

Università degli Studi di Torino  
Facoltà di Scienze Matematiche, Fisiche e Naturali  
Corso di Laurea in Fisica

TESI DI LAUREA

STUDIO E SVILUPPO DELLO STUDIO DI  
INGRESSO CON SISTEMA VLSI PER UN  
RIVELATORE FINALIZZATO A MISURE  
DOSIMETRICHE IN ADROTERAPIA

Relatore: Prof.ssa Maria Itala Ferrero

*Mi Ferrero*

Correlatore: Dott. Flavio Marchetto

*Flavio Marchetto*

Candidato: Walter Allasia

*Walter Allasia*

Anno Accademico 1994/1995

# Indice

Introduzione	I
<b>1 Radioterapia oncologica</b>	<b>1</b>
1.1 Introduzione	1
1.1.1 Basi biologiche della radioterapia	2
1.1.2 Parametri clinici	3
1.2 Radiazioni convenzionali	7
1.2.1 Fotoni	7
1.2.2 Elettroni	11
1.3 Adroterapia	12
1.3.1 Particelle cariche pesanti: protoni e ioni	13
1.3.2 Neutroni	19
1.4 Il progetto TERA	23
1.4.1 Specifiche dei fasci terapeutici dei protoni	24
1.4.2 Metodo di diffusione passiva	26
1.4.3 Metodo di scansione magnetica (raster scan)	27
1.4.4 Calcolo del flusso di protoni	28
<b>2 Misura di dose in adroterapia</b>	<b>30</b>
2.1 Introduzione	30
2.2 Misura della dose $D(x,y,z)$	31
2.2.1 Misura on-line della dose	33
2.2.2 Misura di microdosimetria: track structure	34
2.3 Calcolo dell'intensità dei protoni nei fasci di trattamento	36
2.4 Rivelatori	37
2.4.1 Rivelatori a gas	38
2.4.2 Il problema della ricombinazione	41

2.4.3	Fibre plastiche scintillanti . . . . .	42
<b>3</b>	<b>Specifiche del front-end</b>	<b>44</b>
3.1	Derivazione delle correnti . . . . .	44
3.1.1	Range dinamico . . . . .	47
3.1.2	Stima della capacità di ingresso . . . . .	47
3.2	Soluzioni possibili del front-end . . . . .	48
3.2.1	Integratore ciclico . . . . .	49
3.2.2	APM: Analog Pipeline Memory . . . . .	50
3.3	APC128 . . . . .	51
3.4	HARP . . . . .	53
3.5	Configurazione possibile dello stadio di ingresso . . . . .	58
3.5.1	Soluzione possibile per l'acquisizione dei dati . . . . .	59
<b>4</b>	<b>Introduzione ai sistemi VLSI</b>	<b>62</b>
4.1	Fabbricazione dei dispositivi attivi in processo MOS . . . . .	65
4.2	Il transistor MOS . . . . .	66
4.2.1	Modello per piccoli segnali . . . . .	72
4.3	Semplici applicazioni dei dispositivi MOS . . . . .	76
4.3.1	L'amplificatore a singolo stadio . . . . .	76
4.3.2	Risposta in frequenza di amplificatori a singolo stadio . . . . .	78
4.3.3	Resistore attivo . . . . .	82
4.3.4	Semplice generatore di corrente . . . . .	83
4.4	Disegno di circuiti integrati . . . . .	85
<b>5</b>	<b>Analisi e modifica del chip HARP</b>	<b>89</b>
5.1	Analisi del circuito . . . . .	93
5.1.1	L'amplificatore operazionale . . . . .	93
5.1.2	Descrizione generale del circuito OTA . . . . .	96
5.1.3	Simulazione dell'OTA . . . . .	100
5.2	Modifica dell'OTA . . . . .	103
5.2.1	Risposta in frequenza e stabilità degli amplificatori controreazionati . . . . .	109
5.2.2	Compensazione dell'OTA . . . . .	118
5.3	Simulazione di una cella di memoria . . . . .	127



## Introduzione

In questi ultimi anni stiamo assistendo ad una progressiva applicazione in campo medico della più avanzata tecnologia utilizzata in esperimenti di fisica nucleare. Basti pensare alla CT, PET e NMR <sup>1</sup>, che costituiscono ormai uno strumento indispensabile della diagnostica, o. per ciò che riguarda la terapia, alla ormai comune "radioterapia convenzionale", che utilizza fasci di fotoni ed elettroni per uccidere le cellule cancerogene. La fisica nucleare ci insegna però che i fasci di adroni, cioè particelle più pesanti (quali protoni, neutroni e ioni leggeri), sono più efficaci di quelli convenzionali per combattere neoplasie profonde o vicine ad organi critici.

La fondazione TERA <sup>2</sup> si è proposta di introdurre per prima in Italia l'adroterapia, cioè questa nuova tecnica di radioterapia che utilizza fasci di particelle pesanti. Un aspetto fondamentale di qualsiasi tecnica radioterapica è la conoscenza della distribuzione della dose di radiazione assorbita dal volume investito.

Il lavoro svolto in questa tesi riguarda lo studio dell'elettronica necessaria ad acquisire i segnali forniti dal rivelatore utilizzato per misurare la dose fornita al volume tumorale. In collaborazione con il C4I <sup>3</sup>, si è quindi iniziato a sviluppare un opportuno dispositivo integrato con sistema VLSI <sup>4</sup>, dato l'elevato numero di canali di acquisizione e le piccole correnti di segnale da rivelare.

Nel **Capitolo 1** vengono esaminate le proprietà dei fasci di radiazione convenzionali e le prospettive aperte dall'utilizzo dei fasci di adroni.

Nel **Capitolo 2** si introducono i concetti fondamentali di dosimetria ed i rivelatori adatti a misurare la dose irradiata.

Nel **Capitolo 3** si affronta lo studio delle possibili configurazioni dell'elettronica di "front-end", cioè dell'elettronica che si trova immediatamente dopo l'uscita del

---

<sup>1</sup>CT, Computerized Tomography; PET, Positron Emission Tomography; NMR, Nuclear Magnetic Resonance.

<sup>2</sup>TERA, Fondazione per l'Adroterapia Oncologica, si veda il primo capitolo.

<sup>3</sup>C4I, Centre de Competence en Conception de Circuits Integres.

<sup>4</sup>VLSI, Very Large Scale Integration.

rivelatore, e che deve dunque acquisirne i segnali.

Il **Capitolo 4** presenta, a grandi linee, i sistemi VLSI, in quanto la progettazione di qualsiasi dispositivo elettronico per il front-end del dosimetro verrà comunque (come spiegato nel Capitolo 3) eseguita con sistemi molto integrati.

Nel **Capitolo 5** viene presentato il preliminare lavoro di progettazione e sviluppo di una memoria analogica "pipeline", svolto in collaborazione con il C4I.

# Capitolo 1

## Radioterapia oncologica

### 1.1 Introduzione

Nella società industriale circa una persona su tre è destinata ad ammalarsi di cancro ed una su cinque morirà a causa di questo [1].

I metodi di cura utilizzati oggi dagli oncologi sono:

la chirurgia (rimozione del tessuto tumorale), la radioterapia (distruzione delle cellule cancerogene mediante radiazioni), la chemioterapia (farmaci), l'immunoterapia (potenziamento dei meccanismi di difesa peculiari dell'organismo) e l'ipertermia (trattamento mediante calore).

Dal rapporto *Verso la coordinazione della ricerca sul cancro in Europa* [2] si può vedere che al momento in Europa il 45% dei pazienti vengono guariti, definendo come guarigione un periodo di sopravvivenza asintomatico superiore ai 5 anni. Si può anche notare che il 90% dei pazienti guariti (cioè il 40% del totale) sono guariti grazie ad un controllo loco-regionale del tumore primario.

La chirurgia e la radioterapia da sole hanno successo rispettivamente nel 22% e nel 12% dei casi. Aggiungendo i trattamenti combinati queste percentuali aumentano di un ulteriore 6%, cosicché la radioterapia viene impiegata nella quasi metà dei trattamenti curativi di tipo loco-regionale.

La radioterapia si fonda sugli effetti biologici delle radiazioni ionizzanti, che sono in grado di uccidere le cellule per cessione di energia [1]. La situazione ideale è quella di depositare una grande quantità di energia nel volume tumorale senza danneggiare i tessuti sani circostanti. Ciò significa che, in linea di principio, si deve sia evitare qualsiasi diffusione laterale del fascio di radiazione, che ottenere una ben definita distribuzione dell'energia in profondità [3].

### 1.1.1 Basi biologiche della radioterapia

Per spiegare come agisce la radioterapia vengono spesso invocati i cosiddetti 5 R della radiobiologia [4]: *Radiosensibilità. Ricupero. Riossigenazione. Ripopolazione e Ridistribuzione* delle cellule nel ciclo cellulare.

Accanto a questi fattori vi sono fattori fisici come l'intensità della dose (cioè l'energia trasferita dalla radiazione al tessuto vivente), la qualità delle radiazioni (LET), la distribuzione della dose ed i fattori dose-tempo (numero, grandezza ed intervallo tra le frazioni di tempo e di trattamento globale).

Non vi sono in realtà differenze rilevanti tra le *Radiosensibilità* per l'uccisione di tessuti normali e tumorali. Analogamente, sono poche le differenze di velocità di *Riparazione*; è tuttavia possibile che, siccome molti tumori sono ipossici in vivo od anche temporaneamente anossici e poiché i meccanismi di ricupero sono dipendenti dall'ossigeno, il danno accumulato nei tessuti tumorali può essere maggiore del danno subito dai tessuti normali. La *Riossigenazione* è un fenomeno importante che si verifica in gran parte dei tumori ma non nei tessuti normali. È noto che con l'aumento in grandezza dei tumori si accresce il loro apporto ematico. La distanza di diffusione dell'ossigeno nei tessuti è di circa  $150\mu\text{m}$ , cosicché le cellule di diametro superiore nei capillari o in altri vasi sanguigni diventeranno anossiche dopo aver metabolizzato tutto l'ossigeno disponibile. Minute aree di cellule morte e morenti, "focolai ipossici", sono probabilmente presenti in gran parte dei tumori di diametro superiore a 2 mm. La percentuale di cellule ipossiche od anche anossiche può essere del 20% o più. Dopo irradiazione le cellule ossigenate (radiosensibili) saranno uccise preferenzialmente e la percentuale di cellule ipossiche può anche raggiungere il 100%. Tuttavia, quando le cellule uccise dalle radiazioni vengono eliminate mediante fagocitosi il tumore si riduce, ma i vasi sanguigni del tumore, vale a dire l'apporto di ossigeno, saranno relativamente poco interessati dall'irradiazione, e così le cellule tumorali residue diventeranno improvvisamente ben ossigenate e nuovamente radiosensibili a successive dosi.

La *Ricrescita* o ripopolazione dei tessuti può verificarsi nell'intervallo tra le dosi spopolanti di uno schema di frazionamento. Se l'entità di questa proliferazione cellulare fosse maggiore nei tessuti normali che nei tessuti tumorali, questi ultimi regredirebbero gradualmente ed il tessuto normale potrebbe mantenere la sua normale cellularità.

Infine esiste la possibilità di una *Ridistribuzione* delle cellule del ciclo cellulare, durante il quale la radiosensibilità cambia: le dosi frazionate uccidono preferenzialmente le cellule nelle fasi più sensibili. Una nuova selezione differenziale delle cellule tumorali in rapporto alle cellule normali potrebbe avere l'effetto di spopolare preferenzialmente il tumore.

### 1.1.2 Parametri clinici

Per valutare gli effetti reali o potenziali delle radiazioni ionizzanti si sono introdotte delle quantità misurabili fisicamente, che prendono il nome di quantità dosimetriche [5]. La quantità dosimetrica fondamentale è la dose assorbita,  $D$ , definita come

$$D = \lim_{m \rightarrow 0} \frac{\bar{\epsilon}}{m} \quad (1.1)$$

dove  $\bar{\epsilon}$  è il valore medio dell'energia depositata dalla radiazione ionizzante alla materia di massa  $m$ .

Il nome per l'unità di dose assorbita è il *Gray* ( $1 \text{ Gy} = 1 \text{ J}\cdot\text{kg}^{-1}$ ).

Nella figura 1.1 è riportata una tipica curva di sopravvivenza cellulare in funzione della dose assorbita [1]. Un'altra quantità dosimetrica importante è il **LET**, cioè il "trasferimento lineare di energia" ( $LET = \text{Linear Energy Transfer}$ ) ad un mezzo da parte di una particella carica, definito come il rapporto:

$$LET = \frac{dE}{dl} \quad (1.2)$$

dove  $dE$  è l'energia media persa dalla particella, dovuta alle collisioni con gli elettroni, attraversando una distanza  $dl$ .

L'unità di misura sarà  $\text{J}\cdot\text{m}^{-1}$  o  $\text{eV}\cdot\text{m}^{-1}$ .

Idealmente la distribuzione della dose è conosciuta in tutti i punti del volume di trattamento e nei tessuti circostanti; la sua misura è chiamata dosimetria e gli strumenti utilizzati dosimetri.

La ricerca di una accurata dosimetria è di fondamentale importanza per poter controllare la dose somministrata al tumore e ai tessuti critici circostanti, e per dare la possibilità di scambiare informazioni cliniche tra i vari istituti di terapia.

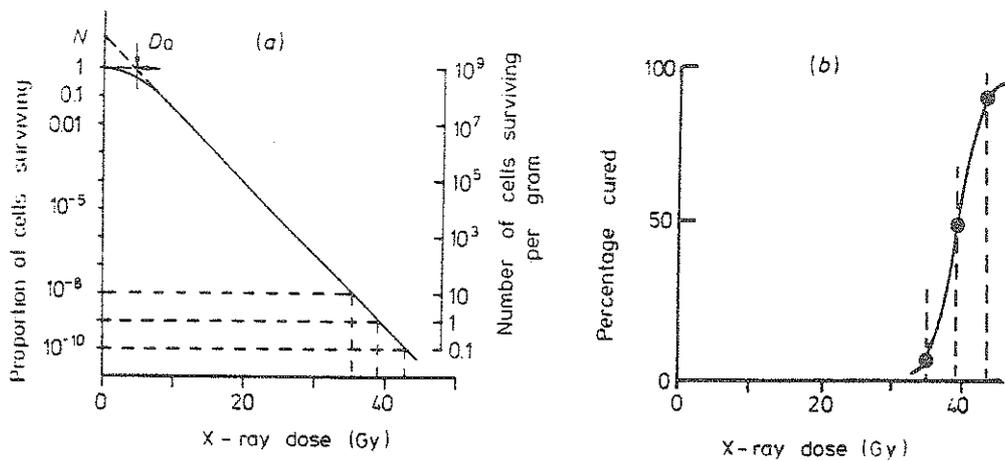


Figura 1.1: (a) Tipica curva di sopravvivenza cellulare, in cui si ha la percentuale di cellule sopravvissute (scala logaritmica verticale) in funzione della dose di radiazione a cui sono esposte. La scala di destra riporta il numero di cellule sopravvissute per grammo di tessuto o tumore, assumendo  $10^9$  cellule per grammo prima della ionizzazione (da Henk e Fowler 1976). (b) Percentuale di tumori curati, corrispondenti al numero di cellule maligne sopravvissute a una dose di radiazione X. Si è assunto che una cellula sopravvissuta fornisce il 50% di probabilità di cura [4].

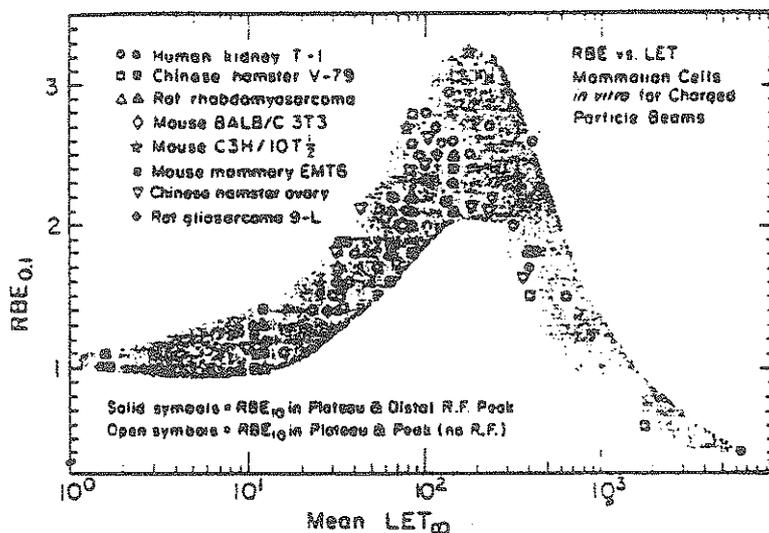


Figura 1.2: Grafico dell'efficacia biologica relativa (RBE) in funzione del LET. Non esiste una semplice relazione tra i due, sebbene il LET sia comunemente utilizzato per descrivere le differenze del danno da radiazione da parte di vari ioni leggeri. I dati provengono da un numero di esperimenti utilizzando diversi tipi di ioni, energie e tipi cellulari. L'area ombreggiata mostra l'andamento generale dei dati.

Sperimentalmente si dimostra che la radiazione ad alto LET è biologicamente più efficace della radiazione a basso LET. L'efficacia biologica delle radiazioni possono poi essere diverse rispetto a quella dei raggi X (RBE). Se una dose  $D$  di un dato tipo di radiazione produce uno specifico effetto biologico, all'ora l'RBE (Relative Biological Effectiveness) è definito come il rapporto:

$$RBE = \frac{D_X}{D} \quad (1.3)$$

dove  $D_X$  è la dose di raggi X necessaria a produrre, nelle stesse condizioni, lo stesso effetto biologico [7]. L'RBE varia generalmente in funzione dei diversi tipi di cellule, dei livelli di sopravvivenza, e delle diverse energie o tipi di radiazione. La figura 1.2 riporta i risultati degli esperimenti biologici eseguiti al *Lawrence Berkeley Laboratory* con un fascio di particelle cariche pesanti [6]. L'ossigeno disciolto nei tessuti si comporta da agente radiosensibilizzante. Questo così detto

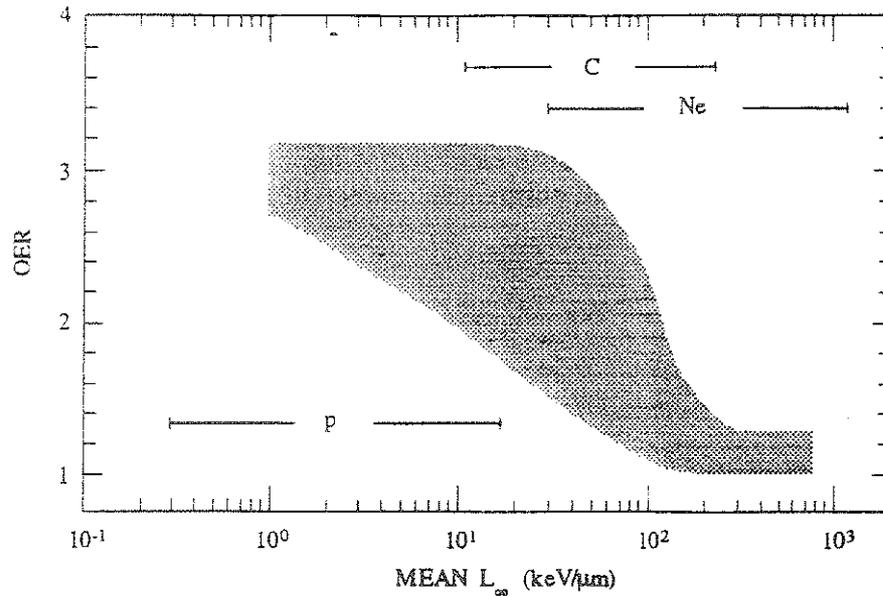


Figura 1.3: Riassunto dei risultati sperimentali degli Oxygen Enhancement Ratio misurati in funzione di  $L_{\infty}$  (LET). L'area ombreggiata ricopre i valori di OER misurati per vari tipi di radiazioni e linee cellulari. Sono inoltre indicati gli intervalli di LET ottenibili con protoni, ioni carbonio e neon a energie di interesse radiobiologico [3].

*effetto ossigeno*, che è quasi sempre osservato in radiobiologia, è illustrato in figura 1.3 e può essere espresso quantitativamente dal rapporto di potenziamento dell'ossigeno, l'**OER** (Oxygen Enhancement Ratio) [7]. L'OER è definito [9] come il rapporto tra la dose richiesta in condizioni ipossiche e quella in condizioni di normale ossigenazione, per produrre lo stesso livello di effetto biologico. Dato che le cellule tumorali si dividono rapidamente, esse hanno uno scarso rifornimento di sangue e ossigeno e quindi l'*effetto ossigeno* ha importanti implicazioni in radioterapia. La maggior parte del potenziamento apportato dall'ossigeno è ottenuto a concentrazioni considerabilmente al di sotto della saturazione. Alcuni successi sono stati raggiunti aumentando il livello di ossigeno disciolto nelle persone prima di sottoporle a radioterapia. In questo modo, sia i tessuti normali che il tumore, aumentano il loro livello di  $O_2$ , il potenziamento relativo dell'efficacia della radiazione nel tumore è più grande, a causa del suo basso livello iniziale [7].

## 1.2 Radiazioni convenzionali

La radiazione utilizzata nella cura del cancro può provenire sia da fonti esterne al corpo del paziente, sia da sorgenti radioattive immesse nel paziente stesso .

Le prime radiazioni esterne utilizzate in terapia sono stati i raggi X, scoperti nel 1895 (esattamente cento anni fa) da Roentgen in Wuerzburg, Germania. Fino al 1970 molti ospedali hanno usato raggi gamma prodotti da isotopi radioattivi, come il  $^{60}\text{Co}$  , o raggi X prodotti da betatroni (acceleratori circolari) [6]. Ora, i moderni ospedali usano soprattutto fotoni prodotti da elettroni di alta energia, accelerati da acceleratori lineari e fatti collidere su bersaglio fisso (in questo modo il processo di perdita di energia degli elettroni si traduce in emissione di fotoni, fenomeno noto col nome di *bremstrahlung*).

Inoltre gli elettroni di alta energia sono essi stessi direttamente utilizzati per trattare i pazienti.

Dato che sia i fotoni che gli elettroni sono stati ampiamente utilizzati in radioterapia, sono chiamati *radiazioni convenzionali*.

Nei capitoli successivi si esamina l'interazione di queste radiazioni con la materia.

### 1.2.1 Fotoni

I fotoni, o quanti di luce, sono elettricamente neutri e la distanza percorsa nella materia è governata statisticamente da una probabilità di interazione per unità di cammino percorso che dipende dallo specifico mezzo attraversato e dall'energia dei fotoni incidenti.

Quando un fotone interagisce può essere assorbito e scomparire o può essere diffuso e cambiare la sua direzione, con o senza perdita di energia [7].

I principali meccanismi di deposizione di energia nella materia da parte dei fotoni sono l'*effetto fotoelettrico*, l'*effetto Compton*, la *produzione di coppie*. Un fascio di fotoni che attraversa uno spessore di materia viene attenuato nell'intensità in modo esponenziale [10]:

$$-\frac{dI}{I} = \mu dx \quad (1.4)$$

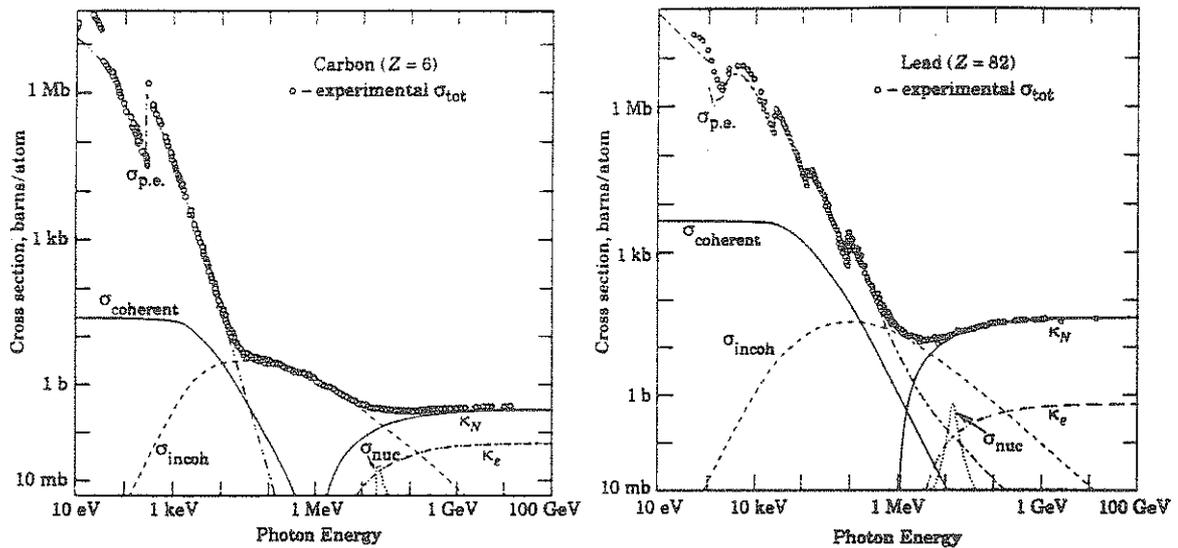


Figura 1.4: Sezione d'urto totale dei fotoni in funzione dell'energia in Carbonio e Piombo; sono inoltre riportati i contributi dei differenti processi:  $\sigma_{p.e.}$  = effetto fotoelettrico;  $\sigma_{coherent}$  = diffusione coerente (Rayleigh scattering);  $\sigma_{incoherent}$  = diffusione incoerente (Compton scattering);  $\kappa_n$  = produzione di coppie nel campo nucleare;  $\kappa_e$  = produzione di coppie nel campo elettronico;  $\sigma_{nuc}$  = assorbimento fotonucleare (solitamente seguito dalla emissione di un neutrone o altre particelle) [8].

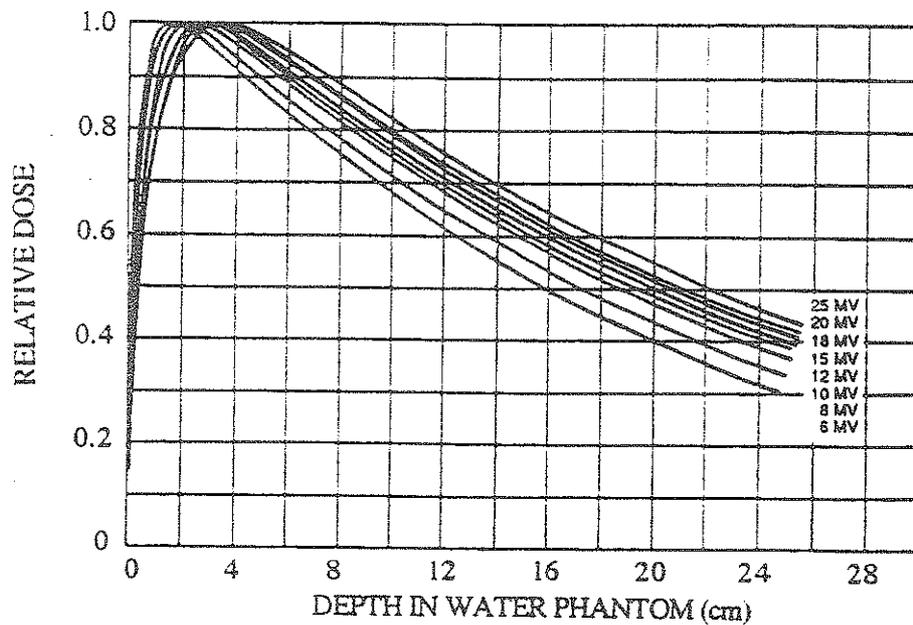


Figura 1.5: Curve dose-profondità in acqua per fasci di fotoni aventi energie massime nell'intervallo tra 6 e 25 Mev [3].

in cui  $I$  è l'intensità della radiazione primaria,  $\mu$  il coefficiente di assorbimento (direttamente legato alla sezione d'urto totale), e  $dx$  lo spessore attraversato. La probabilità totale che un fotone interagisca nella materia è la somma delle sezioni d'urto dei tre processi:

$$\sigma_{tot} = \sigma_{foto} + Z\sigma_{Compton} + \sigma_{coppie} \quad (1.5)$$

dove si sono trascurati gli altri tipi di interazione e si è moltiplicato per il numero atomico la sezione d'urto per effetto Compton poiché proporzionale al numero di elettroni.

Nella figura 1.4 sono riportati gli andamenti delle sezioni d'urto per i vari processi, mentre nella figura 1.5 è presentato il profilo di dose nell'intervallo energetico di interesse radioterapico. Come si può notare i fasci di fotoni sono caratterizzati da un massimo di assorbimento (raggiunto a 3.5 cm per energia massima di 25

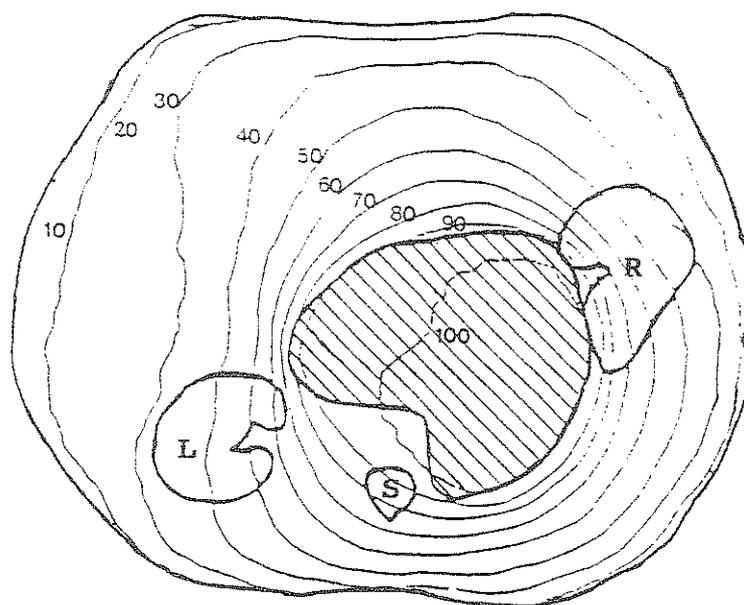


Figura 1.6: Curve isodosi per l'irradiazione di un liposarcoma retroperitoneale adiacente al midollo spinale: irradiazione dinamica (5 campi) con fotoni di 10 MeV di energia massima. Sono riportati il rene destro (R) e sinistro (L) e la colonna vertebrale (S). La conformazione della dose è ottenuta con una rotazione della testata isocentrica attorno ad un solo asse e con un collimatore a apertura variabile (*multi-leaf collimator*) [3].

MeV) corrispondente al percorso massimo degli elettroni secondari prodotti dai fotoni primari negli strati più superficiali del tessuto irraggiato.

Nella figura 1.6 sono riportate le *curve di isodose* bidimensionali calcolate per un'irradiazione tipica di un liposarcoma retroperitoneale. Queste curve mostrano chiaramente come nella radioterapia convenzionale si ottenga un livello di *irradiazione conformazionale* (adattato cioè alla forma del tumore) buono ma non ottimale. Questo costituisce un fattore limitante nel trattamento di molti tipi di tumori, specialmente in prossimità di organi critici, che sono particolarmente soggetti ad essere danneggiati dalla radiazione [3].

## 1.2.2 Elettroni

La perdita di energia degli elettroni avviene fondamentalmente per ionizzazione (in quanto particelle cariche), e per radiazione elettromagnetica (*bremstrahlung*) emessa nelle violente decelerazioni che si verificano durante gli urti (in quanto gli  $e^-$  sono particelle leggerissime rispetto ai nuclei).

La perdita di energia per ionizzazione per elettroni non relativistici è [10]:

$$\left(-\frac{dE}{dx}\right)_{ion} = \frac{4\pi e^4 N}{m_e v^2} \cdot \left(\log \frac{m_e v^2}{2I} - \frac{1}{2} \cdot \log 2 + \frac{1}{2}\right) \quad (1.6)$$

dove:

$m_e$  è la massa dell'elettrone;

$v$  è la velocità dell'elettrone incidente;

$N = N_A \cdot \frac{Z}{A} \cdot \rho$  è il numero di elettroni per centimetro cubo del materiale assorbitore;

$I$  è il potenziale di ionizzazione medio dell'assorbitore.

La *bremstrahlung* si verifica quando un elettrone è deflesso nel campo elettrico di un nucleo (e in maniera meno intensa, nel campo di un elettrone atomico). Ad alte energie, la radiazione è emessa maggiormente nella direzione di arrivo degli elettroni [7]. Quantitativamente si ha [11]:

$$\left(-\frac{dE}{dx}\right)_{brem} = 4N_A \alpha \frac{Z^2}{A} r_0^2 E \log \frac{183}{Z^{1/3}} = \frac{E}{X_0} \quad (1.7)$$

con

$N_A$  numero di Avogadro;

$\alpha = 1/137$  costante di struttura fine;

$r_0$  raggio classico dell'elettrone;

$Z$  numero atomico del materiale assorbitore;

$E$  è l'energia dell'elettrone incidente.

L'ultimo membro definisce la grandezza  $X_0$  detta lunghezza di radiazione di una sostanza che può essere interpretata come lo spessore di materia che riduce (per radiazione di frenamento) l'energia di un elettrone di un fattore  $e$  in media [11].

La figura 1.7 mostra il confronto tra *bremstrahlung* e ionizzazione in piombo. Si può notare come la prima prevalga per alte energie mentre la seconda per basse energie.

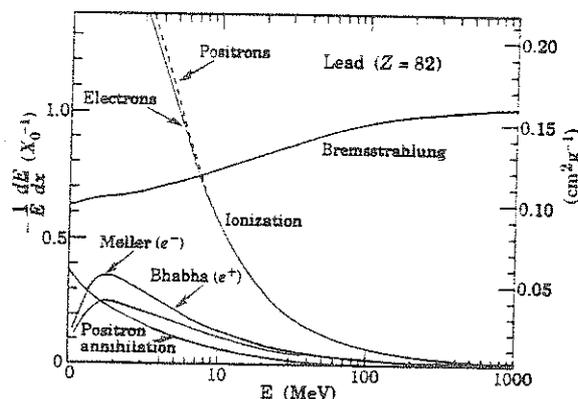


Figura 1.7: Frazione dell'energia persa per lunghezza di radiazione in funzione dell'energia di elettroni e positroni (particelle  $\beta$ ). Si è assunto  $X_0(\text{Pb}) = 6.4 \text{ g/cm}^2$  [8].

Nella figura 1.8 si può vedere che i fasci di elettroni sono caratterizzati da un percorso massimo nel tessuto che dipende dalla energia iniziale del fascio. La coda di bassa intensità è dovuta ai fotoni di bremsstrahlung [3]. Il percorso massimo, espresso in centimetri, è circa uguale a metà dell'energia iniziale del fascio, espressa in MeV. Per queste caratteristiche i fasci di elettroni sono adatti per il trattamento di focolai superficiali o semiprofondi (qualche centimetro dalla superficie cutanea).

### 1.3 Adroterapia

Nella fisica moderna, la parola adroni è usata per tutte le particelle soggette a interazioni forti, che sono mesoni e barioni [10].

Oggi l'adroterapia utilizza fasci di neutroni o sorgenti schermate di neutroni, fasci di protoni e ioni leggeri, fasci di pioni negativi, o frammenti nucleari originatisi in atomi trasportati da veicoli molecolari che ricercano il tumore, come il radionuclide emettitore  $\alpha$ ,  $^{211}\text{At}$ , o il nuclide stabile  $^{10}\text{B}$  che fornisce particelle molto ionizzanti dopo l'attivazione con neutroni lenti [12]. Queste modalità apportano potenzialità uniche, grazie al fatto che, come si può vedere nella figura 1.9 le caratteristiche fisiche dell'interazione con la materia dei vari adroni cambiano

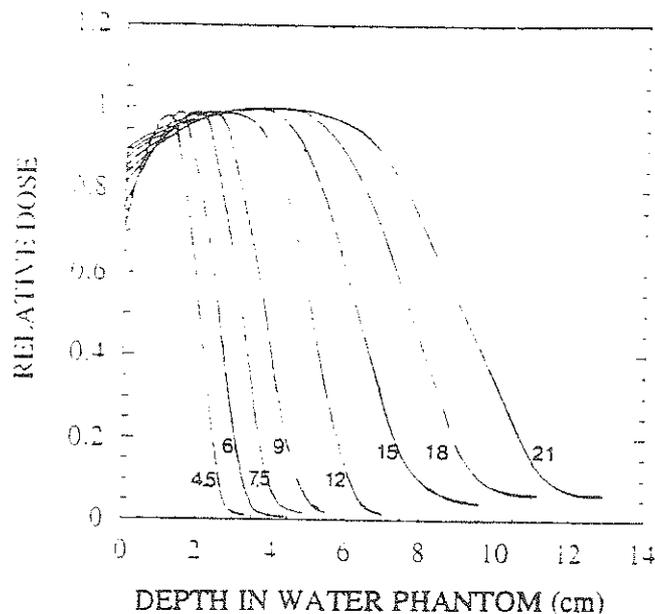


Figura 1.8: Curve dose profondità in acqua per fasci di elettroni nell'intervallo energetico tra 4.5 e 21 MeV [3].

marcatamente e favorevolmente rispetto a quelle delle *radiazioni convenzionali*. Queste infatti possiedono una distribuzione di dose che ha un andamento esponenziale decrescente, rilasciando quindi dosi più alte nei tessuti in cui entra il fascio e più basse nel volume in cui si situa il tumore.

### 1.3.1 Particelle cariche pesanti: protoni e ioni

Le particelle cariche pesanti sono superiori a tutti gli altri tipi di radiazione nella terapia dei tumori [13]. Comparata alla radiazione elettromagnetica i fasci di ioni pesanti presentano una piccola diffusione laterale e un percorso (*range*) finito e modulabile. Il loro profilo dose-profondità è invertito: questo permette di incrementare la deposizione di energia con la profondità di penetrazione. Questi vantaggi sono già stati verificati con i fasci di protoni e sono la base del grande successo della terapia con protoni per la cura del cancro [14].

Quando una particella carica pesante penetra nella materia, perde la sua energia cinetica principalmente per ionizzazione del mezzo. L'energia persa per unità di percorso nel mezzo assorbitore (o energia persa specifica), aumenta col diminuire

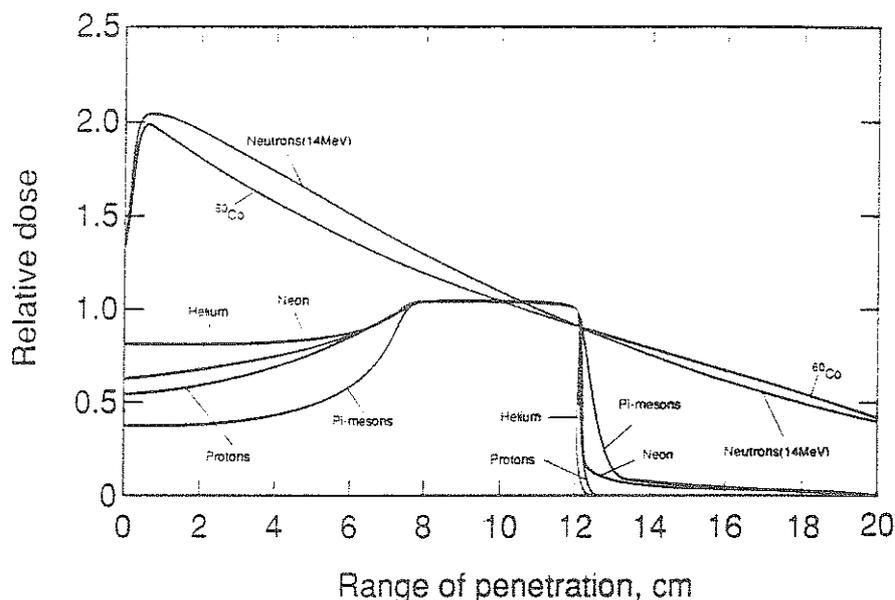


Figura 1.9: Curve dose-profondità per fasci di neutroni, pioni, protoni e ioni di neon comparati con lo "standard d'oro", un fascio di fotoni prodotti da  $^{60}\text{Co}$  [12].

della velocità della particella, dando luogo ad un massimo di ionizzazione vicino alla fine del range, conosciuto come picco di Bragg [6].

La perdita lineare di energia per ionizzazione è data dalla famosa formula di Bethe e Bloch [10] (calcolo già iniziato nel 1913 da Bohr [7]):

$$\left(-\frac{dE}{dx}\right) = \frac{4\pi e^4 N}{m_e c^2} \cdot \frac{z^2}{\beta^2} \cdot \left(\log \frac{2m_e c^2 \beta^2 \gamma^2}{I} - \beta^2\right) \quad (1.8)$$

dove:

$z$  è la carica della particella incidente in multipli della carica dell'elettrone;

$\beta = v/c$  della particella incidente, e gli altri parametri sono già stati definiti.

La formula di Bethe e Bloch riproduce bene i dati sperimentali solo per valori di  $\beta$  maggiori di 0.01. Per velocità inferiori si hanno fenomeni di cattura elettronica da parte della particella incidente, che in tal modo si trova ad avere una carica efficace minore di quella iniziale; quindi la perdita di energia diminuisce col diminuire della velocità.

Come precedentemente accennato, quando un fascio monoenergetico di particelle cariche pesanti entra nel corpo del paziente, la distribuzione dose-profondità è

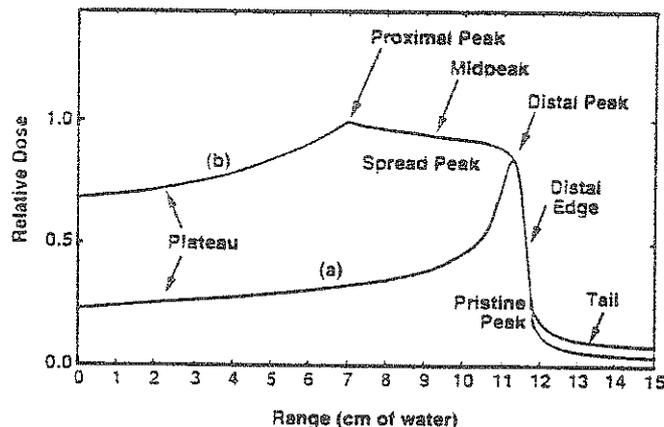


Figura 1.10: Sono mostrate le distribuzioni di dose in funzione della profondità di penetrazione per (a) un fascio primario e (b) per un fascio la cui energia è modulata in modo da allargare il range [6].

caratterizzata da una dose relativamente bassa nella regione di entrata (plateau) vicino alla pelle, e una sottile elevata dose alla fine del range (picco di Bragg). Il range può essere modificato, in modo che il picco di Bragg si verifichi esattamente nel volume in cui si trova il tumore da trattare [6].

Modulando l'energia delle particelle incidenti, si può allargare il picco di Bragg (SOBP, dall'inglese *Spread Out Bragg Peak*), come mostrato in figura 1.10, il che permette di trattare una regione più vasta, ma il rapporto picco-plateau diminuisce [15]; in ogni caso la dose biologicamente efficace nella regione di allargamento del picco può ancora essere più grande della dose di entrata. Si può ottenere una dose biologicamente uniforme, nella regione di allargamento del picco di Bragg, compensando le variazioni nell'efficacia biologica relativa (RBE) della radiazione in funzione della profondità di penetrazione, ottenendo così un "allargamento del picco" (Spread Peak in figura 1.10) pressoché piatto relativo ad una regione di isosopravvivenza cellulare.

Nella figura 1.11 sono disegnate le curve di isodose relative ai protoni, per l'irradiazione dello stesso tumore riportato nella figura 1.6. Si può notare come i protoni siano in grado di danneggiare in maniera estremamente ridotta gli organi critici circostanti (midollo spinale e reni) [3].

Nella curva dose-profondità (si veda figura 1.10) per gli ioni leggeri oltre il picco di Bragg allargato si presenta una coda (Tail) che aumenta con il numero atomico delle particelle utilizzate. Essa è dovuta alla frammentazione dei nuclei incidenti.

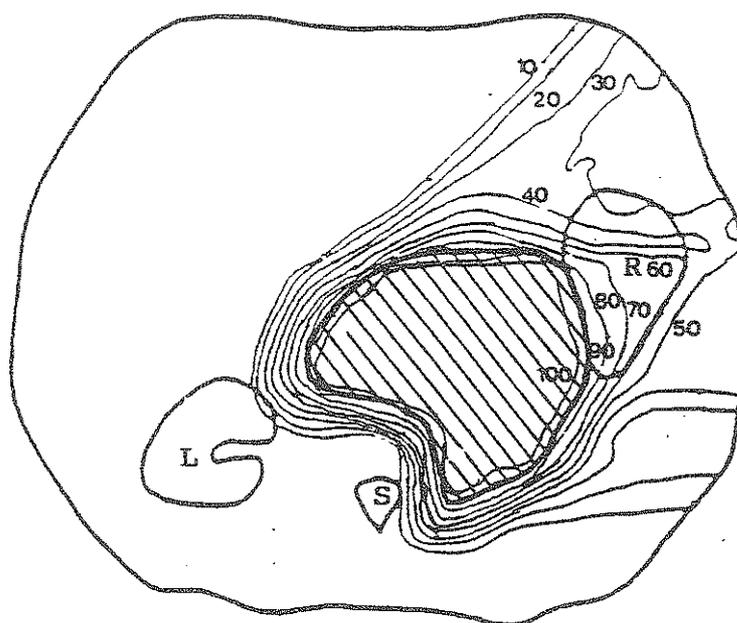


Figura 1.11: Curve di isodose per il trattamento di un liposarcoma retroperitoneale adiacente al midollo spinale: trattamento con due campi di protoni con la tecnica a scansione a punti. Sono riportati il rene destro (R) e sinistro (L) e la colonna vertebrale (S). La zona ombreggiata rappresenta il volume del bersaglio [3]. Si confronti con la figura 1.6.

I frammenti più leggeri hanno un percorso nella materia maggiore rispetto a quello degli ioni progenitori e danno quindi luogo ad un aumento della dose assorbita oltre il picco.

L'incremento percentuale della dose in questa regione dipende dalla massa dello ione: è dell'ordine del 15% per ioni come il carbonio e l'ossigeno, mentre può raggiungere il 30% per ioni di neon. Per questa ragione non è giustificato l'uso di ioni più pesanti dell'ossigeno per una terapia veramente conformazionale [3]. In figura 1.12 sono riportati i risultati di un confronto di sopravvivenza cellulare, misurati con una linea cellulare in vitro di tumore umano, in funzione della profondità di penetrazione del fascio, per protoni e ioni carbonio di 3 cm di allargamento del picco di Bragg [16]. Per uno stesso effetto biologico nella regione di allargamento del picco, i protoni uccidono un numero superiore di cellule nella regione di plateau. Il carbonio viceversa, presenta una dose di uscita superiore a quella dei protoni dovuta alla frammentazione.

Come si era mostrato a proposito dei parametri clinici della radioterapia, le radiazioni ad alto LET (e tali sono gli ioni) aumentano l'efficacia biologica e sono quindi da preferirsi. Per i protoni, comunque, la regione di alto LET nel picco di Bragg è troppo piccola per essere di rilevanza clinica; per ioni più pesanti del neon, invece, si sono trovati valori significativamente elevati di RBE già nella zona di ingresso del fascio e poi saturati nella zona del massimo di Bragg [13].

La ragione dell'incremento dell'efficacia biologica può essere compresa qualitativamente dal confronto della densità locale di ionizzazione prodotta dagli elettroni (raggi  $\delta$ ) che si liberano attorno alla traccia della particella. Nella figura 1.13 [17] sono confrontate le strutture di traccia, calcolate per protoni e ioni carbonio, con le dimensioni della molecola di DNA, bersaglio critico all'interno del nucleo cellulare. Ovviamente, un numero più grande di rotture della doppia treccia elicoidale del DNA (*double-strand break*), può essere prodotto al passaggio di un solo ione carbonio rispetto ai protoni o ad altre radiazioni di basso LET.

Per le radiazioni di basso LET, gli elettroni agiscono maggiormente in maniera indipendente, producendo lesioni isolate di DNA, come la rottura di una sola treccia (*single-strand break*). L'utilizzo di ioni pesanti aumenta la probabilità di produrre rotture della doppia treccia, per l'alta densità locale di elettroni nella traccia di particelle pesanti. Questo fornisce un aumento dell'RBE fino a valori di 3-5 nel massimo di Bragg degli ioni leggeri. Comunque, nel caso degli ioni pesanti un ulteriore aumento del danno locale non può fornire un migliore effetto

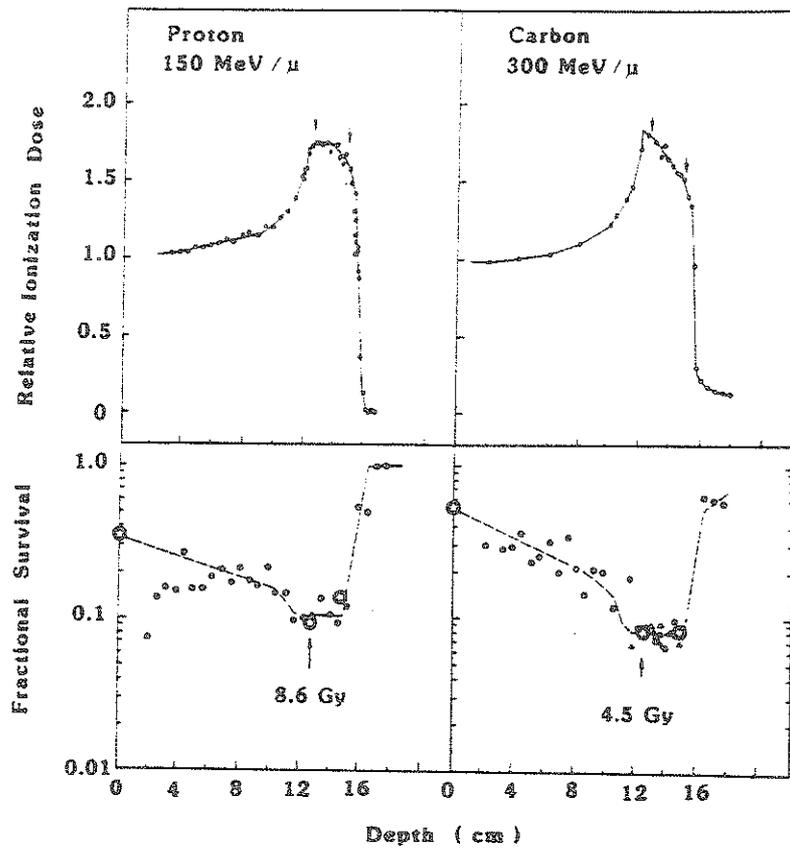


Figura 1.12: Misure di sopravvivenza in funzione della profondità di un tumore umano  $T_1$  per un massimo di dose del picco allargato di Bragg (proximal peak) di 8.6 Gy (protoni) e 4.5 Gy (Carbonio). Sono inoltre riportate le rispettive curve di Bragg. Le stelle sul grafico riportano il valore (dati non mostrati) dei punti di sopravvivenza nel *proximal peak* e nel *distal peak* [16].

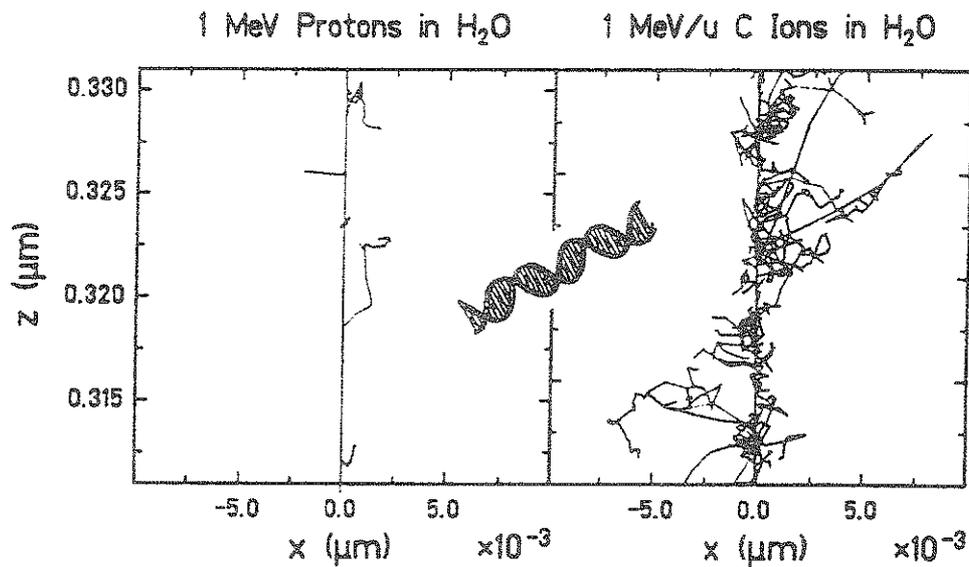


Figura 1.13: Confronto della struttura di traccia di protoni e ioni carbonio, entrambe di 1 MeV/u. con le dimensioni di una molecola di DNA, rappresentata schematicamente. Gli ioni carbonio hanno una probabilità molto più alta di produrre una rottura simultanea delle due opposte trecce [17].

biologico, per gli effetti di saturazione fisica, chimica e biologica (*overkill*) [13].

Nella figura 1.14 è riportato il numero di pazienti trattati con particelle cariche pesanti attualmente nel mondo, mentre nella figura 1.15 sono riportati i progetti futuri dell'adroterapia.

### 1.3.2 Neutroni

L'interazione dei neutroni con la materia, essendo particelle prive di carica elettrica, avviene tramite la sola forza nucleare forte, che ha un range dell'ordine di  $10^{-13}$  cm. L'interazione è quindi piuttosto rara ed i neutroni sono considerati particelle molto penetranti. Dalla figura 1.9 si può notare che un fascio di neutroni monoenergetici viene attenuato in modo esponenziale, mostrando una distribuzione dose-profondità simile a quello dei fotoni. Per la terapia vengono utilizzati neutroni veloci: neutroni cioè, che possiedono una energia cinetica compresa tra 1 MeV e 150 MeV. In questo intervallo i processi di assorbimento sono accom-

PLACE	COUNTRY	TYPE	PERIOD	PATIENTS	DATE
Berkeley 184	CA U.S.A	p	1954-1957	30	
Berkeley	CA U.S.A	He	1957-1992	2054	Jun91
Uppsala	Sweden	p	1957-1976	73	
Harvard	MA U.S.A	p	1961-	6138	June94
Dubna	Russia	p	1967-1974	84	
Moscow	Russia	p	1969-	2550	Oct92
Los Alamos	NM U.S.A.	pion	1974-1982	230	
St Petersburg	Russia	p	1975-	891	Jun94
Berkeley	CA U.S.A	ion	1975-1992	433	Jun91
Chiba	Japan	p	1979-	86	Jun93
TRIUMF	Canada	pion	1979-	314	Jun93
PSI (SIN)	Switzerland	pion	1980-1993	503	Jun93
PMRC, Tsukuba	Japan	p	1983-	393	Mar94
PSI (SIN)	Switzerland	p	1984-	1574	Dec93
Dubna	Russia	p	1987-	31	Apr94
Uppsala	Sweden	p	1989-	34	May93
Clatterbridge	England	p	1989-	513	May94
Loma Linda	CA U.S.A	p	1990-	805	Feb94
Louvain-la-Neuve	Belgium	p	1991-	21	Nov93
Nice	France	p	1991-	338	Dec93
Orsay	France	p	1991-	402	May94
Faure	South Africa	p	1991-	6	Nov93
Bloomington	IN U.S.A	p	1993-	1	Dec93
U C Davis	CA U.S.A	p	1994-	4	Jun94
				1047	pion
				2487	ion
				13974	p
			TOTAL	17508	all

Figura 1.14: Numero di pazienti totale trattato da particelle cariche. (luglio 1994) [18].

PLACE	COUNTRY	TYPE	START	COMMENTS
PSI	Switzerland	p	1994	200 MeV, gantry, dedicated line
HIMAC, Chiba	Japan	ion	1994	first ion beam December 1993
TRIUMF	Canada	p	1994	therapy with existing beams
Berlin	Germany	p	1995	72 MeV cyclotron, eye treatments
Munich	Germany	p	1995?	64 MeV, eye treatments
GSI Darmstadt	Germany	ion	1996	new cave for treatment
ITEP Moscow	Russia	p	1996	2 gantries, H <sup>-</sup>
Jülich (KFA)	Germany	p	1997	COSY storage ring
KVI Groningen	Netherlands	p	1997?	200 MeV, gantry
NPTC (Harvard)	MA U S A	p	1998	new facility to be built
NC Star	NC U S A	p	1999?	70-300 MeV synchrotron, gantry
TERA	Italy	ion	2000?	60-250 MeV protons, H <sup>-</sup>
Novosibirsk	Russia	p	?	180-200 MeV linear accelerator
Proton Dev N A	IL USA	p	?	250 MeV, private facility
Clatterbridge	England	p	?	upgrade linear accelerator
Tsukuba	Japan	p	?	230 MeV
Krakow	Poland	p	?	60 MeV

Figura 1.15: Programmi futuri per il trattamento con particelle cariche, (luglio 1994) [18].

pagnati da emissione di nucleoni quali protoni, neutroni e particelle  $\alpha$  [11]. Si possono quindi considerare radiazione ad altro LET dal momento che le particelle emesse per interazione hanno un alto potere ionizzante. Questo fa sì che l'OER sia più vicino all'unità di quello dei fasci di fotoni: inoltre l'RBE dei neutroni è maggiore per i tumori radioresistenti, e i meccanismi di riparazione cellulare sono meno efficaci [3]. Purtroppo l'andamento esponenziale della distribuzione di dose e la difficoltà di collimare i fasci, limita le possibilità di eseguire una terapia conformazionale.

Più promettente sotto questo aspetto risulta la Terapia per Cattura Neutronica del Boro, la BNCT (Boron Neutron Capture Therapy). Questa terapia utilizza la reazione che si verifica quando un neutrone termico interagisce con un nucleo di  $^{10}\text{B}$  [19]. I neutroni termici possiedono una energia cinetica inferiore o uguale ad 1 eV e interagiscono con la materia in generale con fenomeni di cattura, in cui il neutrone viene assorbito da un nucleo che poi si disintegra emettendo radiazioni [11]. La sezione d'urto per assorbimento di neutroni da parte del boro è molto più elevata di quella dei tessuti umani e quindi si può indurre l'assorbimento portando il boro, con appositi veicoli molecolari che si attaccano alle cellule tumorali, nel volume bersaglio e successivamente irradiare il paziente con neutroni termici. La cattura del neutrone termico da parte del  $^{10}\text{B}$  porta alla formazione di un nucleo instabile ( $^{11}\text{B}$ ) che si disintegra spontaneamente in due particelle:  $^4\text{He}$  ( $\alpha$ ) e  $^7\text{Li}$  [19].

Le due particelle trasferiscono tutta la loro energia alle cellule tumorali dato che il loro range è più corto del raggio cellulare (alcuni  $\mu\text{m}$  contro alcune decine di  $\mu\text{m}$ ) [20]. L'efficacia di questo tipo di trattamento, la cui concezione risale al 1936, dipende ovviamente dalla capacità di creare composti capaci di trasportare una quantità sufficiente di boro nei nuclei delle cellule neoplastiche. In futuro, la BNCT sarà applicabile, date le sue caratteristiche, al trattamento di tumori di estensione limitata che non possono essere trattati efficacemente da altri tipi di terapia [19].

## 1.4 Il progetto TERA

Il progetto TERA, dove TERA sta per TERapia Adronica, si propone di portare in Italia le più moderne tecniche di radioterapia che utilizzano i fasci di **protoni**, **neutroni** e **ioni leggeri** fino all'**ossigeno** [3].

Nasce nel 1991 nell'ambito dell'Istituto Nazionale di Fisica Nucleare (INFN) dalla collaborazione tra Ugo Amaldi (Università di Milano e CERN) e Giampiero Tosi (Istituto Europeo di Oncologia).

È stato calcolato che ogni anno a circa 4500 italiani sono diagnosticati tumori o lesioni che trarrebbero vantaggio da un trattamento con fasci di protoni aventi una energia massima di 200-250 MeV. Protoni di queste energie sono molto più difficili da produrre degli elettroni da 25 MeV della terapia convenzionale, ma soltanto così si riesce a penetrare per 25-30 cm il corpo del paziente e raggiungere tumori profondi. Vi sono inoltre circa 6000 pazienti all'anno che hanno tumori radioresistenti cioè insensibili sia alle radiazioni convenzionali che ai protoni e che quindi possono giovare di un trattamento con ioni.

La collaborazione TERA è formata da più di centocinquanta medici, fisici, ingegneri ed informatici che provengono da una quarantina di università, ospedali, centri oncologici e laboratori di ricerca italiani e vogliono portare per l'anno 2000 l'Italia all'avanguardia in Europa nel campo della terapia dei tumori con le radiazioni, passando la frontiera oggi segnata dagli Stati Uniti e dal Giappone, che sono i soli ad avere Centri di adroterapia di alta energia installati in complessi ospedalieri.

Per raggiungere questo scopo si lavora su tre linee complementari [21].

- La creazione di un *Centro Nazionale di Adroterapia Oncologica* che permetterà di irradiare pazienti in cinque sale in parallelo. Il Centro, il cui progetto è fatto da una collaborazione INFN-TERA-CNR-CERN, sarà dotato di un sincrotrone di 18 metri di diametro che sarà installato, con le sale e i servizi, in un bunker sotterraneo di 3500 metri quadrati. Il Centro tratterà a partire dal 2000, circa 100 pazienti all'anno. Le autorità pubbliche e le autorità private della città di Novara hanno formato un consorzio che finanzierà tutte le infrastrutture. Per la costruzione di tutte le parti di alta tecnologia (compreso il sincrotrone), sono necessari su cinque anni (1995-1999) circa 82 miliardi di lire, che sono stati richiesti alla Regione Piemonte e al Ministero della Sanità.

- Il progetto e la costruzione di un prototipo di un nuovo tipo di *acceleratore compatto*, che possa essere installato, in soltanto 300 m<sup>2</sup> di superficie, in ospedali ove esiste già un buon servizio di fisica sanitaria. Lo studio è fatto in collaborazione con l'ENEA.
- La realizzazione della *Rete Italiana Trattamenti Adroterapici* (RITA), che collegherà al Centro Nazionale, con moderni mezzi informatici multimediali, tutti gli Istituti, Ospedali, Cliniche pubbliche e private che vorranno avvalersi dell'Adroterapia. Si potranno scambiare immagini diagnostiche e i medici dei Centri Associati potranno pianificare in dettaglio il trattamento dei loro pazienti, che saranno irradiati in uno dei Centri ove i fasci di adroni sono disponibili.

#### 1.4.1 Specifiche dei fasci terapeutici dei protoni

A livello clinico, il *Comitato Patologie e Trattamenti* ha stabilito che il rateo di dose media deve essere di 30-40 Gy/min per volumi di trattamento inferiori a 50 cm<sup>3</sup>, e di 2-10 Gy/min per volumi di trattamento tra i 50 e 2000 cm<sup>3</sup>; inoltre la distribuzione della dose nel volume tumorale deve avere una precisione maggiore o uguale a 2 mm [3].

A livello fisico, l'acceleratore dovrà fornire fasci di protoni di energia compresa tra 60 e 250 MeV, con una variazione di energia minore o uguale a 0.8 MeV, corrispondente ad un passo di 0.2 cm: la varianza della distribuzione in energia del fascio (considerando una distribuzione gaussiana nello spread energetico) dovrà essere inferiore agli 80 KeV. Questi valori limiti dell'energia del fascio corrispondono ad un percorso nella materia che va da 3.1 g/cm<sup>2</sup> a 38.3 g/cm<sup>2</sup>. In realtà il percorso effettivo per i trattamenti sarà più piccolo a causa della perdita di energia del fascio nei materiali che precedono il paziente. L'intensità del fascio al termine della linea di fascio deve essere maggiore di 9 nA, corrispondente ad un tasso di dose di 1 Gy/min su un campo di 25×25 cm<sup>2</sup>, e potrà arrivare fino a ~ 17 nA per il massimo tasso di dose richiesto, cioè di 2 Gy/min per il massimo campo (si è assunta una efficienza del 90 % per il sistema di trasporto del fascio). Si richiede inoltre un tempo di estrazione di 250 ms con un duty cycle maggiore del 50 % a cui corrisponde quindi una frequenza di 2 Hz.

La variazione di energia del fascio si ottiene o mediante modulazione del percorso dei protoni nella materia, oppure cambiando l'energia del sincrotrone.

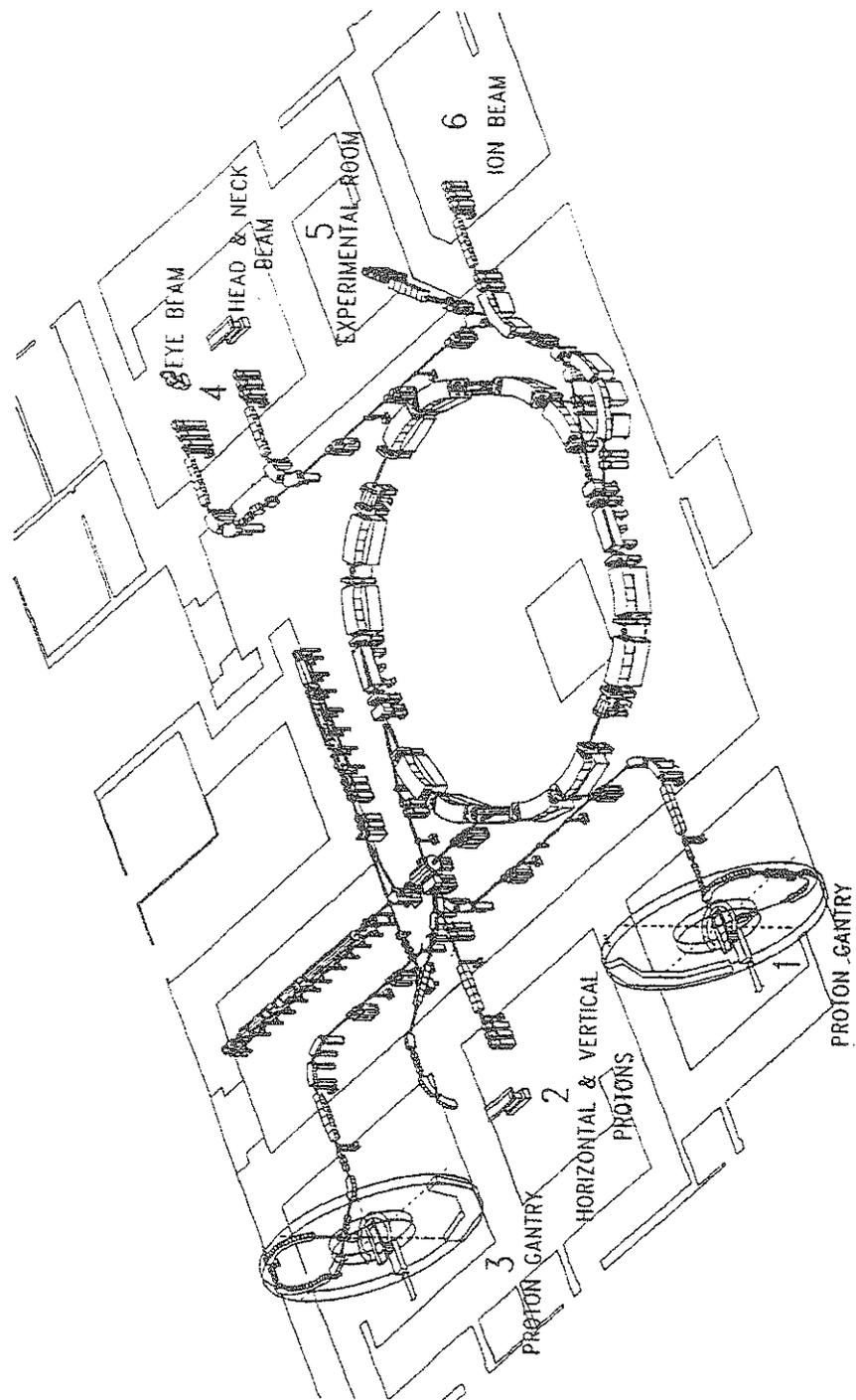


Figura 1.16: Vista della struttura ospedaliera complessiva: layout del bunker che contiene il sincrotrone e le sei stanze di trattamento [3].

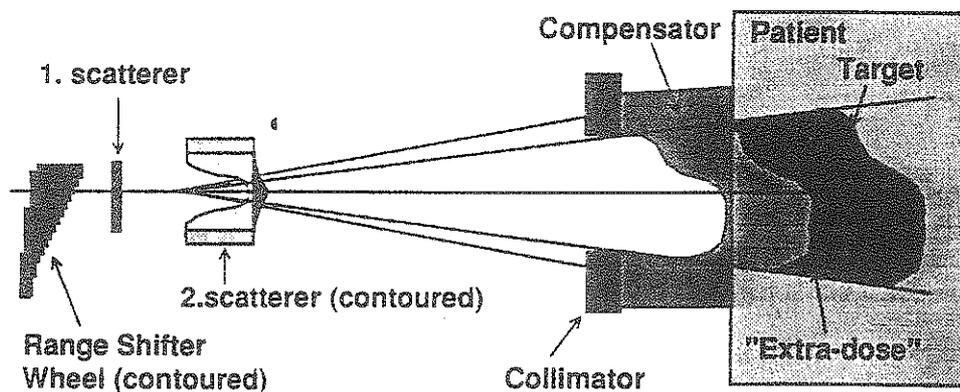


Figura 1.17: Principio alla base della tecnica di diffusione passiva del fascio [24].

Praticamente si utilizzano o un *sistema di trasmissione passivo* o un *dispositivo a scansione magnetica* (raster scan) [3].

### 1.4.2 Metodo di diffusione passiva

L'uso di un **diffusore passivo** rappresenta oggi il metodo più utilizzato in terapia. La figura 1.17 mostra il principio alla base del metodo [24].

Il fascio è diffuso 4 o 5 metri prima del paziente con un sistema di *scattering* disegnato in modo da produrre una buona uniformità (illuminazione costante) del fascio nella posizione del paziente. Vi è poi un collimatore per confinare la dose nella direzione laterale, mentre un sistema di modulazione rotante (Range Shifter Wheel) permette di cambiare il range del fascio in modo da distribuire la dose uniformemente in profondità (SOBP). Per rendere il profilo della dose esattamente corrispondente al profilo del bersaglio (target) si utilizza un mezzo di compensazione (compensator bolus) tra il collimatore ed il paziente, che avrà uno spessore variabile nel piano trasversale al fascio. Esso viene costruito con le informazioni fornite dalla Tomografia Computerizzata (CT). Purtroppo una dose non voluta si deposita nella regione prossima al volume bersaglio (extra dose). I vantaggi nell'utilizzo di un diffusore passivo del fascio sono la sicurezza e realizzabilità, semplicità ed economicità. D'altra parte occorre costruire un collimatore e un compensatore per ogni paziente; inoltre la distanza tra il diffusore ed il paziente è dell'ordine di 5 o 6 metri. È poi importante notare che il passivatore

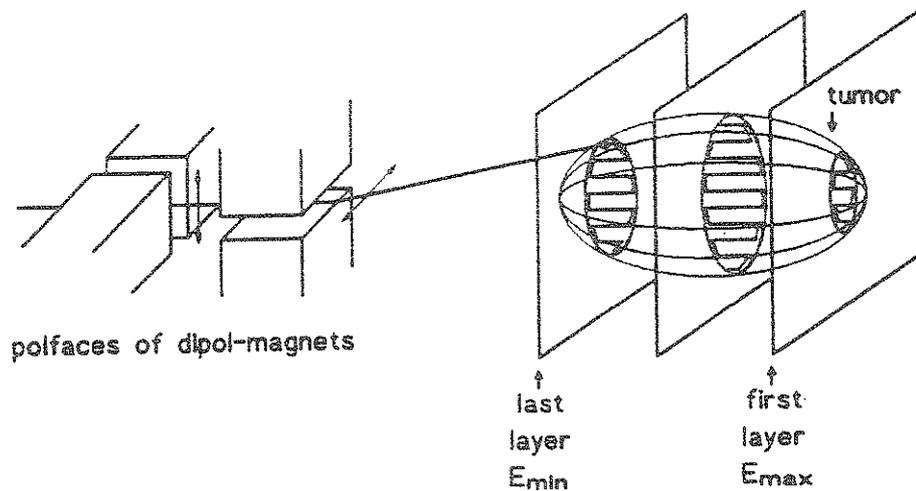


Figura 1.18: Vista schematica della procedura di irradiazione per un trattamento conforme utilizzando un sistema di scansione magnetica [13].

non può essere utilizzato con i fasci di ioni leggeri poiché indurrebbe il fenomeno della frammentazione.

### 1.4.3 Metodo di scansione magnetica (raster scan)

Quando si utilizza un sistema di **scansione magnetica (raster scan)**, il volume bersaglio viene diviso in sezioni corrispondenti ad eguale range delle particelle ed ogni sezione è trattata indipendentemente distribuendo il fascio su tutto la superficie della sezione stessa [13]. Dalla figura 1.18 si può notare che quando una sezione è stata trattata, l'energia del fascio, e quindi il range delle particelle, viene ridotta e si tratta la prossima sezione allo stesso modo. In funzione della geometria del volume bersaglio, le parti prossime al tumore sono preirradiate quando si trattano le parti più distanti. Quindi si rende necessaria una distribuzione non omogenea delle particelle sulla sezione per poter produrre effetti biologici omogenei.

Sono stati proposti due sistemi per produrre una distribuzione inomogenea di particelle: il pixel scanning e il raster scanning. La loro differenza risiede solo nella realizzazione dei magneti del sistema di scansione in quanto nel pixel scanning il beam è spento durante il movimento da un punto ad un altro mentre nel raster

scanning il fascio è sempre acceso. Considerando per semplicità il solo raster scanning, il fascio è mosso continuamente su un percorso a zig-zag (raster-like pattern) per ogni singola sezione e la velocità di scrittura è controllata dall'intensità del fascio e dalle richieste cliniche. Il percorso del fascio deve essere diviso in un numero molto grande di coordinate individuali che devono essere immagazzinate nella memoria di un calcolatore, ed il fascio si deve muovere in piccoli passi da una coordinata all'altra e la distanza tra un punto ed il successivo è piccola rispetto al diametro del fascio. A titolo di esempio dato che la velocità massima che si riesce a raggiungere è di circa 10 m/s, con un fascio di area circolare pari a 1 cm<sup>2</sup> occorrono circa 100 ms per coprire una superficie di 10×10 cm<sup>2</sup>. Per trattare un volume tridimensionale il fascio deve avere energia variabile. Ciò è possibile in un moderno sincrotrone, poiché si può controllare l'energia di ogni impulso. L'uso della scansione magnetica è oggi indispensabile per il trattamento con ioni leggeri. La tecnica è molto sofisticata e costosa. Rispetto alla diffusione passiva però, non occorre costruire compensatori e collimatori speciali per ogni paziente ed inoltre il paziente può essere disposto immediatamente dopo la macchina. Una versione digitale del sistema raster scanning (cioè una versione ibrida tra raster e pixel scanning) è utilizzata attualmente al laboratorio di biofisica del GSI, Darmstadt [13], dove si utilizza la terapia con ioni.

#### 1.4.4 Calcolo del flusso di protoni

La massima corrente che il sincrotrone deve raggiungere per un tasso di dose di 2 Gy/min ad un campo di (25×25) cm<sup>2</sup> ed un range R di 25 cm, con un fascio di protoni di 200 MeV, si può calcolare nel seguente modo:

$$Intensità_{protoni nell'unità di tempo} = I_p = \frac{2}{60} \cdot \phi \cdot A$$

dove il fattore 2 è stato introdotto per il fatto che il duty-cycle è del 50%, il fattore 60 per trasformare i secondi in minuti, mentre A è l'area in cm<sup>2</sup> e  $\phi$  è il flusso:

$$\phi = \frac{N}{A} = D \frac{R \cdot \rho}{E \cdot 1.6 \cdot 10^{10}}$$

dove il fattore  $1.6 \cdot 10^{10}$  è stato introdotto per poter esprimere la dose D in Gy.

l'energia dei protoni  $E$  in MeV, il range  $R$  in centimetri, la densità  $\rho$  in  $\text{gr}/\text{cm}^3$ .  
Si ottiene quindi una corrente di circa 5 nA di protoni a cui corrisponde un flusso di circa  $3 \cdot 10^{10}$  protoni al secondo per centimetro quadrato.

## Capitolo 2

# Misura di dose in adroterapia

### 2.1 Introduzione

L'energia impartita alla materia in un determinato volume da una radiazione ionizzante è la quantità fondamentale in dosimetria [23], sulla base della quale si è introdotta la quantità "dose assorbita" per attenersi alle richieste di distribuzione spaziale di energia rilasciata ad un mezzo irradiato.

Un rivelatore di radiazione risponde all'irradiazione con un segnale che è strettamente legato all'energia depositata al volume sensibile del rivelatore. Il segnale può inoltre dipendere dalla distribuzione microscopica nello spazio e nel tempo dell'energia depositata.

L'energia impartita in un processo base di deposizione di energia è data da [22]:

$$\delta\epsilon = T_b - \sum_i T_{a,i} + Q \quad (2.1)$$

dove

$T_b$  è l'energia cinetica ( $T = E - m_0c^2$ ) della particella ionizzante, immediatamente prima dell'interazione (in un processo spontaneo come il decadimento nucleare,  $T_b = 0$ );

$\sum_i T_{a,i}$  è la somma dell'energia cinetica di tutte le particelle ionizzanti create nel processo (inclusa l'energia cinetica residua della particella primaria se rimane una particella ionizzante dopo l'interazione);

$Q$  è il rilascio di energia di massa da parte dei nuclei e particelle elementari nei processi in cui si verificano trasformazioni nucleari.

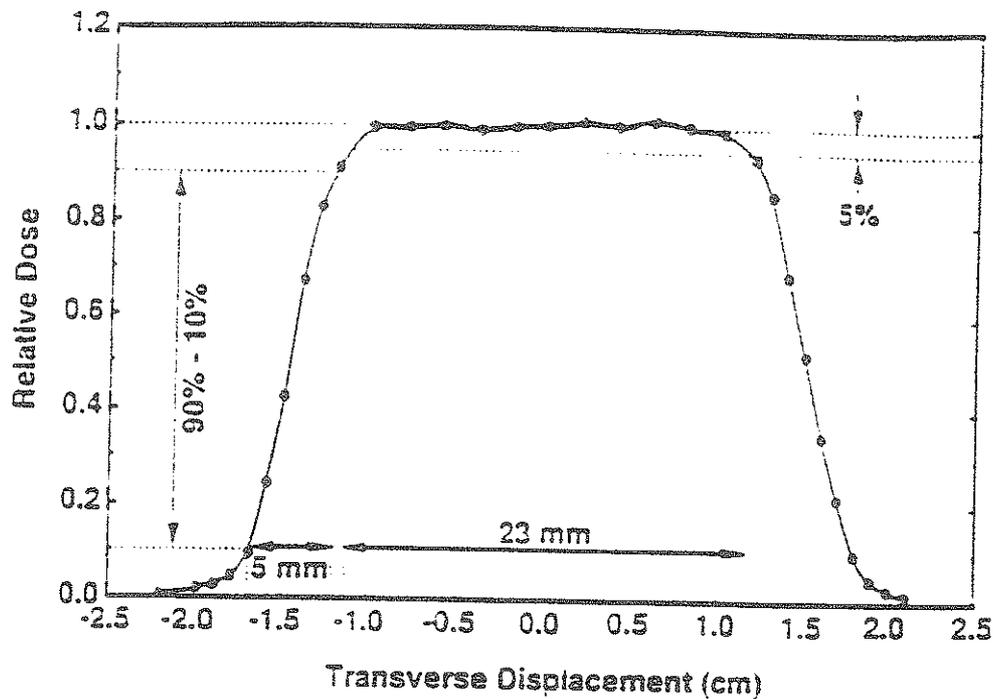


Figura 2.1: Profilo trasversale di un campo di trattamento con ioni He. Si può notare la caduta laterale della dose e l'uniformità del campo di trattamento nella zona tumorale. nel picco della distribuzione [6].

## 2.2 Misura della dose $D(x,y,z)$

Per poter attenersi ai piani di trattamento è di primaria importanza conoscere la deposizione di energia ad ogni istante in funzione della posizione trasversale e profondità del volume tumorale da trattare. Di particolare interesse clinico sono sia l'uniformità della radiazione (sul volume trattato) che la decrescita della dose alle estremità del campo di trattamento [6].

Lo spessore della decrescita laterale della dose (falloff, si veda la figura 2.1), spesso chiamata penombra laterale, è di grande importanza clinica poiché determina l'esposizione alla radiazione dei tessuti sani vicini alla zona tumorale. La caduta laterale della dose è definita come la distanza in cui la dose decresce dal 90% al 10% del valore centrale della dose.

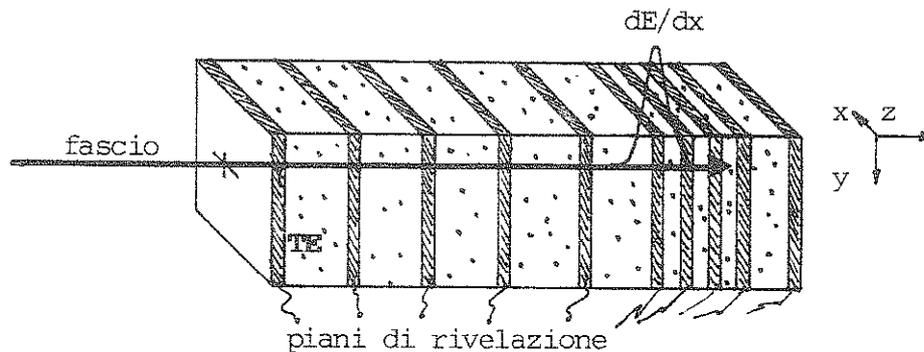


Figura 2.2: Vista schematica dell'ipotetico sistema di rivelazione

Le richieste terapeutiche del *Progetto Adroterapia* esigono una penombra laterale di 2 mm [3] e sono specificate per la superficie di ingresso nel tessuto, in modo da considerare solo i contributi dovuti alla macchina (acceleratrice) e non alla diffusione coulombiana o alla eventuale frammentazione del nucleo. Occorre dunque monitorare la perdita di energia per ionizzazione.

Si può pensare di utilizzare un calorimetro cubico delle dimensioni tipiche dei piani di trattamento, ad esempio  $(10 \times 10 \times 10) \text{ cm}^3$ , costituito da un materiale il più vicino possibile al tessuto umano (Tessuto Equivalente, TE), in cui vengono interposti dei rivelatori che siano in prima approssimazioni di dimensioni trascurabili rispetto all'insieme. Se si utilizzano delle camere proporzionali a gas di qualche millimetro di spessore, rapportato al TE si avrà:

$$\text{mm} \times \frac{\rho_{\text{gas}}}{\rho_{\text{TE}}} \sim \mu\text{m}$$

dove  $\rho$  è la densità, e si è considerata la superficie costante. Quindi qualche  $\mu\text{m}$  è perfettamente trascurabile rispetto ai circa 10 cm complessivi del cubo considerato.

Dalla tabella 2.1 si può notare che l'acqua può essere tranquillamente utilizzata

	muscolo ICRU	A-150 plastica	A-150 gas	metano gas TE	H <sub>2</sub> O
H	10.2	10.2	10.2	10.2	11.1
C	12.3	76.8	16.8	45.6	-
O	72.9	5.9	5.9	40.7	88
N	3.5	3.6	3.6	3.5	-
Ca	0.007	1.8	-	-	-
F	-	1.7	1.7	-	-
Ar	-	-	1.8	-	-
altro	1.093	-	-	-	-

Tabella 2.1: Composizione elementare in peso percentuale di differenti materiali TE (tessuto equivalente) e del tessuto muscolare ICRU.

come tessuto equivalente TE.

Nell'ambito del *Progetto TERA* (si veda il paragrafo 1.4), un gruppo di ricercatori dell'Università di Torino (INFN) sta operando per la realizzazione di tale rivelatore che ha preso il nome **cubo magico** e, sostanzialmente, di un sistema compatto e semplice di monitoraggio dei fasci di terapia.

I rivelatori nel cubo magico saranno interposti da una quantità di acqua variabile in funzione del trattamento, in quanto si vuole avere la massima acquisizione, cioè la massima risoluzione, nell'intorno dello Spread Out Bragg Peak (SOBP, si veda il paragrafo 1.3.1).

Si può controllare così l'energia rilasciata e la sua struttura dimensionale.

### 2.2.1 Misura on-line della dose

Per motivi di sicurezza e di conformità ai piani di trattamento occorre monitorare la dose in funzione della posizione trasversale  $x,y$  in modo continuo (on-line), di modo che si conosca esattamente l'energia rilasciata al paziente. Logicamente questa misura non potrà essere fatta con il cubo magico poiché in esso il fascio perde tutta la sua energia.

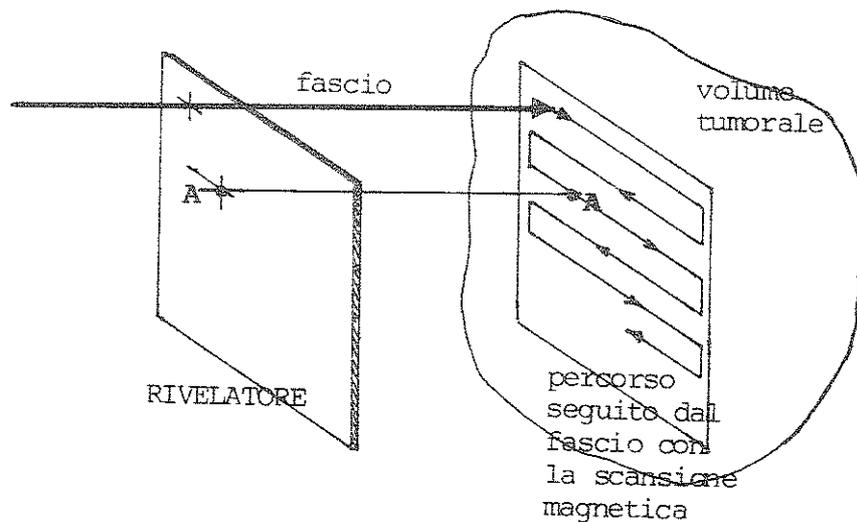


Figura 2.3: Vista schematica dell'ipotetico sistema di misura on-line della dose nel caso del raster scan.

Ipotizzando un trattamento che utilizzi il raster scan è fondamentale la misura on-line non solo della dose ma anche della posizione del fascio.

Dalla figura 2.3 si può notare che nel caso in cui i magneti del sistema di scansione non funzionassero bene per qualche motivo e il fascio si fermasse nel punto A, una semplice misura della dose totale impartita al paziente non sarebbe in grado di salvaguardarlo. Occorre cioè conoscere anche la posizione del fascio e bloccare la macchina quando questa per un difetto si fermi in un punto fisso. Questo tipo di monitoraggio può comunque, in linea di principio, essere fatto con lo stesso rivelatore utilizzato dal cubo magico.

### 2.2.2 Misura di microdosimetria: track structure

Come si è spiegato nel primo capitolo, per provocare un certo danno biologico nel tessuto tumorale, non è solo importante la dose rilasciata ma anche il tipo di radiazione utilizzata, cioè l'efficacia biologica relativa della medesima (RBE) e, del tipo di tessuto cellulare bersaglio. La dose, essendo definita come una quantità macroscopica, non tiene conto delle fluttuazioni, inomogeneità, e natura stocas-

Sezione A-A

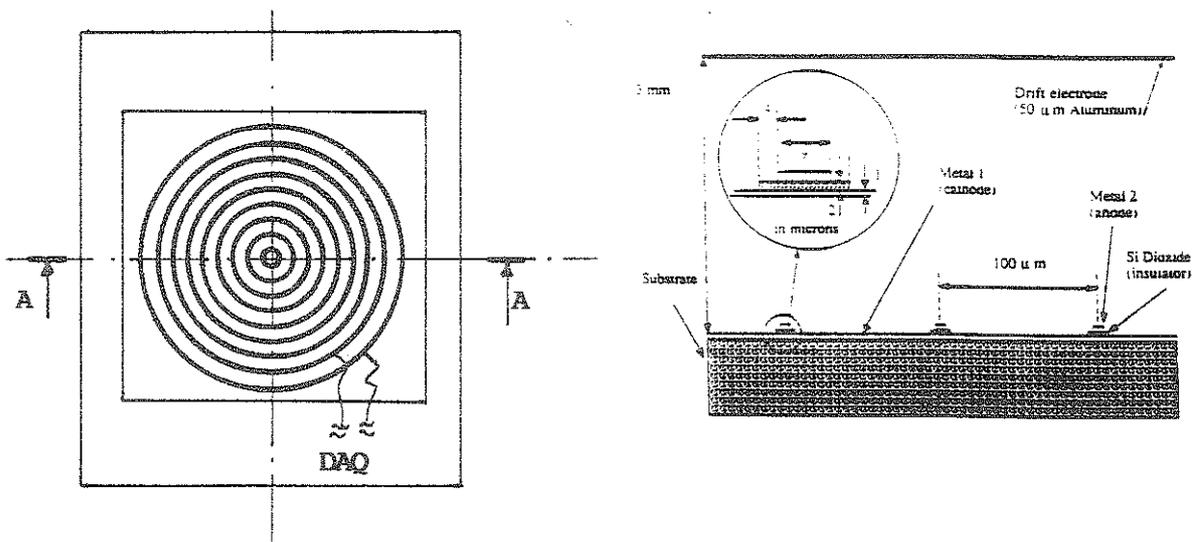


Figura 2.4: Disegno schematico di una ipotetica Micro-Gap-Gas-Chamber per la ricotruzione della struttura della traccia [25].

tica degli effetti della radiazione nella materia. sebbene questi siano importanti per la comprensione degli effetti biologici della radiazione. Per capire gli effetti di natura strettamente statistica della deposizione di energia in una cellula, occorre avere un rivelatore di dimensioni comparabili con quelle cellulari o quantomeno, occorre riuscire a ricostruire la struttura della traccia (track-structure).

Per gli ioni, la conoscenza dell'RBE verso il LET (Linear Energy Transfer, si veda il paragrafo 1.2) è di grande importanza. poiché il fenomeno della frammentazione comporta notevoli problemi di valutazione esatta della dose equivalente impartita. Ad esempio, a parità di dose rilasciata da un atomo primario di un fascio di ioni  $^{+6}\text{C}$  e dai figli secondari (es:  $^{+3}\text{Li} - ^{+3}\text{Li}$  ;  $^{+4}\text{Be} - ^{+2}\text{He}$  ;  $^{+5}\text{B} - \text{p}^+$ ), l'efficacia biologica relativa del genitore è estremamente superiore. In queste condizioni, con un rivelatore che fornisce la sola dose  $D(x,y,z)$  non è possibile stabilire il danno ai tessuti.

Uno dei problemi della microdosimetria è appunto la valutazione della distribuzione di energia rilasciata in funzione della distanza dalla traccia [27].

Modelli teorici tendono a descrivere il comportamento dell'RBE partendo dai dati sperimentali. In effetti le misure sperimentali sono limitate a poche situazioni e puntano ad una distribuzione di energia che è proporzionale a  $\frac{1}{d^2}$  [28].

È evidente che la costruzione di una camera a microgap [25] con anodi circolari

concentrici (come mostrato in figura 2.4), potrebbe fornire risultati che provano la correttezza dei modelli attuali.

Rimane sottinteso che l'uso di camere a strip con dimensioni caratteristiche  $\leq 100\mu\text{m}$  permette di scendere rapidamente a misure di  $\sim 0.1\mu\text{m}$  in TE (Tessuto Equivalente). Un ulteriore guadagno si può ottenere operando le camere a pressione inferiore a quella atmosferica. Per esempio con pressioni dell'ordine di 0.1 bar si possono raggiungere distanze dell'ordine di 10 nm in TE [26].

## 2.3 Calcolo dell'intensità dei protoni nei fasci di trattamento

Dalle specifiche delle intensità tipiche dei piani di trattamento del Progetto Tera (si veda il paragrafo 1.4.1) si possono identificare due situazioni interessanti:

- dose impartita di intensità di  $2\div 10$  Gy/min su di un volume di  $2000\text{ cm}^3$
- dose impartita di intensità di  $30\div 40$  Gy/min su di un volume di  $50\text{ cm}^3$

Assumendo che la dose sia rilasciata principalmente nello Spread Out Bragg Peak e che il rapporto tra la dose impartita in questo rispetto al plateau sia  $f$  ( $f \sim 3$ , si veda la figura 1.10), si ha:

$$Dose/min = f \cdot P \cdot SP \cdot \rho \cdot X \cdot \frac{K}{M} \cdot d \quad (2.2)$$

dove

$P$  è il numero di protoni al secondo;

$SP$  è il potere frenante (Stopping Power) per i protoni espresso in  $\frac{\text{MeV}}{\text{gr/cm}^2}$ ;

$M = \rho \cdot X \cdot A$  è la massa irradiata;

$K = 1.6 \cdot 10^{-10} \cdot 60$  è un fattore di conversione per poter esprimere la dose  $D$  in Gy/min; infatti:

$$1 \frac{\text{MeV}}{\text{grsec}} = 1.6 \cdot 10^{-19} \cdot 10^9 \frac{\text{J}}{\text{Kgsec}} \text{ e quindi } 1 \frac{\text{MeV}}{\text{grsec}} = 1.6 \cdot 10^{-10} \cdot 60 \frac{\text{Gy}}{\text{min}};$$

$d$  è il duty cycle dell'acceleratore:  $d = 0.5$  con una frequenza di 2 Hz.

Considerando un fascio di protoni di 200MeV, lo stopping power  $SP$  è di circa

4.5  $\frac{MeV}{gr/cm^2}$  in TE (plastica A-150) [33]. Il numero di protoni per unità di tempo sarà quindi dato da:

$$P = \frac{D}{f \cdot SP \cdot K} \cdot \frac{A}{d} = \frac{D}{f \cdot SP \cdot K} \cdot \frac{V}{X} \cdot \frac{1}{d} \quad (2.3)$$

dove

$V$  è il volume trattato in  $cm^3$  e  $X$  è il range di trattamento.

Per un range di  $X = 20$  cm si ottengono i seguenti risultati:

- con una dose di 2 Gy/min su  $2000cm^3$  si hanno:  
 $P = 3.08 \cdot 10^9$  protoni/sec mentre il flusso  
 $\phi_p = 3.08 \cdot 10^7$  protoni/sec· $cm^2$ ;
- con una dose di 10 Gy/min su  $2000cm^3$  si hanno:  
 $P = 1.54 \cdot 10^{10}$  protoni/sec mentre il flusso  
 $\phi_p = 1.54 \cdot 10^8$  protoni/sec· $cm^2$ ;
- con una dose di 30 Gy/min su  $50cm^3$  si hanno:  
 $P = 1.14 \cdot 10^9$  protoni/sec mentre il flusso  
 $\phi_p = 4.62 \cdot 10^8$  protoni/sec· $cm^2$ ;
- con una dose di 40 Gy/min su  $50cm^3$  si hanno:  
 $P = 1.54 \cdot 10^9$  protoni/sec mentre il flusso  
 $\phi_p = 6.16 \cdot 10^8$  protoni/sec· $cm^2$ ;

## 2.4 Rivelatori

Molti e diversi tipi di rivelatori sono ormai oggi utilizzati per misurare le caratteristiche dei fasci di radiazione. I rivelatori in cui le particelle perdono la loro energia, come i contatori proporzionali a gas o i rivelatori al silicio, sono utilizzati per misurare il LET del fascio di particelle. Altri tipi di rivelatori, come i rivelatori a stato solido al germanio o i rivelatori a tempo di volo (*time-of-flight*), sono utilizzati per misurare l'energia totale delle particelle. Inoltre diversi di questi rivelatori sono stati integrati in sistemi per la misura delle proprietà della

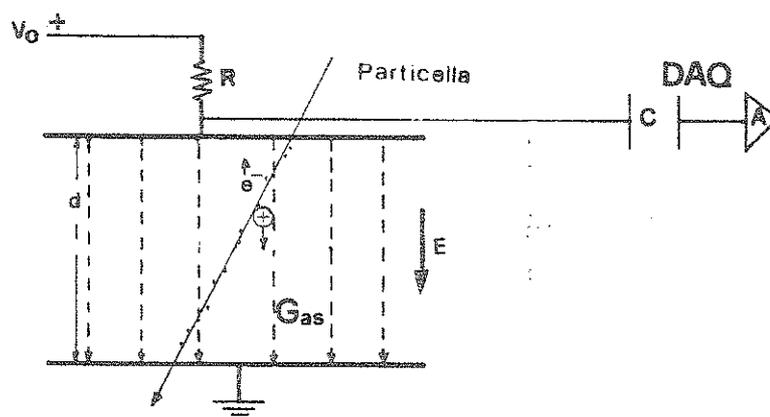


Figura 2.5: Schema di una camera a ionizzazione.

radiazione sotto svariate condizioni. per valutare modelli biologici della risposta cellulare alla radiazione [6].

### 2.4.1 Rivelatori a gas

I rivelatori a gas sono basati sulla misura delle cariche liberate per ionizzazione all'interno di una opportuna miscela gassosa. Il primo rivelatore utilizzato per la misura della dose è stata la camera a ionizzazione, che è il più semplice rivelatore a gas. In riferimento alla figura 2.5, si può constatare che questo rivelatore è costituito da un volume gas in un campo elettrico generato da due elettrodi piani a distanza  $d$  tenuti ad una differenza di potenziale  $V_0$ . Quando una particella attraversa il volume di gas, cede una energia  $\Delta E$  e libera un numero di coppie elettrone-ione pari a:

$$n_{coppie} = \frac{\Delta E}{W} \quad (2.4)$$

dove  $W$  è l'energia media necessaria a creare una coppia  $e^-$ -ione ed è dell'ordine dei 30eV ( ad es.  $W = 36\text{eV}$  per il Ne,  $24.6\text{eV}$  per il  $\text{C}_2\text{H}_6$ ) [34].

La ionizzazione del gas contenuto nel rivelatore è proporzionale all'energia ri-

lasciata dalla radiazione, che a sua volta è proporzionale alla dose assorbita dal rivelatore. Questa dose può essere messa in relazione con la dose assorbita da un altro mezzo, come il tessuto umano, conoscendo il rapporto tra il potere frenante (stopping power) del gas e quello del mezzo considerato.

La quantità di carica  $Q$  (Coulomb) che si può raccogliere agli elettrodi della camera è direttamente proporzionale all'energia rilasciata nel gas.

In queste condizioni la dose assorbita dal rivelatore è data dall'equazione di Bragg-Gray:

$$D = \frac{QW}{\rho V} \quad (2.5)$$

dove  $\rho$  ( $\text{kg}/\text{m}^3$ ) è la densità e  $V$  ( $\text{m}^3$ ) è il volume del gas.

Come si è discusso nel paragrafo precedente, il flusso di protoni è dell'ordine dei  $10^7 \div 10^8$   $\text{p}/\text{s}\cdot\text{cm}^2$ , ed è troppo alto per essere sopportato in una camera proporzionale (come una Multi Wire Proportional Chamber). In questa infatti, si verrebbero a creare delle cariche spaziali che distorcerebbero il potenziale in prossimità degli elettrodi: il numero di ioni creati in regime di moltiplicazione diverrebbe così alto che si produrrebbe una nube di potenziale positivo in grado di modificare le linee di campo in cui è immerso il gas. Inoltre le strutture a fili invecchiano molto velocemente e comportano quindi delle variazioni di guadagno nel tempo [11].

Un modo per poter superare questo problema è quello di utilizzare due camere: una camera a ionizzazione con uno o comunque pochi elettrodi per la misura della dose, a cui si associa trasversalmente una MWPC in regime di ionizzazione per la misura della posizione, come mostrato in figura 2.6. Si introduce così un problema non indifferente, cioè l'accoppiamento di due rivelatori. L'ideale sarebbe utilizzare un unico rivelatore per i due scopi: misura della dose e posizione  $x,y$ . Tutto ciò è possibile utilizzando camere a ionizzazione a strip o a pixel.

Quest'ultima non è altro che una camera a ionizzazione già descritta nella figura 2.5, in cui uno degli elettrodi (piani in figura) è diviso in tanti piccoli pixel (dell'ordine di  $(1 \times 1)$   $\text{mm}^2$ ); volendo leggere gli elettroni liberati nella ionizzazione del gas da parte della radiazione incidente occorre suddividere il piano tenuto a potenziale più alto (anodo).

Nella camera a ionizzazione a strip invece, (si veda la figura 2.7) un piano è costituito da strip di elettrodi adiacenti, spaziate da una distanza trascurabile

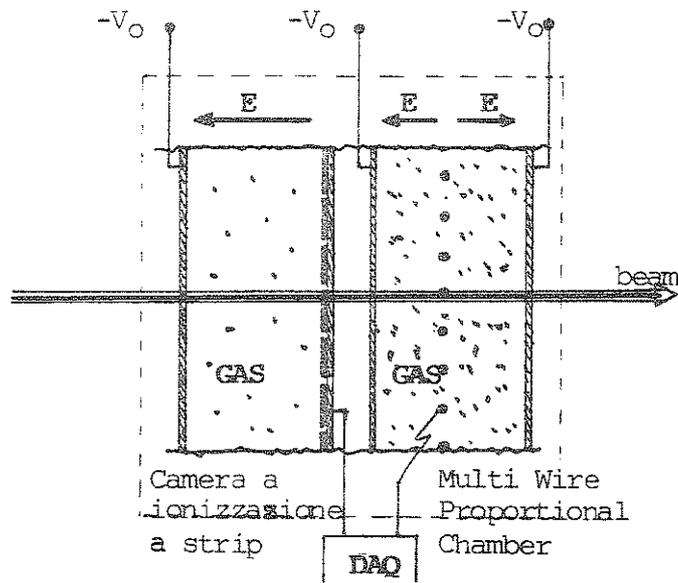


Figura 2.6: Vista schematica di una possibile soluzione per la misura della dose  $D(x,y,z)$ .

rispetto alla larghezza della strip. Per essere più precisi possiamo dare le misure approssimative della camera a strip da utilizzare nel cubo magico.

La larghezza delle strip dovrà essere dell'ordine del millimetro (1 o 2 mm) in quanto tale è la risoluzione richiesta dalle specifiche dei trattamenti (si veda paragrafo 2.1). Lo spazio tra una strip e l'altra sarà dell'ordine della decina di  $\mu\text{m}$  in quanto tale è il limite tecnologico per la deposizione di strip su vetrino, in ogni caso è trascurabile rispetto alla larghezza.

Il gap, ossia la distanza tra il piano catodico e le strip dovrà essere sicuramente inferiore al centimetro per non dare problemi nella costruzione del fantoccio di rivelazione, e sufficientemente grande da poter dare la possibilità alla radiazione di generare coppie elettrone-ione tali da fornire un segnale elettrico rivelabile con l'attuale elettronica di acquisizione; sarà dimostrato più oltre che un gap di 3 mm è sufficiente a dare una corrente dell'ordine del nA e quindi ragionevole. Dato che ci interessa avere l'integrale della dose impartita nel tempo non è importante rivelare il singolo evento. Si potrebbe e quindi far "driftare" gli elettroni per una distanza anche maggiore. Non ci interessa qui scegliere un valore ottimale di gap per avere un determinato rate rivelabile ( $rate \sim \frac{w_e}{gap}$  dove  $w_e$  è la velocità di deriva degli elettroni, tipicamente  $\sim 50\text{mm}/\mu\text{s}$ ).

La tensione da applicare agli elettrodi dovrà essere tale da garantire, in base al

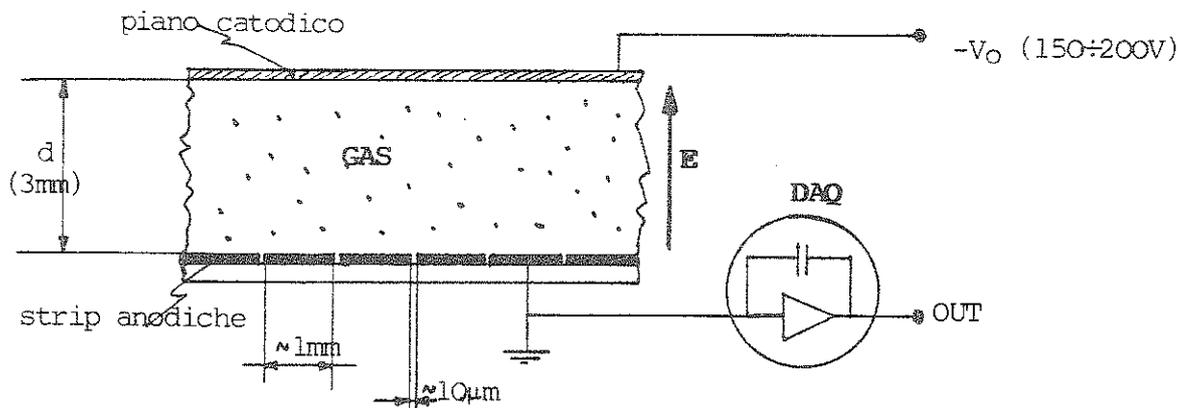


Figura 2.7: Vista schematica di una camera a ionizzazione a strip

gap scelto, il regime di ionizzazione e quindi non superare il valore di  $1000 \text{ V/cm}$ ; una tensione quindi di  $150 \div 200 \text{ V}$  su  $d = 3 \text{ mm}$  sarà più che sufficiente.

Il numero delle strip dovrà essere tale per cui si riesca a coprire i  $(10 \times 10) \text{ cm}^2$  di superficie sensibile richiesta dalle specifiche. Si dovranno quindi avere 100 strip da  $1 \text{ mm}$  per ogni camera ed utilizzare due camere accoppiate ruotate di  $90^\circ$ , l'una rispetto all'altra, per avere le due coordinate del piano. Infine più camere accoppiate disposte sullo stesso asse forniscono la coordinata di profondità ( $D(x,y,z)$ ). Supponendo di utilizzare 10 coppie di camere da 100 strips ciascuna, si avranno in totale  $10 \times 2 \times 100 = 2000$  strips totali.

#### 2.4.2 Il problema della ricombinazione

Nel loro moto di deriva verso gli elettrodi, le cariche possono essere neutralizzate in seguito a processi di urto che si risolvono con la ricombinazione di cariche di segno diverso.

L'effetto complessivo può essere separato in due componenti:

i) *Ricombinazione iniziale* dovuta agli ioni generati dalla stessa traccia e che si ricombinano prima che il campo elettrico applicato possa separarli. Questa componente è indipendente dall'intensità del fascio ed è trascurabile per particelle a basso LET.

ii) *Ricombinazione generale* dovuta alla ricombinazione degli ioni generati da tracce differenti. Le tracce molto vicine spazialmente e temporalmente, hanno una alta probabilità di ricombinazione. Quindi l'entità del problema dipende dall'intensità del fascio.

I parametri che fanno variare l'efficienza di raccolta degli ioni sono:

- le condizioni di lavoro e struttura geometrica della camera, quali la pressione e il tipo di gas utilizzato per riempire la camera. Per una data densità di carica la probabilità di ricombinazione è inversamente proporzionale alla mobilità degli ioni, che per l'aria a pressione atmosferica è quasi uguale per ioni positivi e negativi, in quanto gli elettroni non sono catturati dagli atomi per formare anioni pesanti. Per gas inerti come Ar, Ne, la mobilità degli elettroni è invece tre ordini di grandezza superiore.

Inoltre la ricombinazione è inversamente proporzionale al rapporto  $\frac{V}{d^2}$ .

Per fasci di alta intensità occorre porre molta attenzione a fare in modo che l'efficienza di raccolta sia la più alta possibile, cioè prossima all'unità.

- la struttura temporale del fascio ed in particolare la lunghezza di estrazione  $t_s$ . Nella letteratura si considerano due regioni estreme rispetto al tempo  $\mathcal{T}$  di raccolta degli ioni. Da un lato, la radiazione è impartita in un pacchetto temporale trascurabile rispetto a  $\mathcal{T}$  ( $t_s \ll \mathcal{T}$ : radiazione impulsata); dall'altro lato la ricombinazione e la raccolta raggiungono l'equilibrio ( $t_s \gg \mathcal{T}$ ).

### 2.4.3 Fibre plastiche scintillanti

Possono essere ottenute con sezioni circolare o quadrata di dimensioni tipiche di un millimetro, e decine di centimetri di lunghezza. Una particella carica, quando attraversa una fibra, ne eccita gli atomi che, quando ritornano al loro stato naturale, emettono fotoni di lunghezza d'onda dell'ordine dei nanometri. Una grande frazione della luce è intrappolata nella fibra che internamente è totalmente riflettente, e viene trasmessa alle sue estremità. Ad una di queste (o ad entrambe, raddoppiando però il numero di canali di elettronica), un trasduttore converte la quantità di luce raccolta in una carica elettrica. Data la composizione chimica e la forma geometrica della fibra l'uscita del trasduttore è proporzionale alla sola energia depositata. Allineando le fibre su di un sottile supporto plastico si può

ricoprire un'area delle dimensioni tipiche di un campo di trattamento. Il vantaggio di questa tecnica risiede nella compattezza meccanica e nella semplicità. Inoltre la quantità di luce è linearmente proporzionale all'energia depositata e non è soggetta ad effetti di ricombinazione, già accennati a riguardo della camera a ionizzazione. Occorre però considerare il fatto che il rivestimento delle fibre (che è indispensabile) e lo spazio tra una fibra e l'altra, non sono sensibili alla radiazione. Quindi occorre calibrare le fibre per valutare un opportuno fattore di correzione che tenga conto di questo effetto. Per la lettura delle fibre si possono poi utilizzare *fotomoltiplicatori multi anodo*, in cui le fibre sono poste sul fotocatodo come una matrice ed ogni sezione del catodo è mappata in una dell'anodo, oppure utilizzare altri sistemi (*contatori di fotoni di luce visibile*, ecc.).

La radiazione assorbita cambia la lunghezza di attenuazione della fibra e conseguentemente la luce ottenibile. Tipicamente la lunghezza di attenuazione di una fibra è dell'ordine di circa 5 m, che scende a circa 2 m dopo l'esposizione ad 1 Mrad di radiazione.

Considerando una dose media per trattamento di circa 1 Gy (= 100 rad), dopo 100 trattamenti (circa 1 giorno) la dose assorbita dalla fibra sarà dell'ordine della decina di krad e la luce ottenibile diminuirà dell'1%. Questo implica una calibrazione periodica (una volta al giorno) del sistema.

## Capitolo 3

### Specifiche del front-end

L'elettronica di front-end riveste un ruolo di primaria importanza nella precisione e attendibilità della misura della dose. Le camere a ionizzazione, come la maggior parte dei rivelatori usati in dosimetria, raccolgono ai loro elettrodi una carica elettrica che è proporzionale all'energia depositata nel volume sensibile del rivelatore (equazione 2.5 ).

Dato che i fasci clinici generati da acceleratori hanno una struttura temporale, misurare la dose significa misurare la carica totale come integrale di una corrente fluttuante intorno ad un valore medio. La corrente media può cambiare di vari ordini di grandezza secondo i piani di trattamento e quindi l'elettronica deve essere disegnata in modo che sia garantito il funzionamento nelle condizioni estreme del fascio.

#### 3.1 Derivazione delle correnti

Come spiegato nel paragrafo 2.4.1 il rivelatore *cubo magico*, che dovrà essere utilizzato per il controllo di routine del fascio di trattamento, avrà in totale circa 2000 strips e quindi un egual numero di canali di lettura.

Inoltre nei paragrafi 1.4.2 e 1.4.3 si sono citate due possibili configurazioni per investire il bersaglio (zona tumorale) con il fascio di radiazione: la diffusione passiva e la scansione magnetica (raster scan). Si ricorda che mentre per i protoni sono possibili entrambe le soluzioni, per gli ioni può essere utilizzata la sola scansione magnetica, per evitare il fenomeno indesiderato di frammentazione.

Si procederà ora al calcolo del segnale in corrente che ci si aspetta per ogni strip

(canale) del cubo magico ipotizzando che questo sia investito da un fascio di protoni di 200 MeV (energia tipica di trattamento richiesta dalle specifiche del *Progetto Tera*), avente un diametro di 1 cm. Si considereranno inoltre le due possibili configurazioni di trattamento sopra citate (passivatore e raster scan). L'energia persa in gas dovuta alla ionizzazione da parte di protoni di 200MeV è di 3.5 KeV/cm [33]. Al picco di Bragg l'energia persa è sei volte più alta, cioè di 21 KeV/cm.

Assumendo, con buona approssimazione, che l'energia media richiesta per creare una coppia sia di 35 eV, il numero di coppie elettrone-protoni prodotte in regime di ionizzazione sarà:

$$n_{coppie} = 3.5 \cdot 10^3 K \cdot d \cdot \frac{1}{35}$$

dove  $d$  è lo spessore di gas della camera (gap  $\sim 3$  mm):

$K$  vale 1 per il plateau e 6 per il picco di Bragg.

Se tutte le coppie prodotte raggiungono gli elettrodi (cioè se l'efficienza di raccolta  $\times$  guadagno della camera  $\sim 1$ ), la carica generata al secondo nella camera, che è quindi la corrente fornita dalle strip, è data da:

$$i_{strip} = n_{coppie} \times \phi_{protoni} \times A$$

dove  $\phi_{protoni}$  è il flusso di protoni, cioè il numero di protoni al secondo per  $\text{cm}^2$ , e  $A$  è l'area di intersezione tra il fascio di protoni incidente e la singola strip considerata. Entrambe queste due grandezze cambiano se il sistema utilizzato per impartire la dose al paziente è la *diffusione passiva* o la *scansione magnetica*. Dato un piano di trattamento (cioè definiti il campo  $l \times l$  e la dose  $D$ ), e assumendo che il tempo di irradiazione del paziente sia lo stesso per tutti e due i sistemi di scansione, si ha che il numero di protoni al secondo,  $P$ , forniti dall'acceleratore al paziente, deve rimanere lo stesso nei due casi (per il calcolo di  $P$  si veda il paragrafo 2.3).

Utilizzando un sistema di diffusione passivo si ha quindi:

$A = l \times spessore_{strip}$  (assumendo che il campo di trattamento sia più piccolo delle dimensioni del rivelatore), mentre:

$$\phi_{protoni} = P/(l \times l):$$

Quindi posto  $l = 10$  cm e  $spessore_{strip} = 0.1$  cm si ottiene:

Dose	2Gy/min su 2000cm <sup>3</sup> l=10cm	10Gy/min su 2000cm <sup>3</sup> l=10cm	30Gy/min su 50cm <sup>3</sup> l=1.5cm	40Gy/min su 50cm <sup>3</sup> l=1.5cm
$P \times 10^9$ (protoni/sec)	3.08	15.4	1.14	1.54
Diffusione passiva $i_{strip}$ nA	0.88	4.44	2.16	2.96
Scansione magnetica $i_{strip}$ nA	11.30	56.52	4.16	5.64

Tabella 3.1: Valori di corrente che si ottengono per il sistema a *diffusione passiva* e quello a *scansione magnetica* per i diversi piani di trattamento limite richiesti:  $K = 6$  (picco di Bragg).

$$i_{strip} = 3.5 \cdot 10^3 K \frac{0.3}{35} \times \frac{P}{100} \cdot 10 \cdot 0.1$$

Utilizzando un sistema di scansione magnetica si ha invece:

$A \sim Diametro_{fascio} \times spessore_{strip}$ , mentre:

$$\dot{\phi}_{protoni} = P / (Area_{fascio}).$$

e quindi posto il fascio di diametro 1 cm si ottiene:

$$i_{strip} = 3.5 \cdot 10^3 K \frac{0.3}{35} \times \frac{P}{\pi(0.5)^2} \cdot 1 \cdot 0.1$$

Sostituendo e ricordando le intensità calcolate nel paragrafo 2.3 si trovano i valori riportati nella tabella della tabella 3.1.

### 3.1.1 Range dinamico

Dalla tabella 3.1 si possono leggere i valori delle massime correnti da rivelare che sono:

$$i_{max} = 4.44 \text{ nA per la } \textit{diffusione passiva}$$

$$i_{max} = 56.52 \text{ nA per la } \textit{scansione magnetica}$$

La minima corrente la si può calcolare nel seguente modo.

Ipotizzando che si richieda una precisione del 2.5% al picco di Bragg, si calcola dunque questa percentuale per il minimo valore riportato in tabella per i due sistemi di scansione del fascio, e lo si divide per un fattore che terrà conto del rapporto tra rilascio di energia nello Spread Out Bragg Peak e nel plateau. Infatti fino ad ora si era considerato un fascio di protoni monoenergetico e quindi si aveva un rapporto di  $K = 6$  tra la dose rilasciata al picco di Bragg ed al plateau. Durante il trattamento però, questo rapporto diminuisce poichè il picco viene allargato (SOBP) come riportato in figura 1.10. Possiamo, approssimando per difetto i calcoli, assumere che ora  $K = 2$ . Si ha dunque:

$$i_{min} = 0.88 \times 0.025 / 2 = 0.010 \text{ nA per la } \textit{diffusione passiva};$$

$$i_{min} = 4.16 \times 0.025 / 2 = 0.052 \text{ nA per la } \textit{scansione magnetica}.$$

Si può quindi calcolare il range dinamico  $R$  che sarà pari a:

$$R = \frac{4.44}{0.010} \sim 440 \text{ per la } \textit{diffusione passiva};$$

$$R = \frac{56.52}{0.052} \sim 1080 \text{ per la } \textit{scansione magnetica}.$$

Possiamo quindi dire che il range dinamico a cui deve soddisfare lo stadio di ingresso accoppiato al sistema di acquisizione, per soddisfare le richieste cliniche, deve coprire approssimativamente tre ordini di grandezza ( $R \sim 10^3$ ).

### 3.1.2 Stima della capacità di ingresso

Come già spiegato il *cubo magico* sarà costituito da camere a ionizzazione a strip <sup>1</sup>. Il gap sarà di  $\sim 3$  mm (distanza tra gli elettrodi) e le strip saranno larghe  $\sim 1$  mm e lunghe 10 cm, spaziate l'una dall'altra dalla minima distanza possibile, imposta dalla tecnologia utilizzata (qualche  $\mu\text{m}$  se si utilizza il processo SACMOS per la metallizzazione del vetrino di supporto delle strip).

---

<sup>1</sup>Altresì si sta esaminando la possibilità di suddividere la camera in pixel (picture elements). Affinchè il numero di canali non diventi troppo alto, l'area dei pixel dovrebbe essere dell'ordine di qualche decina di  $\text{cm}^2$ . Si noti comunque che le conclusioni di questo lavoro sostanzialmente non cambierebbero.

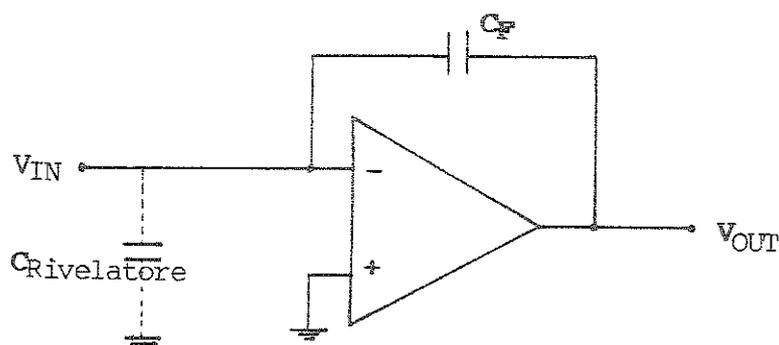


Figura 3.1: Schema di un circuito integratore di carica.

Ogni strip costituisce allora un condensatore a facce piane e parallele, per cui la capacità sarà [35]:

$$C = \varepsilon \cdot \frac{S}{d} + \delta$$

dove:

$\varepsilon = \varepsilon_0 \cdot \varepsilon_r \sim \varepsilon_0 = 8.85 \cdot 10^{-12}$  F/m. è la costante dielettrica del vuoto;

$S$  e  $d$  sono la superficie e la distanza tra gli elettrodi;

$\delta$  è un fattore introdotto per tener conto del fatto che inevitabilmente le strip dovranno essere unite a delle piste di collegamento di metallo, le quali posseggono una capacità parassita (tuttavia la si trascurerà in quanto dipenderà dal sistema di costruzione dei supporti di collegamento). Si ottiene dunque:

$$C = 8.85 \cdot 10^{-12} \cdot (0.01 \cdot 0.1)/0.03 \sim 0.3 \text{ pF.}$$

## 3.2 Soluzioni possibili del front-end

L'elettronica di front-end dovrà essere in grado di misurare la corrente che arriva dalle strip della camera a ionizzazione, la quale è pressochè continua e dell'ordine del nA. Dovrà quindi possedere un range dinamico di  $10^3$ , come precedentemente calcolato.

Il primo stadio di preamplificazione di potrebbe essere un integratore di carica o un dispositivo sensibile alla corrente (amperometro). Si consideri per esempio un integratore di carica come mostrato in figura 3.1.

La corrente che arriva dalla strip, vede l'ingresso dell'amplificatore operazionale, che idealmente ha resistenza infinita, e poi vede una capacità (siveda la figura 3.2) che per effetto Miller [36] è:

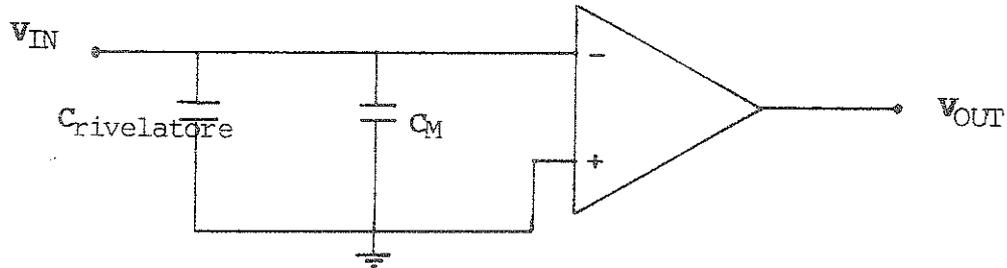


Figura 3.2: Circuito equivalente.

$$C_M = (1 + A_0) \cdot C_F$$

dove  $A_0$  è il guadagno ad open loop dell'amplificatore operazionale (generalmente dell'ordine di  $10^3$ ). Gli elettroni che vengono liberati dalla ionizzazione del gas, raggiungono le strips e la loro carica viene quasi totalmente trasferita su  $C_F$ , essendo la capacità del rivelatore trascurabile rispetto a  $C_F$  stessa.

Si ha così in uscita una tensione [36]:

$$V_{out}(t) = -\frac{1}{C} \int_0^t i_{strip}(t) dt + V_{out}(0) \quad (3.1)$$

Caricando la capacità durante il tempo di trattamento, la tensione ai capi di questa sarà proporzionale alla dose assorbita.

È molto difficile però creare un sistema che sia allo stesso tempo lineare e di buona precisione su diversi ordini di grandezza: si avranno infatti inevitabilmente problemi di saturazione dell'elettronica o di soglia.

Ci sono almeno due modi per risolvere il problema:

- Integratore ciclico (Recycling Integrator)
- APM (Analog Memory Pipeline, traducibile in memoria analogica vettorizzata)

### 3.2.1 Integratore ciclico

**Recycling integrator o Charge balancing current integrator.**

Un tipo di integratore di carica particolarmente indicato per applicazioni cliniche è quello che fornisce un impulso per ogni incremento fissato di carica in ingresso

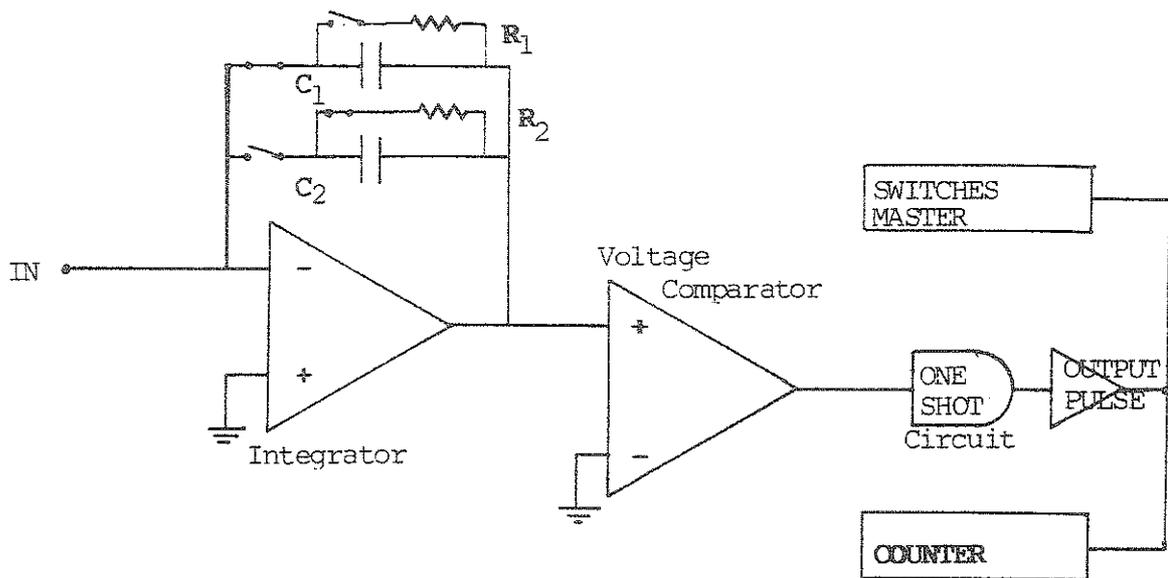


Figura 3.3: Circuito standard di integratore ciclico

[32]. Dalla figura 3.3 si può capire il funzionamento. È conveniente scegliere questa unità di misura così piccola che ogni carica residua all'inizio o alla fine della misura (del trattamento) possa essere ignorata [6].

Tutte le volte che la tensione sulla capacità di integrazione raggiunge un valore di soglia predeterminato, viene dato un impulso in uscita, e immediatamente o si sottrae una carica di riferimento dalla capacità (così funziona il classico circuito degli Analog Digital Converter), o si carica un'altra capacità di integrazione mentre si scarica la prima (come mostrato in figura 3.3).

L'unità di carica è piccola, solitamente tra 1 e 10 pC e il range dinamico che si viene a raggiungere può andare da  $10 \mu\text{A}$  a  $1 \text{ pA}$  ( $R \sim 10^4$ ) di corrente in ingresso [32].

### 3.2.2 APM: Analog Pipeline Memory

L'APM (potrebbe tradursi con "memoria analogica vettorizzata") non è altro che un integratore dove le capacità di carica sono molte e vengono caricate una dopo l'altra, sequenzialmente.

In sostanza si ha una matrice di memorie analogiche (capacità) indirizzabili, per l'immagazzinamento temporaneo dei dati.

La corrente in ingresso carica le capacità per un tempo regolabile con un impulso di clock. Le stesse verranno poi lette serialmente e scaricate per essere pronte per una nuova scrittura.

Si ha quindi bisogno sostanzialmente di tre segnali di clock: uno per la scrittura, uno per la lettura ed uno per l'azzeramento o scarica delle capacità.

In questo modo è possibile variare il range dinamico semplicemente cambiando la lunghezza del periodo di scrittura delle celle di memoria.

Per la costruzione del *cubo magico*, si è pensato di adottare questa soluzione e si sono prese in considerazione due diverse logiche di APM:

- APC128 [30]
- HARP [29]

### 3.3 APC128

L'APC128 [30] è una memoria analogica vettorizzata che contiene 128 canali, ognuno dei quali possiede 32 celle (capacità di memoria, *deep* della *pipeline*).

Ogni canale consiste di un integratore di carica come quello già descritto, ma con due possibili valori della capacità di carica, selezione che permette di incrementare il range dinamico del sistema. La corrente in ingresso viene dunque integrata su queste capacità e successivamente la loro tensione viene trasferita alle capacità di memoria, che possono quindi essere più piccole, riducendo la superficie occupata sul chip (si veda la figura 3.4).

Infatti per le *proprietà del nodo di somma*, quando l'amplificatore operazionale è inserito in un circuito con reazione negativa e se il suo guadagno  $A_0$  (ad anello aperto) è molto grande, per un valore finito della tensione di uscita la tensione di ingresso deve tendere a zero, poichè:

$$V_{in} = -\frac{V_{out}}{A_0}$$

Dunque in prima approssimazione  $V_{in} \sim 0$  e la carica integrata su  $C_1$  (o  $C_1 + C_2$ , si veda figura 3.4) produce una tensione che può essere direttamente trasferita (con gli switches  $SR$ ) su  $C_p$  la quale ha un capo collegato a terra.

Ogni cella di memoria  $C_p$  è connessa a tre transistori di switch. Due di questi ( $SR$  e  $\bar{S}\bar{R}$ ) sono utilizzati per connettere l'intera pipeline all'output o all'input dell'amplificatore operazionale. Il terzo transistor è utilizzato come switch sia per la scrittura che per la lettura delle celle  $C_p$ . In questo modo si assicura che gli effetti dovuti alla iniezione di carica (per le capacità parassite dei transistor)

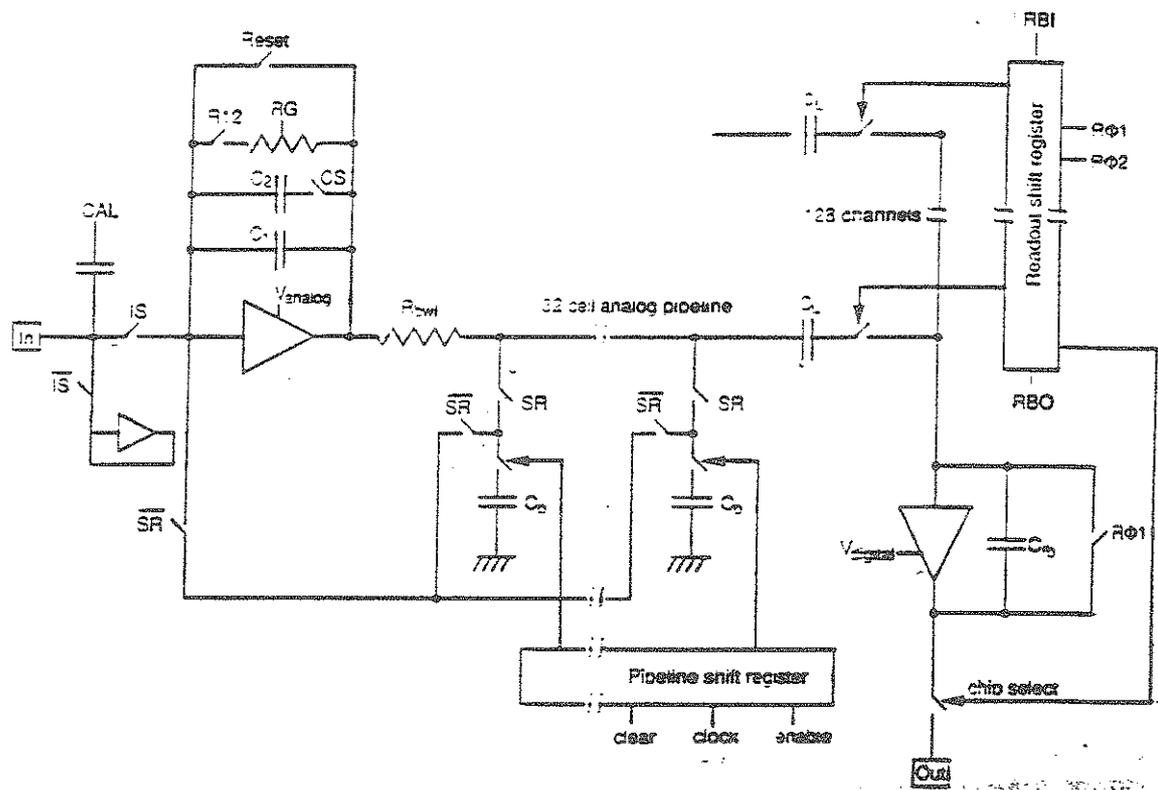


Figura 3.4: Circuito schematico dell'APC128 [30].

siano gli stessi nei due casi.

L'apertura o la chiusura degli switches è eseguita da un meccanismo di controllo digitale che è implementato sul chip stesso. Occorre semplicemente fornire dall'esterno i clock di scrittura, lettura e azzeramento. Inoltre l'amplificatore in ingresso è lo stesso sia per la scrittura che per la lettura, ed essendo in feedback invertente si ha una doppia inversione del segnale e quindi si annullano (in prima approssimazione) le piccole tensioni di offset dovute all'iniezione di carica nei transistor di switch.

Il controllo della pipeline è capace di indirizzare parecchie celle (buffers) allo stesso tempo, e quindi permette di scrivere e leggere parallelamente; leggere la tensione di due celle contemporaneamente equivale a sommare i segnali; è quindi possibile compiere operazioni logiche sui segnali.

L'uscita è "multiplexata" nel senso che una unica uscita permette la lettura dei 128 canali  $\times$  32 celle sequenzialmente.

La velocità del chip è stata adattata al suo utilizzo nell'esperimento HERA ed il chip è stato testato fino a 10MHz in scrittura e 2.5MHz in lettura [30]. Il noise (rumore) è di  $675e^- + 28e^-/pF$  di capacità in ingresso.

Nella figura 3.5 si può vedere il chip APC128 con i contatti di bonding del package che lo contiene (Pin Grid Array). L'APC128 è stato prodotto con un processo SACMOS2 (Self Aligning Contact CMOS); questo processo (VLSI) ha una lunghezza minima di gate di  $2 \mu\text{m}$  (che permette di rendere compatto il layout del circuito); il preamplificatore, così come la pipeline, hanno uno spessore (nel layout) di appena  $44 \mu\text{m}$ . Ogni cella di memoria ( $C_p = 0.46 \text{ pF}$ ) richiede una profondità di  $50 \mu\text{m}$ .

Come si può vedere nella figura 3.6 l'intero chip è lungo  $6.3 \text{ mm}$  e largo  $3.5 \text{ mm}$ .

### 3.4 HARP

L'HARP [29] è una memoria analogica vettorizzata (APM), simile all'APC128, proposta per l'elettronica di front-end degli esperimenti di collisione adronica (LHC al CERN) in cui si hanno rate di  $66 \text{ MHz}$ . Nella sua veste definitiva ha 128 celle di memoria per 32 canali, mentre all'Università di Torino abbiamo un prototipo sperimentale di soli 4 canali per chip.

Lo schema di base di un singolo canale è riportato nella figura 3.7. Si può vedere che le capacità della memoria (disposte in linea) sono tutte collegate in feedback agli amplificatori di lettura e scrittura. Quindi sono le stesse capacità di memoria ad integrare la corrente che arriva dal rivelatore (architettura differente rispetto all'APC128). Inoltre si hanno due amplificatori distinti per la lettura e la scrittura che permettono l'esecuzione di queste due operazioni indipendentemente.

Ogni cella di memoria consta di una capacità del valore di  $0.8 \text{ pF}$  e da 5 transistor che fungono da switches per indirizzare la scrittura, la lettura e l'azzeramento della cella. Dato che la corrente di leakage degli switches (che fa scaricare le capacità di memoria) è molto bassa, il segnale può essere letto dopo tempi molto lunghi (molti  $\mu\text{s}$  dopo la scrittura).

Il transistor di azzeramento cancella la carica appena prima che inizi una nuova integrazione. Il range dinamico dipende dal valore della capacità di memoria, dall'escursione della tensione ai suoi capi (che dipende dalla corrente in ingresso e dalle alimentazioni dell'amplificatore operativo) e dal rumore di fondo (noise) del sistema. I dati immagazzinati vengono (come nell'APC128) multiplexati e inviati serialmente (canale  $\times$  profondità) all'uscita del chip. Le celle di memoria occupano un'area di

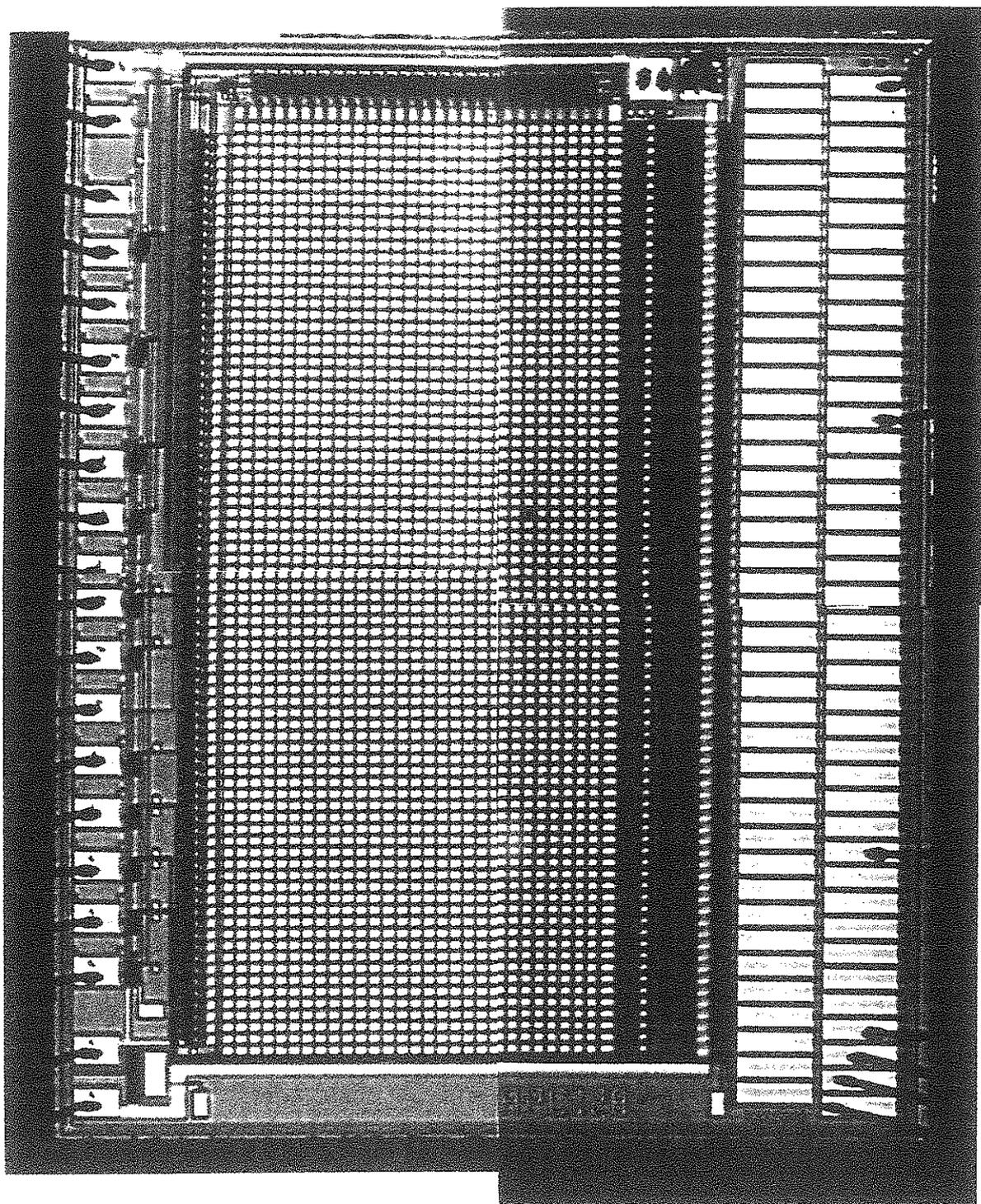


Figura 3.5: Layout del chip APC128.

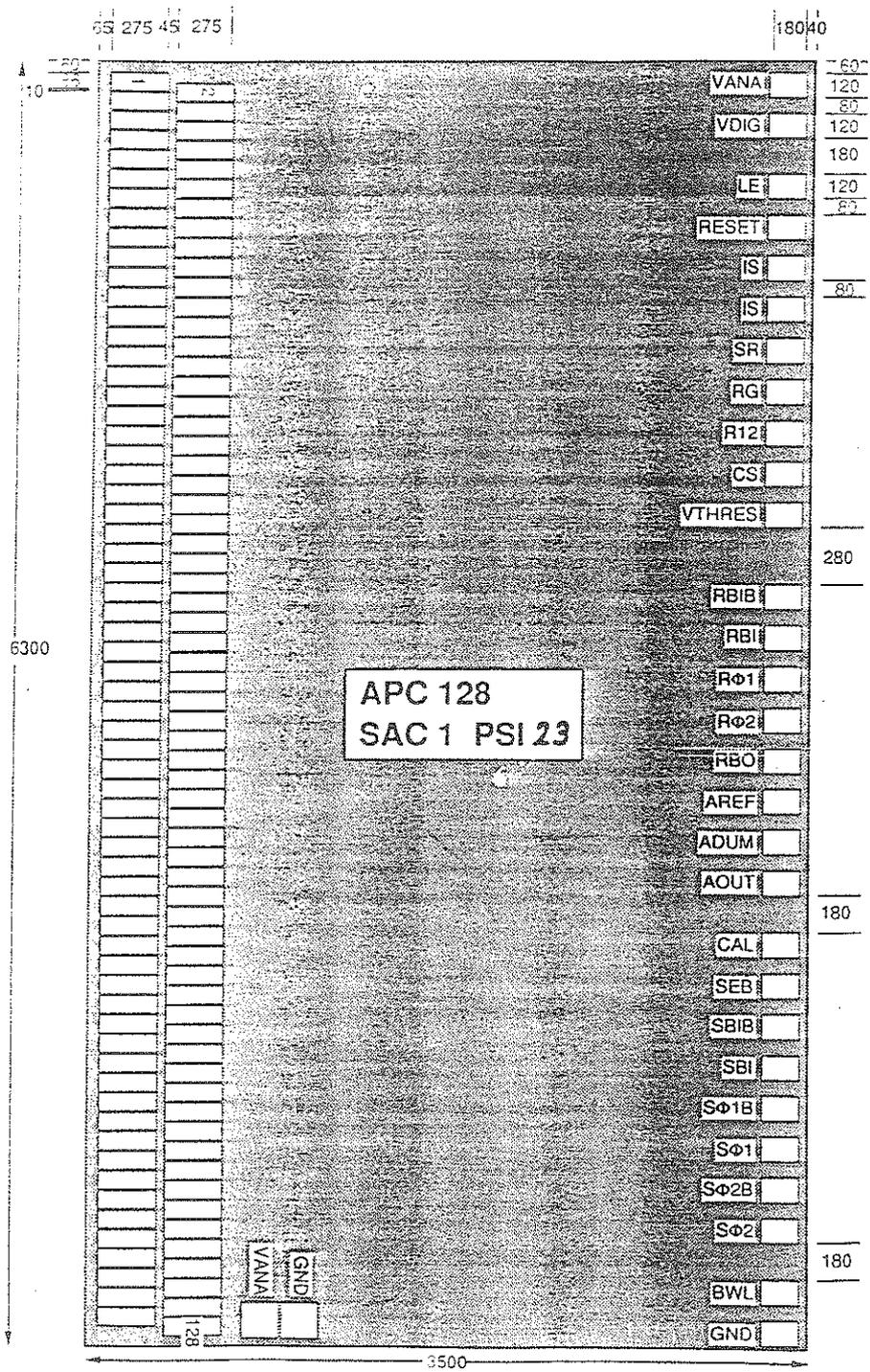


Figura 3.6: Layout del chip APC128: nomenclatura dei contatti e dimensioni

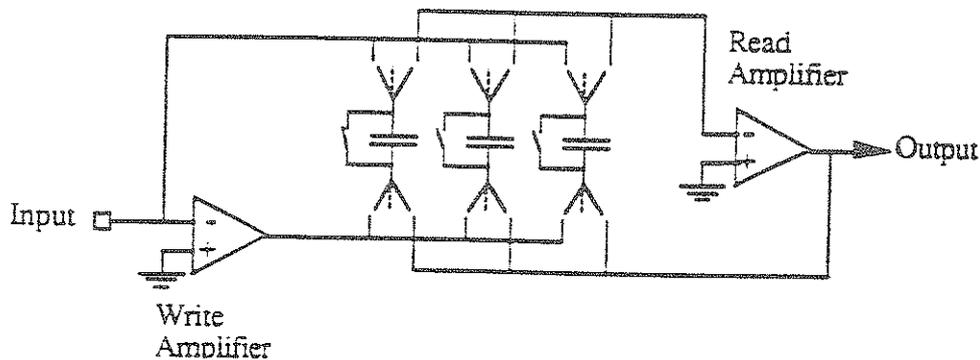


Figura 3.7: Schema circuitale di base del chip HARP [29] (sono riportate solo 3 celle ma in realtà sono 32 per ogni canale).

$60\mu\text{m} \times 100\mu\text{m}$  (più  $60\mu\text{m}$  di lunghezza per i 5 NMOS di switch per ogni cella), avendo utilizzato un processo CMOS  $1.5\mu\text{m}$  in tecnologia Mietec, che ha due layer di polisilicio e di metallizzazione.

Il chip è stato testato [29] con un clock di scrittura di 66 MHz e di 2.5 MHz di lettura. In queste condizioni si hanno  $\sim 2000 e^-$  rms di rumore ed il consumo per ogni canale è di  $\sim 2.5$  mW.

Per l'utilizzo dell'HARP nel front-end del *cubo magico* si sono fatti dei test a bassa frequenza (kHz) e si è verificata la linearità, il rumore e la conservazione del segnale per lungo tempo [31]. Dalla figura 3.8 si vede che si ha una buona linearità tra  $-500$  mV e  $+880$  mV. Si è trovato inoltre che il noise per tempi di integrazione fino a 1 ms è all'incirca  $4000e^-$  (rms).

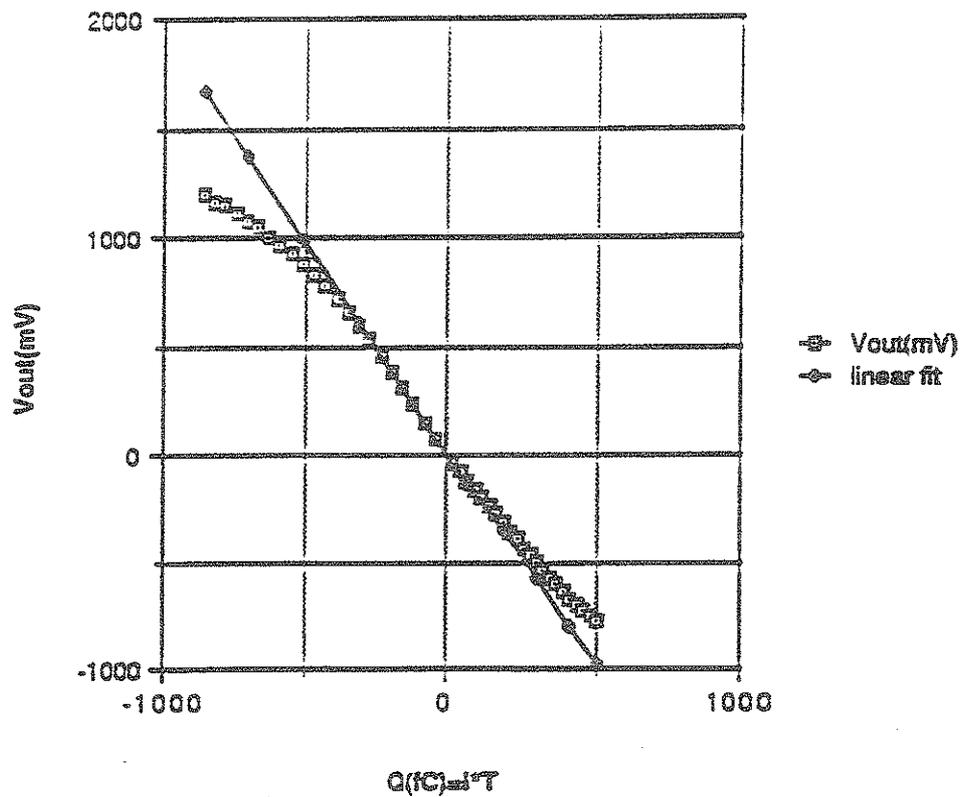


Figura 3.8: Grafico riportante  $V_{out}$  in funzione di  $Q_{in} = I \cdot \Delta t$  dove  $I$  è la corrente in ingresso e  $\Delta t$  è il tempo di integrazione [31].

### 3.5 Configurazione possibile dello stadio di ingresso

Dato che il rivelatore manda una corrente molto piccola in uscita, occorre minimizzare le possibilità di leggere del rumore da parte dell'elettronica di front-end, altrimenti le misure saranno deteriorate. L'amplificatore dovrà essere il più vicino possibile al rivelatore e non si può quindi pensare di utilizzare usuali schedine di amplificazione. Occorre passare ad un sistema molto integrato (VLSI) per poter avere tutta l'elettronica necessaria ad una camera di rivelazione a qualche centimetro di distanza. Come già derivato nel paragrafo 3.1.1 il range dinamico dovrà essere di  $\sim 10^3$  ed una soluzione possibile è la memoria analogica pipeline. Ipotizzando di utilizzare una configurazione simile a quella del chip HARP, proviamo a delinearne le principali caratteristiche.

Assumendo che il preamplificatore della pipeline sia lineare tra 40 e 800 mV e che il rumore sia molto più piccolo di 40 mV (assunzione ragionevole, infatti con l'HARP si sono misurati al massimo 15 mV di noise), si ha che il range dinamico è almeno  $800/40 = 20$ . Dato che si richiede un range di  $10^3$ , il rimanente fattore 50 del range può essere ottenuto variando il gate di scrittura delle celle di memoria. Solitamente gli switches (transistor) di scrittura, lettura e azzeramento della cella vengono azionati con un segnale (sul gate) che è sensibile ad un fronte (di salita o di discesa) del clock rispettivo. Ciò significa che il periodo del clock determina il tempo dell'operazione.

Si otterrebbe una configurazione più elastica se si potesse azionare gli switches di scrittura con entrambi i fronti, salita e discesa del write-clock. Questo permetterebbe di aprire l'integrazione per una percentuale variabile (e non per la totalità) del periodo del clock di scrittura.

Se la pipeline dell'ipotetica APM avesse 32 celle di memoria (come nella configurazione dell'APC128), la lunghezza massima del tempo di integrazione sarebbe di  $250\text{ms}/32 \sim 8\text{ms}$ . Questo perchè, come si era visto nel paragrafo 1.4.1 le specifiche dei fasci terapeutici per protoni richiedono un tempo di estrazione di 250ms con un duty cycle del 50%. Dato che occorre recuperare un range dinamico di 50 dalla variazione dei tempi di integrazione, si ha che il tempo minimo di lettura non deve essere superiore a  $8\text{ms}/50 = 160\mu\text{s}$ .

Le capacità delle celle di memoria si possono allora calcolare come segue. Il tempo più lungo di integrazione di 8ms verrà utilizzato quando si ha da rivelare

la corrente più piccola che, come si può leggere dalla tabella 3.1, vale 0.88 nA per il sistema a *diffusione passiva* e 4.16 nA per la *scansione magnetica*. Assumendo che il segnale in uscita sia di 800mV, la capacità della cella dovrà dunque essere:

$$C_{cella} = \frac{Q}{V} = \frac{i \cdot \Delta t}{V} = \frac{i \cdot 8ms}{800mV}$$

Sostituendo si ha:

$C_{cella} \sim 8.8$  pF per la *diffusione passiva* e

$C_{cella} \sim 40$  pF per la *scansione magnetica*.

Quest'ultimo valore è però da ridimensionare. Infatti in un trattamento che utilizzi la *scansione magnetica*, il fascetto di protoni si muove ad una velocità  $v_{fascio}$  di circa 10m/s. Si ha dunque che il fascio illumina ogni strip per il tempo  $\sim \text{Diametro}_{fascio} / v_{fascio} \sim 1ms$ . Confrontato con gli 8ms di integrazione possiamo dire che la carica totale che si dovrà accumulare sulla cella di memoria è  $\sim 8$  volte più piccola di quella calcolata precedentemente e quindi:  $C_{cella} \sim 5$  pF, non molto differente dal sistema a *diffusione passiva*.

### 3.5.1 Soluzione possibile per l'acquisizione dei dati

Pensando al *cubo magico* come ad un rivelatore composto da una ventina (10 piani bidimensionali) di camere a strip a ionizzazione, ciascuna composta di 100 canali, si hanno in totale  $\sim 2000$  canali da acquisire. Come precedentemente detto, il tempo di integrazione minimo dovrebbe essere di  $160 \mu s$  e quindi nella situazione più sfavorevole si hanno 2000 segnali da leggere in  $160 \mu s$  per poter campionare l'intero spill (che è lungo 250 ms) senza tempi morti. Dato che un convertitore analogico digitale (ADC) trasforma ogni segnale in una decina di bits, saranno necessari almeno 2 bytes (16bits) per ogni canale e quindi il massimo numero di bytes da leggere per ogni spill sarà:

$$N_{bytes_{max}/spill} = 2000 \cdot 2 \cdot \frac{250ms}{160\mu s} \sim 6.3Mb/spill$$

mentre il minimo numero di bytes per spill si avrà con il massimo tempo di integrazione, cioè di 8 ms:

$$N_{bytes_{min}/spill} = 2000 \cdot 2 \cdot \frac{250ms}{8ms} = 125kb/spill$$

Pensando di dover trasferire queste informazioni nel tempo di interspill, e cioè in 250 ms (circa uguale alla lunghezza dello spill), si ha:

- $N_{bytes_{max}}/sec = 25.2 \text{ Mb/s}$ .

- $N_{bytes_{min}}/sec = 500 \text{ kb/s}$ .

Per compattare il sistema *cubo magico*, ogni chip dovrebbe essere sufficiente ad una camera di rivelazione e dovrà quindi avere circa 100 canali, ognuno composto da 32 celle di memoria (in questa ipotetica configurazione). Questa matrice  $100 \times 32$  dovrà poi essere multiplexata all'interno del chip che possiederà dunque una unica uscita seriale (come le APM considerate). Questa sarà direttamente collegata ad un convertitore analogico digitale (ADC) molto veloce.

In commercio vi sono Flash ADC (FADC) da 10 bits a canale che funzionano a 40 MHz. Per ogni FADC si avranno al massimo:

$$100 \times \frac{250ms}{160\mu s} \times 2 = 312500$$

conversioni al secondo, ognuna delle quali richiederà 2 bytes come già spiegato. Saranno dunque necessari 20 canali di FADC comandati da dei processori di dati digitali (DSP) che possono controllarne 4. La quantità di dati che saranno gestiti dal DSP sarà:

$$312500 \times 2(\text{bytes/conv}) \times 4 \simeq 2.5 \text{ Mb/s}$$

che è dunque un valore ragionevole se confrontato con il limite di 20 Mb/s degli usuali DSP.

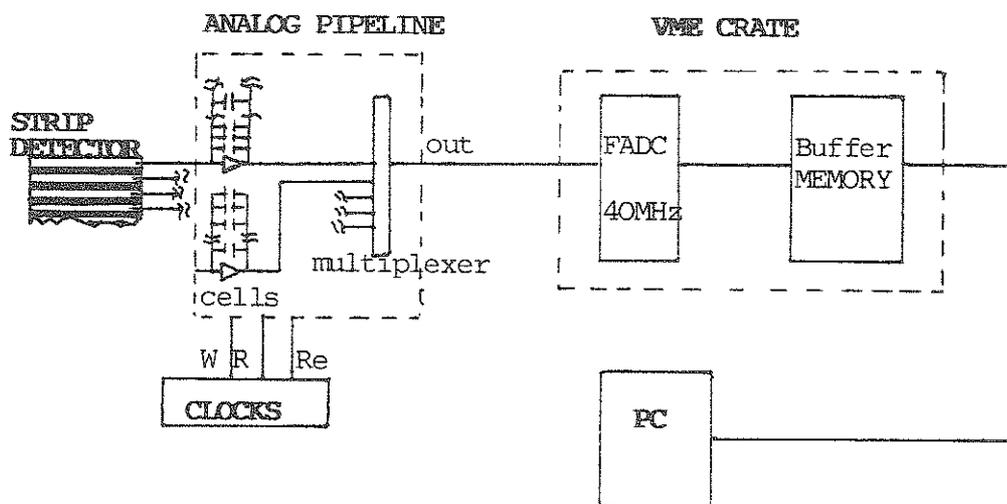


Figura 3.9: Schema complessivo dell'acquisizione dei dati.

## Capitolo 4

# Introduzione ai sistemi VLSI

Ciò che oggi chiamiamo comunemente tecnologia VLSI ha avuto una evoluzione storica caratterizzata da parecchi eventi. Fino agli anni '50 la tecnologia dei dispositivi elettronici attivi era dominata dai tubi al vuoto e, sebbene avesse avuto inizio una certa misura di integrazione circuitale, la tecnologia non permetteva una miniaturizzazione come la si concepisce oggi.

L'inizio pratico dell'industria microelettronica si deve all'invenzione del transistor a giunzione bipolare (BJT), nel 1947, nei laboratori della Bell.

Nel 1958 la Texas Instruments inventa il primo circuito integrato, e dagli anni '60 ad oggi possiamo già contare quattro generazioni di circuiti integrati.

Il numero di transistors commercialmente integrati su un singolo chip è passato da 2 a più di 500000 in sole due decadi, come si può vedere nella figura 4.1 [41].

Il disegno e la simulazione di circuiti, contenenti un grande numero di dispositivi, è così complesso che ha potuto svilupparsi solo grazie all'aiuto del calcolatore (Computer Aided Design: CAD).

Nel campo della microelettronica si possono trovare molti "processi" differenti. Nella figura 4.2 sono riportati i processi che hanno ricevuto un ragionevole grado di utilizzo [42].

La prima divisione tra i processi si ha tra i substrati attivi ed inerti, dove i primi vengono utilizzati tipicamente per circuiti integrati di grande volume, e gli altri nella maggior parte dei circuiti integrati ibridi. Il substrato attivo è costituito generalmente da silicio (o silicio drogato) sebbene si sia compiuto uno sforzo considerevole della ricerca per l'utilizzo dell'arseniuro di gallio GaAs.

Due processi di fondamentale importanza utilizzano substrati attivi: il processo bipolare, che utilizza il BJT come dispositivo attivo di base, e il processo MOS.

Year	1947	1950	1961	1966	1971	1980	1985	1990
Technology	<i>Invention of the transistor</i>	<i>Discrete components</i>	<i>SSI</i>	<i>MSI</i>	<i>LSI</i>	<i>VLSI</i>	<i>ULSI*</i>	<i>GSII†</i>
Approximate number of transistors per chip in commercial products	1	1	10	100-1000	1000-20,000	20,000-500,000	> 500,000	> 10,000,000
Typical products	—	Junction transistor and diode	Planar devices Logic gates Flip-flops	Counters Multiplexers Adders	8 bit micro-processors ROM RAM	16 and 32 bit micro-processors Sophisticated peripherals	Special processors Real time image processing	?

\* Ultra large-scale integration

† Giant-scale integration

Figura 4.1: Evoluzione della Microelettronica. Si tenga presente che il passaggio da una tecnologia all'altra richiede una nuova metodologia di disegno e simulazione [41].

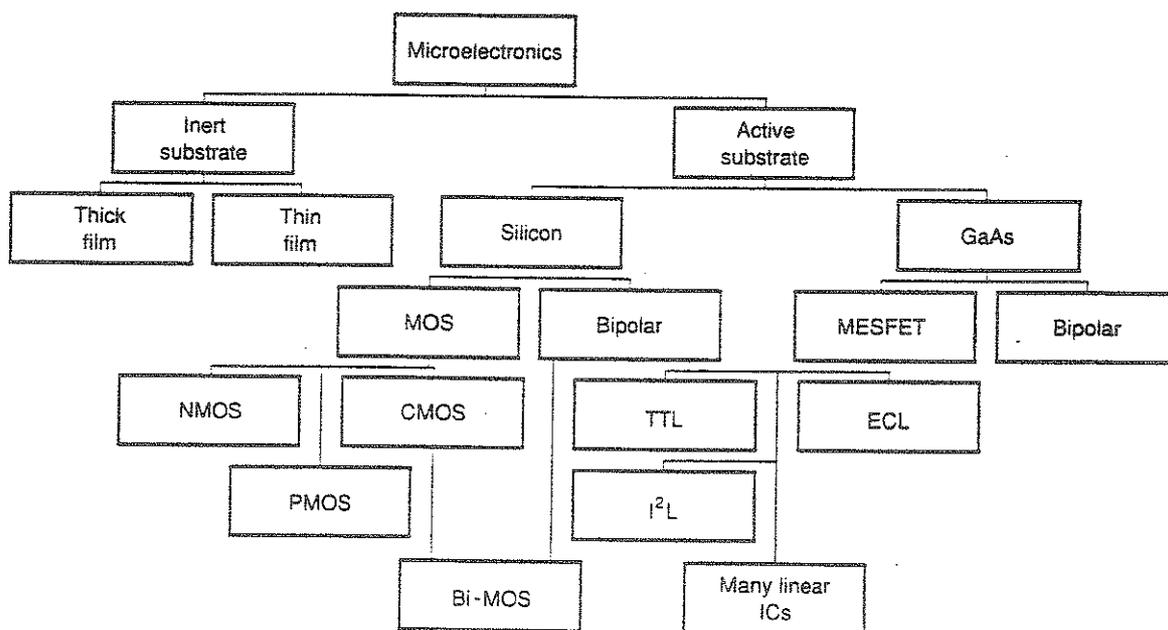


Figura 4.2: Processi maggiormente utilizzati nella fabbricazione di circuiti integrati [42].

che utilizza il MOSFET. Metallo Ossido Semiconduttore Transistor ad Effetto Campo. (a volte chiamato IGFET. Insulated Gate Field Effect Transistor) come dispositivo attivo di base.

Il processo bipolare è stato il più importante tra gli anni '60 e '70. È caratterizzato dalla possibilità di lavorare ad alte frequenze e possiede elevati valori di transconduttanza, che costituisce un vantaggio in molte applicazioni. La dissipazione termica è però abbastanza elevata e la densità dei dispositivi non è così alta come quella che si può raggiungere con il processo MOS.

Il processo MOS è spesso suddiviso in tre categorie: NMOS, PMOS e CMOS. Gli elementi base del processo MOS sono i transistor MOSFET a canale  $n$  e  $p$ . Il termine PMOS si riferisce ad un processo MOS che utilizza solo transistori a canale  $p$ , mentre il termine NMOS si riferisce ad un processo che utilizza solo transistori a canale  $n$ .

Entrambi i dispositivi possono essere ad arricchimento (enhancement), su cui si agisce creando il canale di carica ( $n$  per l'NMOS e  $p$  per il PMOS), o a svuotamento (depletion) su cui si agisce svuotando il canale di carica.

Dato che la mobilità delle cariche elettriche nei materiali di tipo  $n$  è molto maggiore rispetto a quelli di tipo  $p$ , il processo PMOS è tutt'oggi raramente utilizzato. Viceversa il processo NMOS è caratterizzato da una eccellente densità (di dispositivi) e buone prestazioni.

Con il termine CMOS (complementary MOS) si indica il processo MOS che fornisce entrambi i dispositivi, a canale  $n$  e  $p$ , il che offre una grande flessibilità nella progettazione dei circuiti. Nelle applicazioni digitali, la possibilità di avere dispositivi complementari permette un consumo energetico statico molto basso.

Il processo MOS è utilizzato per la maggior parte dei circuiti ad Largesima Scala di Integrazione (VLSI). Esso trova applicazione nel disegno di memorie, interfacce, microprocessori, etc. .

A scopo di completezza occorre aggiungere che recentemente si sono compiuti molti sforzi per poter combinare assieme i dispositivi bipolari (di cui si hanno ormai pacchetti logici insostituibili) e quelli MOS in un unico processo: il processo Bi-MOS, ancora complesso e costoso, ma che trova ormai sempre più diffuse applicazioni.

## 4.1 Fabbricazione dei dispositivi attivi in processo MOS

La fabbricazione di circuiti integrati in processi MOS, parte dal *wafer* (o *slice*) che è l'unità fisica base del processo, generalmente circolare, del diametro tipico di  $110 \div 150$  mm (4.5 e 6 pollici) e spesso  $\sim 0.4$  mm, su cui si disegnano, fotolitograficamente, parecchi circuiti integrati identici [42].

Per semplicità si consideri il processo NMOS, in cui si voglia creare un transistor a canale  $n$  ad arricchimento.

In riferimento alla figura 4.3 [41] il processo di fabbricazione può essere spiegato come segue:

1. Anzitutto il *wafer* viene tagliato da un unico cristallo di silicio di elevata purezza nel quale sono state introdotte le richieste impurezze  $p$  (ad esempio Boro con una concentrazione dell'ordine di  $10^{15} \div 10^{16}$  atomi/cm<sup>3</sup>) durante la formazione del cristallo stesso.
2. Si fa crescere, sulla superficie del *wafer*, un sottile strato di ossido di silicio, SiO<sub>2</sub>, tipicamente dell'ordine di  $1 \mu\text{m}$ , per proteggerla dal processo di drogaggio e per creare un substrato isolante su cui possono essere disposti successivi strati.
3. La superficie viene ora ricoperta di "fotoresist".
4. Il foglio di fotoresist viene quindi esposto alla luce ultravioletta attraverso una maschera che definisce le regioni in cui si vuole la diffusione insieme al canale del transistor (che sarà all'interno della regione di diffusione). Le aree esposte alla luce UV si polimerizzano, indurendosi, mentre le aree richieste per la diffusione rimangono inalterate.
5. Queste aree di fotoresist protette dalla luce UV, vengono ora asportate insieme al sottile strato di SiO<sub>2</sub> sottostante, riportando alla luce la superficie del *wafer*.
6. Si rimuove il rimanente fotoresist e si fa crescere uno strato molto sottile, dell'ordine di  $0.1 \mu\text{m}$ , di ossido di silicio sull'intera superficie del chip. Poi si deposita il polisilicio su questo ossido, per formare la pista di *gate* del transistor. Quest'ultimo strato consiste di polisilicio molto drogato depositato da vapori chimici (CVD).
7. Quindi si esegue un nuovo processo di mascheramento con fotoresist e successivamente si rimuove il sottile strato di ossido. Ora, quindi, si possono diffondere con impurità di tipo  $n$ , le zone che dovranno creare il *source* ed il *drain* del tran-

sistor. La diffusione si esegue riscaldando il *wafers* ad elevata temperatura ed immergendolo in un gas contenente l'impurità desiderata. Occorre notare che il polisilicio ed il sottostante ossido, agiscono da maschera durante il processo di diffusione (il processo è auto-allineante, Self Aligning).

8. Un sottile strato di ossido è fatto crescere su tutta la superficie, viene mascherato con il fotoresist e poi rimosso in modo da esporre le aree in cui si dovranno eseguire le connessioni con il gate, drain e source.

9. L'intero chip viene ricoperto con il metallo (alluminio) di collegamento, di uno spessore di circa  $1\ \mu\text{m}$ . Questo foglio di metallo viene poi mascherato e rimosso per dare origine al circuito desiderato.

L'intero processo è dunque basato sulla formazione o deposito di tre strati, separati dall'ossido di isolamento: lo strato di diffusione all'interno del substrato, quello del polisilicio e quello del metallo.

La struttura che si viene a creare è riportata in figura 4.4(i), che si riferisce all'istante descritto nel punto 7. del processo di fabbricazione [44].

## 4.2 Il transistor MOS

Il dispositivo NMOS ad arricchimento riportato in figura 4.4 conduce in modo apprezzabile tra drain e source solo quando sotto il gate si forma un canale  $n$ , da cui la denominazione di MOS a *canale n*. Il termine *ad arricchimento*, come spiegato all'inizio di questo capitolo, si riferisce al fatto che si agisce sul dispositivo creando questo canale  $n$ , e quindi non si ha conduzione per tensione di gate nulla ( $V_{GS} = 0$ )<sup>1</sup>.

Facendo riferimento alla figura 4.5 si possono ricavare le caratteristiche di trasferimento, per grandi segnali, di questo dispositivo [39].

Quando tutti i generatori di tensione sono spenti (cortocircuitati), le regioni di source e drain sono separate tra loro da giunzioni  $pn$  contrapposte. Tali giunzioni sono create tra le zone  $n$  di source e drain ed il substrato  $p$ , e danno luogo ad una resistenza molto alta ( $\sim 10^{12}\Omega$ ) tra drain e source quando il dispositivo è

---

<sup>1</sup>Oltre a questo dispositivo, nel processo NMOS si ha anche il transistor detto a *svuotamento*, che possiede un canale "impiantato" tra drain e source, cosicché si ha conduzione anche per  $V_{GS} = 0$ ; si agisce sul dispositivo *svuotando* il canale impiantato. Usualmente trova applicazione nei circuiti come carico attivo.

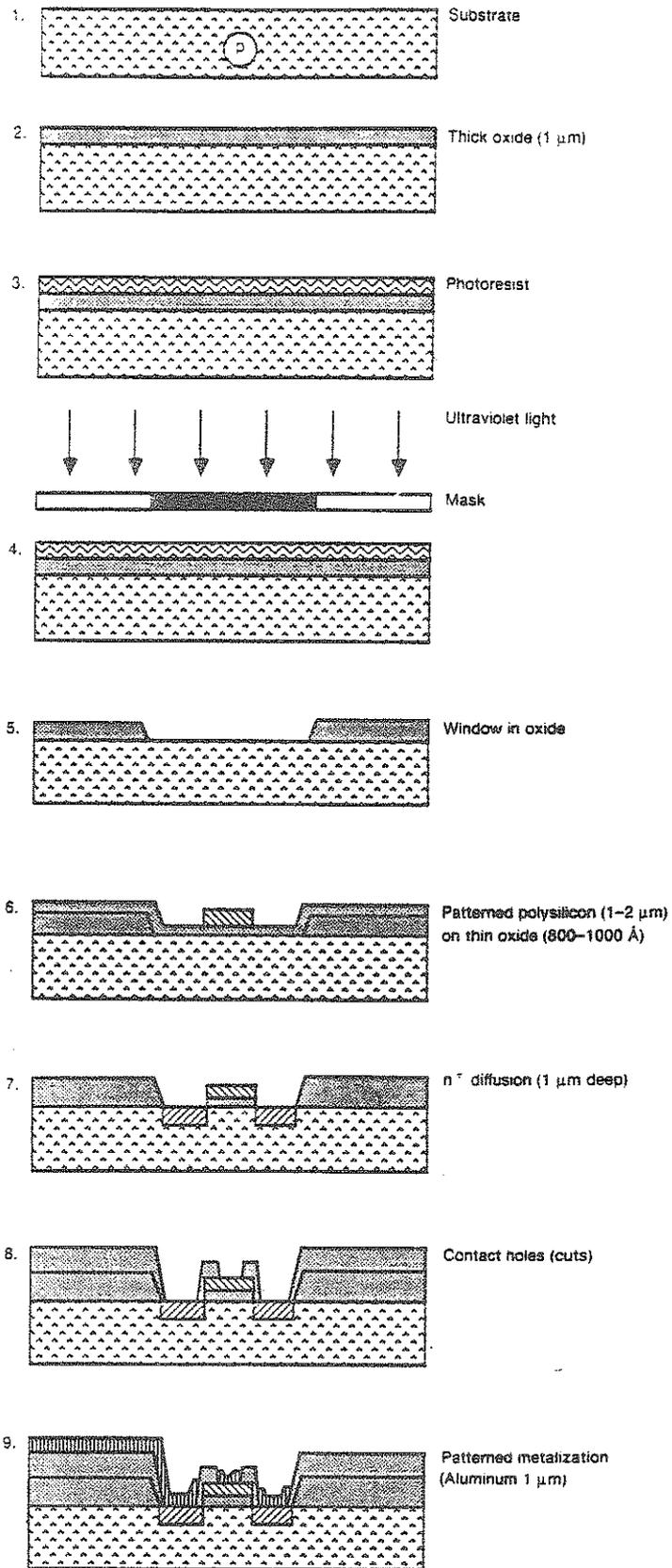


Figura 4.3: Processo di fabbricazione di MOS di un transistor a canale  $n$  ad arricchimento [41].

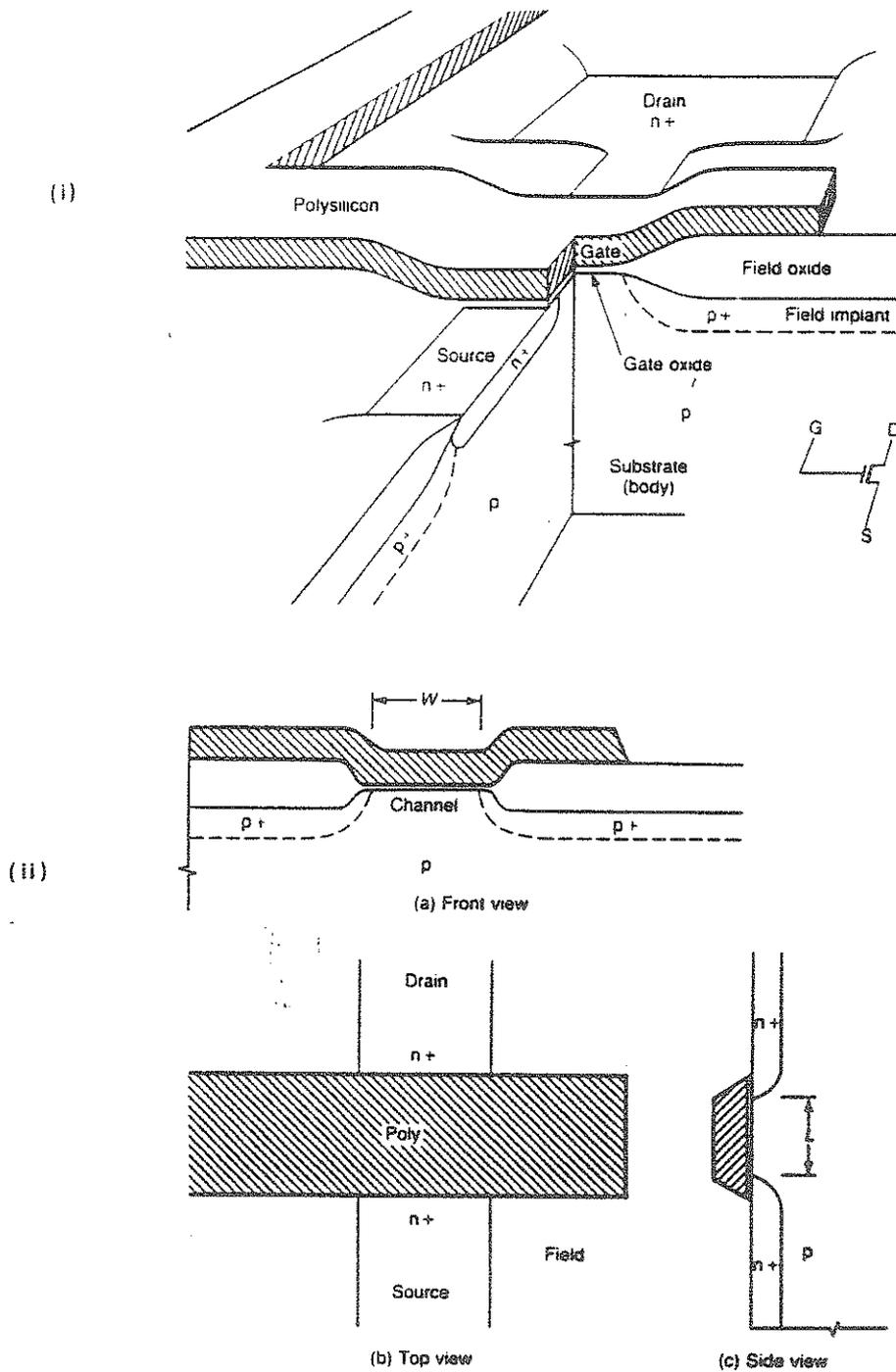


Figura 4.4: (i). Spaccato di un transistor MOS a canale  $n$  riferito al punto 7. del processo di fabbricazione riportato in figura 4.3: a destra è riportato il simbolo circuitale del dispositivo. (ii). Tre viste di un transistor MOS a canale  $n$ : (a) frontale. (b) dall'alto. (c) laterale [44].

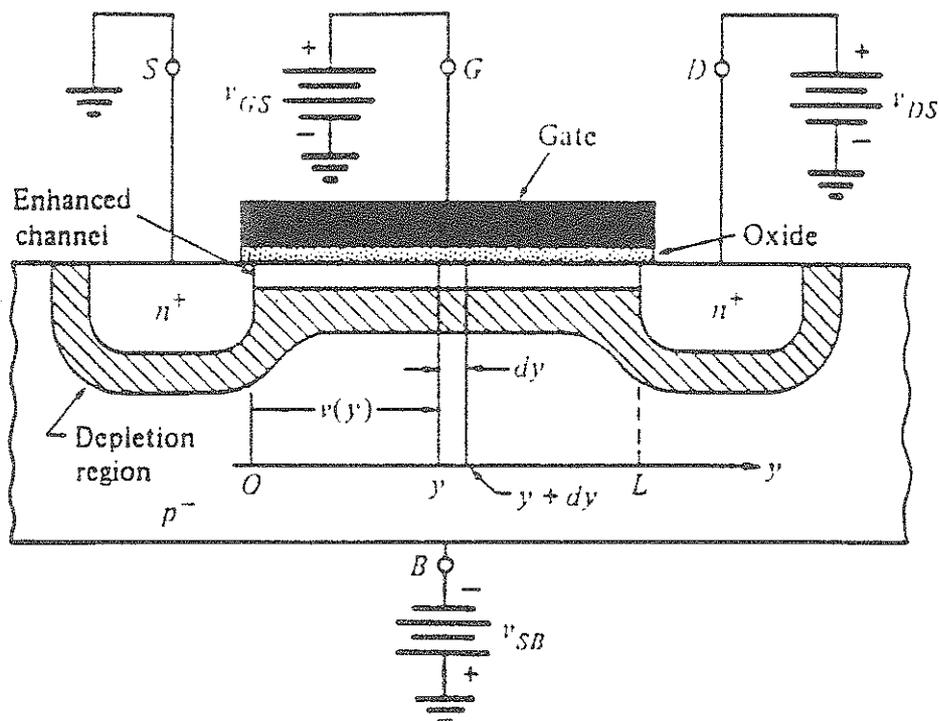


Figura 4.5: Sezione di un transistor MOS a canale  $n$  con piccola tensione  $V_{DS}$ , e  $V_{GS} > V_{TH}$  [39].

interdetto.

Si applichi ora una tensione positiva  $V_{GS}$  tra gate e source. Gate e substrato costituiscono le armature di un condensatore piano con lo strato di  $\text{SiO}_2$  che funge da dielettrico. La carica positiva si accumula sul gate e quella negativa nel substrato. Inizialmente la carica negativa nel substrato di tipo  $p$  si manifesta con la formazione di una zona di svuotamento: gli elettroni vanno a "colmare" le lacune in modo che non si hanno più cariche libere. Quando però il potenziale nel silicio raggiunge un valore critico pari al doppio del livello di Fermi  $\phi_f \simeq 0.3 \text{ V}$ , si manifesta un fenomeno noto col termine di *inversione*. Un ulteriore aumento della tensione di gate non produce alcun cambiamento nello spessore dello strato di svuotamento, e viene invece indotto un sottile strato di elettroni nella zona di svuotamento immediatamente al di sotto dell'ossido. Si forma così una zona di tipo  $n$  continua che collega il drain al source e che costituisce il canale di conduzione. Questo canale può quindi venire modulato aumentando o diminuendo la tensione di gate.

La tensione necessaria a creare uno strato di inversione è chiamata *tensione di soglia*  $V_{TH}$ , ed è così definita <sup>2</sup>:

$$V_{TH} = V_{T0} + \gamma(\sqrt{2\phi_f + V_{SB}} - \sqrt{2\phi_f}) \quad (4.1)$$

<sup>2</sup>La derivazione teorica della  $V_{TH}$  (come di molte altre grandezze esaminate) non viene qui eseguita poiché esula dal contesto di questa discussione.

dove:

$V_{T0}$  è la tensione di soglia con tensione di source-bulk nulla.  $V_{SB} = 0$ , e viene di solito regolato durante il processo di fabbricazione impiantando impurezze aggiuntive nella regione di canale:  $V_{T0}$  è compresa nell'intervallo 0,5÷1,5 V per transistor a canale  $n$  ad arricchimento:

$\gamma$  è il parametro di soglia del substrato (bulk threshold parameter) e dipende dalla densità dei droganti del substrato (nel nostro caso di tipo  $p$ ), e dalla densità di capacità dell'ossido  $C'_{ox}$ :

$$C'_{ox} = \frac{\epsilon_{ox}}{t_{ox}} \quad (4.2)$$

essendo  $\epsilon_{ox}$  e  $t_{ox}$  rispettivamente la costante dielettrica e spessore dell'ossido.

Per avere una idea dell'ordine di grandezza di queste variabili si faccia riferimento alle tabelle 5.1 e 5.2, dove  $C'_{ox}$  è espressa in Farad per unità di area, e  $C_{GX0}$  in Farad per unità di lunghezza di gate.

Definita la tensione di soglia, possiamo ora calcolare la corrente che scorre nel canale arricchito di figura 4.5. Se dunque  $V_{GS}$  è maggiore di  $V_{TH}$  si forma un canale di conduzione e  $V_{DS}$  fa fluire una corrente  $I_{DS}$  tra drain e source. La presenza della tensione  $V_{DS}$ , fa sì che la tensione inversa tra drain e substrato (diodo) aumenti più di quanto non faccia tra source e substrato, e di conseguenza al terminale di drain la zona di svuotamento è più ampia. Per semplicità si assuma che la  $V_{DS}$  sia piccola, di modo che la larghezza dello strato di svuotamento si possa considerare costante.

Secondo la figura 4.5, sia  $V(y)$  la tensione rispetto al source ad una distanza  $y$  lungo il canale. Quindi la tensione gate-canale nel punto  $y$  sarà  $V_{GS} - V(y)$ . Assumendo che questa sia sempre maggiore della tensione di soglia (altrimenti non ci sarebbe il canale), la carica indotta per unità di area nel canale è data da:

$$Q_i(y) = C_{ox}(V_{GS} - V(y) - V_{TH}) \quad (4.3)$$

La resistenza del canale sarà data dall'usuale equazione per il calcolo della resistenza di conduttori di conduttività  $\sigma$  lunghi  $L$  e di sezione  $S$ , dove  $S$  nel nostro caso sarà la larghezza  $W$  moltiplicata il piccolo spessore del canale arricchito  $t_{ch}$ :

$$dR = \frac{1}{\sigma} \frac{dL}{S} = \frac{1}{\sigma W t_{ch}} dL \quad (4.4)$$

Possiamo esprimere la conduttività come il prodotto della mobilità  $\mu_n$  degli elettroni, per la loro concentrazione  $n_n$  e per la carica  $q$  del singolo elettrone [43]:

$$\sigma = \mu_n \cdot n_n \cdot q = \mu_n \cdot \rho_n$$

dove  $\rho_n$  è la densità di carica volumetrica. In prima approssimazione scriveremo

$$\rho_n \approx \frac{Q_i}{t_{ch}}$$

e quindi sostituendo nell'equazione 4.4 si ottiene:

$$dR = \frac{dy}{W\mu_n Q_i(y)} \quad (4.5)$$

La caduta di tensione lungo un tratto  $dy$  di canale non è altro che  $dV = I_{DS}dR$ , che sostituita nell'equazione 4.5 ed integrando su tutta la lunghezza del canale:

$$\int_0^L I_{DS} dy = \int_0^{V_{DS}} W\mu_n C_{ox} (V_{GS} - V - V_{TH}) dV \quad (4.6)$$

si ottiene (*zona resistiva*):

$$I_{DS} = \frac{K_n W}{2 L} (2(V_{GS} - V_{TH})V_{DS} - V_{DS}^2) \quad (4.7)$$

dove  $K_n = \mu_n C_{ox}$ . è il parametro di transconduttanza.

Al crescere della tensione  $V_{DS}$  il canale indotto sotto il gate si restringe in corrispondenza del terminale di drain, poichè aumenta la tensione di inversione, e si verifica uno stato di saturazione: ulteriori aumenti di  $V_{DS}$  provocano piccoli cambiamenti della corrente  $I_{DS}$ .

Allora per  $V_{DS} \geq (V_{GS} - V_{TH})$  si ha (*zona di saturazione*):

$$I_{DS} = \frac{K_n W}{2 L} (V_{GS} - V_{TH})^2 \quad (4.8)$$

Si può dimostrare sia teoricamente che sperimentalmente, che la corrente di drain nella regione di saturazione, aumenta leggermente, in modo approssimativamente lineare, all'aumentare di  $V_{DS}$  [42]. Questo è fisicamente dovuto ad una leggera restrizione del canale nei pressi della regione di drain, in cui la tensione inversa tende ad aumentare la zona di svuotamento. Definendo  $\lambda$  (channel length modulation

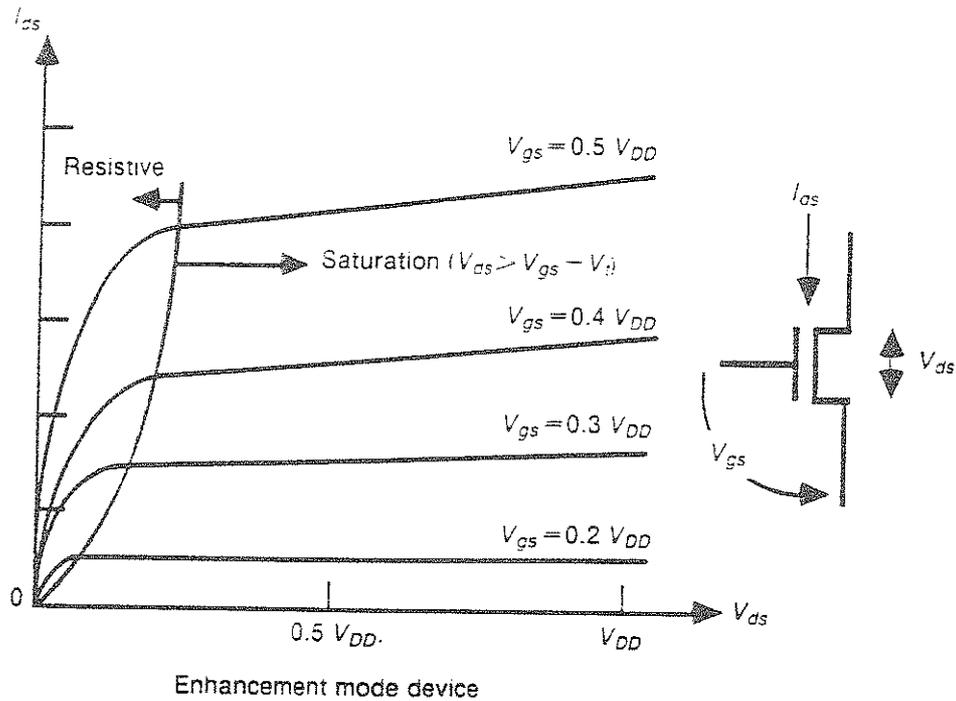


Figura 4.6: Caratteristiche di un dispositivo MOS a canale  $n$  ad arricchimento [41].

parameter) come il coefficiente che rappresenta la dipendenza lineare di  $I_{DS}$  con  $V_{DS}$  una espressione più accurata della 4.8, nella zona di saturazione, è fornita da:

$$I_{DS} = \frac{K_n W}{2 L} (V_{GS} - V_{TH})^2 (1 + \lambda V_{DS}) \quad (4.9)$$

dove  $V_{DS} > V_{GS} - V_{TH} > 0$ .

Il parametro  $\lambda$  è considerevolmente importante solo nel caso di transistori molto corti. Le curve caratteristiche  $I_{DS}(V_{DS})$  del transistor a canale  $n$  ad arricchimento, sono riportate in figura 4.6.

Prima della saturazione il dispositivo si comporta come un resistore non lineare controllato di tensione (resistive zone), mentre dopo, il dispositivo approssima il comportamento di un generatore di corrente controllato di tensione.

#### 4.2.1 Modello per piccoli segnali

Dalle equazioni finora ricavate, valide per il funzionamento dei transistor MOS con grandi segnali (dc model), si può ora ottenere il modello per piccoli segnali di questo dispositivo.

Come si è visto nel paragrafo precedentemente, esistono tre zone in cui può lavorare il transistor: una zona interdotta quando  $V_{GS} < V_{TH}$ , una zona resistiva, quando  $V_{DS} < (V_{GS} - V_{TH})$ , ed una zona di saturazione, quando  $V_{DS} > V_{GS} - V_{TH} > 0$ .

Queste regioni corrispondono ad altrettanti modelli per piccoli segnali del transistor. Non considerando la zona di interdizione in cui la corrente di drain è essenzialmente nulla, rimangono la regione resistiva e quella di saturazione. Solitamente il MOSFET non è polarizzato per lavorare nella prima regione con piccoli segnali, date le sue limitate prestazioni [42]. La maggior parte delle applicazioni per piccoli segnali utilizza il MOSFET polarizzato in modo che sia in saturazione. Come si vedrà nel capitolo 5, infatti, i transistori utilizzati comunemente in circuiti di amplificazione, lavorano tutti in saturazione. In questa regione vale dunque l'equazione 4.9. Si definisce *transconduttanza*  $g_m$  del MOSFET la variazione della corrente di drain in funzione della tensione di gate-source:

$$g_m \equiv \frac{\partial I_{DS}}{\partial V_{GS}} = K_n \frac{W}{L} (V_{GS} - V_{TH})(1 + \lambda V_{DS}) \quad (4.10)$$

Se non consideriamo la variazione  $L_s$  della lunghezza  $L$  dovuta allo svuotamento vicino alla regione di drain (cioè non consideriamo  $\lambda$ ) si ha:

$$g_m = \sqrt{2K_n \frac{W}{L} I_{DS}} \quad (4.11)$$

che fornisce la transconduttanza in funzione della corrente di drain.

Si definisce inoltre *conduttanza di uscita*  $g_{ds}$  la variazione della tensione  $V_{DS}$  rispetto alla corrente di drain, che non è altro (per come è stato definito  $\lambda$ ) che [38]:

$$g_{ds} \equiv \frac{\partial I_{DS}}{\partial V_{DS}} = \lambda I_{DS} \quad (4.12)$$

e quindi la resistenza di uscita sarà esattamente l'inverso della 4.12:

$$r_o = \frac{1}{g_{ds}} \quad (4.13)$$

Dall'equazione 4.1 si ha che la tensione source-substrato  $V_{SB}$  influenza la tensione di soglia, e quindi la corrente di drain.

Si definisce quindi una transconduttanza dovuta al bulk  $g_{mb}$  come la variazione

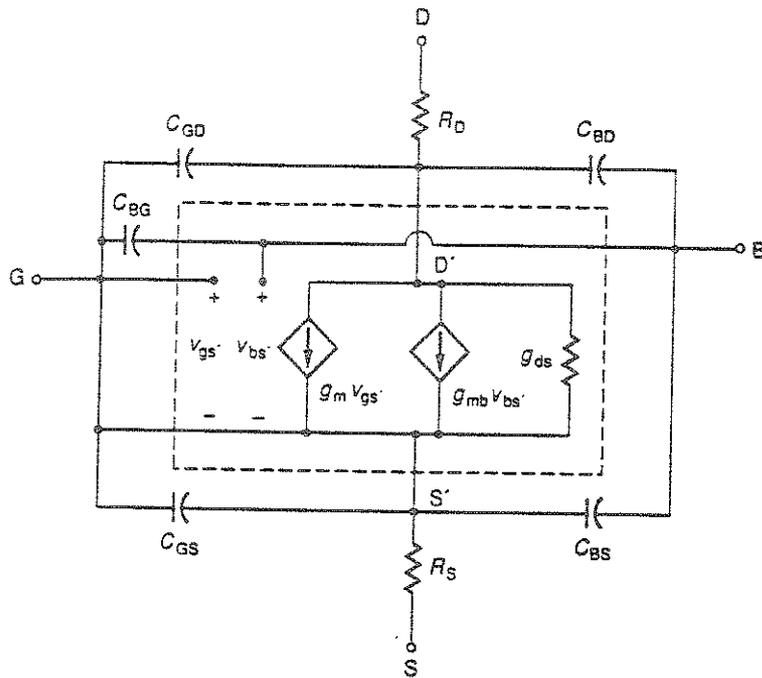


Figura 4.7: Circuito equivalente per piccoli segnali di un transistor MOSFET a canale  $n$  ad arricchimento [42].

della corrente di drain in funzione della variazione della tensione di bulk:

$$g_{mb} \equiv \frac{\partial I_{DS}}{\partial V_{BS}} = \eta g_m \quad (4.14)$$

dove  $\eta = \frac{\gamma}{2\sqrt{2\phi_f - V_{BS}}}$  [42].

Dato che la corrente di drain è funzione sia di  $V_{GS}$  che di  $V_{SB}$ , per creare un modello per piccoli segnali del transistor occorreranno due generatori a controllo di transconduttanza, come mostrato nella figura 4.7. Nel circuito sono inoltre riportate le capacità parassite tra gate-source-bulk-drain, che assumono un'importanza fondamentale quando si utilizza il dispositivo ad alte frequenze. Le resistenze  $R_D$  ed  $R_S$  della figura sono state incluse per tenere conto della resistenza ohmica della regione di diffusione tra le connessioni fisiche di drain e source (D e S), e quelle reali del dispositivo (D' e S').

Il calcolo delle capacità parassite può essere svolto in modo analogo a quello descritto nel paragrafo 3.1.2. Spesso è più conveniente combinare la costante dielettrica  $\epsilon$  con la distanza tra gli elettrodi  $d$  in un singolo parametro  $C_d = \epsilon/d$  che è una densità di capacità, come quella già calcolata nell'equazione 4.2. Allora se  $A$  è la superficie delle facce piane del condensatore si avrà:

	Region		
	Cutoff	Ohmic	Saturation
$C_{GD}$	$C_{OX}WL_d$	$C_{OX}WL_d + \frac{1}{2}WLC_{OX}$	$C_{OX}WL_d$
$C_{GS}$	$C_{OX}WL_d$	$C_{OX}WL_d + \frac{1}{2}WLC_{OX}$	$C_{OX}WL_d + \frac{2}{3}WLC_{OX}$
$C_{BG}$	$C_{OX}WL$	0	0
$C_{BD}$	$C_{BD1}$	$C_{BD1} + \frac{C_{BC1}}{2}$	$C_{BD1}$
$C_{BS}$	$C_{BS1}$	$C_{BS1} + \frac{C_{BC1}}{2}$	$C_{BS1} + \frac{2}{3}C_{BC1}$

Figura 4.8: Valori delle capacità parassite riportate nel circuito equivalente per piccoli segnali di figura 4.7.

$$C = C_d \cdot A$$

dove la  $C_d$  dipende dal processo di fabbricazione, e l'area  $A$  dipende dal dispositivo<sup>3</sup>.

Calcoliamo, a titolo di esempio, la capacità parassita tra gate e drain,  $C_{GD}$ , in regime di saturazione. Il condensatore parassita che si viene a creare è dovuto alla diffusione involontaria sotto il gate, di alcune impurità utilizzate per creare le regioni di drain e source (punto 7. della figura 4.3). Questa diffusione causa una leggera diminuzione dell'effettiva lunghezza  $L$  del transistor. Posto  $L_d$  la distanza della *diffusione laterale*, e  $W$  la larghezza del gate, possiamo scrivere:

$$C_{GD} = C_{ox}WL_d$$

Analogamente si possono calcolare le altre capacità parassite, le cui espressioni

<sup>3</sup>Occorre però notare che la capacità parassita così calcolata, dovrà essere compensata con un termine che tenga conto della variazione delle tensioni nei vari contatti del transistor, le quali determinano la variazione della zona di svuotamento, dello spessore del canale arricchito, etc.

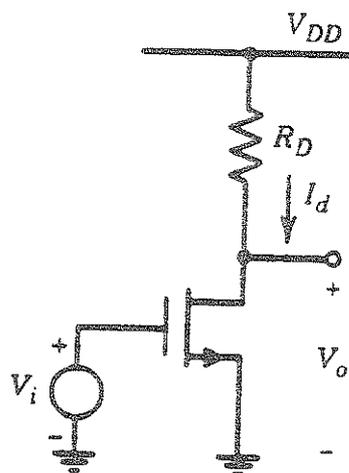


Figura 4.9: Amplificatore a source comune con carico resistivo.

sono riportate in figura 4.8.

## 4.3 Semplici applicazioni dei dispositivi MOS

### 4.3.1 L'amplificatore a singolo stadio

#### Singolo transistor in configurazione a source comune

Nella figura 4.9 è riportato un amplificatore a singolo transistor a canale  $n$  ad arricchimento, in configurazione a source comune. con carico resistivo  $R_D$ ; nella figura 4.10 è invece riportato il circuito equivalente per piccoli segnali. dove la resistenza di uscita  $R_{OUT}$  è la resistenza equivalente Thèvenin vista guardando dentro l'uscita con l'ingresso in cortocircuito:

$$R_{OUT} = R_D // r_o$$

e dove la resistenza di ingresso ( $r_i$ ) per dispositivi MOS è praticamente infinita. Il funzionamento dell'amplificatore considerato può essere descritto nel seguente modo: quando la tensione applicata al gate,  $V_i$ , è nulla, il transistor è interdetto e dunque la corrente di drain,  $I_{DS}$ , è nulla, e la tensione di uscita è uguale a quella

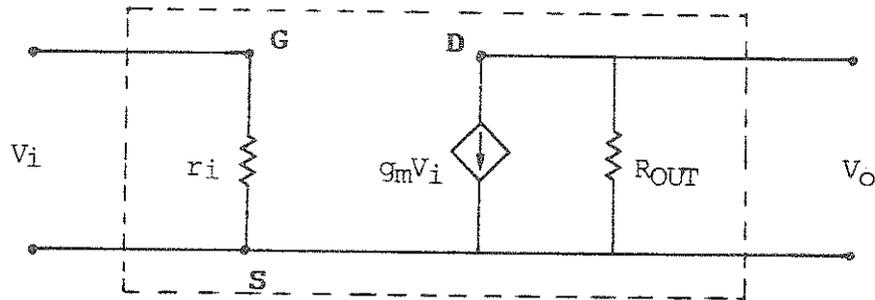


Figura 4.10: Circuito equivalente per piccoli segnali dell'amplificatore a source comune con carico resistivo, riportato in figura 4.9.

di alimentazione,  $V_0 = V_{DD}$ .

Quando la  $V_i$  supera la tensione di soglia  $V_{TH}$ , comincia a scorrere corrente tra drain e source, e si ha:

$$V_0 = V_{DD} - I_{DS}R_D = V_{DD} - \frac{K_n W}{2 L} (V_{GS} - V_{TH})^2 R_D$$

La tensione di uscita diminuisce fino a quando il MOSFET non entra nella *zona resistiva*: la resistenza di uscita diventa molto piccola ed il guadagno di tensione diminuisce drasticamente.

Dal circuito equivalente della figura 4.10 si ha:

$$V_0 = -g_m V_i R_{OUT}$$

e possiamo allora ricavare il guadagno di tensione per piccoli segnali:

$$A_v \equiv \frac{V_0}{V_i} = -g_m R_{OUT} \quad (4.15)$$

Il massimo guadagno dell'amplificatore lo si ottiene quando sono applicati grandi carichi resistivi:

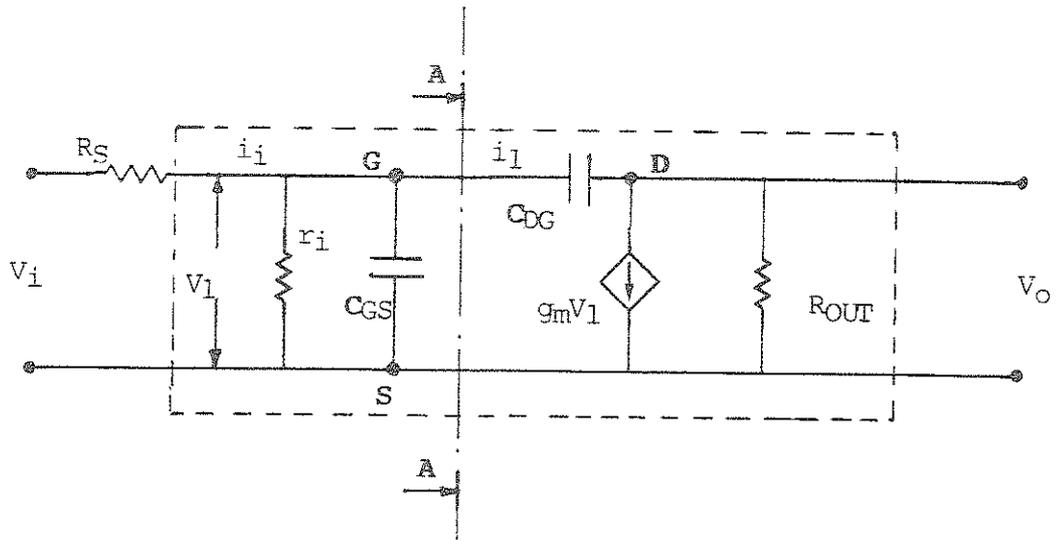


Figura 4.11: Circuito equivalente per piccoli segnali dell'amplificatore a source comune con carico resistivo, per l'analisi in frequenza.

$$A_v(max) = \lim_{R_D \rightarrow \infty} A_v = -g_m r_o$$

### 4.3.2 Risposta in frequenza di amplificatori a singolo stadio

#### Singolo transistor in configurazione a source comune

Si consideri il circuito di figura 4.9 a cui si aggiunga, per completezza, una resistenza di sorgente ( $R_S$ ) tra l'ingresso ( $V_i$ ) ed il gate del transistor. Si ottiene allora il circuito equivalente per piccoli segnali riportato in figura 4.11, in cui si sono evidenziate le capacità parassite del dispositivo MOSFET che, ad alte frequenze, assumono una importanza basilare.

Come nel caso discusso nel paragrafo precedente si ha che la resistenza di ingresso  $r_i$  del MOSFET può considerarsi infinita. Si assuma poi che la resistenza di carico  $R_D$  sia sufficientemente grande, tale per cui:

$$R_{OUT} = r_o // R_D = \left( \frac{1}{r_o} + \frac{1}{R_D} \right)^{-1} \simeq R_D$$

Allora in riferimento al circuito equivalente riportato in figura 4.11 si ha:

$$i_1 = (V_1 - V_0)C_{DG}j\omega$$

ed al contatto di drain D :

$$g_m V_1 + \frac{V_0}{R_D} + (V_0 - V_1)C_{DG}j\omega = 0$$

dove l'ultimo termine rappresenta la corrente fornita direttamente all'uscita attraverso la  $C_{DG}$ , ed è normalmente trascurabile rispetto agli altri termini dell'equazione. Si può quindi ricavare  $V_0$ :

$$V_0 \simeq -g_m R_D V_1$$

Sostituendo nella  $i_1$  si ottiene:

$$i_1 = (1 + g_m R_D)C_{DG}V_1j\omega$$

e l'impedenza vista guardando attraverso il piano A-A sarà data da:

$$X(\text{sez}_{A-A}) = \frac{i_1}{V_1} = (1 + g_m R_D)C_{DG}j\omega$$

L'ultima equazione indica che la  $X$  è puramente capacitiva. Possiamo dunque dire che l'impedenza vista guardando attraverso il piano A-A è una capacità e si definisce *capacità di Miller*:

$$C_M \equiv (1 - A_0)C_{DG} \quad (4.16)$$

dove  $A_0$  è il guadagno di tensione tra il nodo interno di gate e quello di drain (si veda l'equazione 4.15). Dato che questo è solitamente molto maggiore dell'unità,

si ha che la capacità di Miller è  $|C_M/C_{DG}| \gg 1$ . L'origine fisica della capacità di Miller è legata al guadagno in tensione del circuito. Una piccola tensione in ingresso  $V_1$  produce una tensione in uscita molto più grande  $V_0 = -A_0 V_1$  che è di polarità opposta. Dal circuito equivalente si vede che la  $C_M$  si somma direttamente alla  $C_{SG}$  degradando la risposta in frequenza. La capacità totale tra gate e source sarà dunque:

$$C_t = C_{SG} + C_M$$

La tensione ai capi di  $C_t$  è data da:

$$V_1 = \frac{\frac{1}{C_t j\omega}}{\frac{1}{C_t j\omega} + R_S} V_i$$

Ricordando l'espressione scritta più sopra per  $V_0$ , possiamo ora calcolare la funzione di trasferimento:

$$A_v = \frac{V_0}{V_i} = \frac{-g_m R_D}{C_t j\omega} \frac{1}{\frac{1}{C_t j\omega} + R_S} = \frac{A_0}{1 + j\omega R_S C_t} \quad (4.17)$$

Questa analisi indica che il circuito ha un singolo polo:

$$p_1 = -\frac{1}{R_S C_t}$$

in cui il guadagno di tensione  $A_v$  scende di 3 dB al di sotto del valore  $A_0$ , guadagno di tensione a bassa frequenza:  $\omega_{-3dB} = |p_1|$ .

Le prestazioni in frequenza dell'amplificatore sono sintetizzate dunque dalla  $f_{-3dB}$ , che è una misura approssimativa della massima frequenza utile:

$$f_{-3dB} \equiv \frac{1}{2\pi} \frac{1}{(C_{SG} + (1 - A_v)C_{DG})R_S} \quad (4.18)$$

Per quanto riguarda gli amplificatori a singolo stadio trattati, occorre aggiungere che uno dei metodi (comunemente utilizzati nel disegno di circuiti con dispositivi

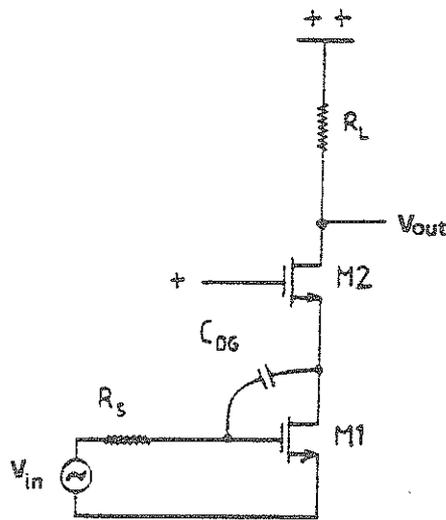


Figura 4.12: Amplificatore a singolo stadio "normal cascode". cioè con il transistor M2 in configurazione a base comune [45].

MOS) per eliminare l'effetto degradante della capacità di Miller, è costituito dall'applicare un secondo transistor, simile al primo, in configurazione a gate comune, usualmente definita *cascode*, come mostrato in figura 4.12 [45].

In questo modo il drain del transistor M1 ora vede la piccola resistenza di ingresso ( $1/g_{m2}$ ) del transistor M2. Quindi, ricordando l'equazione 4.15, se i due dispositivi sono identici, il guadagno in tensione sul transistor M1 è dato da:

$$A_v(M1) = -g_{m1} \frac{1}{g_{m2}} = -1$$

eliminando così l'effetto Miller, mentre il guadagno in tensione totale, tra il gate di M1 ed il drain di M2, sarà sempre dato da:

$$A_v = -g_{m1} R_L$$

La configurazione cascode permette quindi di avere alti guadagni in tensione dell'amplificatore a singolo stadio, mantenendo una buona risposta in frequenza.

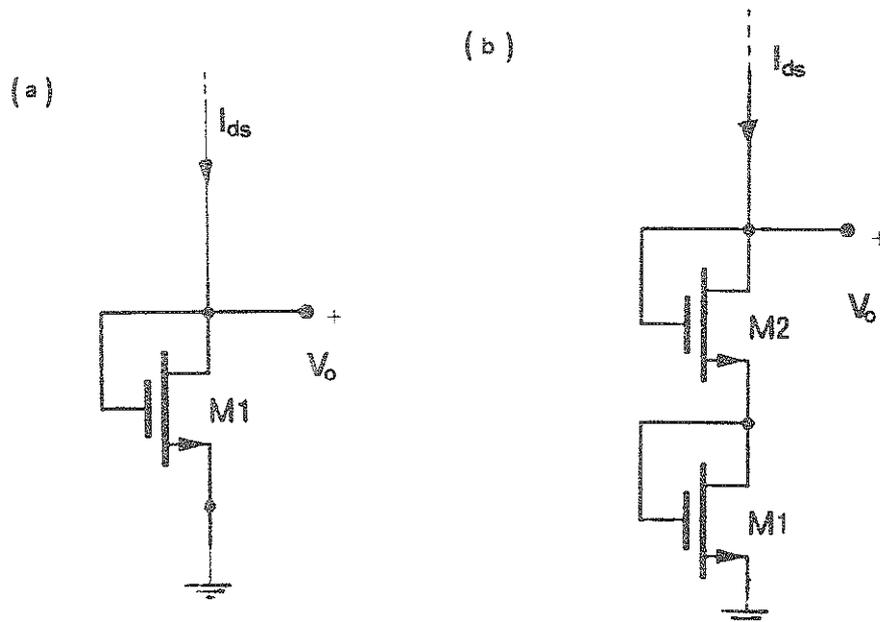


Figura 4.13: Utilizzo di resistori attivi per creare una caduta di tensione.

### 4.3.3 Resistore attivo

I resistori attivi sono utilizzati al posto dei resistori ottenuti da superfici diffuse o di polisilicio, per creare una caduta di tensione, oppure per creare una resistenza per piccoli segnali, che sia lineare su un piccolo range [39]. In molti casi la superficie necessaria per creare una resistenza per piccoli segnali è più importante della linearità. Infatti un piccolo dispositivo MOS può simulare la resistenza di un resistore equivalente diffuso o in polisilicio, in un'area estremamente inferiore. Si crea un resistore attivo semplicemente connettendo il gate al drain del MOSFET. Essendo ora  $V_{GS} = V_{DS}$ , dalla equazione 4.8 si ha che ora la tensione di drain-source controlla la corrente di drain e quindi la transconduttanza del canale  $g_m$  diviene ora una conduttanza del modello per piccoli segnali. Si può infatti facilmente vedere che [39]:

$$r_0 = \frac{1}{g_m + g_{mb} + g_{ds}} \simeq \frac{1}{g_m}$$

Per esempio, si immagina di avere a disposizione una corrente di  $25 \mu\text{A}$ , e di dover creare una caduta di tensione di  $5 \text{ V}$ , cioè una resistenza di  $200 \text{ k}\Omega$ .

Utilizzando un resistore passivo in polisilicio è necessaria una superficie di circa  $10^5 \mu\text{m}^2$  [42]. Utilizzando invece un resistore attivo MOSFET a canale  $n$  ad arricchimento, come mostrato nella figura 4.13.(a) si avrà:

$$I_{DS} = 25\mu A = \frac{K_n W}{2 L} (5 - V_{TH})^2$$

Assumendo (si veda la tabella 5.1) che:  $V_{TH}=0.75$  V e che  $K_n = 2.4E-05$  A<sup>2</sup>/V si ottiene un valore del rapporto  $W/L$  pari a:

$$\frac{W}{L} = \frac{1}{8.7}$$

che con un processo MOS 2  $\mu$ m fornisce una superficie necessaria di appena 17.4  $\mu$ m<sup>2</sup>.

Utilizzando poi due resistore attivi MOSFET a canale  $n$  ad arricchimento, come mostrato nella figura 4.13.(b) si avrà:

$$I_{DS} = 25\mu A = \frac{K_n W}{2 L} \left(\frac{5}{2} - V_{TH}\right)^2$$

e il rapporto  $W/L$  diventa:

$$\frac{W}{L} = \frac{1}{1.47}$$

che con lo stesso processo fornisce una superficie necessaria di 5.88  $\mu$ m<sup>2</sup>, cioè circa 1/3 dell'area richiesta con un solo transistor. Si comprende dunque la necessità del transistor n4d nel circuito schematico riportato in figura 5.4: diminuisce la superficie occupata sul chip dalla polarizzazione.

#### 4.3.4 Semplice generatore di corrente

I generatori di corrente che utilizzano dispositivi attivi sono di uso comune nei circuiti analogici integrati, sia come elementi di polarizzazione, sia come dispositivi di carico negli stadi degli amplificatori. L'uso di generatori di corrente come elementi di polarizzazione può condurre ad una elevata insensibilità delle prestazioni del circuito alle variazioni della sorgente di alimentazione e alla temperatura. Inoltre i generatori di corrente, così come i partitori di tensione (vedi

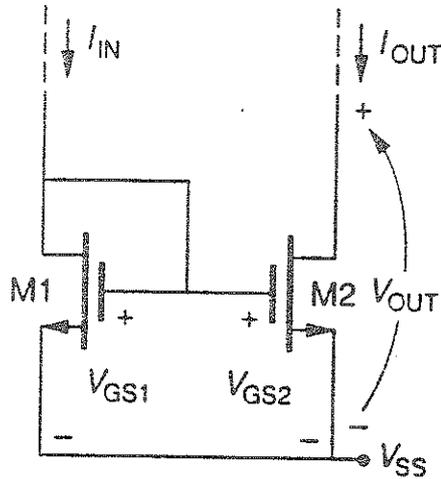


Figura 4.14: Semplice specchio di corrente in processo MOS [42].

paragrafo precedente) sono più economici che non i resistori in termini di area del chip richiesta per fornire una determinata corrente di polarizzazione, soprattutto quando questa è piccola.

Questo paragrafo non pretende di dare una trattazione teorica dei generatori di corrente che si possono creare con i MOSFET. Vuole invece portare a conoscenza del lettore una semplice configurazione di generatore di corrente: lo *specchio di corrente*, così chiamato perché la corrente in ingresso ( $I_{IN}$  della figura 4.14) viene riprodotta in uscita ( $I_{OUT}$ ).

Questo semplice dispositivo è largamente utilizzato nel disegno di circuiti integrati, ed è alla base della polarizzazione del dell'amplificatore operazionale "OTA" riportato in figura 5.4.

Considerando la figura 4.14 si ha che il transistor **M1** è in saturazione, poiché il drain è collegato al gate. Lavora quindi come un carico attivo e automaticamente fissa la tensione di gate del transistor **M2**.

Ricordando l'equazione 4.8. per il transistor **M1** si ha:

$$I_{IN} = \frac{K_n W_1}{2 L_1} (V_{GS1} - V_{TH})^2 \quad (4.19)$$

mentre per il transistor **M2**:

$$I_{OUT} = \frac{K_n W_2}{2 L_2} (V_{GS2} - V_{TH})^2 \quad (4.20)$$

Ora, facendo il rapporto dell'equazione 4.19 con la 4.20 (considerando la tensione

di soglia costante) si ottiene:

$$I_{OUT} = I_{IN} \frac{W_2 L_1}{W_1 L_2} \quad (4.21)$$

Quindi con lo specchio di corrente è possibile ottenere una qualsiasi corrente di polarizzazione semplicemente collegando tra loro i gate dei transistor e modificando opportunamente le dimensioni  $W$  ed  $L$ .

## 4.4 Disegno di circuiti integrati

In generale, l'obiettivo da raggiungere nel disegno di circuiti integrati, è di soddisfare le specifiche richieste spendendo il minore tempo lavorativo e dispositivi fisici possibili, in un periodo temporale limitato. Inoltre si deve ottenere una alta produttività con un processo relativamente semplice e limitare al massimo le superfici morte (quelle inutilizzate del *wafers*, e quelle inevitabilmente perdute nel processo di fabbricazione). La complessità delle informazioni associate al disegno del circuito schematico, del layout ed alla simulazione, è così imponente che l'aiuto del calcolatore, attraverso pacchetti software di disegno (CAD) e di simulazione, è di cruciale importanza.

In riferimento alla figura 4.15 si possono identificare le seguenti fasi (blocchi) del processo convenzionale di disegno di circuiti integrati.

- Il punto di partenza è costituito dall'insieme di specifiche richieste per il circuito integrato.
- Da queste si passa automaticamente ad un disegno preliminare, basato su semplici modelli dei dispositivi o sottocircuiti utilizzati, che ne rappresentano, in modo sommario, il comportamento.
- Si esegue dunque una prima simulazione al calcolatore, con modelli più accurati, per verificare le prestazioni del disegno preliminare. "Buoni" modelli dei dispositivi e sottocircuiti sono di cruciale importanza in questa fase. Un modello può considerarsi "buono" quando predice accuratamente il comportamento sperimentale del dispositivo simulato, ed è sufficientemente sem-

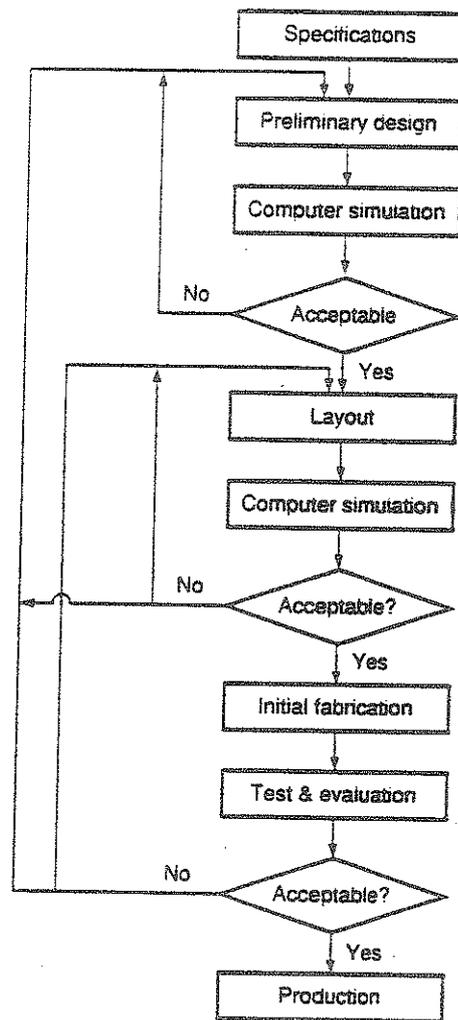


Figura 4.15: Diagramma a blocchi del processo convenzionale di disegno di circuiti integrati [42].

plice da richiedere tempi di calcolo ragionevoli. Queste prime fasi richiedono la maggior parte del tempo di disegno del circuito integrato.

- Una volta che il disegno schematico preliminare è stato eseguito e si è dimostrato accettabile, si può iniziare a disegnarne il "layout", cioè le maschere fotolitografiche che serviranno per la fabbricazione del circuito integrato (si veda il paragrafo 4.1).
- Si esegue quindi una nuova simulazione del circuito. Ora però si terrà conto della configurazione geometrica dei dispositivi ed il simulatore calcolerà le capacità e resistenze parassite che si verrebbero a creare realmente nel processo di fabbricazione. Gli effetti parassiti associati al layout giocano un ruolo fondamentale sia nei circuiti analogici che in quelli digitali. In quelli analogici, gli effetti parassiti tendono a degradare le prestazioni specificate, mentre in quelli digitali causano dannosi ritardi temporali (cambiando le costanti di tempo del circuito) che possono potenzialmente dare luogo a situazioni disastrose. In questa fase, occorre inoltre rispettare le regole di disegno dei vari dispositivi, dettate dal tipo di tecnologia utilizzata, cioè dalla fonderia che produrrà il circuito integrato. Ogni layout è dunque disegnato specificatamente per una fonderia, di cui il simulatore deve essere in grado di creare dei modelli accurati. I risultati della simulazione del layout, necessitano spesso dei cambiamenti nel disegno del circuito schematico, poiché gli effetti parassiti non possono essere completamente risolti cambiando il solo layout.
- Raggiunta una simulazione accettabile dell'intero circuito, il disegno del layout viene mandato alla fonderia, che ne ricaverà le maschere fotolitografiche per la fabbricazione. Per disegni molto complicati, prima di mandare in produzione l'intero circuito, si fabbricano spesso dei sottocircuiti del disegno complessivo, in modo da avere maggiori informazioni per creare dei modelli accurati.
- Una volta fabbricato, il chip (cioè il singolo circuito integrato replicato molte volte sul *wafers*) viene spesso provato dalla fonderia stessa. Un singolo errore nel disegno del circuito o nella simulazione o nel layout, ne compromette il funzionamento in modo parziale o totale.

- Se il test della fabbricazione iniziale del chip ha dato un esito favorevole rispondendo alle specifiche richieste originariamente, allora il chip entra in produzione.

## Capitolo 5

# Analisi e modifica del chip HARP

Il funzionamento generale della *memoria analogica pipeline* HARP è già stato descritto nel paragrafo 3.4. Ci limiteremo ora a prendere in considerazione la struttura analogica del chip, cioè l'amplificatore operazionale e le celle di memoria con i relativi switches. Analizzatone il comportamento potremo in seguito apportare le modifiche più convenienti per il "front-end" dei rivelatori che comporranno il *cubo magico*.

L'HARP è stato costruito in tecnologia Mietec con un processo CMOS di 1.5  $\mu\text{m}$ . Il disegno schematico dell'amplificatore operazionale per la scrittura e lettura delle capacità di memoria è riportato nella figura 5.1, mentre una singola cella di memoria è riportata nella figura 5.2.

Il valore della capacità di integrazione (0.4 pF) è un compromesso tra segnale integrato e rumore da una parte, e tra superficie occupata sul chip e velocità dall'altra.

L'organizzazione degli switches è simmetrica intorno alla capacità per minimizzare gli effetti dovuti alla iniezione di carica e "crosstalk".

L'amplificatore OTA (Operational Transconductance Amplifier) è stato disegnato per avere alte velocità nel limite del consumo energetico accettabile, che, alla corrente nominale di polarizzazione, è di 5.2 mW. La capacità di ingresso è di  $\sim 0.5$  pF.

Dalla figura 5.3 si vede che l'OTA ha un guadagno unitario a 220 MHz con 0.8 pF di carico in uscita [37].

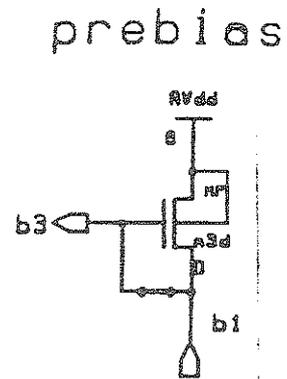
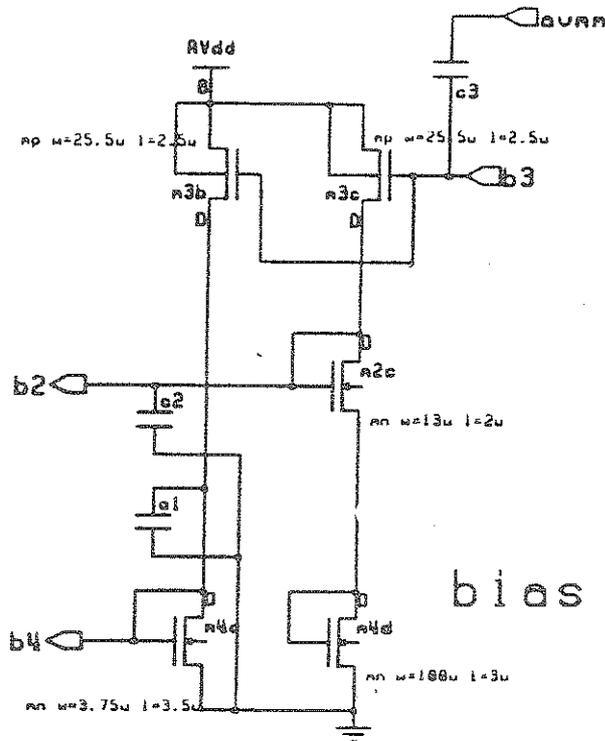
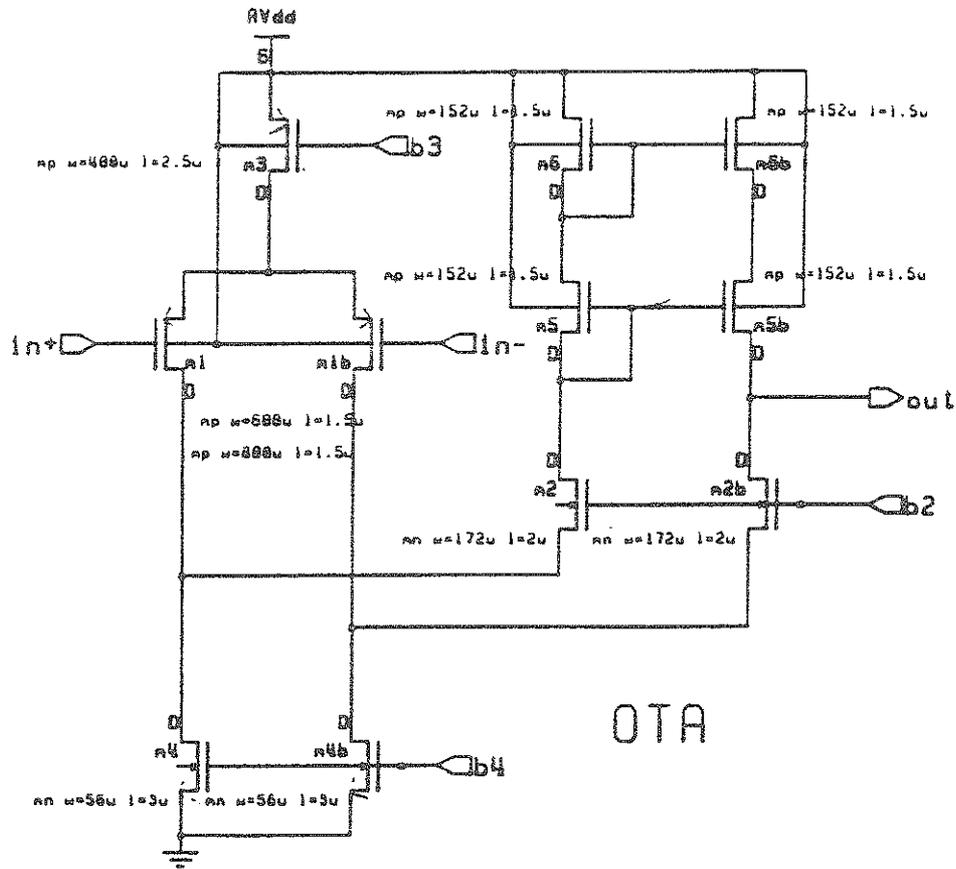


Figura 5.1: Circuito schematico dell'OTA, amplificatore operativo dell'HARP (tecnologia Mietec, CMOS 1.5 $\mu$ m) [37].

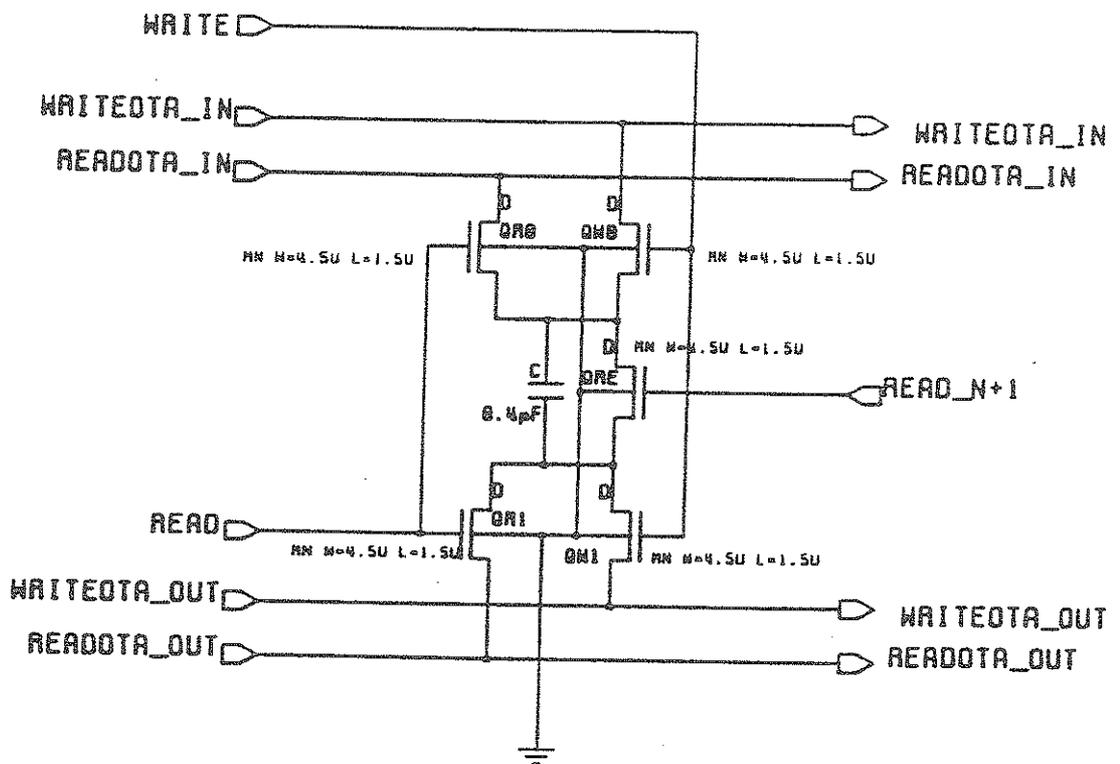


Figura 5.2: Cella di memoria e relativi switches (tecnologia Mietec, CMOS 1.5 $\mu$ m) [37].

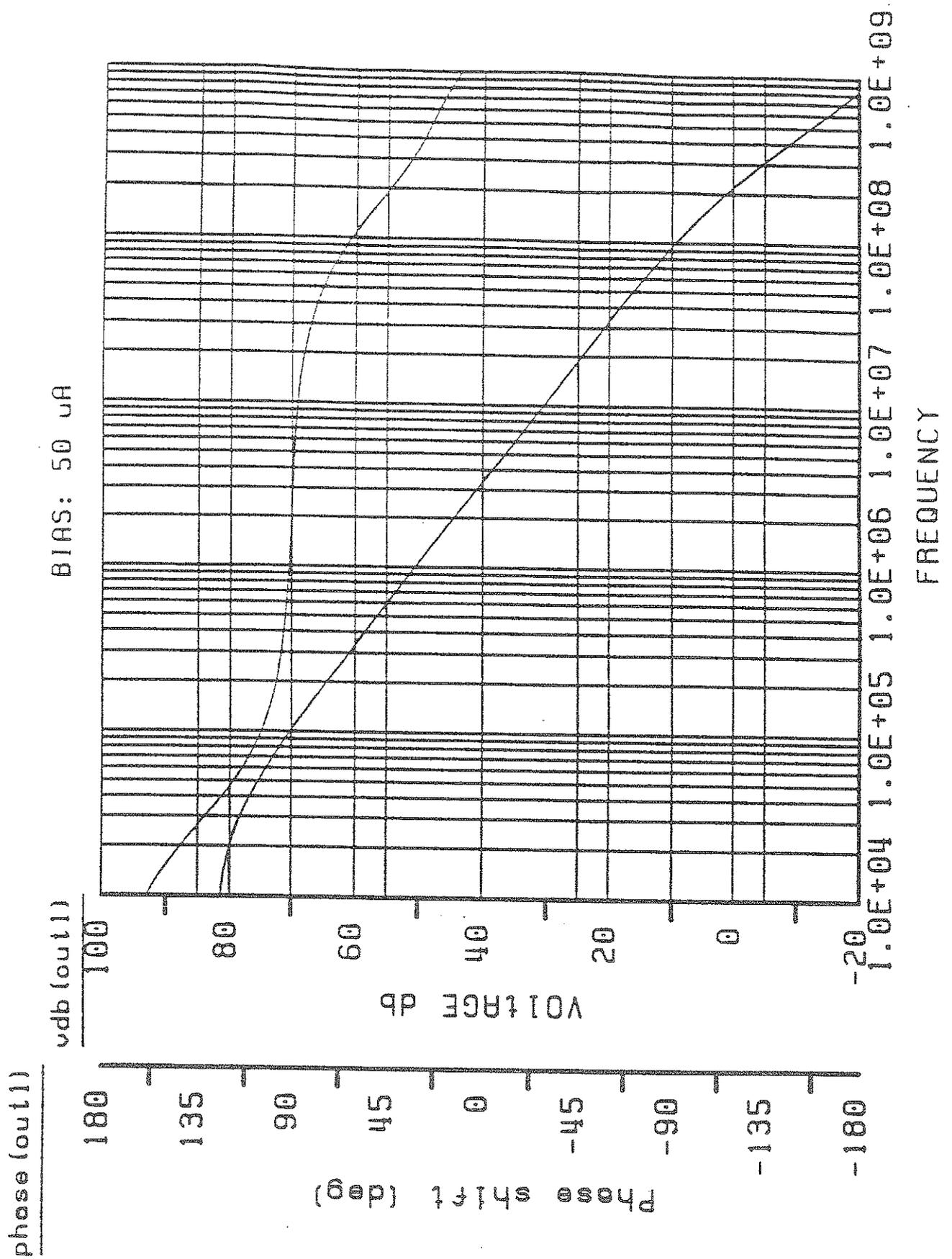


Figura 5.3: Simulazione AC: risposta in frequenza e fase dell'OTA (tecnologia Mietec, CMOS  $1.5\mu\text{m}$ ) [37].

## 5.1 Analisi del circuito

Si è anzitutto implementato su calcolatore il circuito schematico dell'OTA, come si può vedere in figura 5.4.

Si è utilizzato il pacchetto software CADENCE versione 4.3.2 (subver 4.3.2.32) e si sono eseguite le simulazioni con il simulatore ELDO versione 4.3.4.104. Dato che le eventuali modifiche apportate verranno poi eseguite dalla fonderia AMS, si sono utilizzati i files di tecnologia AMS, con un processo CMOS di  $1.2\mu\text{m}$ . Questi sono diversi da quelli utilizzati originariamente per la simulazione riportata nella figura 5.3, e dunque occorrerà sicuramente apportare alcune modifiche per ottenere lo stesso comportamento nella nuova tecnologia. Prima di procedere è forse utile richiamare alcune definizioni.

### 5.1.1 L'amplificatore operazionale

L'amplificatore operazionale è essenzialmente un amplificatore differenziale, in cui il potenziale di uscita dipende dalla differenza di potenziale che si trova tra i due ingressi. Idealmente esso ha una sola uscita ("single-ended"), guadagno infinito, resistenza di ingresso infinita e resistenza di uscita nulla [38].

Considerando la figura 5.5, in prima approssimazione si può assumere che il potenziale di uscita sia una funzione lineare del potenziale di ciascun ingresso [40]:

$$V_s = A^+ \cdot V_e^+ - A^- \cdot V_e^- \quad (5.1)$$

Si possono inoltre definire:

il *potenziale in modo comune* come il potenziale medio applicato ai due ingressi:

$$V_c = (V_e^+ + V_e^-)/2$$

e la *tensione in modo differenziale* come la tensione esistente tra i due ingressi:

$$U_e = V_e^+ - V_e^-$$

Si possono poi esprimere ciascuno dei potenziali di ingresso come il risultato del contributo del modo comune e del modo differenziale  $V_e^\pm = V_c \pm U_e/2$ , ed introducendo queste espressioni nell'equazione 5.1 si ottiene:

$$V_s = \frac{A^+ + A^-}{2} U_e + (A^+ - A^-) \cdot V_c$$

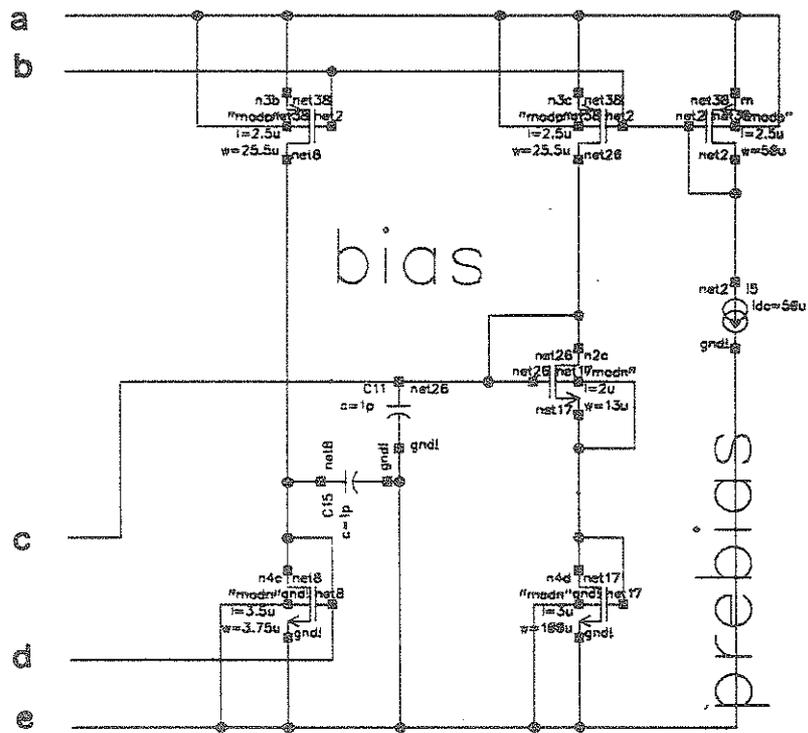
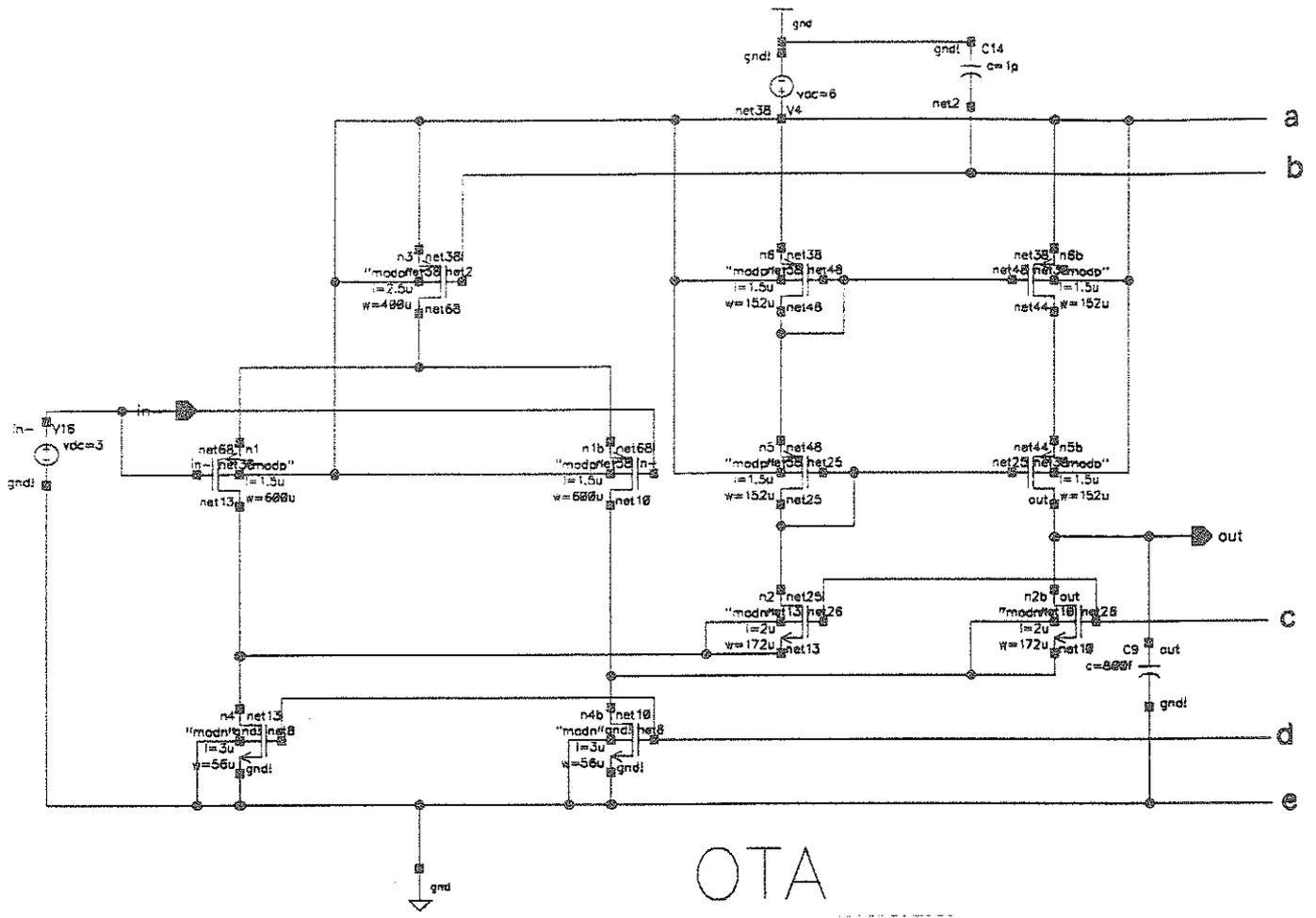


Figura 5.4: Circuito schematico dell'OTA implementato su CADENCE

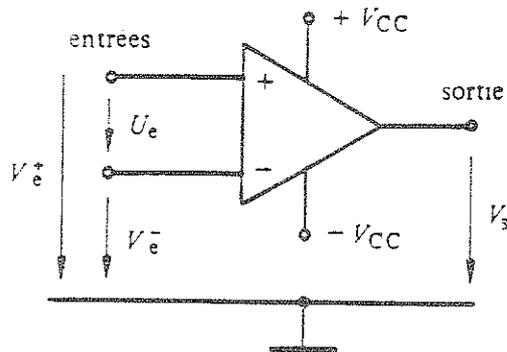


Figura 5.5: Amplificatore operazionale ideale [40].

Si definisce *guadagno in modo differenziale* il rapporto  $A_d$  tra il potenziale di uscita e la tensione in modo differenziale, quando è nullo il potenziale in modo comune:

$$A_d = \left( \frac{V_s}{U_e} \right)_{V_c=0} = (A^+ + A^-)/2$$

Si definisce *guadagno in modo comune* il rapporto  $A_c$  tra il potenziale di uscita e il potenziale in modo comune, quando la tensione in modo differenziale è nulla:

$$A_c = \left( \frac{V_s}{V_c} \right)_{U_e=0} = A^+ - A^-$$

Per mettere in evidenza la natura essenzialmente differenziale dell'amplificatore operazionale si può scrivere il potenziale di uscita nel seguente modo:

$$V_s = A_d \cdot U_e \cdot \left( 1 + \frac{A_c}{A_d} \cdot \frac{V_c}{U_e} \right) \quad (5.2)$$

Si definisce poi *rapporto di reiezione in modo comune* (CMRR) il rapporto  $A_d/A_c$ , che nell'amplificatore differenziale ideale è infinito.

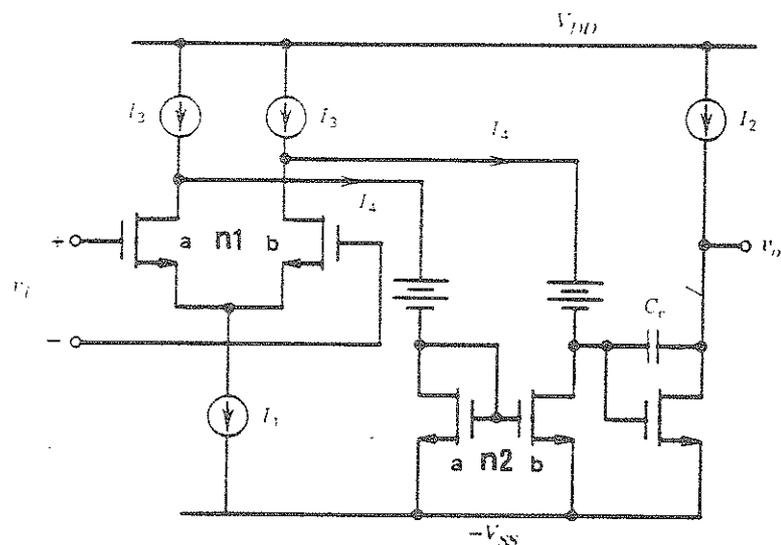


Figura 5.6: Esempio di realizzazione di cascode NMOS con rete di accoppiamento intermedia costituita da generatori di tensione.

### 5.1.2 Descrizione generale del circuito OTA

Dal circuito schematico dell'OTA implementato, riportato in figura 5.4, si può notare che la topologia base è costituita da un amplificatore "folded cascode", letteralmente cascode ripiegato su se stesso: uno stadio di ingresso differenziale ad elevata transconduttanza ( $W/L = 600\mu\text{m}/1.5\mu\text{m}$ ) costituito dai transistori PMOS a source comune **n1** e **n1b** è seguito dai transistori NMOS **n2** e **n2b** che si comportano da cascode: sono cioè in configurazione a gate comune, ed in tal modo incrementano la resistenza di uscita dei precedenti, aumentandone il guadagno in tensione.

Si suole dire che questo circuito è "folded" nel senso che un cascode normale avrebbe avuto dispositivi in ingresso di tipo NMOS a pilotare **n2** e **n2b**; in questo caso però, sarebbero stati necessari dei traslatori di livello tra i contatti di DRAIN, come si può vedere nella figura 5.6.

Questo unico stadio di amplificazione costituito dai transistori **n1**, **n1b**, **n2**, **n2b**, è circondato da una serie di transistori di *bias*, per la corretta polarizzazione dei primi.

I generatori ideali di corrente riportati nella figura 5.6 sono in realtà sostituiti da dispositivi attivi nello schematico dell'OTA (riportato nella figura 5.4), e polarizzano correttamente il cascode in modo che:

$$I_4 = I_3 - \frac{I_1}{2} \gg 0 \quad (5.3)$$

L'utilizzo di una configurazione a gate comune (cascode) è necessaria per poter avere alti guadagni di tensione, senza dover introdurre altri stadi di amplificazione, che apportano inevitabilmente del rumore e consumo maggiore. I transistori di cascode infatti, aumentano la resistenza di uscita dei transistori in ingresso, collegati a source comune, di un fattore  $g_m \cdot r_0$  (si veda paragrafo 4.3.1): si ottiene così un guadagno di tensione tale che un unico stadio è sufficiente.

Il guadagno di tensione del cascode per piccolo segnale alle basse frequenze è dato da [39]:

$$A_v = \frac{g_{m1}}{g_{ds1} + g_{ds2}} \quad (5.4)$$

dove  $g_{m1}$  è la transconduttanza dei transistori in ingresso, e  $g_{ds}$  è la conduttanza di uscita per piccoli segnali (si vedano le equazioni 4.10 e 4.12). Si possono ottenere guadagni dell'ordine di alcune migliaia.

Un importante vantaggio del "folded cascode" è che la capacità di carico, posta direttamente in uscita, attua la funzione di condensatore di compensazione per impedire al circuito di oscillare quando è inserito in un anello di reazione [38]. Purtroppo se si vuole limitare la superficie del chip definitivo, occorrerà limitare i valori delle capacità presenti, soprattutto se queste dovranno essere moltiplicate molte volte (come nel caso di più canali di amplificazione), e sarà quindi necessario cercare altre soluzioni per rendere il circuito compensato, come si leggerà più avanti.

Dalla figura 5.4 si possono identificare tre reti fondamentali: l'OTA vero e proprio, il "bias" ed il "prebias" di polarizzazione: vi è poi l'alimentazione che fornisce una tensione di 6 V al ramo superiore del circuito.

Il "prebias" non è altro che uno specchio di corrente: un generatore ideale di corrente che eroga  $50 \mu A$  continui, viene specchiato dal transistor *m* a tutti i transistor collegati al gate di questo, e cioè *n3b* e *n3c* del bias ed *n3* dell'OTA. Possiamo immediatamente calcolare le correnti  $I_d$  che scorrono tra drain e source dei transistor specchiati. Ricordando l'equazione 4.21 e che tutti i transistori citati hanno la stessa *L* (Lunghezza di gate) si avrà:

$$I_d(n3*) = I_d(m) \cdot \frac{W(n3*)}{W(m)}$$

e quindi:

$$I_d(n3) = 50\mu A \cdot \frac{400\mu m}{50\mu m} = 400\mu A$$

$$I_d(n3b) = I_d(n3c) = 50\mu A \cdot \frac{25.5\mu m}{50\mu m} = 25.5\mu A$$

Il ramo **n3c-n2c-n4d** del "bias" funge invece da partitore di tensione per polarizzare correttamente i transistor **n2** e **n2b** dell'OTA, e cioè fornire la corretta tensione di gate affinché essi funzionino in saturazione.

Infatti i transistor **n2c** e **n4d** hanno il drain collegato al gate e dunque  $V_{DS} = V_{GS}$ ; inoltre il source è collegato al bulk e quindi la tensione di soglia sarà semplicemente la  $V_{T0}$  del dispositivo.

Assumendo  $V_{T0} = 0.75V$  (per dispositivi a canale n) ed il parametro di transconduttanza  $K_n = 7.8 \cdot 10^{-5} A/V^2$ , e ricordando l'equazione 4.8, si calcola la tensione tra drain e source di **n4d** (cioè del nodo **net17**):

$$V_{DS} = V_{GS} = \left( \frac{2 \cdot I_d \cdot L}{K_n \cdot W} \right)^{1/2} + V_{T0}$$

e quindi:

$$V_{DS}(n4d) = \left( \frac{2 \cdot 25.5 \cdot 10^{-6} \cdot 3\mu m}{7.8 \cdot 10^{-5} \cdot 100\mu m} \right)^{1/2} + 0.75 = 0.89 V.$$

mentre:

$$V_{DS}(n2c) = \left( \frac{2 \cdot 25.5 \cdot 10^{-6} \cdot 2\mu m}{7.8 \cdot 10^{-5} \cdot 13\mu m} \right)^{1/2} + 0.75 = 1.06V.$$

La tensione al nodo **net26** che andrà a pilotare i transistor dell'OTA **n2**, **n2b** sarà dunque:

$$V(\text{net26}) = V_{DS}(n4d) + V_{DS}(n2c) = 1.96 V.$$

Si possono confrontare i valori calcolati con quelli riportati nelle simulazioni che seguono.

Il ramo n3b-n4c del "bias" è nuovamente uno specchio di corrente per i transistori n4, n4b dell'OTA. In questo ramo scorre una corrente di  $25.5 \mu\text{A}$  come calcolato ed il transistor n4c forza nel drain di n4, n4b una corrente:

$$I_d(n4) = I_d(n4c) \cdot \frac{W(n4) L(n4c)}{W(n4c) L(n4)}$$

quindi:

$$I_d(n4) = I_d(n4b) = 25.5 \mu\text{A} \cdot \frac{56}{3.75} \frac{3.5}{3} = 444.3 \mu\text{A}.$$

È dunque verificata l'equazione 5.3 e si forza così sui rami n2-n5-n6 ed n2b-n5b-n6b una corrente di:  $(444.3 - 400/2) \mu\text{A} = 244.3 \mu\text{A}$ .

Nell'OTA possiamo quindi identificare i transistori n3, n4 e n4b come generatori di corrente polarizzati dal bias. Per quanto riguarda gli altri transistori possiamo dire che l'amplificatore folded cascode è fondamentalmente costituito dai quattro transistori nei due rami n1-n2 e n1b-n2b: il primo transistor di ogni ramo è il transistor di ingresso configurato a source comune, mentre il secondo è il transistor di cascode, cioè configurato a gate comune.

I transistori n5 n6 n5b n6b rappresentano uno specchio di corrente per poter avere una unica uscita. La corrente che scorre nel cascode n2 dovuta all'ingresso n1, viene specchiata da n5 n6 sul ramo n5b n6b, dove si trova la corrente del cascode n2b dovuta all'altro ingresso n1b. In uscita si avrà quindi la differenza di queste due correnti; da qui l'aspetto puramente differenziale dell'amplificatore operazionale.

Le capacità di 1 pF che compaiono nella rete di bias servono esclusivamente a stabilizzare la polarizzazione ed il loro valore potrebbe anche essere diverso. Il fatto che sia abbastanza grande non deve spaventare poiché queste, così come la polarizzazione, serviranno per tutti i canali del chip e quindi occuperanno una superficie relativamente piccola.

Il circuito è stato alimentato con la tensione più positiva di 6 V e quella più negativa di 0 V. In questo modo, se tutti i dispositivi funzionano al meglio, in ingresso ed in uscita si dovrebbe avere una tensione di equilibrio tra i due valori citati e quindi di 3 V. In seguito saranno riportati i punti di lavoro di ogni dispositivo e l'analisi DC.

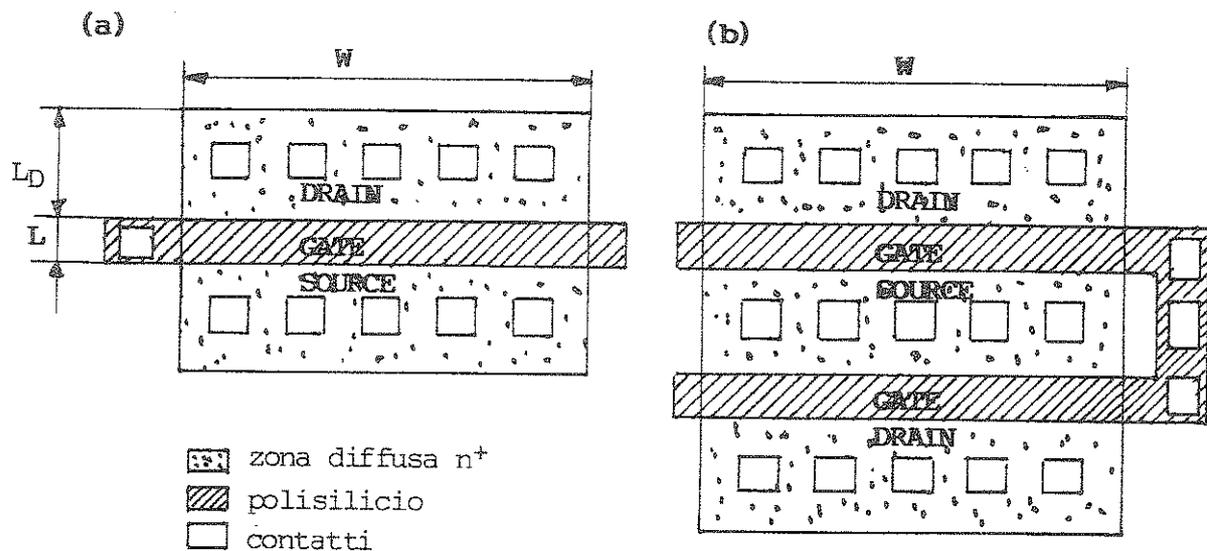


Figura 5.7: Possibili soluzioni di layout di transistor: sono indicate le zone di diffusione per un transistor normale (a), ed uno ad alto rapporto  $W/L$  (b).

### 5.1.3 Simulazione dell'OTA

Una volta "disegnato" il circuito su CADENCE occorre creare una *netlist*, cioè una lista dei dispositivi presenti sul circuito e relative connessioni. A questo punto è possibile far leggere questa netlist ai vari simulatori di circuiti esistenti in commercio (SPICE, SPECTRE, ELDO, cdsSPICE, hSPICE....) i quali trasformano i dispositivi letti in modelli matematici.

Le simulazioni sono state eseguite con il simulatore ELDO v4.3.4.104, Feb-15-1995, e nelle tabelle 5.1 e 5.2, sono riportati i principali parametri utilizzati dal simulatore per la creazione del modello matematico dei transistor MOS.

Per ogni transistor si sono impostati l'area ed il perimetro di diffusione della zona di drain e source. In questo modo le capacità parassite calcolate dal simulatore, che alterano inevitabilmente la risposta in frequenza, saranno il più vicino possibile alla realtà. Utilizzando files di tecnologia AMS (processo CMOS 1.2  $\mu\text{m}$ ), si ha una lunghezza di diffusione di 3.5  $\mu\text{m}$  ( $L_D$ ). Dalla figura 5.7.(a), si può quindi calcolare:

- area di diffusione =  $A_d = W \cdot L_D$ ;
- perimetro di diffusione =  $P_d = (W + L_D) \cdot 2$ .

Non è sempre vero che l'area di diffusione del source sia uguale a quella del drain; infatti nel caso di transistori molto larghi il layout si esegue secondo la figura

Temperatura	27 C
Livello	12
Tipo	NMOS
Tensione di soglia a zero "body-bias" ( $V_{T0}$ )	7.512E-01 V
Parametro di transconduttanza ( $K_n$ )	7.802E-05 A/V <sup>2</sup>
Parametro di "body effect" ( $\gamma$ )	6.355E-01 $\sqrt{V}$
Potenziale di superficie per inversione forte ( $\phi$ )	7.529E-01 V
Velocità di saturazione dei portatori di carica	9.430E+04 m/s
Spessore dell'ossido ( $t_{ox}$ )	2.160E-08 m
Concentrazione di drogaggio del canale	3.110E+16 cm <sup>-3</sup>
Diffusione laterale ( $L_d$ )	7.700E-08 m
Mobilità superficiale dei portatori di carica	4.880E+02 cm <sup>2</sup> /Vs
Capacità dell'ossido ( $C_{ox}$ )	1.598E-03 F/m <sup>2</sup>
Capacità gate-source ( $C_{GSO}$ )	3.300E-01 F/m
Capacità gate-drain ( $C_{GDO}$ )	3.300E-01 F/m
Capacità gate-bulk ( $C_{GBO}$ )	1.500E-01 F/m

Tabella 5.1: Parametri principali del modello matematico per transistori NMOS del simulatore ELDO.

Temperatura	27 C
Livello	12
Tipo	PMOS
Tensione di soglia a zero "body-bias" ( $V_{T0}$ )	-7.597E-01 V
Parametro di transconduttanza ( $K_n$ )	2.389E-05 A/V <sup>2</sup>
Parametro di "body effect" ( $\gamma$ )	6.084E-01 $\sqrt{V}$
Potenziale di superficie per inversione forte ( $\phi$ )	7.484E-01 V
Velocità di saturazione dei portatori di carica	5.310E+04 m/s
Spessore dell'ossido ( $t_{ox}$ )	2.160E-08 m
Concentrazione di drogaggio del canale	2.850E+16 cm <sup>-3</sup>
Diffusione laterale ( $L_d$ )	3.000E-08 m
Mobilità superficiale dei portatori di carica	1.494E+02 cm <sup>2</sup> /Vs
Capacità dell'ossido ( $C_{ox}$ )	1.598E-03 F/m <sup>2</sup>
Capacità gate-source ( $C_{GS0}$ )	3.300E-01 F/m
Capacità gate-drain ( $C_{GD0}$ )	3.300E-01 F/m
Capacità gate-bulk ( $C_{GB0}$ )	1.500E-01 F/m

Tabella 5.2: Parametri principali del modello matematico per transistori PMOS del simulatore ELDO.

5.7.(b) e quindi, mentre l'area di diffusione del drain è sempre ( $W \cdot L_D$ ), quella del source è solo la metà di questa. In questo caso la larghezza effettiva del transistor è  $2W$  (vedi figura) e la capacità parassita del source si è dimezzata. Ponendo in ingresso all'OTA la metà della tensione di alimentazione (cioè 3 V) si è eseguita una analisi dei punti di lavoro di ogni transistor: occorre verificare che tutti quanti funzionano bene e che quindi sono in regime di saturazione:

$$V_{DS} \geq (V_{GS} - V_{TH}) \quad (5.5)$$

Le correnti, tensioni e capacità parassite, calcolate per ogni transistor, sono riportate nella figura 5.8.

## 5.2 Modifica dell'OTA

Come precedentemente spiegato, il circuito schematico dell'HARP riportato in figura 5.1 si riferisce ad una tecnologia differente da quella che si sta utilizzando per le simulazioni. Ci aspettiamo dunque che il circuito implementato nella figura 5.4 si comporti in modo diverso. Quello che ci accingiamo a fare è apportare le modifiche necessarie a riottenere una configurazione stabile da utilizzarsi per il disegno del chip che verrà impiegato nel front-end del cubo magico.

Dalla figura 5.8 si può notare che l'uscita (OUT) si trova ad una tensione di:

$$V_{OUT} = V_{DS}(n4b) + V_{DS}(n2b) \simeq 3 \text{ V},$$

e quindi l'amplificatore è ben equilibrato, avendo una tensione di alimentazione tra 6 V e 0 V. Si può inoltre notare che per quasi tutti i transistori l'equazione 5.5 è soddisfatta, tranne che per n4 e n4b. Per aumentare lievemente la loro tensione  $V_{DS}$  basterà portare ad una tensione più alta il nodo net26 che collega i gate di n2, n2b ed n2c. Infatti la tensione  $V_{GS}(n2) = V_{GS}(n2b)$  è fissata dalla corrente di polarizzazione che scorre tra drain e source e dunque aumentando la loro tensione di gate automaticamente si alza la tensione di source che essendo collegato al drain di n4 e n4b, ne aumenta la  $V_{DS}$ .

Nel ramo di bias n3c-n2c-n4d, essendo la corrente impostata dallo specchio m÷n3c la  $V_{GS}(n2c)$  e la  $V_{GS}(n4d)$  devono rimanere costanti (per l'equazione 4.8). Quindi diminuendo la larghezza  $W$  di n4d si aumenta la sua  $V_{GS}$  che, dato il collegamento, è uguale alla  $V_{DS}$ ; aumenta la tensione di source di n2c ed automaticamente, per quanto detto sopra, quella di gate, cioè la tensione del nodo



net26.

Dalla figura 5.9 si può vedere che cambiando la  $W(n4d)$  da  $100 \mu\text{m}$  a  $80 \mu\text{m}$  i transistori **n4** e **n4b** ora funzionano approssimativamente in saturazione.

Dato che si è verificato che tutti i transistori funzionano correttamente, possiamo iniziare a simulare l'amplificatore. Utilizzando il circuito di simulazione riportato in figura 5.10 si è eseguita una analisi in DC.

L'ingresso positivo dell'OTA è stato posto ad una tensione costante di 3 V, cioè la metà della tensione di alimentazione, mentre l'ingresso negativo è stato connesso ad un generatore di tensione variabile nell'intorno dei 3 V, tensione a cui ci aspettiamo che lavori l'OTA.

Nella figura 5.11 si può osservare la caratteristica di trasferimento statico dell'OTA, cioè la tensione di uscita in funzione di quella di ingresso.

Si possono distinguere tre zone:

- una zona di amplificazione caratterizzata da una relazione lineare [40]:

$$V_{OUT} = A_0 \cdot (V_{in} - V_{i0})$$

dove la pendenza della retta  $A_0$  non è altro che il guadagno ad anello aperto (open loop).

- due zone di saturazione (positiva e negativa) in cui il potenziale di uscita rimane costante per qualsiasi valore della tensione di ingresso.

Dunque l'OTA deve essere fatto funzionare, come tutti gli amplificatori operazionali, con gli ingressi posti a metà della tensione di alimentazione.

Ponendo dunque un DC bias di 3 V a tutti e due gli ingressi, ed un generatore di tensione a frequenza variabile solo sull'ingresso negativo, si è eseguita la simulazione AC, come riportato in figura 5.12. Si è aggiunto un carico capacitivo di  $0.8 \text{ pF}$  per poter ottenere gli stessi risultati riportati nell'articolo di riferimento dell'OTA [37] di figura 5.3.

La risposta in frequenza, cioè il guadagno  $A(j\omega)$  è riportato nella figura 5.13 ed è espresso in dB, mentre la fase del guadagno è riportata nella figura 5.14 ed è espressa in gradi sessagesimali.

Dalla risposta in frequenza possiamo notare che l'OTA possiede fondamentalmente 2 poli: il primo polo del circuito si trova a circa 10 kHz, con un guadagno

MODEL	MN2E	MN4D	MN4C	MN4B	MN4	MN2	M
ED	2.68066E-05	2.68066E-05	2.69220E-05	5.91352E-04	5.91352E-04	3.85485E-04	HOEP
VGS	1.08894E+00	9.09424E-01	1.77990E+00	1.77990E+00	1.77990E+00	1.09342E+00	-5.00000E-05
VDS	1.08894E+00	9.09424E-01	1.77990E+00	9.14949E-01	9.14949E-01	2.04941E+00	-1.26833E+00
VBS	0.00000E+00	0.00000E+00	0.00000E+00	0.00000E+00	0.00000E+00	0.00000E+00	-1.26833E+00
VTH	7.47579E-01	7.63526E-01	7.64948E-01	7.63502E-01	7.63502E-01	7.42513E-01	0.00000E+00
VDSAT	2.90495E-01	1.40418E-01	7.75723E-01	7.69414E-01	7.69414E-01	2.90793E-01	-7.8649E-01
GN	1.36291E-04	2.84271E-04	5.11546E-05	1.11088E-03	1.11088E-03	1.96439E-03	-3.9349E-01
GBS	1.55169E-06	1.77001E-06	5.23649E-07	2.79494E-05	2.79494E-05	1.63853E-05	1.90863E-04
GBE	3.93887E-05	9.06057E-05	1.39209E-05	2.98447E-04	2.98447E-04	5.64723E-04	1.78354E-06
CRD	1.18076E-14	9.46828E-14	3.01078E-15	5.29550E-14	5.29550E-14	1.32680E-13	5.71759E-05
CRS	1.64010E-14	1.26162E-13	4.72926E-15	7.06505E-14	7.06505E-14	2.16998E-13	3.82336E-14
CCS	3.21186E-14	2.92369E-13	1.24222E-14	2.02247E-13	2.02247E-13	4.49607E-13	5.95912E-14
CGD	6.61110E-15	5.27436E-14	1.83246E-15	5.05861E-14	5.05861E-14	7.50369E-14	1.50509E-13
CCB	9.26487E-16	1.33094E-14	6.34584E-16	1.90458E-15	1.90458E-15	8.86692E-15	2.62942E-14
MN3C		MN3B		MN5	MN6		MN1B
MODEL	MODP						
ED	-2.6807E-05	-2.6822E-05	-3.8548E-04	-3.8548E-04	-3.8548E-04	-2.0587E-04	-2.0587E-04
VGS	-1.2683E+00	-1.2683E+00	-1.6657E+00	-1.6657E+00	-1.6657E+00	-1.3261E+00	-1.3261E+00
VDS	-3.9349E+00	-4.2201E+00	-1.6657E+00	-1.6657E+00	-1.6657E+00	-3.4112E+00	-3.4112E+00
VBS	0.00000E+00	0.00000E+00	1.36993E+00	1.36993E+00	1.36993E+00	0.00000E+00	1.67389E+00
VTH	-7.8380E-01	-7.8363E-01	-1.1006E+00	-1.1006E+00	-1.1006E+00	-7.7546E-01	-1.1521E+00
VDSAT	-3.9396E-01	-3.9612E-01	-4.8656E-01	-4.8656E-01	-4.7127E-01	-4.7127E-01	-1.7788E-01
GN	1.02089E-04	1.02503E-04	1.26294E-03	1.26294E-03	1.20477E-03	1.20477E-03	1.92818E-03
GBS	5.11001E-07	4.9892E-07	2.02279E-05	2.02279E-05	2.21347E-05	2.21347E-05	9.63166E-06
GBE	5.0462E-05	3.05417E-05	2.25590E-04	2.25590E-04	3.42133E-04	3.42133E-04	3.33264E-04
CRD	1.27842E-14	1.24026E-14	3.5823E-14	3.5823E-14	1.13523E-13	1.13523E-13	2.70661E-13
CRS	9.08315E-14	3.08915E-14	1.13523E-13	1.13523E-13	1.81157E-13	1.81157E-13	4.20009E-13
CCS	7.85799E-14	7.85799E-14	3.01410E-13	3.01410E-13	3.01410E-13	3.01410E-13	1.19411E-12
CGD	1.00565E-14	9.96966E-15	6.43894E-14	6.43894E-14	6.72527E-14	6.72527E-14	2.26370E-13
CCB	1.41406E-15	1.41379E-15	3.46749E-15	3.46749E-15	3.40705E-15	3.46705E-15	2.55159E-14

Figura 5.9: DC operating point del circuito schematico riportato in figura 5.4 in cui si è cambiata la  $W(n4d)$  da  $100 \mu\text{m}$  a  $80 \mu\text{m}$ .

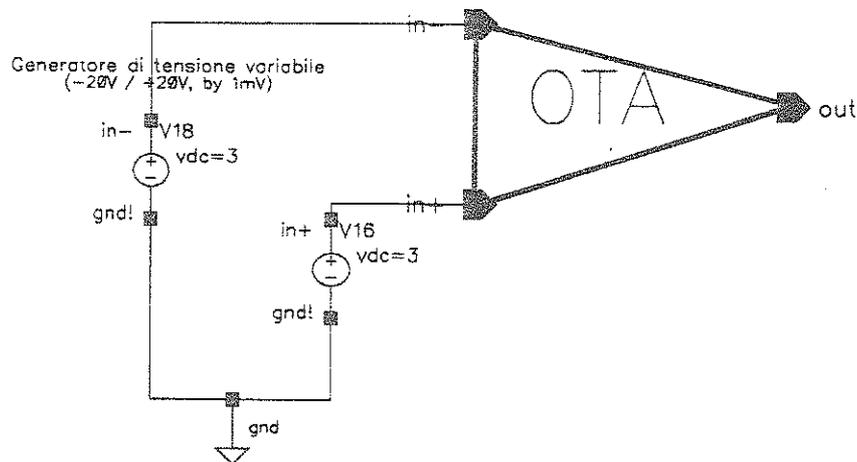


Figura 5.10: Circuito schematico utilizzato per l'analisi in DC.

di 70 dB, il che rende l'OTA un amplificatore relativamente veloce, mentre il secondo polo si trova a circa 100 MHz, con un guadagno di circa 5 dB. Il guadagno unitario, cioè lo zero delle ordinate ( $\text{Log}_{10}(1)=0$ ), si trova un po' al di sotto dei 200 MHz. La differenza, comunque piccola, rispetto ai 220 MHz previsti nella figura 5.3, è da imputare alla differenza di modelli utilizzati, e quindi di capacità parassite calcolate dal simulatore.

La fase del guadagno differisce invece in modo più marcato rispetto alla figura 5.3. Ora infatti, in corrispondenza della frequenza a guadagno unitario, la fase è di  $30^\circ$  invece che di  $45^\circ$ .

Questo valore, che rappresenta il *margin di fase* (si veda figura 5.15), è un indice della stabilità dell'amplificatore, e ci dice che il circuito dell'OTA implementato potrebbe oscillare in una configurazione controreazionata, come quella che si vuole utilizzare per integrare la corrente proveniente dal rivelatore su una pipeline di capacità.

Si cercherà dunque di studiare la stabilità dell'OTA controreazionato ed eventualmente di compensarlo. In quanto segue si utilizzerà il circuito presentato in figura 5.16 dove le alimentazioni sono state portate a +3 V e -3 V per poter lavorare con gli ingressi a terra, e dove si è modificata la larghezza  $W$  del transistor  $n4d$ , come già precedentemente spiegato.

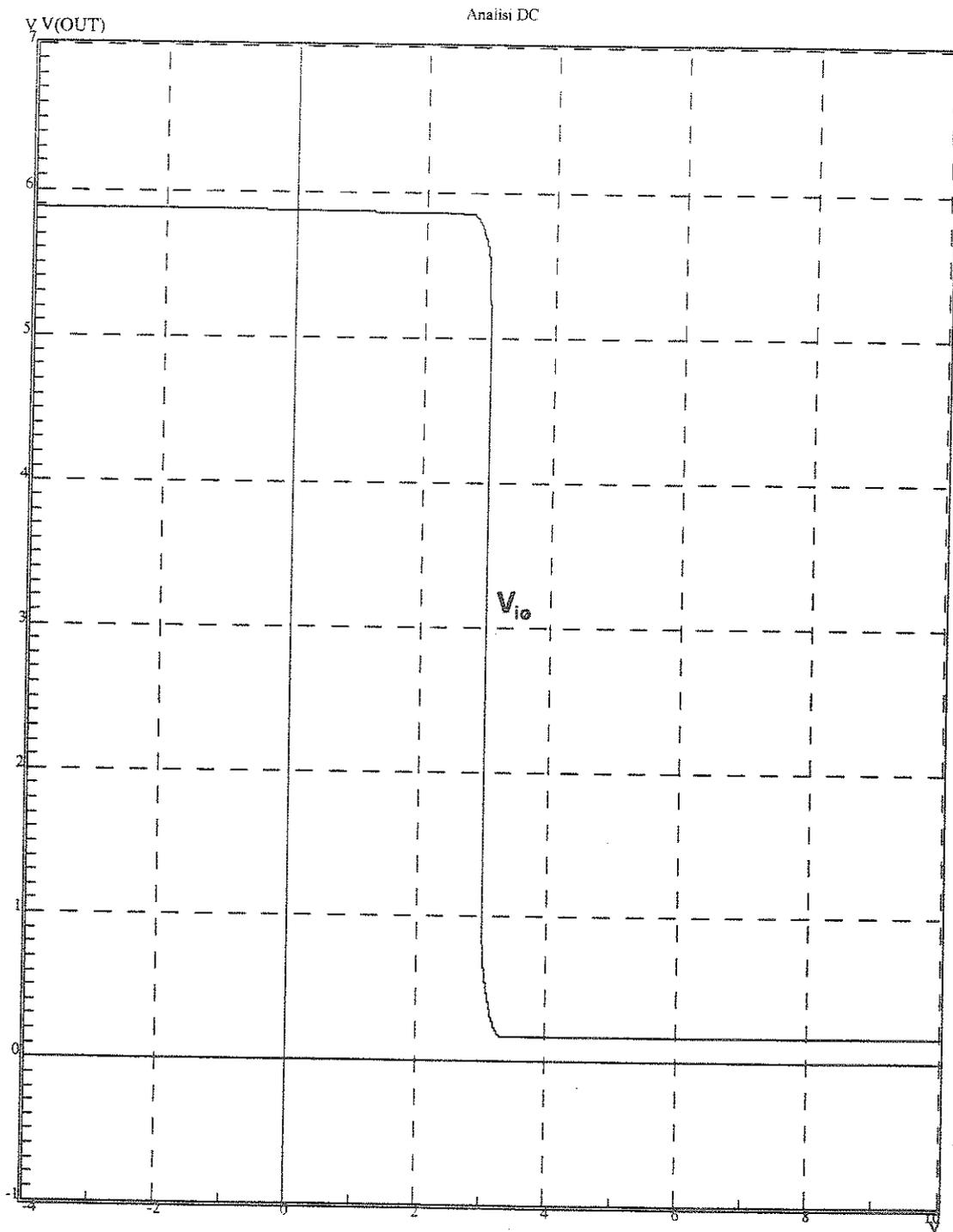


Figura 5.11: Caratteristica di trasferimento statico dell'OTA:  $V_{OUT} = V_{OUT}(V_{in})$ .

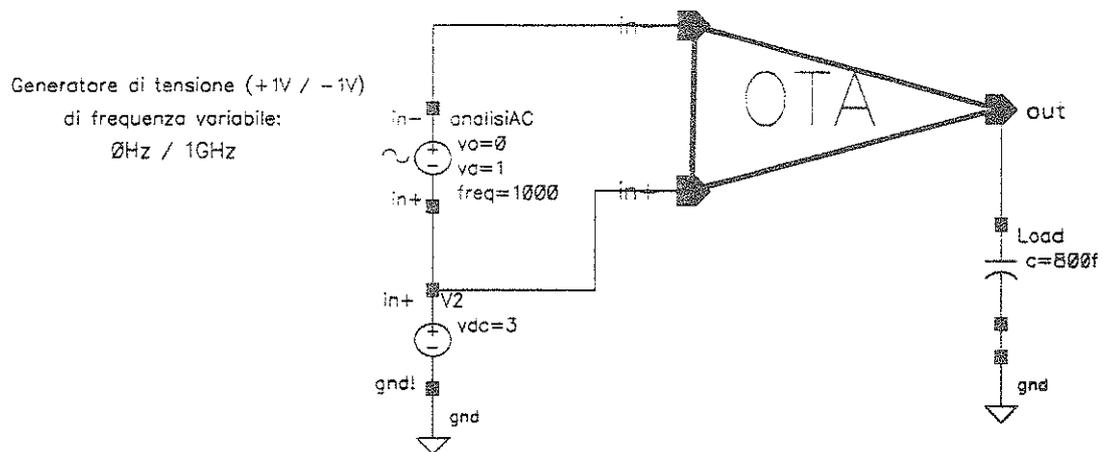


Figura 5.12: Circuito schematico utilizzato per l'analisi in AC.

### 5.2.1 Risposta in frequenza e stabilità degli amplificatori controeazionati

- La retroazione ideale [38].

Nella figura 5.17 è riportata la configurazione della retroazione ideale. La rete di controeazione è normalmente lineare e costituita da elementi passivi (nel nostro caso saranno condensatori); la sua funzione di trasferimento  $f$  fornisce all'ingresso un segnale  $S_{fb}$ . Assumendo che il nodo di ingresso sia un *nodo sottrattore* (come capita negli amplificatori retroazionati negativamente), all'amplificatore di andata, che ha una funzione di trasferimento  $a$ , è dunque fornito il *segnale errore*:

$$S_e = S_i - S_{fb}$$

differenza tra il *segnale di ingresso* ed il *segnale di reazione*. Assumendo inoltre che la rete di controeazione non costituisca un carico per l'amplificatore di andata si ha:

$$S_o = a \cdot S_e$$

$$S_{fb} = f \cdot S_o$$

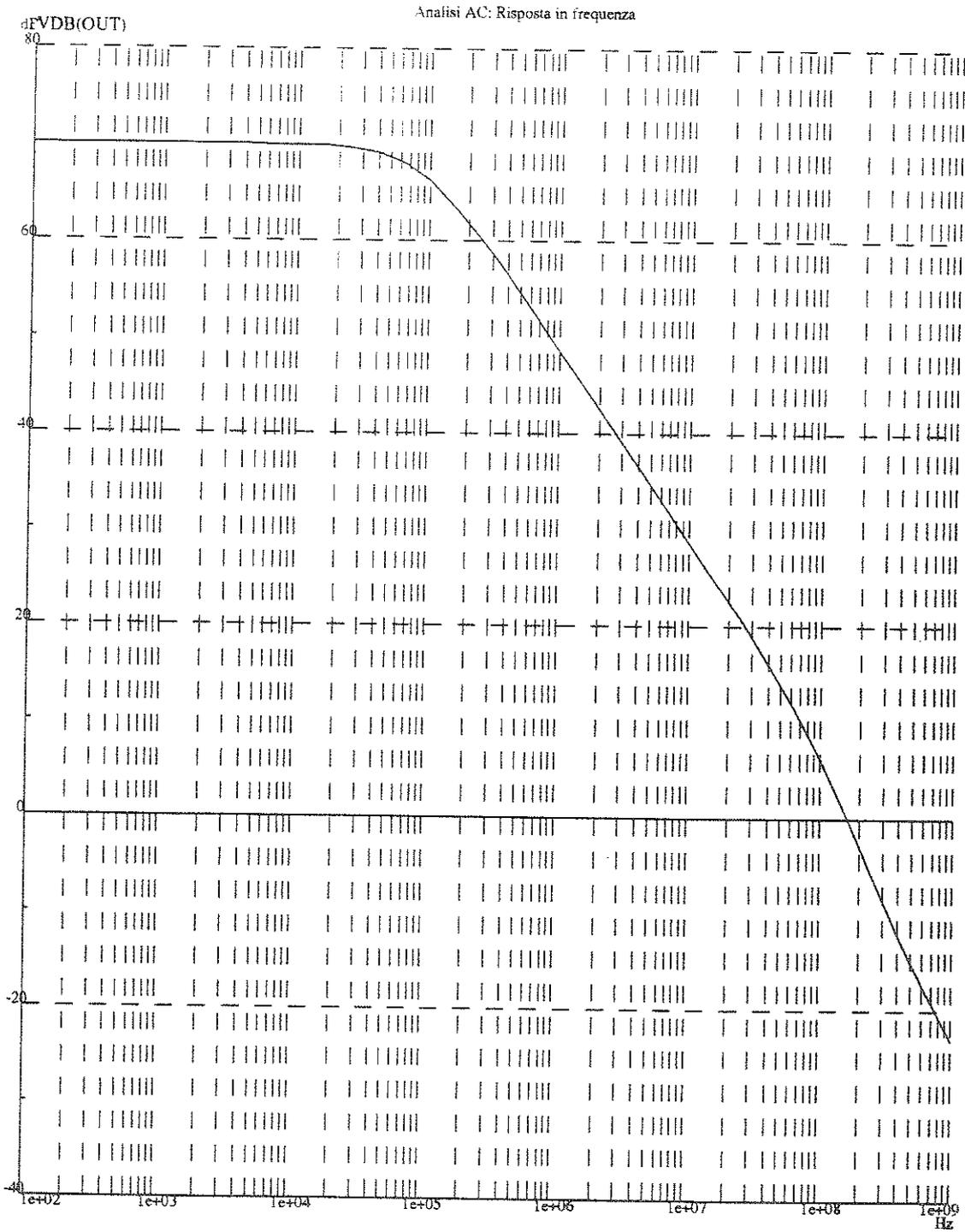


Figura 5.13: Analisi AC: guadagno dell'OTA ( $20 \cdot \text{Log}_{10}(V_{OUT}/V_{in})$ ) in funzione della frequenza, con 0.8 pF di carico.

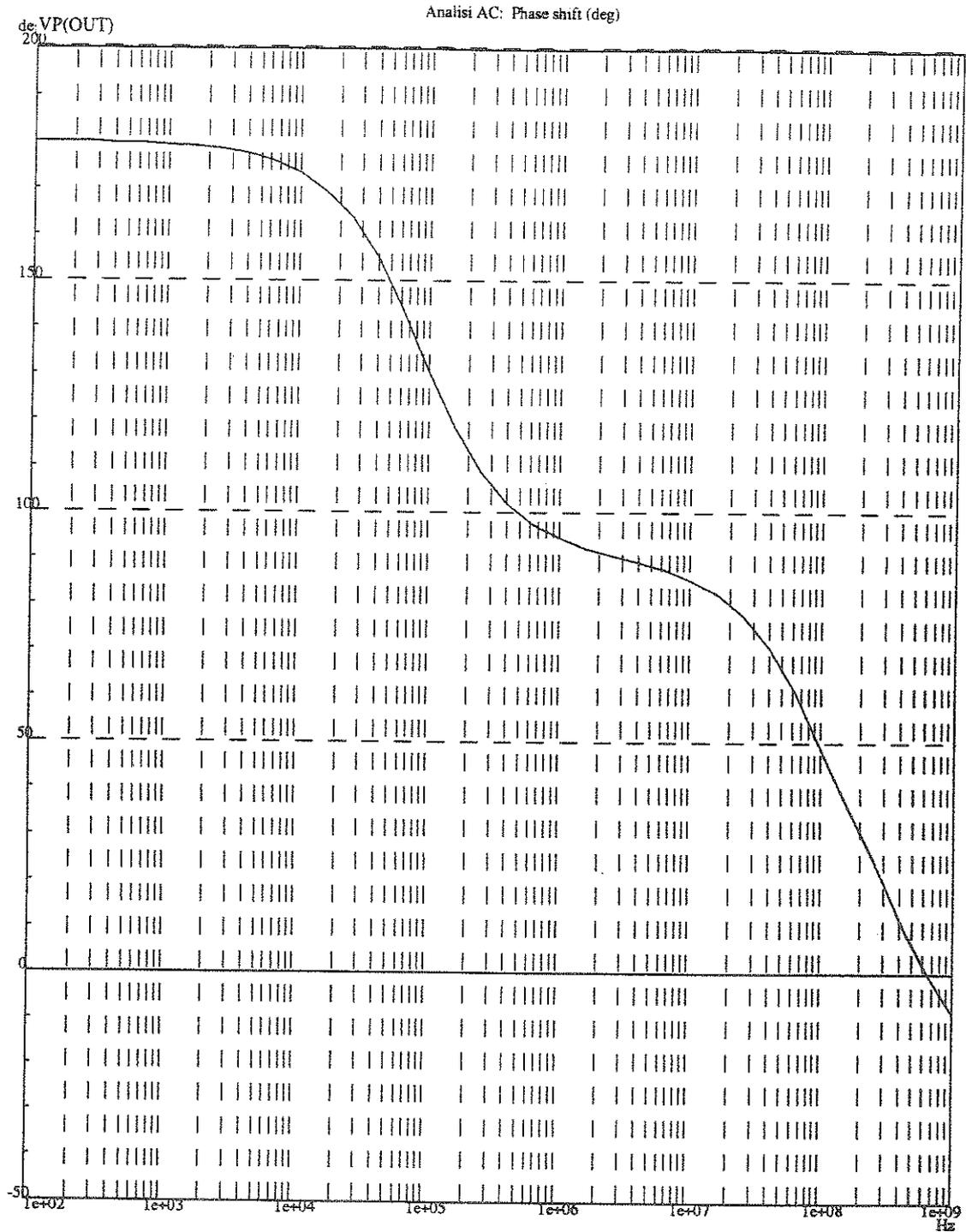


Figura 5.14: Analisi AC: fase del guadagno dell'OTA ( $ph(V_{OUT}/V_{in})$ ) in funzione della frequenza, con 0.8 pF di carico.

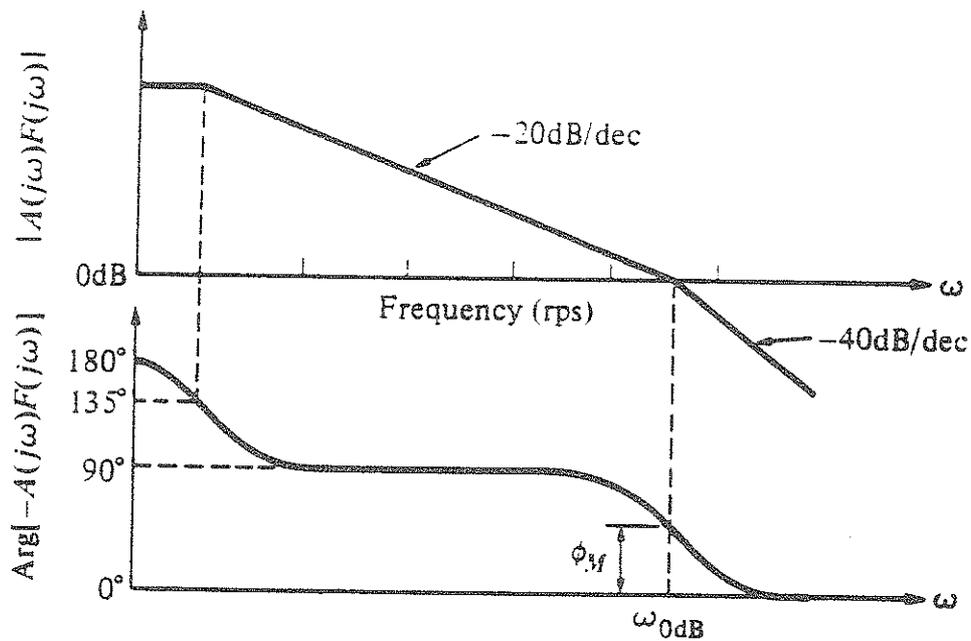


Figura 5.15: Guadagno e fase in funzione della frequenza di un sistema del secondo ordine (a due poli). È indicato il margine di fase  $\phi_M$  [39].

e dunque:

$$S_e = S_i - f \cdot S_0$$

Sostituendo l'ultima equazione scritta nella prima si ha:

$$S_0 = a \cdot S_i - a \cdot f \cdot S_0$$

da cui si ricava il guadagno *ad anello chiuso* di un amplificatore, cioè quando è applicata la retroazione:

$$A = \frac{S_0}{S_i} = \frac{a}{1 + a \cdot f} \quad (5.6)$$

È utile definire il *guadagno d'anello*  $T = a \cdot f$ , che rappresenta il guadagno totale lungo l'anello di reazione. Si ha quindi:

$$A = \frac{S_0}{S_i} = \frac{a}{1 + T} \approx \frac{1}{f} (\text{per } T \gg 1) \quad (5.7)$$

Allora per valori elevati del guadagno d'anello  $T$ , il guadagno dell'amplificatore è determinato dalla funzione di trasferimento  $f$  della retroazione.

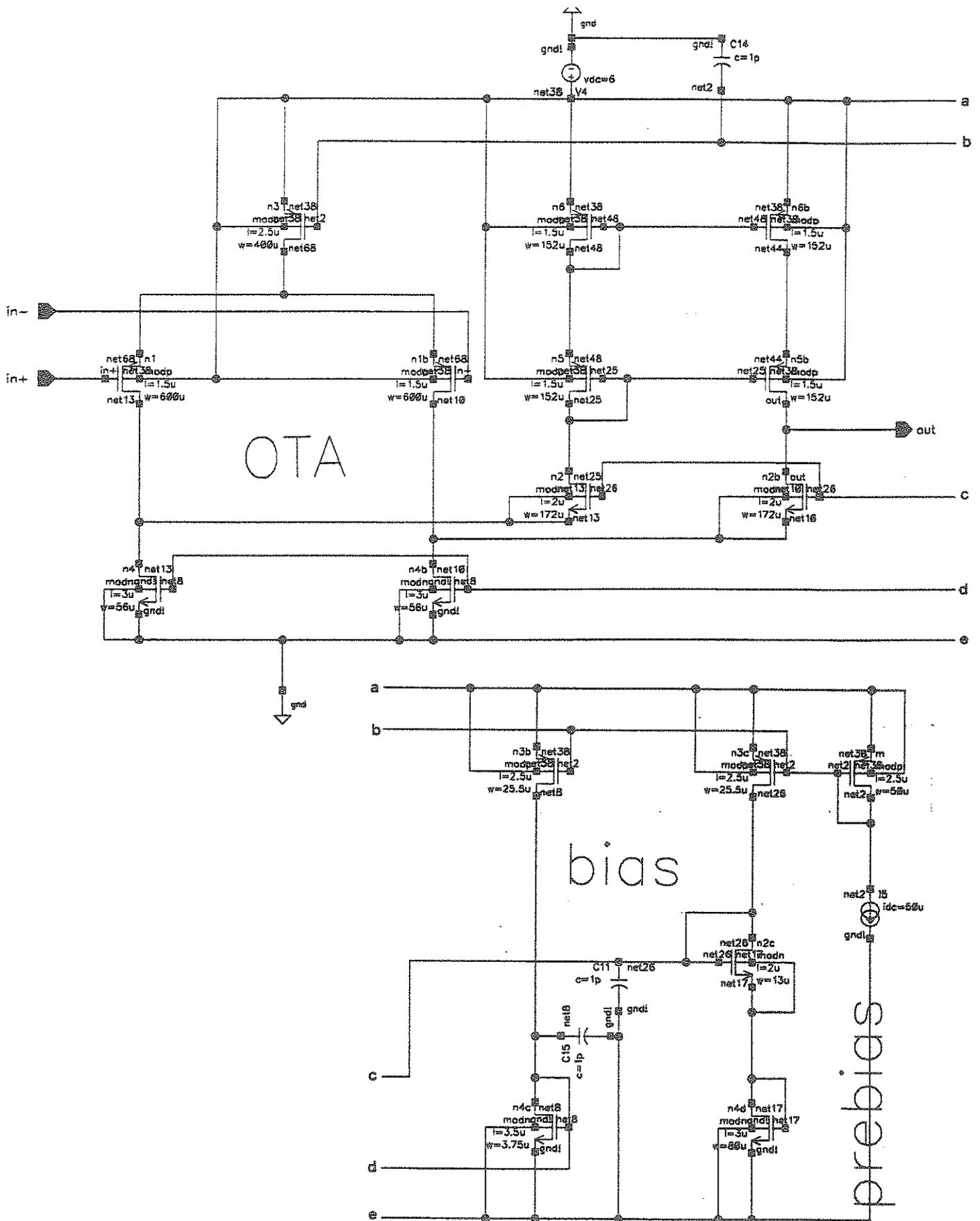


Figura 5.16: Schematico dell'OTA utilizzato per le simulazioni che seguono.

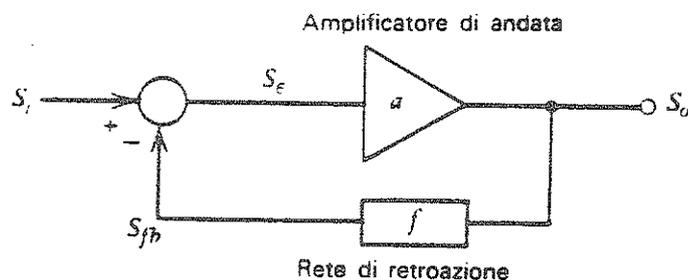


Figura 5.17: Configurazione della retroazione ideale

La rete di retroazione agisce forzando  $S_{fb}$  ad essere quasi uguale a  $S_i$ . Si può infatti ricavare:

$$\frac{S_{fb}}{S_i} = \frac{T}{1+T} \simeq 1 (\text{per } T \gg 1)$$

e quindi il segnale di retroazione replica il segnale di ingresso. Essendo poi  $S_{fb} = f \cdot S_o$ , se  $f \leq 1$  si ha che il segnale in uscita è una replica amplificata del segnale di ingresso, che è lo scopo di un amplificatore retroazionato [38].

• **Instabilità ed il criterio di Nyquist [38].**

Si consideri un amplificatore con una funzione di trasferimento a tre poli:

$$a(j\omega) = \frac{a_0}{\left(1 - \frac{j\omega}{p_1}\right) \left(1 - \frac{j\omega}{p_2}\right) \left(1 - \frac{j\omega}{p_3}\right)} \quad (5.8)$$

dove  $p_1, p_2$  e  $p_3$  sono le frequenze dei poli in radianti al secondo.

Nella figura 5.18 sono riportati il guadagno e la fase in funzione della frequenza, ipotizzando un fattore 10 di separazione tra i poli. A frequenze superiori a quelle del primo polo  $p_1$  si ha che la fase si avvicina a  $-90^\circ$ , dopo il secondo polo  $p_2$  la fase si avvicina a  $-180^\circ$ , mentre dopo il terzo polo  $p_3$  si hanno circa  $-270^\circ$  di fase. La frequenza alla quale la fase vale  $-180^\circ$  ha un significato particolare ed è indicata con  $\omega_{180}$ . Se i poli sono abbastanza distanti tra loro (un fattore 10 o più,

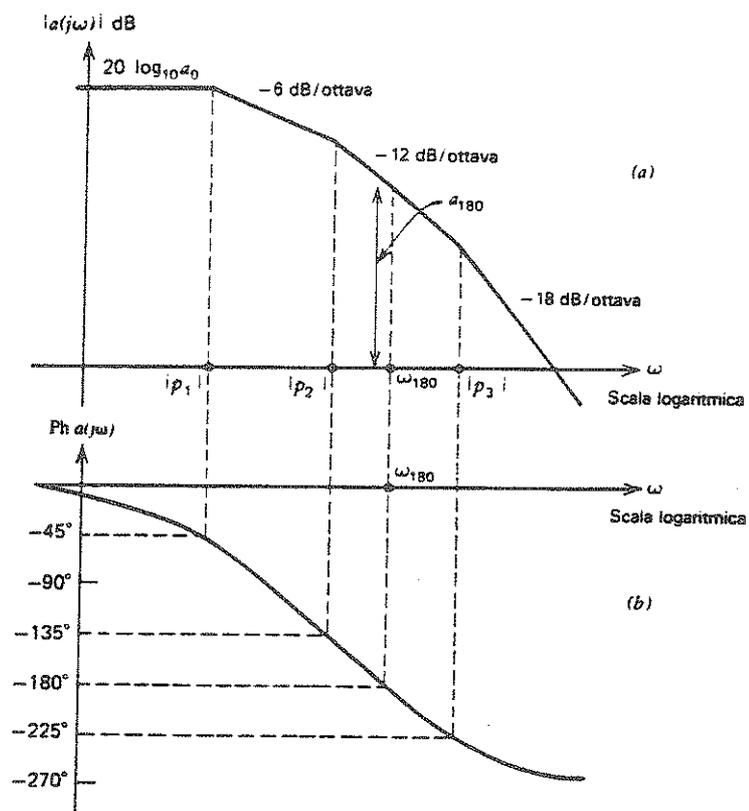


Figura 5.18: Guadagno e fase in funzione della frequenza per un circuito con una funzione di trasferimento a tre poli [38].

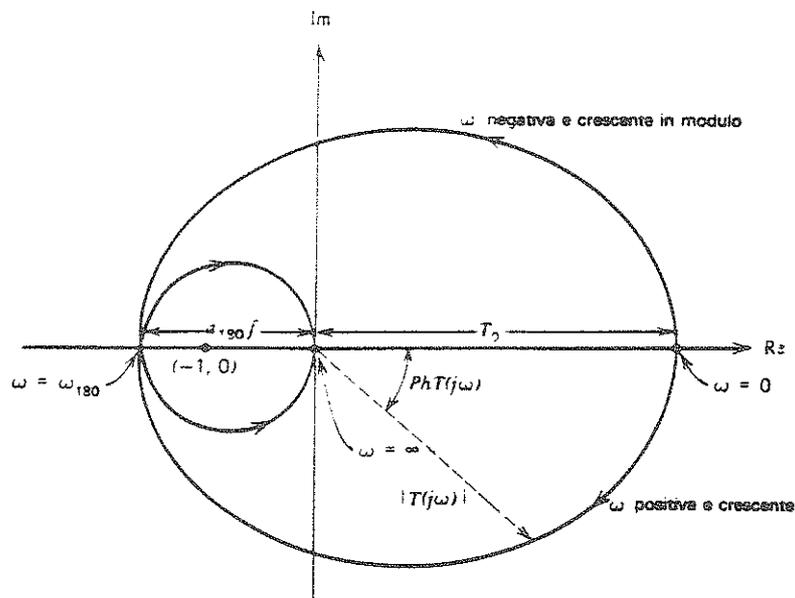


Figura 5.19: Rappresentazione polare di  $T(j\omega)$  in ampiezza e fase (non in scala), detto *diagramma di Nyquist*, riferito alla figura 5.18 [38].

come ipotizzato in questo esempio), gli sfasamenti delle relative frequenze sono approssimativamente  $-45^\circ$  per  $p_1$ ,  $-135^\circ$  per  $p_2$  e  $-225^\circ$  per  $p_3$ .

Si supponga ora che questo amplificatore sia connesso in un anello di reazione come quello di figura 5.17, e che il cammino di reazione sia puramente resistivo, cosicché la funzione di retroazione  $f$  sia una costante. In queste condizioni il guadagno d'anello  $T(j\omega) = a(j\omega)f$  ha lo stesso andamento con la frequenza del guadagno dell'amplificatore  $a(j\omega)$ .

Con i dati riportati nella figura 5.18 si può tracciare un diagramma polare (in ampiezza e fase) del guadagno d'anello. Una rappresentazione di questo tipo prende il nome di *diagramma di Nyquist* ed è riportato nella figura 5.19, riferita all'esempio considerato.

- Per  $\omega=0$  si ha  $T(j\omega) = T_0$  e la fase  $ph(T(j\omega)) = 0^\circ$ . Al crescere di  $\omega$  il modulo del guadagno  $a(j\omega)$  diminuisce e la sua fase diviene negativa: il diagramma si trova nel quarto quadrante.
- All'estremo opposto, quando  $\omega \rightarrow \infty$  si ha che la fase del guadagno  $ph(a(j\omega)) \rightarrow -270^\circ$ , mentre il modulo del guadagno tende a 0. Quindi

in questo punto il diagramma è asintotico all'origine e tangente all'asse immaginario.

- Per  $\omega = \omega_{180}$  si ha che la fase del guadagno  $a(j\omega)$  vale  $-180^\circ$ , e la curva attraversa l'asse reale negativo. Se  $T(j\omega_{180}) \gg 1$  in questo punto il diagramma di Nyquist circonda il punto  $(-1,0)$  (come si vede nella figura 5.19).

Siamo ora in grado di enunciare il criterio di Nyquist per la stabilità dell'amplificatore:

*se il diagramma di Nyquist circonda il punto  $(-1,0)$  l'amplificatore è instabile.*

Per capirne il significato si consideri il caso in cui il diagramma di Nyquist passi per il punto  $(-1,0)$ . In questa condizione si ha che alla frequenza  $\omega_{180}$  il guadagno d'anello vale  $T(j\omega_{180}) = -1$  e quindi per l'equazione 5.6 si ha che  $A(j\omega_{180}) \rightarrow \infty$ . Si ottiene così che l'amplificatore reazionato ha un guadagno infinito e ciò indica l'inizio della instabilità e dell'oscillazione.

Questa situazione corrisponde a poli di  $A(j\omega)$  sull'asse immaginario del piano complesso poiché essendo  $T(j\omega) = a(j\omega)f = -1$ , dove  $f \in \mathcal{R}$ , deve essere  $Im(a) = 0$ ; ricordando l'equazione 5.8, dove  $a_0 \in \mathcal{R}$ , ed esprimendo i poli come  $p_i = \sigma_i \pm j\omega_i$ , si ha che  $\sigma_i = 0$  e dunque i poli si trovano sull'asse  $j\omega$ .

Se  $T_0$  viene ora aumentato facendo crescere  $a_0$  o  $f$ , il diagramma di Nyquist si espande linearmente e circonda il punto  $(-1,0)$ .

Questo corrisponde a poli di  $A(j\omega)$  nel semipiano destro.

Si può ora esprimere in un'altra forma il criterio di stabilità di Nyquist: *se il modulo del guadagno d'anello  $T(j\omega)$  è maggiore di 1 alla frequenza  $\omega_{180}$  a cui la fase  $phT(j\omega) = -180^\circ$ , allora l'amplificatore è instabile.*

Mano a mano che  $mod(T(j\omega_{180}))$  viene fatto avvicinare all'unità, l'amplificatore ha un margine di stabilità minore e questo si può specificare in due modi:

- Il più comune, il *margine di fase* viene definito come:  

$$\text{margine di fase} = \phi_M \equiv 180^\circ + ph(T(j\omega_0))$$
dove  $\omega_0$  è la frequenza a cui il modulo del guadagno d'anello è unitario, cioè la frequenza a cui il diagramma di Bode (risposta in frequenza) interseca l'asse delle ascisse.  $\phi_M$  deve essere maggiore di  $0^\circ$  perché si abbia stabilità.
- Il *margine di guadagno* definito come:  

$$\text{margine di guadagno} \equiv mod(T(j\omega_{180}))$$
e deve essere minore di 0 dB perché si abbia stabilità.

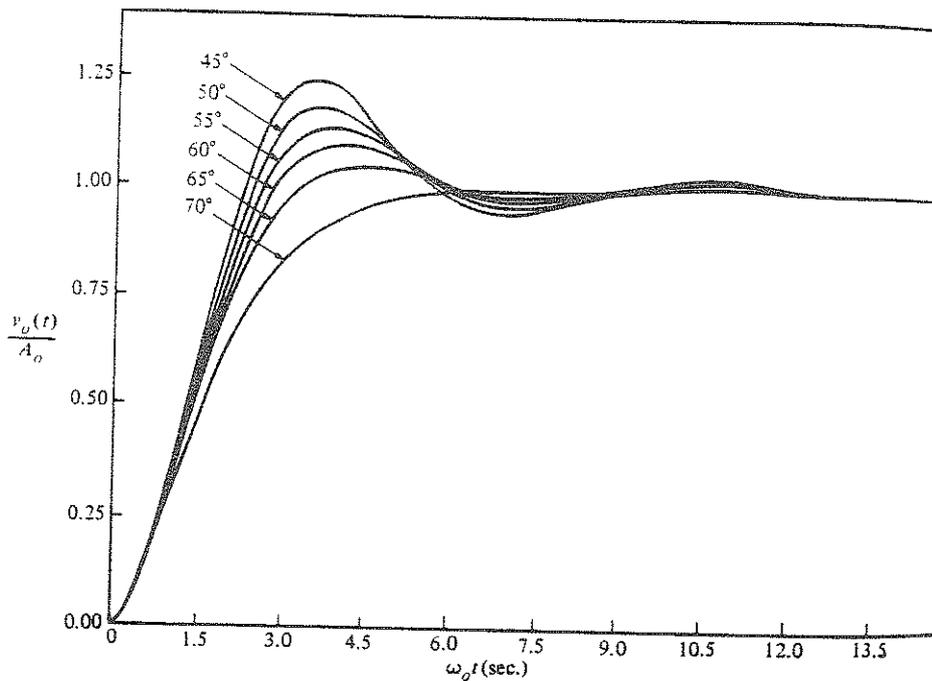


Figura 5.20: Guadagno complessivo normalizzato per amplificatori controreazionati in funzione della frequenza per vari margini di fase  $\phi_M$  [39].

Nella figura 5.20 sono riportati i guadagni normalizzati in funzione della frequenza per vari margini di fase [39]. Le curve sono tracciate nell'ipotesi che la risposta sia dominata dai primi due poli della funzione di trasferimento. Come si può notare più il margine di fase diminuisce, più le oscillazioni aumentano e l'amplificatore controreazionato diviene più instabile.

Quando un amplificatore collegato in "feedback" (cioè controreazionato) oscilla, occorre compensarlo per aumentare il margine di fase.

L'obiettivo della compensazione è raggiungere un margine di fase:  $\phi_M \geq 45^\circ$  [39].

### 5.2.2 Compensazione dell'OTA

Alla luce di quanto esposto nel paragrafo precedente ci accingiamo ora a verificare la stabilità dell'OTA.

Utilizzando il circuito controreazionato riportato in figura 5.21 si è eseguita una analisi AC. L'OTA è collegato in feedback con un cammino di reazione puramente resistivo, da cui si deduce che  $f$  è reale e finita.

I risultati sono riportati nel grafico di figura 5.22. Come si può notare, con-

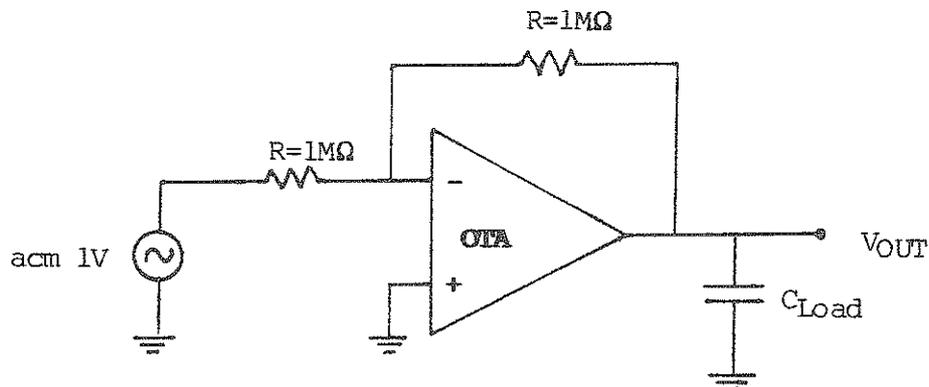


Figura 5.21: Circuito schematico utilizzato per controreazionare l'OTA e verificarne la stabilità.

frontandola con la figura 5.23, che l'OTA è instabile e possiede un margine di fase inferiore ai  $30^\circ$ .

Occorre quindi compensare l'amplificatore OTA. Come precedentemente spiegato, il "folded cascode" si compensa aggiungendo una semplice capacità di carico in uscita. Utilizzando lo schema circuitale riportato nella figura 5.12 si è quindi eseguita una simulazione AC con differenti valori della capacità di carico e si è letto, dal grafico, il valore del margine di fase. La risposta in frequenza riportata in figura 5.24 si riferisce all'analisi AC dell'OTA con  $200 \text{ fF}$  di carico in uscita, mentre la risposta in frequenza riportata in figura 5.25 si riferisce ad un carico di  $1 \text{ pF}$  in uscita.

L'aggiunta di una capacità di carico in uscita sposta verso sinistra l'intera risposta in frequenza. Quindi l'intersezione del diagramma di Bode con l'asse delle ascisse viene spostato a sinistra ed intercetta un margine di fase maggiore. Le simulazioni forniscono un margine di fase di  $\sim 22^\circ$  quando si hanno  $200 \text{ fF}$  di carico, e  $\sim 38^\circ$  quando si ha  $1 \text{ pF}$  di carico.

Per raggiungere un valore di  $\phi_M$  che garantisca la stabilità (cioè  $\geq 45^\circ$ ) occorrerebbe aggiungere valori di capacità di carico ancora più elevati.

Dato che si dovranno avere due amplificatori per ogni canale, e che si avranno almeno un centinaio di canali per ogni chip, la soluzione di compensare l'OTA con una grande ( $\gg 1 \text{ pF}$ ) capacità di uscita non sembra essere la migliore.

Allo scopo di chiarire le idee circa le superfici occupate dai componenti attivi

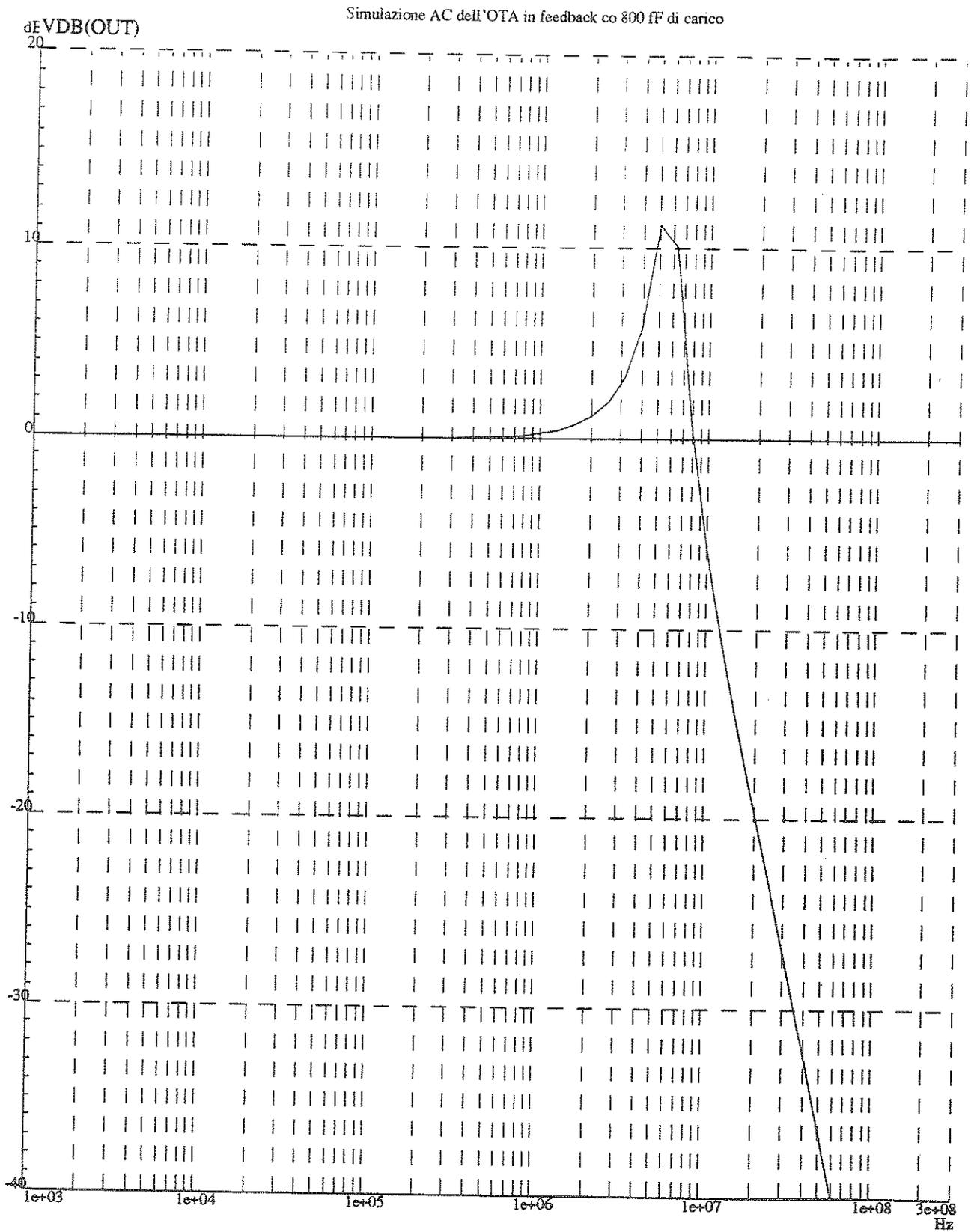


Figura 5.22: Analisi AC: Guadagno normalizzato in funzione della frequenza per il circuito controreazionato riportato in figura 5.21.

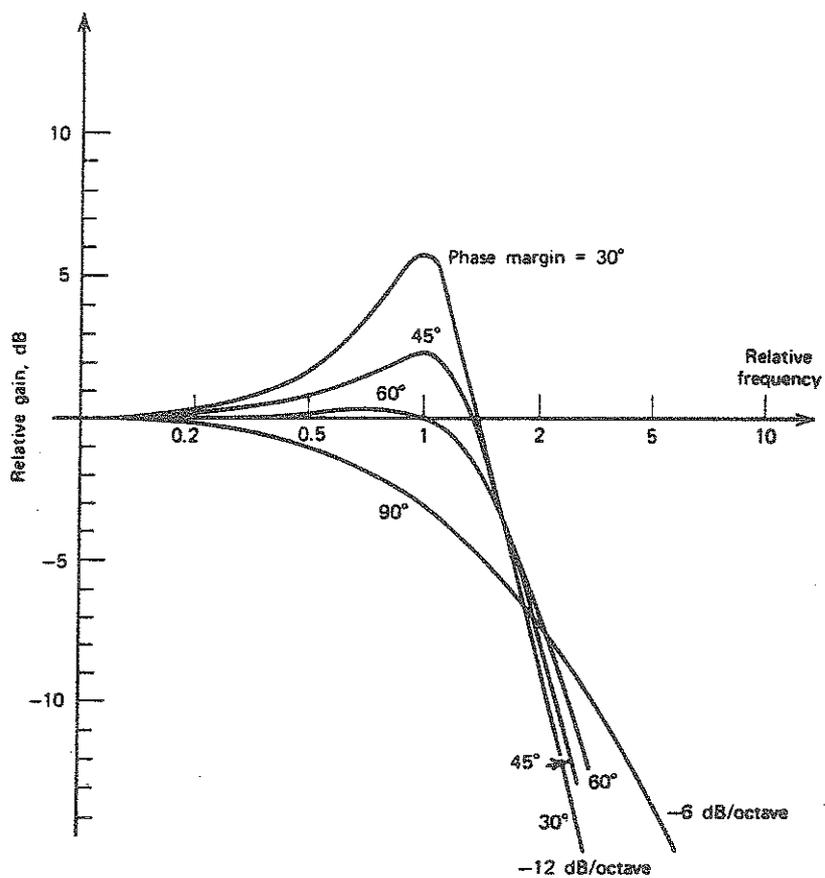


Figura 5.23: Guadagno complessivo normalizzato per amplificatori controreazionati in funzione della frequenza normalizzata per vari margini di fase. La frequenza è normalizzata alla frequenza a cui il guadagno d'anello  $T(j\omega)$  è unitario [38].

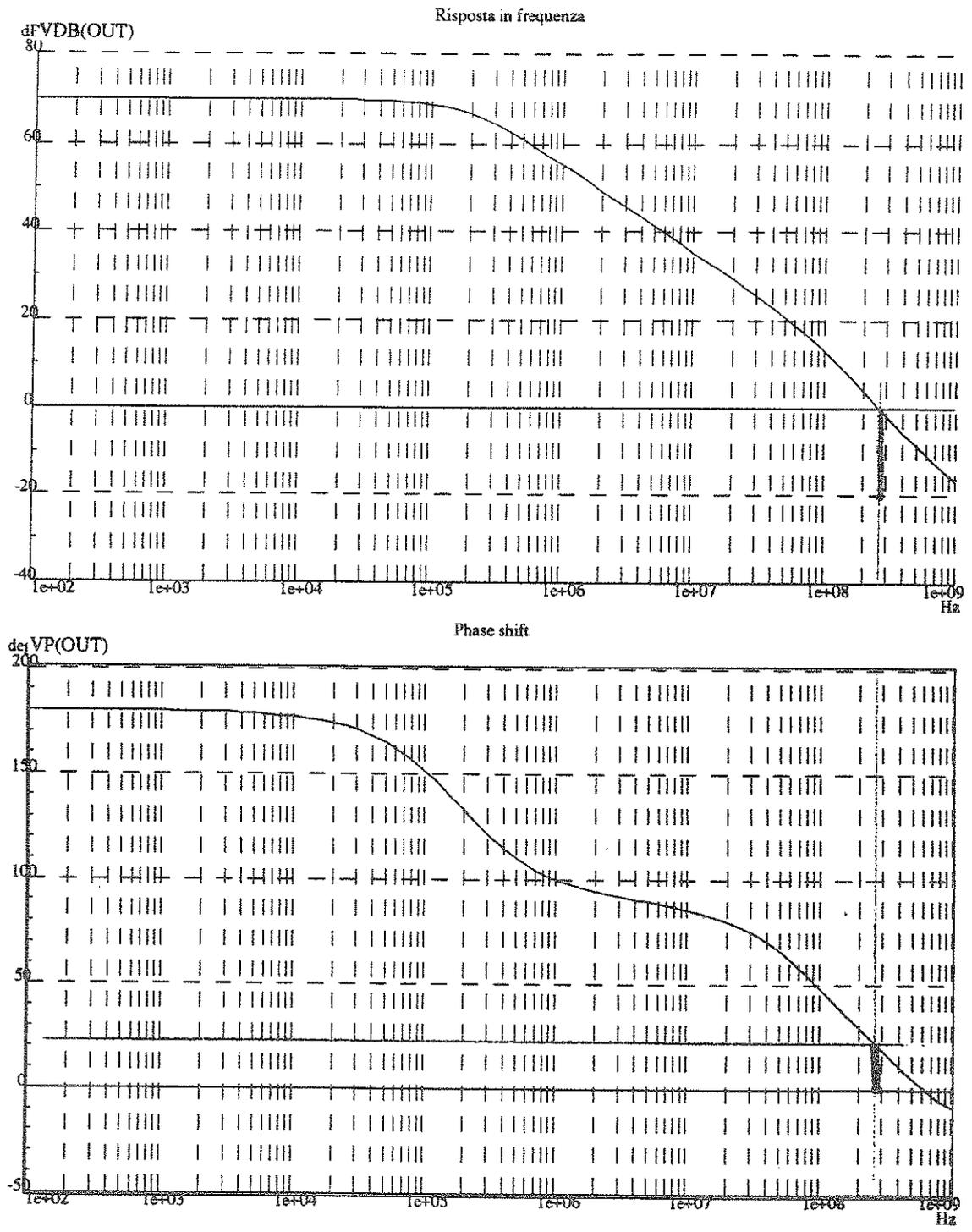


Figura 5.24: Guadagno e fase dell'OTA in funzione della frequenza con una capacità di carico in uscita di 200 fF.

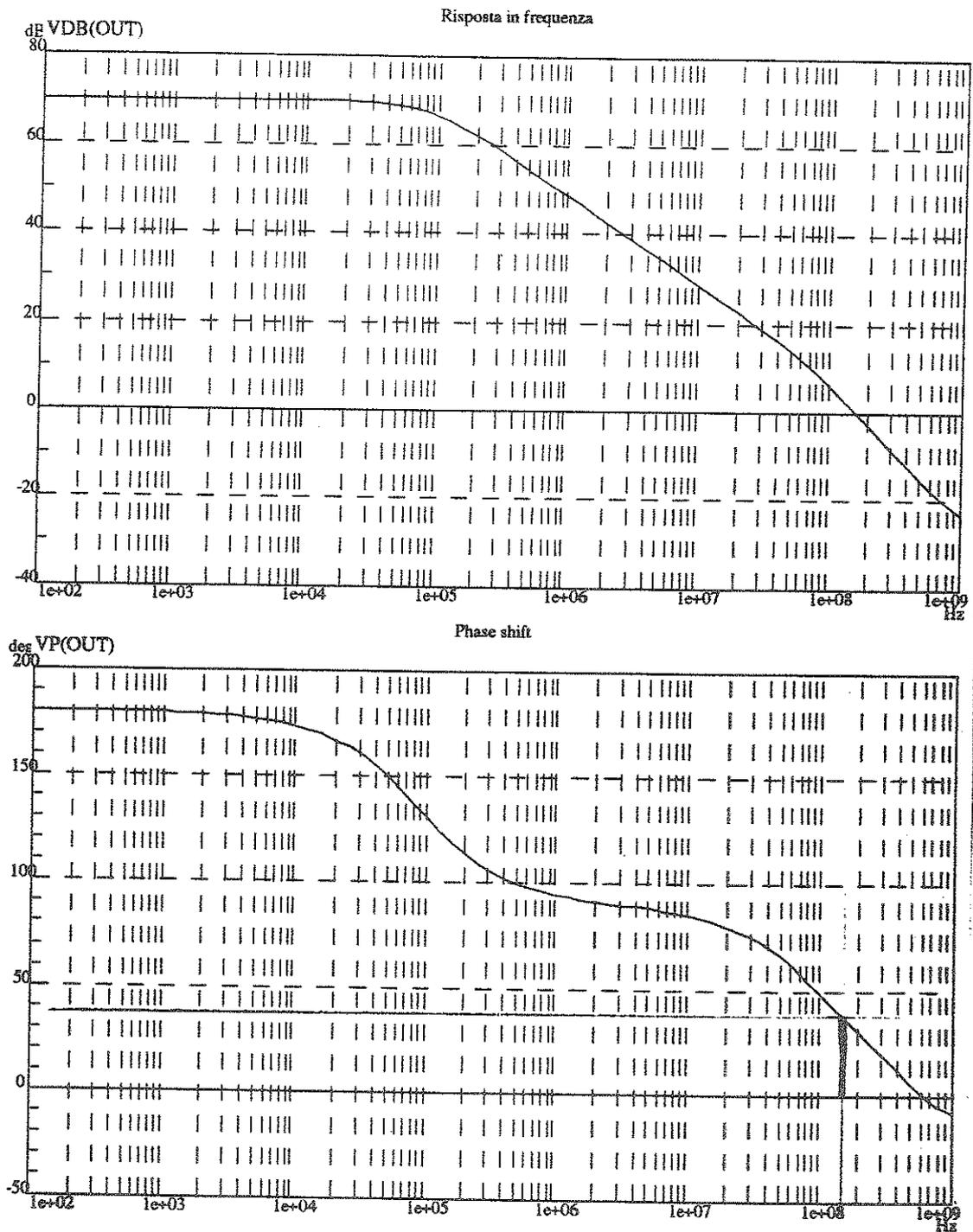


Figura 5.25: Guadagno e fase dell'OTA in funzione della frequenza con una capacità di carico in uscita di 1 pF.

(transistor) e passivi (prevalentemente capacità) su un chip, si riporta nella figura 5.26 lo schematico, il layout e la microfotografia di un comune amplificatore operativo [39].

La tecnologia utilizzata è ormai superata, ma le dimensioni relative dei vari dispositivi possono considerarsi le stesse. Come si può vedere, la capacità di compensazione  $C_C$  occupa gran parte dell'intero chip.

Si è quindi provato a compensare l'OTA anticipando il secondo polo, il che fa diminuire la frequenza di guadagno unitario e conseguentemente avere un aumento del margine di fase: purtroppo non si è riusciti ad incrementare il  $\phi_M$  di molto.

Anche riducendo di poche decine di dB il guadagno, riducendo la corrente che alimenta i transistori di ingresso, non si riesce ad aumentare il margine di fase oltre i 30 gradi.

Rimangono dunque due soluzioni per compensare l'OTA:

*i)* ridisegnare parte della polarizzazione (bias) in modo da avere delle correnti nell'OTA tali da aumentare l'impedenza di uscita;

*ii)* porre un secondo stadio di amplificazione e compensare in modo classico l'operazionale: interponendo cioè una capacità tra il primo ed il secondo stadio.

÷

*i)* attraverso gli specchi di corrente presenti nella polarizzazione si è forzata una corrente di  $100 \mu\text{A}$  nei transistori n2 e n2b, invece dei precedenti  $\simeq 200 \mu\text{A}$ . Essendo:

$$g_m = \sqrt{2K_p \cdot I_{DS} \frac{W}{L}}$$

si ha che la  $g_m$  dei transistori di uscita diminuisce ed automaticamente la resistenza aumenta:  $R = \frac{1}{g_m}$ . In tali condizioni si sono trovati i seguenti valori:

- primo polo: 62 kHz ( si è spostato a sinistra di quasi due decadi);
- secondo polo: 63 MHz;

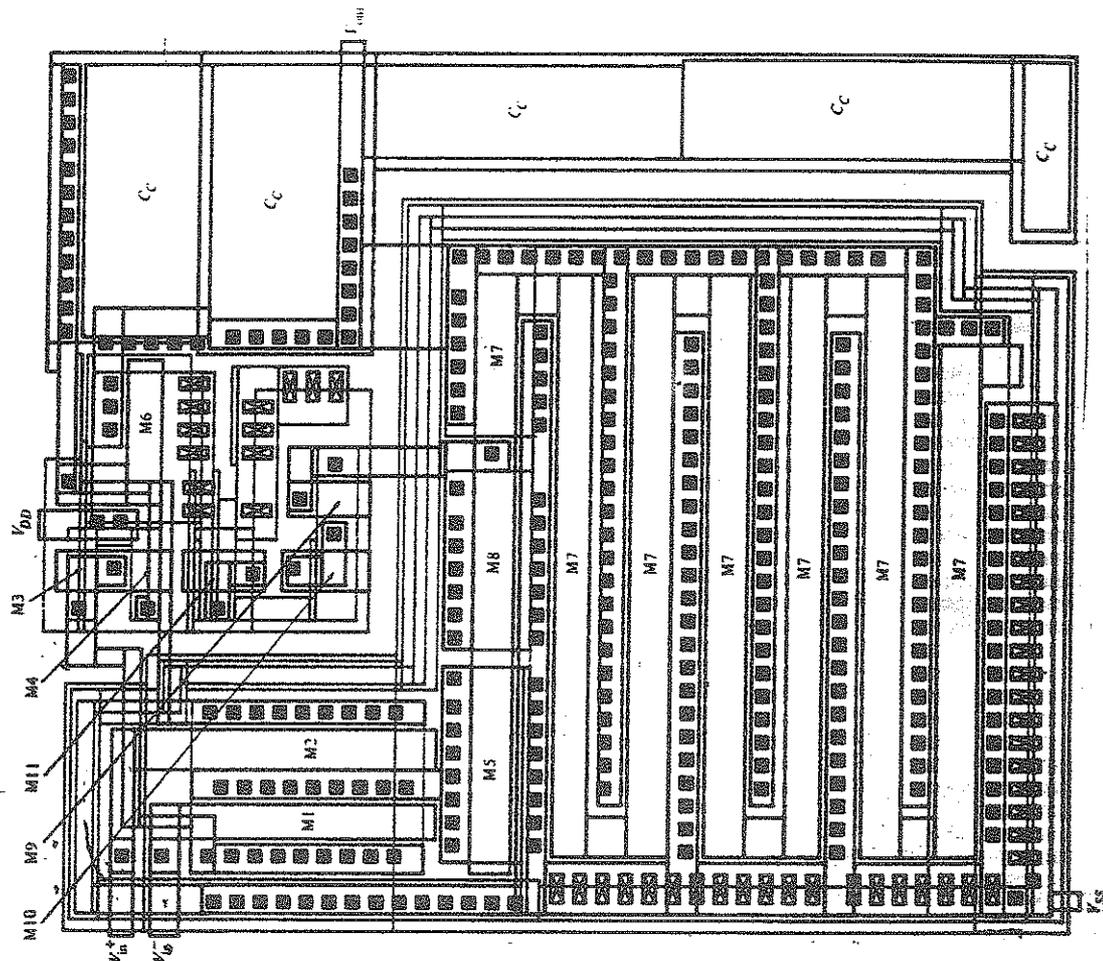
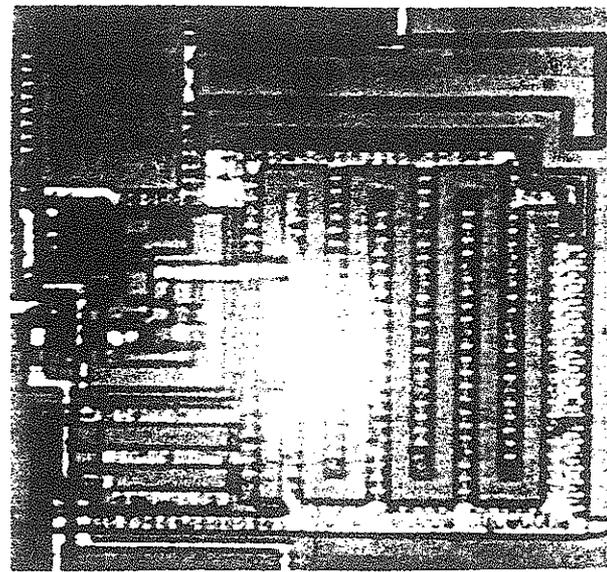
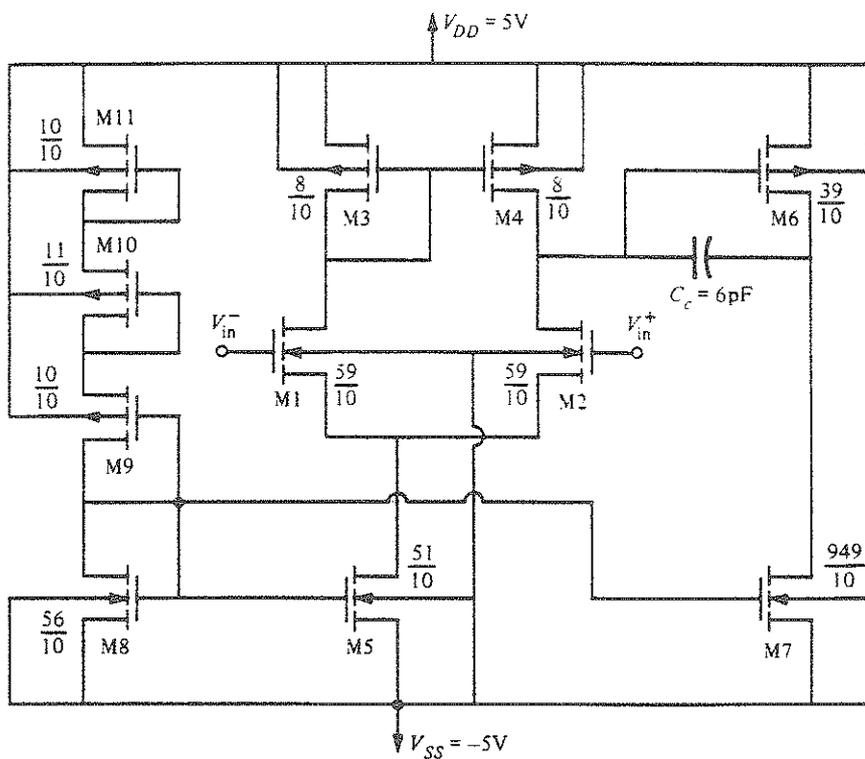


Figura 5.26: Schematico, Layout e Microfotografia di un comune amplificatore operativo in processo CMOS con transistor di ingresso di tipo N. Si noti la superficie, relativamente grande, occupata dalla capacità di compensazione  $C_c$  [39].

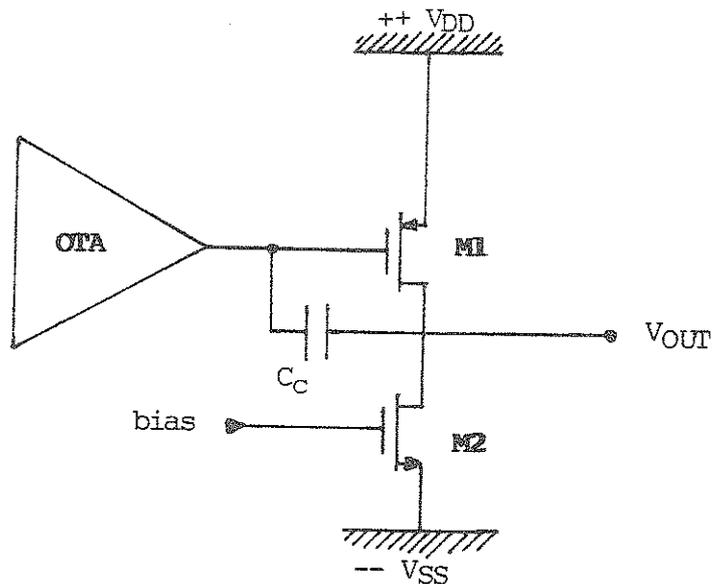


Figura 5.27: Secondo stadio di amplificazione costituito da due transistor in cascata. È visibile la capacità  $C_C$  di compensazione.

• margine di fase:  $23^\circ$ .

Si è dunque spostata a sinistra la curva del guadagno, ma parallelamente si è leggermente spostata anche quella della fase; non si ha quindi nessun incremento del margine di fase.

ii) si è aggiunto all'OTA un secondo stadio, come quello riportato in figura 5.27. Il guadagno di questo nuovo stadio è dato da:

$$A_2 = \frac{g_{m1}}{g_{ds1} + g_{ds2}}$$

dove  $g_{m1}$  è la transconduttanza del transistor M1, mentre  $g_{ds1}$  e  $g_{ds2}$  sono le

conduttanze di uscita di M1 ed M2 rispettivamente [39].

In questo modo la compensazione può essere eseguita con valori molto piccoli della capacità  $C_C$  poiché questa è vista, per il teorema di Miller, come una capacità  $C^*$  di valore:

$$C^* = (1 + A_2) \cdot C_C$$

e dunque  $C_C$  può essere  $A_2$  volte più piccola di quella che si dovrebbe porre in uscita senza il secondo stadio di amplificazione.

Si sono quindi calcolati i valori dei rapporti  $W/L$  ottimali per avere una uscita equilibrata a 0 V. Avendo una alimentazione  $+3V \div -3V$  occorre avere una caduta di tensione  $\text{mod}(V_{DS}) = 3\text{ V}$  per ogni transistor.

In riferimento alla figura 5.27 si sono trovati:

$$W/L(M1) = 14.6,$$

$$W/L(M2) = 33.4.$$

L'analisi dei punti di lavoro dei due transistors è riportata in figura 5.28.

Per verificare la stabilità di questo nuovo amplificatore (con il secondo stadio), si è controreazionato come si può vedere nella figura 5.29 con un cammino di reazione puramente resistivo ( $R = 1\text{ M}\Omega$ ); circuito identico quindi a quello già utilizzato e riportato in figura 5.21, ad eccezione del fatto che ora gli ingressi sono invertiti poiché l'uscita è invertita per la presenza del nuovo stadio (invertente) di amplificazione.

I risultati della simulazione AC sono riportati in figura 5.30. Da questa, per confronto con la figura 5.23, si può concludere che il nuovo amplificatore a due stadi sembra stabile e possiede un margine di fase  $\phi_M \sim 60^\circ$ .

### 5.3 Simulazione di una cella di memoria

Nella figura 5.2 è riportata la cella di memoria e relativi switches per la scrittura, lettura e azzeramento (WRITE, READ and RESET). Si è dunque impacchettato l'OTA riducendolo ad una vista simbolica, e lo si è collegato alla cella di memoria, in modo che la capacità di integrazione venga a trovarsi sul percorso di reazione

M1		M2	
MODEL	MODP	MODEL	MODN
ID	-6.8383E-04	ID	6.83829E-04
VGS	-2.9999E+00	VGS	1.50309E+00
VDS	-2.9325E+00	VDS	3.06752E+00
VBS	0.00000E+00	VBS	0.00000E+00
VTH	-7.6540E-01	VTH	7.04503E-01
VDSAT	-1.4887E+00	VDSAT	5.82531E-01
GM	4.90364E-04	GM	1.60886E-03
GDS	3.64436E-05	GDS	2.75807E-05
GMB	1.00788E-04	GMB	3.73476E-04
GBD	9.96474E-15	GBD	2.72199E-14
GBS	2.08569E-14	GBS	5.04647E-14
GGS	2.80886E-14	GGS	6.96562E-14
GGB	6.51082E-15	GGB	1.49823E-14
GGB	2.66083E-16	GGB	7.67167E-16

Figura 5.28: DC operating point del secondo stadio di amplificazione di figura 5.27.

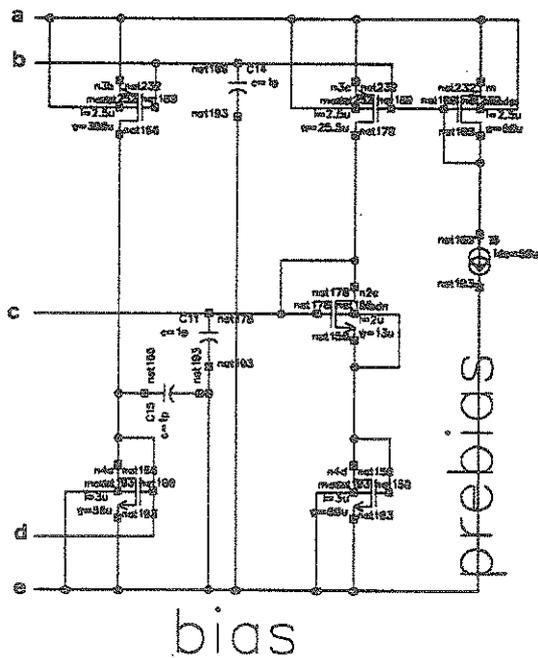
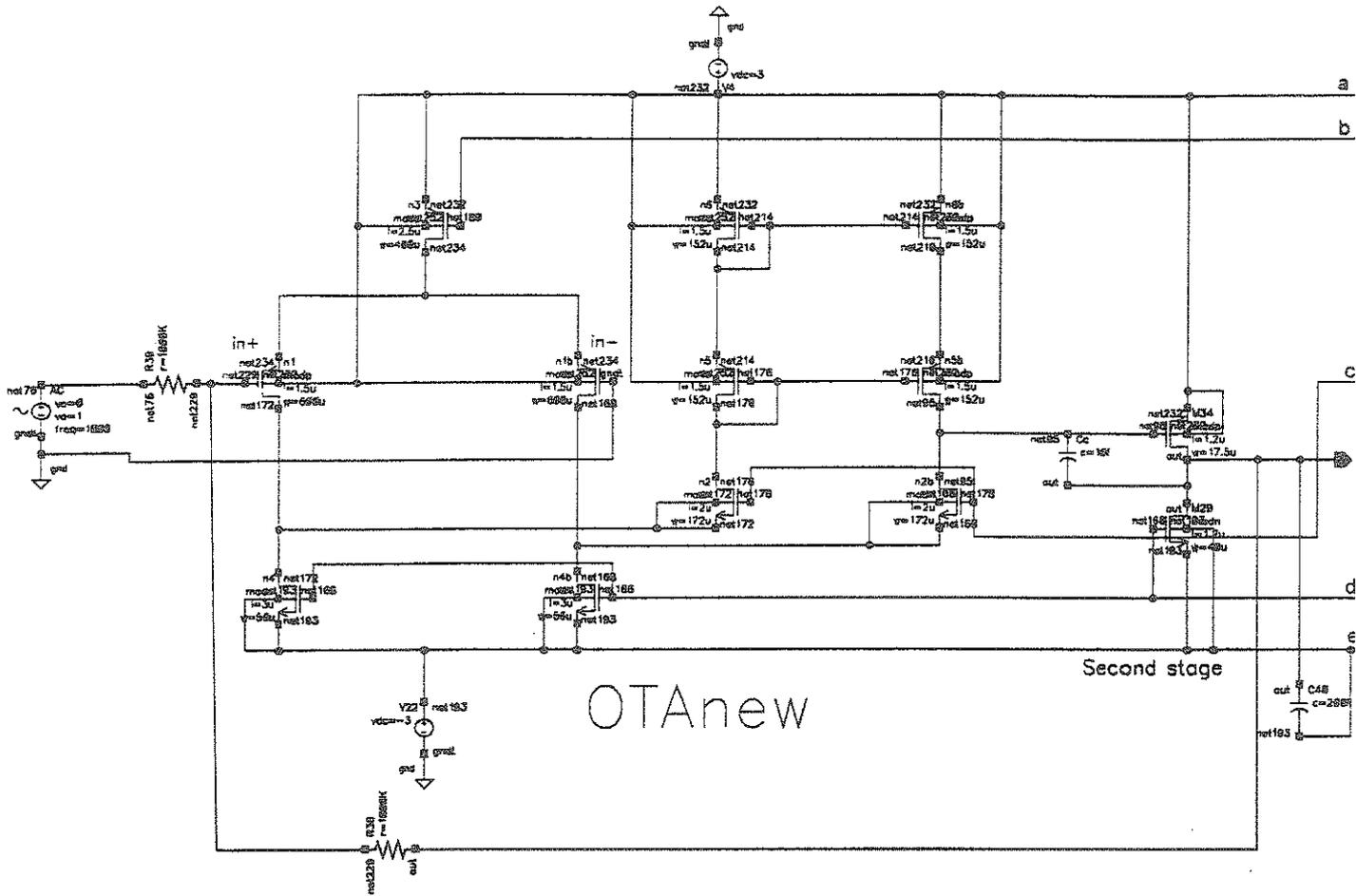


Figura 5.29: Circuito schematico dell'OTA a cui è stato aggiunto il secondo stadio di amplificazione, in configurazione controeazionata per l'analisi della stabilità.

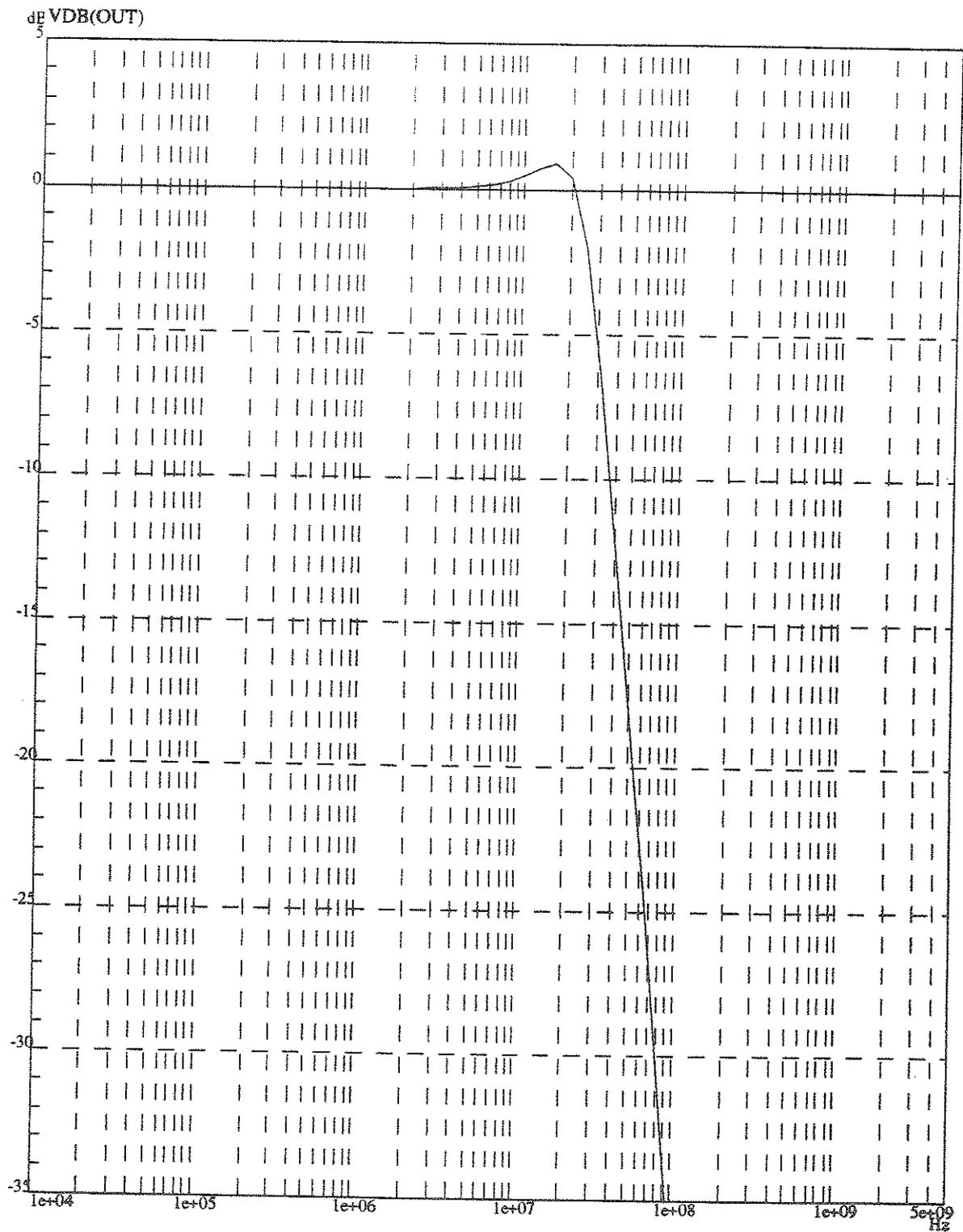


Figura 5.30: Analisi AC: Guadagno normalizzato in funzione della frequenza per il circuito controeazionato dell'OTA a cui è stato aggiunto il secondo stadio di amplificazione, riportato in figura 5.29.

dell'amplificatore (circuito integratore).

Tale configurazione è riportata nella figura 5.31 dove sono inoltre presenti i "clock" per il controllo degli switches ed un generatore di corrente ideale per fornire all'ingresso un segnale da integrare, per simulare il funzionamento della cella.

I substrati (bulks) degli switches sono collegati alla alimentazione più negativa (- 3 V), per limitare al massimo lo scorrere di correnti parassite dai contatti di source e drain al substrato. In questo modo infatti, tra questi viene a crearsi un diodo polarizzato inversamente.

L'OTA è alimentato, come precedentemente detto, alle tensioni + 3 V ÷ - 3 V, e quindi, quando non è collegato ad alcun dispositivo (caso degli switches della cella aperti), le tensioni presenti ai suoi ingressi ed alla uscita possono fluttuare all'interno del range di alimentazione. Dato che l'uscita potrebbe raggiungere circa +3 V prima di chiudere gli switches, e dato che questa è collegata ai source dei transistor di switches, occorre fornire ai gate di quest'ultimi una tensione che permetta loro di superare la soglia:

$$V_G = V_{GS} + V_S > V_{TH} + V_S \simeq 0.75V + 3V \text{ (per dispositivi a canale n).}$$

Per evitare la possibilità che gli switches non conducano, si sono collegati i gate a clock che forniscono una tensione di +6 V nella fase attiva(chiusura degli switches) e 0 V nella fase passiva (apertura degli switches).

In ingresso si è fornito un impulso rettangolare di corrente, come mostrato nel primo grafico di figura 5.32.

Dal tempo  $t = 0s$  al tempo  $t = 20\mu s$  il transistor di reset è chiuso, permettendo alla capacità di azzerare la tensione ai suoi capi, mentre tutti gli altri switches sono aperti, e l'OTA può fluttuare nel range delle tensioni di alimentazione.

Al tempo  $t = 50\mu s$  si chiudono i transistor di scrittura, in attesa che arrivi un segnale in ingresso. Quindi a  $t = 70\mu s$  il generatore ideale di corrente fornisce un impulso rettangolare di corrente, alto  $1\mu A$  e di  $1\mu s$  di durata. Questo segnale viene integrato sulla capacità e la tensione che ci aspettiamo ai suoi capi sarà:

$$V_{OUT} = \frac{Q}{C} \simeq \frac{\Delta i \cdot \Delta t}{C_{cella}}$$

$$\text{e quindi } V_{OUT} = \frac{1\mu s \cdot 1\mu A}{0.4pF} = 2.5 \text{ V.}$$

Al tempo  $t = 150\mu s$  gli switches di scrittura si aprono mentre a  $t = 200\mu s$ , per

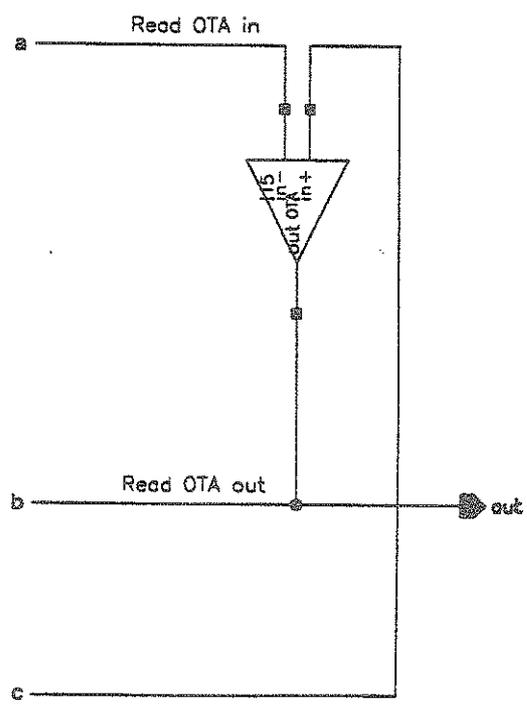
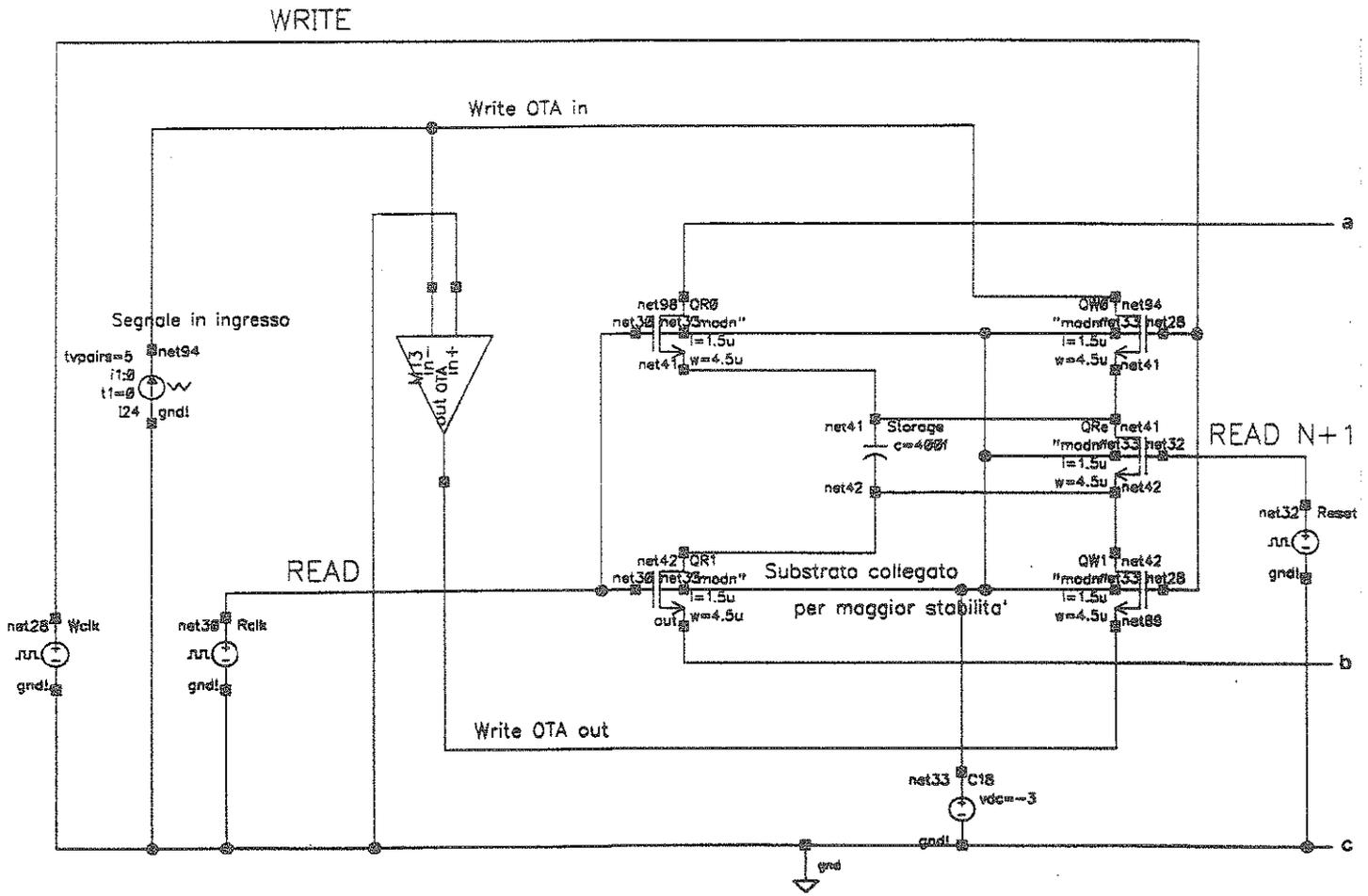


Figura 5.31: Cella di memoria completa: capacit  di integrazione (con gli switches di scrittura, lettura ed azzeramento, ed i relativi clock) collegata in feedback all'OTA.

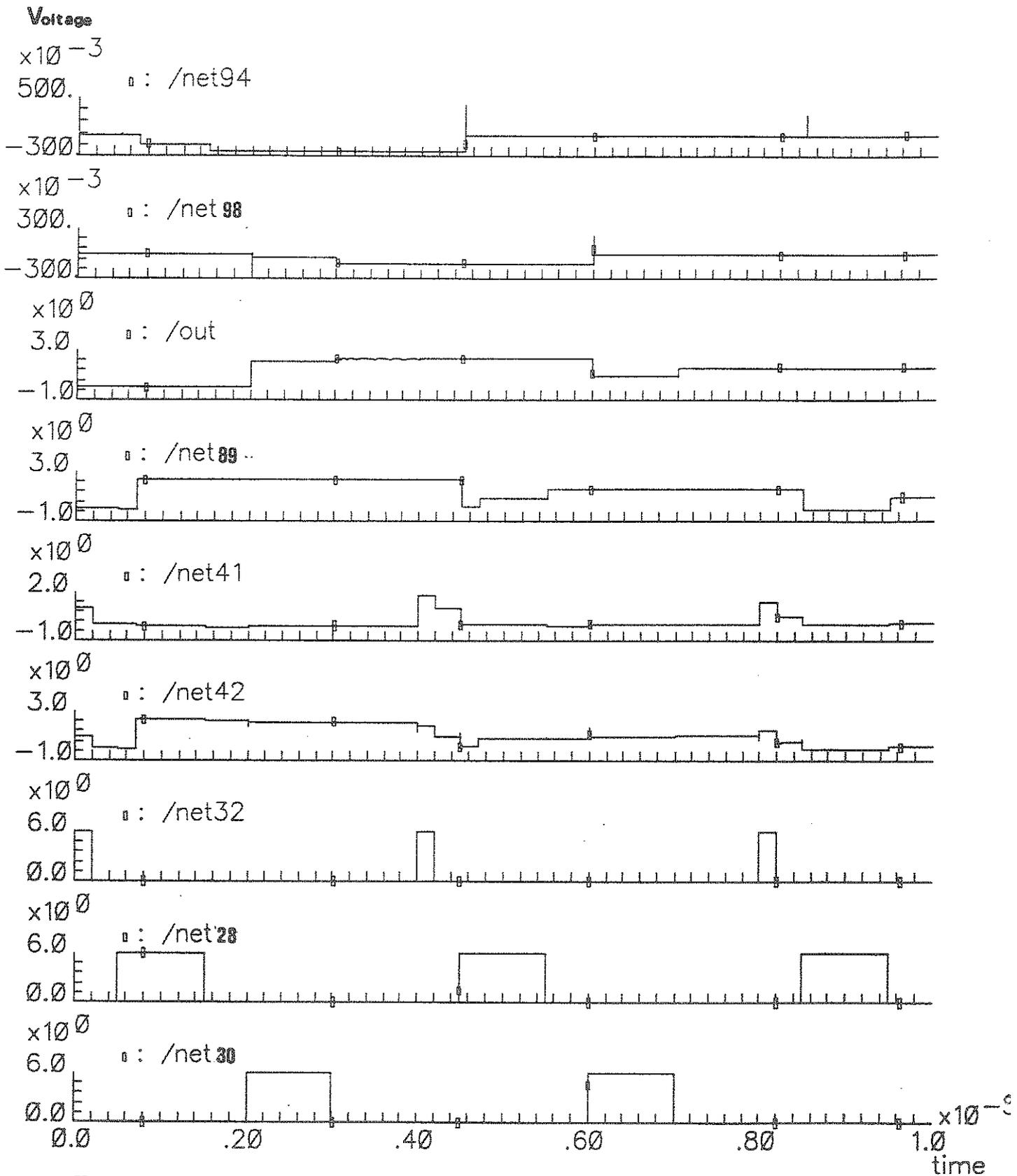


Figura 5.32: Risposta dell'analisi in transiente del circuito schematico di figura 5.31, riportante la cella di memoria completa.

una durata di  $100\mu s$ , si chiudono quelli di lettura, collegando l'OTA di lettura e fornendo effettivamente in uscita i  $2.5\text{ V}$  previsti, come si può leggere nel grafico di  $V_{OUT}$  nella figura 5.32.

Ora si inizia un nuovo ciclo di integrazione e quindi al tempo  $t = 400\mu s$ , per un  $\Delta t = 20\mu s$ , si azzerava la capacità. Al  $t = 450\mu s$  per un  $\Delta t = 100\mu s$  ha luogo un nuovo periodo di scrittura, mentre al tempo  $t = 470\mu s$ , il generatore di corrente fornisce un impulso rettangolare di corrente alto la metà di quello precedente, e cioè di soli  $0.5\mu A$ , per un  $\Delta t = 1\mu s$ . La tensione che ora ci si aspetta in uscita, nel nuovo ciclo di lettura, sarà quindi la metà di quella calcolata nel primo caso, e cioè di  $1.25\text{ V}$ .

In realtà la tensione che si crea ai capi del condensatore (che è uguale alla  $V_{OUT}$ ) nella simulazione, è soltanto il 90% della tensione di uscita e vi è sicuramente una perdita della carica in ingresso dovuta alle capacità parassite degli switches (charge injection). Occorrerà dunque cercare la configurazione migliore affinché la cella di memoria possa avere il minore rumore possibile dovuto alla disposizione ed alimentazione degli interruttori ed amplificatori.

# Bibliografia

- [1] J. F. Fowler. *Nuclear Particles in Cancer Treatment*. Medical Physics Handbook 8. Adam Hilger Ltd. Bristol (1981).
- [2] *Towards Coordination of Cancer Research in Europe*. A.J.M. Vermoken Ed. F.A.T.M. Schermer Co-Ed. Amsterdam-Oxford-Washington-DC-Tokyo. IOS PRESS (1994).
- [3] INFN-Istituto Nazionale di Fisica Nucleare, AIRO-Associazione Italiana Radioterapia Oncologica, CNR-Consiglio Nazionale delle Ricerche, ISS-Istituto Superiore della Sanità', TERA-Fondazione per Adroterapia Oncologica. *The TERA Project and the Centre for Oncological Hadroterapy*, INFN-LNF, U. Amaldi and M. Silari Editors (July 1994).
- [4] J. E. Coggle, The Medical College of St. Bartholomew's Hospital-Londra, *Effetti biologici delle radiazioni*. Edizioni Minerva Medica (1985).
- [5] ICRU REPORT 51. *Quantities and Units in Radiation Protection Dosimetry*, INTERNATIONAL COMMISSION ON RADIATION UNITS AND MEASUREMENTS, 7910 WOODMOONT AVENUE BETHESDA, MARYLAND 20814, U.S.A. (1993).
- [6] W. T. Chu, B. A. Ludewight, and T. R. Renner, Lawrence Berkeley Laboratory, University of California, Berkeley, California 94720, *Instrumentation for treatment of cancer using proton and light-ion beams*. Rev. Sci. Instr. 64. 2055(August 1993).
- [7] James E. Turner. Oak Ridge National Laboratory, *Atoms, Radiation, and Radiation Protection*. PERGAMON PRESS (1986).
- [8] Particle Data Group: M. Aguilar-Benitez et al., *Phys. Rev. D*, 50,(1994).

- [9] ICRU REPORT 30. *Quantitative Concepts and Dosimetry in Radiobiology*, INTERNATIONAL COMMISSION ON RADIATION UNITS AND MEASUREMENTS, 7910 WOODMOONT AVENUE BETHESDA, MARYLAND 20814, U.S.A. (1979).
- [10] Emilio Segrè. *Nuclei e particelle*. Ed. Zanichelli (2<sup>nd</sup> ed. 1982).
- [11] E. Chiavassa. L. Ramello. E. Vercellin. *Rivelatori di particelle*. Appunti dalle lezioni di Fisica dei Neutroni. La Scientifica Editrice - Torino (1991).
- [12] Börje Larsson. Institute for Medical Radiobiology, University of Zurich and the Paul Scherrer Institute. Zurich. Switzerland. *The new radiation in cancer therapy, Hadrontherapy in Oncology*, U. Amaldi and B. Larsson Ed., Elsevier Sciences (1994).
- [13] G. Kraft, W. Becher, K. Blasche, D. Böhne, B. Franczak, Th. Habere, W. Kraft-Weyrather, M. Krämer, B. Langenbeck, G. Lenz, S. Ritter, M. Sholz, D. Schardt, H. Stelzer, P. Strehl and U. Weber, GSI, Darmstadt, Germany, *The Darmstadt Program HITAG: Heavy ion therapy at GSI*, Hadrontherapy in Oncology, U. Amaldi and B. Larsson Ed., Elsevier Sciences (1994).
- [14] Kraft G. *Strahlenther Onkol* 1990; 166:10.
- [15] M. Koehler, R. J. Schneider, and J. M. Sisterson. *Nucl. Instrum. and Methods*, 131, (1975) 437.
- [16] Eleanor A. Blakeley, Life Sciences Division, Lawrence Berkeley Laboratory, Berkeley, California, USA, *Current issue in low- and high-LET medical radiobiology*, Hadrontherapy in Oncology, U. Amaldi and B. Larsson Ed., Elsevier Sciences (1994).
- [17] Kraft G. *Nucl Sci Appl* 1987; 3:1.
- [18] J. Sisterson, *PARTICLES*, Newsletter from the Proton Therapy Cooperative Group, (July, 1994).
- [19] Detlef Gabel<sup>1</sup> and Wolfgang Sauerwein<sup>2</sup>, (1) Department of Chemistry, University of Bremen, Bremen: (2) Strahlenklinik, Universität/Gesamthochschule Essen, Essen, Germany, *Clinical implementation of boron neutron capture therapy in Europe*, Hadrontherapy in Oncology, U. Amaldi and B. Larsson Ed., Elsevier Sciences (1994).

- [20] S. Agosteo<sup>1</sup>, G. Bodei<sup>1</sup>, R. Leone<sup>2</sup> and M. Silari<sup>3</sup>. (1) Dipartimento di Ingegneria Nucleare, Politecnico di Milano, Milano; (2) Fondazione Italiana per la Ricerca sul Cancro, Milano; (3) Consiglio Nazionale delle Ricerche, Istituto Tecnologie Biomediche Avanzate, Milano, Italy, Hadrontherapy in Oncology, U. Amaldi and B. Larsson Ed., Elsevier Sciences (1994).
- [21] TERA. *Comunicato stampa* del giorno 8 settembre 1994. Palazzina di Stupinigi, To.
- [22] Kenneth R. Kase, Bengt E. Bjärngård, Frank H. Attix, *The dosimetry of ionizing radiation*, Volume 1. Academic Press, Inc. (1985).
- [23] ICRU REPORT 31. *Quantitative Concepts and Dosimetry in Radiobiology*, INTERNATIONAL COMMISSION ON RADIATION UNITS AND MEASUREMENTS, 7910 WOODMOONT AVENUE WASHINGTON, D.C. 20014, U.S.A. (april,1979).
- [24] E. Pedroni, *Beam Delivery*, Hadrontherapy in Oncology, U. Amaldi and B. Larsson Ed., Elsevier Sciences (1994).
- [25] R. Bellazzini and M. A. Spezziga, *Electric field, avalanche growth and signal development in Micro-Strip Gas Chamber and Micro-Gap Chamber*, INFN PI/AE - 94/02 (June 24, 1994).
- [26] Efrat Shefer. *Low-pressure operation of microstrip gas chamber*, Thesis for the M.Sc. Degree submitted to the Scientific Council of the Weizmann Institute of Science, Rehovot 76100 - Israel, (11/1994).
- [27] G. Kraft and M. Kramer, *Linear Energy Transfer and Track Structure*, Advances in Radiation Biology, Vol. 17,1-52. (1993).
- [28] Varma, Baum and Kuehner, *Radiat. Res.* 70,511-518. (1977).
- [29] F. Anghinolfi et al., *A 66 MHz, 32 channel analog memory circuit with data selection for fast silicon detectors*. Nucl. Instrum. and Methods. A326 (1993) 100-11.
- [30] R. Horisberger and D. Pitizl. *A novel readout chip for silicon strip detectors with an analog pipeline and digitally controlled analog signal processing*, Nucl. Instrum. and Methods. A326 (1993) 92-99.

- [31] F. Anghinolfi, S. Bouvier and E. Pernigotti, *Test of analog memory pipeline at low frequencies (kHz range)*. INFN-CERN (1995).
- [32] B. Gottschalk. *Charge-balancing current integrator with large dynamic range*, Nucl. Instrum. and Methods. 207 (1983) 417-421.
- [33] ICRU REPORT 49. *Stopping Power and Ranges for Protons and Alpha Particles*, INTERNATIONAL COMMISSION ON RADIATION UNITS AND MEASUREMENTS, 7910 WOODMOONT AVENUE BETHESDA, MARYLAND 20814, U.S.A. (may,1993).
- [34] ICRU REPORT 31. *Average Energy Required to Produce an Ion Pair*, INTERNATIONAL COMMISSION ON RADIATION UNITS AND MEASUREMENTS, 7910 WOODMOONT AVENUE WASHINGTON, D.C. 20014, U.S.A. (may,1979).
- [35] Amaldi, Bizzarri, Pizzella, *Fisica generale - elettromagnetismo - relatività - ottica*, Zanichelli ed. (1994).
- [36] P. Horowitz and W. Hill, *The Art of Electronics*, Cambridge University Press (1980).
- [37] P. Jarron, F. Anghinolfi and E. Heijne, *Fast analog storage and delay line for application in front end electronics system with high rate capability*, CERN.
- [38] R. G. Meyer and P. R. Gray, *Analysis and Design of Integrated Circuit*, John Wiley Sons, Third edition (1993).
- [39] P. E. Allen and D. R. Holberg, *CMOS, Analog Circuits Design*, Saunders College Publishing, (1987).
- [40] J. D. Chatelain et R. Dessoulavy, *Electronique - Tome 1*, Dunod ed.
- [41] D. A. Pucknell and K. Eshraghian, *Basic VLSI Design: Systems and Circuits*, Perentice-Hall, (1988).
- [42] R. L. Geiger, P. E. Allen, N. R. Strader, *VLSI Design Techniques for Analog and Digital Circuits*, McGraw-Hill, (1990).
- [43] J. Millman and C. C. Halkias, *Integrated Electronics: Analog and Digital Circuits and System*, McGraw Hill (1972).

- [44] L. A. Glasser and D. W. Dobberpuhl. *The design and Analysis of VLSI Circuits*. Addison-Wesley (1990).
- [45] W. Sansen. *Integrated Low-Noise Amplifier in CMOS Technology*, Nucl. Instrum. and Methods. A253 (1987) 427-433.

