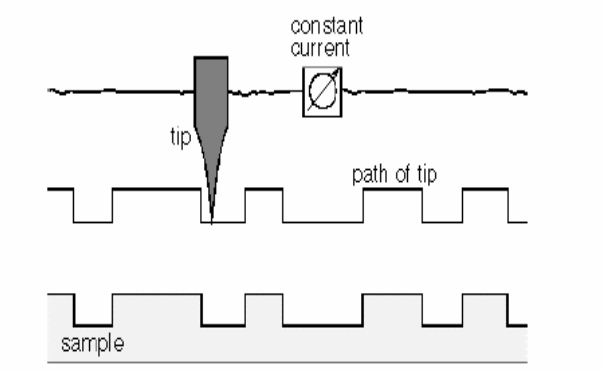

Figura 3.25

Il sistema tipicamente utilizzato consiste di una camera di crescita, una camera di misura STM, una camera di preparazione della punta e una camera di aggancio del carico. In questo modo, i campioni possono essere sottoposti a processi di crescita su di essi, trasferiti alla camera di misura e misurati in condizioni di vuoto spinto (fig. 3.26).

Il sistema della sospensione di un STM è di particolare importanza, perché la distanza tra la punta e il campione è solo di pochi angstrom. Le vibrazioni meccaniche possono far collidere la punta sul campione e renderla inutilizzabile.

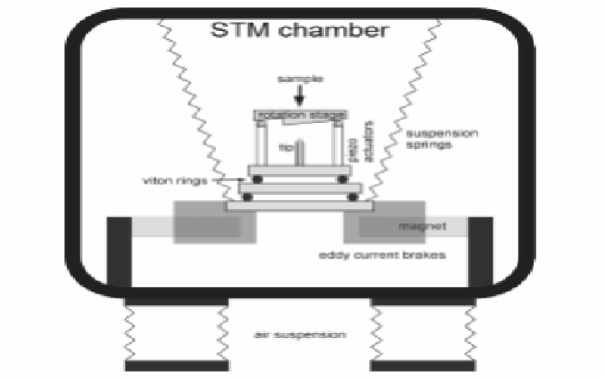

Figura 3.26

I campioni sono montati su un *holder* di molibdeno e introdotti direttamente nella camera di crescita. Dopo il processo di estrazione del gas, della durata di diverse ore a 700°C, i campioni vengono illuminati per 30s a 1250°C al fine di togliere l'ossido naturale e i contaminanti dalla superficie. Il riscaldamento viene realizzato resistivamente pilotando una corrente attraverso il campione. La velocità di crescita è monitorata usando una microbilancia al quarzo vicina al campione. La temperatura del substrato durante la crescita è misurata otticamente attraverso un pirometro. Dopo la deposizione, la temperatura viene ridotta e il campione viene trasferito all'interno stesso del sistema, sotto vuoto spinto, alla camera STM per la misura.

3.9.3 Transmission Electron Microscopy (TEM)

I TEM sono, almeno in principio, molto simili ai microscopi ottici. Tuttavia, il TEM offre una maggiore risoluzione spaziale. La risoluzione s è data da:

$$s = \frac{0.61\lambda}{NA}$$

dove λ è la lunghezza d'onda della sonda e NA l'apertura numerica.

Benché l'apertura numerica sia solo circa 0.01 per il TEM, rispetto a circa 1 di un microscopio ottico (OM), la lunghezza d'onda degli elettroni è molti ordini di grandezza più piccola di quella usata negli OM. Per elettroni con 200keV, la lunghezza d'onda di de Broglie è:

$$\lambda_c = \frac{h}{\sqrt{2em_eU}} = \frac{1.22}{\sqrt{U}} \approx 0.003nm$$

Usando un TEM tipico, è possibile ottenere risoluzioni ben sotto 1nm con magnificazioni che superano diverse centinaia di migliaia.

Un quadro schematico di un TEM è mostrato in fig. 3.27.

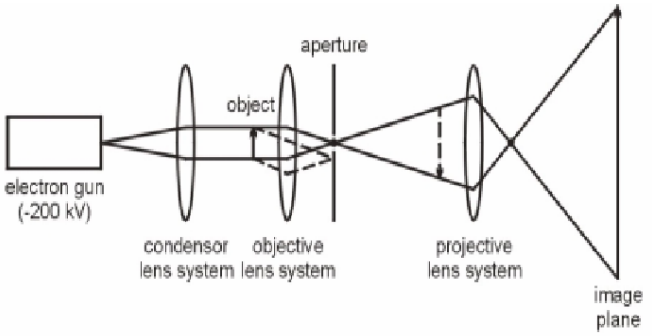

Figura 3.27

Gli elettroni provenienti da un cannone elettronico sono accelerati mediante elevate tensioni di 200kV e focalizzati con un sistema di lenti sull'oggetto. Il campione deve essere sufficientemente sottile (10÷100nm) per essere trasparente agli elettroni. Gli elettroni trasmessi formano un *pattern* di diffrazione nel piano focale posteriore del sistema di lenti. Per la modalità di *imaging* a campo luminoso, viene posizionata un'apertura nel piano focale posteriore per bloccare tutti gli elettroni eccetto il fascio di ordine zero. L'immagine ottenuta è magnificata enormemente con un sistema di lenti proiettive. Nel piano dell'immagine del sistema di lenti di proiezione, viene collocata una camera CCD per raccogliere la radiazione.

Il contrasto dell'immagine non è determinato dall'assorbimento, come negli OM, ma piuttosto attraverso lo *scattering* e la diffrazione degli elettroni nel campione. La sezione trasversale di *scattering* aumenta all'aumentare del numero atomico *Z* dell'elemento. Dunque, gli elementi più leggeri appaiono più luminosi su un film positivo rispetto a quelli più pesanti. Oltre al contrasto dovuto al numero atomico *Z*, un ulteriore contrasto può venir fuori dalle disomogeneità dello spessore del campione.

CAPITOLO 4 – Strumentazione e Misure

In questo capitolo verrano descritte le caratteristiche in frequenza del transistore bipolare e il setup di misura allestito per misurare sperimentalmente il guadagno di corrente e la frequenza di guadagno unitario del dispositivo. Inoltre, si mostrano e si confrontano i dati ottenuti dalle misure effettuate su un SiGe HBT e su un Si BJT.

4.1 Risposta in frequenza del transistore bipolare

Il guadagno in alta frequenza del transistore bipolare è controllato dagli elementi capacitivi del circuito equivalente in figura 4.1.

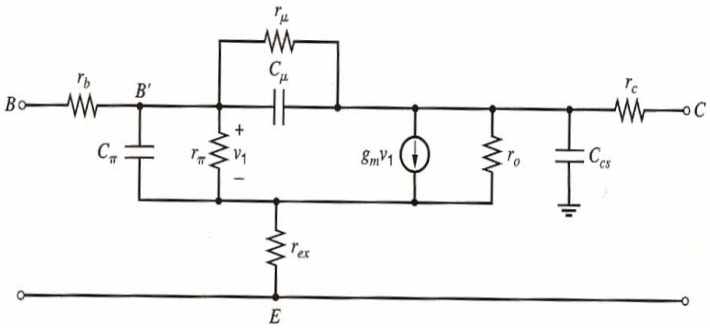

Figura 4.1

Le caratteristiche in frequenza del transistore sono più spesso specificate nella pratica determinando la frequenza a cui l'ampiezza del guadagno di corrente ad emettitore comune e in corto circuito diventa unitaria. Questa è denominata *frequenza di transizione* f_T ed è una misura della massima frequenza utile del transistore quando viene utilizzato come amplificatore. Il valore di f_T può essere misurato e calcolato usando il circuito AC di figura 4.2.

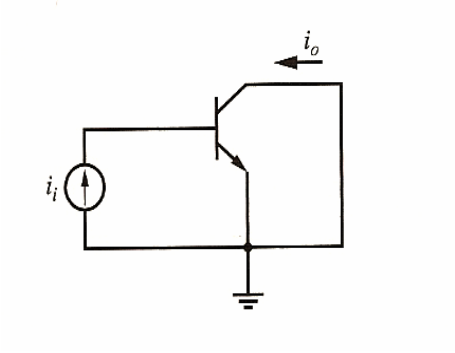

Figura 4.2

Una corrente di piccolo segnale i_i è applicata alla base e la corrente di uscita i_o è misurata con il collettore corto-circuitato per i segnali AC.

Un circuito equivalente per piccoli segnali può essere formato per questa situazione usando il circuito equivalente di figura 4.3.

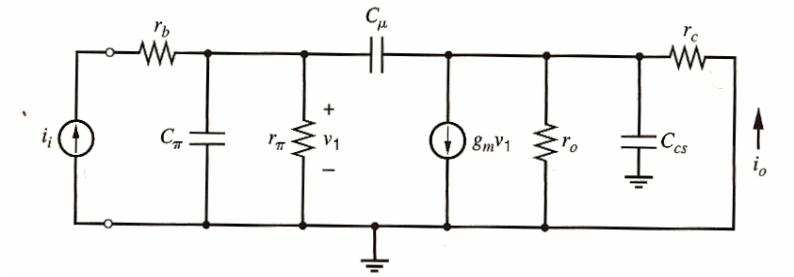

Figura 4.3

Rispetto al modello riportato in figura 4.1, sono state trascurate r_{ex} e r_μ. Se r_c è assunta essere piccola, allora r_o e C_{cs} non hanno influenza e, dunque, si ottiene:

$$v_1 \cong \frac{r_\pi}{1 + r_\pi \left(C_\pi + C_\mu\right)s} i_i \tag{4.1}$$

Se la corrente attraverso C_μ viene trascurata, allora si ha:

$$i_o \cong g_m v_1 \tag{4.2}$$

Sostituendo la (4.1) nella (4.2) si ottiene:

$$i_o \cong i_i \frac{g_m r_\pi}{1 + r_\pi \left(C_\pi + C_\mu\right)s}$$

e, dunque:

$$\frac{i_o}{i_i}(j\omega) = \frac{\beta_0}{1+\beta_0 \dfrac{C_\pi + C_\mu}{g_m} j\omega} \tag{4.3}$$

avendo usato l'espressione $r_\pi = \beta_0 / g_m$.

Ora, se $i_o / i_i(j\omega)$ è scritto come $\beta(j\omega)$ (guadagno di corrente di piccolo segnale ad alta frequenza), allora si ha:

$$\beta(j\omega) = \frac{\beta_0}{1+\beta_0 \dfrac{C_\pi + C_\mu}{g_m} j\omega} \tag{4.4}$$

Ad alte frequenze la parte immaginaria del denominatore nella (4.4) è dominante e si può scrivere:

$$\beta(j\omega) \cong \frac{g_m}{j\omega(C_\pi + C_\mu)} \tag{4.5}$$

Dalla (4.5) si vede che $|\beta(j\omega)| = 1$ quando:

$$\omega = \omega_T = \frac{g_m}{C_\pi + C_\mu} \tag{4.6}$$

e, dunque:

$$f_T = \frac{1}{2\pi} \frac{g_m}{C_\pi + C_\mu} \tag{4.7}$$

Il comportamento del transistore può essere illustrato riportando il grafico di $|\beta(j\omega)|$ mediante l'uso della (4.4), come mostrato in figura 4.4.

La frequenza ω_β è definita come la frequenza dove $|\beta(j\omega)|$ diventa uguale a $\beta_0 / \sqrt{2}$ (3dB sotto il valore DC). Dalla (4.4) si ottiene:

$$\omega_\beta = \frac{1}{\beta_0} \frac{g_m}{C_\pi + C_\mu} = \frac{\omega_T}{\beta_0} \tag{4.8}$$

Dalla figura 4.4 si vede che ω_T può essere determinata misurando $|\beta(j\omega)|$ ad una frequenza ω_x dove $|\beta(j\omega)|$ decresce con pendenza di 20dB/decade e usando l'espressione:

$$\omega_T = \omega_x |\beta(j\omega_x)| \tag{4.9}$$

Questo è il metodo di misura usato nella pratica.

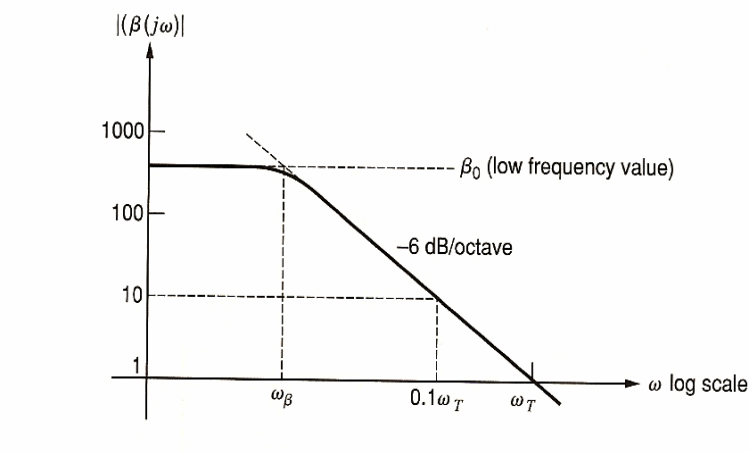

Figura 4.4

E' interessante esaminare la costante di tempo τ_T associata a ω_T. Questa è definita come:

$$\tau_F = \frac{1}{\omega_T} \tag{4.10}$$

e l'uso della (4.6) nella (4.10) dà:

$$\tau_F = \frac{C_\pi}{g_m} + \frac{C_\mu}{g_m} \tag{4.11}$$

Ora, sfruttando le due seguenti relazioni:

$$C_\pi = C_b + C_{je} \tag{4.12}$$

$$C_b = \tau_F g_m \tag{4.13}$$

e sostituendo nella (4.11) si ottiene:

$$\tau_T = \tau_F + \frac{C_{je}}{g_m} + \frac{C_\mu}{g_m} \tag{4.14}$$

Quest'ultima equazione indica che τ_T dipende da I_C (attraverso g_m) e tende al valore costante τ_F per elevate correnti di polarizzazione di collettore. Per bassi valori di I_C, i termini contenenti C_{je} e C_μ predominano e fanno in modo che τ_T cresca e f_T decresca al diminuire di I_C. Questo comportamento è illustrato in figura 4.5 e rappresenta il grafico tipico di f_T in funzione di I_C per un transistore *npn* integrato.

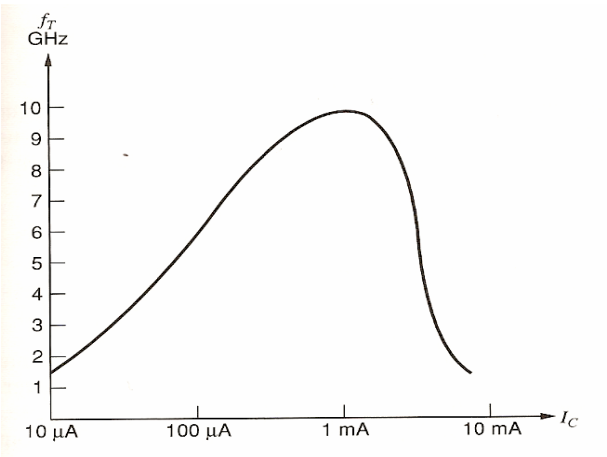

Figura 4.5

Il decremento di f_T per elevate correnti di collettore non è descritto dal semplice modello mostrato ed è dovuto ad un aumento di τ_F causato da alti livelli di iniezione e dall'effetto Kirk ad alte correnti. Questi fenomeni causano la diminuzione di β_F per elevati valori di corrente.

4.2 Misura delle caratteristiche in frequenza degli HBT

Come evidenziato nella (4.9), la caratterizzazione in frequenza degli HBT avviene attraverso una misura dei parametri *h* della rete 2-porte con la quale è

possibile rappresentare il transistore stesso. In generale, i parametri *S*, *h*, *y*, *z* e *ABCD* sono strumenti per caratterizzare una rete n-porte. In base ad uno dei set di parametri indicati e alla conoscenza dei carichi esterni applicati alle porte della rete, è possibile determinare la risposta della rete ad un segnale d'ingresso senza conoscerne la struttura interna.

I set di parametri citati in precedenza sono tutti basati sulle tensioni e correnti totali alle porte della rete.

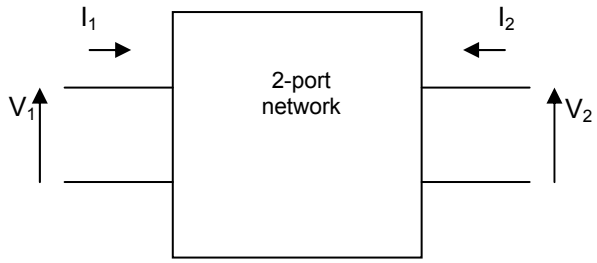

Figura 4.6

Si consideri il set di parametri *h*. Il modello a parametri *h* è tipicamente adatto alla modellizzazione circuitale del transistore e risulta essere importante perché:

- ✓ i suoi valori sono usati sui *data sheet*;
- ✓ è un modello che può essere usato per analizzare il comportamento del circuito;
- ✓ può essere usato per formare la base di un modello del transistore più accurato.

A basse e a medie frequenze, i parametri *h* sono tipicamente valori reali.

Il modello a parametri *h*, relativo alla rete 2-porte di figura 4.6, è definito come segue:

$$\begin{vmatrix} V_1 \\ I_2 \end{vmatrix} = \begin{vmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{vmatrix} \begin{vmatrix} I_1 \\ V_2 \end{vmatrix}$$

da cui:

$$V_1 = h_{11} I_1 + h_{12} V_2 \tag{4.15}$$

$$I_2 = h_{21} I_1 + h_{22} V_2 \tag{4.16}$$

Da queste relazioni è possibile estrarre le definizioni operative dei parametri *h*.

$$h_{11} = \left. \frac{V_i}{I_i} \right|_{V_o=0} \tag{4.17}$$

pertanto, h_{11} ha le dimensioni fisiche di un'impedenza (è l'impedenza d'ingresso quando l'uscita è corto-circuitata, figura 4.7).

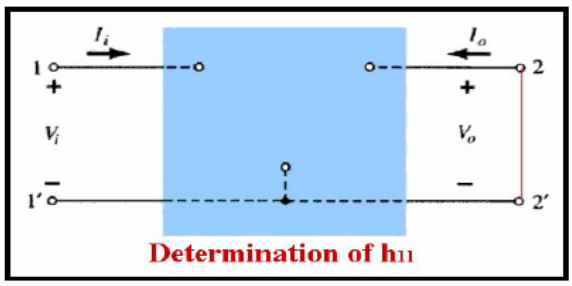

Figura 4.7

$$h_{12} = \left. \frac{V_i}{V_o} \right|_{I_i=0} \tag{4.18}$$

quindi, h_{12} è dimensionale e può essere calcolato aprendo i terminali d'ingresso e applicando una tensione di test alla porta 2 (figura 4.8).

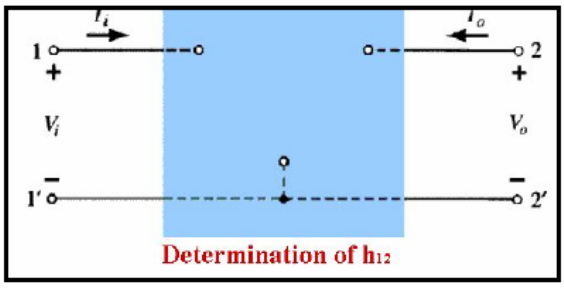

Figura 4.8

$$h_{21} = \left.\frac{I_o}{I_i}\right|_{V_o=0} \tag{4.19}$$

anche h_{21}, pertanto, risulta essere un parametro dimensionale e rappresenta il guadagno di corrente della rete 2-porte, quando l'uscita è corto-circuitata, misurato applicando una tensione di test ai terminali d'ingresso (figura 4.9).

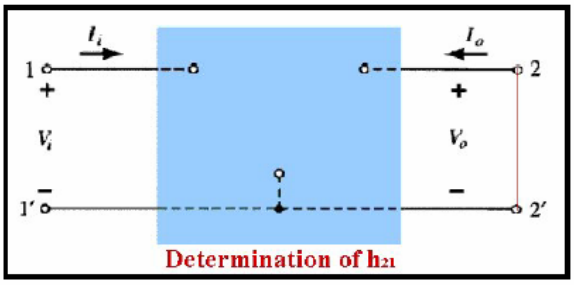

Figura 4.9

$$h_{22} = \left.\frac{I_o}{V_o}\right|_{I_i=0} \tag{4.20}$$

dove h_{22} ha le dimensioni fisiche di un'ammettenza e si calcola operativamente aprendo i terminali d'ingresso e applicando una tensione di test alla porta di uscita (figura 4.10).

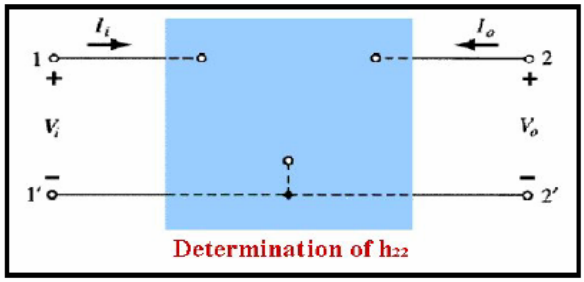

Figura 4.10

Come anticipato, l'obiettivo è quello di caratterizzare la rete ad elevate frequenze. Tuttavia, quando si cerca di effettuare le misure dei parametri indicati a queste frequenze, ci si imbatte in numerosi problemi. I principali possono essere sintetizzati come segue:

- ✓ la strumentazione non è prontamente disponibile per misurare le tensioni e le correnti totali alle porte della rete;
- ✓ i corti circuiti e i circuiti aperti sono difficili da ottenere su un'ampia banda di frequenze;
- ✓ i dispositivi attivi, come i transistori, molto spesso possono entrare in un modo di funzionamento instabile o, addirittura, danneggiarsi in presenza di circuiti aperti o corti circuiti.

Quindi, ad alte frequenze è molto difficile misurare le tensioni e le correnti totali alle porte del dispositivo. Non è possibile semplicemente connettere un voltmetro o una *probe* di corrente ed ottenere misure accurate a causa dell'impedenza delle *probe* stesse e della difficoltà di collocare le *probe* nella posizioni desiderate.

Per queste ragioni, è necessario ricorrere ad un differente modo di caratterizzare la rete a queste frequenze. La soluzione è rappresentata dai parametri *S*. Questi parametri sono definiti in termini di onde di tensione che si

propagano, le quali possono essere misurate ad alte frequenze con un *network analyzer*.

Si consideri la rete 2-porte riportata in figura 4.11.

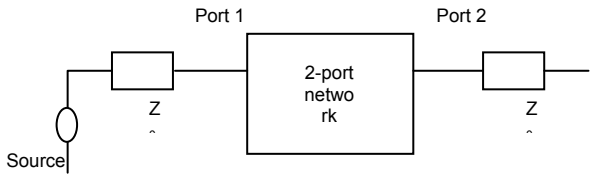

Figura 4.11

Un'onda viaggiante Ei_1, generata alla sorgente, si propaga verso la porta 1 della rete attraverso una linea di trasmissione di impedenza caratteristica Z_0 (si trascurano gli effetti della linea di trasmissione). Quando l'onda raggiunge la rete, vengono prodotte, come effetto, due nuove onde viaggianti. Una appare alla porta 2 e si propaga allontanandosi dalla rete (Er_2) e l'altra appare alla porta 1 che si propaga indietro verso la sorgente (Er_1). I parametri *S* caratterizzano la rete indicando la potenza riflessa dalla rete ad entrambe le porte ($Er_1/\sqrt{Z_0}$, $Er_2/\sqrt{Z_0}$) rispetto alla potenza incidente alle porte medesime ($Ei_1/\sqrt{Z_0}$, $Ei_2/\sqrt{Z_0}$). I parametri *S* per una rete 2-porte sono definiti dalle equazioni 4.21 e 4.22:

$$\frac{Er_1}{\sqrt{Z_0}} = S_{11}\frac{Ei_1}{\sqrt{Z_0}} + S_{12}\frac{Ei_2}{\sqrt{Z_0}} \tag{4.21}$$

$$\frac{Er_2}{\sqrt{Z_0}} = S_{21}\frac{Ei_1}{\sqrt{Z_0}} + S_{22}\frac{Ei_2}{\sqrt{Z_0}} \tag{4.22}$$

Ciascuno dei parametri *S* può essere definito indipendentemente in termini di un *rapporto di onda stazionaria* (SWR), definito come il rapporto dell'onda riflessa e dell'onda incidente, con l'altro segnale incidente posto a 0:

$$S_{11} = \left.\frac{Er_1}{Ei_1}\right|_{Ei_2=0} \tag{4.23}$$

$$S_{12} = \left.\frac{Er_1}{Ei_2}\right|_{Ei_1=0} \tag{4.24}$$

$$S_{21} = \left.\frac{Er_2}{Ei_1}\right|_{Ei_2=0} \tag{4.25}$$

$$S_{22} = \left.\frac{Er_2}{Ei_2}\right|_{Ei_1=0} \tag{4.26}$$

Le tensioni e le correnti totali sono funzioni anche delle onde viaggianti. Nello specifico, la tensione totale ad una porta è uguale alla somma delle due onde di tensione alla porta stessa:

$$V_1 = Ei_1 + Er_1 \tag{4.27}$$

$$V_2 = Ei_2 + Er_2 \tag{4.28}$$

La corrente totale è la differenza delle due onde di corrente:

$$I_1 = Ii_1 - Ir_1 = \frac{Ei_1}{Z_0} - \frac{Er_1}{Z_0} \tag{4.29}$$

$$I_2 = Ii_2 - Ir_2 = \frac{Ei_2}{Z_0} - \frac{Er_2}{Z_0} \tag{4.30}$$

A partire da queste formule possono essere sviluppate altre importanti relazioni.

Il carico alla terminazione della linea di trasmissione può essere relazionato al SWR delle tensioni riflesse ed incidenti (figura 4.12):

$$Z_{LOAD} = \frac{V_L}{I_L} = \frac{Ei_L + Er_L}{Ii_L - Ir_L} = Z_0 \frac{Ei_L + Er_L}{Ei_L - Er_L} = Z_0 \frac{1+SWR}{1-SWR} \tag{4.31}$$

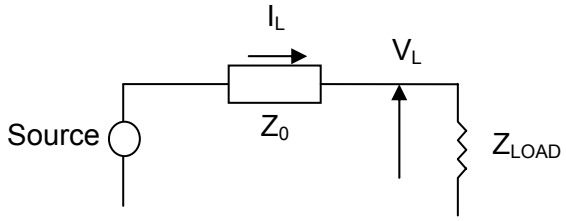

Figura 4.12

A partire dall'equazione (4.31), se $Z_0 = Z_L$ (ovvero, il carico è adattato alla linea di trasmissione) si vede che non esiste onda riflessa. Se il carico è un corto circuito ($Z_L = 0$) allora l'onda riflessa è uguale al segnale incidente ma si propaga nella direzione opposta.

Inoltre, a partire dalle equazioni mostrate precedentemente, è possibile esprimere i parametri *h*, *y*, *z* e *ABCD* in termini di parametri *S*. In particolare, tenendo conto che l'obiettivo è misurare le caratteristiche in frequenza del transistore bipolare, per lo scopo di questo lavoro di tesi risulta interessante esprimere i parametri *h* in funzione dei parametri *S*. Le relazioni sono le seguenti:

$$h_{11} = \frac{(1+S_{11})(1+S_{22}) - S_{12}S_{21}}{(1-S_{11})(1+S_{22}) + S_{12}S_{21}} \tag{4.32}$$

$$h_{12} = \frac{2S_{12}}{(1-S_{11})(1+S_{22}) + S_{12}S_{21}} \tag{4.33}$$

$$h_{21} = \frac{-2S_{21}}{(1-S_{11})(1+S_{22}) + S_{12}S_{21}} \tag{4.34}$$

$$h_{22} = \frac{(1-S_{11})(1-S_{22}) - S_{12}S_{21}}{(1-S_{11})(1+S_{22}) + S_{12}S_{21}} \tag{4.35}$$

4.3 Analisi delle reti

L'analisi delle reti (*network analysis*) è il processo attraverso il quale i progettisti e gli *application engineers* misurano le prestazioni elettriche dei componenti e circuiti usati in sistemi più complessi. Quando questi sistemi vengono impiegati per la trasmissione di segnali con un proprio contenuto informativo, l'obiettivo è quello di trasportare il segnale da un punto ad un altro con la massima efficienza e la minima distorsione. La *vector network analysis* è un metodo per caratterizzare accuratamente tali componenti misurando il loro effetto sull'ampiezza e sulla fase di segnali di test al variare della frequenza e della potenza.

In questo paragrafo, verranno trattati i principi della *vector network analysis.* La discussione riguarda i parametri comuni che possono essere misurati, in particolare il concetto dei parametri di *scattering* (parametri *S*). Sul mercato sono disponibili strumenti per l'analisi scalare e vettoriale delle reti per la caratterizzazione di componenti dalla DC fino a frequenze molto elevate (100GHz e più).

4.3.1 Misure nei sistemi di comunicazione

In ogni sistema di comunicazione occorre considerare l'effetto della distorsione del segnale. In generale, la distorsione è determinata da effetti non lineari (ad esempio, i prodotti di intermodulazione sono generati dalle portanti). Tuttavia, anche i sistemi puramente lineari possono introdurre effetti di distorsione del segnale. I sistemi lineari possono variare la forma d'onda nel tempo dei segnali

che li attraversano modificando le relazioni di ampiezza e di fase delle componenti spettrali che costituiscono il segnale.

Più in dettaglio è possibile esaminare la differenza tra comportamento lineare e non lineare.

I dispositivi lineari impongono variazioni di ampiezza e di fase sui segnali d'ingresso (figura 4.13). Qualunque sinusoide in ingresso sarà presente anche in uscita e alla stessa frequenza. Nessun nuovo segnale viene creato. Sia i dispositivi attivi sia quelli passivi non lineari possono generare uno *shift* nella frequenza del segnale in ingresso o aggiungere altre componenti in frequenza, come armoniche e segnali spuri. Segnali d'ingresso grandi possono normalmente portare i dispositivi lineari in compressione o in saturazione, causando un comportamento non lineare. Per avere una trasmissione lineare priva di distorsione, la risposta in ampiezza del dispositivo sotto test (DUT, *device under test*) deve essere piatta e la risposta di fase deve essere lineare sull'intera banda d'interesse.

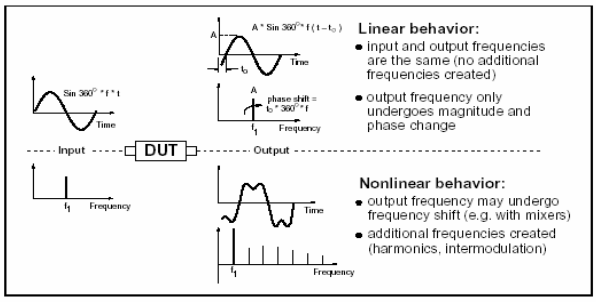

Figura 4.13

Ad esempio, si consideri un segnale ad onda quadra ricco di componenti in alta frequenza che attraversa un filtro passa-banda, il quale lascia passare frequenze selezionate con una piccola attenuazione mentre attenua tutte le frequenze esterne alla banda passante. Anche se il filtro presenta delle prestazioni di fase

lineari, le componenti fuori banda dell'onda quadra saranno attenuate, lasciando un segnale in uscita che, in questo esempio, presenta una forma più simile a quella di un'onda sinusoidale (figura 4.14).

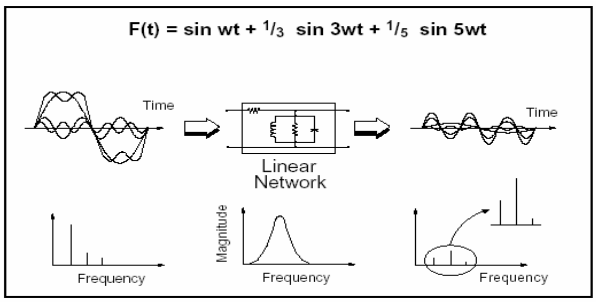

Figura 4.14

Se il medesimo segnale d'ingresso ad onda quadra è trasmesso attraverso un filtro che inverte solamente la fase della terza armonica, ma lascia inalterate le ampiezze delle armoniche, l'uscita sarà più simile ad un impulso (figura 4.15). Mentre ciò è vero per il filtro dell'esempio, in generale, la forma d'onda in uscita si presenterà con una distorsione arbitraria, a seconda delle non linearità di ampiezza e di fase.

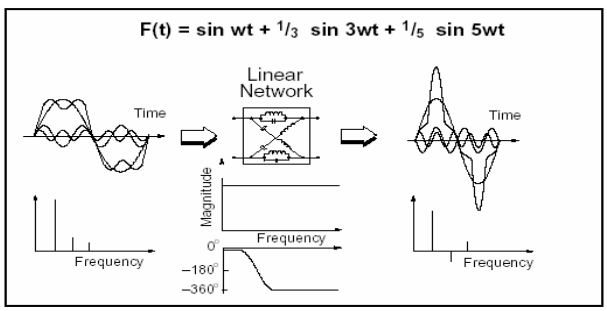

Figura 4.15

Anche i dispositivi non lineari introducono distorsione (figura 4.16). Per esempio, se un amplificatore è sovra-pilotato, il segnale d'uscita non si presenta più come una sinusoide pura e compaiono delle armoniche ai multipli della frequenza d'ingresso. Anche i dispositivi passivi possono esibire un comportamento non lineare ad elevati livelli di potenza e un buon esempio di

questo è il filtro LC che usa induttori con un nucleo magnetico. I materiali magnetici spesso esibiscono effetti di isteresi che sono altamente non lineari.

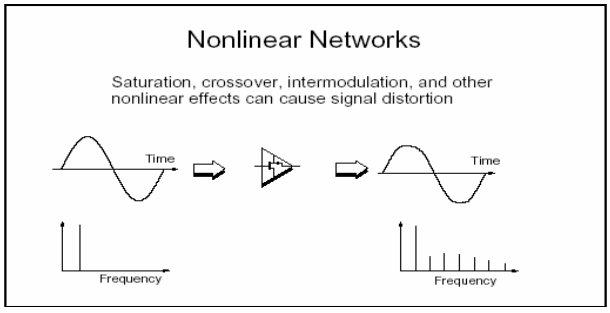

Figura 4.16

4.3.2 Misure vettoriali

La misura di entrambe le componenti d'ampiezza e di fase è importante per diverse ragioni. Primo, sono richieste entrambe le misure per caratterizzare completamente una rete lineare ed assicurare una trasmissione senza distorsione. Per progettare reti con un *matching* efficiente, occorre misurare l'impedenza complessa. I progettisti che sviluppano modelli CAD per i programmi di simulazione dei circuiti richiedono sia i dati di ampiezza sia quelli di fase per un'accurata modellizzazione. Inoltre, la caratterizzazione nel dominio del tempo richiede informazioni sull'ampiezza e sulla fase al fine di calcolare la trasformata inversa di Fourier (IFFT). La correzione di errore del vettore, che migliora l'accuratezza delle misure rimuovendo gli effetti degli errori sistematici intrinseci dovuti al sistema di misura, richiede entrambi i dati in ampiezza e in fase per costruire un modello di errore efficace. Le potenzialità delle misure di fase sono molto importanti anche per misure scalari come la perdita di ritorno, al fine di ottenere un elevato livello di accuratezza.

4.3.3 Teoria delle onde incidenti e riflesse

Nella sua forma fondamentale, un'analisi di rete riguarda le misure delle onde incidente, riflessa e trasmessa che viaggiano lungo una linea di trasmissione. Utilizzando l'analogia con le lunghezze d'onda ottiche, quando la luce colpisce una lente trasparente (energia incidente), una parte dell'onda incidente viene riflessa dalla superficie della lente, ma la maggior parte dell'energia incidente viene trasmessa attraverso la lente (energia trasmessa). Se la lente è caratterizzata da superfici riflettenti, la maggior parte della luce viene riflessa e solo una piccola parte viene trasmessa (figura 4.17).

Le lunghezze d'onda sono chiaramente diverse per segnali RF e a microonde ma il principio risulta essere lo stesso. I *network analyzers* misurano con accuratezza l'energia trasmessa, riflessa ed incidente, ovvero l'energia che si presenta all'ingresso di una linea di trasmissione, riflessa dalla linea verso la sorgente (a causa del *mismatch* d'impedenza) e successivamente trasmessa ad un dispositivo terminale (ad esempio, un'antenna).

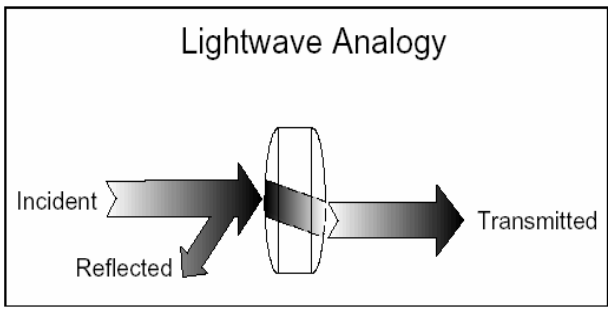

Figura 4.17

4.3.4 Condizioni per il trasferimento di potenza

Data una resistenza di sorgente R_S e una resistenza di carico R_L, alla connessione tra due dispositivi deve esistere una condizione di adattamento

perfetto affinché si abbia il massimo trasferimento di potenza al carico. Questa condizione si verifica quando $R_L = R_S$ ed è vera se lo stimolo è una sorgente di tensione DC o una sorgente di onde sinusoidali RF (figura 4.18). Quando l'impedenza della sorgente non è puramente resistiva, il massimo trasferimento di potenza si verifica quando l'impedenza di carico è uguale al complesso coniugato dell'impedenza della sorgente. Questa condizione è soddisfatta invertendo il segno della parte immaginaria dell'impedenza.

La necessità di un trasferimento di potenza efficiente è una delle principali ragioni per l'uso della linea di trasmissione a frequenze più elevate. A frequenze molto basse (con lunghezze d'onda più grandi), è possibile rappresentare la linea con un semplice conduttore (ovvero una resistenza).

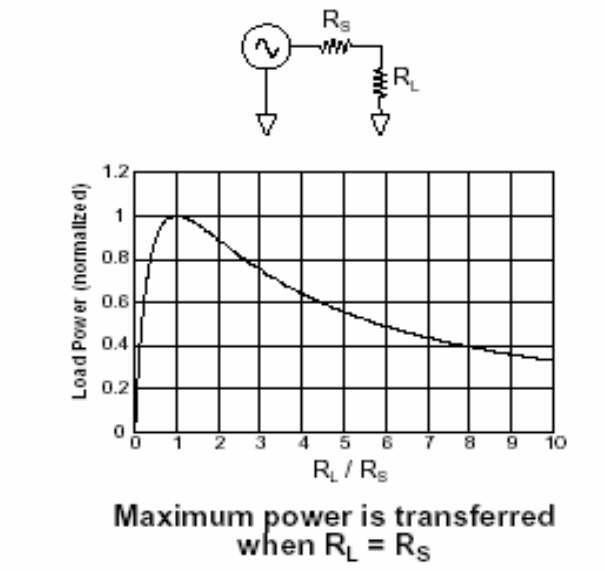

Figura 4.18

La resistenza del conduttore è relativamente bassa ed ha un piccolo effetto sui segnali a bassa frequenza. Tensioni e correnti sono le stesse indipendentemente dalla sezione del conduttore. A frequenze elevate, le lunghezze d'onda sono confrontabili o più piccole della lunghezza dei conduttori e la trasmissione di potenza può essere pensata in termini di onde che si propagano. Quando la

linea di trasmissione è terminata con un'impedenza uguale alla sua impedenza caratteristica, la massima potenza viene trasferita al carico. Quando la terminazione non è uguale all'impedenza caratteristica, quella parte del segnale che non è assorbita dal carico viene riflessa verso la sorgente.

Se una linea di trasmissione è terminata con un'impedenza uguale alla sua impedenza caratteristica, non c'è segnale riflesso poiché tutta la potenza trasmessa è assorbita dal carico (figura 4.19). Guardando all'inviluppo del segnale RF in funzione della distanza lungo la linea di trasmissione, si vede che non ci sono onde stazionarie poiché, in assenza di riflessioni, l'energia fluisce solo in una direzione.

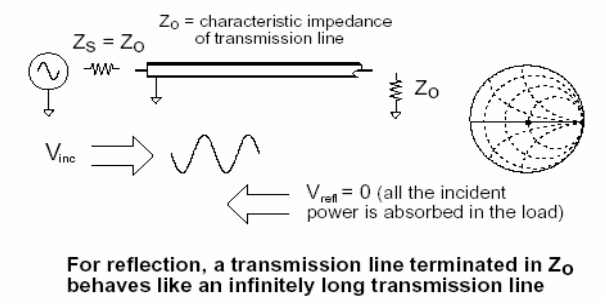

Figura 4.19

Quando la linea di trasmissione è terminata in un corto circuito (tensione nulla e, dunque, nessuna dissipazione di potenza), si verifica la riflessione di un'onda verso la sorgente (figura 4.20). L'onda di tensione riflessa deve essere uguale in ampiezza all'onda di tensione incidente ed essere sfasata di 180° sul piano del carico. Le onde incidente e riflessa sono uguali in ampiezza ma viaggiano nelle direzioni opposte.

Se la linea di trasmissione è terminata con un circuito aperto (assenza di corrente), l'onda di corrente riflessa sarà sfasata di 180° rispetto all'onda di corrente incidente, mentre l'onda di tensione riflessa sarà in fase con l'onda di

tensione incidente sul piano del carico. Questo garantisce che la corrente in corrispondenza del carico sarà nulla. Le onde di corrente incidente e riflessa sono uguali in ampiezza ma viaggiano in direzioni opposte. In entrambi i casi di corto circuito e circuito aperto, si genera un *pattern* di onde stazionarie sulla linea di trasmissione. Le valli delle tensioni saranno zero e i picchi di tensione saranno il doppio del livello della tensione incidente.

I moderni *network analyzers* misurano le onde incidente e riflessa direttamente durante uno *sweep* di frequenza e i valori d'impedenza possono essere visualizzati in differenti formati.

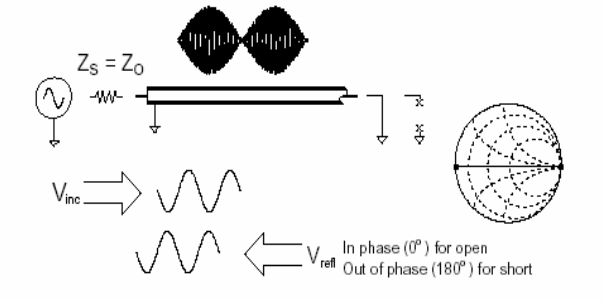

Figura 4.20

4.3.5 Terminologia dell'analisi delle reti

In generale, la terminologia dei *network ana*lyzer denota le misure di onda incidente con R (*reference channel*). L'onda riflessa è misurata con il canale A e quella trasmessa col canale B (figura 4.21).

Con le informazioni di ampiezza e di fase in queste onde, è possibile quantificare le caratteristiche di trasmissione e di riflessione di un DUT. Queste possono essere espresse attraverso un vettore (ampiezza e fase), uno scalare (solo ampiezza) o attraverso quantità di sola fase.

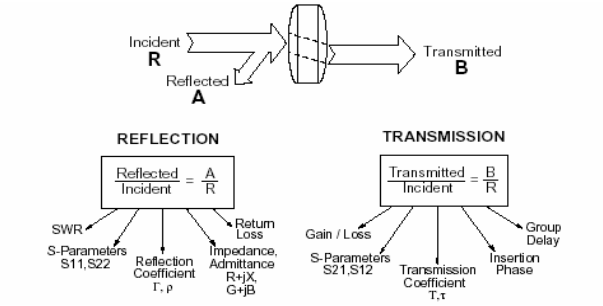

Figura 4.21

Per esempio, la perdita di ritorno è una misura scalare di riflessione, mentre l'impedenza è una misura in riflessione di un vettore. Le misure di rapporti consentono di realizzare misure di riflessione e trasmissione che risultano indipendenti dalla potenza assoluta e dalle variazioni della potenza della sorgente in funzione della frequenza. La riflessione "normalizzata" (ovvero, il rapporto tra l'onda riflessa e quella incidente) è spesso indicata con A/R e la trasmissione "normalizzata" con B/R, con riferimento ai canali di misura dello strumento.

Il termine più generale per la riflessione "normalizzata" è il coefficiente di riflessione complesso Γ (figura 4.22). L'ampiezza di Γ è indicata con ρ. Il coefficiente di riflessione è il rapporto tra il livello in tensione del segnale riflesso con il livello in tensione del segnale incidente. Per esempio, una linea di trasmissione terminata con la sua impedenza caratteristica Z_0 avrà tutta l'energia trasferita al carico e dunque $V_{ref1} = 0$ e $\rho = 0$. Quando l'impedenza del carico Z_L non è uguale all'impedenza caratteristica, l'energia viene riflessa e $\rho > 0$. Quando l'impedenza di carico uguaglia il corto circuito o il circuito aperto, tutta l'energia viene riflessa e $\rho = 1$. Pertanto, il *range* di valori possibili è $0 < \rho < 1$.

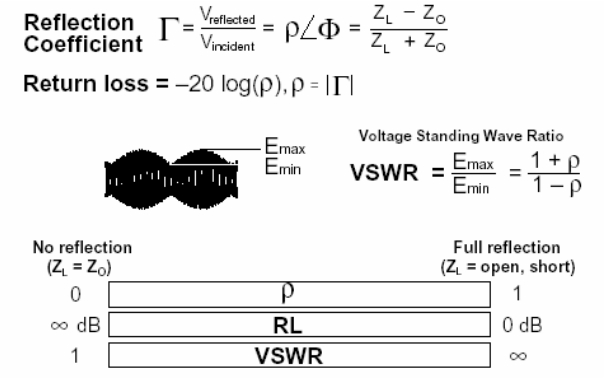

Figura 4.23

La perdita di ritorno è un modo di esprimere il coefficiente di riflessione in termini logaritmici (decibel). Esso rappresenta il numero di decibel per cui il segnale riflesso è sotto il segnale incidente, è sempre espressa con un numero positivo e varia tra infinito, per un carico uguale all'impedenza caratteristica, e 0dB per un circuito aperto o per un corto circuito. Un altro termine comune usato per esprimere la riflessione è il rapporto d'onda stazionaria in tensione (*voltage standing wave ratio*, VSWR) che è definito come il valore massimo dell'inviluppo RF rispetto al valore minimo dell'inviluppo RF:

$$VSWR = \frac{(1+\rho)}{(1-\rho)}$$

Il VSWR varia tra 1 (assenza di riflessione) e infinito (riflessione totale).

Il coefficiente di trasmissione è definito come la tensione trasmessa diviso la tensione incidente (figura 4.24).

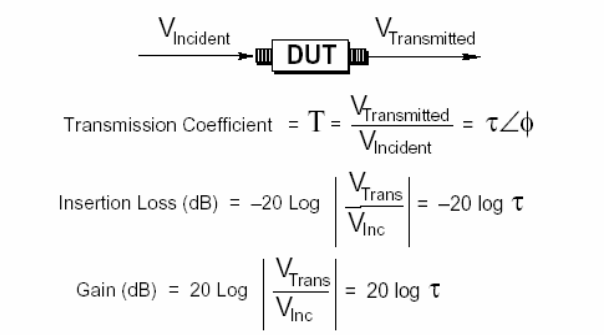

Figura 4.24

Se il valore assoluto della tensione trasmessa è maggiore del valore assoluto della tensione incidente, si dice che il DUT presenta un guadagno. Se il valore assoluto della tensione trasmessa è inferiore al valore assoluto della tensione incidente, si dice che il DUT presenta un'attenuazione o una perdita di inserzione. La fase del coefficiente di trasmissione è denominata "fase d'inserzione".

L'esame diretto della fase d'inserzione usualmente non fornisce informazioni utili. Questo perché la fase d'inserzione presenta una grande (e negativa) pendenza rispetto alla frequenza, dovuta alla lunghezza elettrica del DUT. La pendenza è proporzionale alla lunghezza del DUT. Poiché la distorsione nei sistemi di comunicazione è causata dalla deviazione dalla fase lineare, è di solito opportuno eliminare la parte lineare della risposta di fase per analizzare la parte rimanente non lineare. Questo può essere fatto usando il ritardo elettrico del *network analyzer* per cancellare matematicamente la lunghezza elettrica del DUT. Il risultato è una visualizzazione della distorsione di fase con una più elevata risoluzione (figura 4.25).

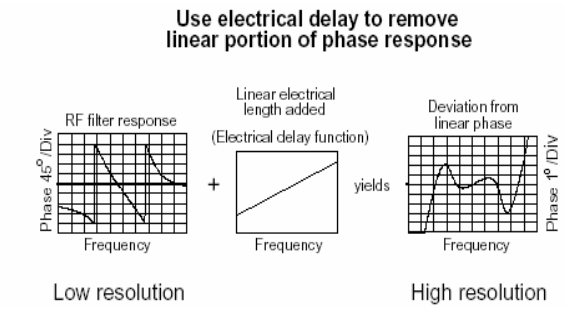

Figura 4.25

4.3.6 Caratterizzazione della rete

Al fine di caratterizzare completamente un dispositivo lineare 2-porte, occorre fare misure in diverse condizioni e calcolare un set di parametri. Questi parametri possono essere usati per descrivere completamente il comportamento elettrico del dispositivo (o rete) anche in differenti condizioni di sorgente e di carico. La caratterizzazione del dispositivo a bassa frequenza è solitamente basata sulle misure di parametri *h*, *y* e *z*. Per fare questo, occorre misurare la tensione e la corrente totali alle porte d'ingresso o di uscita di un dispositivo. Inoltre, le misure devono essere fatte in condizioni di circuito aperto e di corto circuito. Come già discusso nei paragrafi precedenti, alle alte frequenze, essendo difficile misurare le tensioni e le correnti totali, è necessario ricorrere alla misura dei parametri *S* (figura 4.26). Questi parametri si relazionano alle misure di guadagno, perdita e coefficiente di riflessione. Sono relativamente semplici da misurare e non richiedono connessioni del DUT a carichi indesiderati. I parametri *S* di dispositivi multipli possono essere messi in cascata per predire le prestazioni complessive del sistema. Inoltre, questi parametri sono usati nei *tools* di simulazione di circuiti lineari e non lineari e i parametri *h*, *y* e *z* possono essere derivati all'occorrenza dagli stessi parametri *S*, come mostrato in precedenza.

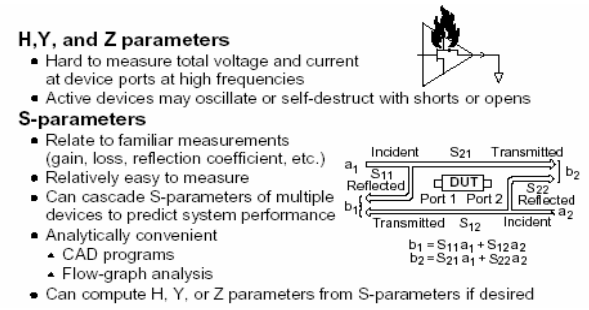

Figura 4.26

Il numero di parametri *S* per un dato dispositivo è uguale al quadrato del numero di porte. Per esempio, un dispositivo 2-porte ha 4 parametri *S*. La convenzione adottata per i parametri *S* è che il primo numero che segue la *S* sia la porta da cui l'energia esce e il secondo numero è la porta a cui l'energia entra. In tal modo, S_{21} rappresenta la misura della potenza uscente dalla porta 2 quando viene applicato uno stimolo RF alla porta 1. Quando i numeri sono gli stessi (ad esempio, S_{11}) viene indicata una misura di riflessione.

I parametri *S* diretti sono determinati misurando l'ampiezza e la fase dei segnali incidente, riflesso e trasmesso quando l'uscita è terminata in un carico che è esattamente uguale all'impedenza caratteristica del sistema di test. Nel caso di una semplice rete 2-porte, S_{11} è equivalente al coefficiente di riflessione complesso in ingresso o impedenza del DUT, mentre S_{21} è il coefficiente di trasmissione complesso diretto. Collocando la sorgente alla porta di uscita del DUT e terminando la porta d'ingresso con un carico, è possibile misurare gli altri 2 (inversi) parametri *S*. Il parametro S_{22} è equivalente al coefficiente di riflessione complesso in uscita o impedenza d'uscita del DUT, mentre S_{12} rappresenta il coefficiente di trasmissione complesso inverso (figura 4.27).

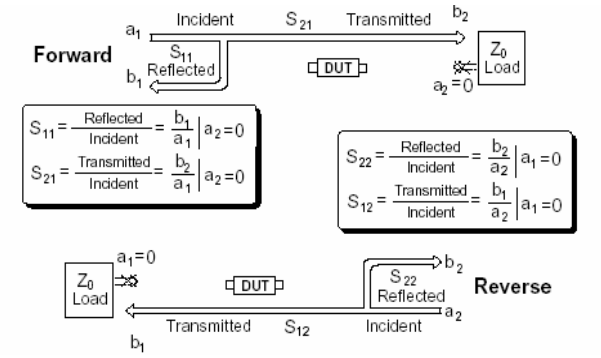

Figura 4.27

4.4 Strumentazione di misura

Il *network analyzer* utilizzato nelle misure dei parametri del transistore bipolare in SiGe può essere pensato come costituito da tre unità distinte ma integrate: il *source/converter*, il *frequency controller*, il *signal processor*.

4.4.1 Source/Converter

Il *source* è il generatore di segnale che determina i segnali sinusoidali di uscita alla porta etichettata con "RF" (figura 4.28). Il livello di potenza del segnale di uscita può essere regolato tipicamente tra +10dBm a -60dBm.

Il *converter* consiste di tre porte d'ingresso a 50Ω etichettate con R, A e B. Lo *switch* del livello d'ingresso individua la massima potenza disponibile per i segnali d'ingresso. Se la potenza in ingresso supera questo livello, si attiva un LED di *overload* per indicare ciò.

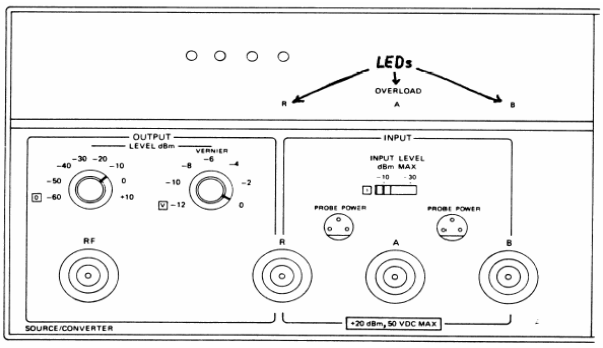

Figura 4.28

4.4.2 Frequency Controller

La funzione principale del *frequency controller* è quella di individuare il *range* di frequenza della risposta e di selezionare il tipo di *scaling* da usare. Ci sono

tre tipi di scale: lineare, logaritmica ed espansa. I *range* delle scale lineare e logaritmica sono selezionati attraverso lo *switch* RANGE (3 impostazioni possibili). La scala lineare espansa consente all'utente di selezionare un più specifico *range* all'interno di ciascuna delle impostazioni generali del *range* medesimo. Nelle misure effettuate, il *range* è stato selezionato attraverso il metodo dello "*start/stop*", ovvero si imposta il *range* desiderato selezionando la frequenza di *start* attraverso il regolatore di frequenza più a sinistra (FA in figura 4.29) e la frequenza di *stop* attraverso il regolatore di frequenza più a destra (FB in figura 4.30). Quando viene selezionato lo *switch* RANGE per espandere la scala, Il *range* di frequenze della risposta del dispositivo va dalla frequenza di *start* alla frequenza di *stop*. Un'altra caratteristica del *frequency controller* sono i *markers*. Attraverso questi è possibile selezionare ogni frequenza all'interno del *range*. Inoltre, il *controller* è dotato anche di una funzione di scansione temporale e di un *trigger control*.

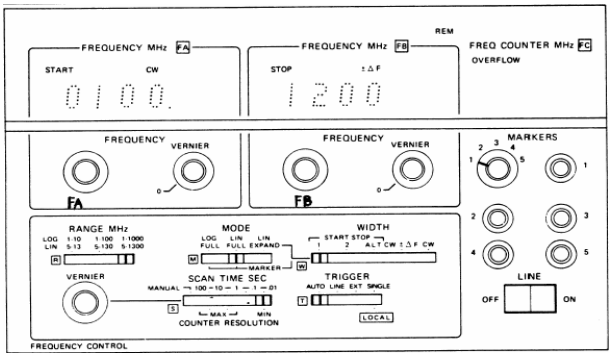

Figura 4.30

4.4.3 Signal Processor

Il *signal processor* permette all'utente di visualizzare la risposta della rete su due canali identici ed indipendenti. Per ciascun canale, l'utente seleziona il segnale da visualizzare: gli input A, B o R o i rapporti A/R o B/R. E' possibile

visualizzare l'ampiezza o la fase della risposta sia in coordinate rettangolari sia in quelle polari. La divisione della scala del display è, inoltre, regolabile.

Un lettore digitale viene utilizzato allo scopo di visualizzare gli offset di riferimento oppure l'ampiezza e la fase della risposta alla frequenza del marker. Il *signal processor* consente, inoltre, la possibilità di aumentare o diminuire la lunghezza elettrica dell'ingresso A o B al fine di eliminare le discrepanze di fase nella misure SWR (A/R o B/R).

4.4.4 Misure dei parametri di *scattering*

Un tipico diagramma a blocchi, che descrive il *setup* operativo di misura impiegato, è riportato in figura 4.31.

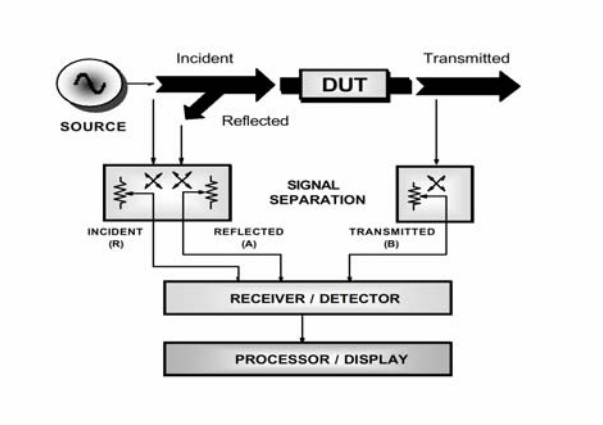

Figura 4.31

Nel diagramma sono mostrate le sezioni di processamento del segnale. Allo scopo di misurare i segnali incidente, riflesso e trasmesso, sono necessarie quattro sezioni:

- ✓ una sorgente per la generazione degli stimoli;
- ✓ i dispositivi di separazione dei segnali;
- ✓ un ricevitore per la *detection*;

- ✓ processore/display per il calcolo e la visualizzazione dei risultati.

Nel paragrafo precedente, è stata già fornita una descrizione delle componenti di un *network analyzer*, tra cui la sorgente di segnale.

Per quanto concerne i blocchi di separazione dei segnali, l'hardware usato per questa funzione viene generalmente chiamato *test set.* Il *test set* è tipicamente un blocco separato dal *network analyzer*. Le funzioni dei componenti di separazione sono sostanzialmente due. La prima consiste nel misurare una frazione del segnale incidente, allo scopo di fornire un riferimento per poter determinare il rapporto dei segnali. Questo può essere fatto utilizzando degli *splitters* o degli accoppiatori direzionali (figura 4.32).

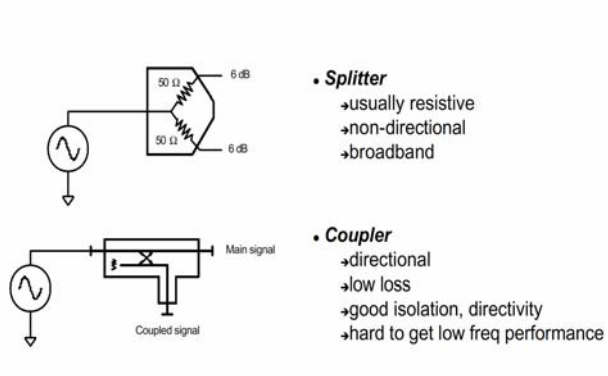

Figura 4.32

Gli *splitters* sono tipicamente resistivi. Non sono dispositivi direzionali e operano su larga banda. Il *trade-off* sta nel fatto che, nel percorso del segnale in cui sono presenti, si ha una perdita di guadagno di 6dB.

Gli accoppiatori direzionali possono essere realizzati con perdite molto basse, con un buon isolamento e un'ottima direttività. Tuttavia, è piuttosto difficile realizzare degli accoppiatori in grado di lavorare bene a basse frequenze.

La seconda funzione dei componenti di separazione del segnale è quella di separare le onde di propagazione incidente e riflessa all'ingresso del DUT.

Spesso, a causa del fatto che risulta essere difficile realizzare accoppiatori realmente a banda larga, vengono impiegati dei circuiti a ponte, i quali lavorano in un ampio *range* di frequenze introducendo, tuttavia, perdite di guadagno. I circuiti a ponte vengono utilizzati per misurare l'onda riflessa.

L'accoppiatore direzionale misura una frazione del segnale che si propaga in una sola direzione. Il segnale che passa nel ramo principale è mostrato con una linea continua, mentre il segnale accoppiato è mostrato con una linea tratteggiata in figura 4.33.

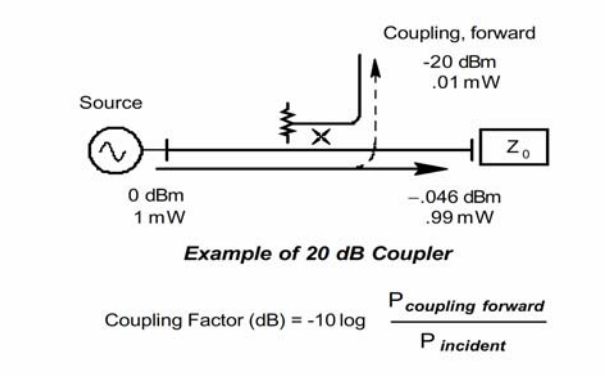

Figura 4.33

Il segnale che arriva alla porta accoppiata si presenta ridotto di una quantità nota come "fattore di accoppiamento". Questo fattore viene misurato collocando l'accoppiatore nella direzione diretta e misurando la potenza alla porta accoppiata, rispetto all'onda incidente. Idealmente, un segnale che si propaga nella direzione inversa dell'accoppiatore non dovrebbe apparire alla porta accoppiata, poiché la sua energia viene assorbita dal carico interno dell'accoppiatore o dalla terminazione esterna del ramo principale. In realtà, una parte dell'energia si perde attraverso il ramo accoppiato, come risultato di un isolamento non perfetto (finito). L'isolamento è definito come la potenza di *leakage* alla porta accoppiata rispetto alla potenza incidente.

Un altro parametro molto importante dei componenti di separazione dei segnali è la direttività, definita come la differenza (in dB) tra il fattore di accoppiamento inverso (isolamento) e il fattore di accoppiamento diretto. Un modo per misurare la direttività degli accoppiatori, che non richiede misure dirette ed inverse, consiste nel collocare un corto circuito alla porta di uscita del ramo principale (l'accoppiatore risulta essere nella direzione diretta). Si normalizza la misura di potenza a questo valore così misurato, fornendo un riferimento a 0dB. Questo passo tiene conto del fattore di accoppiamento. Successivamente, si colloca una terminazione perfetta alla porta principale dell'accoppiatore. Adesso, l'unico segnale che si misura alla porta accoppiata è dovuto al *leakage*. Siccome la misura è stata già normalizzata, il valore così misurato rappresenta la direttività dell'accoppiatore (figura 4.34).

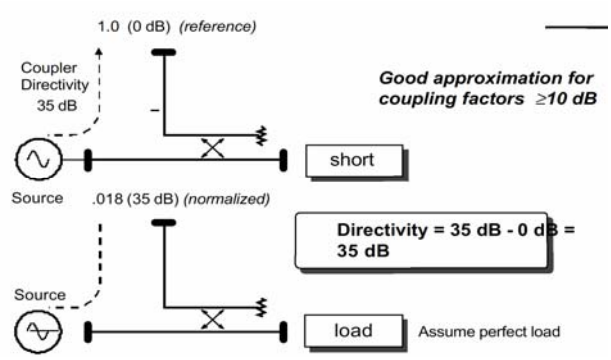

Figura 4.34

Ci sono due tecniche per fornire la *detection* del segnale in un *network analyzers* (figura 4.35). I *detectors* a diodo convertono il livello del segnale RF in un livello DC proporzionale. Se il segnale è modulato in ampiezza (AC *detection*) il diodo sottrae la portante RF dalla modulazione. La *detection* a diodo è intrinsecamente scalare, in quanto l'informazione di fase della portante RF viene persa.

Il *tuned receiver* impiega un oscillatore locale (LO) per convertire il segnale RF in un segnale a frequenza intermedia (IF). Il segnale IF è filtrato passabanda, la qual cosa restringe la banda del ricevitore e migliora notevolmente la sensibilità e il *range* dinamico. Gli *analyzers* moderni usano un ADC (*analog to digital converter*) e un DSP (*digital signal processing*) per estrarre l'informazione di ampiezza e di fase dal segnale IF.

Il vantaggio principale dei *detectors* a diodo è rappresentato dall'ampia copertura di banda, mentre il vantaggio dei *tuned receivers* sta nella maggiore sensitività, nel maggior *range* dinamico e nella capacità di "reiettare" segnali spuri.

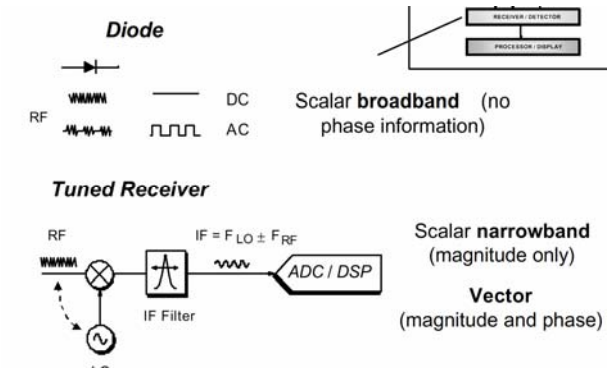

Figura 4.35

In particolare, il *network analyzer* impiegato in questo lavoro di tesi utilizza un *tuned receiver* per la *detection*. In generale, i moderni *network analyzer* sono realizzati secondo lo schema a blocchi più dettagliato, riportato in figura 4.36. I blocchi costituenti sono la sorgente integrata, un campionatore di front-end, un *tuned receiver* che fornisce i dati di fase e di ampiezza con il vettore di correzione degli errori. Il *test set* (ovvero, la parte dello strumento che contiene i dispositivi di separazione del segnale e gli *switches* per direzionare la potenza RF) può essere un *test set* basato su misure T/R (trasmissione/riflessione) o basato su parametri *S*.

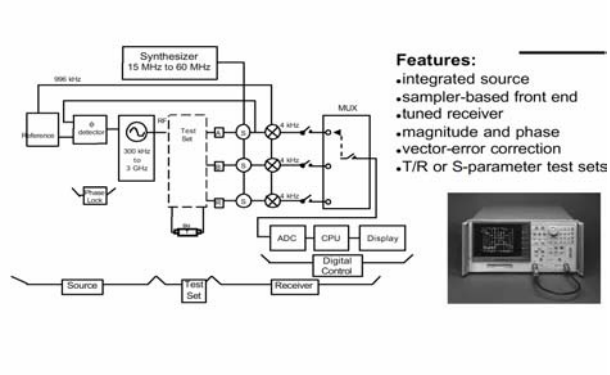

Figura 4.36

Nel *network analyzer* HP8510C, impiegato in questo lavoro, il *test set* è basato sui parametri *S*.

L'ultimo blocco che costituisce l'hardware del *network analyzer* è rappresentato dal processore/display. In questo blocco i dati relativi alla trasmissione e alla riflessione sono formattati in modo che siano di semplice interpretazione. Sono disponibili *sweep* lineari e logaritmici, formati lineari e logaritmici, grafici polari, carte di Smith, etc. Altre caratteristiche sono i *markers* di tracciamento, le linee di delimitazione, etc. Sul display è possibile visualizzare i plot dei parametri *S* misurati in funzione della frequenza.

In questo lavoro di ricerca sono state fatte misure di parametri *S* eseguendo il *probing* direttamente sul wafer (figura 4.37).

Figura 4.37

Il wafer in oggetto viene collocato sulla *wafer station*, al centro dell'apparato di misura. Come si può vedere in figura 4.38, la singola *probe* è costituita da tre morsetti, di cui uno rappresenta il segnale e gli altri due rappresentano la massa (GSG *probes*).

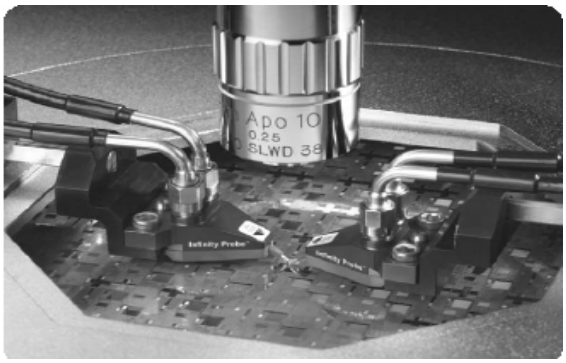

Figura 4.38

Durante le misure, vengono impiegate due *probes*, una per la trasmissione e l'altra per la ricezione, caratterizzate da un *pitch* di 100um (figura 4.39).

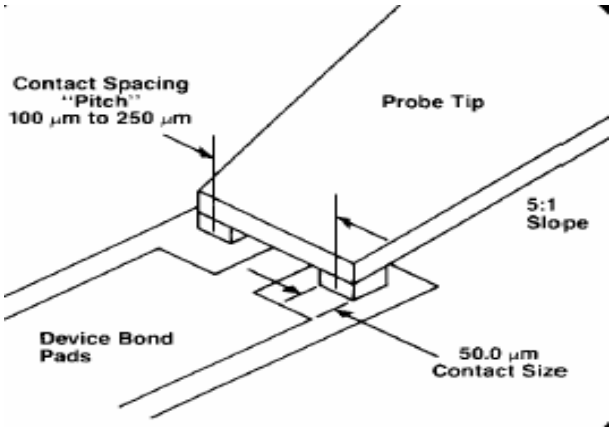

Figura 4.39

Attraverso opportuni connettori, le *probes* sono collegate ciascuna ad un cavo coassiale che porta il segnale da/verso il *network analyzer*. I connettori impiegati sono del tipo 1mm e sono stati sviluppati dalla HP, ora Agilent Technologies. La loro strategia di progetto è focalizzata su interfacce fisiche ad elevata rugosità che consentono di accoppiarsi con le dimensioni dei connettori SMA, mediante l'impiego di opportuni adattatori. La possibilità di eseguire

questo tipo di connessione è di fondamentale importanza poiché le misure effettuate sono di piccolo segnale (AC), ovvero vengono effettuate in un intorno lineare del punto operativo del circuito sotto test. Questo vuol dire che, oltre al segnale RF, occorre fornire anche un segnale di polarizzazione (*bias*) che fissi il punto operativo. La polarizzazione è fornita da un alimentatore esterno, mentre il segnale RF è fornito dalla *source* del *network analyzer*. Tipicamente, viene utilizzato un tipo particolare di adattatore, chiamato *bias T*, a forma di T (da cui il nome), che "miscela" i due segnali, *bias* e RF, e li fornisce, così "miscelati", alla porta d'ingresso del DUT. Il *bias T* è costituito da tre connettori SMA (3.5mm): questa è la ragione per cui risulta essere importante disporre degli adattatori citati in precedenza.

Infine, i cavi coassiali utilizzati possono essere impiegati per il trasporto di segnali fino a diverse decine di gigahertz e sono progettati per lavorare con gli specifici connettori utilizzati. In figura 4.40 è riportata un'immagine del *network analyzer* impiegato nelle misure.

Figura 4.40

Per misurare i parametri di *scattering*, viene impiegata la seguente procedura teorica. La figura 4.41 mostra le interconnessioni tra DUT, *test set* in trasmissione/riflessione e *network analyzer*. Il segnale incidente viene misurato alla porta R, il segnale riflesso alla porta A e il segnale trasmesso alla porta B.

Le linee di trasmissione, i connettori e le porte sono tutti caratterizzati da un'impedenza di 50Ω, pertanto tutti i carichi sono adattati al fine di impedire la presenza di riflessioni indesiderate.

A questo punto, è possibile ottenere la misura dei vari parametri *S* agendo sulle selezioni. Selezionando l'ingresso A/R ai canali 1 e 2, il modo *magnitude* al canale 1 e il modo *phase* al canale 2, è possibile visualizzare l'ampiezza e la fase del parametro S_{11}. Selezionando B/R viene, invece, visualizzato il parametro S_{21}. Scambiando i ruoli delle due porte del DUT, A/R visualizza S_{22} e B/R visualizza S_{12}. Pertanto, in questo modo possono essere visualizzate le caratteristiche complesse di tutti i parametri *S*.

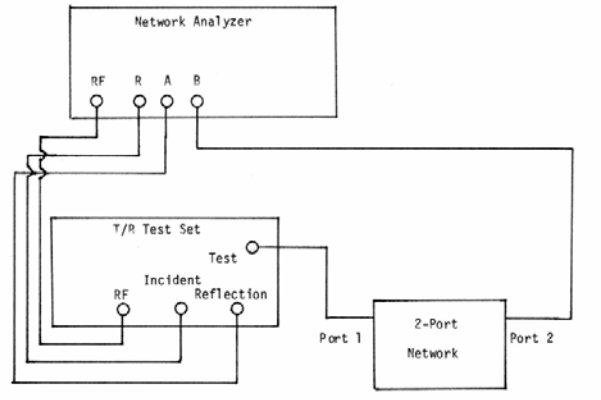

Figura 4.41

Nei moderni *network analyzer*, le operazioni descritte vengono effettuate via software, selezionando direttamente da un menu tipo Windows i parametri che si vogliono misurare. Data una certa banda di frequenze, è possibile selezionare il numero di punti in frequenza, in corrispondenza dei quali misurare i parametri *S*. Durante la misura, è possibile visualizzare sul display dell'*analyzer* i quattro parametri in funzione della frequenza. Una volta eseguite le misure dei parametri *S* in funzione della frequenza, i dati in uscita saranno costituiti da 8 vettori: i vettori di ampiezza e di fase dei quattro

parametri in oggetto, dove ogni elemento del vettore corrisponde al valore dello specifico parametro, misurato in corrispondenza di uno specifico valore di frequenza. Questi dati sono disponibili anche in formato ASCII e possono essere inviati ad un computer per la loro elaborazione.

4.4.5 Calibrazione del setup di misura

Le misure dei dispositivi in alta frequenza mediante l'uso di *network analyzer* presentano delle difficoltà legate alla calibrazione della strumentazione di misura per produrre risultati accurati, rispetto ad un piano di riferimento elettrico noto. Ad esempio, la caratterizzazione di molti componenti a microonde risulta essere difficile poiché i dispositivi non possono essere connessi direttamente a mezzi coassiali. Spesso, il DUT è realizzato in un mezzo non coassiale e, pertanto, la misura dei suoi parametri ad alta frequenza richiede componenti e cavi aggiuntivi per permettere la connessione elettrica al VNA (figura 4.42). Il punto in cui il DUT si connette al sistema di misura è definito "il piano di riferimento" del DUT medesimo. Tuttavia, ogni misura comprende, oltre al contributo del DUT, anche i contributi parassiti dei componenti (*fixtures*) e dei cavi che costituiscono il circuito di misura. In particolare, all'aumentare della frequenza, il contributo elettrico dei componenti e dei cavi diventa sempre più significativo. Inoltre, una serie di limitazioni pratiche del VNA, come un *range* dinamico limitato, l'isolamento, il *mismatch* tra sorgente e carico e altre imperfezioni contribuiscono a creare un errore sistematico nelle misure. Per ridurre sensibilmente il contributo dell'errore sistematico, per rimuovere i contributi parassiti delle *fixtures* e dei

cavi e, dunque, per incrementare l'accuratezza della misura, è necessario applicare una procedura di calibrazione del *setup* di misura attraverso l'impiego di dispositivi standard, i cui parametri elettrici siano noti a priori.

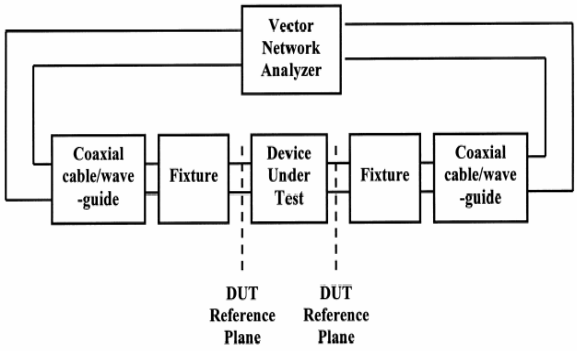

Figura 4.42

Nella figura 4.43 viene riportato un diagramma a blocchi semplificato che illustra la funzionalità del *vector network analyzer* (VNA). Generalmente, un VNA comprende uno *switch* RF in modo da permettere l'applicazione dello stimolo RF ad entrambe le porte 1 e 2 della rete, consentendo così la completa caratterizzazione della rete stessa senza la necessità di disconnettere ogni volta il DUT per invertire le connessioni. Degli accoppiatori RF, connessi all'ingresso e all'uscita della rete, consentono la misura delle tensioni riflesse.

Considerando il segnale RF applicato alla porta 1, una parte dei segnali incidente (a_1 in figura) e riflesso (b_1) vengono portati al ricevitore. Il segnale trasmesso b_2 è, a sua volta, portato al ricevitore. Il ricevitore esegue una demodulazione dei segnali RF (*downconvertion*), trasformandoli in segnali a frequenze più basse (IF), al fine di permetterne la conversione in segnali digitali e, di conseguenza, la loro elaborazione.

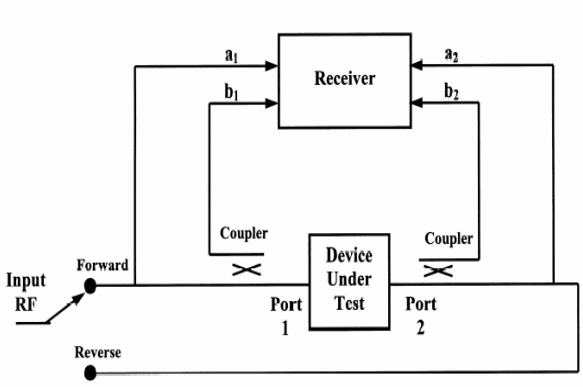

Figura 4.43

Le cause di errore nelle misure con VNA sono fondamentalmente dovute alla presenza di errori sistematici, casuali e di deriva. Le ultime due cause di errore sono impredicibili e, dunque, non possono essere rimosse dalle misure. Essi sono il risultato di fenomeni come il rumore di sistema, la ripetitività dei connettori, le variazioni di temperatura e le variazioni di parametri fisici all'interno del VNA. Gli errori sistematici, invece, sono generati da imperfezioni del VNA, sono ripetibili e possono, pertanto, essere rimossi attraverso un processo di calibrazione. In generale, alla frequenze RF, dei tre tipi di errore, quelli sistematici sono i più importanti. Durante la calibrazione, tali errori sono quantificati misurando le caratteristiche di dispositivi noti, detti "standard". Dunque, una volta quantificati, gli errori sistematici possono essere rimossi dalle misure finali. La scelta degli standard per la calibrazione non è necessariamente univoca. Spesso, è legata a numerosi fattori, tra cui anche la natura del DUT.

E' possibile fornire una descrizione matematica degli errori sistematici mediante l'uso del concetto di "modelli di errore". I modelli di errore vengono utilizzati per rappresentare gli errori sistematici più significativi in un sistema VNA fino al piano di riferimento, ovvero il piano elettrico dove gli standard

sono connessi. Dunque, si tengono in conto i contributi dei cavi e dei componenti del sistema di misura. In figura 4.44 viene riportato un diagramma di flusso che illustra un tipico modello di errore per misure in riflessione a singola porta. Il modello consiste di tre termini, E_{DF}, E_{RF} e E_{SF}. Il termine S_{11M} rappresenta il coefficiente di riflessione misurato dal ricevitore nel VNA. Il termine S_{11} rappresenta, invece, il coefficiente di riflessione del DUT rispetto al piano di riferimento (ovvero, la grandezza desiderata).

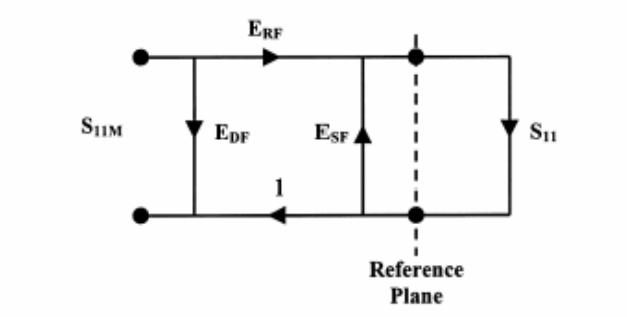

Figura 4.44

I tre termini d'errore provengono da varie sorgenti. Il termine E_{DF} tiene conto della direttività per cui il segnale riflesso misurato non risulta essere costituito interamente dalle riflessioni causate dal DUT, ma contiene anche riflessioni spurie. La limitata direttività degli accoppiatori direzionali e altri percorsi di *leakage* del segnale costituiscono le altre componenti che si combinano settorialmente con il segnale riflesso dal DUT. Il termine E_{SF} rappresenta il *mismatch* della sorgente, per cui al piano di riferimento l'impedenza non coincide esattamente con l'impedenza caratteristica della linea. Il termine E_{RF}, infine, descrive le imperfezioni nel tracciamento della frequenza tra il riferimento e i canali di test.

Nella figura 4.45 viene riportato un diagramma di flusso che illustra un modello di errore tipico per misure di reti 2-porte.

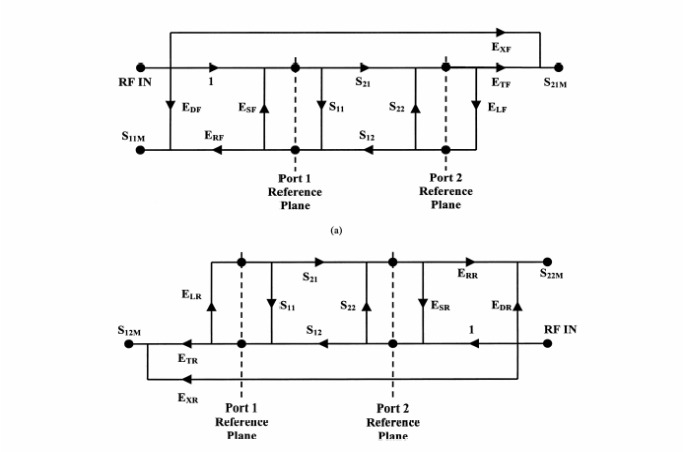

Figura 4.45

In particolare, in figura sono riportati i due modelli di errore, quello per misure dirette e quello per misure inverse. Rispetto al modello per reti ad una porta, sono presenti tre termini di errore aggiuntivi (E_{LF}, E_{TF} e E_{XF} per il percorso diretto; E_{LR}, E_{TR} e E_{XR} per il percorso inverso). Anche in questo caso, i coefficienti misurati dal ricevitore del sistema VNA sono denotati con i pedici M. I coefficienti di riflessione e trasmissione del DUT sono, invece, denotati come al solito (S_{ij}). Il termine E_{LF} tiene conto degli errori di misura dovuti alla terminazione di carico imperfetta. Il termine E_{TF} descrive gli errori di tracciamento della frequenza in trasmissione. Il termine E_{XF} tiene conto del fatto che una piccola componente del segnale trasmesso, che raggiunge il ricevitore, è dovuta all'isolamento non perfetto per cui essa raggiunge il ricevitore senza passare attraverso il DUT. Ovviamente, i coefficienti d'errore per il percorso inverso sono definiti allo stesso modo.

Da quanto esposto finora, è possibile stabilire delle relazioni tra i parametri di scattering non corretti (S_M), i parametri S del DUT e i termini di errore. Per esempio, nel caso di misura su una rete ad una porta, il coefficiente di riflessione del DUT è dato da:

$$S_{11} = \frac{S_{11M} - E_{DF}}{E_{SF}\left(S_{11M} - E_{DF}\right) + E_{RF}}$$

Allo stesso modo, per la rete 2-porte, i parametri *S* del DUT possono essere legati ai termini di errore e ai parametri *S* misurati. I parametri del DUT S_{11} e S_{21} possono essere descritti come funzioni di S_{11M}, S_{12M}, S_{21M} e S_{22M} e dei sei termini di errore diretti. Similmente, i parametri S_{12} e S_{22} sono funzioni dei quattro parametri *S* misurati e dei sei termini di errore inversi. Pertanto, una volta noti i coefficienti di errore, è possibile ricavare i parametri *S* del DUT a partire dai parametri misurati, attraverso le relazioni sopra menzionate.

La calibrazione risulta essere, quindi, quel processo che essenzialmente consente la determinazione dei coefficienti d'errore. Questo processo viene realizzato sostituendo il DUT con un certo numero di standard le cui proprietà elettriche sono note rispetto al piano di riferimento desiderato.

Inoltre, essendo il sistema dipendente dalla frequenza, questa procedura deve essere ripetuta per tutte le frequenze d'interesse. In tal modo, i coefficienti di errore verranno memorizzati nel VNA (in ampiezza e fase) ad ogni frequenza, dopodichè vengono sottratti vettorialmente alle misure per ottenere i parametri *S* del dispositivo.

Per le misure on-wafer, gli standard sono dei wafer sui quali sono implementati un corto circuito (*short*), un circuito aperto (*open*), un carico opportuno (*load*) e una linea di trasmissione avente una data lunghezza elettrica diversa da zero (*through*). La tecnica di calibrazione che impiega questi standard è nota come SOLT (*short*, *open*, *load*, *through*). Nella figura 4.46 sono raffigurate le terminazioni, che rappresentano gli standard, impiegate nella tecnica SOLT. Questi standard sono realizzati su un wafer fornito dalla Casa costruttrice del

network analyzer. Il wafer viene posizionato nella *wafer station* del setup di misura e su di esso vengono eseguite le misure per il calcolo degli errori.

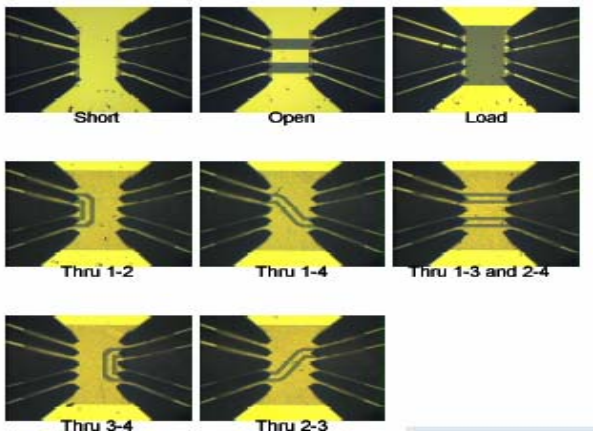

Figura 4.46

Durante il processo di calibrazione, il setup di misura deve essere perfettamente identico a quello che, successivamente, verrà utilizzato nelle misure del DUT. L'unica differenza è rappresentata dal fatto che, essendo i dispositivi da misurare dei carichi particolari (corti circuiti e circuiti aperti), per evitare sovraccarichi, è necessario accendere l'alimentatore ma utilizzare una tensione di *bias* nulla (occorre polarizzare a 0V). Dal display dell'*analyzer* si seleziona il kit di calibrazione in uso, dopodiché si seleziona il tipo di standard che si sta misurando (procedura manuale). Ogni standard è caratterizzato da parametri elettrici definiti e le misure sui quattro standard citati permettono di calcolare tutti i termini d'errore del modello per la rete 2-porte. I coefficienti d'errore, così calcolati, sono memorizzati all'interno dell'*analyzer* stesso, e saranno automaticamente sottratti dalla misure successive. In tal modo, verranno forniti i valori corretti dei parametri misurati sul DUT.

Le definizioni elettriche per gli standard SOLT ideali e privi di perdite (rispetto ai piani di riferimento delle porte 1 e 2) sono riportate in figura 4.47.

Naturalmente, soprattutto ad alte frequenze, è impossibile realizzare standard che siano completamente privi di perdite e che esibiscano coefficienti di riflessione e di trasmissione ben definiti ai piani di riferimento.

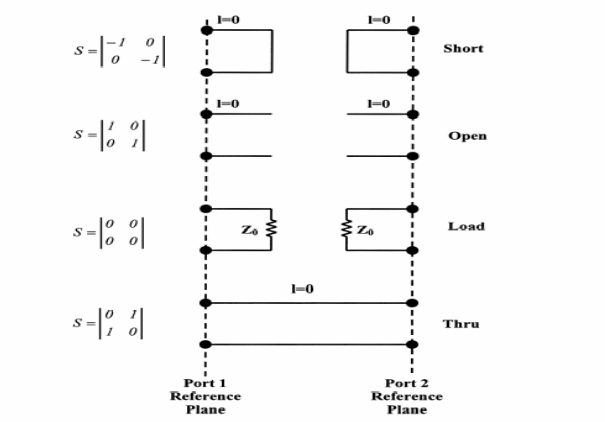

Figura 4.47

Nella realtà, occorre considerare delle linee di trasmissione di lunghezza elettrica non nulla che devono essere associate alle rappresentazioni elettriche degli standard (figura 4.48).

Quindi, per completezza, le caratteristiche della linea di trasmissione devono essere note ed incluse nella definizione dei parametri di ciascuno standard.

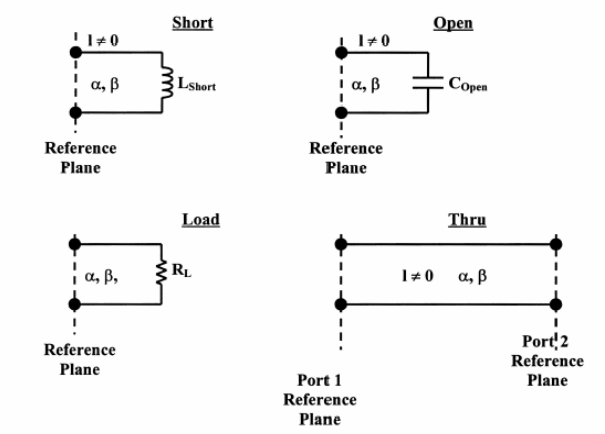

Figura 4.48

Per quanto concerne il *through*, è noto che la propagazione delle onde è descritta mediante la relazione:

$$V(z) = Ae^{-\gamma z} + Be^{\gamma z}$$

dove γ è la costante di propagazione definita come:

$$\gamma = \alpha + j\beta$$

Se si assume che la lunghezza elettrica della linea di trasmissione, associata allo standard, sia piccola, le perdite diventano piccole e il parametro α può essere trascurato, senza nessun degrado significativo dell'accuratezza. Sui wafer del kit di calibrazione sono disponibili linee di trasmissione di varie lunghezze elettriche.

Lo standard *open* esibisce ulteriori imperfezioni poiché il campo elettrico alla terminazione di *open* tende a variare con la frequenza. L'effetto è spesso descritto in termini di una capacità di *fringing*, funzione della frequenza, C_{Open}, espressa come un'espansione polinomiale:

$$C_{Open} = C_0 + C_1 F + C_2 F^2 + C_3 F^3 + \ldots$$

dove C_0, C_1, … sono i coefficienti e F è la frequenza.

Lo standard *load* determina per buona parte i termini di errore di direttività diretta ed inversa (E_{DF} ed E_{DR}). Considerando i modelli di errore riportati graficamente nelle figure 4.44 e 4.45, con lo standard *load* applicato alla porta 1, l'errore di direttività diretta assume la forma seguente:

$$E_{DF} = S_{11M} - \frac{S_{11Load} E_{RF}}{1 - E_{SF} S_{11Load}}$$

dove S_{11Load} rappresenta l'effettivo coefficiente di riflessione dello standard *load*. Idealmente, lo standard *load* dovrebbe esibire un'impedenza pari a Z_0 (impedenza caratteristica) e, dunque, un coefficiente di riflessione nullo. In questo caso, E_{DF} fornisce il valore misurato di S_{11} con lo standard *load*

connesso alla porta 1. A frequenze più alte oppure quando le prestazioni elettriche degli standard *load* risultano essere inadeguate, vengono utilizzate delle terminazioni scorrevoli. Queste impiegano metodi meccanici per regolare la lunghezza elettrica della linea di trasmissione associata allo standard *load*. Trascurando le perdite nella linea di trasmissione, l'espressione precedente forma un cerchio nel piano di misura di S_{11}, al variare della lunghezza della linea. Il centro del cerchio definisce il termine di errore E_{DF} (figura 4.49).

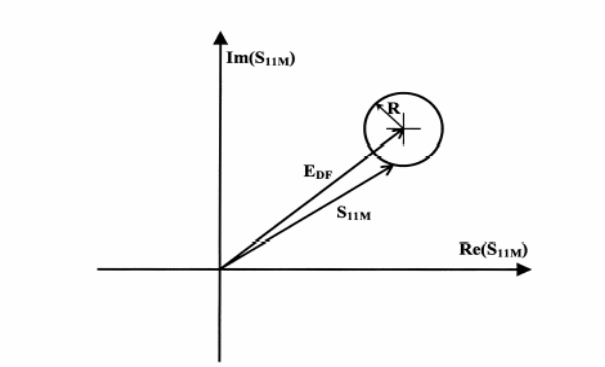

Figura 4.49

4.5 Misure in alta frequenza sui transistori bipolari

L'estrazione dei parametri capacitivi e DC di un SiGe HBT viene fatta utilizzando un sofisticato software di estrazione dei parametri "customizzato" per le specifiche aziendali. Tuttavia, se il dispositivo in questione deve essere impiegato in applicazioni ad alta frequenza, diventa molto importante l'estrazione dei parametri SPICE associati al modello AC per piccoli segnali. A tale scopo, vengono realizzate ed analizzate le misure dei parametri *S*. Come già anticipato, nel setup di misura è stato utilizzato il *network analyzer* 8510C della Agilent Technologies. Questo strumento è in grado di caratterizzare il DUT in una banda di frequenze 45MHz÷110GHz (per le caratteristiche

tecniche dello strumento, si faccia riferimento al *datasheet* il cui riferimento è indicato nella Bibliografia).

Quando un SiGe HBT deve essere impiegato in un circuito ad alta frequenza, è necessario effettuare delle stime accurate dei parametri associati al modello AC del transistore. L'implementazione in SPICE del modello AC per piccoli segnali è, fondamentalmente, una linearizzazione del modello per ampi segnali intorno al suo punto operativo. Pertanto, è fondamentale che le capacità, le resistenze di collettore, di base e di emettitore siano estratte prima di lavorare sui parametri AC. Occorre aggiungere sostanzialmente tre elementi al modello DC del HBT allo scopo di completare il modello AC. Questi elementi aggiuntivi sono la resistenza variabile di base, la capacità di diffusione base-emettitore e la capacità di diffusione base-collettore ed essi aggiungono un ritardo RC alla risposta in frequenza del modello per piccoli segnali. Al fine di caratterizzare questi tre elementi addizionali del modello AC, in genere si determinano il guadagno di corrente AC diretto ed inverso e l'impedenza d'ingresso, attraverso le misure dei parametri *S*.

Di solito, il compito di eseguire misure di parametri *S* risulta essere piuttosto difficile. In particolare, quando si eseguono misure in bande di frequenza dell'ordine delle centinaia di GHz, è necessario utilizzare delle metodologie di calibrazione al fine di "mascherare" le imperfezioni tipiche della strumentazione di misura, delle attrezzature utilizzate per il test e anche delle strutture di test, nel caso in cui vengano eseguite misure di parametri *S* direttamente sul wafer. E' molto importante individuare e quantificare gli errori sistematici legati al *mismatch* d'impedenza, all'isolamento, ai *leakages* e alla risposta in frequenza della strumentazione e delle attrezzature impiegate nel

setup di misura. Inoltre, l'obiettivo è quello di "compensare" questi scostamenti attraverso il processo di calibrazione, descritto nella sezione precedente.

Il SiGe HBT è stato analizzato nella configurazione circuitale ad emettitore comune (figura 4.50).

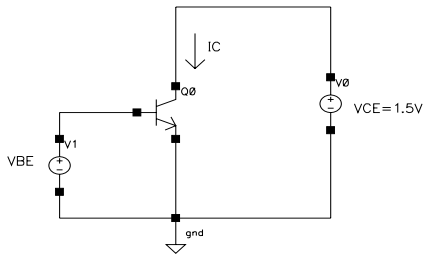

Figura 4.50

Nel setup utilizzato, il software citato sopra gestisce tutte le operazioni di selezione e di *sweep* della polarizzazione, permettendo così di eseguire la misura di tutti e quattro i parametri *S* per ciascuna polarizzazione scelta. La schermata del software usata a tale scopo è rappresentata in figura 4.51.

E' possibile utilizzare differenti opzioni relative allo *sweep* della polarizzazione. Lo schema scelto in questo lavoro consiste nel fare uno *sweep* della tensione di *bias* V_{BE}, mantenendo costante la tensione V_{CE} al valore definito. Queste opzioni vengono selezionate sulla schermata del software relativa alla configurazione elettrica del dispositivo da misurare.

Inoltre, lo stesso software permette di elaborare i dati di misura, in uscita al *network analyzer*, e, mediante procedure interne, di trasformarli nei valori corrispondenti di un altro set di parametri. Nel caso specifico, i parametri *S* misurati vengono convertiti in parametri *h* allo scopo di ricavare l'andamento del guadagno di corrente in funzione della frequenza e in funzione del *bias*.

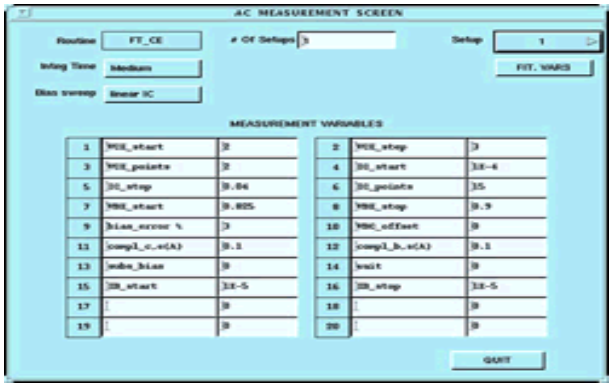

Figura 4.51

Per ogni punto di polarizzazione DC viene estratto un grafico di h_{21} in funzione della frequenza, riportato su scala logaritmica, dopodichè è possibile estrarre, attraverso il software, la frequenza di guadagno unitario f_T . In figura 4.52 viene riportata una schermata tipica che consente la visualizzazione del grafico suddetto.

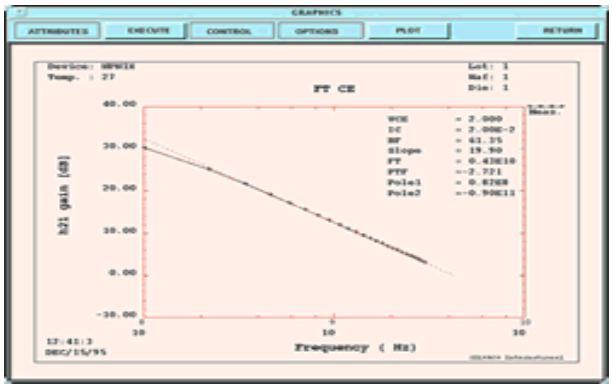

Figura 4.52

Con il software in oggetto, è, inoltre, possibile collezionare i valori delle frequenze di guadagno unitario per ogni punto operativo e, attraverso uno specifico programma, visualizzare il grafico di f_T in funzione di I_C. In questo lavoro, si è preferito utilizzare MATLAB per la raccolta dei dati e la loro visualizzazione, in virtù della sua maggiore flessibilità computazionale e grafica.

In questa attività di ricerca, sono stati considerati due differenti dispositivi. Il primo è un classico BJT in Si, il secondo è un HBT con la regione di base in SiGe. Sono state misurate le caratteristiche in alta frequenza di entrambi i dispositivi al fine di confrontarne le prestazioni e di verificare che l'impiego del SiGe effettivamente determina un incremento nella risposta in frequenza del transistore bipolare rendendolo, pertanto, più veloce.

Ovviamente, i dispositivi confrontati hanno le stesse caratteristiche costruttive. In particolare, l'area di emettitore dei dispositivi misurati è $0.4 \cdot 12\,\mu m^2$ e la tensione collettore-emettitore è stata mantenuta costante al valore $V_{CE} = 1.5V$.

Per il SiGe HBT, il profilo di Ge in base è del tipo riportato in figura 4.53.

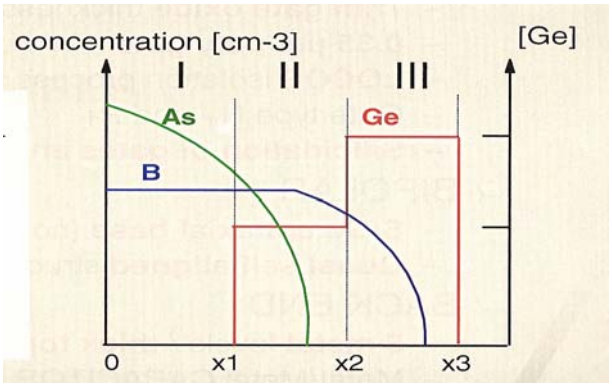

Figura 4.53

Nella figura 4.54 vengono riportati i grafici dei valori misurati di f_T in funzione della corrente di collettore I_C, espressa in ampere (A). In particolare, la curva rossa esegue il *fitting* dei valori misurati sul SiGe HBT, mentre la curva blu rappresenta il *fitting* dei valori misurati sul Si BJT.

Dalla figura si evince che i valori di picco di f_T sono ottenuti in corrispondenza di $I_C \cong 6mA$. Inoltre, il valore di picco per il SiGe HBT risulta essere $f_{T_{SiGe}} \cong 45GHz$, mentre il valore di picco per il Si BJT è $f_{T_{Si}} \cong 26GHz$.

Questo risultato dimostra che l'impiego del silicio-germanio, nella realizzazione della base del transistore bipolare, ne ha migliorato notevolmente le caratteristiche in alta frequenza, aumentandone la banda e, dunque, la velocità di commutazione.

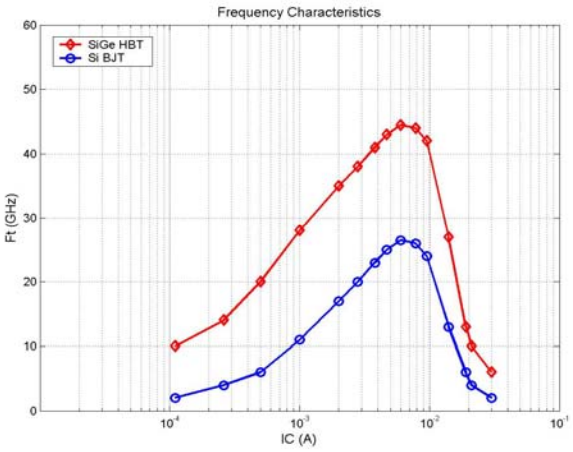

Figura 4.54

Inoltre, sono state eseguite misure del guadagno di corrente (h_{21}) in bassa frequenza per differenti valori della corrente di collettore (quindi, per differenti polarizzazioni). In figura 4.55 sono riportati i grafici che interpolano i dati misurati, sia per il SiGe HBT (curva in rosso) sia per il Si BJT (curva in blu).

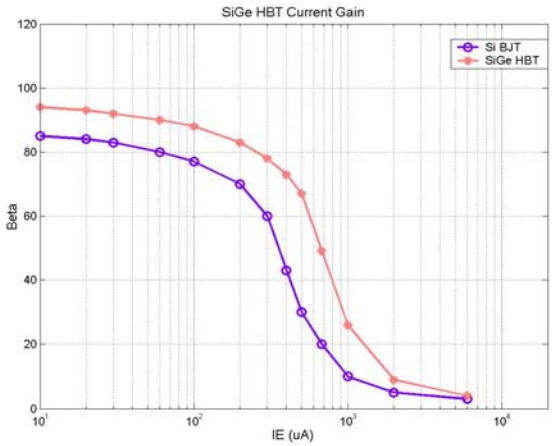

Figura 4.55

I grafici sono stati ottenuti riportando i valori di β, misurati per differenti polarizzazioni, in MATLAB e visualizzando i plot che interpolano i dati stessi. Come si può vedere, il guadagno di corrente risulta essere più elevato, a parità di caratteristiche costruttive, nel caso di SiGe HBT, verificando sperimentalmente l'assunto teorico esposto nei capitoli precedenti.

Nella tabella seguente sono riassunte le caratteristiche elettriche più importanti di un SiGe HBT impiegato nella tecnologia BiCMOS di STM.

PARAMETER	TYPICAL	UNITS
Gain	90	
BVCE0	3.3	V
VEARLY	>60	V
FT @ VCE=1.5V	45	GHz
FMAX @ VCE=1.5V	60	GHz
NFmin @ 2GHz	0.8	dB

Tabella I

CAPITOLO 5 – Conclusioni

L'ambito delle tecnologie elettroniche per i dischi rigidi è estremamente competitivo ed in continuo mutamento. Le richieste del mercato vanno sempre di più nella direzione di un aumento della capacità di immagazzinamento dei dati. Questo comporta che l'elettronica che gestisce le operazioni di lettura e scrittura su disco deve essere sempre più efficiente e veloce.

Al fine di ottenere circuiti elettronici in grado di gestire le sempre più elevate velocità delle operazioni di lettura/scrittura su disco, si seguono tipicamente due strade.

La prima consiste nello studio e nella progettazione di nuove architetture circuitali che consentano, dal lato del circuito di scrittura, di aumentare la velocità di scrittura dei dati riducendo i transitori, i tempi di salita e di discesa; dal lato del circuito di lettura, di aumentare la velocità con cui l'elettronica è in grado di prelevare i dati dal disco rigido e trasformarli in correnti e tensioni elettriche.

La seconda consiste nello studio di nuove tecnologie per la realizzazione dei dispositivi attivi di base che costituiscono l'elettronica degli *hard disk drivers*. In particolare, la progettazione del dispositivo attivo si basa principalmente sulla possibilità di impiegare nuovi materiali semiconduttori con caratteristiche fisiche tali da consentire un miglioramento delle prestazioni del dispositivo stesso, in termini di amplificazione del segnale e di velocità di commutazione. Un guadagno più elevato permette il pilotaggio di carichi capacitivi con *slew rate* decisamente più grandi (quindi, con tempi di salita e discesa più piccoli),

senza essere costretti ad aumentare le dimensioni del dispositivo, la qual cosa determinerebbe la presenza di capacità parassite maggiori e, dunque, una riduzione della velocità operativa del circuito. Una velocità di commutazione più elevata consente la riduzione dei transitori e, dunque, la possibilità di trasmettere e ricevere dati a velocità decisamente più elevate, conservandone il corretto contenuto (ovvero, senza apprezzabili distorsioni dei segnali in oggetto).

In questo lavoro di tesi, è stata presa in considerazione una particolare architettura del dispositivo che prevede l'impiego di una lega silicio-germanio, in luogo del solo silicio, nella realizzazione della regione di base dei transistori bipolari. Sono state analizzate le conseguenze sulla fisica del dispositivo derivanti dall'impiego di questo nuovo materiale semiconduttore e le tecniche che consentono la crescita di strati epitassiali di silicio-germanio, a partire da substrati di silicio puro. La necessità di impiegare strati epitassiali deriva dall'esigenza di utilizzare film SiGe ad elevata purezza, che conservino la struttura monocristallina del silicio sottostante, la qual cosa rappresenta una condizione necessaria per il funzionamento del dispositivo elettronico in oggetto.

Una volta realizzato in tal modo il transistore bipolare, sono state effettuate misure di parametri di *scattering* allo scopo di determinarne sperimentalmente il guadagno di corrente e la frequenza di guadagno unitario. Analoghe misure sono state realizzate su un dispositivo attivo "classico", ovvero interamente in silicio. Il confronto tra le misure sui due dispositivi, a parità di caratteristiche costruttive, ha dimostrato che l'impiego di una lega silicio-germanio nell'architettura del bipolare effettivamente ne migliora le prestazioni, sia in

bassa frequenza, in termini di guadagno di corrente, sia in alta frequenza, in termini di prodotto banda-guadagno.

Ulteriori miglioramenti nelle prestazioni del transistore bipolare, in particolare in alta frequenza, potranno essere senz'altro ottenuti impiegando carbonio nella regione di base in silicio-germanio. Gli atomi di carbonio, infatti, all'interno dello strato epitassiale di silicio-germanio, fungono da centri di aggregazione per i droganti in base, riducendone l'effetto di diffusione laterale nella base stessa. Questo fenomeno consente la riduzione della larghezza elettrica della base, a parità di condizioni operative. Tale caratteristica, come mostrato in questo lavoro di tesi, riduce il tempo di attraversamento in base degli elettroni liberi e, dunque, aumenta la velocità di commutazione del transistore bipolare.

Bibliografia

[1] J. Cressler et al., *Silicon-Germanium Heterojunction Bipolar Transistors*, Artech House

[2] F. Capasso, *Bandgap engineering: from physics and materials to new semiconductor devices*, Science, vol. 235, pp. 172-176, 1987

[3] J. W. Matthews et al., *Defects in epitaxial multilayers-I: misfit dislocations in layers*, J. Cryst. Growth, vol. 27, pp. 118-125, 1974

[4] R. E. Ham et al., *Microwave Measurements – The RF and Microwave Handbook*, CRC Press LLC, 2001

[5] C. Schelling, *Growth and characterization of self-organized and organized Si and $Si_{1-X}Ge_X$ nanostructures*, PhD thesis, Linz, 2000

[6] C. C. Meng et al., *RF Characteristics of BJT Devices with Selectively or Fully Ion-Implanted Collector*, 13th GAAS Symposium, Paris, 2005

[7] T. Swartz, *The development of an automated S-parameter measurement system*, EEAP 399

[8] H. C. Wu et al., *Extended Mextram Model To Wide Frequency Range*, Microwave Component Group, ECTM/DIMES

[9] P. Erratico et al., *Tecnologie RF nel silicio e le loro applicazioni al mondo delle comunicazioni wireless*, STMicroelectronics

[10] Agilent Technologies, *Agilent 8510C – Data Sheet*

[11] Hewlett-Packard, *S-Parameters Techniques for faster, more accurate network design*, http://www.hp.com/go/tmappnotes

[12] Agilent Technologies, *Understanding the Fundamental Principles of Vector Network Analysis*, Application Note

[13] G. L. Patton et al., *SiGe-base heterojunction bipolar transistors: physics and design issues*, IEEE, 1990

[14] L. E. Larson, *SiGe HBT BiCMOS Technology as an Enabler for Next Generation Communications Systems*, 12th GAAS Symposium – Amsterdam 2004

[15] D. Paul, *The Physics, Material and Devices of Silicon Germanium Technology*, Physics World

[16] M. D. Brunsman, *Characterization and Modeling Group CMOS Platform Device Development*, Motorola – Semiconductor Products Sector

[17] A. Lord, *Advanced RF Calibration Techniques*, Cascade Microtech, http://www.cascademicrotech.com

[18] D. Ballo, *VNA basics*, Hewlett-Packard

[19] T. Gaier et al., *On-Wafer Testing of Circuits Through 220GHz*, Jet Propulsion Laboratory, California Institute of Technology

[20] R. F. Scholz et al., *Advanced Technique for Broadband On-Wafer RF Device Characterization*, 63th ARFTG Conference Digest, 2004

[21] Università di Parma, *High Frequency Network Characterization*

[22] Agilent Technologies, *Network Analyzer Basics,* http://www.agilent.com/find/backtobasics

[23] J. Staudinger, *A two-tier method of de-embedding device scattering parameters using novel techniques,* Master Thesis, Arizona State University, 1987

[24] R. Lane, *De-Embedding Device Scattering Parameters*, Microwave J., August 1984

[25] J. Fitwpatrick, *Error Models For Systems Measurement*, Microwave J., May 1978

[26] Hewlett-Packard, Inc., *Operating and Programming Manual For the HP8510 Network Analyzer*, Santa Rosa, CA