

Appunti
di
Elettronica Digitale

Massimo Macucci

La seguente raccolta è tratta dalle dispense redatte dal Prof. Massimo Macucci per il corso di Elettronica Digitale (C.d.L. Triennale in Ing. Informatica) dell'Università di Pisa, reperibili in originale sul sito del corso

<http://brahms.i.et.unipi.it/esd/>

Indice

1	Cenni sulla fisica dei semiconduttori e sul principio di funzionamento del diodo a giunzione	
1.1	Materiali semiconduttori	1
1.2	Diffusione e drift	2
1.3	Semiconduttori drogati	5
1.4	Dipendenza della conducibilità dalla temperatura	6
1.5	La giunzione <i>pn</i> a circuito aperto	6
1.6	La giunzione <i>pn</i> in polarizzazione inversa e diretta	7
1.7	La capacità di transizione	9
1.8	La giunzione <i>pn</i> in breakdown	11
1.9	Capacità di diffusione	12
1.10	Caratteristica I – V dei diodi	13
2	Analisi del comportamento circuitale dei diodi	
2.1	Simboli circuitali	13
2.2	Retta di carico	14
2.3	Circuiti equivalenti del diodo per grandi segnali	15
2.4	Metodi di analisi dei circuiti	16
2.5	Circuito rettificatore	17
2.6	Raddrizzatori a doppia semionda	18
2.7	Circuiti tagliatori e fissatori	20
2.8	Logica a diodi	22
2.9	Regolatore di tensione con diodo Zener	23
3	Analisi dei circuiti a diodi per piccoli segnali	
3.1	Modello del diodo per piccoli segnali	24
3.2	Esempio di applicazione del modello per piccoli segnali	26
3.3	Interruttore a diodo	27
4	Il transistor BJT	
4.1	Generatore di corrente controllato in corrente	30
4.2	Il principio di funzionamento del BJT	32
4.3	Le equazioni di Ebers-Moll	34
4.4	Caratteristiche a emettitore comune	36
5	I transistori a effetto di campo	
5.1	Generatore di corrente controllato in tensione	40
5.2	Il transistor MOS	41
5.3	Il transistor JFET	47
6	Reti di polarizzazione	
6.1	Reti di polarizzazione per transistori BJT	51
6.2	Reti di polarizzazione per transistori a effetto di campo	55

7	Funzionamento linearizzato per piccoli segnali	
7.1	Modello linearizzato per transistori BJT	58
7.2	Modello linearizzato per transistori a effetto di campo	63
8	Analisi delle principali configurazioni circuitali a BJT e FET	
8.1	Concetti generali	65
8.2	Stadio amplificatore a emettitore comune senza resistenza di emettitore	65
8.3	Stadio amplificatore a emettitore comune con resistenza di emettitore	67
8.4	Stadio amplificatore a collettore comune	68
8.5	Stadio amplificatore a base comune	70
8.6	Riepilogo configurazioni a BJT	71
8.7	Comportamento dei transistori PNP	72
8.8	Stadio amplificatore a source comune	73
8.9	Stadio amplificatore a source comune con resistenza di source	74
8.10	Stadio amplificatore a drain comune	75
8.11	Riepilogo configurazioni a FET	76
8.12	Amplificatori multistadio	77
8.13*	Teorema di Miller	78
9	Analisi del comportamento in frequenza degli amplificatori	
9.1	Concetti generali	80
9.2	La trasformata di Laplace	80
9.3	Funzioni di rete e funzioni di trasferimento	82
9.4	La funzione di trasferimento nel dominio della frequenza generalizzata	82
9.5	Il metodo della resistenza vista	83
9.6	Calcolo della funzione di trasferimento	85
9.7	Diagrammi di Bode	87
9.8	Esempio di tracciamento di diagrammi di Bode	90
9.9	Funzione di trasferimento degli amplificatori	93
9.10	Poli non interagenti	96
9.11*	Risposta alle basse e medie frequenze dello stadio a emettitore comune	97
9.12	Circuito equivalente del transistore BJT alle alte frequenze	99
9.13*	Analisi dello stadio a emettitore comune alle alte frequenze	101
9.14	Circuito equivalente alle alte frequenze per transistori FET	103
10	La tecnica della reazione	
10.1	Concetti generali	105
10.2	Teoria semplificata della reazione	106
10.3	Effetto della reazione sulle impedenze di ingresso e di uscita	108
11	Circuiti basati su amplificatori operazionali	
11.1	Amplificatori differenziali	110
11.2	Amplificatori operazionali	112
11.3	Metodo del corto circuito virtuale	113
11.4	Amplificatore invertente	114
11.5	Amplificatore non invertente	115
11.6	Amplificatore differenziale	116
11.7	Integratore di Miller	117
11.8	Sommatore	119
12	Regolatori di tensione	
12.1	Regolatore lineare serie	121
12.2	Regolatore monolitici integrati	122
12.3	Regolatori non lineari a commutazione	123
12.4	Alimentatore a commutazione flyback con trasformatore ad alta frequenza e circuito di regolazione	127

13 Circuiti non lineari a operazionali

13.1	Comparatori	130
13.2	Generatore di forme d'onda quadre e rettangolari	133

14 Il programma di simulazione SPICE

14.1	Scopo dei programmi di simulazione	136
14.2	Funzionalità di SPICE	136
14.3	Struttura della netlist	137
14.4	Statement per la descrizione dei componenti passivi	137
14.5	Statement per la descrizione dei generatori di tensione e di corrente	138
14.6	Statement per la descrizione dei componenti a semiconduttore e dei sottocircuiti	140
14.7	Direttive per l'analisi dei circuiti	141
14.8	Direttive di uscita	142

15 Concetti di base sui circuiti digitali

15.1	Introduzione	144
15.2	L'inverter CMOS	145
15.3	Parametri caratteristici dei circuiti digitali	148
15.4	Sintesi delle porte CMOS	153
15.5	Protezione dalle scariche elettrostatiche	160
15.6	Logica basata sui pass transistor	161
15.7	Famiglie logiche bipolari RTL e DTL	166
15.8	Famiglia logica TTL	169
15.9	Famiglia logica TTL Schottky	175
15.10	Confronto tra le varie famiglie logiche	178

16 Logica sequenziale

16.0	Introduzione	179
16.1	Latch	179
16.2	Flip-flop set-reset	180
16.3	Flip-flop tipo D	183

17 Memorie a semiconduttore

17.1	Introduzione	185
17.2	Architettura delle memorie	186
17.3	RAM statiche (SRAM)	187
17.4	RAM dinamiche (DRAM)	188
17.5	Procedure di lettura e scrittura	189
17.6	Decoder e multiplexer per gli indirizzi	192
17.7	Le memorie ROM	194
17.8	Le memorie PROM, EPROM, EEPROM	195

18 Circuiti monostabili, astabili e generazione di segnali di clock

18.1	Introduzione	201
18.2	Multivibratore monostabile	201
18.3	Multivibratore astabile	204
18.4	Oscillatore ad anello	205
18.5	Oscillatori quarzati	206
18.6	Applicazioni del circuito integrato NE555	211

19 Logica programmabile

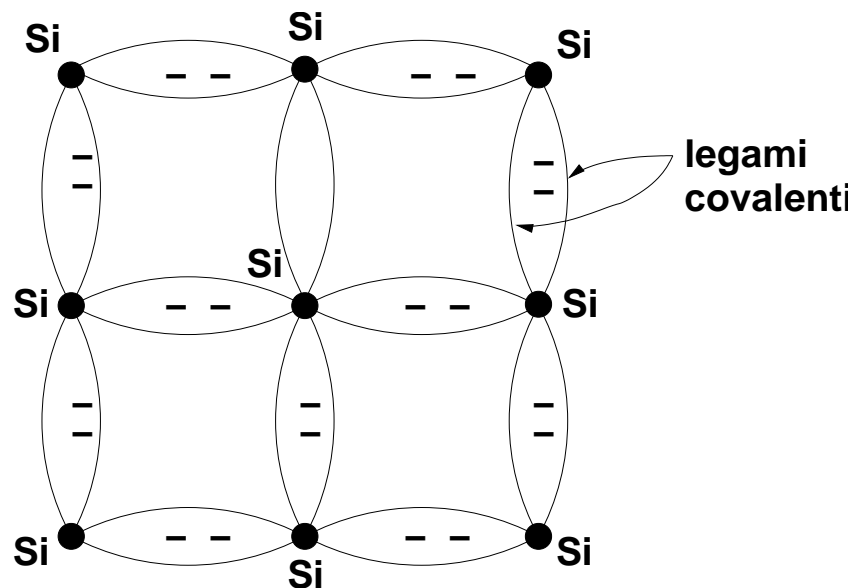
19.1	Introduzione	216
19.2	Programmable Array Logic (PAL)	216
19.3	Programmable Logic Device (PLD)	217
19.4	Application Specific Integrated Circuit (ASIC)	218
19.5	Field Programmable Gate Arrays (FPGA)	219

1. Cenni sulla fisica dei semiconduttori e sul principio di funzionamento del diodo a giunzione

1.1 Materiali semiconduttori

Una distinzione può essere fatta tra i vari materiali sulla base della loro resistività elettrica. I materiali caratterizzati da una conducibilità molto bassa ($\rho > 10^5 \Omega \text{ cm}$) sono classificati come isolanti; quelli che invece presentano una conducibilità molto alta ($\rho < 10^{-2} \Omega \text{ cm}$) sono di solito indicati come conduttori. Si definiscono semiconduttori quei materiali che presentano una conducibilità intermedia ($10^{-2} \Omega \text{ cm} < \rho < 10^5 \Omega \text{ cm}$). I semiconduttori hanno inoltre una serie di caratteristiche fisiche peculiari, che li rendono adatti alla realizzazione di dispositivi elettronici. Anche se vari semiconduttori possono essere utilizzati a questo scopo (tra cui germanio, arseniuro di gallio, fosforo di indio, ecc.), quello di gran lunga più diffuso e sul quale concentreremo la nostra attenzione è il silicio, con il quale sono realizzati la stragrande maggioranza dei dispositivi attualmente sul mercato. Un cristallo di silicio puro (di solito detto intrinseco) ha una struttura ordinata nella quale gli atomi sono collocati in un reticolo e mantenuti in posizione da legami covalenti formati dai quattro elettroni di valenza di ciascun atomo di silicio.

A temperature sufficientemente basse tutti i legami covalenti sono intatti e non sono disponibili elettroni liberi, utilizzabili per la conduzione. La situazione cambia al crescere della temperatura, quando l'energia termica diventa sufficiente a spezzare alcuni di questi legami e a rendere quindi alcuni elettroni liberi di muoversi attraverso il cristallo e di contribuire alla conduzione. Va detto che gli elettroni che si spostano in un cristallo si comportano in modo ben diverso da quello in cui si comporterebbero nel vuoto: essi vedono un potenziale periodico dovuto ai nuclei atomici e agli elettroni a essi legati. Si può dimostrare che il loro comportamento è equivalente a quello di “quasi-particelle” che si spostano come nel vuoto, ma che possiedono una massa diversa da quella nel vuoto, la cosiddetta “massa efficace”. Sulla base di questa considerazione, nel seguito non ci occuperemo della presenza del potenziale periodico e tratteremo gli elettroni come quasi-particelle che si spostano nel vuoto.



Quando un elettrone diventa libero, al suo posto rimane una “lacuna” o “buca”, che equivale in un certo senso (l'esatta trattazione della natura delle lacune richiedereb-

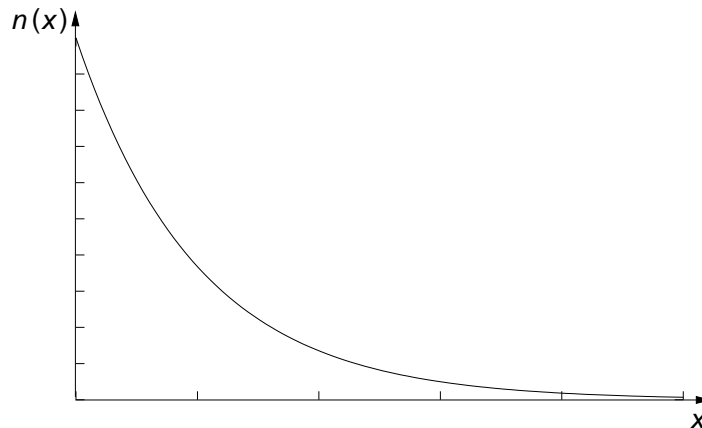
be una discussione piuttosto approfondita di fisica dei semiconduttori) a una carica positiva. Tale lacuna può essere riempita da un elettrone liberatosi da un legame covalente vicino, che, a sua volta, crea una nuova lacuna. In questo modo anche le lacune possono quindi comportarsi come cariche libere e trasportare corrente. Anch'esse possono essere trattate come quasi-particelle con carica opposta a quella di un elettrone e massa efficace in genere un po' diversa da quella degli elettroni.

Questo processo che conduce alla formazione di coppie elettrone-lacuna si definisce ionizzazione termica. Così come le coppie vengono generate, possono anche scomparire se uno degli elettroni liberi va a riempire una lacuna: si parla in tal caso di ricombinazione. All'equilibrio termico il tasso di ionizzazione è pari a quello di ricombinazione, per cui le concentrazioni di elettroni (n) e di lacune (p) sono costanti e pari a un valore n_i che dipende dalle caratteristiche del materiale e dalla temperatura. Nel silicio intrinseco (cioè privo di impurezze aggiunte) a temperatura ambiente ($T \approx 300$ K) $n_i = 1.5 \times 10^{10} \text{ cm}^{-3}$. Se consideriamo che il numero degli atomi di silicio per centimetro cubo è 5×10^{22} , ci rendiamo conto che a temperatura ambiente solo un atomo su circa 3×10^{12} risulta ionizzato. A questo punto si comprende il motivo del comportamento elettrico dei semiconduttori, che conducono meglio degli isolanti (nei quali il numero dei portatori liberi è ridottissimo) e peggio dei conduttori (nei quali il numero dei portatori liberi è comparabile con quello degli atomi).

1.2 Diffusione e drift

Esistono due meccanismi che consentono il trasporto di portatori di carica attraverso un semiconduttore: la diffusione e il "drift". La diffusione è associata con il moto casuale dei portatori dovuto all'agitazione termica. In un cristallo nel quale la concentrazione di elettroni e di lacune è uniforme, il moto termico non porta ad alcun flusso netto di carica. Se, invece, tramite qualche meccanismo, la concentrazione di un tipo di portatori viene resa maggiore in una regione del cristallo rispetto al resto, i portatori migreranno dalla zona a maggior concentrazione verso quella a minore concentrazione, nel tentativo di ristabilire l'equilibrio. Si verrà quindi a creare una corrente netta, che definiamo "corrente di diffusione".

Consideriamo una barra di silicio nella quale un profilo di concentrazione per gli elettroni come quello sotto rappresentato venga creato con una qualche procedura.



Nasce quindi una corrente di diffusione in direzione x , il cui valore risulta proporzionale in ogni punto alla derivata della concentrazione n :

$$J_n = qD_n \frac{dn}{dx},$$

dove J_n è la densità di corrente (corrente per unità di area) in A/cm², q è il modulo della carica dell'elettrone (1.6×10^{-19} C) e D_n è una costante chiamata costante di diffusione degli elettroni. Si noti che nell'esempio presentato dn/dx è negativa e che, come è logico aspettarsi, un flusso di elettroni in direzione positiva dà luogo a una corrente negativa. Una analoga relazione sussiste nel caso sia presente una concentrazione non uniforme di lacune:

$$J_p = -qD_p \frac{dp}{dx}.$$

In questo caso, un flusso di cariche positive in direzione positiva dà luogo a una corrente di segno positivo e compare quindi un segno negativo per compensare quello della derivata della concentrazione. Nel silicio intrinseco valori tipici sono $D_p = 12$ cm²/s e $D_n = 34$ cm²/s.

L'altro meccanismo responsabile per il trasporto di carica nei semiconduttori è il "drift" o "trascinamento". Esso agisce quando è presente un campo elettrico che accelera le cariche libere disponibili all'interno di un semiconduttore. In assenza di meccanismi dissipativi, tali cariche verrebbero accelerate indefinitamente finché si trovano all'interno della zona nella quale il campo elettrico è non nullo. I portatori subiscono però degli urti anelastici dovuti alle fluttuazioni termiche del reticolo cristallino e ad altre cause, per cui raggiungono in modo pressoché immediato una velocità di drift costante, proporzionale al campo elettrico applicato:

$$v_{\text{drift}} = \mu_n E$$

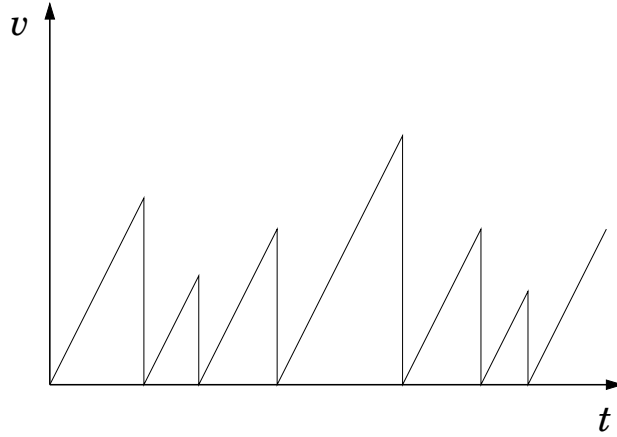
per gli elettroni e

$$v_{\text{drift}} = \mu_p E$$

per le lacune, dove μ_n è una costante definita mobilità degli elettroni ed è espressa in cm²/(V s), μ_p è l'analoga costante per le lacune, v_{drift} è la velocità in cm/s ed E è il campo elettrico in V/cm. Nel silicio intrinseco $\mu_n \approx 1350$ cm²/(V s) e $\mu_p \approx 480$ cm²/(V s). Per capire meglio il concetto di mobilità possiamo tracciare un grafico della velocità di un portatore in funzione del tempo, in presenza di un campo elettrico: sotto l'azione del campo elettrico il portatore tende ad accelerare, con un incremento lineare della velocità. Si verificano tuttavia delle collisioni (scattering), soprattutto con le imperfezioni del reticolo cristallino dovute all'agitazione termica. Tali collisioni sono di tipo anelastico e conducono a una perdita quasi completa dell'energia che il portatore aveva acquisito in conseguenza dell'accelerazione da parte del campo elettrico. Immediatamente dopo ogni collisione si può assumere che l'elettrone abbia una velocità "termica" (dovuta cioè all'energia termica del reticolo cristallino) orientata in una direzione casuale. L'andamento della velocità dell'elettrone nel tempo, depurata della componente termica, che è isotropa e quindi con valor medio nullo, risulta pertanto a dente di sega, come rappresentato nella figura seguente, e il suo valor medio sarà proporzionale al valore del campo elettrico applicato e indipendente dalla lunghezza del conduttore (purché questa sia sufficientemente grande in confronto alla distanza media percorsa tra due eventi di scattering).

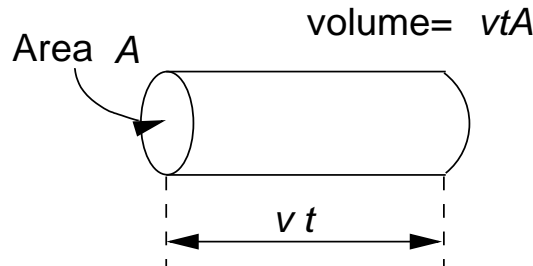
Questo modello, molto semplificato, ma in grado di riprodurre il comportamento ohmico dei conduttori, si definisce "modello di Drude", dal nome del suo ideatore.

Consideriamo ora un cristallo di silicio nel quale siano presenti sia una concentrazione di lacune p sia una concentrazione di elettroni n . In presenza di un campo



elettrico E le lacune si muoveranno (avendo carica positiva) nella stessa direzione di E con una velocità $\mu_p E$. Quindi una densità di carica positiva qp si muove nella direzione del campo con velocità $\mu_p E$. Ne consegue che la sezione di area A sarà attraversata in un secondo da una quantità di carica pari al prodotto del volume V che attraversa tale sezione moltiplicato per la concentrazione di portatori e per la carica di ciascun portatore. Il volume in questione è pari al prodotto dell'area A per la distanza percorsa in un secondo $L = vt$; poiché $t = 1$ s, $L = v = \mu_p E$. Dunque la carica che attraversa la sezione A nell'unità di tempo risulta, come illustrato in figura:

$$Q = qp\mu_p EA.$$



Il flusso di carica attraverso una sezione nell'unità di tempo è proprio la corrente di drift che intendiamo determinare ($I = dQ/dt$, per definizione); possiamo inoltre dividere per l'area A , ottenendo la densità di corrente di drift

$$J_{p\text{-drift}} = qp\mu_p E.$$

Gli elettroni subiranno invece una deriva in direzione opposta a quella di E , dando luogo a una densità di carica $-qn$ che si muove in direzione negativa e quindi con una velocità negativa ($-\mu_n E$). Il risultato è una corrente positiva con densità

$$J_{n\text{-drift}} = qn\mu_n E.$$

La densità della corrente totale di drift risulta pertanto

$$J_{\text{drift}} = q(p\mu_p + n\mu_n)E.$$

Si noti che questa è la rappresentazione puntuale della legge di Ohm $J = \sigma E$, con σ (conducibilità), dato da $\sigma = q(p\mu_p + n\mu_n)$.

1.3 Semiconduttori drogati

Le proprietà del silicio intrinseco finora viste, in particolare quella di avere un uguale numero di lacune e di elettroni liberi, possono essere modificate tramite l'aggiunta, in piccole quantità, di altre specie chimiche definite "droganti". In questo modo si può ottenere il silicio di tipo p (nel quale prevalgono le lacune) o quello di tipo n (nel quale prevalgono gli elettroni). Introducendo atomi droganti di un elemento pentavalente, come il fosforo, si ottiene silicio di tipo n , perché ogni volta che una di tali impurezze sostituisce un atomo di silicio nel reticolo cristallino va a formare quattro legami covalenti con altrettanti atomi di silicio, mentre il quinto elettrone di valenza (il fosforo è un elemento del gruppo V) diviene libero e quindi disponibile per la conduzione. Pertanto il fosforo dona un elettrone libero al cristallo di silicio e viene dunque definito un "donatore". In questo processo non vengono però generate lacune, quindi gli elettroni diventano i portatori maggioritari. Se la concentrazione di atomi donatori è N_D , la concentrazione di elettroni liberi n_{n0} all'equilibrio termico sarà data da

$$n_{n0} \approx N_D,$$

dove il primo pedice (n) indica che stiamo considerando silicio di tipo n e il secondo (0) indica che consideriamo la condizione di equilibrio termico.

Dalla fisica dei semiconduttori si ottiene l'importante risultato che all'equilibrio termico il prodotto delle concentrazioni delle lacune e degli elettroni è costante e pari al quadrato di n_i (la concentrazione intrinseca), indipendentemente dalla concentrazione di droganti:

$$n_{n0}p_{n0} = n_i^2.$$

Questa è definita "legge dell'azione di massa". Quindi la concentrazione di lacune nel silicio n è data da

$$p_{n0} \approx \frac{n_i^2}{N_D}.$$

Poiché n_i è funzione della temperatura, mentre N_D non lo è, la concentrazione dei portatori minoritari dipenderà fortemente dalla temperatura, mentre quella dei maggioritari ne sarà sostanzialmente indipendente.

Per ottenere del silicio di tipo p sarà necessario utilizzare come droganti delle impurezze trivalenti (appartenenti al gruppo III) come il boro. Ciascun atomo di boro dispone di soli tre elettroni di valenza, per cui potrà formare soltanto tre legami covalenti. Nel quarto legame sarà quindi presente una lacuna. Poiché le lacune così formate sono in grado di "accettare" elettroni, i droganti di questo tipo sono detti "accettori".

Ogni atomo accettore dà luogo a una lacuna, per cui il numero delle lacune all'equilibrio termico è approssimativamente uguale a quello degli accettori: $p_{p0} = N_A$. Analogamente a quanto ricavato in precedenza per la concentrazione di lacune nel silicio di tipo n , la concentrazione di elettroni nel silicio p sarà data da

$$n_{p0} \approx \frac{n_i^2}{N_A}.$$

È importante sottolineare il fatto che un pezzo di silicio p o n è elettricamente neutro, perché la carica dei portatori maggioritari è compensata da quella fissa associata agli atomi droganti. Più precisamente, se consideriamo per esempio un pezzo di silicio n , la carica negativa formata dagli elettroni liberi sarà compensata esattamente da

quella positiva costituita dai donatori ionizzati e dalle lacune generate termicamente (si noti che in un semiconduttore di tipo n le lacune sono in numero minore che in un semiconduttore intrinseco, in conseguenza della legge dell'azione di massa, e che ciascun elettrone deriva o dall'ionizzazione di un atomo donatore o dalla formazione di una coppia elettrone-lacuna).

1.4 Dipendenza della conducibilità dalla temperatura

Abbiamo visto che la conducibilità in un semiconduttore è data dall'espressione

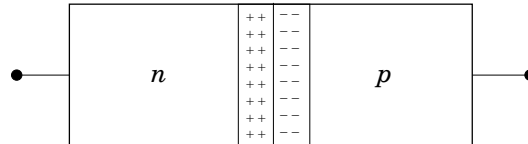
$$\sigma = q(\mu_p p + \mu_n n).$$

La concentrazione di lacune e quella di elettroni aumentano nel silicio intrinseco con la temperatura, ma le mobilità hanno una dipendenza opposta dalla temperatura, almeno per valori della stessa interessanti per le applicazioni elettroniche. L'incremento del numero di portatori prevale peraltro di gran lunga sulla diminuzione di mobilità e quindi la conducibilità aumenta significativamente all'aumentare della temperatura.

Nel silicio drogato, invece, la corrente è prevalentemente trasportata dai portatori maggioritari, la cui concentrazione è sostanzialmente pari a quelle della specie drogante e quindi indipendente dalla temperatura (almeno per la gamma di temperature di normale interesse, all'interno della quale tutti i droganti possono essere assunti ionizzati). In questo caso, quindi, la conducibilità diminuisce all'aumentare della temperatura, a causa del decrescere della mobilità.

Riassumendo, la conducibilità del silicio intrinseco aumenta al crescere della temperatura, mentre quella del silicio drogato diminuisce: queste diverse dipendenze hanno effetti rilevanti sulle proprietà termiche dei dispositivi elettronici.

1.5 La giunzione pn a circuito aperto



Poiché la concentrazione di elettroni dal lato n è molto più grande di quella dal lato p , si avrà una diffusione di elettroni attraverso la giunzione dalla regione n a quella p . In modo simile, le lacune diffonderanno dalla regione p a quella n . Queste due componenti danno luogo a una corrente di diffusione I_D che va dalla zona p alla zona n .

Le lacune che diffondono dalla regione p alla n si ricombinano rapidamente con gli elettroni maggioritari e pertanto scompaiono. Si ha quindi una diminuzione del numero di elettroni liberi nella zona n , quindi alcune delle cariche fisse positive (rappresentate dagli atomi donatori) non saranno più neutralizzate. Poiché la ricombinazione avviene in prossimità della giunzione, esisterà, accanto alla giunzione stessa, una regione svuotata di elettroni e contenente cariche fisse positive non compensate. Nella regione p , accanto alla giunzione, si avrà analogamente una zona svuotata di lacune e contenente cariche fisse negative non compensate. L'insieme di tali zone svuotate di portatori si definisce "zona di svuotamento" o "zona di carica spaziale". In conseguenza delle cariche fisse della zona di svuotamento si viene a stabilire un campo elettrico attraverso la zona di svuotamento stessa, con il lato n positivo rispetto al lato p . Tale campo si oppone a un'ulteriore diffusione di portatori tra la

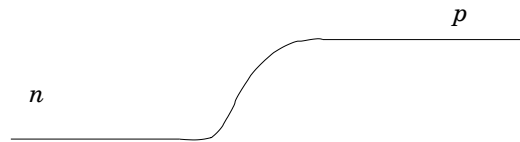
regione p e quella n . La caduta di tensione sulla zona di svuotamento agisce come una barriera di potenziale che deve essere superata dagli elettroni per diffondere nella zona p e dalle lacune per raggiungere la regione n . Questa può essere vista come una condizione di equilibrio dinamico, nella quale la corrente di diffusione I_D è compensata da una corrente uguale e opposta di drift I_S , dovuta al campo elettrico creato dallo squilibrio di carica.

In condizioni di circuito aperto la caduta di potenziale attraverso la giunzione, detta potenziale di contatto, può essere espressa in funzione delle concentrazioni di droganti nella zona p (N_A) e n (N_D), della concentrazione intrinseca di portatori (n_i) e della temperatura T :

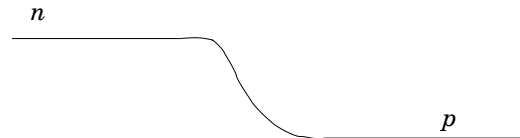
$$V_0 = \frac{kT}{q} \ln \left(\frac{N_A N_D}{n_i^2} \right),$$

dove k è la costante di Boltzmann, pari a 1.38066×10^{-23} J/K. Nel silicio, a temperatura ambiente, V_0 è circa 0.6-0.8 V. La tensione misurata tra i terminali è comunque nulla perché V_0 risulta compensata dai potenziali di contatto metallo-semiconduttore tra le zone p ed n e gli elettrodi esterni. Se così non fosse, sarebbe possibile far passare corrente nel circuito esterno e riusciremmo quindi a trarre energia da una giunzione pn in equilibrio termico con il circuito esterno, violando il secondo principio della termodinamica.

Rappresentiamo ora l'andamento del potenziale visto dagli elettroni e dalle lacune passando dalla regione n a quella p . Poiché esiste una tensione V_0 positiva dal lato n e negativa da quello p , gli elettroni vedranno una barriera in salita andando dal lato n a quello p . Per le lacune la situazione sarà opposta, dato che il segno della loro carica è rovesciato rispetto a quello degli elettroni, quindi per le lacune la barriera sarà in salita andando dal lato p a quello n .



Andamento del potenziale per gli elettroni

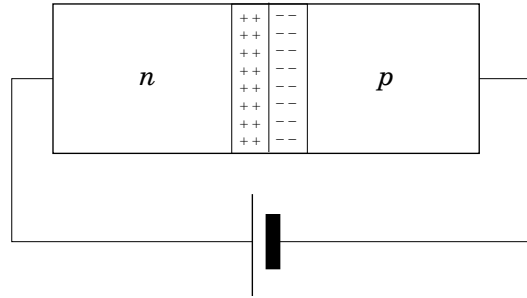


Andamento del potenziale per le lacune

1.6 La giunzione $p - n$ in polarizzazione inversa e diretta

Consideriamo dapprima il caso in cui alla giunzione sia connesso un generatore di tensione continua con il terminale positivo collegato alla regione n , come in figura.

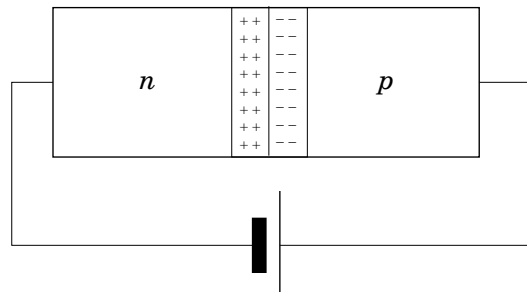
Si parla in questo caso di polarizzazione inversa: la tensione fornita dal generatore esterno tenderebbe a far passare una corrente dalla regione n a quella p , quindi a far fluire lacune dalla regione n a quella p e elettroni dalla regione p a quella n . Nonostante che l'andamento del profilo di potenziale sia favorevole a questi spostamenti, dobbiamo ricordare che nella regione n sono disponibili pochissime lacune e lo stesso



Polarizzazione inversa

discorso vale per gli elettroni nella zona p . La corrente che scorre è quindi estremamente ridotta, perché alimentata dal flusso di portatori minoritari. Tale corrente non è inoltre significativamente dipendente dalla tensione applicata, poiché tutti i portatori minoritari che si trovano in prossimità della zona di svuotamento vengono comunque trascinati via dal campo elettrico favorevole e il loro numero dipende soltanto dalla temperatura. Ne concludiamo che in polarizzazione inversa la corrente è piccola, sostanzialmente indipendente dal valore della tensione applicata e crescente al crescere della temperatura.

Nella condizione di polarizzazione diretta il terminale positivo del generatore esterno è connesso alla regione p : la tensione applicata tende quindi a far circolare una corrente dalla zona p a quella n . Si tratta dunque di una corrente di lacune che va dalla zona p a quella n e di una corrente di elettroni che fluisce dalla zona n a quella p . Si hanno quindi flussi di portatori maggioritari, che sono disponibili in grande quantità, ma questa volta i flussi avvengono in contrasto con l'andamento della barriera di potenziale.



Polarizzazione diretta

Se la tensione applicata è molto piccola, la corrente sarà anche in questo caso trascurabile, perché i portatori non riescono a superare l'ostacolo rappresentato dalla barriera stessa. Poiché l'energia termica posseduta dai portatori è distribuita secondo un andamento di tipo all'incirca Maxwelliano (in realtà l'esatta distribuzione sarebbe quella di Fermi-Dirac, ma questa è simile alla distribuzione di Maxwell per gli aspetti che ci interessano) e quindi ha un andamento di tipo esponenziale (la probabilità $P(E)$ che uno stato di energia E sia occupato risulta data da $P(E) = \exp(-E/kT)$), riesce abbastanza semplice comprendere che al crescere della tensione applicata un numero esponenzialmente crescente di portatori riuscirà a superare la barriera e che la corrente aumenterà in modo esponenziale con la tensione applicata.

Una completa trattazione matematica della conduzione attraverso una giunzione pn esula dagli scopi delle presenti note, ma le considerazioni fenomenologiche finora fornite possono costituire una giustificazione intuitiva della cosiddetta equazione di

Shockley, che dà la corrente in una giunzione pn in funzione della tensione applicata:

$$I = I_S \left(e^{V/V_T} - 1 \right),$$

dove I è la corrente che attraversa il diodo, I_S è la corrente di saturazione inversa, cioè quella che scorre in polarizzazione inversa, V è la tensione applicata tra la regione p e quella n , $V_T = kT/q$ con un valore a temperatura ambiente (300 K) di circa 26 mV. Nel caso di polarizzazione inversa molto maggiore in modulo di V_T il termine esponenziale è trascurabile rispetto all'unità e quindi $I \approx -I_S$, indipendentemente dalla tensione applicata, come prima discusso. Quando ci troviamo in condizione di polarizzazione diretta, invece, non appena V è in modulo significativamente più grande di V_T , il termine esponenziale domina sull'unità e la corrente dipende esponenzialmente dalla tensione applicata.

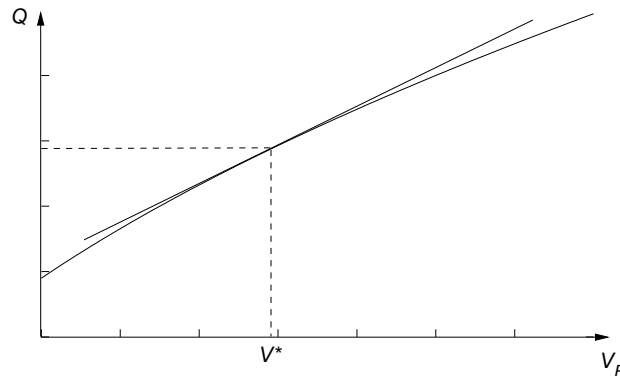
La caratteristica $I - V$ dei diodi al germanio corrisponde in modo praticamente esatto alla legge di Shockley, mentre per i diodi al silicio la caratteristica effettivamente misurata risulta del tipo

$$I = I_S \left(e^{V/(\eta V_T)} - 1 \right),$$

dove η è un fattore di idealità (legato a fenomeni di generazione e ricombinazione) compreso tra 1 e 2.

1.7 La capacità di transizione

Quando si applica una tensione inversa alla giunzione, vengono iniettati, in corrispondenza dei contatti, elettroni nella zona p e lacune nella zona n che vanno a ricombinarsi con i portatori maggioritari nelle corrispondenti zone, portando a una diminuzione degli stessi e quindi a un allargamento della zona di svuotamento, finché non si viene a stabilire una nuova condizione di equilibrio. A regime l'ampiezza della zona di svuotamento risulta tanto più grande (e quindi contenente una carica spaziale maggiore) quanto maggiore è, in modulo, la tensione inversa applicata. Risulta dunque chiara l'analogia tra la zona di svuotamento di una giunzione pn e un condensatore: a un aumento della tensione applicata corrisponde un aumento della carica immagazzinata nella zona di svuotamento. A differenza di quanto accade in un normale condensatore, la dipendenza della carica dalla tensione applicata è non lineare, come illustrato in figura.



È possibile peraltro definire una capacità differenziale, intesa come derivata della carica rispetto alla tensione, intorno a un particolare valore della tensione stessa:

$$C_{\text{diff}} = \left. \frac{dQ}{dV_R} \right|_{V_R=V^*}.$$

Tale capacità differenziale, detta capacità di transizione, potrebbe essere calcolata ricavando prima l'espressione della carica in funzione di V_R e poi derivandola. Più semplicemente, è possibile calcolare la capacità di transizione approssimando la zona di svuotamento come un condensatore a facce piane e parallele.

$$C_{\text{diff}} = \frac{\varepsilon_s A}{W_{\text{dep}}},$$

dove ε_s è la costante dielettrica del silicio, A l'area della sezione trasversale e W_{dep} l'ampiezza della zona di svuotamento. L'ampiezza della zona di svuotamento è proporzionale alla radice quadrata della somma del potenziale di contatto V_0 e del modulo V_R della tensione inversa applicata,

$$W_{\text{dep}} \propto \sqrt{V_0 + V_R},$$

quindi possiamo scrivere

$$\frac{C_{\text{diff}}}{C_{j0}} = \frac{W_{\text{dep}0}}{W_{\text{dep}}},$$

dove con C_{j0} e $W_{\text{dep}0}$ si sono indicate, rispettivamente, la capacità differenziale e l'ampiezza della zona di svuotamento per tensione applicata nulla. Sulla base della dipendenza di W_{dep} da V_R , possiamo ricavare l'espressione di C_{diff} da C_{j0} :

$$C_{\text{diff}} = \frac{C_{j0}}{\sqrt{1 + \frac{V_R}{V_0}}}$$

dove C_{j0} è la capacità differenziale per tensione applicata nulla. Il valore di C_{j0} si può ricavare conoscendo l'espressione completa di W_{dep} :

$$W_{\text{dep}} = \sqrt{\frac{2\varepsilon_s}{q} \left(\frac{N_A + N_D}{N_A N_D} \right) (V_0 + V_R)},$$

quindi, inserendo il valore di W_{dep} per $V_R = 0$, abbiamo:

$$\begin{aligned} C_{j0} &= \frac{\varepsilon_s A}{\sqrt{\frac{2\varepsilon_s}{q} \left(\frac{N_A + N_D}{N_A N_D} \right) V_0}} \\ &= A \sqrt{\frac{q\varepsilon_s}{2} \left(\frac{N_A N_D}{N_A + N_D} \right) \frac{1}{V_0}}. \end{aligned}$$

È evidente che la capacità di transizione diminuisce dunque al crescere della tensione applicata (le estremità della zona di svuotamento si allontanano). Tale proprietà può essere utilizzata per ottenere delle capacità controllabili con una tensione, che

sono indispensabili per la realizzazione di molti circuiti di comune utilizzo, come gli oscillatori a frequenza variabile utilizzati nei sintetizzatori di frequenza.

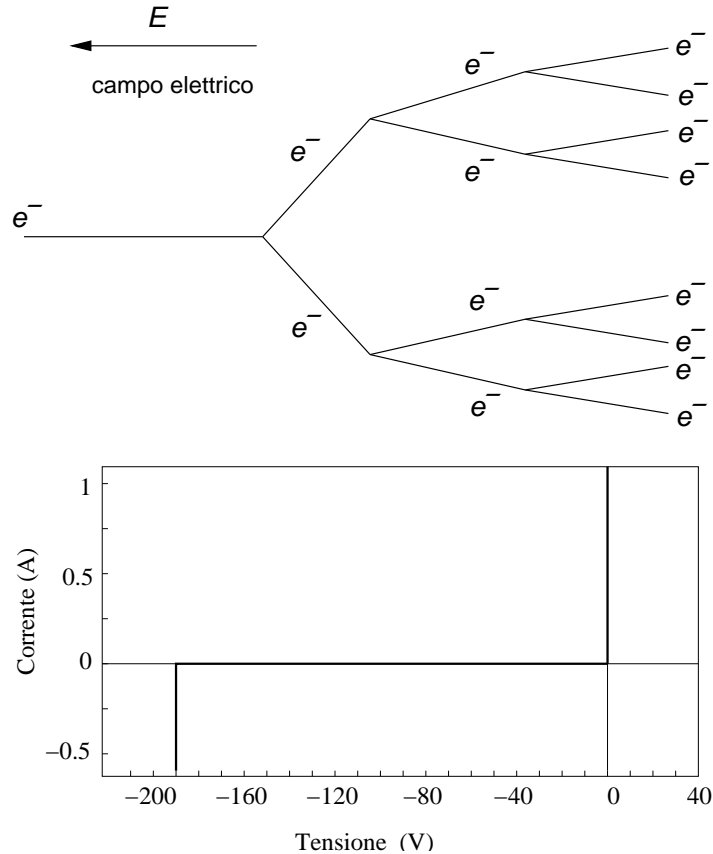
1.8 La giunzione pn in breakdown

Se si applica una tensione inversa sufficientemente grande, entrano in gioco nuovi meccanismi in grado di far transitare una corrente molto più grande di I_S attraverso la giunzione: si parla in tal caso di “breakdown”. Esistono due possibili tipi di breakdown: quello Zener e quello a valanga. La maggior parte dei diodi normalmente utilizzati ha tensioni di breakdown elevate (dell’ordine delle centinaia o addirittura migliaia di volt), in maniera da evitare che il breakdown si verifichi nell’impiego regolare (altrimenti il diodo perderebbe la sua funzione principale, che è quella di far passare la corrente in una sola direzione). Il tipo di breakdown che si verifica in tali diodi è quello a valanga; solo nel caso di applicazioni particolari, come la realizzazione di riferimenti di tensione, si utilizzano diodi la cui tensione di breakdown è stata ridotta, con opportuni accorgimenti costruttivi, a valori tra qualche volt e alcune decine di volt.

Diodi con tensioni di breakdown superiore, in modulo, a 7 V sfruttano comunque l’effetto valanga, mentre quelli con breakdown sotto i 5 V si basano sull’effetto Zener; nel caso che la tensione di breakdown sia compresa tra 5 e 7 V si ha una combinazione dei due meccanismi. Esaminiamo ora l’origine fisica dei due tipi di breakdown. Il breakdown Zener deriva da un fenomeno abbastanza complesso di tunneling interbanda, che, senza conoscere la teoria a bande dei semiconduttori, non possiamo descrivere in dettaglio. Possiamo ricorrere a una rappresentazione molto semplificata, nella quale il breakdown Zener viene descritto come la conseguenza della rottura dei legami covalenti nella zona di svuotamento causata dall’elevato campo elettrico. La rottura di tali legami dà luogo a un gran numero di portatori liberi: gli elettroni generati in questo modo vengono trascinati dal campo elettrico verso il lato n , mentre le lacune vengono trascinate verso il lato p . Per piccoli incrementi della tensione inversa applicata al diodo si avranno forti aumenti della corrente, che a questo punto viene determinata sostanzialmente dal circuito esterno, essendo la caratteristica I - V quasi verticale.

Il breakdown a valanga si verifica invece quando il campo elettrico nella zona di svuotamento può accelerare i portatori minoritari che attraversano la zona stessa fino a una velocità tale da rompere i legami covalenti degli atomi con cui collidono. I portatori così liberati vengono a loro volta accelerati e causano ulteriori rotture di legami. Questo processo si sviluppa quindi a valanga, con un aumento dei portatori in progressione geometrica, tale da dar luogo a qualunque valore di corrente inversa determinato dal circuito esterno. Il processo di breakdown a valanga è illustrato nella figura, per quanto riguarda il processo di moltiplicazione degli elettroni; le lacune danno luogo a un analogo albero, rivolto nella direzione opposta: in questo modo la corrente risulta la stessa in ciascuna sezione trasversale.

Se, in conseguenza del passaggio della corrente di breakdown, la massima dissipazione di potenza ammissibile non viene superata, il breakdown è un fenomeno non distruttivo e il diodo torna a funzionare normalmente non appena la tensione inversa viene riportata sotto il valore di breakdown. Anche nel caso di breakdown a valanga la corrente è determinata sostanzialmente dal circuito esterno, dato che la caratteristica $I - V$ risulta praticamente verticale, una volta che sia iniziato il breakdown, come indicato nella figura seguente (dove la parte di caratteristica per polarizzazione diretta non è in pratica visibile, risultando, su questa scala, sostanzialmente schiacciata sull’asse delle ordinate).



È importante soffermarsi sulla dipendenza dalla temperatura dei due diversi tipi di breakdown: nel caso dell'effetto Zener un aumento della temperatura fa sì che gli elettroni dei legami covalenti possiedano un'energia più alta e che quindi vengano estratti più facilmente per effetto del campo elettrico, producendo un incremento della corrente, a parità di tensione applicata. Nel caso del breakdown a valanga, invece, un aumento di temperatura causa una maggiore probabilità di scattering e quindi i portatori acquistano in media un'energia minore, dando luogo, a parità di tensione applicata, a una minore corrente. Se andiamo a valutare l'effetto della temperatura a parità di corrente di breakdown, nel caso del meccanismo Zener la tensione inversa richiesta diminuisce, mentre in quello del meccanismo a valanga aumenta.

In generale i diodi che presentano tensioni di breakdown volutamente ridotte (al di sotto di qualche decina di volt) sono definiti “diodi Zener”, indipendentemente dall'effettivo meccanismo di breakdown. I diodi con “tensione di Zener” tra 5 e 7 V, in particolare quelli da 5.6 V, presentano la migliore stabilità in temperatura, dato che in essi sono contemporaneamente attivi i due meccanismi di breakdown, le cui dipendenze dalla temperatura si compensano.

1.9 Capacità di diffusione

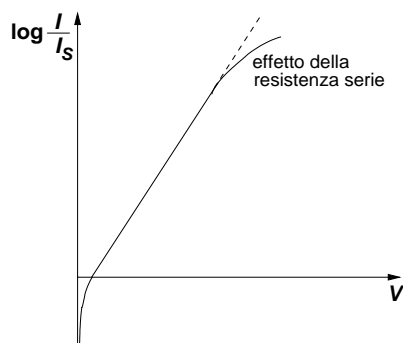
Nel funzionamento della giunzione pn in polarizzazione diretta un certo quantitativo di portatori minoritari in eccesso vengono immagazzinati ai lati della zona di svuotamento. Tale immagazzinamento dà luogo a una capacità, detta capacità di diffusione, decisamente diversa da quella di transizione già vista. Anche questa è una capacità differenziale e il suo valore è dato dall'espressione

$$C_d = \frac{\tau_T}{V_T} I,$$

dove τ_T è il tempo di transito medio attraverso il diodo, che è funzione dei tempi di vita medi (tempo che in media trascorre prima che un portatore minoritario si ricombini con un maggioritario) dei portatori minoritari. Se, come spesso avviene, il drogaggio è fortemente asimmetrico tra le due zone, per cui è prevalente la componente di diffusione in una delle due regioni, τ_T è praticamente coincidente con il tempo di vita medio dei portatori minoritari in tale regione. Si noti che la capacità di diffusione è direttamente proporzionale alla corrente, quindi in polarizzazione inversa essa risulta trascurabile.

1.10 Caratteristica I-V dei diodi

Per correnti elevate la caratteristica $I - V$ delle giunzioni pn comincia a deviare dall'andamento esponenziale previsto dalla legge di Shockley, a causa della presenza di una resistenza serie dovuta ai contatti e alla resistività delle regioni n e p , che causa una caduta di tensione non più trascurabile. Se rappresentiamo la caratteristica $I - V$ in scala logaritmica, questo effetto appare come una deviazione dalla linearità, come mostrato nella figura sottostante.



Una conseguenza importante dell'andamento esponenziale della caratteristica $I - V$ dato dalla legge di Shockley consiste nel fatto che la corrente risulta praticamente trascurabile quando la tensione è inferiore a un valore di soglia, di solito indicato con V_γ e definito "tensione di cut-in". Per diodi al germanio si ha $V_\gamma \approx 0.2$ V, mentre per quelli al silicio $V_\gamma \approx 0.7$ V.

Come abbiamo già detto, la corrente di saturazione inversa I_S dipende fortemente dalla temperatura e aumenta del 7-8% per ogni grado centigrado. Si ha un raddoppio di I_S per ogni incremento di temperatura di 10 °C:

$$I_S(T) = I_S(T_1) \times 2^{\frac{T-T_1}{10}}.$$

Per mantenere la corrente che attraversa un diodo costante all'aumentare della temperatura risulta quindi necessario abbassare la tensione applicata, di circa 2.2 mV/°C, a temperature vicine a quella ambiente.

2. Analisi del comportamento circuitale dei diodi

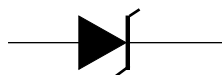
2.1 Simboli circuitali

Il simbolo circuitale utilizzato per rappresentare il diodo è riportato nella figura seguente: assomiglia a una freccia, rivolta nel verso in cui scorre la corrente in polarizzazione diretta. Il terminale posto più a sinistra nel disegno (e corrispondente alla



regione p) si definisce “anodo”, mentre l’altro si indica come “catodo” (e corrisponde alla regione n).

I diodi caratterizzati da tensioni di breakdown volutamente basse vengono, come già detto, indicati genericamente come “diodi Zener”, indipendentemente dal fatto che l’effettivo meccanismo di breakdown sia Zener o a valanga, e sono rappresentati nei circuiti con il seguente simbolo:



2.2 Retta di carico

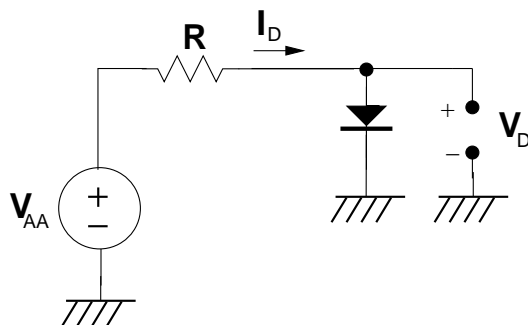
Introduciamo ora il concetto di retta di carico, che risulta molto utile ogni volta che si voglia studiare il comportamento di un bipolo non lineare inserito in un circuito lineare. Consideriamo il semplice circuito riportato nella figura seguente e scriviamo l’equazione che si ricava dall’applicazione della legge di Kirchhoff alle maglie:

$$-V_{AA} + I_D R + V_D = 0,$$

da cui otteniamo

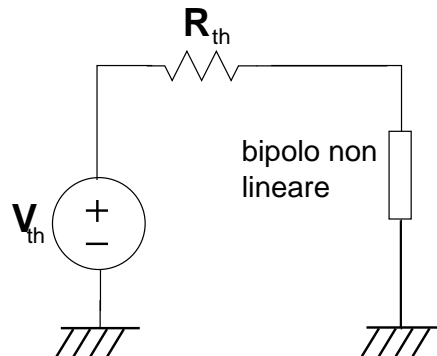
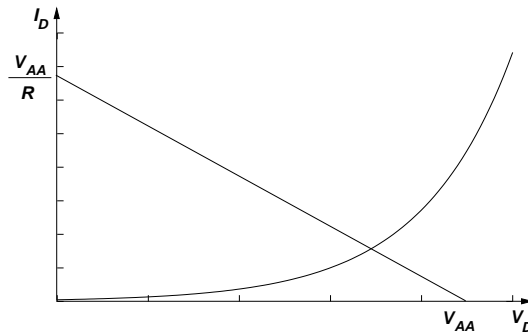
$$I_D = -\frac{1}{R}V_D + \frac{V_{AA}}{R}.$$

L’equazione sopra indicata fornisce I_D in funzione di V_D sotto forma di una relazione lineare che descrive una retta sul piano $V_D - I_D$, con coefficiente angolare e termine noto dipendenti soltanto da V_{AA} e da R . Tale retta, definita “retta di carico”, interseca l’asse delle ordinate in V_{AA}/R e quello delle ascisse in V_{AA} , e indica la relazione tra I_D e V_D imposta dal circuito esterno al diodo.



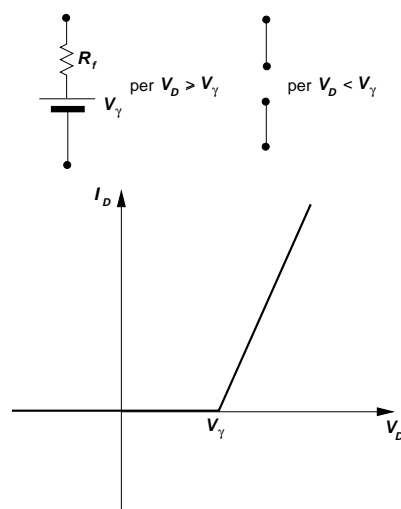
Se la caratteristica $I - V$ del diodo è data in forma grafica, possiamo trovare il punto di lavoro del diodo tracciando sulla stessa la retta di carico e andando a individuare l’intersezione tra le due curve, che corrisponde a soddisfare contemporaneamente i legami tra I_D e V_D imposti dal diodo e dal circuito esterno:

Se variamo il valore di V_{AA} la retta di carico si sposta parallelamente a se stessa, mentre modificando il valore di R essa ruota intorno all’intersezione con l’asse delle ascisse. Si noti che il concetto di retta di carico è del tutto generale e può quindi essere impiegato ogni volta che si abbia un bipolo non lineare inserito in una rete lineare: in tal caso la caratteristica data in forma grafica sarà quella del bipolo non lineare, mentre R e V_{AA} corrisponderanno, rispettivamente, alla resistenza equivalente di Thevenin e al generatore equivalente di Thevenin visti dal bipolo stesso.



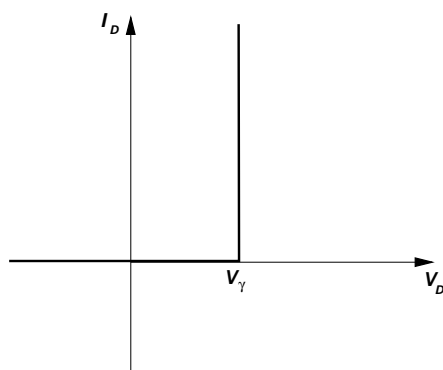
2.3 Circuiti equivalenti del diodo per grandi segnali

Esistono vari tipi di circuiti equivalenti utilizzati per l'analisi dei circuiti a diodi. Un circuito abbastanza completo è quello sotto rappresentato, nel quale il diodo in polarizzazione diretta viene sostituito con un generatore di tensione di valore V_γ in serie con una resistenza R_f , mentre in polarizzazione inversa lo si considera semplicemente un circuito aperto. Questo modello corrisponde ad approssimare la caratteristica $I - V$ del diodo con una spezzata, formata da una semiretta orizzontale coincidente con l'asse delle ascisse per tensioni negative e per tensioni positive fino a V_γ , e da una semiretta con pendenza $1/R_f$ per tensioni superiori a V_γ .

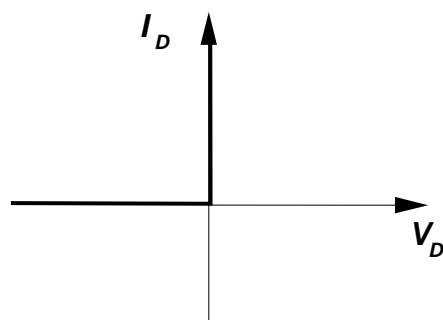


Un circuito equivalente più semplice si ottiene dal primo che abbiamo visto assumendo $R_f = 0$. In tal caso la caratteristica risulta approssimata dalla stessa semiretta già vista per il caso precedente per tensioni inferiori a V_γ , seguita poi da una semiretta

verticale. Si considera quindi il diodo un circuito aperto quando la tensione ai suoi capi è inferiore a V_γ e un generatore di tensione pari a V_γ non appena la tensione ai suoi capi raggiunge V_γ (che non può essere superata nell'ambito di questo modello).



Un circuito equivalente ancor più semplice si ottiene assumendo V_γ trascurabile: in tal caso il diodo viene trattato come un interruttore, che risulta aperto per tensioni applicate tra anodo e catodo negative e chiuse per tensione nulla e corrente che fluisce nel verso della polarizzazione diretta.

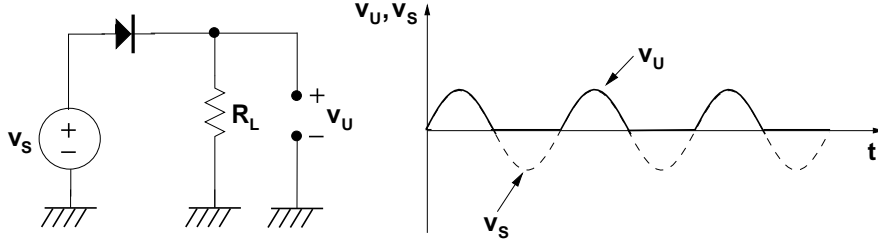


2.4 Metodi di analisi dei circuiti

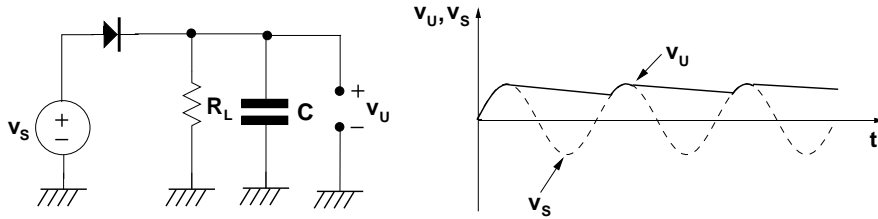
L'analisi dei circuiti a diodi si svolge partendo da un'ipotesi di lavoro per lo stato di conduzione o meno dei vari diodi, ricavata in genere per ispezione. Si risolve poi il circuito sostituendo ai diodi le reti equivalenti consistenti con il modello di diodo adottato e con lo stato assunto per ciascuno di essi. Una volta ricavate le tensioni e le correnti nel circuito, si va a verificare, per ciascun diodo, se queste sono consistenti con lo stato ipotizzato, vale a dire se la tensione che si presenta ai capi dei diodi considerati interdetti è effettivamente inferiore a V_γ (o semplicemente negativa se si considera V_γ trascurabile) e se la corrente nei diodi considerati in conduzione scorre effettivamente nel verso della polarizzazione diretta. Se queste verifiche hanno successo, data l'unicità della soluzione per una rete elettrica in cui siano presenti componenti con caratteristiche $I - V$ monotone, la procedura seguita è corretta e le grandezze trovate sono corrette. Se, invece, anche per un solo diodo, le condizioni non risultano soddisfatte, si deve ripartire da capo, considerando una configurazione diversa per lo stato dei diodi. È chiaro che in un circuito con un numero abbastanza significativo di diodi la scelta delle configurazioni da considerare deve essere guidata da un'ispezione attenta della rete, altrimenti il numero di tentativi necessari per arrivare alla soluzione corretta potrebbe essere troppo elevato.

2.5 Circuito rettificatore

Una delle applicazioni più semplici dei diodi consiste nella realizzazione di circuiti rettificatori, in grado cioè di ricavare un segnale a valor medio diverso da zero (quindi un segnale con componente continua non nulla) a partire da un segnale alternativo (e quindi a valor medio nullo). La versione più elementare di un circuito rettificatore è rappresentata nella figura sottostante.



Se V_S è una tensione alternata, il diodo condurrà soltanto nelle semionde positive (consideriamo V_γ trascurabile), quindi la tensione di uscita sarà costituita dalle sole semionde positive, come rappresentato in figura. Se si vuole ottenere qualcosa di più vicino a una tensione continua, è necessario inserire un condensatore in parallelo alla resistenza, realizzando un rettificatore con filtro capacitivo. Il diodo smette di condurre quando la tensione V_S scende al di sotto della tensione a cui è carico il condensatore. Da questo momento la tensione sul parallelo RC decade esponenzialmente finché non viene nuovamente superata da quella di ingresso e il diodo ritorna a condurre (l'esatto istante in cui termina la conduzione andrebbe calcolato confrontando la derivata della tensione in ingresso e di quella relativa al transitorio sul condensatore: la conduzione finirà quando la prima diventa maggiore in modulo della seconda). Tale decadimento esponenziale, se la costante di tempo RC è molto maggiore del periodo della tensione di ingresso, può essere approssimato da un andamento lineare, come nel grafico sotto riportato.



La tensione di uscita è quindi simile a una continua, ma con una fluttuazione (denominata “ripple”) la cui ampiezza è tanto più piccola quanto più grande è la costante di tempo RC rispetto al periodo di V_S .

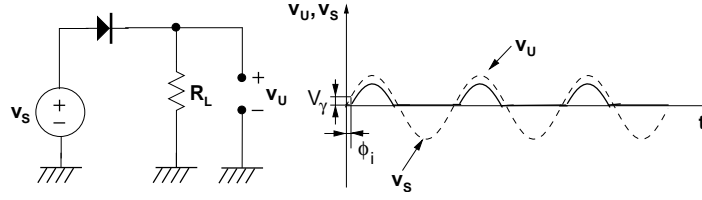
Ritornando al rettificatore senza filtro capacitivo, vediamo che cosa succede nel caso che V_γ non sia trascurabile in confronto alla tensione di ingresso. In tal caso il diodo conduce soltanto nell'intervallo nel quale V_S risulta maggiore di V_γ , per cui in uscita ogni semiperiodo inizia con un ritardo ϕ_i e termina con un anticipo ϕ_i (come indicato nella figura). La differenza di fase ϕ_i può essere valutata semplicemente:

$$v_u = R_L \frac{v_s - V_\gamma}{R_L} = V_M \sin \omega t - V_\gamma,$$

quindi

$$V_M \sin \phi_i - V_\gamma = 0$$

e perciò $\phi_i = \arcsin V_\gamma / V_M$.

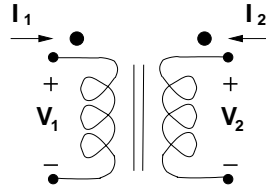


In genere i rettificatori vengono utilizzati per ottenere, a partire dalla tensione alternata a 230 V, 50 Hz della rete di distribuzione, una tensione continua da impiegare per l'alimentazione di apparecchiature elettroniche. A questo scopo è di solito necessario abbassare la tensione e creare una separazione dalla rete stessa: tale risultato si ottiene con un trasformatore, come rappresentato in figura. Il flusso magnetico presente nel nucleo di un trasformatore è pressoché costante e di valore in genere piccolo rispetto al rapporto tra la forza magnetomotrice dovuta alla normale corrente di funzionamento in ciascun avvolgimento e la riluttanza del circuito magnetico (questo perché al crescere della corrente nel primario cresce anche quella nel secondario, che ha un effetto opposto in termini di flusso). Prenderemo in considerazione soltanto i trasformatori ideali, per i quali il flusso magnetico totale può essere considerato trascurabile, per cui

$$N_1 I_1 + N_2 I_2 = 0 \rightarrow \frac{-I_2}{I_1} = \frac{N_1}{N_2}.$$

Si può anche semplicemente dimostrare che

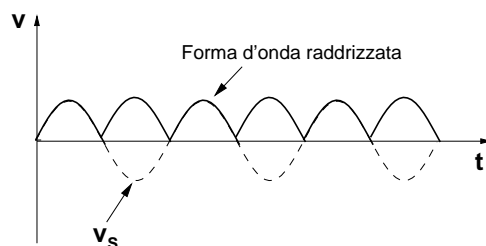
$$\frac{V_2}{V_1} = \frac{N_2}{N_1}.$$



I pallini indicano il verso relativo degli avvolgimenti, nel senso che se la corrente è entrante su tutti i terminali contrassegnati con pallini, i flussi prodotti saranno tutti concordi. È quindi evidente che, nel normale funzionamento del trasformatore, se la corrente entra nel pallino del primario, sul secondario la corrente dovrà uscire dal terminale contrassegnato con il pallino, dovendo essere il flusso totale nullo. Inoltre, se la tensione viene applicata al primario con polarità positiva sul terminale con pallino, per ottenere il verso della corrente sul secondario prima indicato si dovrà avere una tensione sul secondario con polarità positiva in corrispondenza del terminale con pallino.

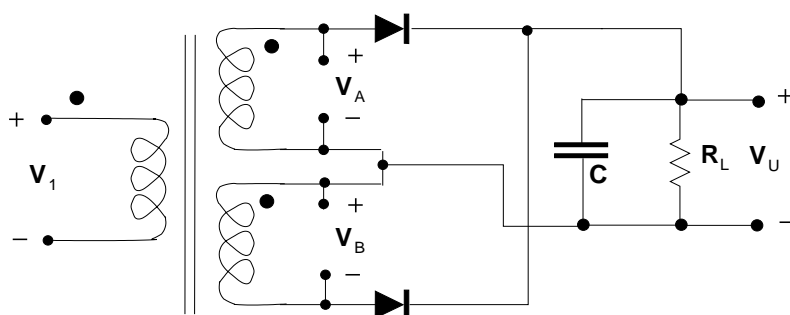
2.6 Raddrizzatori a doppia semionda

Allo scopo di diminuire il ripple in uscita risulta conveniente “raddrizzare” piuttosto che “rettificare” una tensione alternata, vale a dire, oltre a selezionare le semionde



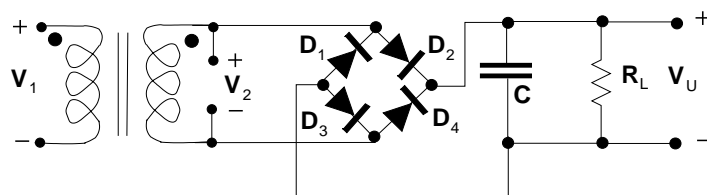
positive, rendere positive anche quelle negative, come indicato nella figura sottostante.

Tale risultato può essere ottenuto utilizzando un trasformatore con due secondari e rettificando le tensioni di uscita, prelevate in opposizione di fase tra loro. Con riferimento allo schema sotto riportato, durante la semionda positiva della tensione di ingresso conduce il diodo superiore, dando luogo a una semionda positiva in uscita, mentre durante quella negativa conduce il diodo inferiore, dando comunque luogo a una semionda positiva in uscita.



Con una soluzione di questo tipo, a parità di costante di tempo RC , si ottiene un ripple ridotto, perché il periodo della tensione pulsante ottenuta risulta dimezzato. Vediamo qual è il valore di tensione inversa che i diodi impiegati devono essere in grado di sopportare. Consideriamo per esempio il caso di semionda positiva in ingresso: il condensatore si può considerare sempre carico al valore massimo V_M della tensione sul secondario, quindi ai capi del diodo inferiore (D_2) sarà presente una tensione $V_{D2} = V_B - V_M$. Poiché la tensione V_B raggiunge il valore minimo $-V_M$, la massima tensione inversa sul diodo D_2 sarà, in modulo, di $2V_M$. Lo stesso ragionamento si può ripetere per il diodo superiore, ottenendo lo stesso valore di massima tensione inversa. Andranno quindi scelti dei diodi caratterizzati da una tensione di breakdown abbastanza più grande di $2V_M$ (di solito si impiegano diodi con tensione di breakdown almeno doppia rispetto a $2V_M$).

È possibile ottenere un raddrizzamento di una tensione alternata anche senza ricorrere a un trasformatore con doppio secondario, utilizzando quattro diodi connessi nella configurazione indicata in figura, denominata “ponte di Graetz”.



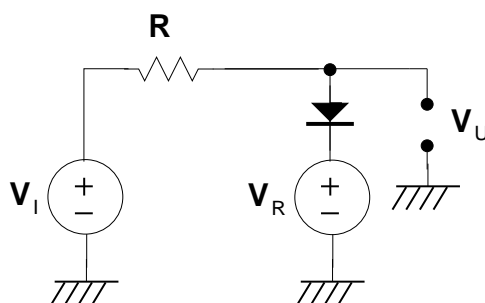
Per l'analisi del funzionamento supponiamo, per semplicità, che non ci sia il condensatore di filtro, il quale introdurrà lo stesso effetto di livellamento della tensione di

uscita già visto in precedenza. Facciamo l'ipotesi, in base a un'ispezione della rete, che durante la semionda positiva della V_2 conducano i diodi D_2 e D_3 e siano interdetti D_1 e D_4 . Procediamo all'analisi del circuito supponendo di sostituire dei corto circuiti ai diodi in conduzione e dei circuiti aperti a quelli interdetti. Verifichiamo il verso delle correnti nei diodi supposti in conduzione e quello delle tensioni ai capi dei diodi supposti interdetti: la corrente scorre attraverso il diodo D_2 e da questo raggiunge la resistenza di carico, ritornando poi al trasformatore tramite D_3 . Quindi il verso della corrente su D_2 e D_3 è coerente con le ipotesi fatte. La tensione ai capi di D_1 e D_4 , misurata tra anodo e catodo, risulta pari alla tensione di uscita del trasformatore cambiata di segno, ed è quindi negativa, confermando l'interdizione. Un ragionamento del tutto analogo porta a concludere che durante la semionda negativa condurranno i diodi D_1 e D_4 , mentre saranno interdetti D_2 e D_3 .

Se si tiene in considerazione anche l'effetto del condensatore, l'unica variante consiste nel fatto che, per esempio nella semionda positiva, i diodi D_2 e D_3 condurranno soltanto nei brevi intervalli necessari a ricondurre la tensione ai capi del condensatore al valore di picco V_M della V_2 e i diodi D_1 e D_4 saranno sottoposti a una polarizzazione inversa pari alla tensione sul condensatore e pari quindi a V_M . Concludiamo quindi che in un raddrizzatore a ponte di Graetz si possono utilizzare diodi di costo inferiore rispetto al caso visto prima del raddrizzatore con trasformatore a doppio secondario, dato che questi devono sopportare una tensione massima inversa pari solo a V_M invece di $2V_M$.

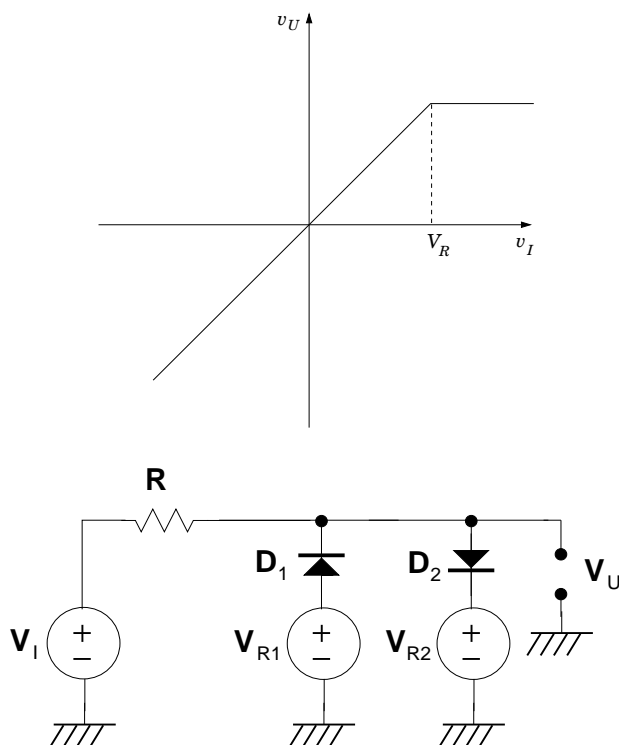
2.7 Circuiti tagliatori e fissatori

I circuiti tagliatori sono utilizzati per selezionare parti di una forma d'onda che stanno al di sopra o al di sotto di un livello di riferimento. In questo senso anche il circuito rettificatore potrebbe considerarsi un tagliatore che seleziona la parte del segnale di ingresso superiore a zero. Lo schema più semplice di un circuito tagliatore è quello sotto riportato:



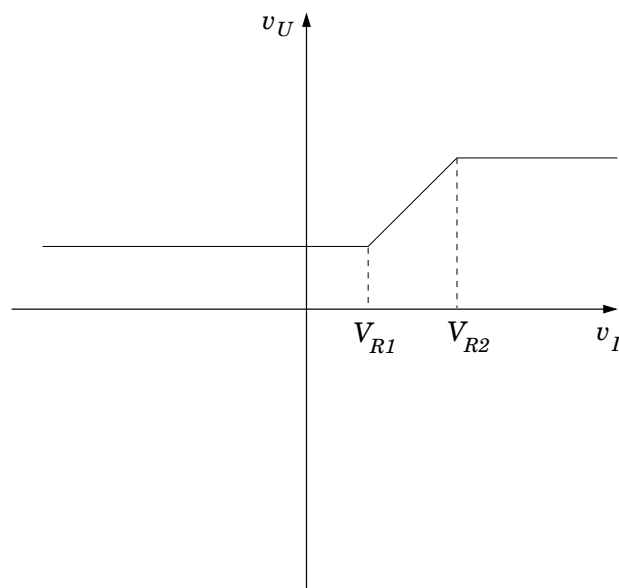
Per l'analisi del circuito consideriamo il modello di diodo più semplice, quello con $V_\gamma = 0$ e $R_f = 0$: se la tensione di ingresso v_I è minore di V_R , il diodo risulta interdetto e la tensione di uscita v_U è pari a v_I . Quando invece la tensione di ingresso supera V_R , il diodo conduce e la tensione di uscita risulta pari a V_R . Possiamo tracciare un grafico, detto caratteristica di trasferimento, della v_U in funzione della v_I : per $v_I < V_R$ si tratta semplicemente di una semiretta con coefficiente angolare unitario; poi, per $v_I \geq V_R$, diventa una semiretta orizzontale.

Se V_γ non fosse trascurabile, la tensione di uscita sarebbe pari a $V_R + V_\gamma$ per $v_I > V_R + V_\gamma$ e pari a v_I per gli altri valori di v_I : il grafico della caratteristica di trasferimento rimarrebbe quindi inalterato, eccetto che per la sostituzione di V_R con $V_R + V_\gamma$.



È anche possibile realizzare un doppio tagliatore, nel quale sono presenti due generatori di tensione V_{R1} e V_{R2} , connessi tramite i diodi D_1 e D_2 , come indicato nella figura sottostante.

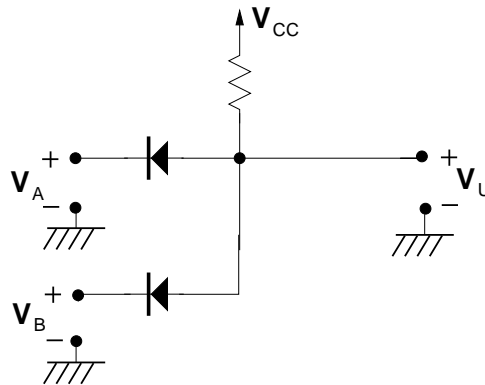
Per evitare il passaggio di una corrente di corto circuito nella maglia formata da V_{R1} , D_1 , D_2 , V_{R2} , deve risultare $V_{R2} > V_{R1}$. Nell'ipotesi di V_γ trascurabile, se la tensione di ingresso è inferiore a V_{R1} , il diodo D_1 conduce, D_2 è interdetto e la tensione di uscita risulta pari a V_{R1} ; per $V_{R1} \leq v_I \leq V_{R2}$ ambedue i diodi sono interdetti e la tensione v_U è pari a quella d'ingresso; per $v_I > V_{R2}$ il diodo D_1 è interdetto, D_2 conduce e la tensione di uscita è pari a V_{R2} . La caratteristica di trasferimento può quindi essere rappresentata come una spezzata costituita da due semirette orizzontali connesse da un segmento con pendenza unitaria:



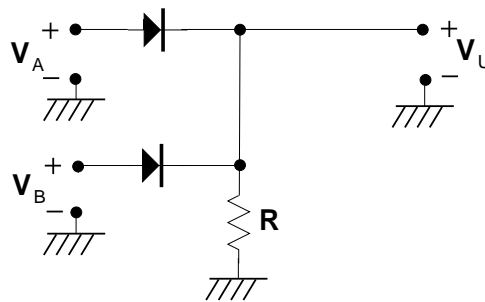
2.8 Logica a diodi

È possibile, utilizzando diodi, resistori e generatori di tensione, realizzare alcune semplici funzioni logiche. Esamineremo due esempi di porte logiche e discuteremo i problemi che la logica a diodi presenta. Nell'analisi che segue sarà utilizzato il modello di diodo per grandi segnali più semplice, quello con $V_\gamma = 0$ e $R_f = 0$, che corrisponde a un interruttore ideale. Associamo il valore logico 1 a una tensione di 5 V e il valore logico 0 a una tensione di 0 V.

In tale contesto il circuito sotto riportato svolge la funzione di una porta logica AND: l'uscita è a livello logico 1 solo se entrambe gli ingressi lo sono. È infatti sufficiente che soltanto V_A o soltanto V_B siano a tensione 0 perché il corrispondente diodo conduca e anche l'uscita vada a 0.

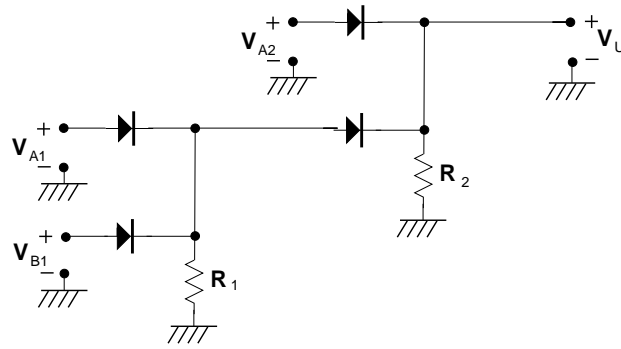


È anche possibile realizzare la funzione OR, il cui circuito, sotto rappresentato, risulta ancor più semplice, richiedendo soltanto due diodi e un resistore. È sufficiente che uno degli ingressi si trovi alla tensione di 5 V perché anche l'uscita venga portata alla stessa tensione, in conseguenza dell'entrata in conduzione del diodo corrispondente. Se anche l'altro ingresso viene portato a 5 V, il diodo relativo condurrà anch'esso; altrimenti risulterà interdetto.



Questo approccio alla implementazione delle porte logiche presenta però dei problemi nel caso in cui si vogliano realizzare dei circuiti più complessi. Consideriamo l'esempio di due porte OR in cascata, come rappresentato nella figura seguente.

Supponiamo di porre $V_{B1} = 5$ V e $V_{A1} = 0$ V. L'uscita del primo OR sarà alta (a 5 V) e tale sarà anche quella del secondo OR. In ingresso a V_{B1} si dovrà quindi fornire sia la corrente che scorre in R_1 sia quella che scorre in R_2 . Se si aggiungono ulteriori porte in cascata, la corrente richiesta diverrà ben presto estremamente elevata e porrà seri problemi pratici. Un altro inconveniente è rappresentato dal fatto che in realtà i diodi non sono interruttori ideali e che quindi hanno $V_\gamma \simeq 0.7$ V (se si tratta di diodi realizzati su silicio). Quindi, se $V_{B1} = 5$ V, avremo $V_{B2} = 4.3$ V e $V_U = 3.6$ V.

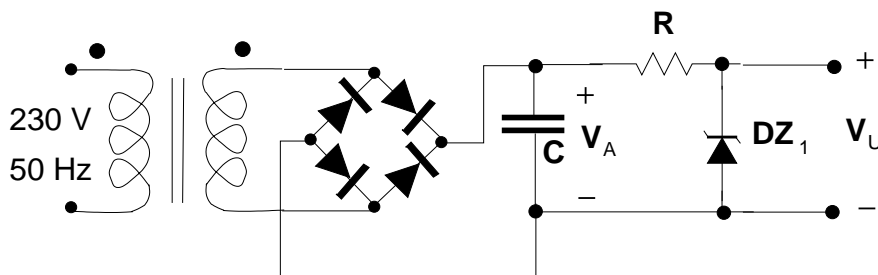


Si assisterà quindi a un progressivo degrado dei livelli logici, che rende impossibile la realizzazione di circuiti con un numero significativo di porte in cascata. È questa la ragione principale per cui le porte logiche a soli diodi sono utilizzate solo per applicazioni estremamente semplici, mentre in tutti gli altri casi si ricorre a circuiti più complessi, utilizzando componenti in grado di fornire un guadagno di potenza.

Un ulteriore, molto grave inconveniente della logica a diodi consiste nel fatto che con essa non è possibile realizzare una delle funzioni logiche fondamentali, quella della porta NOT, che è invece indispensabile per l'implementazione di una generica rete combinatoria.

2.9 Regolatore di tensione con diodo Zener

Abbiamo precedentemente discusso l'utilizzo dei diodi per la rettificazione e il rad-drizzamento di tensioni alternate, utilizzati in circuiti aventi lo scopo di ottenere una tensione continua utile per l'alimentazione di circuiti elettronici. Un problema che si presenta nella realizzazione di circuiti alimentatori, cioè in grado di trasformare la tensione alternata di rete in una tensione continua da utilizzare come sorgente di alimentazione, è rappresentato dalle variazioni che la tensione di uscita subisce al variare della corrente assorbita dal carico e come conseguenza delle inevitabili fluttuazioni della tensione di rete. Per porre rimedio a questi inconvenienti, si introducono circuiti di regolazione, la cui funzione è quella di mantenere costante la tensione d'uscita. Il più semplice circuito di regolazione che si possa concepire è rappresentato da un resistore e un diodo Zener, come mostrato nella figura sottostante, collocati in uscita al circuito.



Per l'analisi di questa rete supponiamo, in prima approssimazione, che le variazioni della tensione V_A ai capi del condensatore di livellamento siano trascurabili rispetto alla differenza tra V_A e V_Z , la tensione di breakdown del diodo DZ_1 . Supponiamo inoltre che V_Z sia minore di V_A : il diodo DZ_1 funzionerà quindi in breakdown e la tensione ai suoi capi, fornita poi al carico, sarà proprio V_Z . Nell'ipotesi fatta di trascurabilità delle variazioni della V_A , il resistore R sarà attraversato da una corrente

costante $I = (V_A - V_Z)/R$. Il massimo assorbimento da parte del carico per cui questo circuito continua a funzionare nel modo desiderato (vale a dire con $V_U = V_Z$) è pari a I : per correnti sul carico al di sopra di I la tensione sul diodo Zener cade al di sotto di V_Z , il diodo si interdice e non svolge più alcuna funzione. Se invece il carico assorbe una corrente nulla, tutta la I va nel diodo Zener, il quale deve essere quindi in grado di dissipare una potenza $V_Z I$. Risulta dunque chiaro che un circuito regolatore di questo tipo può essere utilizzato soltanto per carichi di potenza limitata, che assorbano, al più, qualche centinaio di milliampere, altrimenti si dovrebbe accettare, nel funzionamento a vuoto, una eccessiva dissipazione di potenza sul diodo Zener. Se prendiamo in considerazione anche il fatto che la V_A non è costante, ma variabile in funzione della tensione di rete e dell'assorbimento del carico, l'analisi del circuito non varia di molto: la corrente che attraversa il resistore R non sarà esattamente costante, ma il limite per la corrente massima fornibile al carico senza che il diodo Zener esca dalla zona di breakdown sarà comunque rappresentato da $(V_{A\min} - V_Z)/R$ (dove con $V_{A\min}$ si è indicato il valore minimo che V_A può raggiungere). Inoltre la dissipazione massima sul diodo sarà data da $(V_{A\max} - V_Z)V_Z/R$ (dove $V_{A\max}$ è il massimo valore di tensione raggiungibile da V_A). Fino a questo punto abbiamo assunto la caratteristica $I - V$ del diodo Zener esattamente verticale in corrispondenza della tensione di breakdown; per caratteristiche reali, nelle quali la resistenza differenziale nella regione di breakdown è non nulla, il comportamento del circuito regolatore è leggermente diverso e la tensione di uscita dipende, anche se di poco, dall'assorbimento di corrente del carico.

3. Analisi dei circuiti a diodi per piccoli segnali

3.1 Modello del diodo per piccoli segnali

Fino a questo punto abbiamo preso in considerazione il comportamento del diodo per “grandi segnali”, per i quali tale comportamento risulta fortemente e palesemente non lineare. Abbiamo tuttavia introdotto ugualmente dei modelli linearizzati, che consentono un'analisi semplificata, sotto opportune condizioni, dei circuiti a diodi. Se, per esempio, si è interessati a conoscere la corrente che scorre in un diodo in polarizzazione diretta alimentato da un generatore di tensione tramite una resistenza, l'errore che si commette sostituendo al diodo un corto circuito è trascurabile, purché la tensione di uscita del generatore sia molto più grande di V_γ . Tuttavia, se vogliamo conoscere con buona precisione la variazione di tensione ai capi del diodo in conseguenza di una variazione della tensione del generatore utilizzato per l'alimentazione, l'errore percentuale che commettiamo sarà notevolissimo, anche utilizzando il modello più completo visto finora, quello con il generatore V_γ e la resistenza R_f . Risulta quindi necessario disporre di modelli più precisi del diodo e se, in particolare, si vuole valutare in maniera accurata la risposta a piccoli segnali che costituiscono variazioni intorno al punto di riposo, è possibile ricorrere a modelli linearizzati che consentono comunque valutazioni estremamente affidabili. Si parla in questo caso di “modelli per piccoli segnali”.

Il modello linearizzato per piccoli segnali che considereremo è valido quando la componente variabile nel tempo delle grandezze applicate ai capi del diodo è molto piccola rispetto al valor medio a essa sovrapposto e risulta comunque abbastanza piccola da rendere valida la procedura di linearizzazione utilizzata.

Introduciamo innanzitutto la notazione che useremo poi estesamente nel seguito: si indicano con lettera minuscola e pedice minuscolo le componenti variabili nel tempo

delle grandezze (per esempio v_i), si utilizza invece una lettera maiuscola con pedice maiuscolo per le grandezze medie (per esempio V_I) e, infine, la grandezza totale viene indicata tramite una lettera minuscola con pedice maiuscolo ($v_I = V_I + v_i$). Si impiega invece la lettera maiuscola con pedice minuscolo (V_i) per rappresentare le grandezze espresse come fasori nel dominio trasformato della frequenza f o in quello della frequenza generalizzata s .

Esaminiamo ora il concetto di linearizzazione in generale. Se applichiamo una tensione v a un bipolo non lineare, questo sarà attraversato da una corrente i che è una funzione di v :

$$i = f(v).$$

Supponiamo di operare intorno a una tensione di riposo V_Q e di essere interessati alle variazioni intorno a tale punto di riposo. In questo caso possiamo sviluppare in serie di Taylor la funzione $f(v)$ intorno al punto V_Q :

$$i(V_Q + \delta v) = i(V_Q) + \left. \frac{di}{dv} \right|_{v=V_Q} \delta v + \frac{1}{2} \left. \frac{d^2 i}{dv^2} \right|_{v=V_Q} \delta v^2 + \dots + \frac{1}{n!} \left. \frac{d^n i}{dv^n} \right|_{v=V_Q} \delta v^n.$$

Se δv è abbastanza piccolo e se la funzione $f(v)$ è sufficientemente “well behaved”, i termini in cui compaiono potenze di δv superiori alla prima possono essere trascurati e abbiamo quindi

$$i(V_Q + \delta v) \simeq i(V_Q) + \left. \frac{di}{dv} \right|_{v=V_Q} \delta v.$$

Tutto ciò equivale ad aver effettuato una linearizzazione del problema, dato che la dipendenza di i da v è ora espressa, per valori piccoli di δv , tramite una relazione lineare. Possiamo anche riscrivere la relazione linearizzata nella forma

$$i(V_Q + \delta v) \simeq i(V_Q) + \frac{\delta v}{\left. \frac{dv}{di} \right|_{v=V_Q}} = i(V_Q) + \frac{\delta v}{r_d|_{v=V_Q}},$$

dove è stata introdotta r_d (resistenza differenziale del diodo), che corrisponde alla derivata della tensione ai capi del diodo rispetto alla corrente che lo attraversa, valutata nel punto di lavoro V_Q . Se la caratteristica del diodo è disponibile in forma grafica, dobbiamo innanzitutto individuare il punto di lavoro tramite l'intersezione con la retta di carico, dopodiché la resistenza differenziale può essere ricavata come reciproco del coefficiente angolare della retta tangente alla caratteristica in tale punto. Se la caratteristica è invece fornita in forma analitica, possiamo usare un metodo iterativo per la determinazione del punto di lavoro, o una delle approssimazioni già viste, dopodiché la resistenza differenziale può essere ricavata dalla derivata della relazione $I - V$:

$$\begin{aligned} i &= I_S \left(e^{v/(\eta V_T)} - 1 \right) \\ g_d &= \left. \frac{di}{dv} \right|_{v=V_Q} = \frac{1}{\eta V_T} I_S e^{v/(\eta V_T)} \simeq \frac{1}{\eta V_T} i|_{v=V_Q} = \frac{1}{\eta V_T} I_Q. \\ r_d &= 1/g_d = \frac{\eta V_T}{I_Q} \end{aligned}$$

Quindi la resistenza differenziale è inversamente proporzionale alla corrente di riposo. Sottolineiamo che esiste una differenza sostanziale rispetto al modello linearizzato per

grandi segnali: in quel caso avevamo considerato un unico valore di resistenza costante al variare del punto di riposo, che costituiva una grossolana rappresentazione della realtà, dato che r_d è, come abbiamo appena dimostrato, fortemente dipendente dal punto di lavoro scelto.

Svolgiamo una semplice valutazione dell'ampiezza del segnale dipendente dal tempo per cui l'approssimazione consistente nel modello linearizzato del diodo risulta ragionevolmente applicabile. La derivata seconda della corrente nel diodo rispetto alla tensione ai suoi capi risulta

$$\frac{d^2 i}{dv^2} = \frac{1}{(\eta V_T)^2} i(v),$$

e, più in generale, la derivata n -esima risulta

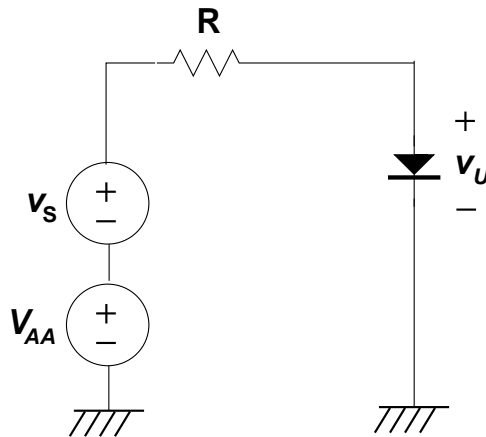
$$\frac{d^n i}{dv^n} = \frac{1}{(\eta V_T)^n} i(v).$$

Se confrontiamo il termine di ordine 1 dello sviluppo in serie di Taylor con quello di ordine 2, notiamo che se δv è minore di $\eta V_T/10$, il termine di ordine 2 è minore di un ventesimo di quello di ordine 1. Pertanto un termine di paragone ragionevole per decidere se il modello linearizzato è applicabile può consistere proprio in $V_T/10$: considerando che il fattore di idealità η è compreso tra 1 e 2, se δv è minore di $V_T/10$ i termini dello sviluppo in serie di ordine superiore al primo sono decisamente piccoli, anche se, a rigore, non del tutto trascurabili. Se poi $\delta v \ll V_T/10$ possiamo essere del tutto certi della bontà dell'approssimazione lineare. Si noti che non è necessario che le componenti variabili nel tempo delle tensioni presenti nel circuito siano minori $V_T/10$: è sufficiente che la variazione nel tempo da esse causata δv della tensione ai capi del diodo sia minore di $V_T/10$. Spesso si applica la procedura di linearizzazione purché risulti $\delta v < V_T/5$, dato che la soluzione del problema non lineare è realisticamente affrontabile solo con metodi numerici.

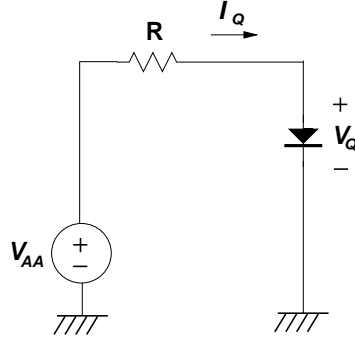
3.2 Esempio di applicazione del modello per piccoli segnali

Prendiamo in considerazione il circuito sotto rappresentato, costituito da due generatori di tensione, di cui uno (V_{AA}) costante e uno ($v_s(t)$) variabile nel tempo, un resistore R e un diodo D . Supponiamo che l'andamento nel tempo della tensione di $v_s(t)$ sia sinusoidale:

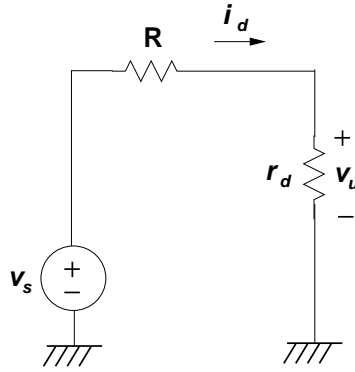
$$v_s(t) = V_M \sin(\omega t).$$



Per prima cosa si deve determinare il punto di lavoro, tramite uno dei metodi precedentemente visti, ricavando in tal modo V_Q, I_Q e, conseguentemente, r_d . Per determinare il punto di riposo si utilizza il cosiddetto “circuitto in continua”, che contiene soltanto il generatore V_{AA} , il resistore R e il diodo D .



Una volta determinati i parametri sopra citati si passa allo studio del cosiddetto “circuitto dinamico” o “circuitto in alternata” nel quale compaiono il generatore $v_s(t)$ e la resistenza R , e il diodo è rappresentato per mezzo della sua resistenza differenziale r_d .



Il valore della componente alternativa in uscita può ora essere calcolato:

$$v_u(t) = v_s(t) \frac{r_d}{R + r_d} = V'_M \sin(\omega t),$$

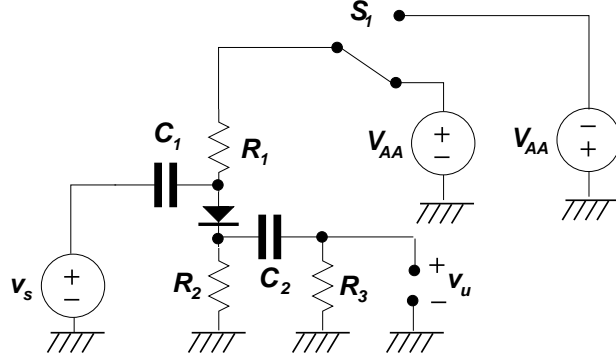
dove $V'_M = V_M r_d / (R + r_d)$. La tensione totale presente sull'uscita può infine essere ottenuta sommando la componente continua e quella variabile nel tempo:

$$v_U = V_U + v_u = V_Q + V'_M \sin \omega t.$$

3.3 Interruttore a diodo

Sfruttando le proprietà dei diodi che abbiamo esaminato è possibile ottenere un interruttore comandato per un segnale variabile nel tempo. Si verifica spesso la necessità di controllare il passaggio di un segnale dipendente dal tempo: tale controllo potrebbe essere ottenuto tramite un normale interruttore, ma in alcune situazioni l'azionamento deve avvenire da una posizione lontana dall'effettivo percorso del segnale. Si potrebbe deviare il percorso del segnale utilizzando dei cavi, ma ciò porterebbe facilmente a introdurre dei disturbi che degraderebbero la qualità del segnale.

stesso. È peraltro possibile, utilizzando un diodo, controllare il passaggio del segnale in questione tramite una tensione continua, che può essere trasportata anche a distanze significative senza creare problemi. Un esempio di un circuito in grado di svolgere questa funzione è riportato di seguito.



Nel circuito compaiono anche due condensatori, C_1 e C_2 , che considereremo di valore tendente all'infinito, e quindi assimilabili a corto circuiti per le componenti variabili nel tempo e a circuiti aperti per le componenti continue. Quando il deviatore S_1 è nella posizione bassa, il diodo risulta polarizzato direttamente e consente quindi il passaggio della corrente, presentando inoltre una resistenza differenziale abbastanza bassa, che costituisce un corto circuito per il segnale alternativo, il quale viene pertanto trasferito in uscita. Quando, invece, il deviatore S_1 è nella posizione alta, il diodo risulta polarizzato inversamente e si comporta sostanzialmente come un circuito aperto, impedendo il passaggio del segnale $v_s(t)$ verso l'uscita.

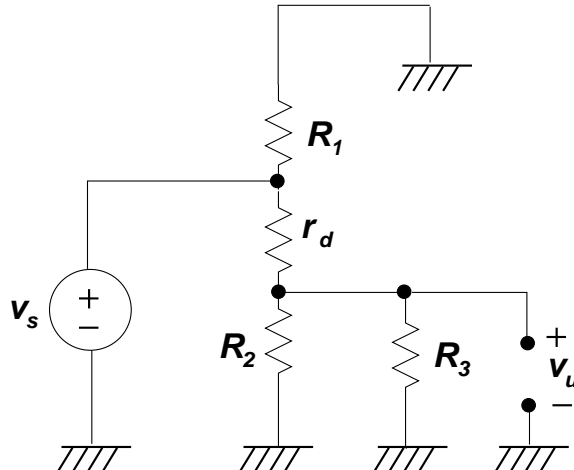
Svolgiamo ora un'analisi quantitativa approssimata. Iniziamo dal caso in cui il deviatore sia nella posizione in basso: il diodo risulta polarizzato direttamente, con una corrente I_Q che può essere valutata rapidamente da

$$I_Q = \frac{V_{AA} - V_\gamma}{R_1 + R_2}.$$

Possiamo poi ricavare r_d da

$$r_d = \frac{\eta V_T}{I_Q}.$$

A questo punto possiamo tracciare il circuito dinamico, che risulterà:

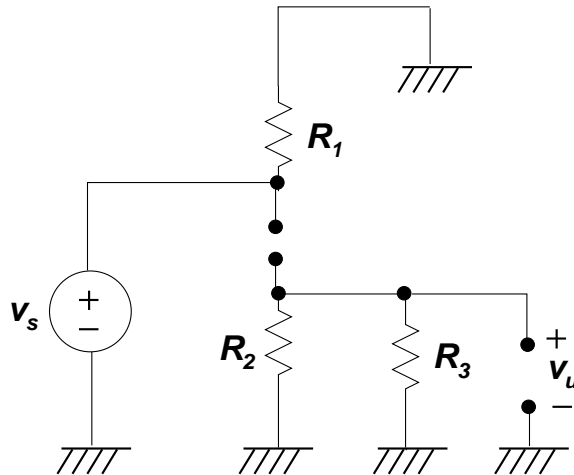


Da tale circuito valutiamo la componente dinamica in uscita:

$$v_u(t) = \frac{R_2 // R_3}{r_d + (R_2 // R_3)} v_s(t).$$

Quindi, se $r_d \ll R_2 // R_3$, la tensione di uscita è praticamente coincidente con quella di ingresso. È importante notare che la corrente scorre nel diodo sempre nel verso della polarizzazione diretta, anche quando $v_s(t)$ risulta negativa, perché la corrente totale nel diodo è il risultato della somma della componente continua I_Q e di quella dovuta a $v_s(t)$. Perché il diodo rimanga sempre in conduzione, la componente dinamica della corrente che attraversa il diodo deve essere quindi in ogni momento minore della corrente di riposo. Requisiti ancor più stringenti devono essere soddisfatti perché sia valido il trattamento linearizzato del circuito dinamico: il modulo della corrente dovuta a $v_s(t)$ dovrà essere molto minore di I_Q in qualsiasi istante, altrimenti l'ipotesi di r_d costante non sarebbe più valida. In questo caso particolare il comportamento del diodo rimarrà più o meno lineare finché la componente variabile nel tempo della tensione ai suoi capi avrà ampiezza minore di V_T (26 mV a temperatura ambiente) o, comunque, dello stesso ordine di grandezza.

Nel caso in cui il deviatore si trovi nella posizione più in alto, il diodo è polarizzato inversamente e il circuito dinamico risulta il seguente.

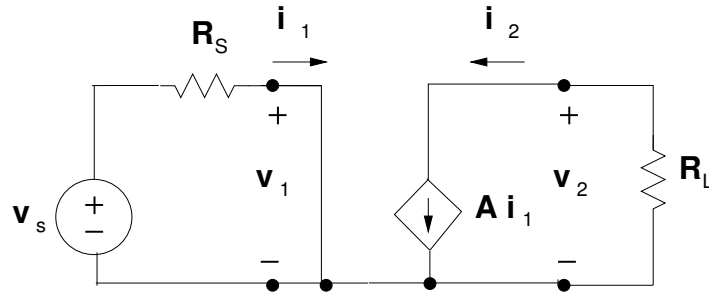


L'uscita è pertanto nulla, dato che non esiste un percorso per il segnale tra la porta di ingresso e quella di uscita.

4. Il transistor BJT

4.1 Generatore di corrente controllato in corrente

Una funzione essenziale nella maggior parte dei circuiti elettronici è rappresentata dal generatore controllato. Un generatore controllato è un dispositivo a due porte, quindi con quattro terminali, che si riducono tuttavia nella maggior parte dei casi a tre, perché due sono in comune tra ingresso e uscita. Prendiamo inizialmente in considerazione il generatore di corrente controllato in corrente, perché questo può essere approssimato tramite il transistor bipolare a giunzione (Bipolar Junction Transistor, BJT). In forma ideale tale generatore può essere rappresentato con lo schema che segue.



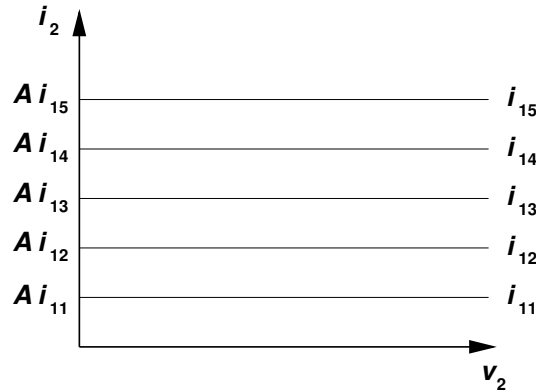
La quantità A che rappresenta il rapporto tra la corrente di uscita e quella di ingresso viene definita “guadagno di corrente”. Notiamo che l’impedenza di ingresso è stata scelta volutamente nulla in modo che la corrente di ingresso sia massima. Notiamo anche che il dispositivo è unidirezionale, cioè che le grandezze elettriche eventualmente applicate all’uscita non influenzano in alcun modo quelle in ingresso. Possiamo facilmente calcolare il rapporto tra la tensione v_s e quella v_2 di uscita:

$$i_1 = \frac{v_s}{R_s} \quad v_2 = -A i_1 R_L = -\frac{A R_L}{R_s} v_s.$$

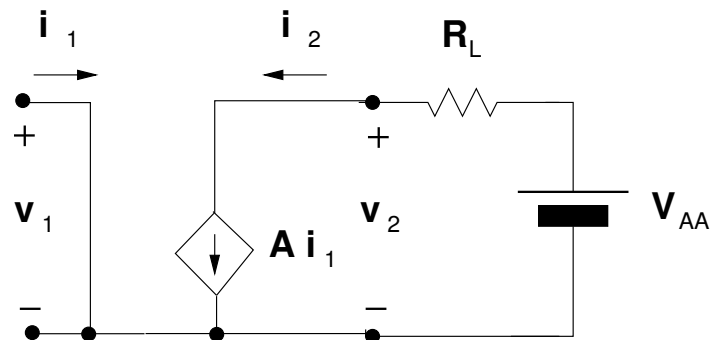
Se $A R_L / R_s$ è maggiore dell’unità $|v_2| > |v_s|$ e si ha quindi un’amplificazione di tensione. Se inoltre $A > 1$, come di solito accade in pratica, si ha anche un’amplificazione di corrente. Se il prodotto del guadagno di tensione per quello di corrente, $A^2 R_L / R_s$ è maggiore dell’unità si ha un’amplificazione di potenza: significa quindi che è possibile controllare con un piccolo segnale un segnale di potenza maggiore. Proprio questo effetto rappresenta la caratteristica peculiare dei cosiddetti “componenti attivi”, dei quali il BJT sarà il primo esempio che esamineremo. È importante sottolineare fino da ora che la potenza in più che si presenta all’uscita rispetto a quella in ingresso proviene dalla sorgente di alimentazione, che è sempre necessaria per il funzionamento dei dispositivi attivi. L’amplificazione di potenza è di fondamentale importanza non solo nel trattamento dei segnali analogici, ma anche in campo digitale, perché, insieme con la proprietà di unidirezionalità, consente la realizzazione di interruttori comandati che controllano correnti più grandi di quella di comando e quindi di reti che non danno luogo a degrado dei livelli logici.

Una descrizione molto utile del comportamento di un generatore comandato si ottiene tramite la rappresentazione grafica delle caratteristiche di ingresso e di uscita, sotto la forma di famiglie di curve. Le caratteristiche di ingresso vengono tracciate sul piano v_1 - i_1 e rappresentano il legame tra tali grandezze in funzione di una delle grandezze di uscita. Nel caso particolarmente semplice considerato non ha

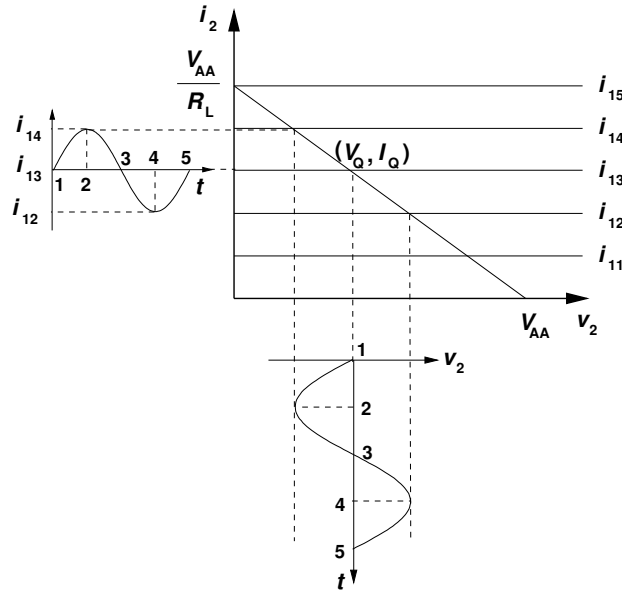
significato tracciare le caratteristiche di ingresso, dato che v_1 è sempre nulla. Per quanto riguarda invece le caratteristiche di uscita, queste vengono rappresentate sul piano v_2-i_2 , in funzione di una delle due grandezze di ingresso, in questo caso la i_1 (per un numero discreto di valori della i_1). Per il generatore di corrente ideale controllato in corrente, le caratteristiche di uscita sono molto semplici: in corrispondenza di ogni valore della i_1 si ha una retta parallela all'asse delle ascisse e da esso distante Ai_1 , visto che la corrente di uscita è proporzionale a i_1 e indipendente dal valore di v_2 .



Se prendiamo in considerazione un circuito come quello di seguito rappresentato, in cui sull'uscita è collegato un generatore di tensione V_{AA} tramite un resistore R_L , possiamo indicare la relazione tra v_2 e i_2 imposta dalla rete esterna per mezzo di una retta di carico, che può essere riportata sul piano delle caratteristiche. La presenza del generatore V_{AA} può sembrare a questo stadio artificiosa, dato che comporta semplicemente l'aggiunta di un valor medio alla tensione di uscita. Vedremo più avanti come, nel caso di un dispositivo reale, questa sia invece indispensabile per fornire l'energia necessaria al suo funzionamento.



Cerchiamo ora di determinare l'andamento della tensione di uscita in funzione della corrente i_1 . Al variare di i_1 il punto di lavoro si sposta lungo la retta di carico: rappresentiamo la corrente i_1 in funzione del tempo a lato della caratteristica di uscita e, per ogni valore di i_1 , andiamo a individuare la corrispondente caratteristica I_2-V_2 e la relativa intersezione con la retta di carico. In questo modo otteniamo il valore all'istante corrispondente della v_2 e lo possiamo riportare in un grafico posto sotto alle caratteristiche di uscita. Notiamo che la forma d'onda d'uscita è una replica amplificata, invertita di fase e traslata di quella di ingresso.



4.2 Il principio di funzionamento del BJT

Il transistorore bipolare a giunzione consiste di due giunzioni pn poste una di seguito all'altra e orientate in senso inverso: si tratta quindi di tre regioni consecutive, una p , una n e una p nel caso di un dispositivo $pn p$ e una n , una p e una n nel caso dei dispositivi npn . La caratteristica fondamentale che distingue il transistorore da due giunzioni pn in serie “back-to-back” è rappresentata dallo spessore estremamente ridotto della zona centrale (non rappresentato correttamente in scala nel disegno sottostante), che, come vedremo, dà luogo all'interazione tra le due giunzioni, alla base dell'“effetto transistor”.

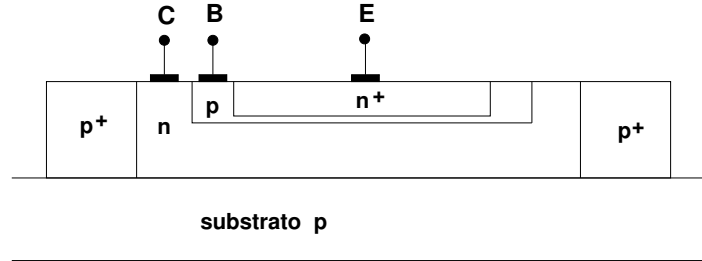


L'elettrodo centrale viene definito base, mentre gli altri due sono denominati emettitore e collettore. Il drogaggio delle due regioni esterne non è uguale: in un transistorore la zona di emettitore è significamente più drogata di quelle di base e di collettore; si indica infatti con p^+ nei transistori $pn p$ e con n^+ nei transistori npn . Nella trattazione del modello fisico del transistorore tutte le correnti vengono di solito scelte con verso entrante. I transistori vengono indicati negli schemi circuitali con i simboli sottoindicati, che differiscono, tra $pn p$ e npn , solo per il verso della freccia dell'emettitore.



Dal punto di vista costruttivo i BJT non vengono realizzati connettendo insieme pezzi separati di silicio con diverso drogaggio, anche perché in tal modo sarebbe ben difficile riuscire a ottenere gli spessori di base estremamente ridotti che sono richiesti per il corretto funzionamento. Un processo tecnologico standard per la fabbricazione dei

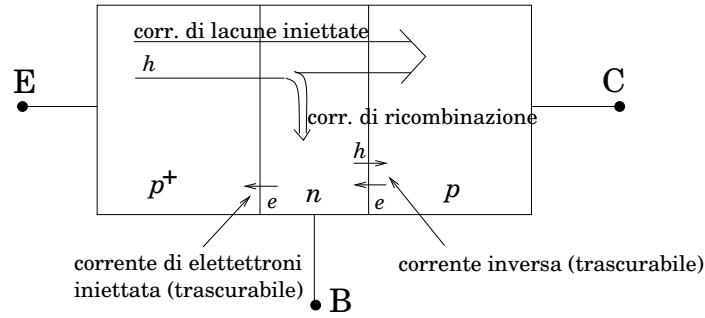
transistori è quello planare, illustrato di seguito e basato sulla successiva realizzazione di strati con diverso drogaggio. L'esempio preso in considerazione è quello di un transistor $n p n$.



Si parte da un substrato di silicio monocristallino di tipo p sul quale viene ottenuto uno strato n che rappresenta il collettore. Si realizza poi uno strato estremamente sottile di tipo p , che costituisce la base e infine viene ottenuta una regione n^+ che rappresenta l'emettitore. Lateralmente vengono realizzate delle regioni p^+ che hanno funzione di isolamento dai transistori adiacenti. La definizione laterale delle varie regioni viene ottenuta tramite tecniche litografiche. Con opportune procedure si ottengono anche dei contatti elettrici che raggiungono le tre regioni del transistor e rappresentano la connessione con il resto del circuito.

Passiamo ora a intraprendere uno studio più dettagliato del funzionamento del BJT. Nella nostra trattazione supporremo che, eccetto che nelle zone di svuotamento, la concentrazione di portatori nelle tre regioni sia sufficiente per poter considerare nulla la caduta di tensione attraverso ciascuna delle regioni stesse: le correnti che attraversano il transistor sono dunque tutte correnti di diffusione. Consideriamo la modalità di funzionamento più tipica, cioè quella nella quale la giunzione base-emettitore viene polarizzata direttamente, mentre quella base-collettore risulta polarizzata inversamente. In conseguenza della polarizzazione diretta tra base ed emettitore, una corrente di lacune viene iniettata dall'emettitore nella base, mentre una corrente di elettroni passa dalla base all'emettitore. Dato che, come già sottolineato, l'emettitore è molto più drogato della base, la corrente di lacune iniettata in base sarà molto più grande di quella di elettroni iniettata dalla base nell'emettitore, che può essere trascurata. Se la regione di base fosse lunga, la corrente di lacune iniettata dall'emettitore, darebbe luogo a una concentrazione in eccesso di lacune, che decaderebbe esponenzialmente con la distanza dalla zona di svuotamento, in conseguenza della progressiva ricombinazione con gli elettroni. Essendo però la base corta rispetto alla lunghezza di ricombinazione per le lacune (la distanza media sulla quale una lacuna si ricombina), solo poche lacune riescono a ricombinarsi, mentre la maggior parte raggiunge la zona di svuotamento tra base e collettore, dove le lacune vengono trascinate verso il collettore dal campo elettrico favorevole. Una grossa parte della corrente di lacune iniettata nella base dall'emettitore raggiunge quindi il collettore, mentre solo una piccola frazione dà luogo a ricombinazione in base.

La corrente di base è quindi sostanzialmente costituita dal flusso di elettroni che danno luogo alla ricombinazione delle lacune. Essendo la frazione di lacune iniettate dall'emettitore che si ricombina in base molto piccola, la corrente di base è molto più piccola di quelle di emettitore e di collettore e a esse proporzionale. Una variazione percentuale della corrente di base si ripercuote dunque in modo proporzionale sulla corrente di collettore, dando luogo a un comportamento corrispondente a quello di un generatore di corrente controllato in corrente.

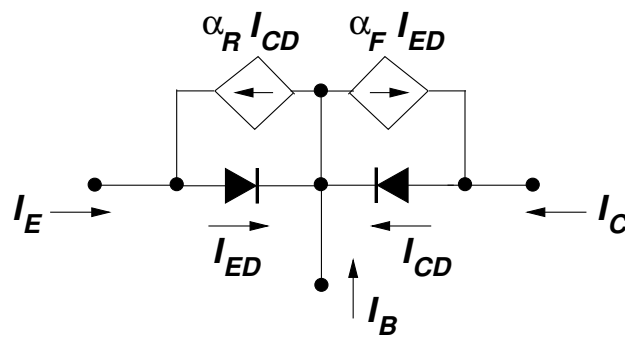


Le correnti all'interno di un transistor *pnp* possono dunque essere elencate nel modo seguente:

- corrente di lacune iniettata dall'emettitore nella base
- corrente di elettroni iniettata dalla base nell'emettitore (trascurabile)
- corrente di elettroni in base, corrispondente alla ricombinazione di una frazione costante delle lacune.
- corrente inversa attraverso la giunzione base-collettore (trascurabile)
- corrente di lacune inviata nel collettore, corrispondente alla frazione della corrente proveniente dall'emettitore che non si ricombina in base.

4.3 Le equazioni di Ebers-Moll

È possibile definire un modello matematico che descrive il comportamento per grandi segnali del transistor BJT, includendo tutte le componenti di corrente sopra menzionate. Tale modello, definito modello di Ebers-Moll dai nomi di coloro che lo proposero, è necessariamente non lineare e rappresenta il transistor come costituito da due giunzioni *pn* contrapposte con l'aggiunta di generatori di corrente comandati, che descrivono la porzione della corrente di emettitore trasferita al collettore oppure quella della corrente di collettore trasferita all'emettitore, nel caso che il transistor venga fatto funzionare con la giunzione base-emettitore in polarizzazione inversa e quella base-collettore in polarizzazione diretta.



Dallo schema sopra rappresentato, valido per un transistor *pnp*, otteniamo le cosiddette equazioni di Ebers-Moll:

$$\begin{cases} I_E = I_{ED} - \alpha_R I_{CD} = I_{ES} \left(e^{\frac{V_{EB}}{V_T}} - 1 \right) - \alpha_R I_{CS} \left(e^{\frac{V_{CB}}{V_T}} - 1 \right) \\ I_C = I_{CD} - \alpha_F I_{ED} = -\alpha_F I_{ES} \left(e^{\frac{V_{EB}}{V_T}} - 1 \right) + I_{CS} \left(e^{\frac{V_{CB}}{V_T}} - 1 \right) \end{cases},$$

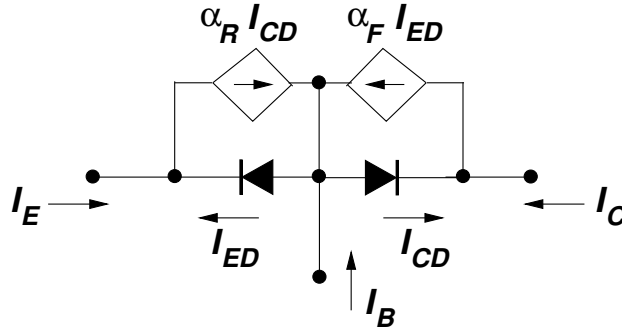
dove α_F rappresenta la frazione della corrente di emettitore che raggiunge il collettore, mentre α_R rappresenta la frazione della corrente di collettore (in caso di polarizzazione diretta della giunzione base-collettore) che viene trasferita all'emettitore. Sia α_R sia α_F risultano minori dell'unità (rappresentando una frazione di una quantità totale) e sono legati tra loro dalle cosiddette condizioni di reciprocità per il transistor:

$$\alpha_F I_{ES} = \alpha_R I_{CS}.$$

Per i transistori normalmente realizzati, nei quali la regione di emettitore è molto più drogata delle altre due regioni, si ha $0.98 \leq \alpha_F \leq 0.998$ e $0.4 \leq \alpha_R \leq 0.8$. Le correnti I_{ES} e I_{CS} sono dell'ordine di 10^{-15} A nei transistori al silicio. La corrente I_B può essere immediatamente ricavata dalle equazioni di Ebers-Moll applicando la legge di Kirchhoff ai nodi:

$$I_B = -(I_E + I_C).$$

Nel caso di un transistor *npn* il circuito equivalente di Ebers-Moll deve essere modificato, in modo da tenere in considerazione il diverso segno delle tensioni e delle correnti:



Le equazioni di Ebers-Moll per un transistor *npn* diventano:

$$\begin{cases} I_E = -I_{ES} \left(e^{\frac{-V_{EB}}{V_T}} - 1 \right) + \alpha_R I_{CS} \left(e^{\frac{-V_{CB}}{V_T}} - 1 \right) \\ I_C = \alpha_F I_{ES} \left(e^{\frac{-V_{EB}}{V_T}} - 1 \right) - I_{CS} \left(e^{\frac{-V_{CB}}{V_T}} - 1 \right) \end{cases}.$$

Sulla base del modello di Ebers-Moll possiamo definire le quattro zone di funzionamento possibili per il transistor:

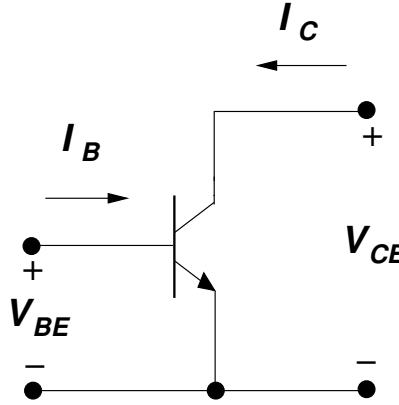
- Zona attiva diretta, con la giunzione base-emettitore polarizzata direttamente e quella collettore-base polarizzata inversamente. Questa è la condizione di funzionamento più utilizzata e quella in cui si hanno buone prestazioni dal punto di vista dell'amplificazione.
- Zona attiva inversa, con la giunzione base-emettitore polarizzata inversamente e quella collettore-base polarizzata direttamente. Dato che il transistor non è simmetrico, le prestazioni sono in questo caso molto degradate rispetto a quelle in zona attiva diretta e pertanto i transistori non vengono quasi mai utilizzati in zona attiva inversa.
- Zona di interdizione, con ambedue le giunzioni polarizzate inversamente. In questo caso le correnti che attraversano il transistor sono estremamente piccole e quindi lo stesso può considerarsi come un circuito aperto. Questo modo di

funzionamento può essere utilizzato in applicazioni di tipo digitale per emulare un interruttore aperto.

- d) Zona di saturazione, con ambedue le giunzioni polarizzate direttamente. In questo caso la caduta di tensione tra collettore ed emettitore è molto piccola (pochi decimi di volt) e il comportamento del transistor assomiglia a quello di un corto circuito. Questo modo di funzionamento può essere impiegato in circuiti digitali per emulare un interruttore chiuso.

4.4 Caratteristiche a emettitore comune

Ci occuperemo ora della determinazione delle caratteristiche $I - V$ del transistor *npn* quando questo sia connesso nella configurazione a emettitore comune, vale a dire con l'emettitore a comune tra ingresso e uscita, la base sulla porta di ingresso e il collettore su quella di uscita. In tale configurazione le grandezze di ingresso sono la V_{BE} e la I_B , mentre quelle di uscita sono la V_{CE} e la I_C .



Le caratteristiche di uscita a emettitore comune descriveranno quindi la relazione tra I_C e V_{CE} in funzione di I_B e saranno costituite da una famiglia di curve sul piano $V_{CE} - I_C$, ciascuna per un diverso valore della corrente I_B .

In zona di funzionamento attiva diretta, osservando le equazioni di Ebers-Moll notiamo che la corrente di emettitore I_E è sostanzialmente pari a $-I_{ES} \exp(V_{BE}/V_T)$, data la condizione di polarizzazione diretta della giunzione BE , che rende trascurabile l'unità rispetto all'esponenziale, e di polarizzazione inversa della giunzione BC , che rende trascurabile il termine contenente il contributo della I_{CS} . La corrente di collettore I_C è invece sostanzialmente corrispondente al termine dovuto all'iniezione di elettroni dall'emettitore in base, $\alpha_F I_{ES} \exp(V_{BE}/V_T)$, quindi possiamo anche scrivere che $I_C = -\alpha_F I_E$.

L'espressione della corrente di base risulta quindi, dall'applicazione del principio di Kirchhoff ai nodi,

$$I_B = -(I_E + I_C) = -(1 - \alpha_F)I_E.$$

Sostituendo nella precedente equazione a I_E la sua espressione in funzione di I_C otteniamo

$$I_B = (1 - \alpha_F) \frac{I_C}{\alpha_F},$$

quindi

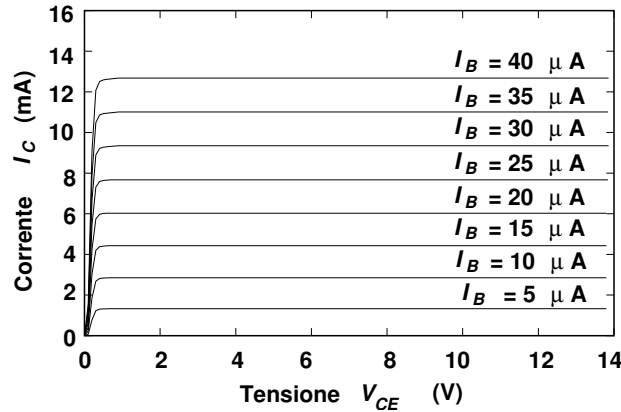
$$I_C = \frac{\alpha_F}{1 - \alpha_F} I_B = \beta_F I_B,$$

dove β_F è definito come

$$\beta_F = \frac{\alpha_F}{1 - \alpha_F}$$

ed è di solito indicato dai costruttori come h_{FE} (si noti che il pedice FE è maiuscolo e quindi questo parametro non deve essere confuso con h_{fe} che sarà introdotto più avanti e avrà tutt'altro significato).

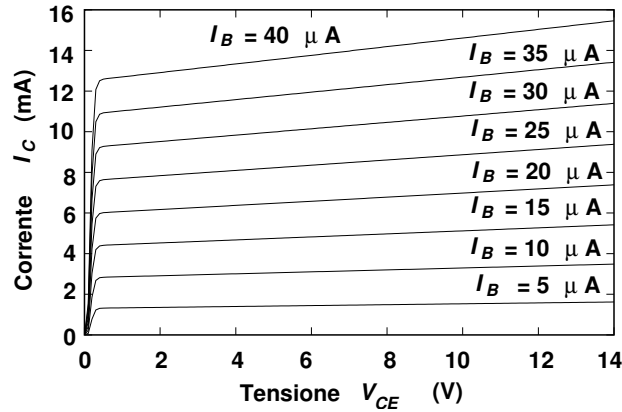
Proviamo a costruire il grafico delle caratteristiche di uscita, nell'ipotesi di β_F costante. In realtà β_F ha una certa dipendenza da I_C , per cui le caratteristiche non sono equispaziate.



È importante ricordare che, poiché $V_{CE} = V_{CB} + V_{BE}$, se V_{BE} viene mantenuta al valore V_γ dal circuito di polarizzazione di base, quando V_{CE} scende al di sotto di V_γ , la giunzione CB comincia a essere polarizzata direttamente, per cui usciamo dalla zona attiva diretta propriamente detta. Al decrescere di V_{CE} al di sotto di V_γ inizialmente non viene osservata una significativa variazione di comportamento della corrente di collettore (poiché la giunzione CB , pur essendo già polarizzata direttamente, non conduce ancora in modo significativo), ma quando V_{CE} risulta minore di 0.3 - 0.4 V, la conduzione della giunzione BC fa sì che il relativo termine nell'equazione di Ebers-Moll non sia più trascurabile (siamo quindi decisamente in zona di saturazione) e la corrente di collettore scende al di sotto del valore $-\alpha_F I_E$. Per valori ancor minori di V_{CE} , quando si raggiunge il cosiddetto valore V_{CEsat} (di solito assunto pari a 0.2 V, ma in realtà dipendente dal tipo di transistor e variabile tra 0.1 e 0.25 V), le diverse caratteristiche collapsano l'una sull'altra e si perde quindi il controllo della corrente di collettore da parte di quella di base. In saturazione il transistor si comporta dunque come l'equivalente di un interruttore chiuso connesso tra emettitore e collettore, visto il piccolissimo valore di tensione (V_{CEsat}) che sussiste tra questi due elettrodi.

Nelle caratteristiche di un transistor BJT reale, come quelle riportate di seguito, notiamo subito un'importante differenza rispetto a quanto visto finora: la corrente di collettore in zona attiva diretta non rimane costante al crescere della V_{CE} , ma subisce un incremento, che è la conseguenza del cosiddetto effetto Early. Da un punto di vista grafico tale incremento risulta visibile nella forma di un'inclinazione verso l'alto delle caratteristiche, per tutta la zona attiva diretta.

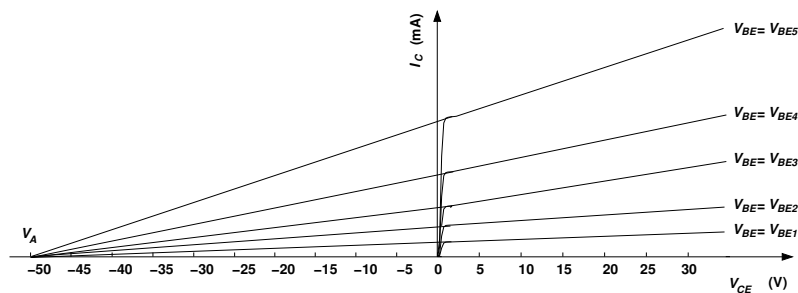
Tale effetto è il risultato dell'allargamento della regione di svuotamento della giunzione base-collettore quando V_{CE} (e conseguentemente la polarizzazione inversa di tale giunzione, V_{CB}) viene aumentata. A causa di tale allargamento il tratto della base all'interno del quale può effettivamente avvenire la ricombinazione dei portatori



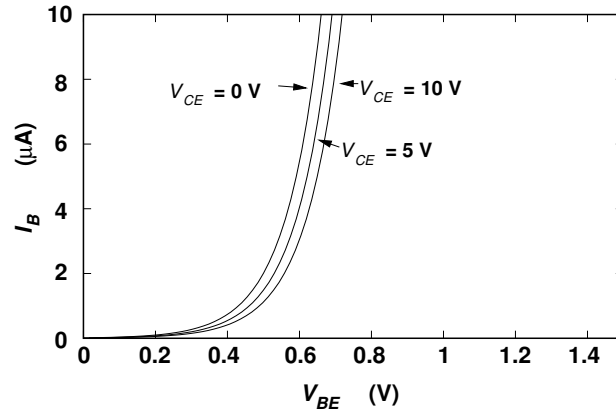
iniettati dall'emettitore (lunghezza efficace di base) risulta ridotto. Pertanto una percentuale più piccola di portatori (elettroni in un transistor nnp) si ricombina in base e una più grande riesce a raggiungere il collettore e a formare una I_C che è pertanto maggiore.

Dunque, per valori di V_{CB} (e quindi di V_{CE}) più elevati la corrente di collettore aumenta a parità di corrente di base e quindi ciascuna delle curve corrispondenti alle caratteristiche di uscita risulta inclinata verso l'alto.

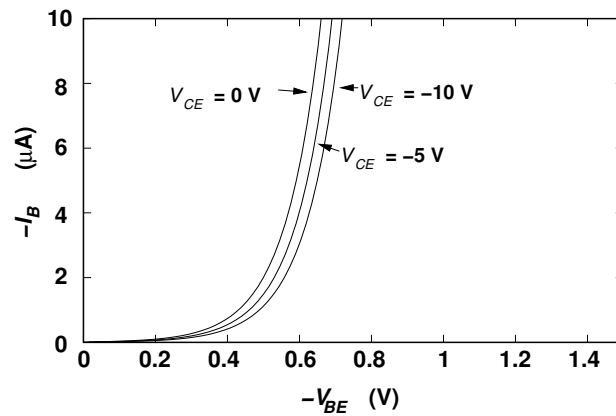
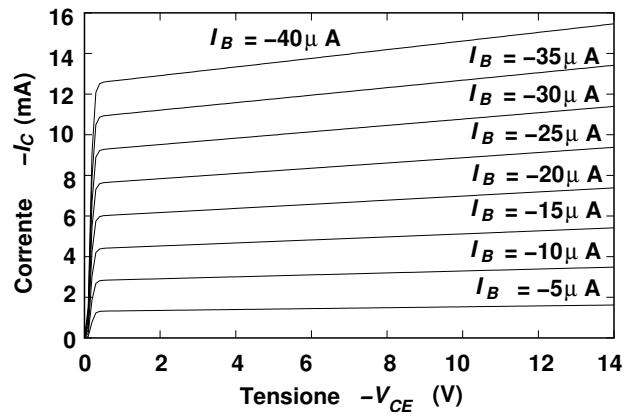
Se tracciamo le caratteristiche di uscita in modo inconsueto, scegliendo come parametro la tensione V_{BE} invece della I_B , osserviamo che anch'esse mostrano l'effetto Early e che prolungando verso sinistra con delle semirette il tratto rettilineo delle caratteristiche si ottiene un'unica intersezione comune con l'asse delle ascisse, in corrispondenza della cosiddetta "tensione di Early" (V_A), in genere compresa tra -50 e -100 V. In realtà, se come effettivamente avviene, l'effetto della tensione collettore-emettitore sulle caratteristiche di ingresso (di cui parleremo di seguito) è estremamente piccolo, c'è una relazione pressoché indipendente da V_{CE} tra V_{BE} e I_B , per cui anche i prolungamenti delle caratteristiche di uscita a I_B costante (che in questo caso non sono altro che una particolare scelta delle caratteristiche a V_{BE} costante) si intersecano nello stesso punto.



Le caratteristiche di ingresso a emettitore comune forniscono la relazione tra I_B e V_{BE} in funzione della V_{CE} . Esse rappresentano il comportamento della giunzione base-emettitore e hanno quindi un andamento esponenziale simile a quello di un diodo; risentono inoltre dell'effetto Early. In particolare, se si incrementa V_{CE} , la lunghezza di base efficace diminuisce, causando una minore ricombinazione e, di conseguenza, una diminuzione della corrente di base. Nella figura seguente sono rappresentate le caratteristiche di ingresso a emettitore comune per un transistor nnp ed è possibile vedere chiaramente il risultato dell'effetto Early, accentuato per renderlo più evidente.



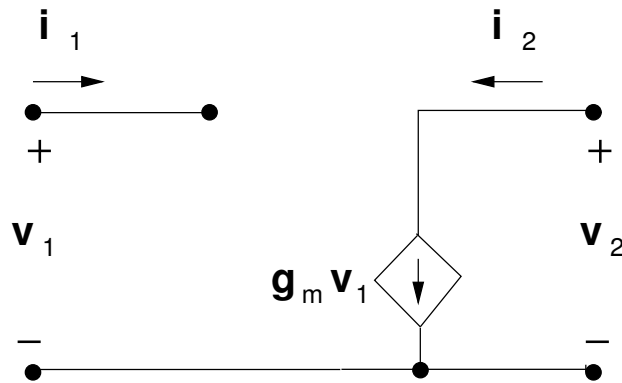
Le caratteristiche di ingresso e di uscita dei transistori *pn*p sono del tutto analoghe (con gli opportuni cambiamenti di segno) a quelle degli *np*n. Tali caratteristiche, riportate di seguito, si ottengono da quelle presentate per i transistori *np*n sostituendo I_B con $-I_B$, V_{BE} con $-V_{BE}$, V_{CE} con $-V_{CE}$, I_E con $-I_E$ e I_C con $-I_C$.



5. I transistori a effetto di campo

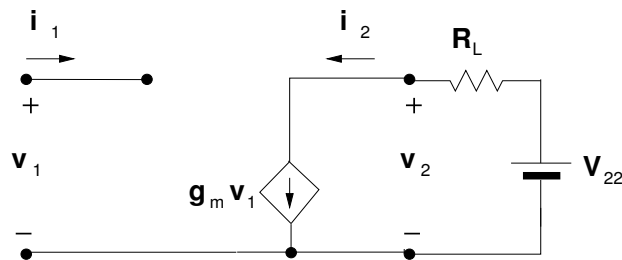
5.1 Generatore di corrente controllato in tensione

Nei transistori a effetto di campo, come suggerisce il nome stesso, il flusso di corrente viene controllato tramite un campo elettrico e, quindi, attraverso il valore della tensione applicata a un opportuno elettrodo di comando. Pertanto, così come nel caso dei BJT avevamo introdotto una rappresentazione idealizzata del funzionamento per mezzo di un generatore di corrente controllato in corrente, per lo studio dei transistori a effetto di campo introduciamo il concetto di generatore ideale di corrente controllato in tensione, la cui rappresentazione circuitale è indicata in figura.



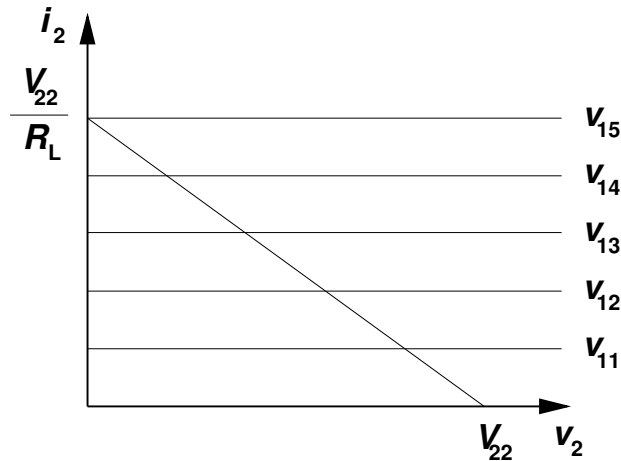
La quantità g_m rappresenta il rapporto tra la corrente di uscita e la tensione in ingresso. Dimensionalmente è quindi l'inverso di una resistenza ed è espressa in A/V o, più comunemente, in mA/V.

Possiamo anche in questo caso rappresentare le caratteristiche di uscita, sul piano v_2 - i_2 , ottenendo una famiglia di curve in funzione del parametro costituito dalla tensione di ingresso. Tali curve altro non sono che rette parallele all'asse delle ascisse, poiché la corrente in uscita non dipende dalla tensione v_2 , ma soltanto dalla v_1 . Per questo tipo di generatore comandato possiamo pensare a una configurazione in cui sono presenti un generatore di alimentazione esterno V_{22} e una resistenza di carico R_L , come quella rappresentata nella figura sottostante.



È possibile tracciare sul piano delle caratteristiche di uscita una retta di carico, la cui pendenza corrisponde a $-1/R_L$ e che interseca l'asse delle ascisse in corrispondenza di V_{22} , come indicato sulle caratteristiche di seguito rappresentate.

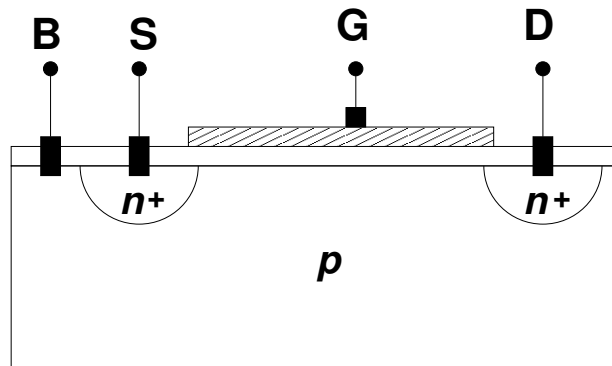
Al variare della tensione v_1 applicata in ingresso, il punto di lavoro si sposterà quindi lungo la retta di carico, dando luogo a una variazione della tensione di uscita v_2 che rappresenta una replica amplificata e invertita di fase della tensione di ingresso. Se le caratteristiche fossero equispaziate come quelle rappresentate in figura, la tensione di uscita presenterebbe una relazione perfettamente lineare con quella di ingresso;



vedremo che in realtà i dispositivi a effetto di campo sono caratterizzati da una dipendenza quadratica della corrente di uscita dalla tensione di ingresso, per cui il loro comportamento può considerarsi lineare solo per piccoli segnali.

5.2 Il transistor MOS

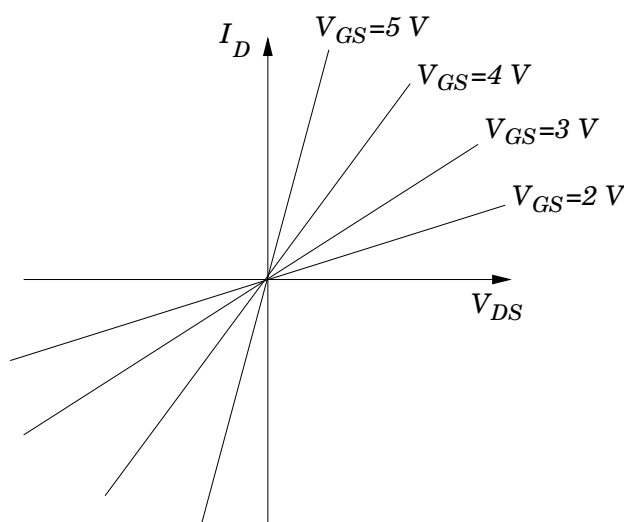
L'idea alla base del funzionamento del transistor MOS (Metal-Oxide-Semiconductor) è abbastanza semplice: si crea uno strato di cariche mobili in prossimità della superficie di un semiconduttore, tramite l'applicazione di un campo elettrico per mezzo di un elettrodo di gate metallico, che "attira" le cariche verso la superficie stessa. Una descrizione appropriata del funzionamento del transistor MOS richiederebbe concetti avanzati di meccanica quantistica che esulano da questa trattazione, quindi sfrutteremo una descrizione intuitiva, ancorché non rigorosa. La struttura di un tipico transistor MOS a canale n è rappresentata nella figura seguente: in un blocco di silicio p vengono realizzate, a una certa distanza tra loro, due diffusioni di tipo n^+ che rappresentano gli elettrodi di source e di drain; sopra la superficie del silicio viene ottenuto un sottile strato di ossido di silicio (dell'ordine della decina di nanometri o, più recentemente, dei nanometri) che funge da isolante; al di sopra dello strato isolante e tra source e drain viene realizzato uno strato metallico (o più recentemente di silicio policristallino fortemente drogato, in modo da renderlo conduttore) che funge da gate.



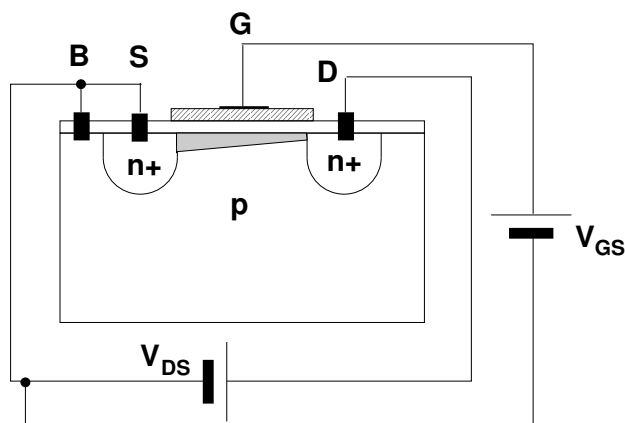
Se si applica una tensione positiva al gate rispetto al silicio p sottostante, vengono richiamati elettroni in prossimità della superficie, cosicché si ha, in un sottilissimo strato subito sotto la superficie stessa, un fenomeno di inversione: il silicio diventa

localmente di tipo n , dando quindi luogo alla formazione di un canale conduttore che mette in connessione source e drain. Il transistor MOS ha perciò 4 elettrodi: source, drain, gate e il quarto, denominato bulk, che corrisponde al silicio di substrato, quello nel quale si forma il canale. Nei dispositivi MOS non integrati l'elettrodo di bulk è di solito internamente collegato al source.

Il canale di portatori minoritari comincia a formarsi alla superficie quando la tensione tra gate e bulk supera un valore di soglia indicato con V_T . Poiché nei dispositivi discreti il bulk è di solito collegato con il source (come sopra accennato), la tensione tra gate e bulk corrisponde con la V_{GS} e come tale la indicheremo nel seguito. È importante sottolineare che la tensione V_T per i MOS non ha nulla a che fare con la V_T che abbiamo considerato nello studio dei BJT (pari a kT/q): la coincidenza dei nomi è del tutto fortuita. Per valori piccoli (inferiori a qualche decimo di volt) della V_{DS} , il canale del transistor MOS ha un andamento uniforme tra source e drain, con uno spessore, e conseguentemente una conducibilità, che cresceranno al crescere della V_{GS} . In questa condizione il MOS può quindi essere utilizzato come una resistenza variabile, controllata tramite la tensione V_{GS} , dando luogo a caratteristiche lineari sul piano $V_{DS} - I_D$.

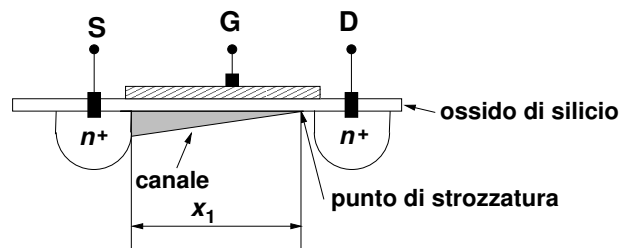


Se V_{DS} è maggiore di qualche decimo di volt, il comportamento del canale non è più quello di una resistenza lineare, poiché l'ampiezza verticale dello stesso verrà a dipendere dalla posizione considerata tra source e drain, come rappresentato nella figura seguente.



L'effettiva tensione tra gate e substrato dipende dalla posizione lungo il canale: sarà pari a V_{GS} in corrispondenza del source e diminuirà avvicinandosi al drain. Se $V_{GS} - V_{DS} < V_T$ esisterà un punto in cui la tensione tra gate e substrato risulta pari a V_T (la tensione di soglia per la formazione del canale) e oltre tale punto risulta inferiore a V_T , quindi il canale scompare o, come si suol dire, viene strozzato. Per $V_{GS} - V_{DS} = V_T$ la strozzatura si trova in corrispondenza dell'estremità di drain; all'aumentare di V_{DS} si sposta verso sinistra, ma di una quantità piccola relativamente alla lunghezza del canale, almeno per i valori di V_{DS} normalmente utilizzati. Quindi si ha un tratto di canale di lunghezza x_1 pari quasi alla distanza tra drain e source, ai capi del quale, una volta superata la V_{DS} per cui $V_{GS} - V_{DS} < V_T$, esiste una tensione pari a $V_{GS} - V_T$, sostanzialmente indipendente da ulteriori variazioni di V_{DS} . La corrente che lo attraversa è quindi anch'essa pressoché indipendente da V_{DS} e raggiunge la diffusione di drain attraverso la zona strozzata. Sul piano delle caratteristiche di uscita questo corrisponde ad avere caratteristiche pressoché orizzontali al di sopra del valore di V_{DS} per cui avviene lo strozzamento. Questa zona delle caratteristiche viene definita “zona di saturazione”; si noti che l'uso del termine “saturazione” in questo caso non ha nulla a che fare con quello che si ha nel caso del transistor BJT. La saturazione nel caso del BJT indica una condizione di funzionamento in cui tutte e due le giunzioni sono polarizzate direttamente e la corrente di base non ha quasi più alcuna influenza su quella di collettore; nel caso del transistor MOS la saturazione corrisponde alla condizione di funzionamento attivo che consente di ottenere un'amplificazione.

Vediamo ora una rappresentazione grafica di quanto è stato fin qui descritto.



All'aumentare di V_{DS} , la distanza x_1 si riduce, anche se di poco, e questo fatto dà luogo all'inclinazione delle caratteristiche di uscita (ai capi di un canale di lunghezza ridotta e quindi di resistenza ridotta è presente la stessa tensione V' , pressoché pari a $V_{GS} - V_T$). Dato che tale riduzione è piuttosto piccola e che la tensione V' è praticamente costante, la corrente I_{DS} si può considerare, almeno in prima approssimazione, indipendente da V_{DS} nella zona di saturazione.

In condizioni di saturazione la corrente è data da

$$I_D = k \left(\frac{W}{L} \right) (V_{GS} - V_T)^2,$$

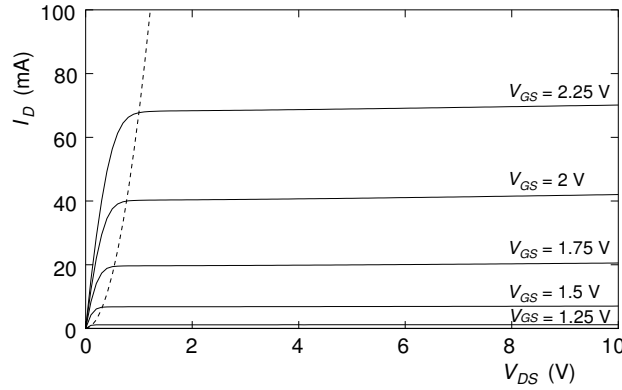
dove k rappresenta un parametro di processo, W è la larghezza in direzione perpendicolare al disegno e L è la lunghezza del canale, cioè la distanza tra drain e source. Il parametro di processo k ha le dimensioni di una corrente divisa per una tensione al quadrato e può essere espresso come

$$k = \mu_n \frac{C_{ox}}{2},$$

dove C_{ox} è la capacità dell'ossido per unità di superficie ($C_{ox} = \varepsilon/T_{ox}$), con ε pari alla costante dielettrica dell'ossido e T_{ox} pari allo spessore dell'ossido) e μ_n è la mobilità dei portatori interessati alla conduzione tra source e drain, in questo caso elettroni. Tale espressione è valida non solo per $V_{DS} = V_{GS} - V_T$, ma anche per valori superiori di V_{DS} , se facciamo l'ipotesi di corrente in zona di saturazione indipendente da V_{DS} . Poiché all'entrata in saturazione $V_{DS} = V_{GS} - V_T$, possiamo ricavare facilmente l'espressione della curva che separa sul piano delle caratteristiche di uscita, la regione cosiddetta "triode" (per V_{DS} minore di quella di saturazione) e quella di saturazione: si tratta di una parabola descritta dall'equazione

$$I_D = k \left(\frac{W}{L} \right) V_{DS}^2.$$

Le caratteristiche di uscita così ottenute sono rappresentate nella figura seguente, nella quale la parabola appena citata è stata tracciata con una linea tratteggiata.



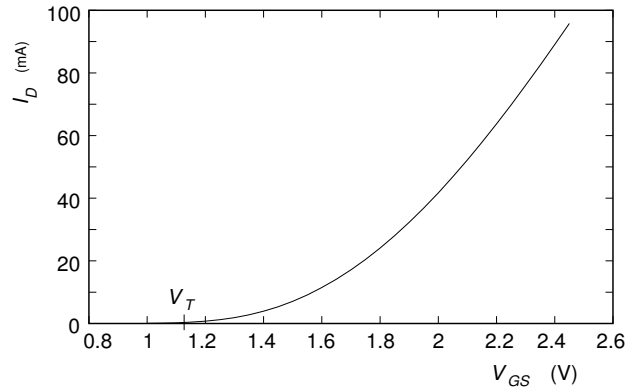
Se, nell'espressione della corrente in saturazione, vogliamo includere anche l'effetto di accorciamento del canale all'aumentare della V_{DS} , che dà luogo a una leggera inclinazione verso l'alto delle caratteristiche di uscita, possiamo usare l'espressione di tipo fenomenologico

$$I_D = k \left(\frac{W}{L} \right) (V_{GS} - V_T)^2 (1 + \lambda V_{DS}),$$

dove il parametro λ rappresenta l'inverso di una tensione, che viene di solito chiamata tensione di Early, in analogia a quanto accade nei BJT, anche se in questo caso il fenomeno fisico che porta all'inclinazione delle caratteristiche di uscita è completamente diverso da quello della riduzione della lunghezza efficace di base dei BJT.

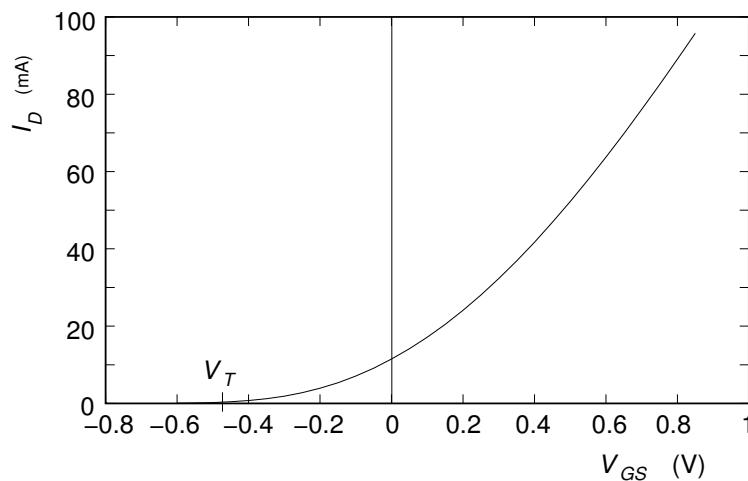
Sulla base delle espressioni finora derivate è anche possibile tracciare la transcaratteristica ingresso-uscita per un transistor MOS: si tratta di una parabola che inizia per $V_{GS} = V_T$. Di solito si trascura la dipendenza della transcaratteristica dalla V_{DS} e si traccia un'unica curva, indipendentemente dal valore di V_{DS} .

Per i transistori MOS esistono versioni complementari a canale p , nelle quali il substrato è di tipo n e le diffusioni di source e di drain sono di tipo p^+ . In tali transistori il trasporto tra source e drain è ottenuto tramite una corrente di lacune, che vengono indotte alla superficie tramite l'applicazione di una tensione negativa tra gate e bulk. Pertanto le caratteristiche di un transistor a canale p sono equivalenti a quelle di un transistor a canale n , purché si scambino i segni



di correnti e tensioni. I MOS a canale p hanno prestazioni inferiori, a parità di altre caratteristiche, rispetto ai corrispondenti transistori a canale n , in conseguenza della ridotta mobilità delle lacune rispetto a quella degli elettroni. Si può compensare peraltro tale ridotta mobilità con un aumento della larghezza di canale W , cosa che si fa molto frequentemente, dato che la disponibilità di transistori MOS complementari è alla base della tecnologia CMOS, che rappresenta, come vedremo più avanti, uno dei motori trainanti dell'attuale industria microelettronica.

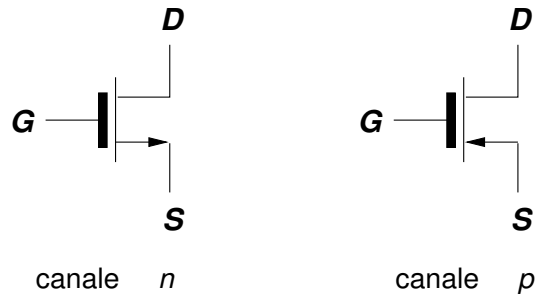
I transistori MOS finora visti sono di tipo ad arricchimento (o enhancement), poiché per tensione di gate nulla il canale non esiste e la sua formazione è una conseguenza dell'arricchimento di portatori del tipo opportuno operato tramite l'applicazione di una tensione di gate. Esistono anche transistori di tipo diverso, nei quali il canale è preesistente, ottenuto tramite impiantazione di droganti nello strato superficiale e lo si può far scomparire con l'applicazione di un'opportuna tensione di gate. Tali transistori si definiscono a svuotamento (o depletion). Nel caso del transistor depletion a canale n la tensione di soglia V_T risulta quindi negativa e la transcaratteristica ingresso-uscita risulta traslata all'indietro rispetto a quella dell'equivalente enhancement:



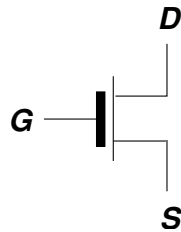
Le caratteristiche di uscita sono praticamente equivalenti a quelle del transistor enhancement, con l'unica variante dei valori di tensione V_{GS} corrispondenti a ciascuna di tali caratteristiche, che risultano traslati verso il basso.

Rovesciando i segni di correnti e tensioni possono facilmente ottenersi le caratteristiche dei transistori a canale p , sia enhancement sia depletion.

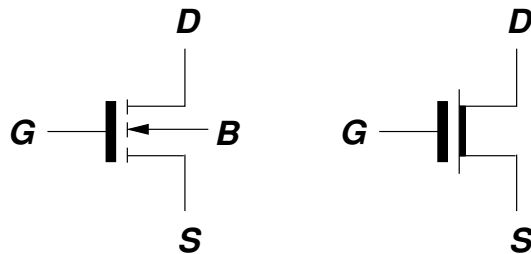
Esistono diversi simboli circuitali per i transistori MOS. Quelli più comunemente usati non danno informazioni sul tipo (ad arricchimento o a svuotamento), ma indicano soltanto se il canale è p o n :



Si utilizza talvolta anche un simbolo ancor più generale, valido sia per MOS a canale n sia per MOS a canale p :

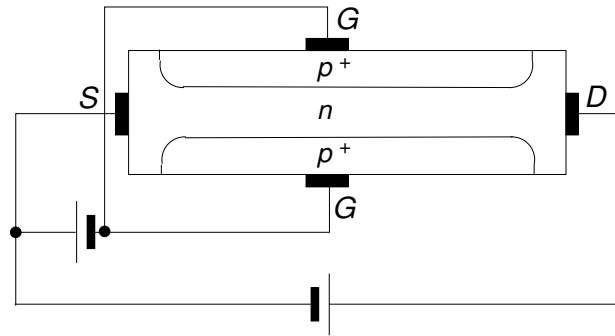


Esistono infine dei simboli, molto poco usati, che distinguono tra MOS ad arricchimento (a sinistra) e MOS a svuotamento (a destra):

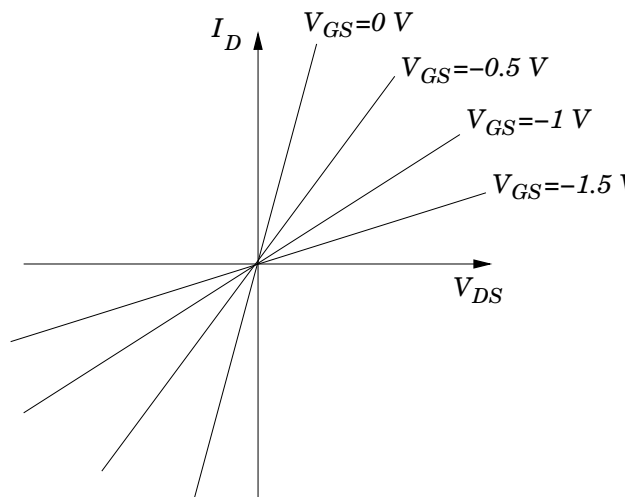


5.3 Il transistor JFET

Il transistor JFET (Junction Field Effect Transistor) è il primo transistor a effetto di campo che fu realizzato e contiene, come è chiaro dal nome, una giunzione p - n . Prendiamo dapprima in considerazione il transistor JFET a canale n , il cui funzionamento può essere compreso in base allo schema di principio riportato nella figura che segue.

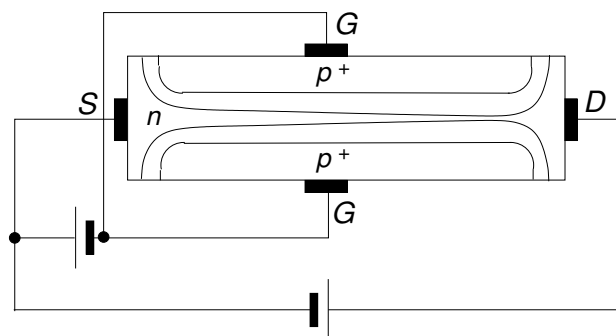


La corrente scorre tra l'elettrodo di source e quello di drain attraverso un pezzo di semiconduttore drogato n , nel quale sono realizzate due diffusioni di tipo p^+ , che costituiscono l'elettrodo di gate. Essendo le diffusioni di gate più drogate, la zona di svuotamento si estende prevalentemente nella regione n . Se supponiamo di applicare una tensione molto piccola tra drain e source e una tensione negativa V_{GS} tra gate e source, in modo da polarizzare inversamente la giunzione, la corrente scorrerà in un canale con dimensione verticale variabile al variare di V_{GS} , compreso tra i bordi delle due zone di svuotamento dovute alle diffusioni p^+ . Per piccoli valori della tensione V_{DS} applicata tra drain e source, il JFET si comporta quindi in modo simile a quanto già visto per il MOS, come una resistenza il cui valore può essere variato per mezzo della V_{GS} : in questo caso tanto più grande è in modulo il valore della V_{GS} tanto maggiore risulta il valore di resistenza presente tra drain e source. Tutto ciò può essere rappresentato tramite le caratteristiche di uscita nella regione per V_{DS} piccola, che sono sostanzialmente delle rette che confluiscono nell'origine, come illustrato nella figura seguente. La pendenza della caratteristica corrisponde all'inverso della resistenza e diminuisce all'aumentare in modulo della V_{GS} .



Per V_{GS} abbastanza negativa le due zone di svuotamento si congiungono e il canale scompare, per cui la corrente si annulla: il valore di V_{GS} per il quale questo avviene si definisce tensione di pinch-off e si indica con V_P . Per tensioni V_{GS} in modulo maggiori di V_P il transistor JFET si comporta come un interruttore aperto.

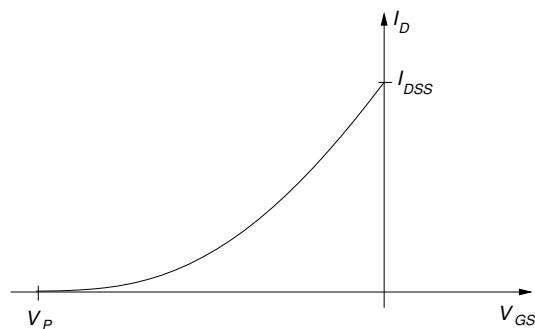
Quando si applica invece una V_{DS} non trascurabile (superiore a qualche decimo di volt), il comportamento del dispositivo non può più essere assimilato a quello di una resistenza lineare: per una V_{DS} significativa, la differenza di potenziale tra drain e canale risulta maggiore in modulo di quella tra source e canale, per cui le zone di svuotamento che definiscono il canale stesso assumono una forma asimmetrica, risultando più ampie in corrispondenza della regione di drain.



Al crescere di V_{DS} , il canale diviene sempre più stretto fino a risultare strozzato in corrispondenza del drain. In questa condizione il canale si estende dall'estremità di source fino al punto dove inizia la strozzatura. Si ha in tal caso un fenomeno di saturazione del tutto analogo a quello già visto per i transistori MOS. Si ha quindi un comportamento che approssima quello del generatore ideale di corrente controllato in tensione. Le caratteristiche corrispondenti a valori equispaziati della tensione V_{GS} non sono però equispaziate: si può dimostrare che in saturazione la corrente di drain I_D in un JFET ha una dipendenza quadratica dalla tensione V_{GS} , secondo la seguente equazione:

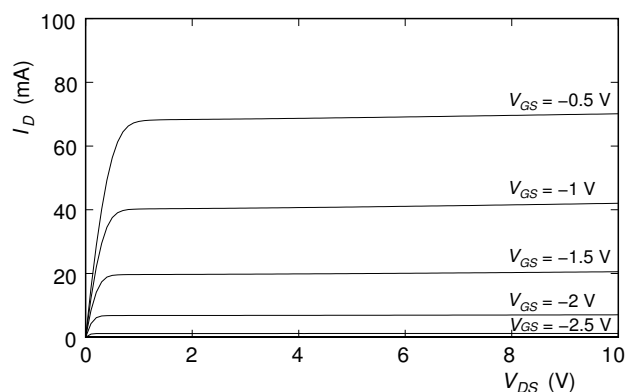
$$I_D = I_{DSS} \left(1 - \frac{V_{GS}}{V_P} \right)^2,$$

dove I_{DSS} è la corrente di drain che si ha per V_{GS} nulla. Tale relazione rappresenta la transcaratteristica ingresso-uscita e può essere riportata graficamente sul piano V_{GS} - I_D .

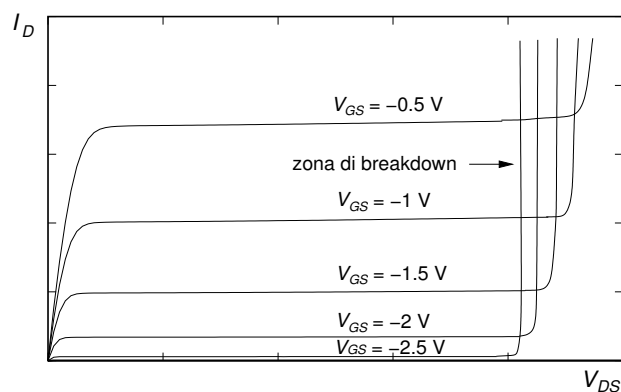


In prima approssimazione possiamo assumere che la transcaratteristica sia indipendente dal valore di V_{DS} (cosa che faremo in tutte le applicazioni numeriche), che

corrisponde ad assumere, come abbiamo fatto in precedenza, orizzontali le caratteristiche di uscita in saturazione. In realtà queste non sono perfettamente orizzontali, a causa del fatto che il punto in cui inizia la strozzatura si sposta leggermente verso il source all'aumentare della tensione V_{DS} , per cui la stessa tensione V_C (tensione tra il punto di strozzamento e il source) viene applicata a un canale di lunghezza minore, dando così luogo a una corrente di drain maggiore. L'andamento delle caratteristiche di uscita in zona di saturazione è perciò inclinato, in modo simile a quanto accade nei BJT a causa dell'effetto Early. Le caratteristiche di uscita di un JFET a canale n hanno perciò l'aspetto sotto rappresentato.



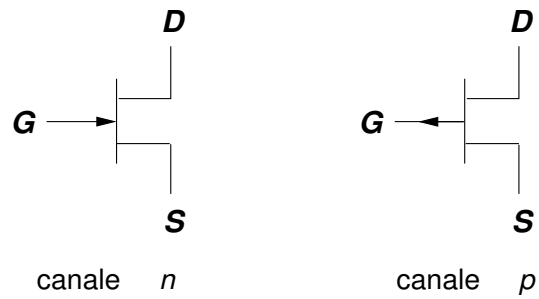
La giunzione tra gate e canale è polarizzata inversamente, quindi viene attraversata da una corrente molto piccola di valore compreso tra le decine di picoampere e qualche nanoampere, a seconda dell'area di giunzione e della temperatura. Per V_{DS} elevata la tensione inversa applicata sulla giunzione tra gate e canale può superare il valore di breakdown e si assiste in tal caso a un rapido aumento della corrente di drain. È da notare che il breakdown avviene per valori di V_{DS} tanto minori quanto più grande è, in modulo, la V_{GS} . Questo perché la tensione effettivamente presente ai capi della giunzione nell'area di drain risulta pari a $V_{DS} - V_{GS}$.



Finora abbiamo parlato di transistori JFET con canale n e gate realizzato con una diffusione p^+ ; è possibile avere anche una realizzazione di tipo complementare, con canale p e diffusioni di gate di tipo n^+ . Il funzionamento è del tutto analogo, con la sola differenza che tutte le tensioni e le correnti avranno segno invertito: la tensione V_{GS} dovrà essere positiva, mentre la V_{DS} sarà negativa. Le prestazioni dei transistori JFET a canale p differiscono, a parità di drogaggi e di caratteristiche geometriche, da quelle dei JFET a canale n , a causa della diversa mobilità delle lacune rispetto

agli elettroni. In genere le prestazioni dei JFET a canale n sono dunque migliori, ed è questo il motivo per cui essi risultano molto più utilizzati.

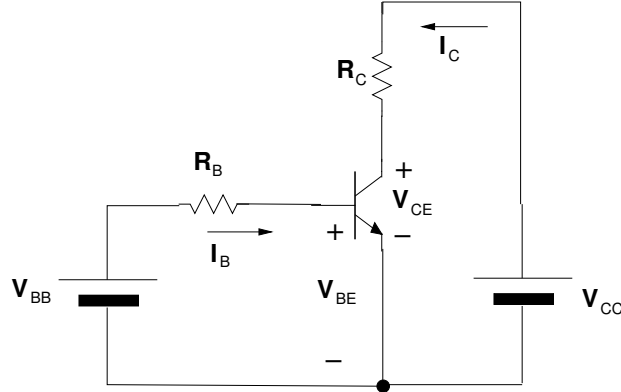
Il simbolo circuitale del JFET è sotto rappresentato, sia per il tipo a canale n sia per quello a canale p .



6. Reti di polarizzazione

6.1 Reti di polarizzazione per transistori BJT

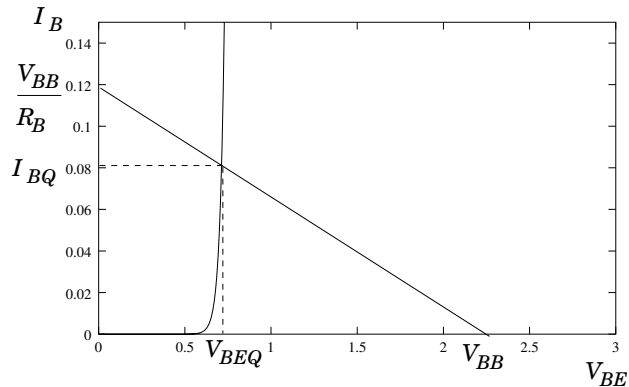
La rete di polarizzazione più semplice possibile da un punto di vista concettuale è quella di seguito riportata:



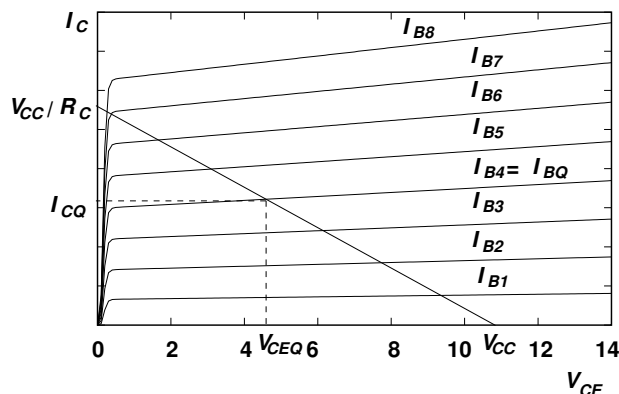
Le quantità incognite sono V_{CE} , I_C , V_{BE} , I_B . Abbiamo quindi la necessità di quattro equazioni che legano tra loro queste quantità. Tali equazioni possono essere ricavate dalle relazioni tra le varie grandezze imposte dal circuito esterno e dalle relazioni imposte dal transistor stesso.

$$\begin{aligned} V_{BE} &= g(I_B, V_{CE}) \\ I_C &= f(I_B, V_{CE}) \\ V_{BB} &= R_B I_B + V_{BE} \\ V_{CC} &= R_C I_C + V_{CE}, \end{aligned}$$

dove le prime due equazioni rappresentano le caratteristiche del transistor (rispettivamente di ingresso e di uscita) e le altre sono semplicemente le equazioni alle due maglie dalle quali il circuito è costituito. Se le caratteristiche di ingresso corrispondenti a diversi valori di V_{CE} non differiscono significativamente tra loro, possiamo determinare il valore di riposo I_{BQ} della I_B tracciando sul grafico delle caratteristiche di ingresso stesse la retta di carico definita dalla terza equazione.



A questo punto possiamo costruire la retta di carico corrispondente alla Eq. (4) sul piano delle caratteristiche di uscita e individuare i valori di riposo della I_{CQ} della I_C e V_{CEQ} della V_{CE} dall'intersezione di tale retta con la caratteristica relativa alla I_{BQ} prima determinata, completando così la determinazione del punto di lavoro del transistor:

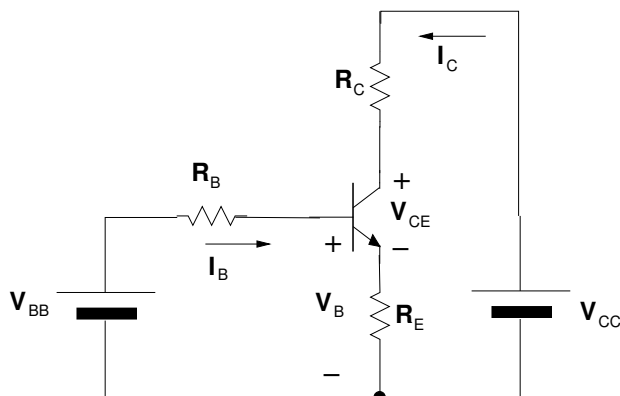


La procedura seguita può essere semplificata notando che in zona attiva diretta la V_{BE} non si discosta da V_γ più di 0.1-0.15 V, per cui possiamo assumere, senza commettere errori significativi, $V_{BE} \simeq V_\gamma$. In tal caso non è più necessario ricorrere alle caratteristiche di ingresso. Dalla equazione delle cadute di tensione sulla maglia di ingresso otteniamo direttamente

$$I_B = \frac{V_{BB} - V_\gamma}{R_B},$$

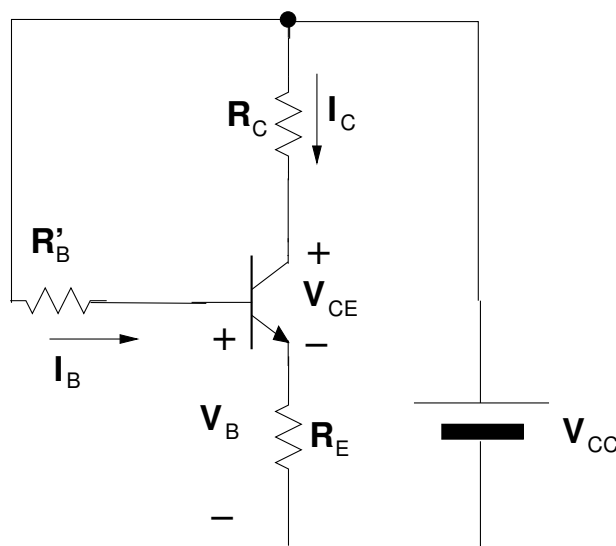
e per il resto si procede nel modo già visto.

Il circuito di polarizzazione esaminato presenta il vantaggio della facilità di analisi, ma ha anche gravi inconvenienti, tali da renderlo raramente utilizzato in pratica. Gli inconvenienti consistono nel fatto che il punto di lavoro è fortemente legato ai parametri del dispositivo e, soprattutto, varia al variare della temperatura. Infatti, se per esempio la temperatura aumenta, diminuisce la V_{BE} necessaria per avere una data I_B . Di conseguenza aumenta la corrente di base e quindi anche quella di emettitore, causando un ulteriore incremento della temperatura della giunzione. Si ha quindi un meccanismo che conduce a forti scostamenti dal punto di riposo e può portare anche alla distruzione del dispositivo. Inserendo una resistenza in serie all'emettitore si ottiene un notevole miglioramento, perché un eventuale incremento di corrente porta a un aumento della caduta sulla resistenza di emettitore e quindi a una diminuzione della V_{BE} , che contrasta l'originario aumento di corrente:



Il calcolo del punto di riposo in presenza di una resistenza di emettitore è più complesso, poiché le due maglie, quella di uscita e quella di ingresso, non sono più disaccoppiate, quindi non è più possibile calcolare la I_B indipendentemente dalla conoscenza della I_C . Per una valutazione esatta del punto di riposo bisogna ricorrere a procedure grafiche abbastanza complesse o a procedure numeriche iterative, che non prenderemo in esame. Considereremo invece un caso particolarmente semplice, che corrisponde alla maggioranza delle situazioni che si incontrano nella pratica.

Innanzitutto va detto che nei circuiti di polarizzazione visti finora è necessario disporre di due generatori distinti: V_{BB} e V_{CC} . Ciò rappresenterebbe in pratica un grosso problema e si cerca invece di avere un unico generatore di alimentazione per molti transistori. Vediamo come si può passare da due generatori di alimentazione a uno soltanto nel circuito appena considerato: basta utilizzare V_{CC} al posto di V_{BB} e variare R_B opportunamente, in modo da ottenere la stessa corrente di base che si sarebbe ottenuta con V_{BB} . Definiamo R'_B la nuova R_B .

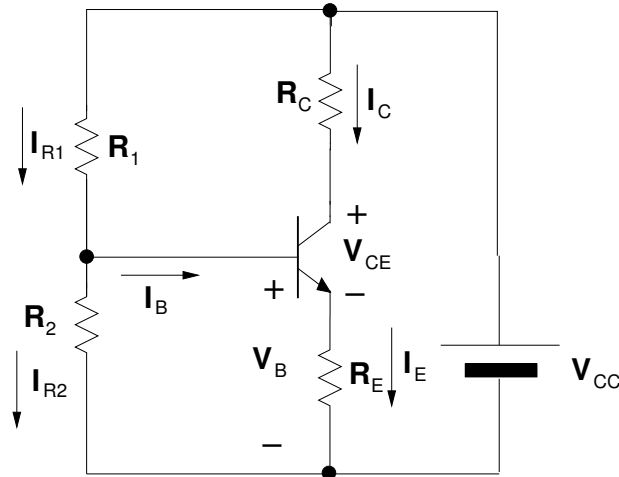


L'altro inconveniente prima citato consiste nel fatto che il punto di lavoro dipende fortemente dalle caratteristiche intrinseche del transistor, le quali possono variare notevolmente, a causa della dipendenza da parametri costruttivi come la larghezza di base che non possono essere controllati con grande precisione. Il circuito che più spesso si utilizza per la polarizzazione dei BJT risolve anche questo problema e si basa sull'impiego di un partitore "pesante" per fissare la tensione di base.

Per partitore pesante si intende un partitore realizzato con resistenze tali da far passare una corrente molto maggiore (almeno 20 volte) di quella che da tale partitore viene derivata, in questo caso particolare la corrente di base. In una tale situazione la tensione che si ottiene con il partitore è sostanzialmente determinata dal rapporto delle resistenze. Perché questa approssimazione sia valida è necessario che la corrente nelle resistenze sia almeno venti volte più grande di quella di base. Di solito, nella risoluzione dei circuiti, si fa l'ipotesi di partitore pesante e poi si verifica al termine se la I_B trovata è consistente con l'ipotesi stessa.

Nell'ipotesi partitore pesante possiamo dunque scrivere che la tensione V_B della base rispetto a massa è data da

$$V_B = V_{CC} \frac{R_2}{R_1 + R_2}.$$



Se supponiamo che la tensione V_{BE} sia pari a V_γ , possiamo subito determinare V_E : $V_E = V_B - V_\gamma$. Quindi

$$I_E = \frac{V_B - V_\gamma}{R_E}.$$

Si noti che in questo caso I_E è stata scelta come uscente dall'emettitore, in modo che abbia segno positivo, per una maggiore comodità nei calcoli. Facciamo ora l'ipotesi, anch'essa da verificare al termine, che $I_B \ll I_C$. In tal caso $I_E \simeq I_C$ e quindi possiamo calcolare la caduta su R_C utilizzando il valore di I_E che abbiamo ottenuto:

$$V_C = V_{CC} - R_C I_C \simeq V_{CC} - R_C I_E.$$

Dunque

$$V_{CE} = V_{CC} - R_C I_C - V_B + V_\gamma.$$

Essendo a questo punto note V_{CE} e I_C , possiamo determinare I_B dalle caratteristiche di uscita a emettitore comune, individuando il punto di riposo identificato da tali valori e ottenendo la I_B corrispondente per interpolazione. Dobbiamo ora verificare che siano effettivamente soddisfatte le disuguaglianze relative alle ipotesi precedentemente fatte:

$$I_B \ll I_{R1}$$

$$I_B \ll I_C.$$

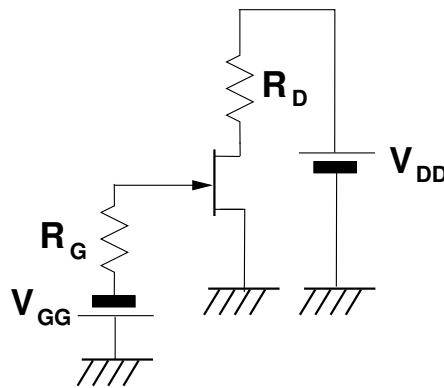
Si noti che per soddisfare la prima delle due disuguaglianze non si può incrementare arbitrariamente I_{R1} , diminuendo il valore delle resistenze del partitore, perché altrimenti avremmo un eccessivo consumo di corrente e un'eccessiva dissipazione di potenza sul partitore. Sottolineiamo anche il fatto che con l'ipotesi di partitore pesante sovrastimiamo la corrente di base (poiché ipotizziamo un valore di V_B maggiore di quello reale, che, assumendo comunque una V_{BE} pari a V_γ , porta a una V_E più grande, conseguentemente a una maggiore I_E e dunque I_C , ottenendo pertanto un punto di lavoro con I_B maggiore), quindi non c'è il rischio che la verifica a posteriori ci tragga in inganno (l'effettiva corrente di base sarà comunque minore di quella stimata).

Dai risultati ottenuti nell'analisi del circuito di polarizzazione con partitore pesante risulta evidente come il punto di lavoro sia fissato, eccetto che per la I_B , dai valori delle quattro resistenze utilizzate e dipenda molto poco dalle caratteristiche intrinseche del transistor, risolvendo quindi il problema che era stato precedentemente messo in evidenza, legato alla dispersione dei parametri dei dispositivi.

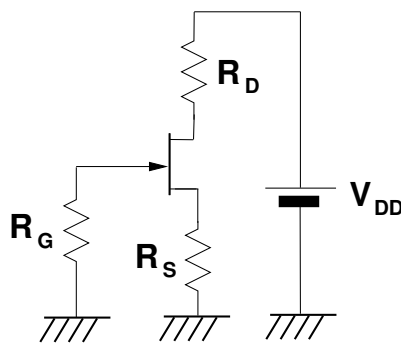
6.2 Reti di polarizzazione per transistori a effetto di campo

La polarizzazione dei transistori a effetto di campo si ottiene con reti molto semplici, grazie al fatto che l'assorbimento di corrente da parte del gate è trascurabile e che il problema della fuga termica non si pone. Prendiamo innanzitutto in esame quest'ultimo punto: avevamo visto che nel caso dei BJT la resistenza di emettitore era indispensabile per ottenere una stabilizzazione del punto di lavoro del transistor. Questa era la conseguenza del fatto che nel BJT la corrente è trasportata prevalentemente da portatori minoritari, la cui concentrazione dipende fortemente dalla temperatura. Nel caso dei transistori a effetto di campo la corrente viene trasportata da portatori maggioritari la cui concentrazione è praticamente indipendente dalla temperatura e la cui mobilità diminuisce all'aumentare della temperatura. Pertanto nei transistori a effetto di campo la corrente tende a diminuire all'aumentare della temperatura, con una conseguente stabilizzazione termica intrinseca.

Per polarizzare un transistor JFET sarebbe quindi sufficiente uno schema del tipo riportato in figura, dove V_{DD} fornisce la tensione di polarizzazione necessaria per il drain e V_G fornisce l'opportuna tensione negativa al gate (la presenza della resistenza R_G è necessaria per evitare di cortocircuitare l'eventuale generatore di segnale connesso al gate).



Un tale circuito non risulta però conveniente dal punto di vista pratico, perché contiene due generatori di tensione, uno dei quali deve fornire una tensione negativa rispetto a massa e non può quindi essere realizzato con una semplice partizione della tensione di alimentazione V_{DD} . Per tale motivo questo circuito non è quasi mai utilizzato in pratica e al suo posto si realizza quello riportato di seguito, consistente nella cosiddetta autopolarizzazione del JFET:



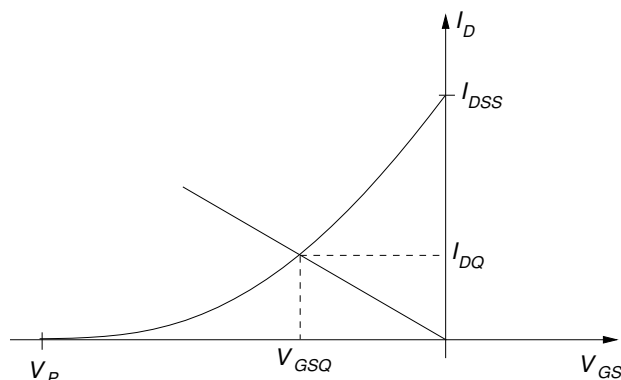
In questo circuito la resistenza R_S dà luogo a una caduta di tensione corrispondente alla V_{GS} necessaria per la polarizzazione (si parla proprio per questo motivo di

autopolarizzazione), mentre la R_G ha la funzione di fissare il gate al potenziale di massa, senza peraltro cortocircuitare il gate stesso a massa. La R_G può anche avere dei valori piuttosto elevati, dell'ordine dei megaohm, senza che su di essa si manifesti una caduta di tensione misurabile, dato il piccolissimo valore della corrente di gate (che ricordiamo corrisponde alla corrente attraverso una giunzione polarizzata inversamente). Possiamo scrivere un'equazione molto semplice:

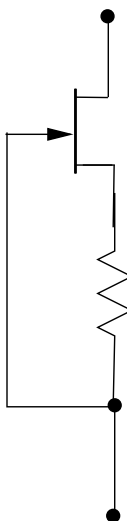
$$I_D = -\frac{V_{GS}}{R_S},$$

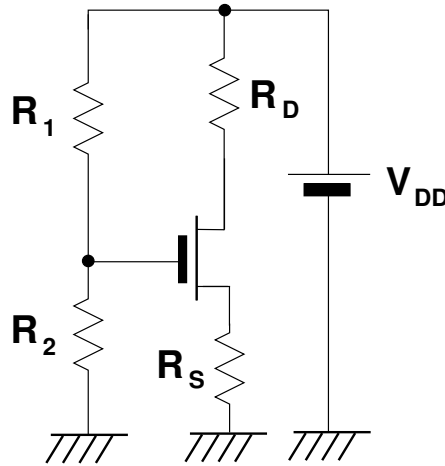
che è rappresentata, sul piano della transcaratteristica ingresso-uscita, da una retta, passante per l'origine e con pendenza pari a $-1/R_S$. Il punto di lavoro corrisponderà all'intersezione tra tale retta e la transcaratteristica stessa. Per quanto riguarda la maglia di uscita è possibile scrivere un'altra semplice equazione, che ci permette di calcolare la V_{DS} :

$$V_{DD} = I_D(R_D + R_S) + V_{DS}.$$



È quindi possibile realizzare un circuito che approssima un generatore ideale di corrente utilizzando semplicemente un transistor JFET e una resistenza, come nello schema di seguito riportato. Il valore della corrente erogata dipenderà dalla transcaratteristica del dispositivo e dal valore della resistenza di source.

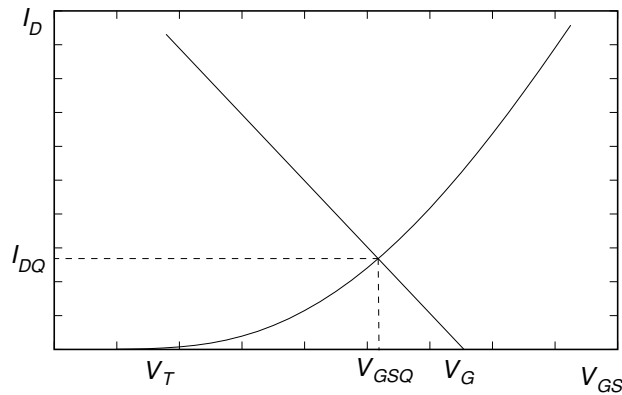




Per la polarizzazione dei transistori MOS si ricorre a schemi di analogia semplicità. Consideriamo per esempio il caso di un transistor MOS a canale n ad arricchimento. Il partitore formato da R_1 e R_2 consente di applicare al gate del transistor MOS una tensione V_G pari a $V_{DD}R_2/(R_1 + R_2)$. Poiché $V_{GS} = V_G - I_D R_S$, avremo l'equazione

$$I_D = -\frac{1}{R_S}V_{GS} + \frac{V_G}{R_S},$$

che rappresenta una retta sul piano della transcaratteristica ingresso-uscita. L'intersezione tra tale retta e la transcaratteristica individuerà il punto di riposo cercato.



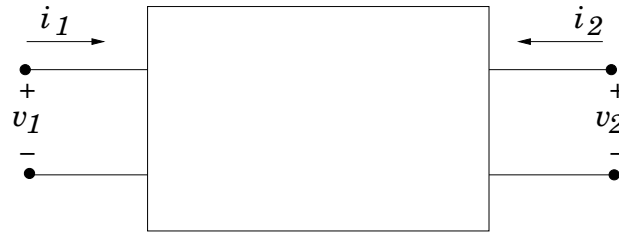
È chiaro che, almeno dal punto di vista del funzionamento in continua, si può fare a meno senza problemi della resistenza di source R_S , dato che la tensione V_{GS} deve comunque essere positiva per questo tipo di dispositivo e non esistono problemi di stabilizzazione termica.

7. Funzionamento linearizzato per piccoli segnali

7.1 Modello linearizzato per transistori BJT

Come nel caso già visto per il diodo, è possibile sviluppare un modello linearizzato anche per i componenti attivi e, in particolare, anche per i transistori BJT. Questo modello è valido soltanto per piccoli spostamenti intorno al punto di lavoro, tali da poter approssimare il comportamento del transistor considerando i soli termini del primo ordine dello sviluppo in serie (nel caso del transistor si tratta, a differenza del diodo, dello sviluppo in serie di funzioni di più variabili).

Il transistor è un quadripolo, quindi il suo modello linearizzato dovrà essere rappresentabile con un quadripolo lineare. Il comportamento elettrico di un quadripolo lineare può essere completamente definito tramite le relazioni tra quattro grandezze: i_1 , v_1 , i_2 , v_2 , le quali sono scelte con i versi illustrati nella figura seguente:



Eccetto che in casi particolari, è di solito possibile esprimere due a scelta delle quattro grandezze in funzione delle due rimanenti. Le relazioni tra tali grandezze risulteranno anch'esse lineari. Scegliamo di esprimere v_1 e i_2 in funzione di i_1 e v_2 (questa è la scelta di solito fatta per la rappresentazione linearizzata dei transistori). Otterremo:

$$\begin{cases} v_1 = h_{11}i_1 + h_{12}v_2 \\ i_2 = h_{21}i_1 + h_{22}v_2. \end{cases}$$

I parametri h_{ij} che definiscono questo modello sono detti parametri ibridi (perché le grandezze indipendenti sono di tipo tra loro diverso: una tensione e una corrente) e si indicano con la lettera h da “hybrid”.

Esaminiamo in dettaglio ciascuno dei parametri h , individuando le relative dimensioni:

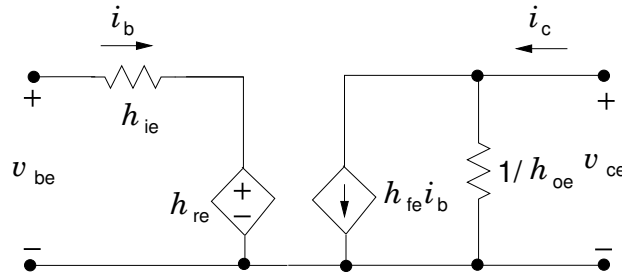
$$\begin{aligned} h_{11} = h_i(\text{input}) &= \left. \frac{v_1}{i_1} \right|_{v_2=0} \quad [\Omega] \\ h_{12} = h_r(\text{reverse}) &= \left. \frac{v_1}{v_2} \right|_{i_1=0} \quad \text{rapporto adimensionale} \\ h_{21} = h_f(\text{forward}) &= \left. \frac{i_2}{i_1} \right|_{v_2=0} \quad \text{rapporto adimensionale} \\ h_{22} = h_o(\text{output}) &= \left. \frac{i_2}{v_2} \right|_{i_1=0} \quad [\Omega^{-1}] \end{aligned}$$

La notazione con pedici letterali indica l'effetto connesso a ciascuno dei parametri h : h_i esprime il rapporto tra la tensione e la corrente in ingresso, quindi una quantità relativa alla sola maglia di ingresso; h_r indica l'azione della tensione di uscita su quella di ingresso, quindi un effetto di tipo inverso; h_f rappresenta il rapporto tra la corrente nella maglia di uscita e quella nella maglia di ingresso, dunque l'effetto

diretto e desiderato; h_o infine esprime il rapporto tra la corrente e la tensione di uscita ed è quindi relativo alla sola maglia di uscita. Di solito viene preso in considerazione il quadripolo corrispondente al montaggio del transistor a emettitore comune, vale a dire con l'emettitore a comune tra l'ingresso e l'uscita. In questa configurazione $v_1 = v_{be}$, $v_2 = v_{ce}$, $i_1 = i_b$ e $i_2 = i_c$. Per specificare che i parametri h si riferiscono al montaggio a emettitore comune si aggiunge un secondo pedice e :

$$\begin{cases} v_{be} = h_{ie}i_b + h_{re}v_{ce} \\ i_c = h_{fe}i_b + h_{oe}v_{ce} \end{cases}$$

Poiché le relazioni appena viste definiscono un quadripolo lineare, possiamo rappresentare tale quadripolo con un circuito equivalente, cercando di identificare il significato fisico dei parametri ibridi:



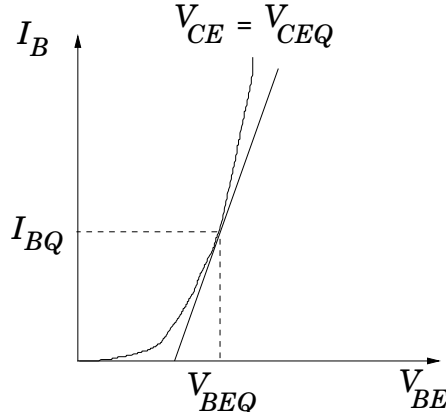
Fino a ora abbiamo rappresentato i parametri ibridi come rapporti tra le grandezze dinamiche, vale a dire come rapporti tra piccole variazioni delle grandezze totali. Possiamo quindi anche indicarli e calcolarli come limite dei rapporti incrementali delle grandezze totali e, di conseguenza, come derivate parziali di tali grandezze (nelle equazioni che seguono useremo, per conformità con l'uso comune, la notazione con simbolo e pedice maiuscolo per indicare le grandezze totali, anche se sarebbe più corretto indicarle con simbolo minuscolo e pedice maiuscolo, come si fa in generale nell'analisi dei circuiti). Per esempio, possiamo scrivere h_{fe} come:

$$h_{fe} = \lim_{\delta I_B \rightarrow 0} \frac{\delta I_C}{\delta I_B} \bigg|_{V_{CE}=V_{CEQ}} = \frac{\partial I_C}{\partial I_B} \bigg|_{V_{CE}=V_{CEQ}},$$

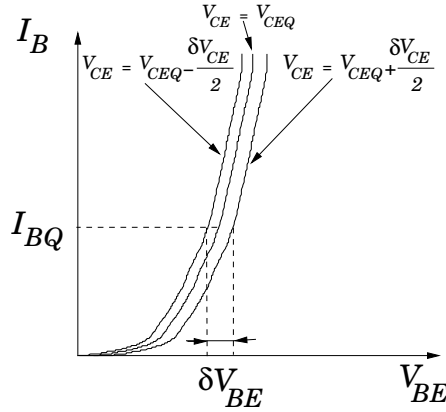
dove la derivata parziale viene valutata per V_{CE} costante e pari al valore nel punto di riposo, che corrisponde ad assumere una variazione nulla rispetto al punto di riposo, e quindi una v_{ce} nulla.

Vediamo ora come si determinano i parametri ibridi a emettitore comune a partire dalle caratteristiche del transistor, facendo riferimento, per semplicità, al caso di un transistor NPN. Consideriamo dapprima una procedura esclusivamente grafica, basata sulle caratteristiche di ingresso e di uscita.

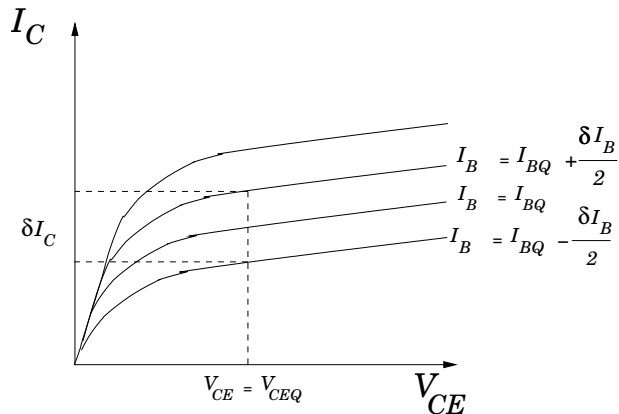
Dalla definizione, h_{ie} può essere approssimato con il rapporto tra la variazione della V_{BE} e quella della I_B , per V_{CE} costante e pari al valore di riposo. Si tratta quindi di prendere in considerazione la caratteristica di ingresso a emettitore comune per $V_{CE} = V_{CEQ}$ e misurare la pendenza della tangente nel punto corrispondente a V_{BEQ} . L'inverso di tale tangente corrisponde proprio a h_{ie} .



Dalla definizione, h_{re} può essere approssimato con il rapporto tra la variazione della V_{BE} e quella della V_{CE} , per I_B costante e pari al valore di riposo. Si tratta quindi di prendere in considerazione le caratteristiche di ingresso a emettitore comune e valutare la differenza tra le V_{BE} che si ottengono sommando o sottraendo $\delta V_{CE}/2$ a V_{CEQ} . Il rapporto tra tale variazione δV_{BE} e quella della tensione di uscita δV_{CE} ci fornisce il valore di h_{re} .

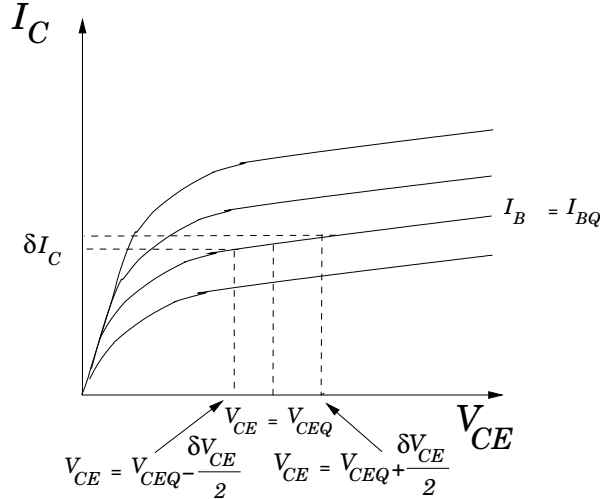


Il parametro h_{fe} può essere approssimato con il rapporto tra la variazione della I_C e quella della I_B , per V_{CE} costante e pari al valore di riposo. Dobbiamo quindi prendere in considerazione le caratteristiche di uscita a emettitore comune e valutare la differenza tra le I_C che si ottengono sommando o sottraendo δI_B a I_{BQ} . Il rapporto tra tale variazione δI_C e quella della corrente di ingresso δI_B ci fornisce il valore di h_{fe} .



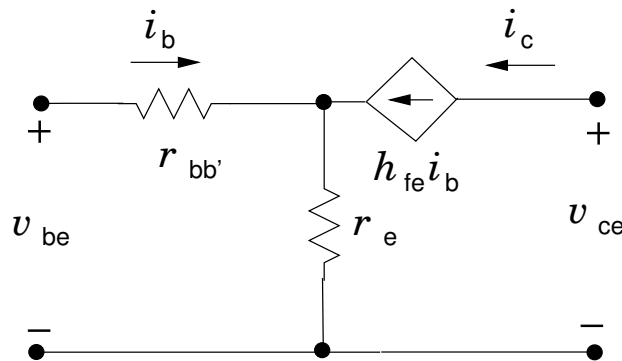
Infine il parametro h_{oe} può essere approssimato con il rapporto tra la variazione della

I_C e quella della V_{CE} , per I_B costante e pari al valore di riposo. Dobbiamo quindi prendere in considerazione le caratteristiche di uscita a emettitore comune e valutare la pendenza della tangente alla caratteristica per $I_B = I_{BQ}$ nel punto di riposo. Il valore di h_{oe} corrisponderà proprio a tale pendenza. Di solito la tangente è quasi orizzontale, dato il piccolo valore di h_{oe} , quindi conviene considerare la variazione di I_C su un intervallo molto ampio, per ottenere una precisione accettabile.



Nella pratica h_{oe} e h_{fe} si valutano effettivamente con la procedura vista, h_{re} si pone di solito pari a zero e h_{ie} si valuta invece in modo diverso, perché le caratteristiche di ingresso sono raramente disponibili.

Innanzitutto dobbiamo precisare che h_{ie} è costituita da due componenti in serie, $r_{bb'}$ e $r_{b'e}$. La componente $r_{bb'}$ corrisponde a una resistenza indesiderata, dovuta alla porzione di base che va dal contatto di ingresso alla zona attiva (base intrinseca). Tale resistenza parassita può avere valori compresi tra pochi ohm e pochi kilohm, a seconda delle caratteristiche costruttive del transistor (di solito è minore nei transistori di potenza maggiore). La $r_{b'e}$ è invece l'effettiva resistenza associata con il funzionamento del transistor. Per calcolare $r_{b'e}$ consideriamo un modello per le variazioni per il transistor in cui introduciamo una resistenza differenziale r_e a comune tra la maglia di ingresso e quella di uscita:



Tale resistenza differenziale è quella della giunzione base-emettitore, che può essere valutata a partire dalle equazioni di Ebers-Moll:

$$I_E = I_{ES} \left(e^{V_{B'E}/V_T} - 1 \right) \simeq I_{ES} e^{V_{B'E}/V_T}.$$

Il termine relativo al generatore comandato dalla corrente nel diodo di collettore non è stato indicato data la polarizzazione inversa della giunzione collettore-base. Si può inoltre trascurare l'unità rispetto a $e^{V_{B'E}/V_T}$, dato che $V_{B'E} \simeq V_\gamma$ e quindi $e^{V_{B'E}/V_T} \gg 1$.

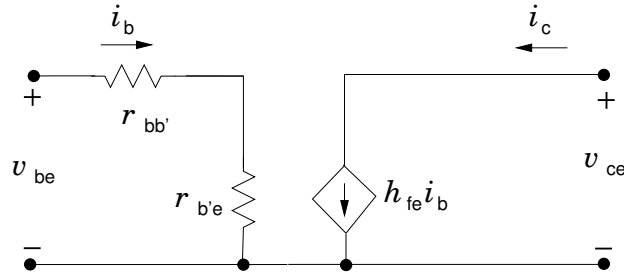
Possiamo quindi ottenere il reciproco di r_e derivando la corrente di emettitore rispetto alla tensione base intrinseca-emettitore:

$$\frac{1}{r_e} = \left. \frac{\partial I_E}{\partial V_{B'E}} \right|_{V_{B'E}=V_{B'E_Q}} = \frac{I_{EQ}}{V_T},$$

dove I_{EQ} è la corrente di emettitore nel punto di lavoro. Inoltre, poiché $I_{EQ} \simeq I_{CQ}$,

$$r_e = \frac{V_T}{I_{EQ}} \simeq \frac{V_T}{I_{CQ}}.$$

Uguagliamo la v_{be} nel circuito appena esaminato a quella nel circuito equivalente a parametri ibridi, semplificando trascurando h_{re} e h_{oe} :



Per il circuito con r_e abbiamo:

$$v_{be} = [r_{bb'} + r_e(h_{fe} + 1)]i_b,$$

mentre per il circuito a parametri ibridi otteniamo:

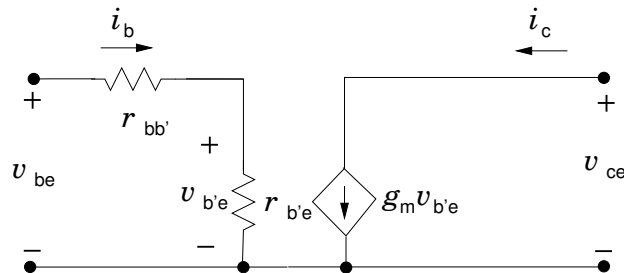
$$v_{be} = (r_{bb'} + r_{b'e})i_b.$$

Dal confronto tra queste due equazioni si ricava che

$$r_{b'e} = r_e(h_{fe} + 1) \simeq \frac{V_T}{I_{CQ}}(h_{fe} + 1) \simeq \frac{V_T}{I_{CQ}}h_{fe}.$$

L'ultima equazione ci permette di calcolare $r_{b'e}$ a partire dal valore di riposo della corrente di collettore e da quello di h_{fe} . Per trovare h_{ie} sarà sufficiente sommare $r_{b'e}$ e $r_{bb'}$, supponendo quest'ultima non dipendente dal punto di riposo, dato che è legata soltanto a parametri costruttivi del transistor.

Talvolta il generatore di corrente comandato in uscita viene rappresentato, invece che come un generatore di corrente comandato in corrente, come un generatore di corrente comandato in tensione, secondo lo schema sotto rappresentato.



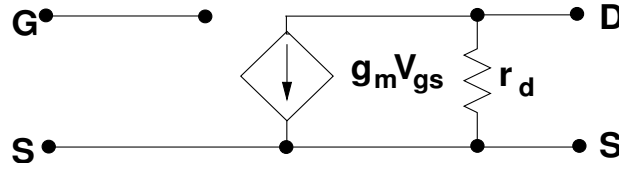
La tensione di controllo è $v_{b'e}$, corrispondente alla differenza di potenziale presente ai capi della $r_{b'e}$. È semplice determinare la relazione tra g_m e i parametri precedentemente considerati, imponendo l'uguaglianza delle correnti di collettore nei due circuiti:

$$h_{fe}i_b = g_mv_{b'e}.$$

Dato che $v_{b'e} = r_{b'e}i_b$, otteniamo che $g_m = h_{fe}/r_{b'e}$. Il parametro g_m ha le dimensioni dell'inverso di una resistenza, quindi Ω^{-1} o S (siemens). Spesso si utilizza come unità di misura per g_m l'ampere su volt (A/V) o il mA/V (dimensionalmente equivalenti all'inverso di una resistenza).

7.2 Modello linearizzato per transistori a effetto di campo

Si utilizza lo stesso modello linearizzato, molto semplice, per tutti i transistori a effetto di campo (sia MOSFET sia JFET). Esso è costituito da un generatore di corrente comandato in tensione, con in parallelo una resistenza r_d , che rappresenta l'effetto dovuto al fenomeno di accorciamento del canale al crescere di V_{DS} .



Il valore della transconduttanza g_m (dell'ordine di qualche mA/V per la maggior parte dei transistori a effetto di campo) si può ricavare dalla transcaratteristica fornita in forma grafica determinando la pendenza della tangente alla stessa nel punto di lavoro, che corrisponde proprio a g_m . Se la caratteristica è invece fornita in maniera analitica, g_m si può ottenere derivando I_D rispetto a V_{GS} . Per un JFET otteniamo

$$g_m = \frac{-2I_{DSS}}{V_P} \left(1 - \frac{V_{GS}}{V_P}\right).$$

Questa formula è valida per JFET sia a canale p sia a canale n e dà luogo a un valore di g_m sempre positivo. È importante sottolineare che se la caratteristica è fornita in modo grafico è opportuno utilizzare il metodo della tangente per ricavare g_m , invece di determinare V_P e I_{DSS} per poi procedere con l'espressione analitica. Quest'ultimo metodo, assai sconsigliabile, può portare a risultati significativamente diversi da quello grafico perché la transcaratteristica non è esattamente una parabola (l'espressione analitica è soltanto un'approssimazione).

Il valore della resistenza differenziale r_d si ricava andando a valutare la pendenza della tangente alle caratteristiche di uscita in corrispondenza del punto di lavoro: r_d corrisponde all'inverso del coefficiente angolare di tale tangente.

Per quanto riguarda i transistori MOS, se la transcaratteristica è fornita in modo grafico si procede in modo analogo a quanto già visto per i JFET. Se, invece, la transcaratteristica è data in forma analitica, la si deriva rispetto a V_{GS} , ottenendo

$$g_m = 2k \left(\frac{W}{L}\right) (V_{GS} - V_T).$$

La r_d dei transistori MOS si ricava come già visto per i JFET se le caratteristiche di uscita sono disponibili in forma grafica. Se sono invece disponibili nella

forma dell'espressione fenomenologica che include la tensione di Early, si ottiene come inverso della derivata di tale espressione rispetto a V_{DS} :

$$r_d = \left(\frac{dI_D}{dV_{DS}} \right)^{-1} = \frac{1 + \lambda V_{DS}}{\lambda I_D}.$$

Se $\lambda V_{DS} \ll 1$, abbiamo $r_d \simeq 1/(\lambda I_D)$.

8. Analisi delle principali configurazioni circuitali a BJT e FET

8.1 Concetti generali

Esamineremo nel seguito i principali montaggi che si utilizzano per i circuiti a transistori, che differiscono tra loro sostanzialmente per la scelta dell'elettrodo che risulta a comune tra l'ingresso e l'uscita. Per il BJT si parla quindi di configurazione a emettitore comune (CE), collettore comune (CC) e base comune (CB), mentre il FET può essere utilizzato a source comune (CS), drain comune (CD) e gate comune (CG).

Per ciascuna configurazione determineremo i quattro parametri caratterizzanti, rappresentati dal guadagno di corrente A_i , che è pari al rapporto tra la corrente di uscita e quella di ingresso

$$A_i = \frac{i_o}{i_i},$$

il guadagno di tensione A_v , corrispondente al rapporto tra la tensione di uscita e quella di ingresso

$$A_v = \frac{v_o}{v_s},$$

la resistenza di ingresso R_i , che è uguale al rapporto tra la tensione e la corrente in ingresso

$$R_i = \frac{v_s}{i_i},$$

e la resistenza di uscita, corrispondente al rapporto tra la tensione posta in uscita tramite un generatore di prova e la corrente di uscita, per tensione di ingresso nulla

$$R_o = \left. \frac{v_o}{i_o} \right|_{v_s=0}.$$

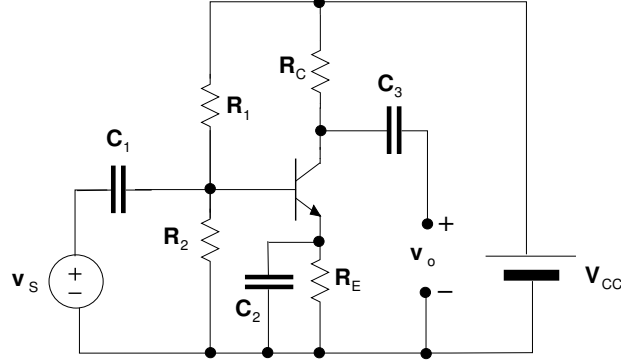
L'analisi di ciascuno stadio sarà condotta supponendo di aver già individuato il punto di lavoro e di conoscere quindi il valore di tutti i parametri del circuito equivalente per le variazioni del transistor. In particolare, per i BJT considereremo nulli sia h_{re} sia h_{oe} , allo scopo di semplificare i calcoli. Nella maggior parte delle situazioni che si incontrano nella pratica h_{re} è effettivamente trascurabile, ma h_{oe} può non esserlo e deve essere preso in considerazione per ottenere risultati accurati.

In questa prima analisi supporremo di operare a “media frequenza”, vale a dire in una condizione per cui tutti gli eventuali condensatori presenti possono essere considerati corto circuiti dal punto di vista del funzionamento dinamico del circuito (mentre vengono chiaramente considerati dei circuiti aperti per quanto riguarda la determinazione del punto di riposo). Se sono presenti delle induttanze, queste saranno considerate corto circuiti nell'analisi in continua e circuiti aperti nell'analisi per le variazioni.

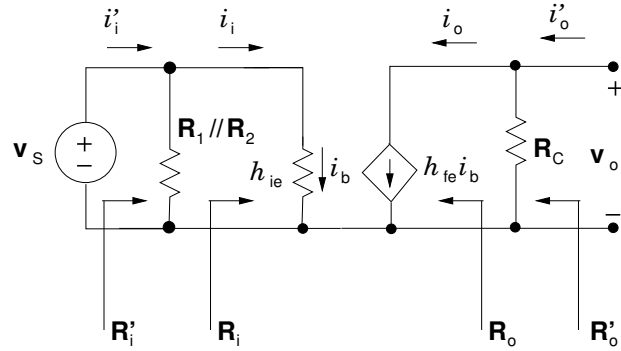
8.2 Stadio amplificatore a emettitore comune senza resistenza di emettitore

In pratica uno stadio senza resistenza di emettitore non si incontra mai, a causa dei già discussi problemi che si verificano in assenza di tale resistenza, ma il suo comportamento è identico a quello di uno stadio, molto comune, in cui la resistenza di emettitore è presente, ma è cortocircuitata, dal punto di vista delle variazioni, da un condensatore posto in parallelo alla stessa.

Di seguito viene riportato uno schema “tipico” di un amplificatore CE con resistenza di emettitore cortocircuitata per le variazioni. Notiamo che anche all’uscita è stato posto un condensatore, allo scopo di disaccoppiare, dal punto di vista della componente di polarizzazione continua, l’eventuale carico connesso in uscita.



Possiamo ora tracciare il circuito dinamico, nel quale i condensatori sono stati sostituiti da corto circuiti, così come il generatore di tensione continua V_{CC} . In tale situazione le resistenze che formano il partitore di ingresso vengono a trovarsi tra loro in parallelo.



A seconda della convenzione scelta per definire le correnti di ingresso e di uscita, possiamo considerare come tali i_i e i_o oppure i'_i e i'_o , e conseguentemente definire come resistenze di ingresso e di uscita R_i e R_o oppure R'_i e R'_o , rispettivamente. Per quanto riguarda il guadagno di corrente di solito ci si riferisce al rapporto i_o/i_i :

$$A_i = \frac{i_o}{i_i} = \frac{h_{fe}i_b}{i_b} = h_{fe}.$$

Anche il guadagno di tensione può essere calcolato facilmente, osservando che $v_o = -h_{fe}i_bR_C$:

$$A_v = \frac{v_o}{v_i} = \frac{-h_{fe}i_bR_C}{h_{ie}i_b} = -\frac{h_{fe}R_C}{h_{ie}}.$$

Per quanto riguarda la resistenza di ingresso otteniamo

$$R_i = \frac{v_s}{i_i} = h_{ie},$$

oppure

$$R'_i = R_1 // R_2 // R_i = R_1 // R_2 // h_{ie},$$

nel caso in cui si voglia considerare la resistenza vista a monte del partitore di base. La resistenza di uscita è data da

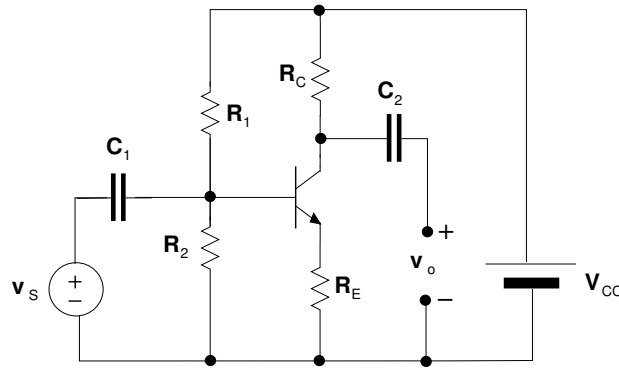
$$R_o = \left. \frac{v_o}{i_o} \right|_{v_s=0} = \infty,$$

dato che dall'uscita si vede soltanto un generatore di corrente controllato, la cui grandezza di controllo non dipende dalla tensione applicata sull'uscita stessa e risulta nulla per $v_s = 0$. Se vogliamo prendere in considerazione la resistenza vista a valle della R_C , otteniamo

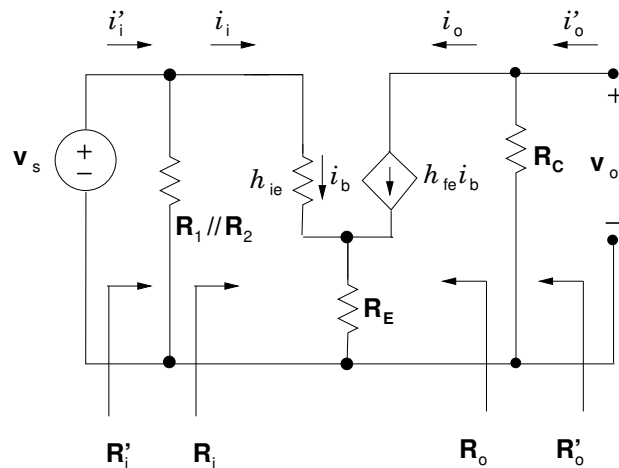
$$R'_o = R_o // R_C = R_C.$$

8.3 Stadio amplificatore a emettitore comune con resistenza di emettitore

Esaminiamo ora come variano i risultati precedentemente ottenuti nel caso in cui venga aggiunta una resistenza in serie all'emettitore del BJT. I calcoli sono leggermente più complessi perché la resistenza di emettitore introduce un accoppiamento tra la maglia di ingresso e quella di uscita.



Rappresentiamo il circuito dinamico e cerchiamo di individuare un approccio che consenta di trovare i guadagni di tensione e di corrente, oltre alle resistenze di ingresso e di uscita, senza bisogno di ricorrere alla trattazione per nodi o per maglie e alla risoluzione di un sistema lineare.



La tensione v_s può essere espressa in funzione della i_b osservando che essa risulta pari alla somma della caduta di tensione sulla resistenza h_{ie} , attraversata da i_b , più quella sulla R_E , attraversata da una corrente $(h_{fe} + 1)i_b$:

$$v_s = h_{ie}i_b + (h_{fe} + 1)i_b R_E.$$

Da questa espressione, dividendo per i_b , si ricava subito il valore di R_i :

$$R_i = h_{ie} + (h_{fe} + 1)R_E.$$

Questo è un risultato molto importante: la resistenza vista sulla base di un transistor BJT è pari alla somma di h_{ie} e della resistenza totale che si trova in serie all'emettitore, moltiplicata per $h_{fe} + 1$. Tale risultato sarà di comune utilizzo nella analisi dei circuiti a transistori. La resistenza R'_i che si vede a monte del partitore di base sarà

$$R'_i = R_1 // R_2 // [h_{ie} + (h_{fe} + 1)R_E].$$

La tensione di uscita v_o risulta $v_o = -h_{fe}i_b R_C$ come nel caso precedentemente esaminato, quindi il guadagno di tensione può essere facilmente determinato

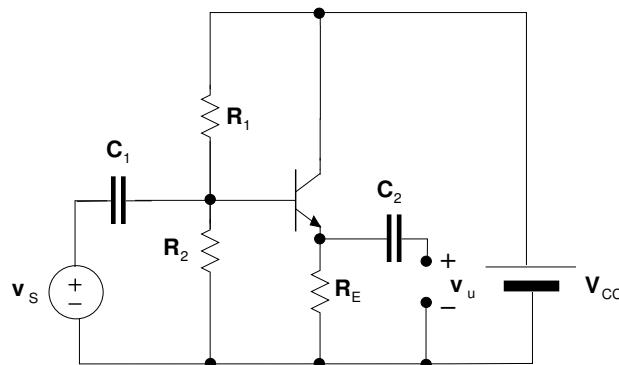
$$A_v = \frac{v_o}{v_s} = \frac{-h_{fe}R_C}{h_{ie} + (h_{fe} + 1)R_E}.$$

Notiamo che il guadagno di tensione risulta ridotto rispetto al caso dell'amplificatore a emettitore comune senza resistenza di emettitore, perché in questo caso il denominatore è aumentato della quantità $(h_{fe} + 1)R_E$. Tale quantità risulta spesso molto maggiore di h_{ie} , essendo $h_{fe} \gg 1$ e R_E frequentemente dello stesso ordine di grandezza di h_{ie} . In tal caso il guadagno di tensione può essere approssimato come $A_v \simeq -R_C/R_E$. In molti casi la riduzione di guadagno dovuta alla presenza di R_E non è accettabile, per cui si ricorre al condensatore di "bypass" in parallelo alla R_E , che abbiamo visto nel precedente paragrafo.

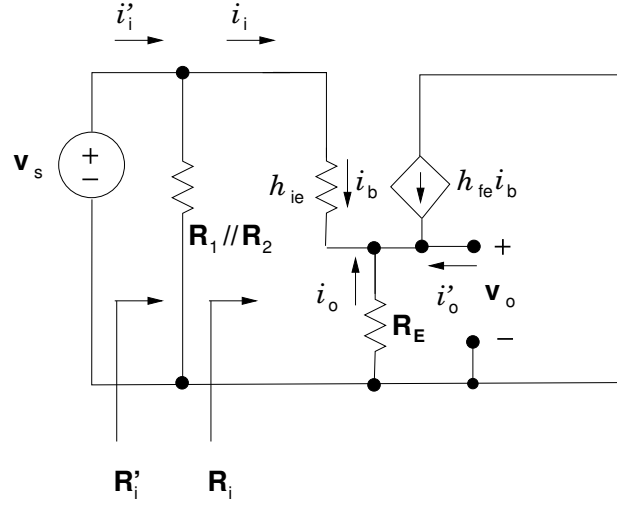
Il guadagno di corrente è lo stesso determinato per la configurazione precedente, essendo anche in questo caso pari al rapporto tra la corrente di collettore e quella di base: $A_i = h_{fe}$. La resistenza di uscita R_o è anche questa volta infinita, per i motivi visti nel paragrafo precedente e la R'_o risulta pari alla R_C .

8.4 Stadio amplificatore a collettore comune

Nello stadio a collettore comune, il collettore è l'elettrodo a comune tra ingresso e uscita, il segnale di ingresso viene inviato alla base e quello di uscita viene prelevato dall'emettitore. Vediamone innanzitutto lo schema.



Il circuito equivalente per le variazioni è ottenuto secondo le solite regole, cortocircuitando generatori di tensione continua e condensatori e sostituendo il transistor con il suo equivalente a parametri ibridi.



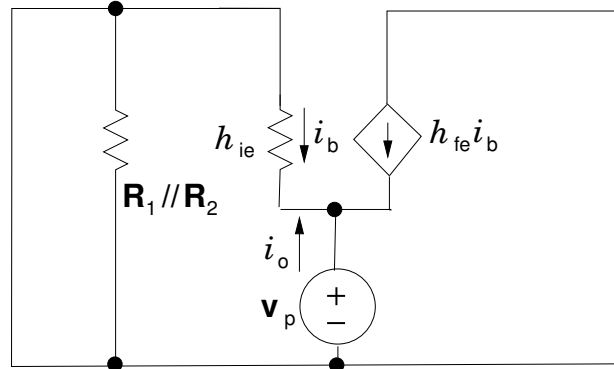
La relazione che sussiste tra la tensione di ingresso v_s e la corrente di ingresso $i_i = i_b$ è esattamente la stessa del circuito trattato nel paragrafo precedente: $v_s = i_b[h_{ie} + (h_{fe} + 1)R_E]$, quindi

$$R_i = h_{ie} + (h_{fe} + 1)R_E$$

e, per quanto riguarda la resistenza vista a monte del partitore di base,

$$R'_i = R_1 // R_2 // R_i.$$

Per quanto riguarda invece la resistenza di uscita R_o , questa risulta data dal rapporto tra la tensione v_o e la corrente i_o entrante nell'emettitore. Calcoliamo il valore di questa resistenza ponendo un generatore di tensione di prova v_p sull'uscita e calcolando la i_o risultante (abbiamo tolto la resistenza R_E , dato che stiamo calcolando R_o e non R'_o).



La corrente erogata dal generatore V_p risulta pari a $-(h_{fe} + 1)i_b$, mentre i_b si ottiene facilmente notando che ai capi di h_{ie} è presente una tensione $-v_p$:

$$i_b = -\frac{v_p}{h_{ie}},$$

quindi

$$R_o = \frac{v_p}{i_o} = \frac{v_p}{\frac{v_p}{h_{ie}}(h_{fe} + 1)} = \frac{h_{ie}}{h_{fe} + 1}.$$

Quindi la resistenza vista sull'emettitore di un transistor BJT risulta pari alla resistenza che si trova sul ramo di base (questa volta semplicemente h_{ie} , ma in generale tutto ciò che si trova in serie sul ramo di base) divisa per $h_{fe} + 1$. Anche questo è un risultato importante e di uso generale nell'analisi dei circuiti a transistori. La resistenza di uscita R'_o vista dalla porta di uscita risulta pari al parallelo di R_o e di R_E :

$$R'_o = R_E // \frac{h_{ie}}{h_{fe} + 1}.$$

Quindi la resistenza di uscita di uno stadio CC è in genere molto bassa: nel parallelo prevale di solito $h_{ie}/(h_{fe} + 1)$, che è di qualche decina di ohm contro le centinaia di ohm che rappresentano un valore comune per R_E .

Il guadagno di corrente risulta semplicemente pari al rapporto tra la corrente di emettitore e quella di base, quindi abbiamo:

$$A_i = \frac{i_o}{i_i} = -(h_{fe} + 1).$$

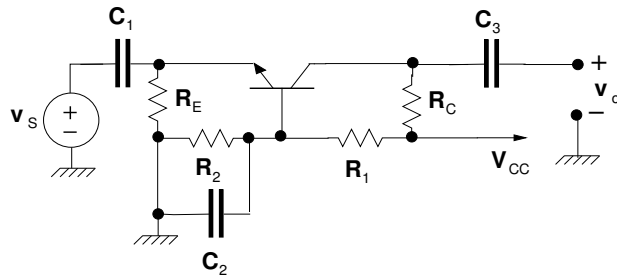
Determiniamo infine il guadagno di tensione: la tensione di uscita è data da $v_o = (h_{fe} + 1)i_b R_E$, quindi

$$A_v = \frac{v_o}{v_s} = \frac{(h_{fe} + 1)i_b R_E}{[h_{ie} + (h_{fe} + 1)R_E]i_b} = \frac{(h_{fe} + 1)R_E}{h_{ie} + (h_{fe} + 1)R_E}.$$

Osserviamo che il guadagno di tensione è minore dell'unità, poiché il denominatore è sempre maggiore del numeratore. Osserviamo inoltre che se $(h_{fe} + 1)R_E \gg h_{ie}$, $A_v \simeq 1$. Il guadagno è dunque circa pari all'unità e con segno positivo. Pertanto la tensione di uscita sull'emettitore "insegue" quella sulla base. È questo il motivo per cui lo stadio CC è spesso indicato come "inseguitore di emettitore" o "emitter follower".

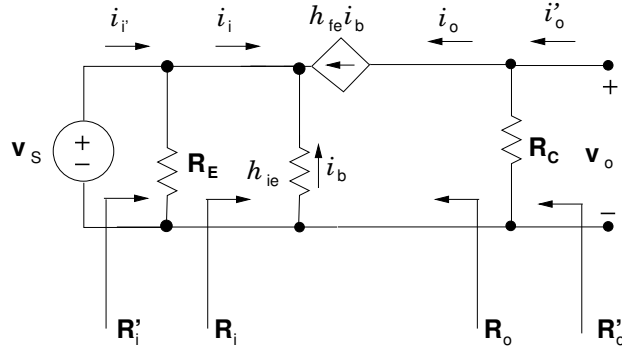
8.5 Stadio amplificatore a base comune

Nello stadio a base comune, la base si trova a comune tra ingresso e uscita, ma non viene collegata direttamente a massa, per evitare di complicare il circuito di polarizzazione. Dal punto di vista dinamico essa è comunque connessa a massa tramite un condensatore di "bypass", indicato come C_2 nello schema.



In questo caso il segnale di ingresso viene applicato all'emettitore e il segnale di uscita viene prelevato sul collettore. Esaminando il circuito equivalente per le variazioni, riportato di seguito, valutiamo la resistenza di ingresso R_i vista a valle di R_E . Poiché $i_i = -(h_{fe} + 1)i_b$ e $i_b = -v_s/h_{ie}$,

$$R_i = \frac{v_s}{i_i} = \frac{h_{ie}}{h_{fe} + 1}.$$



Quindi la resistenza di ingresso di uno stadio a base comune risulta particolarmente bassa. Il guadagno di corrente, definito come rapporto di i_o e i_i , risulta

$$A_i = \frac{i_o}{i_i} = \frac{h_{fe}i_b}{-(h_{fe} + 1)i_b} = \frac{-h_{fe}}{h_{fe} + 1}.$$

Se, come di solito avviene, $h_{fe} \gg 1$, $A_i \simeq -1$. Per quanto riguarda il guadagno di tensione avremo

$$A_v = \frac{v_o}{v_s} = \frac{-h_{fe}R_C i_b}{-i_b h_{ie}} = h_{fe} \frac{R_C}{h_{ie}}.$$

Il guadagno di tensione risulta quindi positivo e potenzialmente molto maggiore dell'unità (se R_C è dello stesso ordine di grandezza di h_{ie} o maggiore). La resistenza di uscita si valuta analogamente al caso del CE, con R_o , valutata senza considerare R_C , infinita e $R_o' = R_C$.

8.6 Riepilogo configurazioni a BJT

Possiamo dunque concludere che la configurazione a emettitore comune è l'unica che può fornire allo stesso tempo un guadagno di tensione e un guadagno di corrente molto maggiori dell'unità. Essa è anche caratterizzata da un guadagno di tensione di segno negativo; si dice quindi che lo stadio CE è uno stadio invertente. La resistenza di ingresso dello stadio CE può variare in un intervallo molto ampio, a seconda della presenza o meno della resistenza di emettitore.

Lo stadio a collettore comune presenta una resistenza di ingresso analoga a quella del CE con resistenza di emettitore, mentre la resistenza di uscita è molto bassa e il guadagno di tensione è circa unitario (comunque mai maggiore dell'unità). Si utilizzerà tale configurazione quando si dispone di una sorgente che non è in grado di fornire molta corrente e si vuole pilotare un carico che assorbe invece una corrente significativa.

Lo stadio a base comune presenta una resistenza di ingresso molto bassa e un guadagno di tensione potenzialmente elevato, ma il guadagno di corrente è circa unitario.

Riassumiamo i risultati ottenuti in una tabella riepilogativa, ricordando che essi sono validi nell'ipotesi semplificativa di $h_{re} = h_{oe} = 0$. In caso contrario, alcune delle espressioni ricavate sarebbero state significativamente più complicate.

	CE	CE con R_E	CC	CB
A_i	h_{fe}	h_{fe}	$-(1 + h_{fe})$	$\frac{-h_{fe}}{1+h_{fe}}$
A_v	$-\frac{h_{fe}R_C}{h_{ie}}$	$-\frac{h_{fe}R_C}{h_{ie}+R_E(h_{fe}+1)}$	$\frac{(1+h_{fe})R_E}{h_{ie}+R_E(h_{fe}+1)}$	$h_{fe}\frac{R_C}{h_{ie}}$
R_i	h_{ie}	$h_{ie} + R_E(h_{fe} + 1)$	$h_{ie} + R_E(h_{fe} + 1)$	$\frac{h_{ie}}{1+h_{fe}}$
R_o	∞	∞	$\frac{h_{ie}}{h_{fe}+1}$	∞

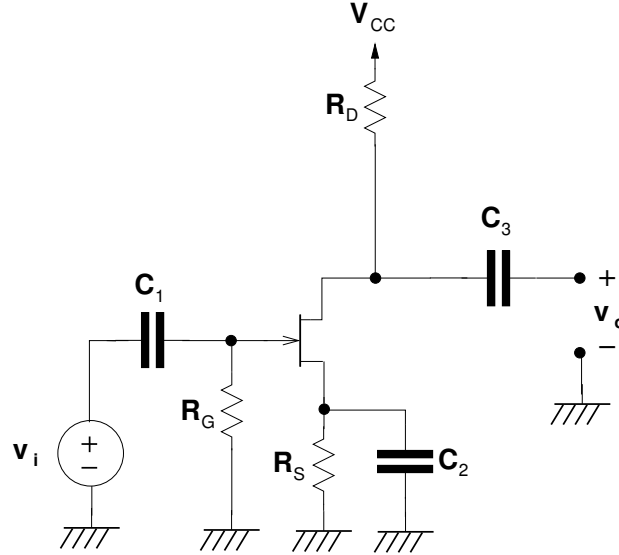
8.7 Comportamento dei transistori PNP

Tutti gli esempi che abbiamo visto fino a questo punto sono stati per amplificatori utilizzando transistori NPN. È facile convincersi che nel caso siano presenti dei transistori PNP il circuito equivalente per le variazioni rimane del tutto invariato, mentre invece le polarità delle tensioni di polarizzazione devono essere rovesciate nel circuito per il funzionamento in continua.

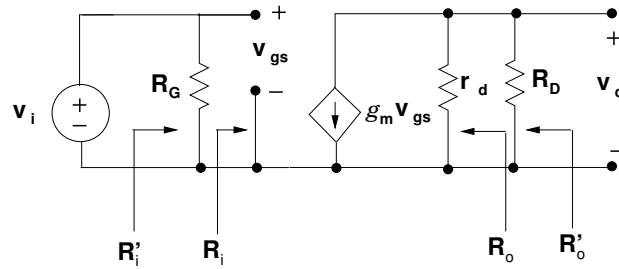
Vediamo un semplice ragionamento che può essere utile per convincersi della completa equivalenza dal punto di vista delle variazioni tra transistori PNP e transistori NPN. Tra parentesi riporteremo le quantità relative a un transistor PNP, mentre le altre sono relative a un transistor NPN. Consideriamo le correnti totali i_B e i_C entranti nel collettore e nella base per ambedue i tipi di transistor. Avremo $i_B > 0$ ($i_B < 0$), $i_C > 0$, ($i_C < 0$). Consideriamo in ambedue i casi una variazione Δv_{BE} di v_{BE} positiva, corrispondente a un segnale positivo in ingresso. Avremo allora $\Delta v_{BE} > 0$ ($\Delta v_{BE} > 0$) e $\Delta i_B > 0$, $\Delta|i_B| > 0$ ($\Delta i_B > 0$, $\Delta|i_B| < 0$), poiché, in conseguenza della variazione positiva di v_{BE} applicata, nel caso NPN la giunzione base-emettitore risulta polarizzata in diretta da una tensione di modulo maggiore, mentre nel caso PNP tale tensione risulta, in modulo, minore. La variazione di i_B indicata ha come conseguenza la seguente variazione di i_C : $\Delta|i_C| > 0$ ($\Delta|i_C| < 0$), la quale, essendo $i_C > 0$ ($i_C < 0$) implica $\Delta i_C > 0$ ($\Delta i_C > 0$). Quindi per una variazione positiva della v_{BE} e, conseguentemente, della i_B si ha in tutti e due i casi una variazione positiva della i_C . Ne consegue che il circuito per le variazioni è lo stesso (in particolare ha gli stessi versi delle correnti e delle tensioni) per transistori PNP e NPN.

8.8 Stadio amplificatore a source comune

Anche i transistori a effetto di campo possono essere utilizzati in tre diverse configurazioni: source comune (CS), drain comune (CD) e gate comune (CG). Prenderemo in esame soltanto le prime due poiché sono le più comunemente utilizzate. Iniziamo dalla configurazione a source comune:



Questa volta abbiamo indicato il generatore di segnale in ingresso con v_i , allo scopo di evitare possibili confusioni con la tensione di source, anch'essa convenzionalmente indicata con v_s . La resistenza di source, necessaria per l'autopolarizzazione del FET è presente, ma dal punto di vista dinamico scompare, cortocircuitata dal condensatore C_2 .



Poiché il source è a massa, $v_s = 0$ e quindi la tensione v_{gs} coincide con la v_g , che è a sua volta pari alla v_i . Calcoliamo il valore della tensione di uscita v_o :

$$v_o = -g_m v_g r_d // R_D = -g_m v_i r_d // R_D,$$

quindi

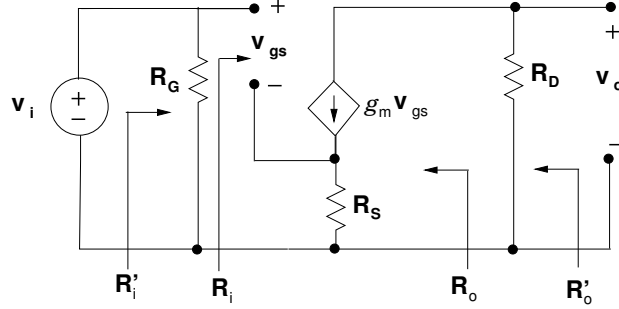
$$A_v = \frac{v_o}{v_i} = -g_m r_d // R_D.$$

Non ha significato parlare del guadagno di corrente, dato che questo sarebbe comunque infinito, essendo nulla la corrente di ingresso.

La resistenza di ingresso R_i risulta dunque anch'essa infinita, mentre $R'_i = R_i // R_G = R_G$. Per quanto riguarda la resistenza di uscita, essa risulta $R_o = r_d$, $R'_o = r_d // R_D$.

8.9 Stadio amplificatore a source comune con resistenza di source

Consideriamo ora il caso in cui non si ponga un condensatore in parallelo alla R_S , la quale compare quindi anche nel circuito per le variazioni. Lo schema complessivo è lo stesso già visto nel paragrafo precedente, con l'unica variazione che il condensatore C_2 in parallelo alla R_S è rimosso. Possiamo dunque tracciare il circuito per le variazioni, nel quale consideriamo r_d infinita allo scopo di semplificare i calcoli.



In questo caso non si verifica più l'uguaglianza tra v_g e v_{gs} , quindi dovremo ricavare l'espressione che lega queste due quantità. Possiamo scrivere v_s in funzione di v_{gs} :

$$v_s = g_m v_{gs} R_S.$$

Poiché $v_{gs} = v_g - v_s$, possiamo ricavare v_{gs}

$$\begin{aligned} v_{gs} &= v_g - g_m v_{gs} R_S \\ v_{gs}(1 + g_m R_S) &= v_g \\ v_{gs} &= \frac{v_g}{1 + g_m R_S}. \end{aligned}$$

Abbiamo così ottenuto una relazione che lega v_g a v_{gs} e possiamo dunque calcolare v_o in funzione di v_g :

$$v_o = -g_m v_{gs} R_D = -\frac{g_m v_g}{1 + g_m R_S} R_D = -\frac{g_m v_i}{1 + g_m R_S} R_D.$$

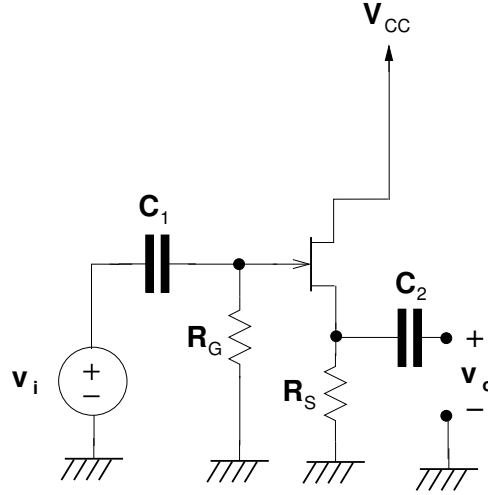
Quindi

$$A_v = \frac{v_o}{v_i} = -\frac{g_m R_D}{1 + g_m R_S}.$$

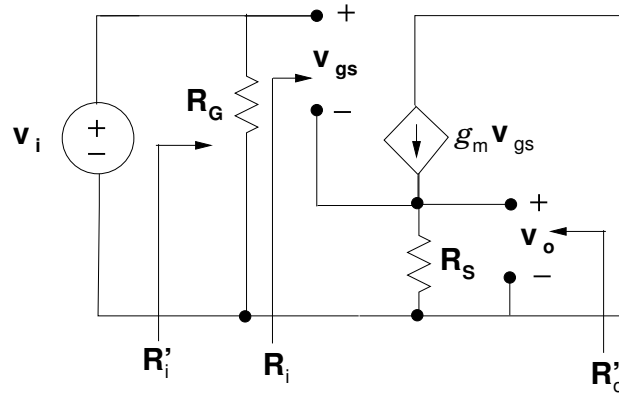
Per quanto riguarda le resistenze di ingresso e di uscita abbiamo gli stessi valori trovati nel paragrafo precedente, con l'unica variante che questa volta, essendo stata considerata r_d infinita, anche R_o è infinita, mentre $R'_o = R_D$. Inoltre R_i è infinita e $R'_i = R_G$.

8.10 Stadio amplificatore a drain comune

Lo schema dello stadio a drain comune è riportato di seguito: in questo caso l'uscita viene prelevata sul source, mentre l'ingresso è sul gate.



Il circuito per le variazioni risulta invece lo stesso già visto nel paragrafo precedente, con l'unica variante della porta di uscita.



Valutiamo la tensione di uscita $v_o = v_s$. Abbiamo precedentemente determinato che $v_s = g_m v_{gs} R_S$ e che $v_{gs} = v_g / (1 + g_m R_S)$. Combinando queste due relazioni otteniamo

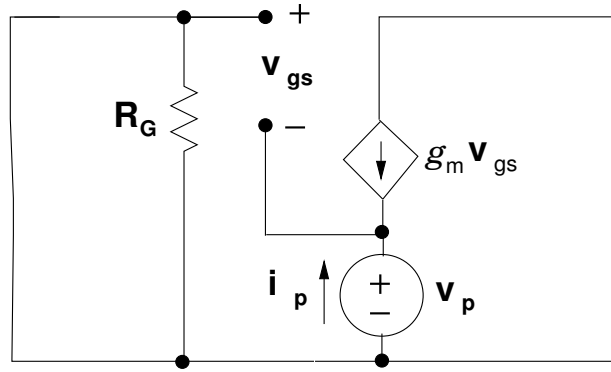
$$v_s = \frac{g_m v_g R_S}{1 + g_m R_S}$$

e quindi che

$$A_v = \frac{v_o}{v_i} = \frac{g_m R_S}{1 + g_m R_S}.$$

Notiamo che A_v risulta positivo e minore dell'unità. Inoltre, se (come di solito avviene) $g_m R_S \gg 1$, $A_v \simeq 1$. Abbiamo quindi un comportamento simile a quello dell'inseguitore di emettitore e si parla infatti di inseguitore di source (source follower).

La resistenza di ingresso è la stessa delle configurazioni viste precedentemente: R_i risulta infinita e R'_i è pari a R_G . Valutiamo ora la resistenza di uscita: a questo scopo consideriamo un generatore di prova v_p posto sull'uscita, come nel circuito dinamico che segue (la resistenza è stata tolta e sarà poi inclusa nel risultato finale).



Poiché $v_g = 0$ (essendo il generatore v_i cortocircuitato nella misura della resistenza di uscita), $v_{gs} = -v_s = -v_p$. Inoltre la corrente erogata da v_p risulta

$$i_p = -g_m v_{gs} = g_m v_p.$$

Pertanto

$$R_o = \frac{v_p}{i_p} = \frac{v_p}{g_m v_p} = \frac{1}{g_m}.$$

Questo è un risultato importante, da ricordare perché torna utile nello studio della maggior parte dei circuiti contenenti transistori a effetto di campo. Si noti come in questo caso la resistenza vista su un generatore comandato sia del tutto diversa da quella che si vede sull'altro terminale dello stesso generatore (sul drain si sarebbe vista una resistenza infinita): la differenza è dovuta al fatto che nel caso presente la tensione applicata agisce sulla grandezza di controllo del generatore. Vista dalla porta di uscita, includendo R_S , la resistenza di uscita R'_o risulta $R'_o = (1/g_m) // R_S$.

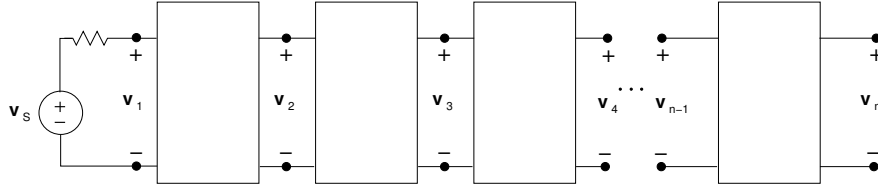
8.11 Riepilogo configurazioni a FET

Abbiamo preso in considerazione le due principali configurazioni a FET, quella a source comune e quella a drain comune, notando le profonde analogie esistenti con le configurazioni a emettitore comune e a collettore comune dei BJT. Come nel caso dei BJT, la configurazione a source comune consente di ottenere un significativo guadagno di tensione e determina un'inversione di fase del segnale; la configurazione a drain comune fornisce invece un guadagno di tensione positivo e circa unitario, ma offre una ridotta resistenza di uscita. In tutti i casi non abbiamo preso in considerazione il guadagno di corrente, perché questo, nell'ipotesi fatta di corrente nulla di gate, risulterebbe infinito. Possiamo riassumere i risultati ottenuti in una tabella:

	CS	CS con R_S	CD
A_v	$-g_m R_D$	$\frac{-g_m R_D}{1+g_m R_S}$	$\frac{g_m R_S}{1+g_m R_S}$
R_o	∞	∞	$\frac{1}{g_m}$

8.12 Amplificatori multistadio

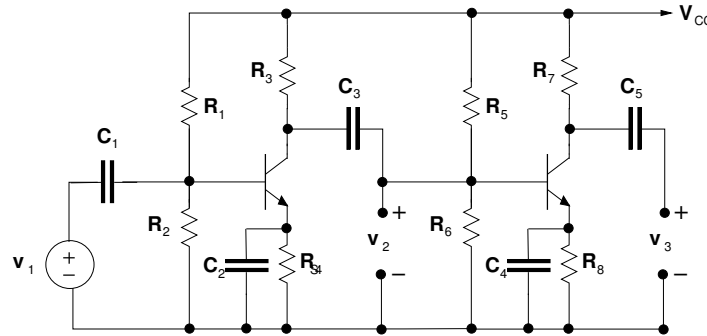
Nel caso in cui si vogliano ottenere guadagni superiori a quelli realizzabili con un singolo transistor, si ricorre ad amplificatori costituiti da più stadi in cascata, che possono essere rappresentati come una catena di quadripoli, il cui guadagno di tensione risulta pari al prodotto dei guadagni di tensione dei vari stadi, calcolati però considerando connesso in uscita a ciascuno stadio un carico corrispondente alla resistenza di ingresso dello stadio successivo.



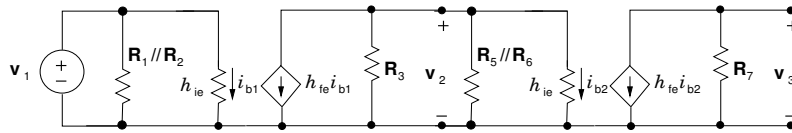
$$A_v = \frac{v_n}{v_1} = \frac{v_n}{v_{n-1}} \frac{v_{n-1}}{v_{n-2}} \dots \frac{v_4}{v_3} \frac{v_3}{v_2} \frac{v_2}{v_1} = A_{v_{n-1}} A_{v_{n-2}} \dots A_{v_3} A_{v_2} A_{v_1}.$$

È importante sottolineare il fatto che i guadagni che compaiono nel prodotto non sono quelli calcolati per i singoli stadi isolati, ma devono tenere conto dell'interazione tra gli stadi stessi.

Vediamo un esempio al riguardo, considerando un circuito formato dalla cascata di due stadi a emettitore comune.



Per analizzare il guadagno ricorriamo alla rappresentazione nella forma di circuito equivalente per le variazioni, sostituendo i transistori con i loro circuiti equivalenti a parametri ibridi.



Notiamo subito che in questo caso il guadagno del primo stadio ($A_1 = v_2/v_1$) non è più dato semplicemente da $-h_{fe}R_3/h_{ie}$, ma va tenuta in considerazione anche la resistenza vista all'ingresso del secondo stadio, che risulta $R_5//R_6//h_{ie}$. Pertanto

$$A_1 = -\frac{h_{fe}R_3//R_5//R_6//h_{ie}}{h_{ie}}.$$

Tale guadagno risulta dunque significativamente inferiore rispetto a quello del singolo stadio isolato.

Da questa semplice analisi possiamo dedurre una regola di utilità generale: il guadagno di tensione di due stadi in cascata è uguale al prodotto dei guadagni di tensione dei singoli stadi, calcolati con gli stadi stessi isolati, solo se la resistenza di ingresso dello stadio a valle è infinita o quella di uscita dello stadio a monte è nulla. Infatti in ambedue questi casi lo stadio a valle non “carica” quello a monte.

Anche il guadagno di corrente non è in generale pari al prodotto dei guadagni di corrente dei singoli stadi calcolati isolatamente, perché la corrente di ingresso di uno stadio non è uguale a quella di uscita dello stadio precedente, a causa delle partizioni che avvengono tra le varie resistenze.

Una relazione importante, che utilizzeremo nel seguito, è quella che ci fornisce il guadagno di tensione di uno stadio in funzione del suo guadagno di corrente. Possiamo scrivere il guadagno di tensione A_v nella forma (usiamo i moduli per evitare di preoccuparci della scelta dei versi delle correnti di ingresso e di uscita)

$$|A_v| = |A_i| \frac{R'_L}{R_i},$$

dove R'_L è la resistenza effettivamente vista in uscita, pari, per esempio nel caso dello stadio a emettitore comune, alla resistenza di collettore con in parallelo la resistenza di ingresso dello stadio successivo, e R_i è la resistenza di ingresso.

Consideriamo ora il problema della scelta della configurazione circuitale da adottare per i vari stadi. Se vogliamo ottenere un guadagno di tensione maggiore dell'unità, è evidente che non possiamo utilizzare più stadi a collettore comune connessi in cascata, dato che il guadagno di ciascuno di essi è minore dell'unità. Anche gli stadi a base comune non si prestano bene al collegamento in cascata, visto che il guadagno di tensione risultante è al più pari a quello del solo ultimo stadio, come possiamo facilmente dimostrare nel caso di stadi CB uguali tra loro. Infatti la resistenza complessiva vista in uscita R'_L è il risultato del parallelo tra la resistenza di collettore e quella complessiva di ingresso dello stadio successivo, che è minore (se è presente un partitore di polarizzazione, per esempio) o uguale della R_i dello stadio considerato. Quindi $R'_L < R_i$ e, ricordando che per lo stadio CB $|A_i| < 1$, possiamo concludere che $|A_v| = |A_i| R'_L / R_i < 1$. Quindi ponendo in cascata più stadi a base comune si ottiene un guadagno inferiore all'unità.

Rimangono quindi solo gli stadi a emettitore comune, i quali possono essere combinati per ottenere un guadagno più grande di quello di un singolo stadio poiché sia il loro guadagno di tensione sia quello di corrente possono essere molto maggiori dell'unità. Stadi a base comune o a collettore comune possono essere utilizzati all'ingresso o all'uscita di una catena multistadio nel caso sia necessario ottenere particolari valori per le resistenze di ingresso e di uscita.

Quanto detto vale anche per gli amplificatori a FET, per i quali si possono realizzare amplificatori multistadio basati sulla configurazione a source comune.

8.13* Teorema di Miller

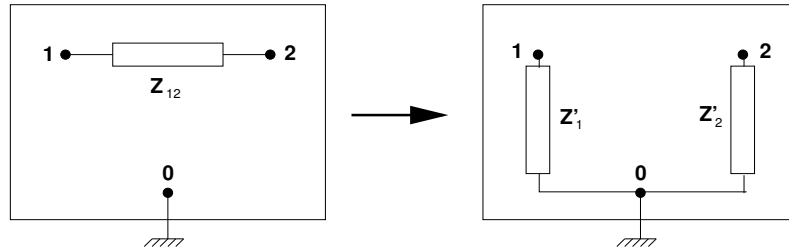
Nello studio degli amplificatori a transistori può tornare utile in alcuni casi (soprattutto nell'analisi, che vedremo più avanti, del comportamento in alta frequenza) l'applicazione del teorema di Miller, che ci consente di sostituire un'impedenza connessa tra due nodi con due impedenze connesse tra tali nodi e la massa.

Consideriamo una generica rete elettrica, nella quale identifichiamo tre nodi: i nodi 1 e 2, tra i quali è connessa un'impedenza Z_{12} e un nodo di riferimento, indicato con 0. Se conosciamo il rapporto k tra le tensioni V_1 e V_2 misurate, rispettivamente,

tra i nodi 1 e 2 rispetto al nodo 0, possiamo definire una rete elettrica equivalente, nella quale l'impedenza Z_{12} è stata rimossa e sono state invece inserite due impedenze Z'_1 e Z'_2 , rispettivamente tra il nodo 1 e il nodo 0 e tra il nodo 2 e il nodo 0, il cui valore è dato da

$$Z'_1 = \frac{Z_{12}}{1 - k}$$

$$Z'_2 = \frac{Z_{12}k}{k - 1}.$$



Questo teorema va usato con attenzione, notando che l'equivalenza tra le due reti è legata alla conoscenza del rapporto k , che deve essere valutabile sulla rete originaria. In particolare, come vedremo nel seguito, il teorema di Miller risulterà utile quando potremo supporre che k sia sostanzialmente indipendente dalla frequenza nella gamma di frequenze di nostro interesse, e pari al valore k_0 a bassa frequenza.

9. Analisi del comportamento in frequenza degli amplificatori

9.1 Concetti generali

Nel seguito prenderemo in esame il comportamento in frequenza degli amplificatori, che fino a ora abbiamo studiato nel cosiddetto limite delle “medie frequenze”, in cui i condensatori che compaiono nel circuito vengono considerati corto circuiti dal punto di vista delle variazioni e circuiti aperti per il funzionamento in continua. Analogamente le induttanze sono state considerate circuiti aperti per il segnale variabile nel tempo e corto circuiti per il funzionamento in continua. Un simile comportamento corrisponderebbe però alla realtà soltanto se tutti i componenti reattivi utilizzati avessero valore tendente all’infinito e se i componenti elettronici avessero davvero caratteristiche del tutto indipendenti dalla frequenza, come abbiamo finora supposto.

In realtà la presenza di componenti reattivi (condensatori e induttanze) determina una dipendenza dalla frequenza di tutti i parametri caratteristici di un circuito elettronico, in particolare del modulo e della fase del guadagno. Per lo studio di tale dipendenza risulta conveniente ricorrere a un formalismo basato su un concetto più ampio di frequenza, la cosiddetta “frequenza generalizzata”, indicata mediante la variabile complessa s , nell’ambito della Trasformata di Laplace. Tale trasformata sarà molto brevemente introdotta nel seguito, trattandone soltanto gli aspetti necessari per il nostro studio degli amplificatori.

Introdurremo, senza dimostrazione, alcuni risultati di base della teoria delle reti elettriche, tra cui le proprietà delle funzioni di trasferimento, dopodiché presenteremo le regole di base per il tracciamento dei diagrammi di Bode e le definizioni generali per lo studio della risposta in frequenza degli amplificatori.

9.2 La trasformata di Laplace

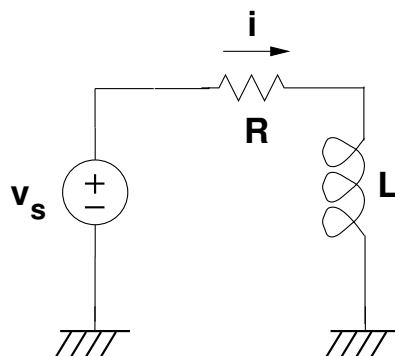
È noto dall’Elettrotecnica che il comportamento nel tempo di una rete elettrica lineare può essere descritto da un sistema di equazioni integro-differenziali, che corrispondono all’imporre, istante per istante, i principi di Kirchhoff alle maglie o ai nodi. Consideriamo, per esempio, il caso delle equazioni alle maglie: la relazione tra la corrente e la tensione ai capi di ciascun elemento circuitale può essere scritta (senza entrare nel dettaglio delle condizioni iniziali e delle variabili di stato):

$$\begin{aligned} v &= Ri && \text{per le resistenze} \\ v &= L \frac{di}{dt} && \text{per le induttanze} \\ v &= \frac{1}{C} \int i dt && \text{per i condensatori.} \end{aligned}$$

Per esempio, se volessimo studiare un circuito a una sola maglia, come quello rappresentato in figura, potremmo scrivere una equazione corrispondente a imporre che la somma delle cadute di tensione sulla maglia sia nulla:

$$-v_s + Ri + L \frac{di}{dt} = 0.$$

Un circuito comunque complesso sarà rappresentato da un sistema di equazioni sia differenziali (per le induttanze) sia integrali (per i condensatori). Tale sistema può essere trasformato sempre in un sistema di equazioni differenziali a coefficienti costanti, se deriviamo ciascun termine rispetto al tempo, eliminando così i termini integrali.



La risoluzione di un tale sistema di equazioni differenziali è in generale di una certa complessità. Risulta quindi molto utile la tecnica della trasformata di Laplace, che ci permette di passare da un sistema di equazioni differenziali del secondo ordine a un sistema di equazioni algebriche lineari.

Così come la trasformata di Fourier ci permette di rappresentare una generica funzione del tempo come sovrapposizione lineare (tramite l'integrale di Fourier) di funzioni sinusoidali e cosinusoidali della frequenza, la trasformata di Laplace ci consente di rappresentare una generica funzione del tempo come sovrapposizione lineare di funzioni di una variabile s , definita frequenza generalizzata. La frequenza generalizzata $s = \sigma + j\omega$ è complessa e rappresenta una classe di eccitazioni elementari della rete più vasta di quella descritta dalla pulsazione reale ω . Mentre nella trasformata di Fourier le eccitazioni elementari (funzioni di base) sono del tipo $e^{j\omega t}$, nella trasformata di Laplace le funzioni di base sono del tipo

$$e^{st} = e^{(\sigma + j\omega)t} = e^{\sigma t} e^{j\omega t} = e^{\sigma t} [\cos(\omega t) + j\text{sen}(\omega t)].$$

Corrispondono pertanto, oltre che alle funzioni sinusoidali e cosinusoidali della base di Fourier (quando $\sigma = 0$) a oscillazioni smorzate (per $\sigma < 0$), a oscillazioni di ampiezza crescente esponenzialmente (per $\sigma > 0$) e a esponenziali crescenti o decrescenti (per $\omega = 0$).

Matematicamente la trasformata di Laplace è definita dal seguente integrale

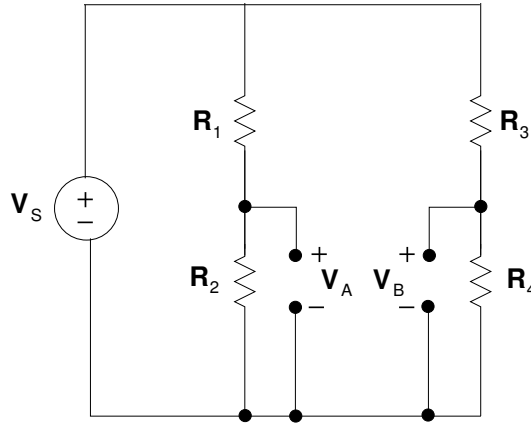
$$F(s) = \int_0^\infty f(t) e^{-st} dt,$$

dove $f(t)$ è una generica funzione del tempo. È possibile dimostrare facilmente che se $F(s)$ è la trasformata di Laplace di $f(t)$, l'integrale di $f(t)$ avrà come trasformata $(1/s)F(s)$ e la derivata di $f(t)$ avrà come trasformata $sF(s)$.

È proprio quest'ultima importante proprietà che ci consente, tramite la trasformata di Laplace, di trasformare un sistema di equazioni differenziali a coefficienti costanti in un sistema di equazioni algebriche lineari. In particolare, si ricava una regola operativa particolarmente semplice: ogni induttanza viene sostituita da un'impedenza di valore Ls e ogni condensatore da un'impedenza di valore $1/(Cs)$, dopodiché le equazioni alle maglie o ai nodi si scrivono seguendo le stesse regole incontrate per la risoluzione dei circuiti puramente resistivi, dove al posto delle resistenze abbiamo impedenze di valore complesso. È anche possibile introdurre facilmente il valore delle condizioni iniziali, ma non ci occuperemo di questo aspetto.

9.3 Funzioni di rete e funzioni di trasferimento

Si definisce funzione di rete il rapporto tra due generiche grandezze elettriche presenti in due punti di una rete, espresse nel dominio della frequenza generalizzata: può trattarsi delle tensioni su due porte della rete, delle correnti su due rami o, più in generale, anche del rapporto tra una corrente e una tensione o viceversa. Trattandosi di due grandezze elettriche qualunque, non è detto che esista tra le stesse una relazione causale. Si parla invece di funzione di trasferimento (un concetto più restrittivo di quello di funzione di rete) nel caso del rapporto tra una grandezza elettrica in un punto della rete e un'altra grandezza elettrica che ne è la causa, ambedue espresse nel dominio della frequenza generalizzata. Un tipico esempio di funzione di trasferimento è il rapporto tra la tensione di uscita e quella di ingresso di un amplificatore: il segnale in uscita è la diretta conseguenza di quello in ingresso, infatti si annulla se il segnale di ingresso viene rimosso. Vediamo un esempio di un circuito molto semplice in cui è possibile individuare delle funzioni di trasferimento e delle funzioni di rete che non sono funzioni di trasferimento (tutte le funzioni di trasferimento sono anche funzioni di rete).



Sono funzioni di trasferimento V_A/V_S e V_B/V_S , poiché V_S è la causa sia di V_A sia di V_B , mentre V_A/V_B o V_B/V_A sono funzioni di rete, visto che V_B non è la causa che produce V_A e V_A non è la causa che produce V_B .

9.4 La funzione di trasferimento nel dominio della frequenza generalizzata

L'andamento in s di una qualsiasi grandezza elettrica $X_u(s)$ del circuito in risposta all'azione di un'altra grandezza elettrica $X_i(s)$ applicata in un qualunque punto della rete può essere calcolato risolvendo un sistema lineare con la regola di Kramer. È possibile dimostrare che la funzione di trasferimento $A(s)$ tra la grandezze $X_u(s)$ e $X_i(s)$ può essere sempre espresso come rapporto di polinomi in s a coefficienti reali, con la proprietà ulteriore, derivante da ragioni di realizzabilità fisica, che il grado del numeratore di tale rapporto è sempre minore o uguale di quello del denominatore (questa limitazione non esiste nel caso di una funzione di rete che non sia una funzione di trasferimento, perché non si pongono problemi di realizzabilità fisica, dato che la grandezza al denominatore non è la causa di quella al numeratore).

$$A(s) = \frac{a_m s^m + a_{m-1} s^{m-1} + \cdots + a_1 s + a_0}{s^n + b_{n-1} s^{n-1} + \cdots + b_1 s + b_0}, \quad m \leq n.$$

Mentre il numeratore della funzione di trasferimento dipende dalla scelta della grandezza di ingresso e di quella di uscita, il denominatore è invece indipendente da tale scelta ed è caratteristico della rete (corrisponde infatti, nella regola di Kramer, al determinante della matrice che descrive la rete). Inoltre, il grado del polinomio al denominatore è pari al numero di elementi reattivi indipendenti che compaiono nella rete (per esempio, due condensatori non sono indipendenti tra loro se formano una maglia impropria, vale a dire una maglia formata di soli condensatori).

La funzione di trasferimento può anche essere scritta nella forma

$$A(s) = a_m \frac{(s - s_{z_1})(s - s_{z_2}) \cdots (s - s_{z_m})}{(s - s_{p_1})(s - s_{p_2}) \cdots (s - s_{p_n})},$$

dove gli s_{z_i} sono le radici del polinomio al numeratore (dette zeri, perché si ha $A(s_{z_i}) = 0$ e gli s_{p_j} sono le radici del polinomio al denominatore, dette poli. Come abbiamo già detto, il polinomio al denominatore non dipende dalla scelta delle grandezze di ingresso e di uscita e il suo grado è pari al numero di elementi reattivi indipendenti, quindi i poli sono pari anch'essi al numero di tali elementi e sono caratteristici della rete, vale a dire gli stessi per ogni funzione di trasferimento relativa a quella rete.

Poiché sia il polinomio a numeratore sia quello a denominatore hanno coefficienti reali, gli zeri e i poli potranno essere reali o complessi coniugati. La presenza di poli con parte reale positiva indica che il sistema è in grado di produrre in uscita un segnale di ampiezza non limitata, in corrispondenza a un segnale di ingresso limitato; in altre parole corrisponde a dire che il sistema è instabile. I circuiti dei quali ci occuperemo saranno di solito stabili, quindi tutti i loro poli avranno parte reale negativa.

Spesso sia i poli sia gli zeri sono reali e negativi (anche se nulla vieta, anche per un sistema stabile, che uno o più zeri abbiano parte reale positiva), quindi, ogni volta che si hanno poli o zeri reali, si è soliti fare una sostituzione puramente formale: al posto di $-s_{z_i}$ si pone ω_{z_i} e al posto di $-s_{p_j}$ si pone ω_{p_j} , ottenendo

$$A(s) = a_m \frac{(s + \omega_{z_1})(s + \omega_{z_2}) \cdots (s + \omega_{z_m})}{(s + \omega_{p_1})(s + \omega_{p_2}) \cdots (s + \omega_{p_n})}.$$

È importante ribadire che si tratta di una sostituzione puramente formale, dato che il contenuto informativo della rappresentazione rimane inalterato.

Un'operazione con un significato tutt'altro che formale è invece quella che consiste nel sostituire s con $j\omega$. In tal caso si passa dalla variabile complessa s , corrispondente alla frequenza generalizzata, alla variabile immaginaria $j\omega$ dove ω corrisponde alla pulsazione "fisica" ($\omega = 2\pi f$, dove f è la frequenza "fisica"). Questa operazione corrisponde anche al passaggio dalla trasformata di Laplace alla trasformata di Fourier.

$$A(j\omega) = a_m \frac{(j\omega + \omega_{z_1})(j\omega + \omega_{z_2}) \cdots (j\omega + \omega_{z_m})}{(j\omega + \omega_{p_1})(j\omega + \omega_{p_2}) \cdots (j\omega + \omega_{p_n})}.$$

Si perde quindi l'informazione relativa alla parte reale di s , dato che rimane solo la parte immaginaria: $A(j\omega)$ descrive il comportamento del circuito in regime sinusoidale.

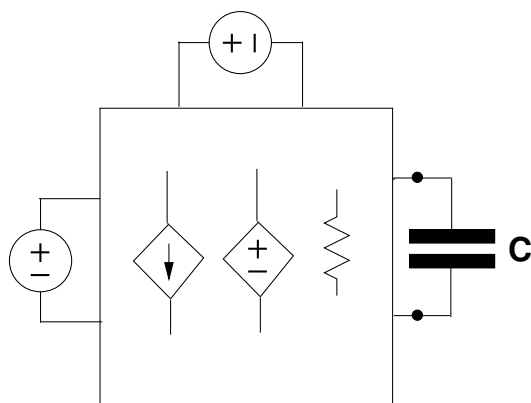
9.5 Il metodo della resistenza vista

Nel caso in cui si abbia una rete elettrica nella quale è presente un solo elemento reattivo o, più in generale, una porzione di una rete elettrica identificabile come non

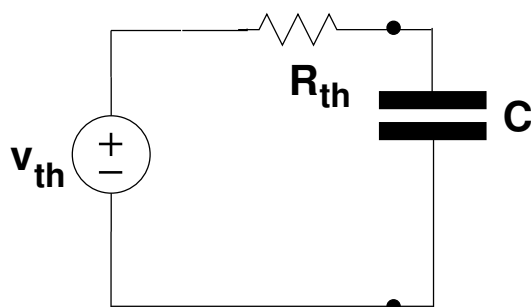
interagente con il resto della rete e contenente un solo elemento reattivo, il polo può essere calcolato con un metodo estremamente semplice, detto metodo della resistenza vista.

Innanzitutto ricordiamo che i poli di una rete sono caratteristici della rete stessa e quindi rimarranno gli stessi per tutte le funzioni di trasferimento di tale rete e delle reti a essa equivalenti.

Consideriamo una generica rete contenente generatori indipendenti, generatori dipendenti, resistenze e un solo elemento reattivo che supponiamo essere un condensatore (potremmo ripetere la stessa dimostrazione, in modo del tutto analogo, considerando un'induttanza invece del condensatore). Poniamo in evidenza i nodi tra i quali è connesso il condensatore C :



Dal punto di vista della porta alla quale è connesso il condensatore, la rete può essere rappresentata con il suo equivalente di Thevenin, costituito da un generatore di tensione di Thevenin V_{th} , di valore pari alla tensione misurata a vuoto sulla porta a cui è connesso C e da una resistenza di Thevenin R_{th} , di valore pari alla resistenza vista dalla porta a cui è connesso C , quando tutti i generatori indipendenti della rete sono disattivati (cortocircuitati i generatori indipendenti di tensione e aperti i generatori indipendenti di corrente).



Il polo di questa rete può essere facilmente calcolato andando a valutare la funzione di trasferimento tra V_{th} e V_c , la tensione ai capi del condensatore.

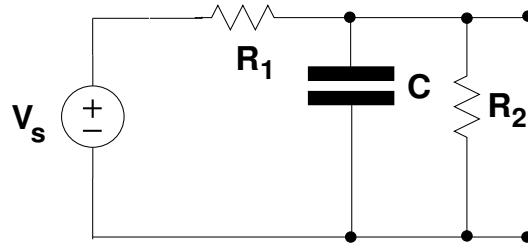
Notiamo che nel dominio della frequenza generalizzata rappresentiamo le grandezze elettriche con lettere maiuscole e pedice minuscolo. Nel seguito adotteremo tale convenzione nello studio di tutti i circuiti linearizzati, per i quali considereremo sempre di effettuare un'analisi nel dominio della frequenza generalizzata.

Sulla base della procedura indicata nel paragrafo 9.2, possiamo scrivere

$$V_c = V_{th} \frac{\frac{1}{Cs}}{R_{th} + \frac{1}{Cs}} = V_{th} \frac{\frac{1}{R_{th}C}}{s + \frac{1}{R_{th}C}},$$

quindi il polo vale $s_p = -1/(R_{th}C)$, $\omega_p = 1/(R_{th}C)$. Questa espressione si può quindi usare in tutte le situazioni in cui si ha una rete con un solo condensatore; di solito, R_{th} si definisce resistenza vista dal condensatore e si indica con R_{VC} .

Vediamo un semplice esempio di applicazione del metodo della resistenza vista. Prendiamo in esame il circuito sotto riportato, nel quale è presente un solo elemento reattivo e determiniamone il polo.



La resistenza vista da C risulta $R_{VC} = R_2 // R_1$, dato che per la sua valutazione dobbiamo considerare V_s cortocircuitato. Quindi

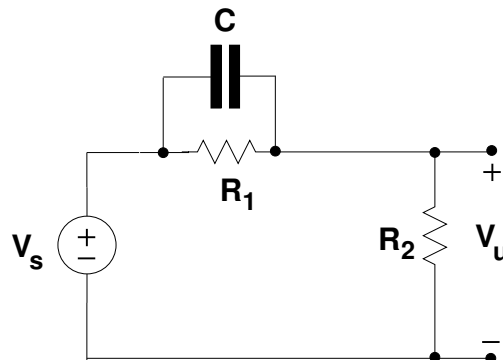
$$\omega_p = \frac{1}{R_{VC}C} = \frac{1}{CR_1 // R_2}.$$

9.6 Calcolo della funzione di trasferimento

Per determinare la funzione di trasferimento di un circuito si può procedere sostituendo agli elementi reattivi le opportune impedenze (in funzione della frequenza generalizzata) e poi svolgere formalmente i calcoli come se si stesse trattando un circuito puramente resistivo. Questa procedura, tuttavia, risulta particolarmente tediosa e soggetta a frequenti sviste anche nei casi più semplici. Cercheremo quindi di evitarla ogni volta che sarà possibile.

Vedremo ora un esempio relativo a un circuito elementare e calcoleremo la funzione di trasferimento sia con il metodo, sconsigliato, sopra descritto sia con una procedura più semplice, basata su una tecnica di “ispezione” della rete.

Il circuito che considereremo è un partitore formato da una resistenza con in serie un parallelo resistenza-condensatore:



Procedendo con la tecnica standard otteniamo

$$\begin{aligned}
 A(s) = \frac{V_u}{V_s} &= \frac{R_2}{R_1 // \frac{1}{Cs} + R_2} = \frac{R_2}{\frac{R_1/(Cs)}{R_1+1/(Cs)} + R_2} \\
 &= \frac{R_2}{\frac{R_1/(Cs) + R_2 R_1 + R_2/(Cs)}{R_1+1/(Cs)}} = \frac{R_2(R_1 Cs + 1)}{R_1 + R_2 + R_2 R_1 Cs} \\
 &= \frac{R_1 R_2 C \left(s + \frac{1}{R_1 C} \right)}{R_1 R_2 C \left(s + \frac{R_1 + R_2}{R_2 R_1 C} \right)} = \frac{s + \frac{1}{R_1 C}}{s + \frac{R_1 + R_2}{R_2 R_1 C}}.
 \end{aligned}$$

Quindi, anche per un circuito così semplice, si è dovuto svolgere un calcolo abbastanza complicato.

Tale risultato si sarebbe potuto ricavare in modo molto più semplice utilizzando il metodo della resistenza vista e dell'analisi dei guadagni a zero e all'infinito. Infatti il sistema ha un solo polo, avendo un solo elemento reattivo (C), e basterà quindi valutare la resistenza vista da tale condensatore. Quindi

$$R_{VC} = R_1 // R_2 = \frac{R_1 R_2}{R_1 + R_2}.$$

Quindi possiamo scrivere che

$$\omega_p = \frac{1}{C R_{VC}} = \frac{R_1 + R_2}{R_1 R_2 C}.$$

Il polo trovato coincide con quello che avevamo determinato con la precedente procedura. Osserviamo che per frequenza tendente all'infinito $A(j\omega)$ tende all'unità perché il condensatore diventa un corto circuito ($A(\infty) = 1$). Se il valore all'infinito della funzione di trasferimento è diverso da zero, il numero di zeri deve essere uguale al numero dei poli (se invece il numero di zeri fosse minore di quello dei poli, la funzione di trasferimento tenderebbe a zero per frequenza infinita). Quindi, l'espressione completa della funzione di trasferimento risulta:

$$A(s) = A(\infty) \frac{s + \omega_z}{s + \omega_p},$$

dove il coefficiente moltiplicativo di fronte alla frazione è proprio $A(\infty)$, essendo la frazione scritta nella forma con coefficienti unitari dei termini in s .

Notiamo inoltre che, per frequenza nulla,

$$A(0) = A(\infty) \frac{\omega_z}{\omega_p},$$

quindi esiste una relazione che lega tra loro il valore della funzione di trasferimento a zero, quello a infinito e i valori del polo e dello zero. In particolare, possiamo determinare il valore dello zero a partire da quello del polo e dai due guadagni (a zero e a infinito):

$$\omega_z = \omega_p \frac{A(0)}{A(\infty)} = \frac{R_1 + R_2}{R_1 R_2 C} \frac{R_2}{R_1 + R_2} = \frac{1}{R_1 C},$$

che coincide con quanto trovato con il metodo visto precedentemente. Abbiamo quindi tutti gli elementi per scrivere l'espressione della funzione di trasferimento, determinati in modo molto più semplice che attraverso la procedura algebrica vista prima. È importante sottolineare che la differenza di complessità tra la procedura “per ispezione” appena vista e quella algebrica precedente diventa molto più grande non appena il numero di componenti nel circuito aumenta e questo comprende anche generatori comandati.

Vediamo infine un metodo per determinare lo zero, alternativo a quello basato sul rapporto tra il guadagno a zero e quello all'infinito. Osserviamo che in corrispondenza dello zero la funzione di trasferimento deve annullarsi: questo si verifica se l'impedenza formata dal parallelo di R_1 e C assume valore infinito. Ciò non può verificarsi per alcun valore della frequenza fisica, ma avviene per un opportuno valore della frequenza generalizzata s :

$$Z(s) = R_1 // \frac{1}{Cs} = \frac{R_1}{R_1Cs + 1},$$

quindi avrà valore infinito quando il denominatore tende a zero, $R_1Cs + 1 = 0$, da cui si ricava che $s_z = -1/(R_1C)$ e, dunque, che $\omega_z = 1/(R_1C)$.

La scelta dell'una o dell'altra tecnica per la determinazione dello zero dipende dal circuito considerato: talvolta può essere molto semplice calcolare i valori del guadagno a zero e all'infinito e complicato invece valutare lo zero per ispezione; altre volte la situazione può essere rovesciata. In alcuni casi può essere semplice calcolare lo zero per ispezione e i guadagni a zero e all'infinito: in tal caso si può fare a meno di calcolare la resistenza vista e valutare il polo a partire dallo zero e dai due guadagni. Inoltre, se ci sono più zeri e più poli, può essere utile applicare ambedue i metodi, in modo da determinare tutti gli zeri fuorché uno per ispezione e l'ultimo sfruttando la conoscenza dei poli e del guadagno a zero e a infinito.

9.7 Diagrammi di Bode

In questo paragrafo presenteremo una trattazione semplificata dei diagrammi di Bode, che verranno poi esaminati più in dettaglio in altri corsi. I diagrammi di Bode offrono una tecnica conveniente per la rappresentazione dell'andamento del modulo e della fase di una funzione di trasferimento, in funzione della frequenza “fisica”. Nella nostra trattazione considereremo soltanto il caso di poli e zeri reali, trascurando quindi lo studio del caso, peraltro abbastanza comune nella pratica, di poli o zeri complessi coniugati.

Il modulo della funzione di trasferimento viene espresso in decibel, che corrispondono a dieci volte il logaritmo in base dieci del rapporto tra il quadrato del valore del modulo stesso e un valore di riferimento, che nel caso della funzione di trasferimento corrisponde all'unità. Il motivo per cui si considera il modulo quadrato della funzione di trasferimento sta nel fatto che si indicano in decibel i rapporti di potenza e in questo caso ci si riferisce a una potenza convenzionale, proporzionale al quadrato dell'ampiezza delle grandezze (tensione o corrente) di ingresso e di uscita. Si parla di potenza convenzionale perché la potenza reale dipenderebbe dalle effettive resistenze sulle quali le tensioni (o correnti) di ingresso e di uscita sono applicate, e quindi anche i valori delle resistenze comparirebbero nel rapporto. Quanto detto può essere sintetizzato nella semplice espressione

$$|A(j\omega)|_{\text{dB}} = 10 \log_{10} |A(j\omega)|^2 = 20 \log_{10} |A(j\omega)|,$$

che ci fornisce la relazione tra il valore del modulo della funzione di trasferimento e la sua rappresentazione in decibel.

Data la struttura precedentemente indicata per $A(j\omega)$, si ha, ricordando che il logaritmo di un prodotto di due o più fattori corrisponde alla somma dei logaritmi di ciascuno di tali fattori:

$$\begin{aligned} |A(j\omega)|_{(\text{dB})} = & 20 \log_{10} |a_m| + 20 \log_{10} \sqrt{\omega^2 + \omega_{z_1}^2} + 20 \log_{10} \sqrt{\omega^2 + \omega_{z_2}^2} + \dots \\ & + 20 \log_{10} \sqrt{\omega^2 + \omega_{z_m}^2} - 20 \log_{10} \sqrt{\omega^2 + \omega_{p_1}^2} \\ & - 20 \log_{10} \sqrt{\omega^2 + \omega_{p_2}^2} - \dots - 20 \log_{10} \sqrt{\omega^2 + \omega_{p_n}^2}. \end{aligned}$$

Con un'ulteriore manipolazione questa espressione può essere scritta nella forma

$$\begin{aligned} |A(j\omega)|_{(\text{dB})} = & 20 \log_{10} |a_m| + 20 \log_{10} \frac{|\omega_{z_1}| |\omega_{z_2}| \dots |\omega_{z_m}|}{|\omega_{p_1}| |\omega_{p_2}| \dots |\omega_{p_n}|} \\ & + 20 \log_{10} \sqrt{1 + \frac{\omega^2}{\omega_{z_1}^2}} + 20 \log_{10} \sqrt{1 + \frac{\omega^2}{\omega_{z_2}^2}} + \dots + 20 \log_{10} \sqrt{1 + \frac{\omega^2}{\omega_{z_m}^2}} \\ & - 20 \log_{10} \sqrt{1 + \frac{\omega^2}{\omega_{p_1}^2}} - 20 \log_{10} \sqrt{1 + \frac{\omega^2}{\omega_{p_2}^2}} - \dots - 20 \log_{10} \sqrt{1 + \frac{\omega^2}{\omega_{p_n}^2}}. \end{aligned}$$

È perciò possibile tracciare il grafico del modulo di $A(j\omega)$ (in scala logaritmica, e quindi lineare in dB) come somma di contributi tutti dello stesso tipo, della forma

$$K_i(j\omega) = 20 \log_{10} \sqrt{1 + \frac{\omega^2}{\omega_i^2}}.$$

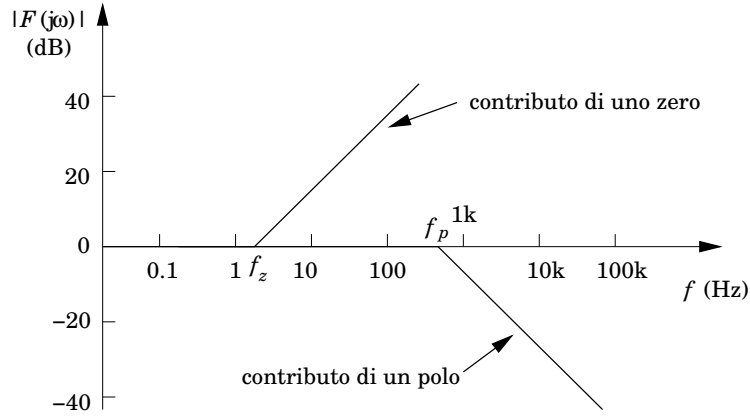
Notiamo che per $\omega \ll \omega_i$ il secondo termine sotto radice è trascurabile rispetto all'unità, quindi $K_i(j\omega) \simeq 0$. Se invece $\omega \gg \omega_i$, l'unità è trascurabile e il termine $K_i(j\omega)$ può essere approssimato nel modo seguente

$$K_i(j\omega) \simeq 20 \log_{10} \frac{|\omega|}{|\omega_i|}.$$

Poiché i diagrammi di Bode si tracciano su una scala doppiamente logaritmica (logaritmica sia sull'asse delle ascisse sia su quello delle ordinate), il termine $K_i(j\omega)$ darà, per $\omega \gg \omega_i$, una semiretta con pendenza di 20 dB/dec, cioè con un'inclinazione tale da determinare un aumento di 20 dB per ogni variazione della frequenza corrispondente a un fattore 10. Per $\omega \ll \omega_i$, invece, il termine $K(j\omega)$ sarà rappresentato da una semiretta orizzontale, coincidente con l'asse delle ascisse. Complessivamente quindi il diagramma asintotico di Bode (detto asintotico perché tendente a essere esatto nei due limiti di $\omega \gg \omega_i$ e $\omega \ll \omega_i$) è costituito da una spezzata, formata da due semirette che si incontrano nel punto $\omega = \omega_i$. I termini corrispondenti ai poli avranno un analogo diagramma asintotico, con l'unica differenza che la pendenza per $\omega \gg \omega_i$ sarà negativa invece che positiva:

Il termine

$$20 \log_{10} |a_m| + 20 \log_{10} \frac{|\omega_{z_1}| |\omega_{z_2}| \dots |\omega_{z_m}|}{|\omega_{p_1}| |\omega_{p_2}| \dots |\omega_{p_n}|}$$



rappresenta una semplice traslazione verticale del diagramma risultante dai contributi dei vari poli e zeri. Notiamo che nel punto angoloso corrispondente a $\omega = \omega_i$ l'effettivo valore del termine $K(j\omega)$ risulta

$$K(j\omega_i) = 20 \log_{10} \sqrt{1 + \frac{\omega_i}{\omega_i}} = 20 \log_{10} \sqrt{2} = 3 \text{ dB}.$$

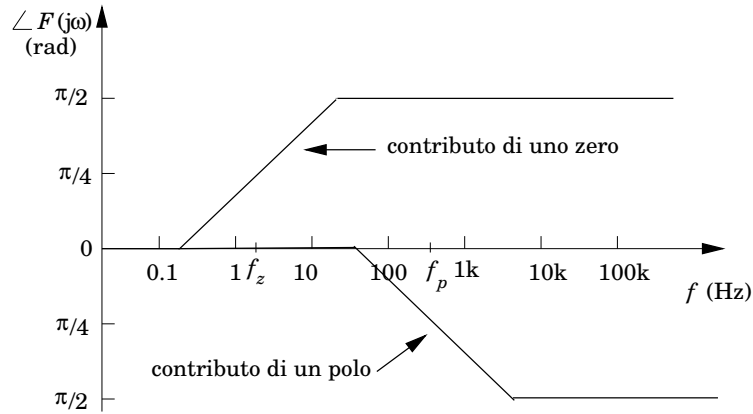
Quindi, in corrispondenza delle frequenze di zero il diagramma reale si trova 3 dB al di sopra di quello asintotico e, analogamente, in corrispondenza delle frequenze di polo il diagramma reale si trova 3 dB al di sotto di quello asintotico.

Per quanto riguarda invece il diagramma di fase, rappresentato in scala logaritmica per le frequenze, ma lineare per gli angoli, espressi in radianti, avremo innanzitutto un termine additivo π nel caso in cui il coefficiente a_m sia negativo e ciascun termine del tipo $(j\omega + \omega_i)$ darà un contributo corrispondente alla propria fase: $\arctg(\omega/\omega_i)$. In realtà dovremmo considerare una arcotangente a quattro quadranti; se utilizziamo una arcotangente come quella comunemente impiegata, definita nel primo e nel quarto quadrante, dobbiamo aggiungere un termine π per ogni zero o polo positivo. Complessivamente, quindi, abbiamo

$$\begin{aligned} \angle A(j\omega) = & \frac{1 - \text{sgn}(a_m)}{2} \pi + \arctg\left(\frac{\omega}{\omega_{z_1}}\right) + \arctg\left(\frac{\omega}{\omega_{z_2}}\right) + \dots + \arctg\left(\frac{\omega}{\omega_{z_m}}\right) \\ & - \arctg\left(\frac{\omega}{\omega_{p_1}}\right) - \arctg\left(\frac{\omega}{\omega_{p_2}}\right) - \dots - \arctg\left(\frac{\omega}{\omega_{p_n}}\right) + \pi(m' + n'), \end{aligned}$$

dove m' è il numero di zeri nel semipiano positivo (corrispondenti a ω_z negative) e n' è il numero di poli nel semipiano positivo (corrispondenti a ω_p negative). Nel diagramma di fase di Bode, ciascuno dei contributi dovuti a un termine del tipo $\arctg(\omega/\omega_i)$ può essere rappresentato come nullo per $\omega \ll \omega_i$ e pari a $\pi/2$ per $\omega \gg \omega_i$. Il diagramma asintotico si traccia considerando il contributo nullo fino a $\omega_i/10$ e pari a $\pi/2$ al di sopra di $10\omega_i$; nell'intervallo tra $\omega_i/10$ e $10\omega_i$ si assume una variazione lineare con pendenza dunque di $(\pi/4)$ rad/dec. Il contributo di fase dei poli è esattamente analogo, con l'unica differenza che il valore asintotico per $\omega \gg \omega_i$ è di $-\pi/2$. Quindi il contributo di fase di eventuali zeri a parte reale positiva è uguale a quello dei poli a parte reale negativa.

Se la funzione di trasferimento presenta uno o più zeri nell'origine, la funzione di trasferimento stessa risulta nulla nell'origine e il diagramma di ampiezza inizia con una pendenza di $n \times 20$ dB/dec, dove n è il numero di zeri nell'origine. In



questo caso si dovrà tracciare il diagramma a ritroso, partendo da una frequenza in corrispondenza della quale il diagramma è piatto e se ne conosce il valore. Dal punto di vista della fase, ogni zero nell'origine dà un contributo pari a $\pi/2$ rad, da aggiungere alla fase totale.

9.8 Esempio di tracciamento di diagrammi di Bode

Prendiamo in considerazione qualche semplice caso per il tracciamento dei diagrammi di Bode. Supponiamo di avere una funzione di trasferimento del tipo

$$A(s) = a_m \frac{s + \omega_z}{s + \omega_p},$$

con $a_m = 50$, $f_z = \omega_z/(2\pi) = 100$ Hz e $f_p = \omega_p/(2\pi) = 2$ kHz. Per determinare la posizione di f_p sull'asse delle ascisse basta calcolare il rapporto tra f_p e il valore di frequenza corrispondente alla suddivisione immediatamente inferiore a f_p , e farne poi il logaritmo in base 10:

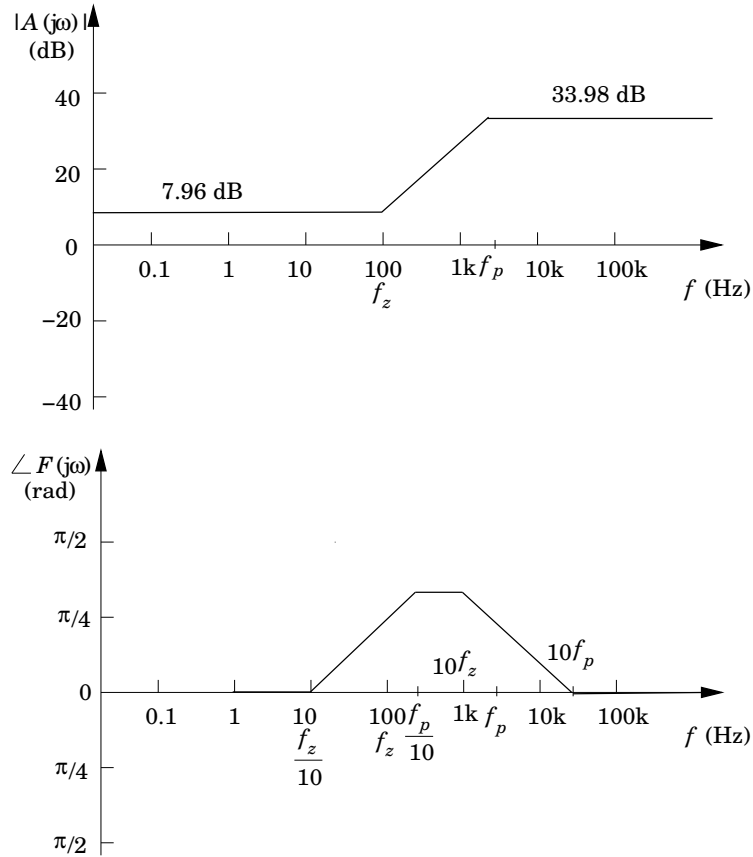
$$x = \log_{10} \frac{f_p}{1 \text{ kHz}} = 0.3.$$

Questo significa che f_p si troverà a una frazione 0.3 della distanza tra la tacca a 1 kHz e quella a 10 kHz. Il guadagno a frequenza infinita è pari ad a_m e quindi risulta, in dB, $|A(\infty)|_{\text{dB}} = 33.98$ dB. Possiamo quindi iniziare a tracciare il diagramma a partire dalle frequenze alte, spostandoci verso il basso: tra la frequenza di zero e quella di polo abbiamo una pendenza di 20 dB/dec, corrispondente al contributo dello zero, che viene poi neutralizzato da quello del polo. Al di sotto dello zero il diagramma è nuovamente piatto e ha un valore $|A(0)|$ che può essere facilmente calcolato:

$$|A(0)| = |A(\infty)| \frac{|\omega_z|}{|\omega_p|} = 2.5.$$

Quindi $|A(0)|_{\text{dB}} = 7.96$ dB.

Tracciamo ora il diagramma di fase: possiamo partire dal fatto che a basse frequenze la fase è nulla, finché non raggiungiamo un decimo della frequenza di zero. A questo punto la fase comincia a crescere finché non si arriva a un decimo della frequenza di polo, dove la fase si stabilizza, perché l'azione del polo compensa quella dello zero. Tale compensazione sussiste fino a una frequenza pari a dieci volte quella di zero, al di

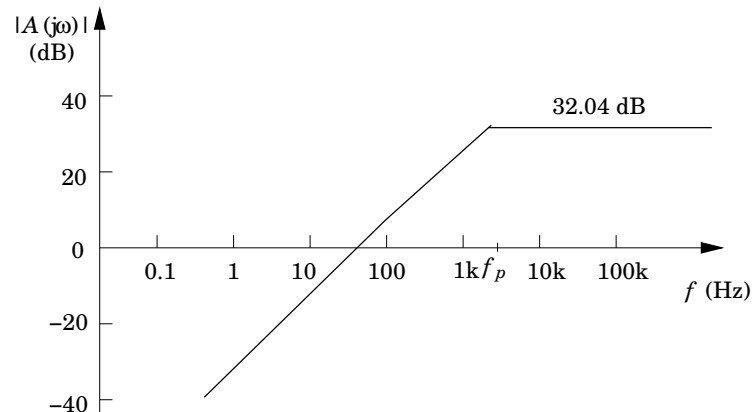


sopra della quale la fase comincia a scendere con una pendenza di $-(\pi/4)/\text{dec}$, finché non si esaurisce anche l'azione del polo, a una frequenza pari a dieci volte quella di polo e il diagramma torna a zero, come è facile prevedere dall'analisi della struttura di questa particolare funzione di trasferimento.

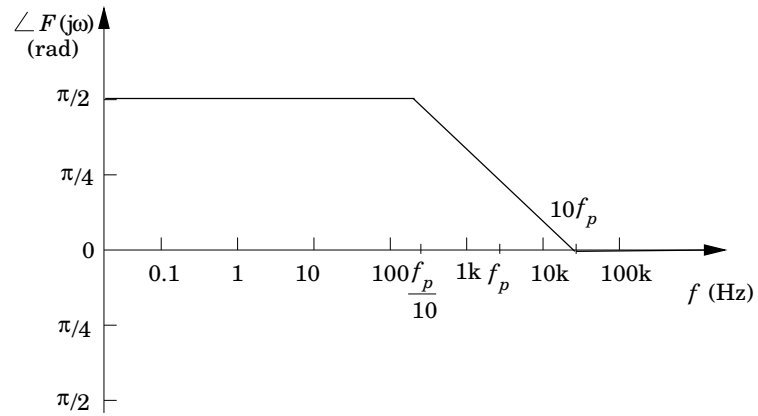
Consideriamo ora il caso di una funzione di trasferimento con uno zero nell'origine:

$$A(s) = a_m \frac{s}{s + \omega_p},$$

con $a_m = 40$, $f_p = \omega_p/(2\pi) = 2$ kHz. Procediamo in modo analogo a prima: poiché conosciamo il valore del modulo della funzione di trasferimento per frequenza infinita, iniziamo a tracciare il diagramma a ritroso, partendo da $|A(\infty)|_{\text{dB}} = 32.04$ dB e procedendo poi con una pendenza di 20 dB/dec al di sotto di f_p .

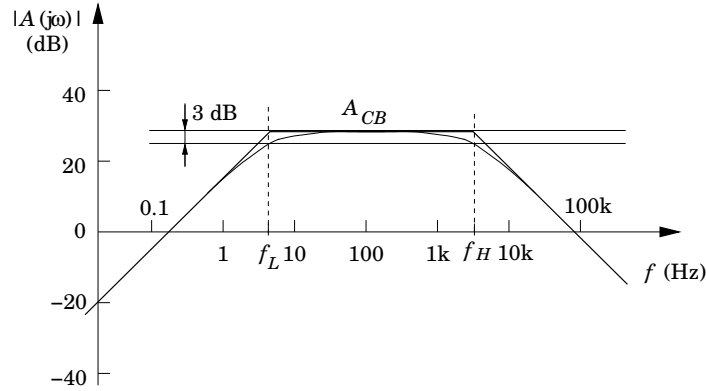


Per quanto riguarda la fase, possiamo partire dalle frequenze basse, dove sappiamo che lo zero nell'origine dà un contributo pari a $\pi/2$, poi, in corrispondenza di $f_p/10$, inizia l'azione del polo, che fa scendere la fase con una pendenza di $-(\pi/4)/\text{dec}$ fino a $10f_p$, dove si annulla e rimane nulla per tutte le frequenze superiori.



9.9 Funzione di trasferimento degli amplificatori

Nella maggior parte degli amplificatori utilizzati in pratica il guadagno risulta costante su un certo intervallo di frequenze, definito “banda passante” dell’amplificatore stesso. Il guadagno all’interno di tale banda si definisce guadagno a centro banda. Il limite inferiore di banda (f_L) si definisce come la frequenza in corrispondenza della quale il guadagno si discosta di 3 dB da quello a centro banda e analogamente viene definito il limite superiore di banda (f_H).



Se i poli e gli zeri che determinano il comportamento dell’amplificatore ad alta frequenza sono ben separati (almeno un paio di decadi) da quelli che caratterizzano il comportamento a bassa frequenza, la funzione di trasferimento dell’amplificatore può essere scritta nella forma

$$A(s) = A_M F_L(s) F_H(s),$$

dove $F_L(s)$ e $F_H(s)$ sono funzioni che rappresentano il comportamento di $A(s)$, rispettivamente, alle basse e alle alte frequenze. Per frequenze molto maggiori di f_L , la funzione $F_L(s)$ è approssimativamente pari all’unità, mentre lo stesso accade per $F_H(s)$ a frequenze molto minori di f_H . Quindi per $f_L \ll f \ll f_H$ il guadagno è sostanzialmente pari a A_M .

Alle basse frequenze possiamo dunque esprimere il guadagno nella forma $A_L(s) \simeq A_M F_L(s)$, mentre alle alte frequenze questo è approssimato da $A_H(s) \simeq A_M F_H(s)$. Tale suddivisione delle bande è valida per tutti quegli amplificatori che presentano un centro banda abbastanza esteso. Per tali amplificatori $F_L(s)$ è in genere determinata dai condensatori di accoppiamento e di bypass, mentre $F_H(s)$ è determinata dai condensatori interni dei dispositivi attivi (che esamineremo in dettaglio nel seguito).

Quindi per la determinazione di A_M si considerano i condensatori di accoppiamento e di bypass come corto circuiti e i condensatori interni dei dispositivi come circuiti aperti (a centro banda $F_H(s) \simeq 1$ e $F_L(s) \simeq 1$). Per l’analisi di $F_L(s)$ consideriamo invece l’azione dei condensatori di accoppiamento e di bypass, supponendo che quelli interni dei dispositivi siano circuiti aperti (alle basse frequenze $F_H(s) \simeq 1$). Infine, per l’analisi di $F_H(s)$ consideriamo l’azione dei condensatori interni dei dispositivi, considerando quelli di accoppiamento e di bypass come corto circuiti (alle alte frequenze $F_L(s) \simeq 1$).

Se vogliamo scrivere la funzione di trasferimento nella forma di un fattore moltiplicativo corrispondente all’amplificazione a centro banda che moltiplica $F_L(s)$ e $F_H(s)$, dovremo indicare queste ultime in una forma tale da soddisfare le condizioni prima indicate, in particolare il fatto che ad alta frequenza $F_L \simeq 1$ e a bassa frequenza $F_H \simeq 1$.

È immediato verificare che se scriviamo $F_L(s)$ nella forma di un rapporto di polinomi con coefficiente unitario per i termini di grado massimo al numeratore e al denominatore, la condizione $F_L \simeq 1$ alle alte frequenze sarà immediatamente soddisfatta:

$$F_L(s) = \frac{(s + \omega_{z_1})(s + \omega_{z_2}) \cdots (s + \omega_{z_{m'}})}{(s + \omega_{p_1})(s + \omega_{p_2}) \cdots (s + \omega_{p_{n'}})},$$

dove m' e n' sono, rispettivamente, il numero di zeri e il numero di poli della $F_L(s)$. Si noti che, se $F_L(s)$ deve soddisfare la condizione che per frequenza molto grande risulti circa unitaria, dovrà sempre essere $m' = n'$, cioè la $F_L(s)$ avrà tanti zeri quanti poli. Succede talvolta (abbastanza di frequente nella pratica) che un polo sia molto più alto di tutti gli altri poli e degli zeri: in tal caso la $F_L(s)$ può essere approssimata nella forma

$$F_L(s) \simeq \frac{s}{s + \omega_{p_1}},$$

dove ω_{p_1} è la pulsazione relativa al polo più alto (tale relazione si trova in modo immediato, considerando tutti gli altri poli e zeri trascurabili rispetto a s). In tal caso si parla di approssimazione a polo dominante e il limite inferiore di banda (che, ricordiamo, corrisponde alla frequenza per cui il guadagno si discosta di 3 dB da quello a centro banda) risulta proprio pari alla frequenza di tale polo, $f_L \simeq f_{p_1}$, poiché in presenza di uno zero nell'origine e di un unico polo il diagramma di Bode reale si trova 3 dB sotto quello asintotico proprio in corrispondenza del polo. Una regola empirica per decidere se un polo sia o meno dominante nella risposta alle basse frequenze consiste nel verificare che tutti gli altri poli e gli zeri della $F_L(s)$ distino almeno di due ottave dal polo più alto. In altri termini, si può considerare un polo dominante nella risposta alle basse frequenze se tutti gli altri poli e gli zeri si trovano a frequenze minori di un quarto della frequenza di tale polo.

Se la condizione sopra detta di polo dominante non è verificata, possiamo utilizzare una formula approssimata per determinare il limite inferiore di banda a partire dalla conoscenza dei poli e degli zeri della $F_L(s)$:

$$f_L \simeq \sqrt{f_{p_1}^2 + f_{p_2}^2 + \cdots + f_{p_{n'}}^2 - 2f_{z_1}^2 - 2f_{z_2}^2 - \cdots - 2f_{z_{m'}}^2}.$$

In tale formula si può chiaramente fare a meno di inserire poli e zeri molto bassi rispetto agli altri, dato che il loro peso relativo sarebbe comunque trascurabile.

Nello scrivere, invece, la $F_H(s)$ come rapporto di polinomi dobbiamo tener conto che per frequenza tendente a zero tale rapporto dovrà risultare unitario. È quindi necessario che i termini di grado zero siano unitari, in modo che per s nulla si abbia $F_H(0) = 1$:

$$F_H(s) = \frac{\left(1 + \frac{s}{\omega_{z'_1}}\right) \left(1 + \frac{s}{\omega_{z'_2}}\right) \cdots \left(1 + \frac{s}{\omega_{z'_{m''}}}\right)}{\left(1 + \frac{s}{\omega_{p'_1}}\right) \left(1 + \frac{s}{\omega_{p'_2}}\right) \cdots \left(1 + \frac{s}{\omega_{p'_{n''}}}\right)},$$

dove le $\omega_{z'_i}$ sono gli m'' zeri e le $\omega_{p'_i}$ sono gli n'' poli ad alta frequenza. Notiamo che in questo caso non è necessario che sia $n'' = m''$, quindi l'unica condizione è quella di fisica realizzabilità ($m'' \leq n''$) e di solito $m'' < n''$, poiché al tendere della frequenza all'infinito la risposta della maggior parte degli amplificatori tende a zero.

Anche per la $F_H(s)$ può sussistere una condizione di polo dominante, analoga a quella che abbiamo incontrato per la $F_L(s)$: in questo caso un polo si dice dominante

se è molto più basso di tutti gli altri poli e gli zeri ad alta frequenza (vale a dire ne dista almeno di due ottave). In altre parole, un polo si dice dominante se tutti gli altri poli e gli zeri sono a frequenza almeno quadrupla di tale polo. Nella condizione di polo dominante la $F_H(s)$ può essere approssimata nella forma

$$F_H(s) \simeq \frac{1}{1 + \frac{s}{\omega_{p'_1}}},$$

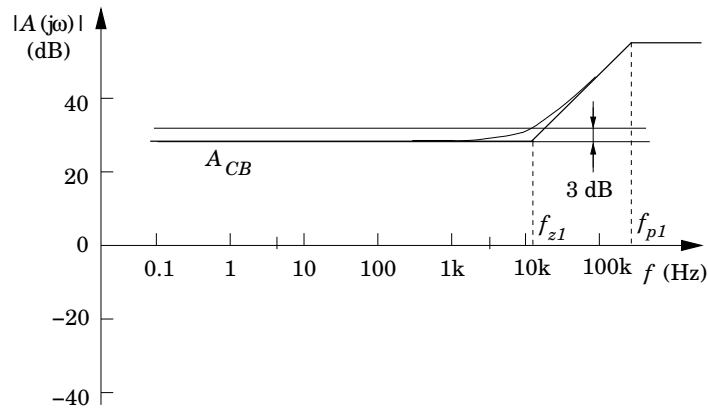
dove $\omega_{p'_1}$ è la pulsazione associata al polo dominante. In questo caso, in analogia a quanto visto per il comportamento a bassa frequenza, il limite superiore di banda si trova in corrispondenza del polo dominante: $f_H = f_{p'_1}$.

Nel caso in cui non vi sia un polo dominante, vale a dire che gli altri poli e gli zeri non distino dal polo più basso almeno due ottave, si può calcolare il limite superiore di banda con una formula approssimata:

$$f_H \simeq \left[\sqrt{\frac{1}{f_{p'_1}^2} + \frac{1}{f_{p'_2}^2} + \dots + \frac{1}{f_{p'_{n''}}^2} - \frac{2}{f_{z'_1}^2} - \frac{2}{f_{z'_2}^2} - \dots - \frac{2}{f_{z'_{m''}}^2}} \right]^{-1}.$$

Se vi sono poli o zeri molto alti rispetto al polo più basso della $F_H(s)$, questi possono essere trascurati nel valutare l'espressione sopra riportata, dato che, comunque, il loro contributo non sarebbe significativo.

È importante precisare che le formule indicate per il calcolo dei limiti di banda sono in generale approssimate e sono applicabili soltanto nel caso di una funzione di trasferimento con guadagno massimo a centro banda e decrescente alle basse e alle alte frequenze. Se si ha, per esempio, una funzione di trasferimento come quella indicata nella figura seguente, il limite superiore di banda è rappresentato dallo zero f_{z_1} (e non dal polo!), dato che in corrispondenza dello stesso il modulo della funzione di trasferimento è 3 dB al di sopra del valore di centro banda (i limiti di banda si hanno in corrispondenza delle frequenze per cui il modulo della funzione di trasferimento si discosta di 3 dB, in più o in meno, dal valore di centro banda). Se il polo f_{p_1} fosse più vicino di due ottave a f_{z_1} , sarebbe necessario svolgere il calcolo esatto del modulo per ottenere l'effettivo limite superiore di banda, dato che gli effetti del polo e dello zero non sarebbero più separabili.



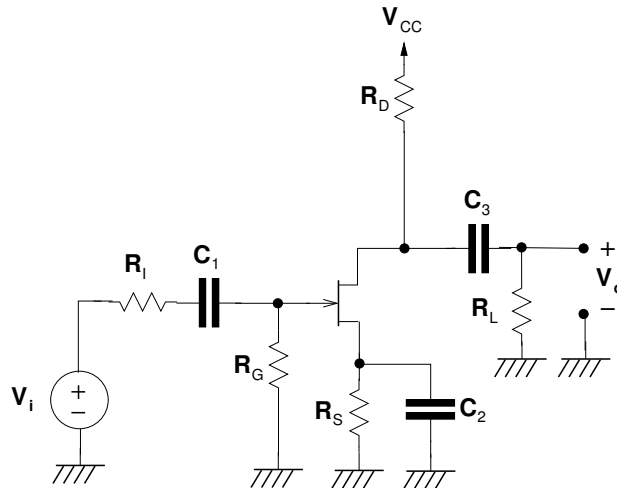
In particolare, se nell'esempio della figura il polo e lo zero sono molto vicini, tali che $f_{z_1}/f_{p_1} < \sqrt{2}$, non c'è alcun limite superiore di banda, dato che il modulo della funzione di trasferimento non si discosta, per alcuna frequenza, più di 3 dB dal valore di centro banda.

9.10 Poli non interagenti

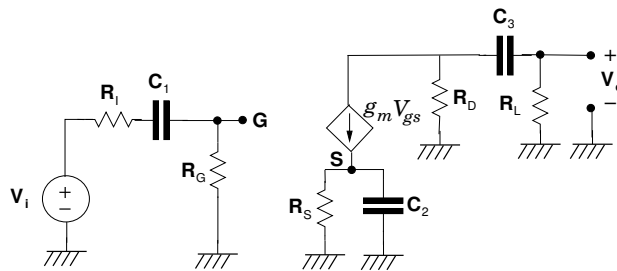
Abbiamo già visto che se un circuito contiene un solo elemento reattivo il polo può essere determinato in modo molto semplice utilizzando il metodo della resistenza vista. Nel caso in cui ci siano più condensatori, il metodo della resistenza vista può tornare ugualmente utile, purché tali condensatori siano “non interagenti”. Due condensatori si definiscono non interagenti quando la resistenza vista da uno dei due non varia se all’altro si sostituisce un corto circuito o un circuito aperto. Una definizione alternativa e equivalente potrebbe essere quella che un condensatore non è interagente con altri quando l’impedenza vista dallo stesso è puramente resistiva; infatti se l’impedenza vista è puramente resistiva significa che nel calcolo della stessa non appaiono in alcun modo gli altri condensatori.

È abbastanza facile immaginare che se i condensatori sono non interagenti il polo associato a ciascuno di essi dipende soltanto dalla resistenza vista e ciascun polo può essere calcolato trattando solo la parte di circuito da esso “vista”, come se si trattasse di un circuito a sé stante. Quanto detto semplifica notevolmente l’analisi di molti circuiti, perché ci permette di calcolare i poli con un semplice procedimento di ispezione.

Vediamo un esempio di applicazione del metodo di ispezione, nella determinazione della funzione di trasferimento alle basse e medie frequenze di un amplificatore a source comune, di cui riportiamo di seguito lo schema.



Il circuito equivalente per le variazioni alle basse e medie frequenze si ottiene sostituendo il transistor JFET con il suo equivalente dinamico e cortocircuitando il generatore di alimentazione:



Notiamo subito che i tre condensatori sono tra loro non interagenti, infatti C_1 è separato dal resto della rete a causa dell’impedenza infinita vista sul gate del FET, mentre

C_3 risulta anch'esso separato dal resto perché sull'uscita del generatore comandato di corrente si vede un'impedenza anch'essa infinita. Possiamo quindi procedere al calcolo dei poli utilizzando il metodo della resistenza vista.

La resistenza vista da C_1 è pari a $R_G + R_I$, quindi $\omega_{p_1} = 1/[C_1(R_G + R_I)]$. Il condensatore C_2 vede la resistenza R_S in parallelo con quella vista sul source del FET, che, come sappiamo, è pari a $1/g_m$, quindi $R_{V_{C_2}} = R_S // 1/g_m$, e $\omega_{p_2} = 1/[C_2(R_S // 1/g_m)]$. Infine, la resistenza vista da C_3 è pari a $R_D + R_L$, quindi $\omega_{p_3} = 1/[C_3(R_D + R_L)]$.

Per ispezione possiamo anche determinare gli zeri: ci sono due zeri nell'origine, dato che C_1 e C_3 si trovano in serie sul percorso del segnale. Il numero totale di zeri deve essere pari a quello dei poli, dato che a frequenze tali da poter considerare tutti i condensatori come corto circuiti il guadagno, corrispondente al guadagno a centro banda, è diverso da zero. Il terzo zero può essere determinato semplicemente, osservando che l'uscita si annulla per il valore di frequenza generalizzata per il quale l'impedenza formata dal parallelo di R_S e C_2 diventa infinita, dato che in tale condizione la corrente $g_m V_{gs}$ diventa nulla. Quindi $\omega_{z_3} = 1/(R_S C_2)$.

Il guadagno a centro banda può essere calcolato considerando tutti i condensatori in corto circuito. Notiamo che, essendo l'impedenza di ingresso del JFET infinita, possiamo partizionare con vantaggio la funzione di trasferimento nel prodotto di due funzioni di trasferimento distinte:

$$A(s) = \frac{V_o}{V_i} = \frac{V_o}{V_g} \frac{V_g}{V_i}.$$

Dato che $V_o = -g_m V_g R_D // R_L$,

$$\frac{V_o}{V_g} = -g_m R_D // R_L.$$

Inoltre V_g è una partizione di V_i :

$$\frac{V_g}{V_i} = \frac{R_G}{R_G + R_I}.$$

Quindi il guadagno a centro banda risulta

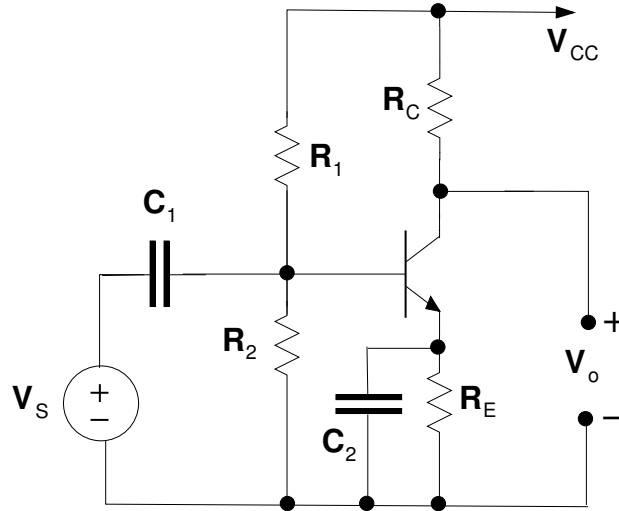
$$A_{CB} = -\frac{R_G}{R_G + R_I} g_m R_D // R_L.$$

Infine, la funzione di trasferimento complessiva può scriversi nella forma

$$A(s) = A_{CB} \frac{s^2(s + \omega_{z_3})}{(s + \omega_{p_1})(s + \omega_{p_2})(s + \omega_{p_3})}.$$

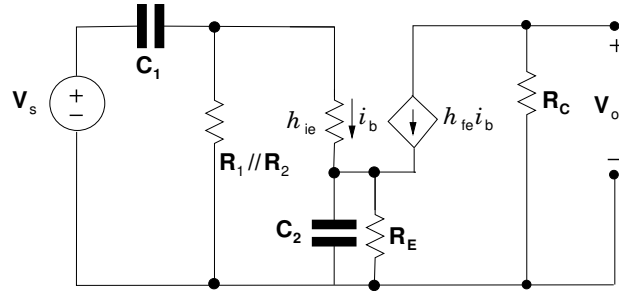
9.11* Risposta alle basse e medie frequenze dello stadio a emettitore comune

Nel caso del circuito a emettitore comune con resistenza di emettitore, riportato nello schema sottostante, possiamo trovarci di fronte a una situazione nella quale il condensatore di accoppiamento C_1 e quello di bypass C_2 non sono non interagenti, infatti la resistenza vista, per esempio, da C_1 vale $R_1 // R_2 // h_{ie}$ se consideriamo C_2 un corto



circuito e, invece, $R_1 // R_2 // [h_{ie} + R_E(h_{fe} + 1)]$ se consideriamo C_2 un condensatore aperto.

È quindi evidente che non possiamo applicare in modo rigoroso il metodo della resistenza vista. Se però immaginiamo che i valori dei due condensatori siano dello stesso ordine di grandezza e notiamo che le resistenze viste, indipendentemente dal fatto che siano calcolate con l'altro condensatore aperto o in corto circuito, differiscono notevolmente tra i due condensatori, possiamo fare un'ipotesi semplificativa che ci consente di utilizzare, anche se in modo approssimato, il metodo della resistenza vista. Consideriamo il circuito equivalente, di seguito rappresentato e valutiamo la resistenza vista da C_2 : se consideriamo C_1 un corto circuito, la resistenza vista è semplicemente $R_E // [h_{ie} / (h_{fe} + 1)]$; se invece consideriamo C_1 un circuito aperto, la resistenza vista da C_2 risulta $R_E // [(h_{ie} + R_1 // R_2) / (h_{fe} + 1)]$. Quindi, comunque si consideri di sostituire C_1 , la resistenza vista da C_2 è dell'ordine delle decine di ohm, mentre quella vista di C_1 , comunque si consideri la situazione di C_2 , risulta dell'ordine di alcuni kilohm o di alcune decine di kilohm.



È dunque ragionevole, se i condensatori sono effettivamente dello stesso ordine di grandezza, supporre che la frequenza di polo relativa a C_1 sia molto più bassa di quella associata a C_2 . Possiamo perciò ipotizzare che alla frequenza di polo associata a C_1 il condensatore C_2 possa essere sostituito con un circuito aperto, perché è ancora ben lontano dall'agire; viceversa alla frequenza di polo associata a C_2 il condensatore C_1 possa essere considerato un corto circuito, dato che il relativo polo interviene a frequenza molto più bassa.

Sulla base di queste considerazioni, il polo relativo a C_1 sarà approssimabile con

$$\omega_{p1} = \frac{1}{C_1 \{R_1 // R_2 // [h_{ie} + R_E(h_{fe} + 1)]\}}$$

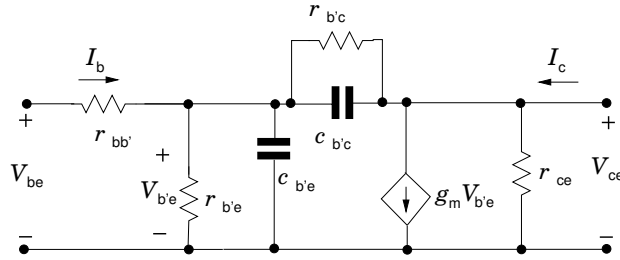
e quello relativo a C_2 con

$$\omega_{p_2} = \frac{1}{C_2 \{R_E // [h_{ie} / (h_{fe} + 1)]\}}.$$

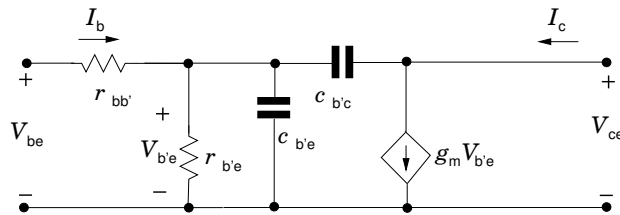
Si possono poi confrontare i valori numerici così ottenuti per verificare se l'ipotesi iniziale $\omega_{p_1} \ll \omega_{p_2}$ è soddisfatta. Se questo non accade, certamente la procedura seguita non è valida; nel caso invece che tale condizione sia verificata non abbiamo però la certezza assoluta che l'approssimazione fatta sia corretta, perché non conosciamo a priori il segno e l'entità dell'errore commesso sui poli in conseguenza dell'approssimazione stessa. Da un punto di vista pratico possiamo dire che negli stadi a emettitore comune di solito realizzati l'approssimazione descritta è in genere valida, purché i condensatori abbiano valori dello stesso ordine di grandezza.

9.12 Circuito equivalente del transistor BJT alle alte frequenze

Alle alte frequenze il comportamento di un transistor BJT è riprodotto con soddisfacente approssimazione da un circuito equivalente un po' più complesso di quelli visti fino a questo punto e contenente le capacità associate alle giunzioni pn . In particolare, abbiamo una capacità di diffusione $c_{b'e}$ relativa alla giunzione base-emettitore polarizzata direttamente e una capacità di transizione $c_{b'c}$ relativa alla giunzione base-collettore polarizzata inversamente. Lo schema risultante è quello detto a π ibrido o di Giacoletto



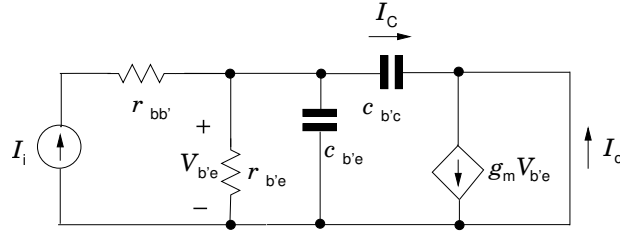
Di solito si trascurano $r_{b'c}$ e r_{ce} , ottenendo quindi il seguente circuito semplificato:



Calcoliamo l'amplificazione di corrente di uno stadio CE con l'uscita in corto circuito. È evidente dallo schema di seguito riportato che a frequenza nulla il rapporto $A_I(0)$ tra la corrente I_o e la corrente I_i è pari a h_{fe} ; infatti

$$I_o = g_m V_{b'e} = g_m r_{b'e} I_i = h_{fe} I_i.$$

Notiamo anche che per frequenza tendente all'infinito la tensione $V_{b'e}$ viene cortocircuitata dal condensatore $c_{b'e}$, determinando quindi la disattivazione del generatore



$g_m V_{b'e}$: in questo caso la I_o risulta dunque data da una partizione della I_i sul partitore formato dai due condensatori $c_{b'c}$ e $c_{b'e}$. Abbiamo quindi (si noti il cambiamento di segno rispetto a $A_I(0)$):

$$A_I(\infty) = -\frac{c_{b'c}}{c_{b'c} + c_{b'e}}.$$

Notiamo inoltre che i due condensatori $c_{b'c}$ e $c_{b'e}$ sono in parallelo tra loro, per cui la funzione di trasferimento A_I avrà soltanto un polo. Deve inoltre avere anche uno zero, dato che $A_I(\infty) \neq 0$. Il polo può essere calcolato con il metodo della resistenza vista: la resistenza vista dal condensatore risultante dal parallelo dei due è semplicemente $r_{b'e}$. Quindi

$$\omega_p = \frac{1}{(c_{b'c} + c_{b'e})r_{b'e}}.$$

Possiamo immediatamente ricavare lo zero tramite i valori del polo e dei guadagni a zero e all'infinito:

$$\omega_z = \omega_p \frac{A_I(0)}{A_I(\infty)} = -\frac{g_m}{c_{b'c}}.$$

Lo zero risulta quindi a parte reale positiva ($s_z = -\omega_z$), in conseguenza del fatto che il segno del guadagno si inverte passando dalle basse alle alte frequenze.

Consideriamo valori numerici realistici per le varie grandezze che compaiono nel circuito: $c_{b'e} = 200$ pF, $c_{b'c} = 5$ pF, $r_{b'e} = 3$ k Ω , $g_m = 50$ mA/V. Con tali valori $\omega_z = -10 \times 10^9$ rad/s e $\omega_p = 1.626 \times 10^6$ rad/s. Quindi lo zero si trova a frequenza molto più alta del polo (in effetti talmente alta che il circuito equivalente perde di validità e quindi questo è uno zero “puramente matematico”) e pertanto la funzione di trasferimento può essere approssimata nella forma a polo dominante:

$$A_I \simeq \frac{h_{fe}}{1 + \frac{s}{\omega_p}}.$$

Per $\omega \gg \omega_p$ possiamo approssimare il modulo di A_I nella forma

$$|A_I(j\omega)| = h_{fe} \frac{\omega_p}{\omega}.$$

Il guadagno diventa pertanto unitario per una pulsazione ω_T pari a $\omega_T = h_{fe}\omega_p$:

$$f_T = \frac{\omega_T}{2\pi} = \frac{h_{fe}}{2\pi r_{b'e}(c_{b'c} + c_{b'e})}.$$

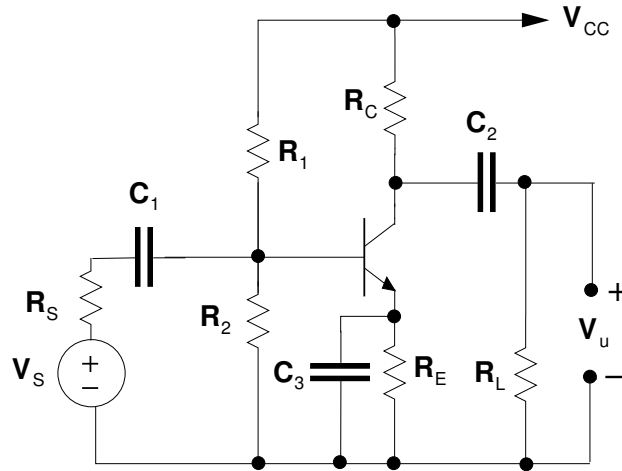
In effetti f_T , detta anche frequenza di taglio del transistor, rappresenta ciò che viene definito il prodotto guadagno-banda: il guadagno di corrente ottenibile a qualunque frequenza è pari all'incirca alla f_T divisa per tale frequenza.

Dal punto di vista pratico la f_T viene utilizzata per la determinazione di $(c_{b'e} + c_{b'c})$. Il costruttore misura f_T e la fornisce nelle caratteristiche, insieme con un grafico della $c_{b'e}$ in funzione della V_{CB} . Pertanto dal valore della f_T si calcola $(c_{b'e} + c_{b'c})$:

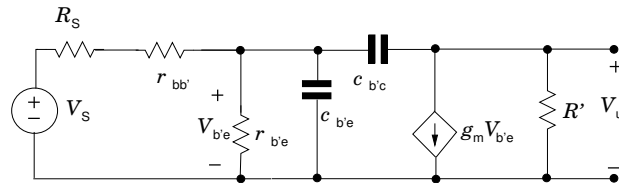
$$c_{b'e} + c_{b'c} = \frac{h_{fe}}{2\pi r_{b'e} f_T},$$

dopodiché il valore di $c_{b'e}$ si ottiene sottraendo quello di $c_{b'c}$. Il circuito a π ibrido è valido fino a una frequenza pari a circa $f_T/3$; oltre è necessario ricorrere a rappresentazioni più complesse del transistor.

9.13* Analisi dello stadio a emettitore comune alle alte frequenze

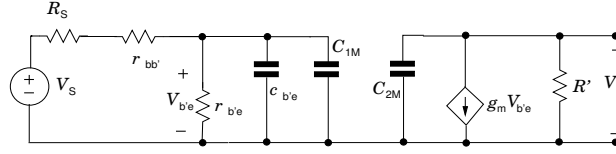


Alle alte frequenze possiamo considerare i condensatori di disaccoppiamento e di bypass come corto circuiti. Quindi R_C , nel circuito dinamico, risulta in parallelo a R_L ; definiamo $R' = R_C // R_L$. Supponiamo inoltre, per rendere più semplici i calcoli, che il parallelo tra R_1 e R_2 sia molto maggiore di h_{ie} e possa quindi essere trascurato. Il circuito dinamico risulta:



Attribuiamo dei valori numerici: $R_S = 1 \text{ k}\Omega$, $r_{bb'} = 900 \text{ }\Omega$, $r_{b'e} = 3 \text{ k}\Omega$, $g_m = 50 \text{ mA/V}$, $R' = 10 \text{ k}\Omega$, $c_{b'e} = 200 \text{ pF}$, $c_{b'c} = 5 \text{ pF}$. Utilizziamo il teorema di Miller, allo scopo di rimuovere il condensatore ponte $c_{b'c}$. Il k di Miller è rappresentato dal rapporto tra V_u e $V_{b'e}$. Si noti che in questo caso k è, oltre che una funzione di rete, anche una funzione di trasferimento, dato che V_u dipende esclusivamente da $V_{b'e}$. Possiamo quindi calcolare k ponendo un generatore di prova in parallelo alla $r_{b'e}$. Facciamo l'ipotesi che k sia reale almeno fino a una decade al di sopra del polo più basso; pertanto calcoliamo soltanto il valore di k a frequenza nulla, k_0 . Giustificaremo più avanti la validità di tale ipotesi. Abbiamo dunque

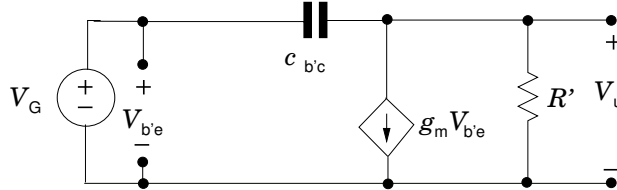
$$k_0 = -g_m R' = -500.$$



Il circuito, dopo l'applicazione del teorema di Miller, risulta:

Quindi, in parallelo a $c_{b'e}$ compare una capacità $C_{1M} = c_{b'e}(1 - k_0)$, che corrisponde sostanzialmente alla $c_{b'e}$ moltiplicata per il modulo di k_0 . Dunque C_{1M} risulta prevalente rispetto a $c_{b'e}$, dato l'elevato valore di $|k_0|$. La capacità totale C presente sulla maglia di ingresso risulta $C = c_{b'e} + c_{b'e}(1 + g_m R') = 2.7 \text{ nF}$ (quindi $C \gg c_{b'e}$). Questo effetto di moltiplicazione della capacità ponte è detto “effetto Miller” ed è spesso sfruttato quando si vogliono ottenere capacità equivalenti abbastanza elevate a partire da capacità reali di relativamente piccolo valore. Il condensatore C_{2M} risulta invece all'incirca pari a $c_{b'e}$. Dato che le resistenze viste da C e da C_{2M} sono dello stesso ordine di grandezza, mentre $C \gg C_{2M}$, ci aspettiamo che il polo più basso sia determinato da C . Calcoliamo quindi la resistenza vista da C : $R_{VC} = r_{b'e} // (R_S + r_{bb'}) = 1.16 \text{ k}\Omega$. Perciò $\omega_p = 1/(R_{VC}C) = 319.28 \text{ krad/s}$, $f_p = 50.8 \text{ kHz}$. Il polo associato alla maglia di uscita risulterebbe $\omega_p^* = 1/(R'C_{2M}) = 19.96 \text{ Mrad/s}$, $f_p^* = 3.18 \text{ MHz}$, tuttavia questo polo non ha alcun significato fisico, perché in corrispondenza della frequenza del polo associato con C_{2M} il k di Miller non sarebbe più reale, nel circuito equivalente da noi disegnato, e quindi non potremmo più sostituirlo con k_0 .

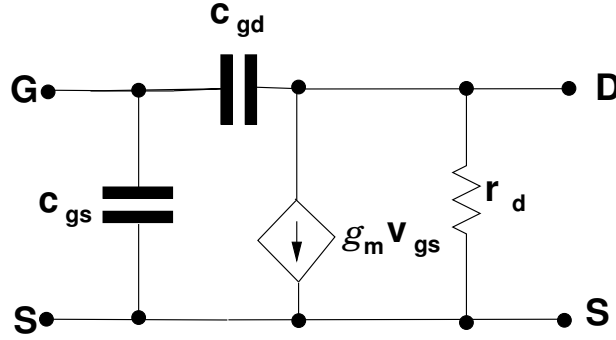
Per quanto riguarda invece ω_p , possiamo svolgere una verifica andando a calcolare l'effettivo polo di k in modo esatto, cosa che in questo caso risulta abbastanza semplice. Come già detto, k è in questo caso una funzione di trasferimento e può essere determinato dal circuito sottostante:



Poiché quando misuriamo la resistenza R_V vista da $c_{b'e}$ il generatore V_G è spento e quindi $g_m V_{b'e}$ è disattivato, $R_V = R'$ e pertanto $\omega_{p_k} = 1/(c_{b'e}R') = 20 \text{ Mrad/s}$. Dunque $\omega_{p_k} \gg \omega_p$, quindi il valore trovato per ω_p è attendibile, dato che a ω_p il k sarà ancora senz'altro reale e pari a k_0 . Se valutassimo i poli del circuito in maniera esatta troveremmo $\omega_{p_1} = 314 \text{ krad/s}$ e $\omega_{p_2} = 274 \text{ Mrad/s}$. Quindi, come ci aspettavamo, il polo più basso trovato con il metodo approssimato è praticamente coincidente con quello esatto, mentre quello più alto è completamente diverso. Se, come spesso accade, ci interessa soltanto determinare il limite superiore di banda, è sufficiente conoscere solo ω_{p_1} , dato che questo agisce da polo dominante. La banda dello stadio a emettitore comune è quindi significativamente limitata dall'effetto Miller sul condensatore ponte.

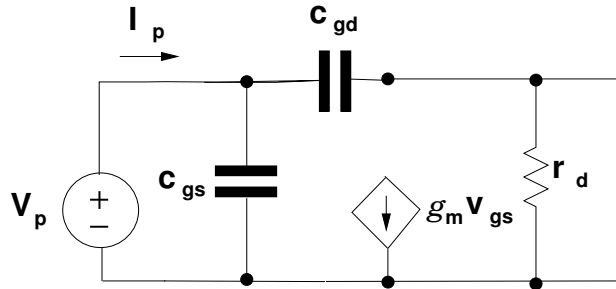
9.14 Circuito equivalente alle alte frequenze per transistori FET

Il comportamento alle alte frequenze dei transistori a effetto di campo, sia JFET sia MOSFET, può essere riprodotto tramite l'aggiunta al circuito equivalente per le variazioni di due capacità interne: c_{gs} tra gate e source e c_{gd} tra gate e drain, come indicato nello schema seguente.



Sia c_{gs} sia c_{gd} sono di solito dell'ordine di qualche picofarad. Sulle caratteristiche fornite dal costruttore vengono dati due valori misurati sperimentalmente: c_{is} , la capacità di ingresso con l'uscita in corto circuito, e c_{rs} , che si ottiene tramite una misura della corrente nel circuito di ingresso, posto in corto, quando viene applicato un segnale di prova all'uscita.

Il circuito di prova per la misura di c_{is} è il seguente:

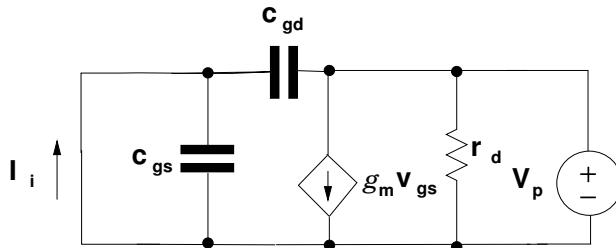


La capacità c_{is} è definita tramite la relazione

$$\frac{V_p}{I_p} = \frac{1}{j\omega c_{is}},$$

quindi, $c_{is} = c_{gd} + c_{gs}$, dato che, in conseguenza del corto circuito in uscita, le due capacità interne risultano in parallelo.

Il circuito per la misura di c_{rs} risulta invece:



La capacità c_{rs} è definita tramite la relazione

$$\frac{V_p}{I_i} = \frac{1}{j\omega c_{rs}},$$

per cui $c_{rs} = -c_{gd}$. Dai valori di c_{rs} e di c_{is} si può dunque risalire facilmente a c_{gs} e c_{gd} .

10. La tecnica della reazione

10.1 Concetti generali

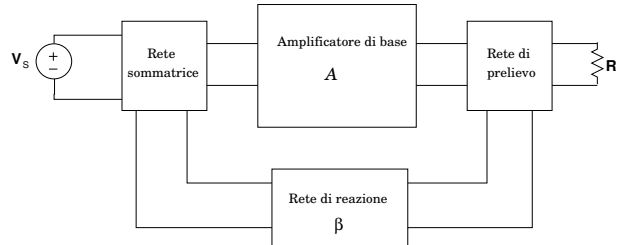
Il principio della reazione consiste nel riportare all'ingresso di un sistema una porzione del segnale in uscita, in modo da modificare le proprietà del sistema stesso. In natura esistono molti esempi di reazione: basta considerare la reazione visiva che ci consente di mantenere l'equilibrio camminando su un asse, il meccanismo di regolazione della temperatura del corpo, l'aggiustamento dell'iride per compensare le variazioni di luce. In tutti questi casi la misura di una grandezza viene utilizzata per correggere dei parametri che influenzano a loro volta la stessa grandezza, in modo da mantenerla costante. Anche nel campo dei controlli la reazione è molto comunemente utilizzata: il regolatore di Watt per mantenere costante la velocità di rotazione di un motore a vapore ne è uno dei primi esempi, così come un altro esempio notevole è rappresentato dal termostato per la regolazione della temperatura in un impianto di riscaldamento.

In campo elettronico abbiamo già incontrato un esempio di reazione, anche se non l'abbiamo identificato come tale: l'inserzione di una resistenza di emettitore nel circuito di polarizzazione di un transistor consente di ottenere una reazione che stabilizza il punto di lavoro. La grandezza di uscita è rappresentata dalla variazione di corrente di emettitore, che viene trasformata tramite la resistenza di emettitore in una variazione di tensione e reinserita, sempre tramite la stessa resistenza e con polarità invertita, in serie alla polarizzazione di base dovuta al partitore. In tutti questi casi nei quali lo scopo è quello di mantenere una grandezza costante, la reazione che si realizza è di tipo negativo, vale a dire che il segnale riportato in ingresso ha segno rovesciato rispetto a quello del segnale di ingresso che lo ha prodotto. In questo modo ogni variazione determina un effetto a essa opposto, che tende a contrastarla. In campo elettronico la reazione utilizzata è di solito negativa, anche se gli scopi per cui viene realizzata sono ben più vari che della semplice regolazione di una grandezza e anche se, storicamente, i primi circuiti in reazione utilizzavano una reazione positiva (allo scopo di incrementare il poco guadagno ottenibile con i primi amplificatori dell'inizio del secolo).

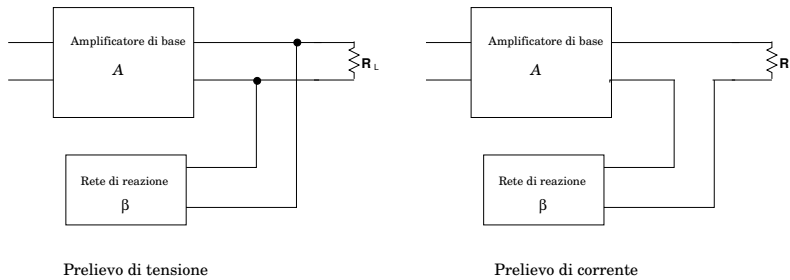
Nel seguito svolgeremo una trattazione semplificata della reazione, che coprirà alcuni dei concetti fondamentali, in particolare l'effetto della reazione sul guadagno e quello sulle resistenze di ingresso e di uscita, che risulteranno utili per la comprensione del funzionamento dei circuiti basati sugli amplificatori operazionali.

10.2 Teoria semplificata della reazione

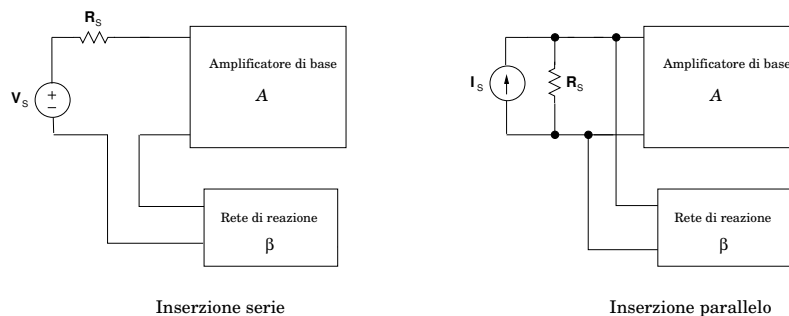
Nell'ambito della teoria semplificata, un circuito in reazione viene descritto sulla base di uno schema a blocchi corrispondente a quello di seguito riportato e costituito da un blocco A amplificatore di base, una rete per il prelievo della grandezza di uscita, una rete di reazione β e una rete sommatrice che consente di sommare il segnale di reazione con quello di ingresso.



Il prelievo in uscita può essere di due tipi: di corrente, nel caso in cui la grandezza prelevata sia la corrente nel carico, oppure di tensione, se la grandezza prelevata è la tensione ai capi del carico. Questi due casi sono illustrati nella figura che segue, indicando i collegamenti per il prelievo.



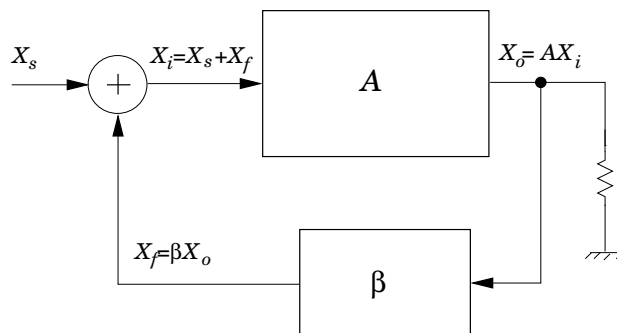
La reinserzione in ingresso può essere anch'essa effettuata in due modi diversi, riportando la grandezza di reazione in serie o in parallelo al generatore posto in ingresso. Si parla in tal caso, rispettivamente, di reazione serie o parallelo. Le due modalità di inserzione sono rappresentate nell'illustrazione che segue.



Si hanno quindi, in totale, quattro tipologie diverse di circuiti in reazione: con reazione serie di tensione, parallelo di tensione, serie di corrente, parallelo di corrente. Vedremo più avanti le importanti differenze tra queste configurazioni dal punto di vista delle impedenze di ingresso e di uscita.

In genere la rete di reazione è realizzata con componenti passivi, ma questa non è una regola generale e si trovano spesso reti di reazione contenenti componenti

attivi. La rete A è di solito costituita da un amplificatore contenente uno o più transistori. Vediamo una rappresentazione generale di una rete in reazione, ove le grandezze sono indicate con X , in modo che possano rappresentare sia tensioni sia correnti. Indichiamo con X_s la grandezza di ingresso proveniente dall'esterno, con X_i la grandezza all'entrata dell'amplificatore di base A , con X_o la grandezza di uscita e con X_f la grandezza di reazione, ottenuta facendo attraversare a X_o la rete di reazione β . Quindi $X_f = \beta X_o$ e, in conseguenza dell'azione del sommatore, $X_i = X_s + X_f$.



Abbiamo inoltre che $X_o = AX_i$. Possiamo dunque scrivere la grandezza di uscita nella forma

$$X_o = A(X_f + X_s) = A\beta X_o + AX_s.$$

Portando a primo membro tutti i termini che contengono X_o , otteniamo

$$X_o(1 - \beta A) = AX_s$$

e quindi

$$\frac{X_o}{X_s} = \frac{A}{1 - \beta A}.$$

Il guadagno della rete in reazione risulta pertanto pari a quello dell'amplificatore di base A diviso per il fattore $(1 - \beta A)$. Il guadagno A si definisce di solito "guadagno ad anello aperto", mentre βA viene indicato come "guadagno d'anello", poiché corrisponde al guadagno subito dal segnale che percorre l'intero anello di reazione. Si parla infine, per il guadagno della rete in reazione, di "guadagno ad anello chiuso".

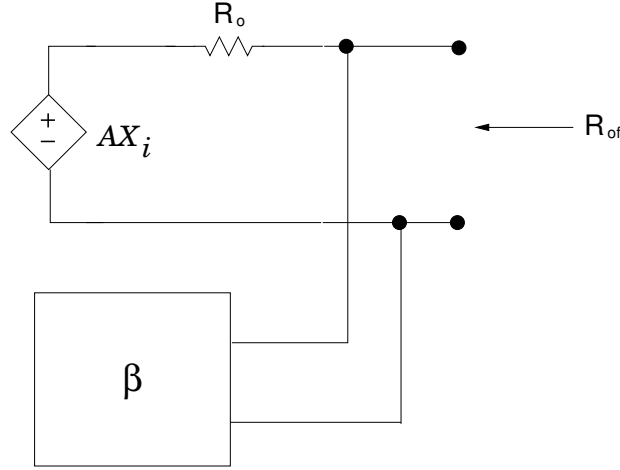
È importante ricordare che il risultato trovato è rigorosamente valido soltanto se sono verificate le condizioni di applicabilità della teoria semplificata della reazione, che indichiamo di seguito:

- 1) la rete di reazione è unidirezionale (il segnale è trasmesso all'uscita solo tramite l'amplificatore di base A);
- 2) l'amplificatore di base A è unidirezionale (il segnale di reazione raggiunge l'ingresso solo attraverso la rete di reazione β);
- 3) il fattore di reazione β è indipendente dalla resistenza di sorgente R_s e da quella di carico R_l .

10.3 Effetto della reazione sulle impedenze di ingresso e di uscita

Svolgiamo i calcoli in modo dettagliato soltanto per il caso della valutazione dell'impedenza di uscita in presenza di reazione di tensione, mentre per gli altri casi daremo semplicemente una breve giustificazione, seguita dal risultato.

Per la reazione di tensione, prendiamo in esame la parte di circuito di uscita, dove la rete β rappresenta la rete di reazione ed è connessa in modo da prelevare la tensione di uscita. La resistenza di uscita può essere ottenuta semplicemente come rapporto della tensione a vuoto (misurata senza alcun carico in uscita) e la corrente di corto circuito (misurata con carico di valore nullo).



Quando si effettua la misura della corrente di corto circuito I_{sc} , l'uscita viene posta in corto e quindi il segnale di reazione diventa nullo. Pertanto, in ingresso non è presente alcun segnale di reazione e si può trovare immediatamente il valore di I_{sc} .

$$I_{sc} = \frac{AX_i}{R_o} = \frac{AX_s}{R_o}.$$

A vuoto, invece, la reazione agisce regolarmente e dà luogo a un'uscita di valore

$$V_o = \frac{A}{1 - \beta A} X_s.$$

Quindi la resistenza di uscita in reazione risulterà

$$R_{of} = \frac{V_o}{I_{sc}} = \frac{A}{1 - \beta A} X_s \frac{R_o}{AX_s} = \frac{R_o}{1 - \beta A}.$$

La reazione di tensione riduce quindi la resistenza di uscita di un fattore $(1 - \beta A)$. Il fatto che una reazione di tensione riduca la resistenza di uscita può essere compreso da un punto di vista intuitivo pensando che tale reazione tende a opporsi alle variazioni della tensione di uscita dovute a variazioni del carico, cosa che corrisponde al comportamento di un generatore con bassa resistenza interna (al limite un generatore di tensione ideale, con resistenza interna nulla, fornirebbe sempre la stessa tensione al carico, indipendentemente dal valore di resistenza di quest'ultimo).

Nel caso di reazione di corrente, invece, possiamo comprendere intuitivamente l'effetto sulla resistenza di uscita pensando che tale reazione tende a opporsi alle

variazioni della corrente di uscita dovute a variazioni del carico, cosa che corrisponde al comportamento di un generatore con elevata resistenza interna (al limite un generatore di corrente ideale, con resistenza interna infinita, fornirebbe sempre la stessa corrente al carico, indipendentemente dal valore di resistenza di quest'ultimo). Più precisamente, la resistenza di uscita in presenza di reazione di corrente risulta $R_o(1 - \beta A)$.

Per quanto riguarda la resistenza di ingresso, nel caso di inserzione serie la grandezza di reazione si oppone al flusso della corrente erogata dal generatore V_S , quindi, a parità di V_S , tale generatore erogherà una corrente minore. Ciò corrisponde a un aumento della resistenza vista in ingresso: se si svolgono i calcoli, si ottiene un valore $R_{in}(1 - \beta A)$.

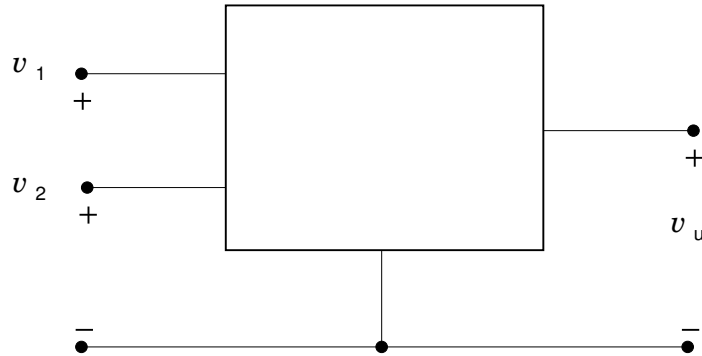
Nel caso invece di reazione parallelo, la corrente di reazione agisce in senso opposto alla corrente del generatore I_S , determinando una riduzione della tensione di ingresso, che corrisponde a una resistenza vista in ingresso ridotta. Il valore ottenuto dai calcoli è $R_{in}/(1 - \beta A)$.

In sostanza, quindi, le resistenze di ingresso e di uscita vengono modificate dalla reazione di un fattore (a moltiplicare o a dividere, a seconda del tipo di reazione) $1 - \beta A$, che può risultare anche molto elevato. È pertanto possibile, utilizzando in modo opportuno la reazione e avendo a disposizione amplificatori con guadagno sufficientemente elevato, ottenere pressoché ogni valore desiderato di resistenza di ingresso e di uscita.

11. Circuiti basati su amplificatori operazionali

11.1 Amplificatori differenziali

In molte applicazioni risulta utile disporre di un amplificatore in grado di fornire in uscita un segnale proporzionale alla differenza tra i segnali applicati a due ingressi e indipendente dalla componente rispetto a massa presente con eguale ampiezza in ciascuno dei segnali. Possiamo rappresentare il generico amplificatore differenziale come una “black-box” con due ingressi e una uscita:



In base alla definizione appena data, per un amplificatore differenziale ideale $V_u = A_d(v_1 - v_2)$, dove A_d è il cosiddetto guadagno differenziale e v_1 e v_2 sono le tensioni sui due ingressi, riferite a massa. Un amplificatore differenziale non ideale presenta anche un “guadagno di modo comune”, fornisce cioè un’uscita non nulla anche in presenza di segnali uguali ai due ingressi e proporzionale al valore di tali segnali rispetto a massa.

Definiamo il segnale a modo differenziale v_d :

$$v_d = v_1 - v_2$$

e il segnale a modo comune

$$v_c = \frac{1}{2}(v_1 + v_2).$$

È possibile esprimere v_1 e v_2 in funzione di v_d e v_c utilizzando le seguenti relazioni, ottenute sommando e sottraendo tra loro le due equazioni utilizzate per definire v_d e v_c

$$\begin{cases} v_1 = \frac{v_d}{2} + v_c \\ v_2 = -\frac{v_d}{2} + v_c. \end{cases}$$

La tensione v_u di uscita può essere scritta come combinazione lineare delle due tensioni di ingresso:

$$v_u = A_1 v_1 + A_2 v_2.$$

Sostituiamo a v_1 e v_2 le espressioni prima trovate in funzione di v_c e di v_d :

$$\begin{aligned} v_u &= A_1 \left(v_c + \frac{1}{2} v_d \right) + A_2 \left(v_c - \frac{1}{2} v_d \right) \\ &= A_1 v_c + \frac{A_1}{2} v_d + A_2 v_c - \frac{A_2}{2} v_d \\ &= (A_1 + A_2) v_c + \left(\frac{A_1 - A_2}{2} \right) v_d \\ &= A_c v_c + A_d v_d, \end{aligned}$$

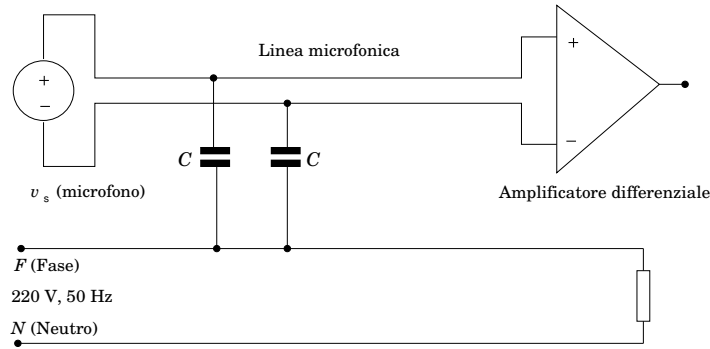
dove abbiamo definito

$$A_c = A_1 + A_2$$

$$A_d = \frac{A_1 - A_2}{2},$$

fornendo così una relazione tra i guadagni a modo differenziale e a modo comune e quelli rispetto all'uno e all'altro ingresso dell'amplificatore. Nel linguaggio degli spazi vettoriali possiamo dire che la rappresentazione delle componenti in ingresso sotto la forma v_1, v_2 o sotto la forma v_c, v_d corrisponde semplicemente a rappresentare i vettori di uno spazio bidimensionale rispetto a due basi differenti. Le relazioni che abbiamo trovato ci permettono di passare facilmente dall'una all'altra rappresentazione.

L'utilizzo di un amplificatore differenziale risulta particolarmente utile in quei casi in cui sono presenti prevalentemente disturbi a modo comune, come quelli, per esempio, indotti da accoppiamenti elettrostatici tra i cavi di rete a 50 Hz e le linee microfoniche. Una rappresentazione semplificata del problema è fornita nella figura sottostante.



La distribuzione dell'energia elettrica, come discuteremo nuovamente nel seguito, viene effettuata connettendo a terra uno dei capi del secondario del trasformatore posto nelle cabine per l'abbassamento della tensione dal valore di media tensione (15 kV) ai 230 V di normale utilizzo. In conseguenza di tale collegamento, uno dei due fili di rete (detto "conduttore neutro") si viene a trovare praticamente al potenziale di terra, mentre l'altro (detto "conduttore di fase") si trova a una tensione alternata di 230 V rispetto alla terra. Il conduttore di fase rappresenta perciò la sorgente dei disturbi a carattere elettrostatico ed è accoppiato ai due conduttori della linea microfonica da capacità praticamente uguali. Se tale linea avesse uno dei due conduttori connesso a terra (linea sbilanciata), il disturbo si accoppierebbe soltanto con l'altro e non si potrebbe fare nulla per eliminarlo. Con la realizzazione di una linea bilanciata come quella in figura (con conduttori elettricamente simmetrici rispetto alla terra), lo stesso disturbo viene iniettato su tutti e due i conduttori: si tratta perciò di un disturbo a modo comune, che può essere eliminato utilizzando un amplificatore differenziale come quello indicato.

Per completare l'analisi degli amplificatori differenziali, possiamo ricavare A_1 e A_2 in funzione di A_c e A_d :

$$\begin{cases} A_c = A_1 + A_2 \\ A_d = \frac{A_1 - A_2}{2} \end{cases}$$

Sommando e sottraendo il doppio della seconda equazione dalla prima otteniamo:

$$\begin{cases} A_c + 2A_d = 2A_1 \\ A_c - 2A_d = 2A_2. \end{cases}$$

Quindi

$$\begin{cases} A_1 = \frac{A_c}{2} + A_d \\ A_2 = \frac{A_c}{2} - A_d. \end{cases}$$

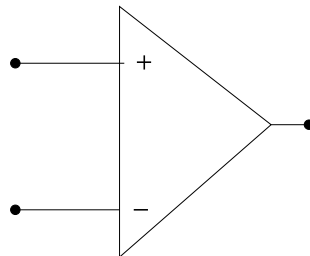
Il rapporto tra il guadagno a modo differenziale e quello a modo comune viene di solito definito “rapporto di reiezione del modo comune” e indicato con la sigla CMRR (Common Mode Rejection Ratio). Spesso il CMRR viene espresso in dB.

11.2 Amplificatori operazionali

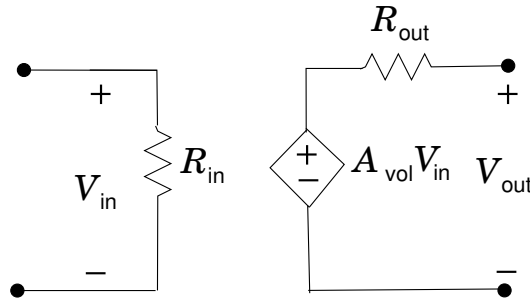
L’amplificatore operazionale è un amplificatore accoppiato in continua (quindi con banda che si estende fino a frequenza nulla), con guadagno molto elevato e con ingresso differenziale. Esso rappresenta il circuito integrato analogico di base ed esiste in moltissime versioni diverse. Il nome “operazionale” deriva dal fatto che esso venne inizialmente concepito per la realizzazione di operazioni di somma e sottrazione tra segnali all’interno di circuiti più complessi, come il calcolatore analogico, che consentiva, proprio grazie all’utilizzo degli amplificatori operazionali, di implementare tramite componenti elettronici le relazioni tra grandezze elettriche e le loro derivate temporali corrispondenti a quelle in una equazione differenziale. In questo modo era possibile, a partire da un insieme di condizioni iniziali definite tramite lo stato di carica dei condensatori che comparivano nel circuito, ottenere la soluzione desiderata della equazione differenziale semplicemente facendo evolvere nel tempo la grandezza elettrica rappresentativa di tale soluzione. Il calcolatore analogico è ormai obsoleto, dato che le sue prestazioni sono di gran lunga superate da quelle ottenibili con metodi numerici su normali calcolatori digitali.

L’amplificatore operazionale, invece, è ancora largamente usato in un gran numero di sistemi analogici, per la semplicità con cui consente di trattare i segnali e di realizzare funzioni anche complesse.

Il simbolo circuitale dell’amplificatore operazionale consiste in un triangolo con un simbolo “+” in corrispondenza dell’ingresso non invertente e un simbolo “-” in corrispondenza di quello invertente.



Il circuito equivalente per le variazioni risulta il seguente, dove il terminale in basso del generatore comandato di tensione va considerato connesso al punto medio o “baricentro” delle tensioni di alimentazione. Infatti gli amplificatori operazionali hanno



di solito due terminali di alimentazione e nessun terminale di massa. Nella maggior parte dei casi tali terminali di alimentazione vengono connessi a due tensioni di alimentazione simmetriche rispetto alla massa, per esempio $+15\text{ V}$ e -15 V , e in tal caso i terminali di alimentazione non vengono nemmeno indicati negli schemi.

In tale situazione il baricentro delle alimentazioni corrisponde proprio alla massa e quindi il terminale basso del generatore comandato di tensione è connesso alla massa stessa. Il coefficiente A_{vol} , che rappresenta il rapporto tra la tensione di uscita e quella di ingresso, è di solito detto “guadagno ad anello aperto”, perché corrisponde al guadagno dell’amplificatore operazionale quando questo è usato ad “anello aperto”, vale a dire senza chiudere l’anello di reazione nel quale viene di solito invece inserito.

Un amplificatore operazionale ideale è caratterizzato dalle seguenti proprietà:

- 1) Resistenza di ingresso R_{in} infinita
- 2) Resistenza di uscita R_{out} nulla
- 3) Guadagno A_{vol} infinito
- 4) Banda infinita
- 5) Amplificazione a modo comune nulla

Gli amplificatori operazionali reali rappresentano soltanto un’approssimazione di queste caratteristiche ideali. In particolare, la banda risulta molto spesso piuttosto limitata, per evitare problemi di stabilità che potrebbero insorgere quando l’amplificatore operazionale viene utilizzato all’interno di circuiti in reazione. È da notare, comunque, che anche se la banda ad anello aperto può risultare di poche decine di hertz o addirittura di pochi hertz, il prodotto guadagno banda, dato l’alto valore di A_{vol} , è sempre molto elevato e, quindi, con un’opportuna scelta della reazione, che ci permette di ampliare la banda a spese del guadagno, si possono ottenere circuiti caratterizzati da banda anche molto ampia.

Un tipico amplificatore operazionale, inizialmente realizzato negli anni ’70, ma ancora talvolta utilizzato, è il $\mu\text{A } 741$, che è caratterizzato da R_{in} di $2\text{ M}\Omega$, R_{out} di $25\ \Omega$, A_{vol} tra 10^5 e 2×10^5 , e una banda tra i 4 e gli 8 Hz. Il rapporto di reiezione del modo comune (CMRR) alle basse frequenze è dell’ordine dei 90 dB. Attualmente sono disponibili amplificatori operazionali con caratteristiche molto superiori, in particolare si possono avere resistenze di ingresso di centinaia di megaohm e prodotti guadagno-banda oltre il gigahertz.

11.3 Metodo del corto circuito virtuale

Dato il valore estremamente alto del guadagno degli amplificatori operazionali e il fatto che la loro tensione di uscita non può comunque superare quella di alimentazione, la tensione presente tra gli ingressi in condizioni di funzionamento lineare è estremamente piccola, tale da poter essere in genere considerata nulla, dal punto di vista pratico, nella maggior parte dei casi. In particolare, la tensione tra gli ingressi è trascurabile nel caso essa sia molto più piccola delle altre tensioni significative

presenti nel circuito. Si può dimostrare che questa condizione è verificata, nei casi di reazione negativa, se il modulo del guadagno di anello ($|\beta A|$) della rete in reazione nella quale l'operazionale è inserito è molto maggiore dell'unità. È facile capire che in tal caso la tensione di ingresso dell'operazionale risulterà molto più piccola di quella riportata in ingresso dalla reazione.

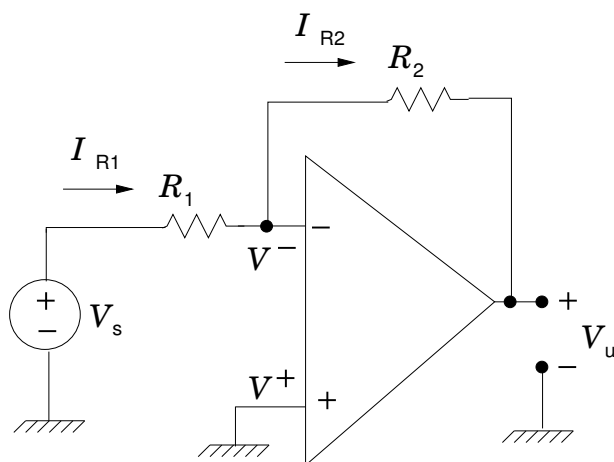
Se la tensione di ingresso può essere considerata nulla, tra i due ingressi sussiste quindi una sorta di corto circuito, di tipo, però, molto particolare, perché non è attraversato da corrente: essendo la tensione ai capi di R_{in} praticamente nulla, anche la corrente che la attraversa deve essere nulla. Si noti che il fatto che la corrente che fluisce negli ingressi dell'operazionale sia trascurabile è una conseguenza diretta dell'ipotesi di corto circuito virtuale e non richiede in alcun modo che la resistenza di ingresso dell'amplificatore sia elevata o, tantomeno, infinita.

Il metodo del corto circuito virtuale rappresenta uno strumento estremamente semplice per l'analisi dei circuiti lineari a operazionali, che consente di calcolare la funzione di trasferimento in modo particolarmente rapido. Dobbiamo però sempre aver presente il fatto che tale approccio è valido solo fintanto che il modulo del guadagno d'anello è molto maggiore dell'unità: questa condizione non è di solito verificata al di sopra di una certa frequenza, a causa della riduzione di guadagno dell'operazionale. Inoltre può non essere verificata, dipendentemente dalla struttura della rete di reazione, alle frequenze molto basse o su una particolare banda frequenziale.

Un'altra condizione nella quale non si può assolutamente applicare il metodo del corto circuito virtuale è costituita dal funzionamento in condizioni di non linearità degli amplificatori operazionali. Se l'uscita, per esempio, raggiunge il valore di saturazione (di solito pari in modulo al valore assoluto della tensione di alimentazione meno circa 1 V), la differenza di potenziale tra gli ingressi può assumere un valore qualunque (e non trascurabile), purché con polarità consistente con quella della tensione di uscita.

11.4 Amplificatore invertente

L'amplificatore invertente è uno dei circuiti più semplici e di più immediata comprensione tra quelli realizzabili tramite amplificatori operazionali. Lo schema risulta:



Applichiamo il metodo del cortocircuito virtuale. In conseguenza dell'uguaglianza tra le tensioni sui due terminali di ingresso, $V^- = V^+ = 0$, quindi risulta nota e

pari a V_s la tensione ai capi della resistenza R_1 . La corrente I_{R_1} può dunque essere calcolata

$$I_{R_1} = \frac{V_s}{R_1}.$$

Poiché, sempre in base al metodo del corto circuito virtuale (c.c.v.) non fluisce corrente negli ingressi dell'amplificatore operazionale, $I_{R_2} = I_{R_1}$. Quindi

$$V_u = -I_{R_2}R_2 + V^- = -I_{R_2}R_2 = \frac{-V_s}{R_1}R_2.$$

Pertanto il guadagno A risulta

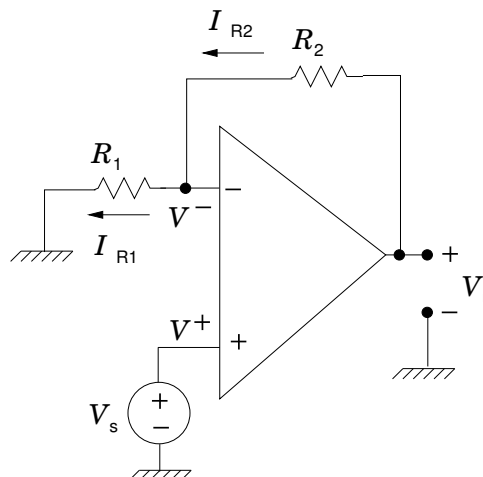
$$A = \frac{V_u}{V_s} = -\frac{R_2}{R_1}.$$

Osserviamo che A è in questo caso, grazie alla reazione, indipendente dalle caratteristiche specifiche dell'elemento attivo (purché questo presenti un guadagno ad anello aperto estremamente elevato) e il suo modulo dipende soltanto dal rapporto di due resistenze, che possono essere anche di notevole precisione.

Notiamo inoltre che siamo di fronte a una reazione di tensione con inserzione parallelo: la resistenza di uscita sarà quindi molto bassa, mentre per quella di ingresso dobbiamo svolgere un breve ragionamento. L'inserzione parallelo avviene ai terminali dell'operazionale, che non coincide con l'ingresso dell'amplificatore; quindi la resistenza di ingresso vista sull'ingresso dell'operazionale risulterà molto bassa, praticamente nulla, ma quella vista dal generatore V_s sarà sostanzialmente pari alla resistenza R_1 che si trova in serie al c.c.v.. Un altro modo di raggiungere la stessa conclusione consiste nell'osservare che la tensione tra l'ingresso invertente e la massa è nulla a causa del corto circuito virtuale, quindi il generatore V_s vede semplicemente una resistenza pari a R_1 connessa verso massa.

11.5 Amplificatore non invertente

Si può realizzare un amplificatore non invertente basato su un operazionale con piccole modifiche rispetto allo schema già visto per l'amplificatore invertente:



Utilizziamo anche per l'analisi di questo circuito il metodo del corto circuito virtuale. La tensione V^- sull'ingresso invertente risulta pari, per il c.c.v., alla tensione sull'altro ingresso che è a sua volta pari a V_s . Dunque la corrente I_{R_1} che scorre in R_1 è data da

$$I_{R_1} = \frac{V_s}{R_1}.$$

Poichè, per il metodo del corto circuito virtuale, assumiamo che la corrente in ingresso all'amplificatore operazionale sia nulla, $I_{R_2} = I_{R_1}$. Otteniamo dunque

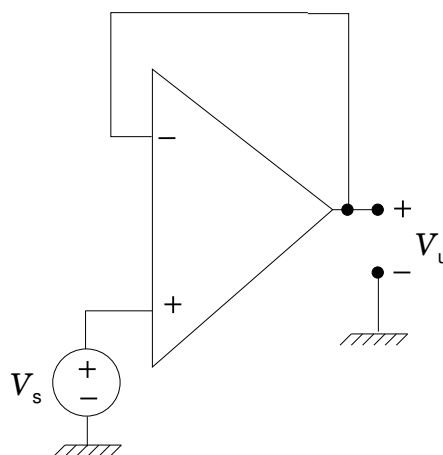
$$V_u = I_{R_2} R_2 + V_1 = I_{R_1} R_2 + V_s = V_s \left(\frac{R_2}{R_1} + 1 \right).$$

Pertanto

$$\frac{V_u}{V_s} = \frac{R_2}{R_1} + 1.$$

Abbiamo quindi ottenuto un amplificatore non invertente con guadagno $R_2/R_1 + 1$. Notiamo che in questo amplificatore si ha una reazione di tensione con inserzione serie, quindi la resistenza di ingresso, anche nel caso di R_{in} dell'operazionale non molto alta, risulta comunque elevatissima, tanto che è di solito determinata dalle resistenze parassite (resistenze di isolamento) che inevitabilmente esistono tra ciascuno degli ingressi e la massa.

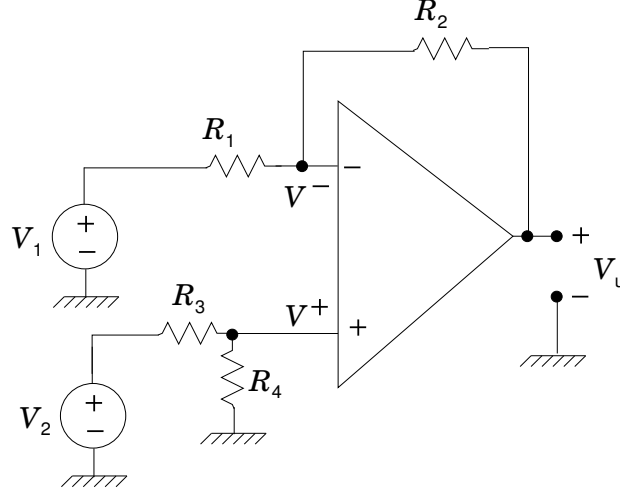
Se poniamo $R_2 = 0$, sostituendola con un corto circuito, il guadagno dell'amplificatore non invertente diventa unitario. In tal caso abbiamo un amplificatore con guadagno pari all'unità, impedenza di ingresso pressoché infinita e impedenza di uscita pressoché nulla. Questo può essere utile ogni volta che risulta necessario collegare un circuito con uscita ad alta impedenza, o comunque non in grado di fornire una corrente significativa, a un carico a bassa impedenza: è per tale motivo che questo amplificatore prende il nome di "buffer". Notiamo che se R_2 è nulla R_1 non svolge più alcuna funzione (qualunque sia la corrente in R_1 la caduta di tensione su R_2 è comunque nulla) e può quindi essere eliminata, ottenendo lo schema tipico del buffer.



11.6 Amplificatore differenziale

Combinando la struttura di un amplificatore invertente con quella di un non invertente si può ottenere un amplificatore differenziale caratterizzato da un guadagno

preciso, che dipende soltanto da un rapporto di resistenze. Si noti che anche un amplificatore operazionale preso da solo, non reazionato, è un amplificatore differenziale, ma non può, nella maggior parte dei casi, essere utilizzato direttamente come tale, a causa dell'eccessivo e non esattamente noto guadagno.



Possiamo studiare questo circuito con il metodo del corto circuito virtuale e il principio di sovrapposizione degli effetti, calcolando la tensione di uscita come somma dei contributi relativi a V_1 e V_2 .

$$\begin{aligned} V_u &= -V_1 \frac{R_2}{R_1} + \frac{V_2 R_4}{R_3 + R_4} \left(\frac{R_2}{R_1} + 1 \right) \\ &= -V_1 \frac{R_2}{R_1} + \frac{V_2 R_4}{R_3 + R_4} \frac{R_2 + R_1}{R_1}. \end{aligned}$$

Se scegliamo i rapporti tra le resistenze in maniera tale che risulti $R_4/R_3 = R_2/R_1$, avremo anche

$$\frac{R_3 + R_4}{R_3} = \frac{R_2 + R_1}{R_1}$$

e quindi, moltiplicando e dividendo il coefficiente di V_2 per R_3 ,

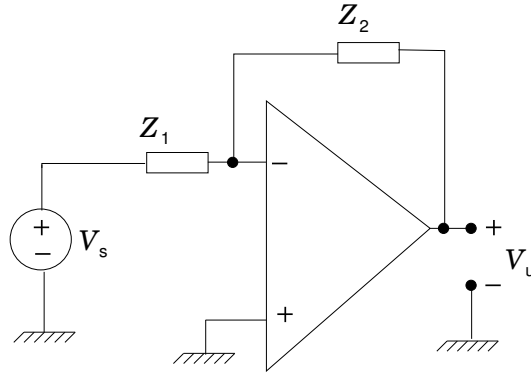
$$\begin{aligned} V_u &= -V_1 \frac{R_2}{R_1} + V_2 \frac{R_4}{R_3} \frac{R_3}{R_3 + R_4} \frac{R_2 + R_1}{R_1} \\ &= -V_1 \frac{R_2}{R_1} + V_2 \frac{R_2}{R_1} = \frac{R_2}{R_1} (V_2 - V_1). \end{aligned}$$

Se la condizione $R_2/R_1 = R_4/R_3$ è soddisfatta, otteniamo quindi un amplificatore differenziale con guadagno $A_d = R_2/R_1$ e rapporto di reiezione del modo comune (CMRR) infinito. Nella pratica il rapporto tra R_4 e R_3 non sarà perfettamente uguale a quello tra R_2 e R_1 , quindi il CMRR risulterà finito.

11.7 Integratore di Miller

Se al posto delle resistenze R_2 e R_1 in un amplificatore invertente abbiamo due generiche impedenze Z_1 e Z_2 , la funzione di trasferimento potrà essere espressa come

$$\frac{V_u}{V_s} = -\frac{Z_2(s)}{Z_1(s)}.$$



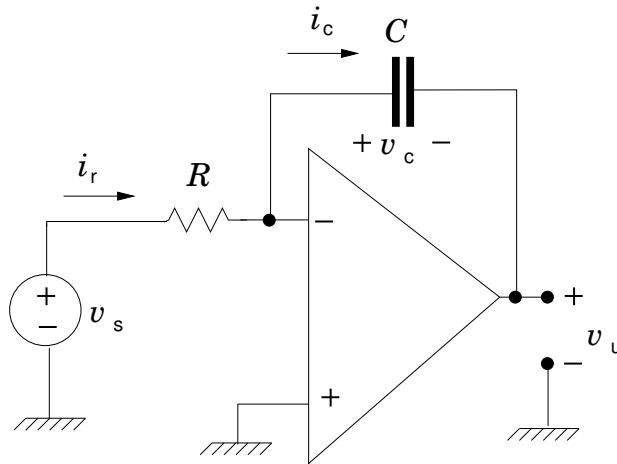
Se, in particolare, consideriamo il caso $Z_1 = R$, $Z_2 = 1/(Cs)$, vale a dire poniamo al posto di R_1 una resistenza R e al posto di Z_2 un condensatore C , la funzione di trasferimento risulterà

$$\frac{V_u}{V_s} = -\frac{1}{RCs},$$

che, ricordando le proprietà della trasformata di Laplace, corrisponde nel dominio del tempo a

$$v_u(t) = -\frac{1}{RC} \int_0^t v_s(\tau) d\tau + v_u(0).$$

Lo stesso risultato può essere ottenuto utilizzando il metodo del corto circuito virtuale nel dominio del tempo.



Poiché la tensione sull'ingresso invertente è nulla a causa del c.c.v., $i_r = v_s/R$. Quindi $i_c = i_r = v_s/R$, dato che non fluisce corrente nell'ingresso dell'operazionale (sempre per il c.c.v.). Poiché tensione e corrente ai capi di un condensatore sono legate dalla relazione

$$i_c = C \frac{dv_c}{dt},$$

otteniamo

$$\frac{dv_c}{dt} = \frac{1}{C} i_c = \frac{v_s}{RC}.$$

Integrando abbiamo

$$v_u = -v_c = -\frac{1}{RC} \int_0^t v_s(\tau) d\tau + v_u(0).$$

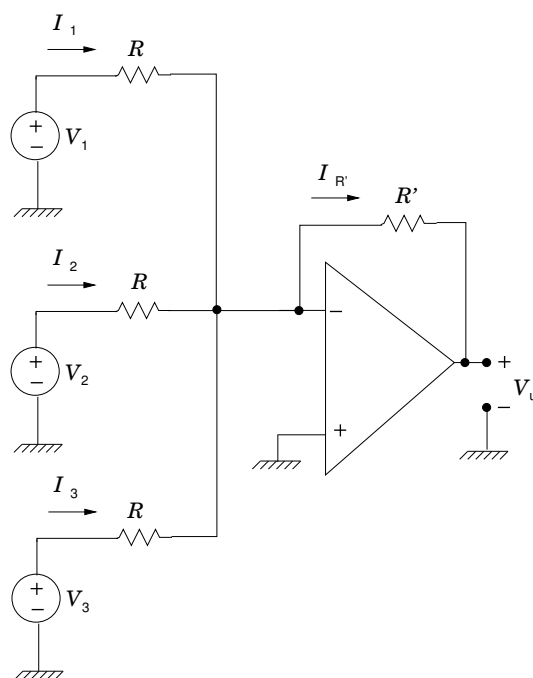
È da notare che l'integratore di Miller ideale ha un polo esattamente nell'origine e non è quindi un sistema stabile sulla base del criterio BIBO (Bounded Input – Bounded Output), infatti è sufficiente porre in ingresso una tensione continua perché in uscita si abbia una rampa che cresce senza limite. Nella realtà, in effetti, la rampa cresce finché non si raggiunge la condizione di saturazione dell'uscita, valore oltre il quale la tensione di uscita non può andare. Non essendo l'integratore un circuito stabile, deve essere sempre impiegato all'interno di un anello di reazione, che provvede a impedirne la deriva.

Dobbiamo inoltre sottolineare il fatto che il comportamento reale dell'integratore di Miller differisce da quello ideale, oltre che per il fenomeno della saturazione, anche perché il modulo di βA non è elevato a tutte le frequenze, a causa sia della caduta del guadagno dell'operazionale al crescere della frequenza sia della diminuzione di β alle basse frequenze. Altri fenomeni di non idealità derivano dal fatto che gli stadi di ingresso hanno bisogno di una certa corrente di polarizzazione, la quale dà luogo a una deriva non trascurabile della tensione di uscita, così come alla stessa contribuiscono effetti di sbilanciamento interni.

È anche possibile ottenere, scambiando tra loro il condensatore e la resistenza, un circuito derivatore, con funzione di trasferimento $A = -RCs$, ma questo non è praticamente mai usato, a causa dell'estrema sensibilità ai disturbi, specialmente quelli ad alta frequenza.

11.8 Sommatore

È possibile, utilizzando un amplificatore operazionale e alcuni componenti passivi, ottenere un circuito sommatore la cui uscita è proporzionale alla somma delle tensioni in ingresso. Il più semplice e utilizzato circuito sommatore è quello invertente di cui riportiamo lo schema per una configurazione a tre ingressi.



Possiamo analizzarne il funzionamento utilizzando il metodo del corto circuito virtuale e il principio di sovrapposizione degli effetti, facendo agire un generatore di

ingresso per volta. Se, per esempio, agisce il generatore V_1 , la corrente che scorre nella R corrispondente risulta $I_1 = V_1/R$, mentre la I_2 e la I_3 sono ambedue nulle, essendo nulla sia la tensione dovuta a V_2 e V_3 , considerati disattivati, sia quella sul terminale invertente dell'operazionale, a causa del corto circuito virtuale. Tutta la corrente I_1 fluisce dunque nella resistenza R' , dando luogo a una tensione di uscita $V_u = -R'/RV_1$. Se ripetiamo lo stesso calcolo per gli altri due ingressi, otteniamo che l'uscita, quando tutti i generatori sono attivi, è data da

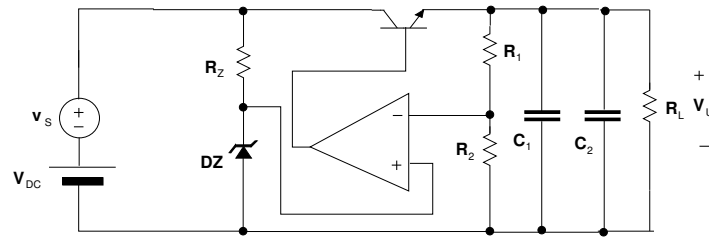
$$V_u = -\frac{R'}{R}(V_1 + V_2 + V_3).$$

L'aspetto importante di questo circuito è che separa completamente tra loro gli ingressi, grazie all'azione del corto circuito virtuale. Infatti la corrente fornita da uno dei generatori di ingresso non può raggiungere gli altri generatori, dato che il nodo a comune tra gli ingressi (corrispondente all'ingresso invertente dell'operazionale) si trova a massa virtuale.

12. Regolatori di tensione

12.1 Regolatore lineare serie

Abbiamo visto in precedenza il funzionamento di regolatori di tensione che sfruttano i diodi zener, ma avevamo osservato che tali circuiti hanno notevoli limitazioni. È possibile superare la maggior parte di tali limitazioni utilizzando un circuito nel quale la corrente fornita al carico viene regolata tramite un transistor (detto “transistore di passo”) controllato da un amplificatore differenziale tra i cui ingressi è presente la differenza tra una partizione della tensione di uscita e una tensione di riferimento. Riportiamo uno schema di principio del regolatore serie:



Il generatore V_{DC} rappresenta la sorgente di alimentazione, che può consistere in un trasformatore con ponte di Graetz e condensatore di livellamento. Il generatore v_s rappresenta le fluttuazioni della tensione in ingresso al regolatore rispetto al valor medio V_{DC} .

Si nota subito che siamo di fronte a un circuito in reazione, in cui la partizione della tensione di uscita ottenuta tramite il partitore formato da R_1 e R_2 rappresenta la grandezza di reazione. Tale grandezza di reazione viene confrontata con una tensione di riferimento ottenuta tramite un diodo zener DZ . La corrente sul diodo zener può essere mantenuta a un valore, pressoché costante di pochi milliampere, dato che in questo caso non preleviamo potenza dalla sorgente di tensione di riferimento. A regime sarà presente una differenza minima tra la tensione sull'ingresso invertente e quello non invertente, che consentirà di ottenere in uscita la tensione desiderata. Tale minima differenza tra la partizione della tensione di uscita e la tensione di riferimento deve esistere, altrimenti anche l'uscita dell'operazionale sarebbe nulla e il transistor di passo risulterebbe interdetto. Dal punto di vista pratico, però possiamo considerare, purché $|\beta A|$ sia sufficientemente grande, le tensioni sui due ingressi dell'amplificatore come coincidenti, per cui

$$V_U \frac{R_2}{R_1 + R_2} = V_Z,$$

dove V_Z è la tensione di breakdown del diodo zener DZ . Quindi la tensione di uscita risulta

$$V_U = V_Z \frac{R_1 + R_2}{R_2}.$$

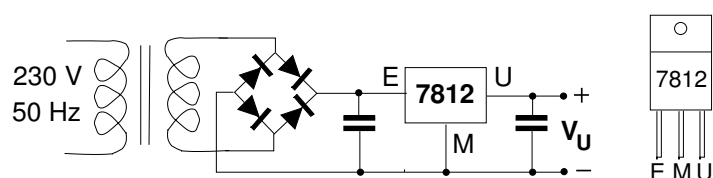
Se, per qualsiasi motivo, V_U tendesse per esempio ad aumentare, si avrebbe una diminuzione della tensione di ingresso dell'amplificatore operazionale (che, come è già stato detto, non è mai esattamente nulla, pur essendo estremamente piccola), che determinerebbe una diminuzione della tensione di uscita dell'operazionale stesso e, conseguentemente, della corrente di base del transistor di passo, traducendosi in una diminuzione della tensione di uscita, che verrebbe riportata al valore desiderato. Lo stesso accadrebbe, in senso inverso, se V_U tendesse a diminuire.

Il circuito del regolatore serie presenta una reazione di tensione, quindi la sua impedenza di uscita è molto bassa, dell'ordine dei pochi milliohm. Questo è vero alle basse frequenze, per le quali il $|\beta A|$ è molto elevato, mentre alle alte frequenze, a causa della diminuzione di $|\beta A|$, l'impedenza di uscita risulta incrementata. Tale fatto può costituire un problema, perché può portare ad accoppiamenti indesiderati tra carichi diversi connessi allo stesso alimentatore, dato che un'eventuale utenza che assorba correnti con rilevanti componenti ad alta frequenza introduce una corrispondente fluttuazione della tensione di uscita dell'alimentatore, che si ripercuote poi sulle altre utenze. Tale problema si risolve introducendo dei condensatori in parallelo all'uscita del regolatore, i quali presentano una reattanza molto bassa alle alte frequenze. Si sono indicati due condensatori perché di solito uno è di grosso valore e di tipo elettrolitico (con il dielettrico ottenuto tramite un processo elettrochimico) e presenta scadenti caratteristiche elettriche a frequenze oltre qualche decina di kHz. L'altro condensatore, non elettrolitico e di valore molto più piccolo, interviene alle frequenze più elevate, garantendo un'impedenza di uscita comunque bassa.

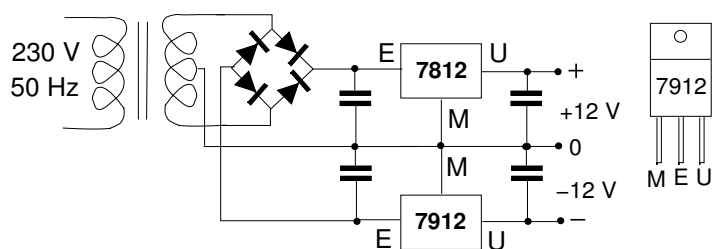
12.2 Regolatore monolitici integrati

Attualmente esistono regolatori serie integrati realizzati su un singolo chip. Tali regolatori sono definiti “regolatori monolitici” e si presentano con un contenitore molto simile a quello di un transistor di potenza, con tre terminali (entrata, uscita, massa). Questi dispositivi sono caratterizzati da una sigla del tipo 78XX, dove le cifre al posto di XX indicano il valore della tensione di uscita regolata. Per esempio, un 7815 regola a 15 V, mentre un 7805 regola a 5 V. Nella maggior parte dei casi è necessaria, per un corretto funzionamento, una caduta di tensione di almeno un paio di volt tra ingresso e uscita.

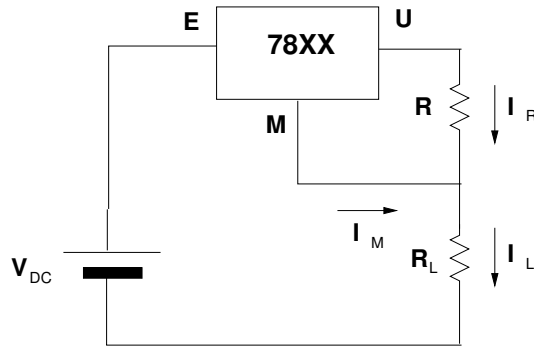
Di seguito rappresentiamo lo schema di un tipico alimentatore basato su un regolatore monolitico.



È possibile realizzare un alimentatore con uscita duale, sfruttando un regolatore monolitico anche per il ramo negativo, della serie 79XX. Come è indicato in figura la piedinatura dei 79XX è diversa da quella dei 78XX.



Si noti l'utilizzo di un unico ponte di Graetz per l'ottenimento di una tensione positiva e una tensione negativa rispetto a massa, sfruttando un trasformatore con presa



centrale. Vediamo infine come è possibile realizzare, a partire da un regolatore monolitico di tensione, un regolatore di corrente, vale a dire un circuito che fornisce in uscita una corrente di valore fissato, indipendente dal carico.

Definiamo V^* la tensione di regolazione del regolatore monolitico utilizzato ($XX = V^*$). Il regolatore manterrà tale tensione tra i terminali U ed M , per cui la corrente I_R risulterà

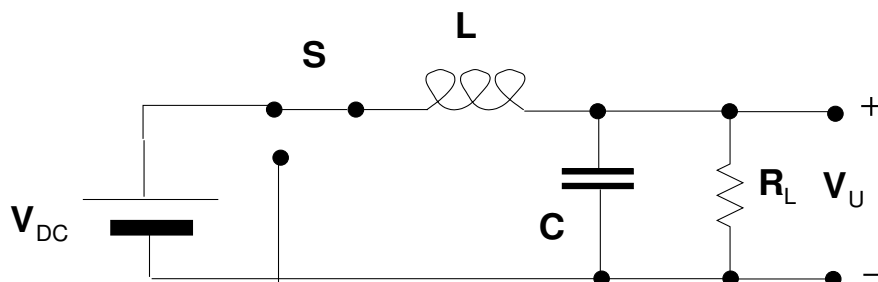
$$I_R = \frac{V^*}{R}.$$

Se tale corrente è almeno di alcune decine di milliampere, la I_M (corrente derivante dal funzionamento del regolatore e pari a pochi milliampere) sarà rispetto a essa trascurabile, per cui $I_L \simeq I_R$, indipendentemente dal valore del carico R_L . Evidentemente tutto ciò si verifica in una gamma limitata di valori di R_L : se R_L fosse di valore troppo elevato, la tensione ai suoi capi, sommata a V^* e alla caduta minima sul regolatore, porterebbe a un valore maggiore di V_{DC} , non consentendo quindi il funzionamento del circuito.

12.3 Regolatori non lineari a commutazione

I regolatori di tensione visti fino a questo punto sono di tipo lineare e presentano un problema comune, consistente nella dissipazione di potenza sull'elemento di passo, il quale è attraversato dalla corrente di uscita e ai capi del quale è presente una caduta di tensione non trascurabile. Questo è indubbiamente un problema, sia perché sprechiamo dell'energia sia perché la dissipazione termica crea dei problemi di smaltimento del calore.

L'inconveniente della dissipazione viene risolto utilizzando i cosiddetti regolatori a commutazione, i quali utilizzano una strategia basata sull'apertura e chiusura, con tempi opportuni, di un interruttore, che è un elemento non dissipativo, poiché la tensione ai suoi capi è nulla o lo è la corrente che lo attraversa. Il principio del regolatore a commutazione è rappresentato nel semplice schema che segue.

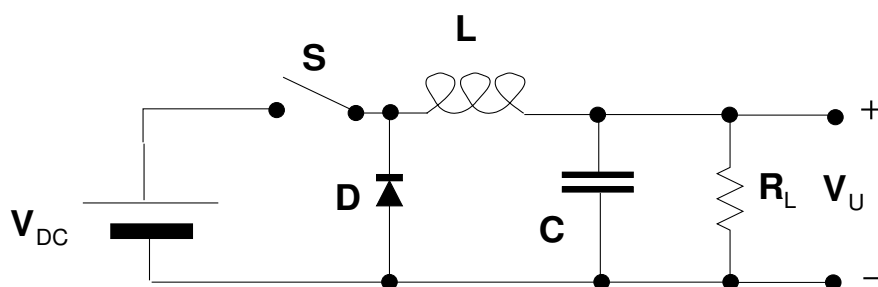


Supponiamo che il deviatore S venga rapidamente e periodicamente commutato tra le due posizioni possibili, quella che corrisponde alla connessione diretta dell'induttanza L al generatore di tensione V_{DC} e quella che corrisponde alla connessione a massa del terminale sinistro di L : la tensione tra tale terminale e la massa sarà quindi un'onda rettangolare, con valore V_{DC} nei periodi di tempo in cui il deviatore è nella posizione "alta" e di valore nullo nei rimanenti periodi di tempo. Si definisce "duty cycle" il rapporto tra il periodo di tempo in cui l'onda rettangolare sta al valore alto e il periodo totale; di solito tale quantità viene espressa sotto la forma di percentuale:

$$\text{duty cycle} = \frac{T_H}{T_H + T_L} \times 100,$$

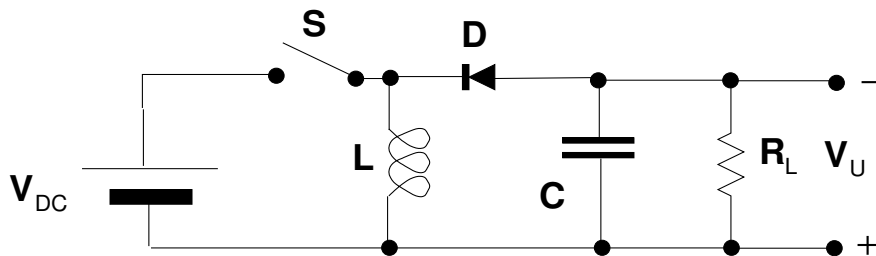
dove T_H è il periodo di tempo in cui l'onda rettangolare si trova al valore alto e T_L quello in cui si trova al valore basso.

L'induttanza L , insieme con il condensatore C , forma un filtro del secondo ordine con due poli, con una risposta quindi che, per frequenza molto maggiore di tali poli, cade con una pendenza di 40 dB per decade. Quindi, se la frequenza fondamentale dell'onda rettangolare è molto maggiore della frequenza di taglio del filtro, sul condensatore e sulla resistenza di carico R_L sarà presente soltanto la componente continua dell'onda rettangolare, il cui valore è proporzionale al duty cycle dell'onda stessa. Dunque abbiamo ottenuto una tensione continua più bassa di quella V_{DC} a disposizione, senza peraltro dissipare potenza nell'operazione di abbassamento, dato che sia il deviatore sia l'induttanza e il condensatore sono elementi non dissipativi. Se abbiamo anche la possibilità di regolare il duty cycle, possiamo ottenere una tensione di uscita del valore desiderato; inoltre rendendo automatica tale regolazione è possibile ottenere un regolatore di tensione che non presenta una significativa dissipazione di potenza. È chiaro che in un tale regolatore il deviatore non può essere di tipo meccanico, a causa delle limitazioni che si avrebbero sulla velocità massima di commutazione e delle difficoltà di controllo dello stesso: vedremo più avanti come tale deviatore può essere realizzato in forma completamente elettronica. Per il momento apportiamo una semplice modifica, che consente di utilizzare, invece di un deviatore, un semplice interruttore, secondo lo schema di seguito riportato.



Quando l'interruttore S è chiuso, il diodo D è polarizzato inversamente e quindi non interviene; quando, invece, l'interruttore S si apre, la corrente tende a mantenersi costante nell'induttanza (data l'inerzialità della stessa): questa volta il diodo risulta polarizzato direttamente e consente alla corrente di continuare a scorrere nel verso che aveva precedentemente (la corrente va dall'induttanza al parallelo C - R_L e poi torna all'induttanza tramite il diodo). Questa è la configurazione tipica del regolatore a commutazione di tipo forward, che è caratterizzato da una polarità di uscita uguale a quella della sorgente di alimentazione e da una tensione di uscita minore o uguale di quella della sorgente di alimentazione.

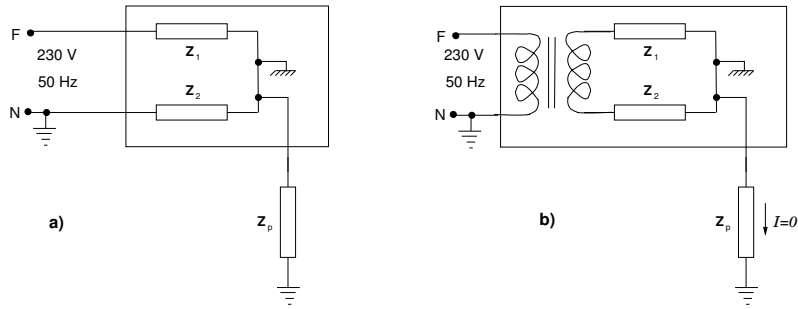
Si può realizzare un diverso tipo di regolatore switching, definito di tipo flyback, che consente di avere una polarità di uscita opposta rispetto a quella di alimentazione e una tensione di uscita anche superiore a quella di ingresso. Lo schema è riportato di seguito.



Quando l'interruttore S è chiuso, il diodo D risulta interdetto e la tensione di alimentazione si trova ai capi dell'induttanza L che si carica, con una corrente che cresce linearmente nel tempo. Quando l'interruttore si apre, la corrente che passava nell'induttanza comincia a circolare in senso antiorario nella maglia formata da L , dal parallelo C - R_L e dal diodo D che risulta ora polarizzato direttamente. In questo modo il condensatore C si carica con polarità positiva in basso e negativa in alto, a una tensione che, a seconda del duty cycle dell'interruttore e dei valori dei componenti, può essere anche maggiore in modulo di quella di alimentazione. Si noti che, mentre nel regolatore forward il carico e il condensatore C sono alimentati dalla sorgente V_{DC} durante l'intervallo in cui l'interruttore è chiuso e dall'induttanza L quando l'interruttore è aperto, nel caso del regolatore flyback la sorgente V_{DC} non alimenta mai direttamente il carico: l'energia viene trasferita da V_{DC} all'induttanza e solo successivamente dall'induttanza al carico. Negli intervalli in cui l'interruttore è chiuso l'energia assorbita dal carico è fornita dal condensatore C , che viene poi ricaricato nell'intervallo in cui l'interruttore è aperto, a spese dell'energia immagazzinata nell'induttanza.

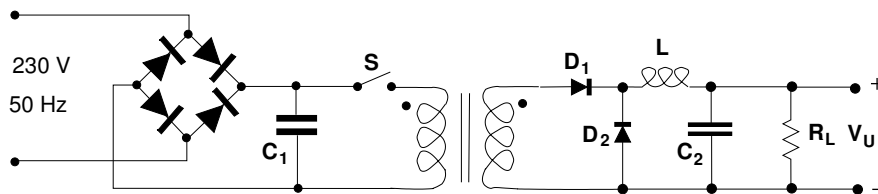
I regolatori switching finora visti necessitano di una sorgente di alimentazione continua: questa può essere ottenuta con un trasformatore (che garantisce l'isolamento galvanico dalla rete di distribuzione dell'energia elettrica) seguito da un ponte di Graetz e da un filtro capacitivo. Tale soluzione funziona ed è in certi casi utilizzata, ma mantiene uno degli svantaggi tipici degli alimentatori tradizionali, vale a dire l'elevato ingombro e l'elevato peso del trasformatore, che, dovendo operare alla frequenza di rete (50 o 60 Hz), deve avere un nucleo in ferro. Si può superare anche questo problema realizzando un alimentatore a commutazione che comprende un trasformatore operante ad alta frequenza, come vedremo tra poco. Si noti che il trasformatore è comunque necessario, altrimenti alcune parti del circuito alimentato potrebbero trovarsi al potenziale del conduttore di fase della rete o a un potenziale intermedio tra questo e la terra e rappresenterebbero pertanto dei potenziali rischi per contatti accidentali. Nella parte a) della figura seguente viene illustrata la situazione di pericolo che si viene a creare quando si tocca una qualunque parte di un'apparecchiatura connessa alla rete direttamente senza trasformatore.

L'impedenza Z_p rappresenta la persona che è sottoposta a scossa elettrica perché viene percorsa da una corrente che si richiude poi a terra, tramite i piedi. Le due impedenze Z_1 e Z_2 presenti all'interno dell'apparecchiatura rappresentano le impedenze misurabili, rispettivamente, tra il conduttore di fase e la massa dell'apparecchiatura, e tra il conduttore neutro e la massa. La massa si verrà quindi a trovare a un potenziale rispetto a terra dipendente dalla partizione tra Z_1 e Z_2 e quindi tale da



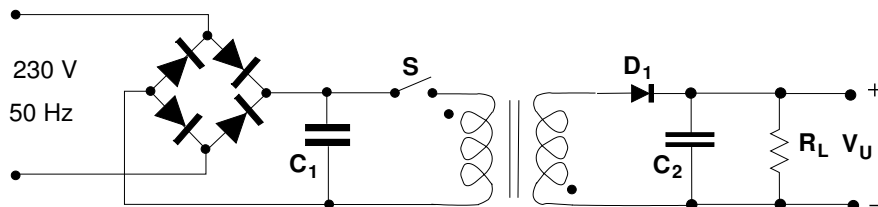
consentire il passaggio di una corrente in Z_p . In presenza di un trasformatore (b), si ha il cosiddetto “isolamento galvanico”: viene interrotto il collegamento che riferiva a terra le tensioni all’interno della apparecchiatura e l’unico modo perché l’utente possa subire una scossa elettrica è che questi tocchi contemporaneamente due punti del circuito. Si notino i due simboli diversi utilizzati per indicare la terra e la massa.

Vediamo ora come può essere combinato un trasformatore ad alta frequenza con lo schema del regolatore switching forward, in modo da ottenere un alimentatore a commutazione senza trasformatore a frequenza di rete. Uno schema possibile è quello di seguito riportato.



La tensione di rete viene raddrizzata dal ponte di Graetz e sul condensatore C_1 è presente una continua circa pari al valore di picco di quella di rete (circa 325 V). Quando l’interruttore S è chiuso, passa una corrente crescente nel primario del trasformatore; per mantenere il flusso nullo all’interno del trasformatore la corrente fuoriesce dal terminale con il pallino del secondario, determinando una polarizzazione diretta di D_1 e inversa di D_2 . La corrente passa quindi dall’induttanza L e raggiunge il carico R_L e il condensatore C_2 . Quando invece l’interruttore è aperto, il trasformatore non è più attraversato da corrente e il diodo D_2 entra in conduzione, consentendo all’induttanza L di scaricarsi sul parallelo C_2 - R_L .

Una soluzione che richiede un numero minore di componenti è quella dell’alimentatore switching di tipo flyback, sempre con trasformatore ad alta frequenza. Tale soluzione è di gran lunga la più utilizzata e lo schema di principio è riportato di seguito.

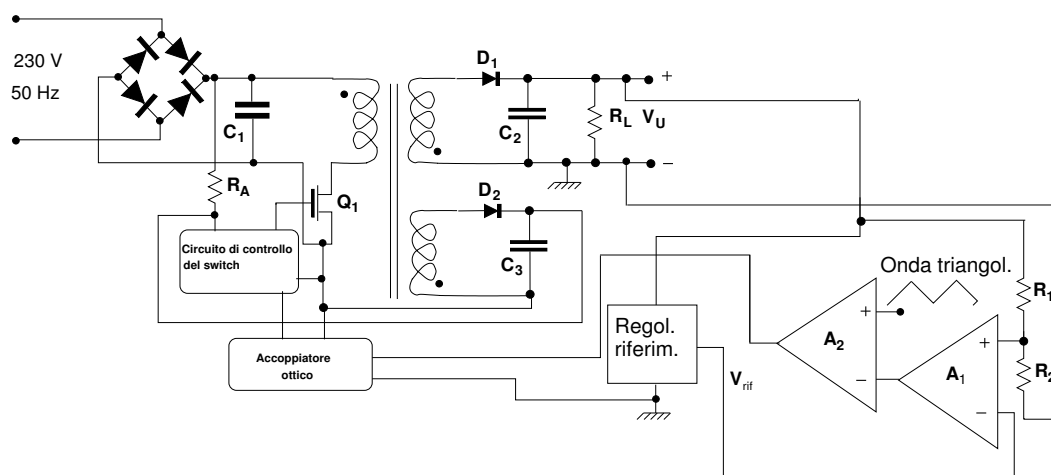


Notiamo innanzitutto che in questo caso il trasformatore non funziona più come tale, poiché, come discuteremo nel seguito, non opera a flusso nullo. Quando l’interruttore S è chiuso, la corrente sale linearmente nel primario, che si comporta a tutti gli effetti come una semplice induttanza, dato che nel secondario la corrente uscirebbe

dal pallino (essendo entrante quella del primario) ma non può farlo, perché il diodo D_1 risulta polarizzato inversamente. Quando poi S si apre, la corrente non può più circolare nel primario e l'unico modo di mantenere il flusso magnetico che era presente subito prima dell'apertura consiste nel far circolare una corrente nel secondario, entrante dal pallino. Questa corrente determina una polarizzazione diretta del diodo D_1 e va ad alimentare il condensatore C_2 e il carico. Quindi, come in tutti gli alimentatori flyback, c'è una fase (S chiuso) in cui viene immagazzinata energia in un elemento induttivo (in questo caso il trasformatore) seguita da una fase (S aperto) in cui l'energia immagazzinata nell'elemento induttivo viene trasferita al carico.

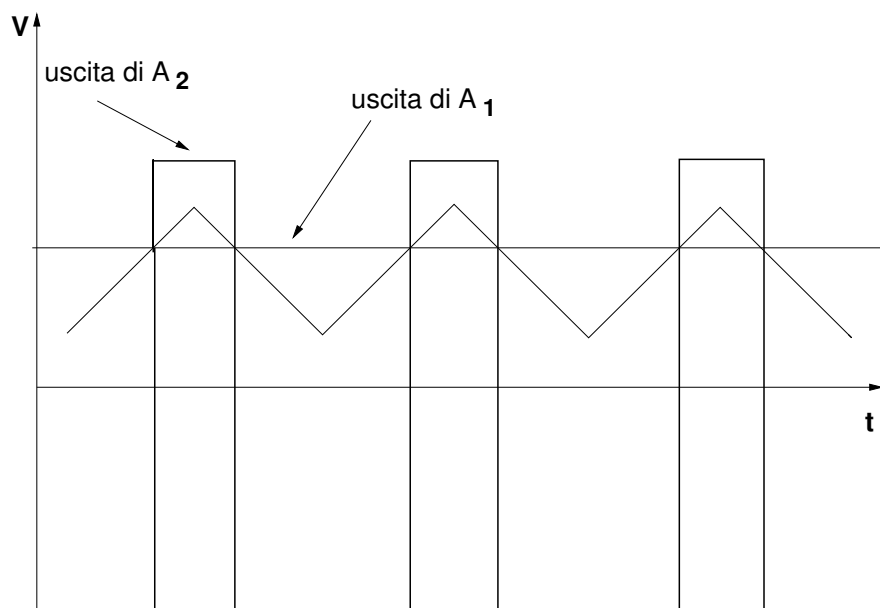
12.4 Alimentatore a commutazione flyback con trasformatore ad alta frequenza e circuito di regolazione

Uno schema di principio dell'alimentatore flyback usato, per esempio, nella maggior parte dei computer è riportato di seguito. Al circuito che abbiamo già visto viene aggiunta una parte che svolge una funzione di regolazione della tensione di uscita.



L'interruttore sul primario è realizzato tramite un transistor MOS di potenza (Q_1) che viene comandato dal circuito di controllo, la cui funzione sarà commentata più in dettaglio nel seguito.

Una porzione della tensione di uscita viene prelevata tramite il partitore formato da R_1 e R_2 e confrontata, tramite l'amplificatore differenziale A_1 , con la tensione prodotta da un riferimento di tensione (che può essere ottenuto con un diodo zener o con un regolatore monolitico di piccola potenza). L'amplificatore A_2 è un amplificatore operazionale utilizzato ad anello aperto, operante come comparatore e quindi in regime di saturazione positiva o negativa: se l'onda triangolare applicata all'ingresso non invertente ha un valore superiore a quello dell'uscita di A_1 , l'uscita di A_2 è al valore di saturazione positiva, altrimenti si trova al valore di saturazione negativa. Lo stato del transistor che opera da switch dipende dal valore di tensione in uscita da A_2 , che viene portato al circuito di controllo tramite l'accoppiatore ottico, che garantisce l'isolamento galvanico: se l'uscita di A_2 è in saturazione positiva, il transistor è in conduzione, altrimenti è in interdizione. Il duty cycle della tensione applicata al primario del trasformatore dipende quindi dal valore della tensione applicata all'ingresso invertente di A_2 , come è possibile comprendere dal grafico che segue.



Vediamo che cosa accade, per esempio, se la tensione di uscita tende a scendere: in tal caso la tensione differenziale all'ingresso di A_1 diminuisce e quindi si abbassa anche il valore della tensione sull'ingresso invertente di A_2 , che determina un incremento del duty cycle e quindi un ripristino della corretta tensione di uscita.

Al posto dell'accoppiatore ottico si potrebbe anche usare un trasformatore per impulsi, che trasferirebbe gli impulsi di comando per il switch. Anch'esso garantirebbe la separazione galvanica tra la parte di circuito connessa alla rete e quella di uscita che poi è connessa alla massa dell'apparecchiatura. Si noti che la capacità necessaria per il filtraggio all'uscita di questo alimentatore può essere molto più piccola che nel caso in cui si lavorasse alla frequenza di rete, dato che il ripple dipende dal rapporto tra il periodo della tensione pulsante e la costante di tempo $R_L C_2$. In genere le frequenze utilizzate sono intorno ai 100 kHz, in corrispondenza delle quali i trasformatori possono essere molto piccoli e leggeri, dato che per ottenere un buon accoppiamento tra primario e secondario è sufficiente un nucleo magnetico in ferrite di dimensioni relativamente ridotte.

Durante il normale funzionamento il circuito di controllo del switch è alimentato dal circuito formato dal secondario più in basso, D_2 e C_3 , che forniscono una tensione senza riferimenti rispetto al secondario principale. Ci sono tuttavia due problemi all'avvio: a) inizialmente C_3 è scarico, quindi il circuito di controllo del switch non può iniziare a funzionare; b) il circuito di regolazione con A_1 e A_2 deve essere alimentato dal secondario principale (e quindi da C_2) oppure da altro secondario ausiliario e su nessun secondario può essere presente tensione prima che inizi la sequenza di impulsi di commutazione. Il problema a) si risolve portando inizialmente l'alimentazione al circuito di controllo del switch direttamente da C_1 , tramite una resistenza di alto valore R_A , mentre il problema b) è più complesso e viene risolto realizzando tale circuito di controllo in modo che, in assenza di impulsi in arrivo dall'accoppiatore ottico, produca comunque una sequenza di comando per il switch che fa caricare i condensatori C_2 e C_3 a una tensione preferibilmente un po' inferiore a quella di regime, attivando così il normale funzionamento prima descritto. Molto spesso gli alimentatori di questo tipo sembrano guastarsi al momento in cui vengono accesi: questo è frequentemente il risultato del guasto di R_A (interruzione) avvenuto in realtà in precedenza; infatti se R_A si interrompe durante il funzionamento, l'alimentatore continua a operare correttamente perché il circuito di controllo del switch è alimentato

dal secondario ausiliario. È solo al momento della riaccensione che il sistema non parte, perché C_3 è scarico e non si può avere una sequenza di commutazioni per l'avvio se il circuito di controllo del switch non è alimentato.

L'alimentatore dei normali personal computer è una variante un po' più complessa di quello appena descritto. È da notare che per soddisfare le normative sulla compatibilità elettromagnetica, in particolare per limitare l'immissione nella rete di distribuzione dell'energia di disturbi ad alta frequenza derivanti dal processo di commutazione, viene inserito all'ingresso di tali alimentatori un filtro passa-basso che comprende due condensatori collegati tra ciascuno dei conduttori di rete e la massa del computer. Il valore di tali condensatori è in genere di 2.2 nF, corrispondenti a una reattanza, a 50 Hz, di circa 1.45 MΩ. È importante che la massa del computer sia collegata a un'efficiente presa di terra, altrimenti si verrebbe a trovare a un potenziale pari a metà della tensione di rete (115 V) rispetto al neutro e quindi alla terra! Tale potenziale non è direttamente pericoloso per le persone perché si ha in serie un'impedenza equivalente di Thevenin (il parallelo dei due condensatori) che è sufficientemente elevata da limitare la corrente a poche centinaia di microampere, ma ugualmente percepibile toccando la massa metallica del computer. Tuttavia può facilmente danneggiare componenti collocati in apparecchiature che vengano connesse al computer.

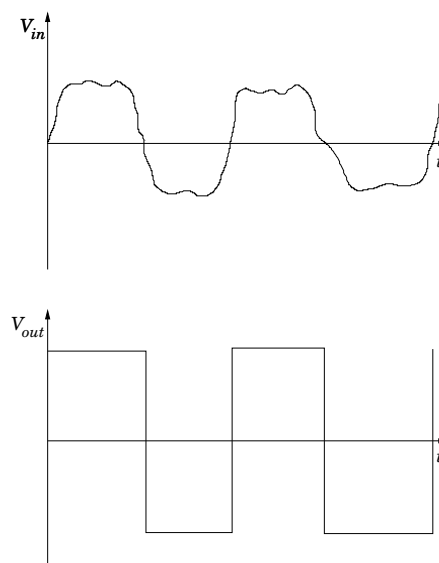
Un altro problema introdotto da tali capacità, questa volta in presenza di un efficiente collegamento di terra, è rappresentato dalla corrente che scorre dal conduttore di fase verso la terra attraverso una delle due capacità (mentre l'altra, quella collegata tra il conduttore neutro e la terra, non è attraversata da una corrente significativa, essendo tali conduttori a potenziali quasi ugali). Tale corrente non ha un corrispettivo nel conduttore neutro e quindi viene vista dagli interruttori differenziali come una componente puramente differenziale: finché si tratta di uno o pochi computer non è un problema, ma se si hanno decine di computer la componente differenziale risultante può superare la soglia di calibrazione dell'interruttore (in genere 30 mA) e determinarne l'intempestivo distacco. Tale inconveniente può essere risolto frazionando l'alimentazione in più gruppi di computer, ciascuno facente capo a un diverso interruttore differenziale oppure creando un isolamento galvanico tramite un trasformatore (soluzione assai più costosa), che rende l'alimentazione svincolata da riferimenti a terra.

13. Circuiti non lineari a operazionali

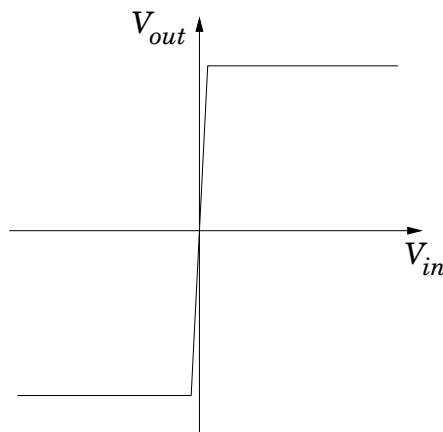
13.1 Comparatori

I circuiti comparatori, che abbiamo già incontrato all'interno degli alimentatori a commutazione, svolgono la funzione di confrontare un segnale di ingresso con un valore di soglia, fornendo un'uscita a livello alto quando il segnale di ingresso supera tale soglia e a livello basso quando è al di sotto della stessa. Sostanzialmente un comparatore è il tipo più elementare di convertitore analogico-digitale: un convertitore a un solo bit.

Un comparatore può essere utilizzato, per esempio, allo scopo di “rigenerare” un segnale digitale affetto da rumore, confrontandolo con una soglia a zero.



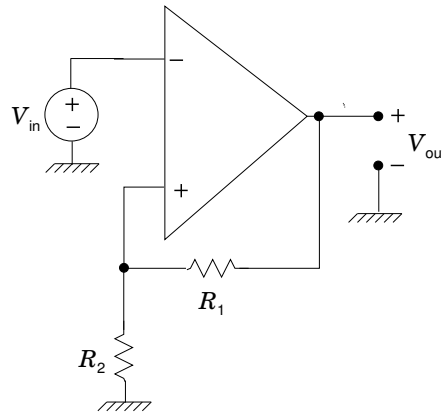
Il comparatore può essere realizzato, per esempio, utilizzando un amplificatore operazionale ad anello aperto, che, in conseguenza dell'altissimo guadagno, ha una caratteristica di trasferimento particolarmente ripida. La caratteristica risulta sostanzialmente lineare per un intervallo molto piccolo di tensioni di ingresso (pari all'escursione tra le due tensioni di saturazione di uscita divisa per il guadagno ad anello aperto) e poi satura bruscamente a valori prossimi alla tensione di alimentazione.



Utilizzando un amplificatore operazionale, però, la velocità con cui l'uscita passa dalla saturazione negativa a quella positiva dipende, nonostante la ripidità della caratteristica, dal modo in cui il segnale di ingresso attraversa lo zero. Questo può, in alcuni casi, essere un problema. Un altro problema può essere rappresentato dal fatto che, se il segnale di ingresso è affetto da un rumore che causa più attraversamenti ravvicinati dello zero, si hanno in uscita commutazioni multiple non desiderate.

Questi problemi possono essere risolti impiegando un circuito comparatore caratterizzato da un comportamento instabile prodotto tramite una reazione positiva con $|\beta A| > 1$. In tal caso la transizione tra il valore basso e quello alto dell'uscita evolve, una volta iniziata, secondo le costanti di tempo proprie dei poli a parte reale positiva, senza più dipendere in modo significativo dall'andamento del segnale in ingresso. Con questo approccio si riesce anche a ottenere una isteresi, tale da risolvere il problema degli attraversamenti multipli indesiderati.

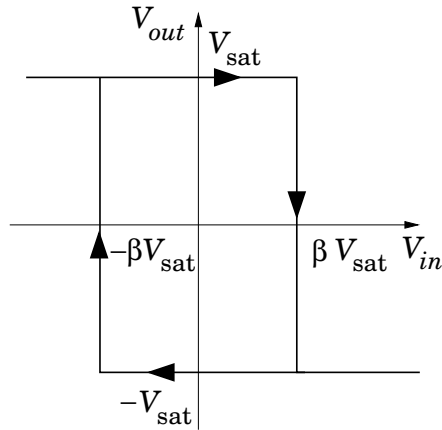
Un circuito tipico con queste caratteristiche è il trigger di Schmitt:



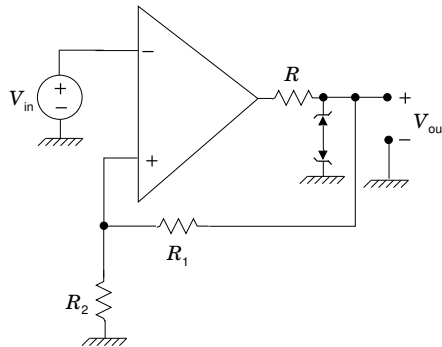
Analizziamone il funzionamento: supponiamo di avere inizialmente in ingresso una V_{in} fortemente negativa; questa determinerà in uscita una V_{out} pari al valore di saturazione positiva V_{sat} . Infatti la tensione sull'ingresso non invertente sarà pari a $R_2/(R_1 + R_2)V_{sat} = \beta V_{sat}$ (dove abbiamo definito $\beta = R_2/(R_1 + R_2)$). Quindi la tensione di ingresso dell'operazionale è significativamente positiva e determina una saturazione dell'uscita a V_{sat} , consistentemente con quanto assunto precedentemente.

Se incrementiamo la tensione di ingresso, non accade nulla finché questa non raggiunge quasi esattamente il valore βV_{sat} , portando l'operazionale in condizioni di funzionamento lineare. Non appena l'amplificatore operazionale esce dalla saturazione, l'uscita evolve in modo instabile verso il valore di saturazione opposto, $-V_{sat}$. Una volta che l'uscita è arrivata a $-V_{sat}$, la tensione sull'ingresso non invertente diventa $-\beta V_{sat}$. Pertanto, se si fa diminuire V_{in} , non si avrà più una commutazione per $V_{in} = \beta V_{sat}$, ma bisognerà aspettare che si sia raggiunto il valore $-\beta V_{sat}$, in modo che si annulli la tensione di ingresso dell'operazionale. A questo punto l'uscita tornerà a V_{sat} . Possiamo rappresentare il ciclo di commutazione con un grafico, riportato nella figura successiva.

Abbiamo quindi un ciclo di isteresi, con il valore di soglia che dipende dal verso di variazione della tensione di ingresso. La presenza del ciclo di isteresi (di ampiezza $2\beta V_{sat}$) permette di evitare commutazioni multiple dovute a fluttuazioni nell'intorno dello zero e comportamenti indesiderati consistenti in rapide commutazioni consecutive che potrebbero verificarsi, in assenza di isteresi, quando il trigger viene inserito in un anello di controllo.

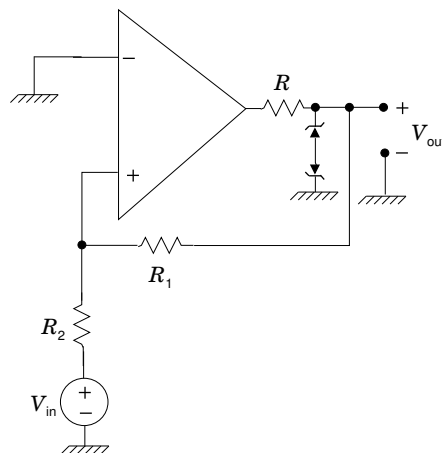


La tensione di saturazione in uscita da un amplificatore operazionale dipende da quella di alimentazione, quindi può variare e, conseguentemente, causare anche una variazione del ciclo di isteresi. Per evitare questo problema si possono inserire due diodi zener, con una opportuna resistenza di caduta:



In questo caso, purché la tensione di saturazione sia in modulo maggiore di $V_\gamma + V_Z$ (dove V_Z è la tensione di breakdown dei diodi), la tensione di uscita assumerà, indipendentemente dalla tensione di alimentazione, i valori $V_0 = V_Z + V_\gamma$ (in caso di saturazione positiva) o $-V_0 = -V_Z - V_\gamma$ (in caso di saturazione negativa).

Quello che abbiamo appena studiato è un trigger di Schmitt invertente, dato che fornisce in uscita un livello alto quando il livello del segnale in ingresso è basso. È possibile, con una semplice modifica, realizzare un trigger di Schmitt non invertente, come rappresentato nella figura seguente.



La tensione di ingresso dell'operazionale corrisponde alla tensione V^+ presente sull'ingresso non invertente (dato che quello invertente è collegato a massa), la quale può essere calcolata con il metodo di sovrapposizione degli effetti. Supponiamo di partire con una tensione di ingresso fortemente negativa; la tensione di uscita sarà quindi al valore $-V_0$, mentre V^+ risulterà

$$V^+ = V_{in} \frac{R_1}{R_1 + R_2} - V_0 \frac{R_2}{R_1 + R_2}.$$

Tale tensione sarà quindi negativa, confermando la consistenza dell'ipotesi fatta riguardo al valore di tensione di uscita. La commutazione avverrà in corrispondenza del valore di V_{in} che fa annullare V^+ :

$$V_{in} = V_0 \frac{R_2}{R_1}.$$

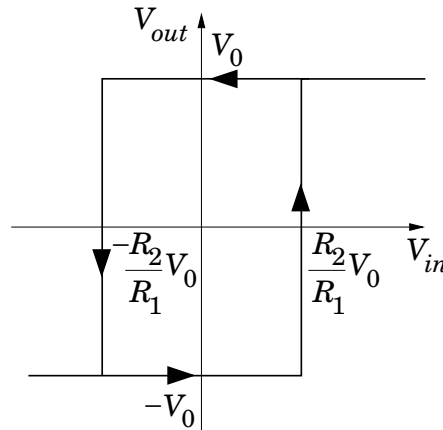
A questo punto l'uscita passa al valore V_0 e la tensione V^+ può essere espressa nella forma

$$V^+ = V_{in} \frac{R_1}{R_1 + R_2} + V_0 \frac{R_2}{R_1 + R_2}.$$

Al decrescere di V_{in} la commutazione inversa si verificherà quindi per

$$V_{in} = -V_0 \frac{R_2}{R_1}.$$

Il ciclo di isteresi avrà dunque una ampiezza $2V_0 R_2 / R_1$, come rappresentato nella figura.

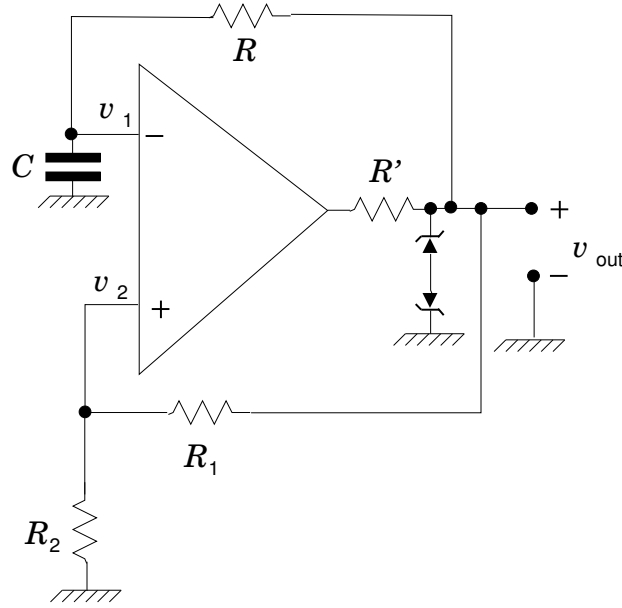


13.2 Generatore di forme d'onda quadre e rettangolari

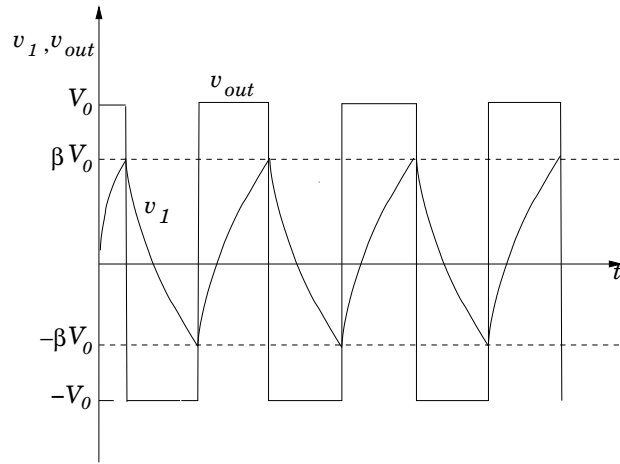
Aggiungendo un condensatore e una resistenza a un trigger di Schmitt si può ottenere un generatore di onde quadre e rettangolari.

Supponiamo il condensatore inizialmente scarico e il trigger in saturazione positiva. In tal caso

$$v_2 = \frac{R_2}{R_1 + R_2} V_0 = \beta V_0 \quad \beta = \frac{R_2}{R_1 + R_2}.$$



Il condensatore C comincia a caricarsi tramite R , tendendo alla tensione V_0 , ma quando v_1 uguaglia βV_0 il trigger scatta e l'uscita passa a $-V_0$. Di conseguenza anche v_2 cambia, passando a $-\beta V_0$ e il condensatore comincia a scaricarsi, tendendo a $-V_0$. Quando v_1 raggiunge il valore $-\beta V_0$ avviene una nuova commutazione del trigger, l'uscita torna a V_0 , v_2 a βV_0 e il condensatore inizia nuovamente a caricarsi, arrivando a βV_0 , quando, con un'ulteriore commutazione, inizia un nuovo ciclo. Rappresentiamo graficamente l'andamento di v_1 e di v_{out} .



Determiniamo ora il periodo dell'onda quadra: possiamo analizzare il semiperiodo durante il quale v_1 va da $-\beta V_0$ a βV_0 . Il transitorio di carica del condensatore può essere scritto come

$$v_1 = v_{\text{finale}} - (v_{\text{finale}} - v_{\text{iniziale}})e^{\frac{-t}{RC}}.$$

Quindi

$$v_1 = V_0 - (V_0 + \beta V_0)e^{\frac{-t}{RC}}.$$

Dopo un semiperiodo:

$$\beta V_0 = V_0 - (V_0 + \beta V_0)e^{\frac{-T}{2RC}}.$$

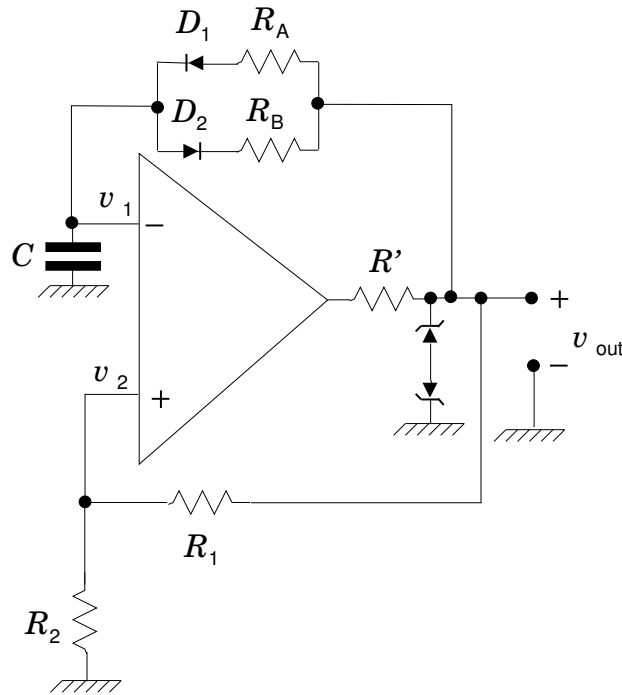
Con qualche passaggio algebrico si ottiene

$$\frac{1 + \beta}{1 - \beta} = e^{\frac{T}{2RC}}.$$

Facendo il logaritmo di ambo i membri e moltiplicando per $2RC$ otteniamo

$$T = 2RC \ln \frac{1 + \beta}{1 - \beta} = 2RC \ln \left(1 + \frac{2R_2}{R_1} \right).$$

Il duty cycle è in questo caso del 50%, quindi i semiperiodi negativi sono di durata uguale a quelli positivi. Se si desidera alterare tale simmetria, è sufficiente far sì che il valore di R sia diverso tra la fase di carica e quella di scarica del condensatore. A questo scopo è sufficiente inserire due resistenze con in serie dei diodi che selezionano l'una o l'altra resistenza a seconda del semiperiodo.



Durante la fase di carica conduce il diodo D_1 , quindi la costante di tempo è $R_A C$, mentre durante quella di scarica conduce D_2 e la costante di tempo è $R_B C$. In questo modo si possono ottenere forme d'onda rettangolari con duty cycle arbitrario (si ricorda che il duty cycle è la percentuale del periodo durante la quale l'onda rettangolare è a valore alto).

14. Il programma di simulazione SPICE

14.1 Scopo dei programmi di simulazione

I programmi per la simulazione svolgono un ruolo di importanza sempre crescente nella progettazione e nello sviluppo dei sistemi elettronici. Essi infatti consentono di ridurre notevolmente il tempo e i costi necessari per la messa a punto di un sistema, in particolare consentendo di ridurre drasticamente il numero di prototipi sperimentali di cui risulta necessaria la realizzazione. La realizzazione, caratterizzazione e conseguente modifica dei prototipi rappresentano spesso la fase più onerosa dello sviluppo di un nuovo prodotto elettronico e quella che più contribuisce all'allungamento del "time to market", particolarmente critico nei settori ad alta tecnologia.

Rivestono quindi grande importanza i programmi di simulazione in grado di fornire previsioni quantitative affidabili per il funzionamento di un circuito, tali da fornire indicazioni equivalenti, tranne che magari per il ciclo di messa a punto finale, a quelle che sarebbero ricavabili da misure su un prototipo effettivamente realizzato. Esiste in effetti una complessa gerarchia di programmi di simulazione: si parte da quelli che trattano il funzionamento del singolo dispositivo e si basano su modelli fisici dettagliati dello stesso (talvolta addirittura su principi primi, cioè senza l'utilizzo di alcun parametro fenomenologico), fornendo una descrizione molto dettagliata del comportamento del dispositivo stesso, fino ad arrivare a programmi che trattano la sintesi di sistemi a partire da circuiti con caratteristiche di ingresso e uscita ben specificate. Rimanendo nell'ambito della simulazione a livello del singolo circuito integrato, blocchi circuitali contenenti fino a qualche migliaio di transistori possono essere analizzati con programmi che svolgono un'analisi del comportamento analogico trattando i componenti con un modello analitico di tipo fenomenologico, ma se si devono prendere in considerazione strutture contenenti centinaia di migliaia o milioni di transistori la simulazione analogica dettagliata diventa proibitiva dal punto di vista del carico computazionale e si deve quindi ricorrere a simulatori logici, i quali considerano l'interazione, a livello di segnali logici, tra blocchi più elementari, ciascuno dei quali è stato esaminato in dettaglio in precedenza con un simulatore circuitali.

Nel presente capitolo ci occuperemo dei simulatori a livello circuitali, quindi intermedio tra quello dei simulatori del comportamento fisico del singolo dispositivo e quello dell'interazione logica tra blocchi di una certa complessità. Il programma di simulazione di gran lunga più utilizzato in questo campo è senz'altro SPICE (Simulation Program with Integrated Circuit Emphasis), che rappresenta ormai uno standard.

14.2 Funzionalità di SPICE

Il programma SPICE, a partire da una descrizione topologica della rete e dai valori e modelli dei componenti, è in grado di svolgere sostanzialmente tre tipi di analisi: a) la determinazione, tramite l'utilizzo di modelli realistici non lineari, del punto di lavoro in continua del circuito, b) l'analisi del comportamento in frequenza di un modello linearizzato del circuito, ricavato in corrispondenza del punto di lavoro, c) lo studio del comportamento non lineare, nel dominio del tempo, per grandi segnali. SPICE contiene quindi modelli sia lineari sia non lineari dei componenti elettronici attivi e consente, quando questo sia necessario, di definire anche resistori, condensatori e induttanze con comportamento non lineare. Nel caso di analisi non lineare nel dominio del tempo è possibile definire una vasta gamma di forme d'onda di ingresso,

che comprendono, per esempio, tutta la classe dei segnali con andamento lineare a tratti.

È importante comprendere che l'analisi linearizzata nel dominio della frequenza può essere svolta sostanzialmente in due modi diversi: l'approccio più tipico, di solito seguito nella pratica, è quello di descrivere il circuito indicando la presenza dei dispositivi attivi che effettivamente si intende utilizzare e far fare a SPICE la determinazione del punto di riposo (a) e la determinazione del circuito equivalente linearizzato da utilizzare per la fase b) di analisi in frequenza. L'altro possibile approccio consiste nel ricavare per altra via, per esempio con calcoli manuali, il punto di riposo dei dispositivi a semiconduttore e ottenerne i parametri differenziali a partire dalle caratteristiche fornite dal costruttore: in questo caso si costruisce un circuito già linearizzato, con generatori comandati, che può essere inserito direttamente nel file di ingresso di SPICE. È chiaro che nei due casi si otterranno risultati diversi, perché il circuito linearizzato ricavato internamente da SPICE sarà in genere più vicino alla realtà, perché ottenuto con tecniche più sofisticate di quelle che si possono applicare con procedure manuali. Evidentemente, nel caso in cui lo scopo dell'analisi con SPICE sia la verifica di un calcolo svolto con procedimento manuale, sarà indispensabile partire dal circuito già linearizzato con tecniche manuali.

Sono possibili ulteriori analisi, come quella del comportamento dal punto di vista del rumore, che non prenderemo in considerazione in questa sede, ma per le quali una documentazione completa può essere trovata nei vari manuali d'uso di SPICE.

14.3 Struttura della netlist

Il file contenente i dati di ingresso del programma SPICE viene di solito indicato come "netlist" e ha una struttura abbastanza semplice nella quale compaiono sostanzialmente uno statement di apertura (contenente il nome del circuito), seguito da degli statement che forniscono la descrizione della rete e poi da delle direttive che indicano i tipi di analisi richiesti e vari aspetti del formato di uscita; il file si conclude sempre con una direttiva `.END`.

In molte versioni commerciali di SPICE esiste un'interfaccia grafica che consente l'inserimento della rete elettrica da analizzare tramite il tracciamento di uno schema vero e proprio. In ogni caso, il front-end grafico produce poi una netlist del tipo che stiamo descrivendo, la cui sintassi è pressoché standard per tutte le versioni di SPICE.

La netlist può essere prodotta con un qualunque editor di testo e commenti possono essere inseriti facendoli precedere dal carattere asterisco (*).

Gli statement di descrizione della rete consistono in una lista di tutti i componenti, con l'indicazione dei loro valori e dei nodi della rete ai quali i terminali del componente sono connessi. I nodi della rete sono identificati tramite una numerazione che può essere attribuita arbitrariamente, con l'unica accortezza che il nodo di massa, quello in riferimento al quale vengono espressi tutti i valori di tensione, deve essere sempre indicato con lo zero.

14.4 Statement per la descrizione dei componenti passivi

In SPICE ciascun componente viene identificato tramite un nome (del quale sono significativi i primi 7 caratteri), la cui iniziale indica univocamente il tipo di componente. Per i componenti passivi (resistori, condensatori, induttanze) è necessario indicare i due nodi tra i quali il componente è connesso, tramite i loro numeri, e il valore del componente stesso. È anche possibile fornire ulteriori informazioni, delle

quali citeremo soltanto le principali, rimandando al manuale per ulteriori informazioni.

Il nome dei resistori deve sempre iniziare con la lettera **R** e il valore può essere espresso in ohm o nei relativi multipli e sottomultipli. Non è necessario specificare l'unità di misura, poiché se si tratta di un resistore devono essere necessariamente ohm; risulta invece necessario indicare, se si considera un multiplo o un sottomultiplo dell'ohm, il relativo prefisso moltiplicatore, secondo quanto indicato di seguito. I prefissi moltiplicatori di uso comune sono:

10^{-12} : P

10^{-9} : N

10^{-6} : U

10^{-3} : M

10^3 : K

10^6 : MEG

Si noti che i megaohm si indicano con **MEG** o **MEGOHM** e non con **M** o **MOHM**, che indicano invece i milliohm!

Lo statement descrittivo di un resistore, che definiamo **RSOURCE**, da $150\ \Omega$ connesso tra i nodi 3 e 6 risulterà quindi

RSOURCE 3 6 150

Il nome dei condensatori deve sempre iniziare con la lettera **C** e il valore viene di default assunto in farad. Per esempio, lo statement descrittivo di un condensatore **C3** da $150\ \text{nF}$ connesso tra i nodi 4 e 5 risulta:

C3 4 5 150N

Per i condensatori si può anche specificare la tensione iniziale all'istante di inizio del transitorio, aggiungendo **IC= valore**, dove con **valore** si indica proprio il valore di tale tensione in volt.

Il nome delle induttanze deve sempre iniziare con la lettera **L** e il valore viene di default assunto in henry. Lo statement per la descrizione di un'induttanza **L1** da $120\ \mu\text{H}$ connessa tra i nodi 8 e 5 risulterà dunque:

L1 8 5 120U

Anche per le induttanze è possibile specificare una condizione iniziale, come per i condensatori, ma in questo caso la grandezza da specificare è la corrente all'istante di inizio del transitorio. Anche in questo caso si aggiunge allo statement descrittivo **IC= valore**.

14.5 Statement per la descrizione dei generatori di tensione e di corrente

I generatori di indipendenti possono essere generatori di tensione, il cui nome deve iniziare con la lettera **V** o generatori di corrente, il cui nome deve iniziare con la lettera **I**. Per i generatori indipendenti si possono specificare una grande varietà di caratteristiche; di queste vedremo soltanto quelle relative ai due casi più semplici, vale a dire i generatori in continua e quelli in alternata utilizzati per la determinazione della risposta in frequenza del circuito.

Per specificare un generatore in continua è sufficiente indicare **DC** seguito dal valore in volt. Per un generatore di tensione continua abbiamo quindi in generale uno statement del tipo

Vnome N+ N- DC valore

dove N+ è il numero del nodo a cui è connesso il terminale positivo e N- è il numero del nodo al quale è connesso il terminale negativo. Nel caso di un generatore di corrente continua il terminale dal lato della “punta” della freccia viene considerato come terminale negativo e quello dal lato della “coda” viene considerato come terminale positivo. Il generico statement per la definizione di un generatore di corrente continua è quindi

Inome N+ N- DC valore

Vediamo qualche esempio: un generatore di tensione continua VCC da 5 V con il terminale positivo sul nodo 5 e quello negativo sul nodo 4, si definisce con uno statement VCC 5 4 DC 5

Nel caso di un generatore in alternata, invece di DC dovremo indicare AC, seguito da un numero che indica il modulo della tensione (o corrente) prodotta e, facoltativamente, da un valore di fase relativa (è chiaro che se abbiamo un solo generatore in alternata non ha significato indicare la fase). Un generatore di corrente alternata IS da 1 mA che “invia” corrente nel nodo 4 e la “estrae” dal nodo 8 si indicherà con

IS 8 4 AC 1M

È possibile, come già detto, specificare tutta una varietà di andamenti della tensione o della corrente in uscita dai generatori indipendenti, utili per l’analisi nel dominio del tempo, come, per esempio, forme d’onda quadre e rettangolari, sinusoidali, esponenziali, lineari a tratti, ma non entreremo nel dettaglio, che può essere trovato in un qualunque manuale di SPICE.

Si utilizzano spesso dei generatori di tensione di valore nullo, con la funzione di amperometri: questi possono essere inseriti in serie a un ramo senza perturbare la funzionalità del circuito ed è poi possibile richiedere che sia inserito tra i dati di uscita il valore della corrente che fluisce dal terminale positivo a quello negativo.

Abbiamo precedentemente discusso la possibilità di svolgere un’analisi tramite SPICE su un circuito già linearizzato, nel quale compaiono quindi generatori di tensione e di corrente comandati. Esistono quattro tipi di generatori comandati, ciascuno dei quali viene identificato in SPICE tramite un nome che inizia per una lettera diversa: E per i generatori di tensione controllati in tensione, H per i generatori di tensione controllati in corrente, F per i generatori di corrente controllati in corrente, G per i generatori di corrente controllati in tensione.

Vediamo lo statement descrittivo generico per un generatore di tensione controllato in tensione:

Enome N+ N- NC+ NC- P1

dove N+ e N- sono i nodi ai quali sono connessi, rispettivamente, i terminali positivo e negativo, NC+ e NC- sono i nodi tra i quali è presente la tensione di controllo, P1 è il coefficiente di proporzionalità tra la tensione di uscita e quella di controllo. Si potrebbero indicare, invece del solo P1, un certo numero di valori, corrispondenti ai coefficienti che descrivono una relazione polinomiale tra la tensione di controllo e quella di uscita. Nel caso in cui siano indicati più numeri, il primo viene considerato come il termine noto del polinomio, il secondo come il coefficiente del termine di ordine 1, il terzo come il coefficiente di ordine 2, ecc. Per esempio, un generatore comandato descritto con

E2 5 6 10 1 2 51 4

impone tra i nodi 5 e 6 una tensione pari a $4v^2 + 51v + 2$, dove v è la tensione tra i nodi 10 e 1. La possibilità di indicare una relazione polinomiale tra la grandezza di ingresso e quella di uscita esiste per tutti i generatori comandati.

Vediamo ora lo statement descrittivo generico per un generatore di tensione controllato in corrente:

Hnome N+ N- VC1 P1

dove **VC1** è il generatore di tensione attraverso il quale scorre la corrente di controllo.

Per un generatore di corrente controllato in corrente abbiamo

Fnome N+ N- VC1 P1

dove **VC1** è ancora il generatore di tensione attraverso il quale scorre la corrente di controllo.

Infine, il generico statement descrittivo del generatore di corrente controllato in tensione risulta

Gnome N+ N- NC+ NC- P1

14.6 Statement per la descrizione dei componenti a semiconduttore e dei sottocircuiti

Il più semplice componente a semiconduttore è il diodo: in SPICE si indica con un nome che inizia sempre con la lettera D e con uno statement del tipo riportato di seguito:

Dnome NA NC model

dove **NA** è il nodo al quale è connesso l'anodo e **NC** quello a cui è connesso il catodo. Con **model** si indica il nome del modello SPICE che deve essere contenuto in una library o definito altrove nella netlist. In genere il nome del modello corrisponde alla sigla commerciale del componente. Per esempio, se abbiamo un diodo 1N4148 connesso con il catodo sul nodo 4 e l'anodo sul nodo 6, useremo lo statement:

DPRIMO 6 4 1N4148

È anche possibile indicare ulteriori parametri, come l'area o la tensione presente ai capi del diodo all'inizio del transitorio, ma non ci soffermiamo su questi aspetti.

I BJT vengono indicati con un nome che inizia per Q e uno statement del tipo

Qnome NC NB NE [NS] modello

dove **NC** è il nodo al quale è connesso il collettore, **NB** è il nodo della base e **NE** quello dell'emettitore. Il nodo **NS** è indicato tra parentesi quadre perché può essere omesso (e di solito lo è), trattandosi del nodo a cui è connesso l'elettrodo di schermo, che è presente solo su alcuni transistori per alta frequenza. Anche in questo caso il modello deve essere presente nella library oppure definito all'interno della netlist stessa e una serie di parametri aggiuntivi possono essere specificati.

I transistori JFET vengono indicati con un nome che inizia per J e con uno statement del tipo

Jnome ND NG NS modello

dove **ND**, **NG**, **NS** sono, rispettivamente, i nodi a cui risultano collegati il drain, il gate e il source.

Analogamente i transistori MOS vengono rappresentati con un nome che inizia sempre con la lettera M e uno statement del tipo

Mnome ND NG NS NB modello

dove **NB** è il nodo a cui è connesso l'elettrodo di bulk (substrato). Per i MOS è possibile specificare un numero molto vasto di parametri addizionali e anche scegliere tra almeno tre modelli con diverso livello di sofisticazione.

Se un particolare circuito compare più volte nella stessa rete, possiamo definirlo come subcircuit e richiamarlo con uno statement del tutto analogo a quello utilizzato per gli altri componenti, indicandolo con un nome che inizia per X:

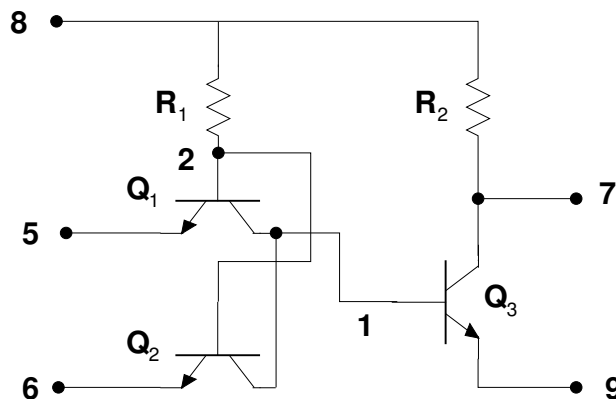
Xname N1 N2 N3 ... **sname**

dove con **sname** si è indicato il nome utilizzato nello statement descrittivo del sottocircuito. Il numero di nodi è arbitrario e dipende chiaramente dalla complessità del subcircuit. La descrizione del subcircuit si fa con una direttiva (le direttive si riconoscono per il fatto che il loro nome ha come primo carattere un punto) che inizia con

.SUBCKT **sname** NA NB NC ...

e poi contiene una normale netlist che descrive il sottocircuito. La direttiva di descrizione si conclude con uno statement **.ENDS**.

Vediamo un esempio di definizione di un subcircuit, per esempio il NAND semplificato il cui schema è riportato di seguito.



La descrizione di questo circuito risulterà

```
.SUBCKT NAND 5 6 7 8 9
Q1 1 2 5 MODTR1
Q2 1 2 6 MODTR1
R1 2 8 3K
R2 7 8 500
Q3 7 1 9 MODTR1
.ENDS
```

dove con MODTR1 si è indicato il modello per il particolare tipo di BJT utilizzato. La numerazione dei nodi all'interno del sottocircuito è del tutto libera e indipendente da quella nella netlist principale, purché non si utilizzi lo 0, che è riservato al nodo di riferimento.

Come già citato, è possibile definire un modello di dispositivo anche all'interno della netlist (quando questo non sia disponibile in una library). In tal caso si utilizza lo statement **.MODEL**, che consente di definire un particolare componente a partire da modelli analitici predefiniti per le varie tipologie di dispositivi. Non ci addentriamo nella descrizione dello statement **.MODEL**, rimandando gli interessati ai manuali di SPICE.

14.7 Direttive per l'analisi dei circuiti

In SPICE esistono una serie di direttive per specificare il tipo di analisi che si intende svolgere sul circuito descritto nella netlist. Nel seguito esamineremo le più comuni tra queste direttive.

La direttiva `.OP` viene utilizzata per il calcolo del punto di riposo: SPICE procede alla determinazione dei valori di tensione continua sui vari nodi del circuito, utilizzando per l'analisi un modello non lineare dei componenti elettronici.

Una volta determinato il punto di riposo, SPICE ricava anche il circuito equivalente lineare per le piccole variazioni. Se si vuole calcolare la risposta in frequenza di tale circuito, è possibile utilizzare la direttiva `.AC` che prevede la seguente sintassi

```
.AC [DEC] [OCT] [LIN] NP fstart fstop
```

dove l'indicazione di `DEC`, `OCT` o `LIN` indica il tipo di risoluzione utilizzato nella determinazione delle frequenze in corrispondenza delle quali viene calcolata la risposta: lineare con `LIN` e logaritmico con `DEC` e `OCT`. Con `NP` si indica il numero di punti totale nel caso di risoluzione lineare o di punti per decade o per ottava se preceduto, rispettivamente da `DEC` o `OCT`. Di solito si usa una risoluzione logaritmica e si utilizza `DEC`, specificando il numero di punti per decade. Per esempio, uno statement del tipo

```
.AC DEC 30 10 100K
```

indica che deve essere effettuata un'analisi in alternata con risoluzione logaritmica con 30 punti per decade da 10 Hz a 100 kHz.

Se vogliamo invece svolgere un'analisi nel dominio del tempo, utilizzando un modello non lineare e valido quindi anche per grandi segnali, dobbiamo usare la direttiva `.TRAN`, per la quale lo statement generale è nella forma

```
.TRAN tstep tstop [tstart] [tmax] [UIC]
```

Il numero `tstep` indica il passo temporale con il quale si vuole che sia rappresentato il transitorio, `tstop` indica l'istante di fine del transitorio. Se `tstart` non è indicato, viene assunto pari a 0. Se si indica un ulteriore numero oltre a `tstart`, questo viene interpretato come il passo massimo che SPICE deve utilizzare per l'integrazione numerica delle equazioni integro-differenziali che descrivono il circuito. Questo è un parametro molto importante, dato che se viene scelto troppo grande il risultato può anche essere completamente sbagliato. Se `tmax` non è indicato, SPICE lo calcola in modo automatico, ottenendo un risultato di solito valido, ma che per circuiti particolari può non essere adeguato. Questo è il motivo per cui è possibile specificarlo manualmente: chiaramente deve essere sempre minore di `tstep` e si può trovare il valore appropriato riducendolo progressivamente finché non si osservano più variazioni nei risultati. Infine, se si include `UIC` nello statement, verranno prese in considerazione le condizioni iniziali eventualmente indicate per i condensatori e per le induttanze, che altrimenti vengono assunte nulle.

14.8 Direttive di uscita

Esistono delle direttive per specificare il tipo di dati che si desidera avere inclusi nel file di uscita. Per esempio è possibile ottenere una tabella dei valori ottenuti nelle varie analisi tramite lo statement `.PRINT`, che ha una sintassi del tipo

```
.PRINT type var1 var2 ...
```

dove `type` indica il tipo di analisi (`AC`, `TRAN`, ecc.). Le variabili indicate corrispondono alle grandezze a cui siamo interessati, che possono essere tensioni su nodi del circuito o correnti che attraversano generatori di tensione. Viene prodotta una tabella nella quale ogni riga contiene un valore di frequenza (nel caso di analisi `AC`) o di tempo (nel caso di analisi `TRAN`), seguito dai relativi valori delle variabili elencate. Per esempio,

```
.PRINT AC V(1), VP(1), I(VCC)
```

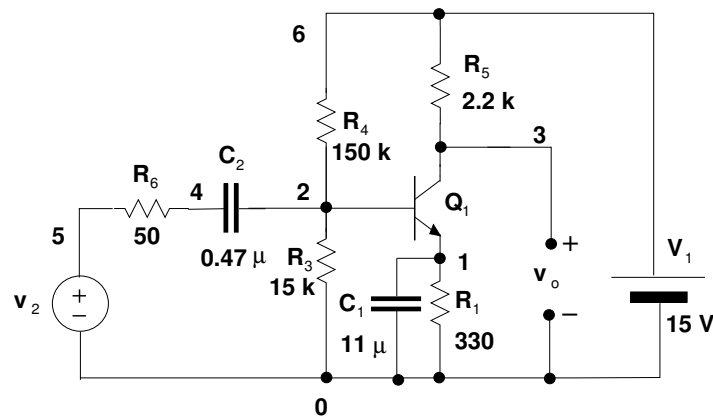
indica che deve essere prodotta una tabella con i valori dell'analisi in frequenza del modulo della tensione sul nodo 1, della fase della tensione sul nodo 1 e della corrente che attraversa il generatore VCC.

Un'altra direttiva di uscita, ormai non più molto utilizzata, vista la disponibilità di tool grafici per l'esame dei risultati della simulazione, consiste nello statement .PLOT, che consente di ottenere un grafico a caratteri ASCII delle grandezze specificate con una sintassi analoga a quella dello statement .PRINT. Per esempio,

```
.PLOT TRAN V(1), V(2), I(VCC)
```

determina il tracciamento del grafico in funzione del tempo delle tensioni sui nodi 1 e 2 e della corrente attraverso il generatore VCC.

Vediamo un esempio di descrizione di un semplice circuito, in particolare un amplificatore a emettitore comune.



Innanzitutto numeriamo i nodi nel modo indicato, attribuendo l'indice 0 al nodo di riferimento. Possiamo poi scrivere la netlist, nella quale possono essere anche inserite righe di commento, precedute da asterisco:

```
CIRCUIT COMMON_EMITTER
```

```
* QUESTO CIRCUITO E' UN AMPLIFICATORE A EMTTITORE COMUNE
```

```
.OP
```

```
.AC DEC 30 10K 10MEG
```

```
.PRINT AC V(3) VP(3)
```

```
.PRINT DC V(3)
```

```
R1 1 0 330
```

```
C1 1 0 11 UF
```

```
R3 2 0 15K
```

```
R4 2 6 150K
```

```
R5 3 6 2200
```

```
C2 2 4 0.47U
```

```
R6 4 5 50
```

```
V2 5 0 AC
```

```
V1 6 0 DC 15
```

```
Q1 3 2 1 BC109C
```

```
.model BC109C NPN(Is=7.049f Xti=3 Eg=1.11 Vaf=28.14 Bf=677 Ise=7.049f
```

```
+ Ne=1.38 Ikf=96.23 Nk=.5 Xtb=1.5 Br=2.209 Isc=250.3p Nc=2.002
```

```
+ Ikr=10.73 Rc=1.433 Cjc=5.38p Mjc=.329 Vjc=.6218 Fc=.5
```

```
+ Cje=11.5p Mje=.2717 Vje=.5 Tr=10n Tf=437.8p Itf=3.097
```

```
+      Xtf=12.85 Vtf=10)
.END
```

È stato inserito anche il modello del BC109C, a scopo esemplificativo. Di solito non è necessario inserire il modello nella netlist, perché questo è già contenuto in una library fornita insieme con SPICE. Una riga che inizia con il simbolo + viene interpretata come continuazione della precedente.

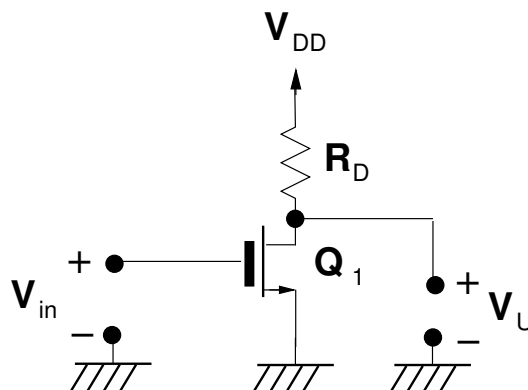
15. Concetti di base sui circuiti digitali

15.1 Introduzione

Come avevamo già visto quando abbiamo trattato i circuiti logici a diodi, la rappresentazione di segnali digitali tramite grandezze intrinsecamente analogiche, come i valori di tensione o di corrente in un circuito, richiede la scelta di un'opportuna convenzione con la quale si associno i valori logici 0 e 1 a intervalli distinti di valori della variabile analogica considerata. Per esempio, possiamo definire come 1 logico qualunque tensione compresa tra una certa soglia e la tensione di alimentazione e come 0 logico qualunque tensione compresa tra lo zero e una data soglia, inferiore a quella relativa all'1. Valori intermedi tra le due soglie corrispondono a un livello logico indeterminato.

Nella logica a diodi avevamo incontrato vari problemi, tra i quali il problema del decadimento dei livelli logici attraverso una catena di porte in cascata, a causa delle cadute di tensione sui diodi, e quello dell'assenza di una porta a diodi in grado di effettuare l'operazione NOT di inversione logica.

Tali inconvenienti possono essere superati utilizzando componenti a tre terminali che presentano un guadagno di potenza. In particolare, soffermiamoci per il momento sul problema della porta NOT o "inverter". Possiamo pensare a una soluzione estremamente semplice come quella dello schema di seguito riportato, in cui, quando la tensione di ingresso è inferiore a quella di soglia del transistor MOS, questo risulta interdetto e la tensione sul drain è pari a quella di alimentazione. Se invece la tensione di ingresso è pari a quella di alimentazione, il transistor MOS si comporta sostanzialmente come un interruttore chiuso e la tensione sul drain è molto vicina allo zero (perché la resistenza di canale del MOS risulta molto più piccola di R_D).

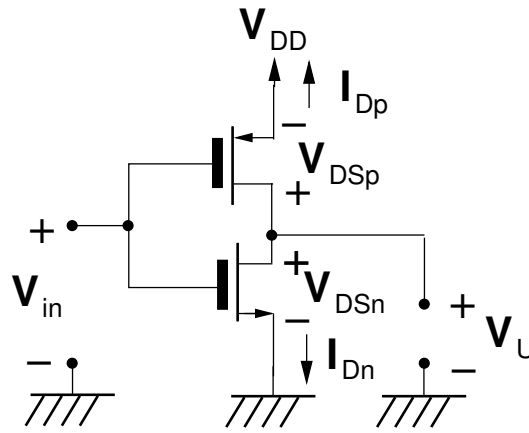


Questo circuito svolge dunque la funzione logica di un inverter, ma presenta un inconveniente significativo: mentre nella condizione corrispondente allo stato logico alto in uscita la dissipazione di potenza è praticamente trascurabile (essendo la corrente pressoché nulla), quando in uscita abbiamo uno stato logico basso una potenza

V_{DD}^2/R_D viene dissipata. In un circuito con centinaia di migliaia o milioni di porte logiche questo può essere un problema significativo. Per questo motivo è stata sviluppata una soluzione diversa, discussa nel paragrafo successivo, che consente di evitare la dissipazione di potenza, eccetto che durante il transitorio di passaggio da uno stato all'altro.

15.2 L'inverter CMOS

La soluzione al problema della dissipazione di potenza in condizioni statiche da parte dell'inverter consiste nell'utilizzo, oltre a un transistor MOS a canale n , anche di un transistor MOS a canale p , realizzando la cosiddetta configurazione MOS complementare (Complementary MOS, CMOS), di seguito illustrata.

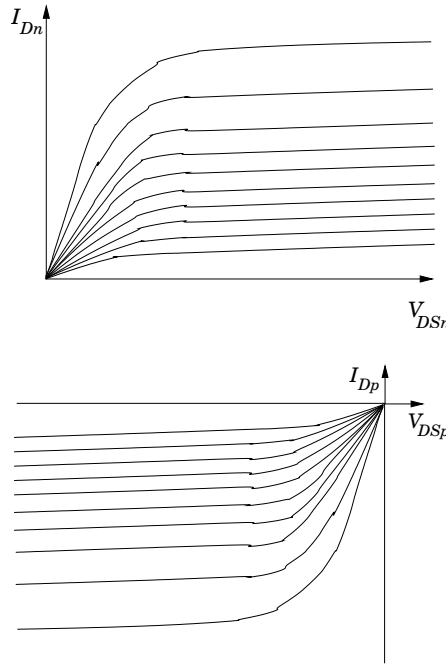


Quando la tensione di ingresso è bassa, il transistor NMOS si comporta come un interruttore aperto, mentre quello PMOS ha una tensione V_{GS} negativa e in modulo pari alla V_{DD} , per cui si comporta come una resistenza di basso valore. Pertanto possiamo considerare il transistor superiore come un interruttore chiuso e quello inferiore come un interruttore aperto, ottenendo una tensione di uscita pari alla V_{DD} . Se invece l'ingresso si trova a tensione V_{DD} , il transistor NMOS si comporta come un interruttore chiuso e quello PMOS, avendo una V_{GS} nulla, come un interruttore aperto. Dunque il comportamento è quello desiderato di un inverter, perché si ha un'uscita alta per un ingresso basso e un'uscita bassa per un ingresso alto. Inoltre la dissipazione di potenza è pressoché nulla in tutti e due gli stati, perché i transistori sono alternativamente attraversati da corrente nulla o hanno una tensione V_{DS} praticamente nulla, per cui il prodotto corrente-tensione risulta trascurabile nei due stati.

Cerchiamo ora di ricavare la caratteristica di trasferimento ingresso-uscita per una porta CMOS, partendo dalle caratteristiche di uscita dei due transistori che la costituiscono e che riportiamo di seguito.

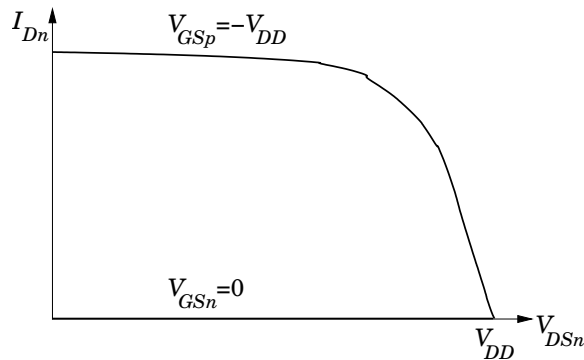
Possiamo scrivere alcune relazioni che legano tra loro le correnti e le tensioni sui due transistori:

$$\begin{cases} I_{Dn} = -I_{Dp} \\ V_{DSn} = V_{DD} + V_{DSp} \\ V_{in} = V_{GSn} \\ V_u = V_{DSn} \end{cases}$$



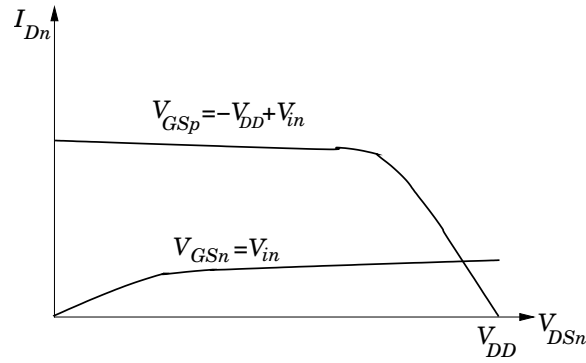
Vediamo come rappresentare contemporaneamente le caratteristiche del transistor p e di quello n sullo stesso grafico. Se assumiamo come coordinate V_{DSn} e I_{Dn} , il grafico delle caratteristiche di uscita del transistor p deve essere innanzitutto ruotato intorno all'asse delle ascisse dato che $I_{Dn} = -I_{Dp}$ e deve essere poi traslato verso destra di V_{DD} , dato che $V_{DSn} = V_{DD} + V_{DSp}$. Per trovare il punto di lavoro dovremo intersecare la caratteristica per il transistor n per $V_{GSn} = V_{in}$ e quella del transistor p per $V_{GSp} = -V_{DD} + V_{in}$.

Iniziamo dalla condizione $V_{in} = 0$: in questo caso $V_{GSn} = 0$ e $V_{GSp} = -V_{DD}$. Tracciamo soltanto le corrispondenti caratteristiche sul piano V_{DSn}, I_{Dn} e individuiamo l'intersezione.

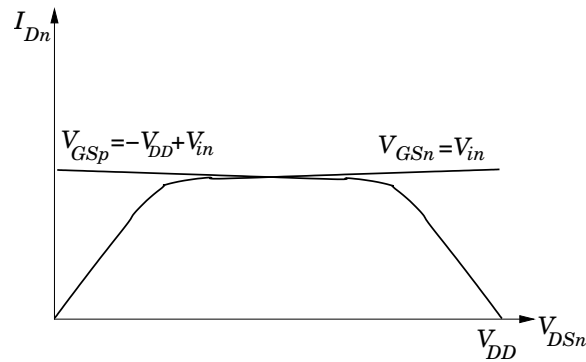


L'intersezione è in corrispondenza del punto ($V_{DSn} = V_{DD}, I_{Dn} = 0$), per cui, come precedentemente discusso, l'uscita si trova a V_{DD} , quindi a livello logico alto. Questa rimane la situazione finché la tensione di ingresso non raggiunge il livello di soglia per il transistor a canale n e tale transistor comincia a condurre, cosicché la relativa caratteristica sul grafico V_{DSn}, I_{Dn} comincia a sollevarsi, mentre quella del transistor a canale p continua a scendere (dato che $V_{GSp} = -V_{DD} + V_{in}$).

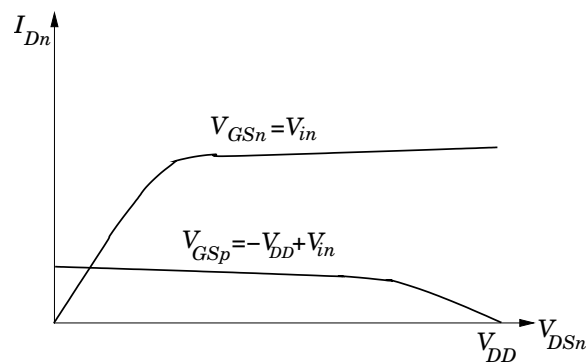
In questa situazione la tensione di uscita comincia a diminuire, il transistor di tipo p è in zona triodo e quello n in saturazione. Se la tensione di ingresso aumenta ulteriormente anche il transistor p raggiunge la condizione di saturazione e la tensione



di uscita varia molto rapidamente, data la quasi orizzontalità delle caratteristiche in zona di saturazione, che porta a notevoli spostamenti del punto di intersezione per piccole variazioni di V_{in} .

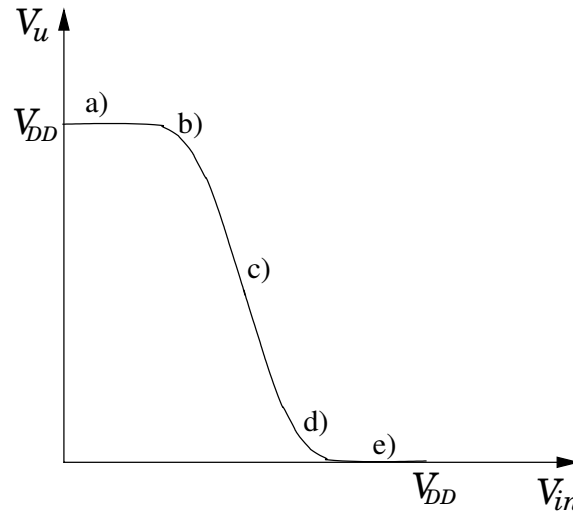
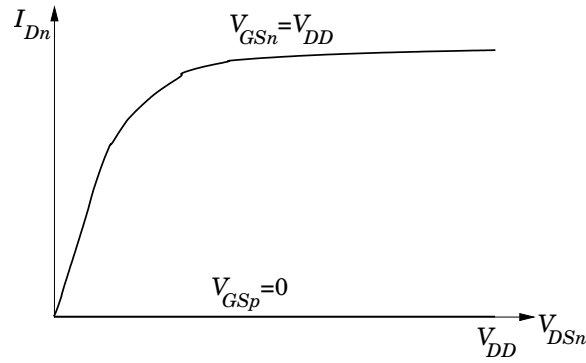


In corrispondenza di $V_{in} = V_{DD}/2$ le due caratteristiche sono simmetriche. Per un ulteriore incremento della tensione di ingresso, il punto di intersezione si sposta più a sinistra e la tensione di uscita scende al di sotto di $V_{DD}/2$. Per un valore abbastanza alto della tensione di ingresso il transistor n si viene a trovare in zona triodo mentre quello p rimane in saturazione.



Infine, quando il modulo di V_{GSp} scende al di sotto di quello della soglia di conduzione del transistor p , quest'ultimo si interdice e la tensione di uscita diventa nulla.

Possiamo quindi tracciare la caratteristica ingresso-uscita dell'inverter CMOS, nella quale individuiamo sostanzialmente cinque zone distinte: una zona a) in cui il transistor n è interdetto e il transistor p è in zona triodo, una zona b) in cui il transistor n è in saturazione e il transistor p è in zona triodo, una zona c) in cui ambedue i transistori sono in saturazione, una zona d) in cui il transistor n è in zona triodo e



il transistor p è in saturazione, una zona e) in cui il transistor n è in zona triodo e il transistor p è interdetto.

La zona c) risulterà tanto più ripida quanto più sarà alta la resistenza differenziale dei transistori nella zona di saturazione. La pendenza della caratteristica ingresso-uscita corrisponde infatti al guadagno del circuito. Vedremo nel paragrafo successivo come la presenza di un guadagno significativo consenta di ottenere la rigenerazione dei livelli logici.

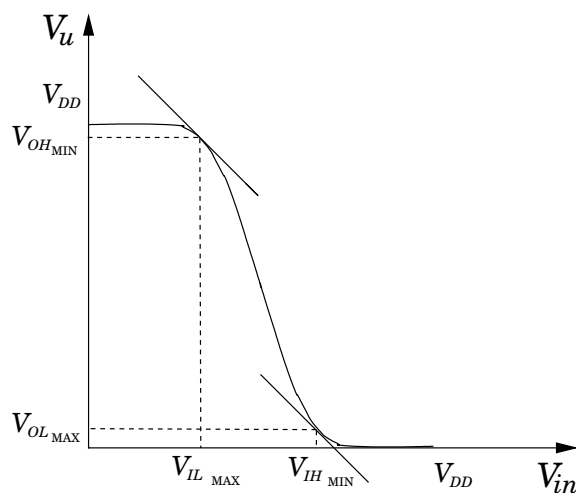
Nella discussione dell'inverter CMOS abbiamo assunto i due transistori tra loro simmetrici: poiché la mobilità delle lacune è significativamente minore di quella degli elettroni, se i transistori fossero costruttivamente simmetrici non lo sarebbero elettricamente. Per questo motivo il transistor a canale p viene realizzato con un rapporto W/L (ricordiamo che W è la larghezza del canale e L è la lunghezza dello stesso, che di solito è uguale per i due transistori) maggiore di quello del transistor n di un fattore corrispondente al rapporto inverso delle mobilità:

$$\frac{\left(\frac{W}{L}\right)_p}{\left(\frac{W}{L}\right)_n} = \frac{\mu_n}{\mu_p}.$$

15.3 Parametri caratteristici dei circuiti digitali

Definiamo ora alcuni dei parametri utilizzati per la caratterizzazione dei circuiti digitali e per determinare l'interoperabilità tra gli stessi. Come abbiamo visto in precedenza, il livello logico 1 viene codificato con una tensione "alta", superiore a una

soglia che definiamo $V_{H_{MIN}}$ e il livello logico 0 con una tensione “bassa”, inferiore a una soglia che definiamo $V_{L_{MAX}}$. Con riferimento alla caratteristica di trasferimento dell’inverter CMOS, identifichiamo i due punti in corrispondenza dei quali la tangente alla caratteristica ha pendenza unitaria: per tensioni di ingresso comprese nell’intervallo tra tali punti il guadagno per piccoli segnali dell’inverter è superiore all’unità, mentre per tensioni di ingresso al di fuori di questo intervallo il guadagno per piccoli segnali risulta minore dell’unità. Ciò significa che al di fuori dell’intervallo in questione variazioni della tensione di ingresso danno luogo a variazioni della tensione di uscita di ampiezza minore. Definiamo $V_{OH_{MIN}}$ la tensione di uscita corrispondente al punto con tangente unitaria più a sinistra e $V_{IL_{MAX}}$ la corrispondente tensione di ingresso. Definiamo altresì $V_{OL_{MAX}}$ la tensione di uscita corrispondente all’altro punto con tangente unitaria e $V_{IH_{MIN}}$ la corrispondente tensione di ingresso.



La tensione di uscita $V_{OH_{MIN}}$ viene assunta come la minima per la quale si considera presente in uscita un 1 logico, mentre $V_{OL_{MAX}}$ è la tensione massima per la quale si considera presente in uscita uno 0 logico. Le tensioni $V_{IH_{MIN}}$ e $V_{IL_{MAX}}$ rappresentano le corrispondenti grandezze in ingresso: si considera presente in ingresso un 1 logico se la tensione è superiore a $V_{IH_{MIN}}$ e si considera invece presente uno 0 logico se la tensione è inferiore a $V_{IL_{MAX}}$.

Notiamo che risultano

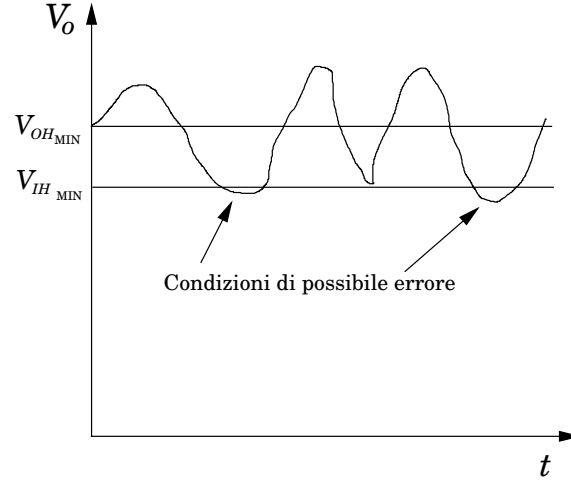
$$V_{OL_{MAX}} < V_{IL_{MAX}},$$

$$V_{OH_{MIN}} > V_{IH_{MIN}},$$

per cui siamo certi che un 1 logico in uscita viene riconosciuto come tale dall’ingresso della porta successiva e che lo stesso accade per uno 0 logico. In sostanza, si ha una “rigenerazione” dei livelli logici, invece di un deterioramento come quello che si verifica nel caso della logica a diodi. Per ottenere tale risultato è necessario che il guadagno del circuito utilizzato per realizzare la porta sia maggiore dell’unità almeno in un intervallo di ampiezza finita di valori della tensione di ingresso, come accade nel caso della porta logica CMOS tra le due tangenti con pendenza unitaria che abbiamo tracciato, infatti una variazione di tensione in ingresso tra gli estremi dell’intervallo di incertezza tra $V_{IL_{MAX}}$ e $V_{IH_{MIN}}$ deve tradursi almeno in una variazione di ampiezza pari al più grande intervallo tra $V_{OL_{MAX}}$ e $V_{OH_{MIN}}$.

Consideriamo ora il problema che deriva dalla presenza di disturbi che possono nascere da accoppiamenti di tipo capacitivo o induttivo con sorgenti esterne. Esaminiamo il caso di due porte in cascata e poniamoci nei due casi peggiori (worst-case)

per i due valori logici di uscita. Nel caso di uscita della prima porta al livello logico 1, supponiamo che questa sia rappresentata da una tensione esattamente pari a V_{OH_MIN} : se il disturbo sovrapposto a tale tensione di uscita ha ampiezza maggiore di $V_{OH_MIN} - V_{IH_MIN}$, non sarà più garantito il fatto che la seconda porta veda in ingresso un 1 logico. Viceversa, se consideriamo il caso di uno 0 logico in uscita dalla prima porta, non sarà garantito che all'ingresso della seconda porta sia visto uno 0, se l'ampiezza dei disturbi supera $V_{IL_MAX} - V_{OL_MAX}$.



Si definiscono quindi due parametri, detti margini di rumore sul livello alto (NM_H) e margine di rumore sul livello basso (NM_L), corrispondenti a

$$NM_H = V_{OH_MIN} - V_{IH_MIN}$$

e

$$NM_L = V_{IL_MAX} - V_{OL_MAX}.$$

Tali margini di rumore rappresentano l'ampiezza massima accettabile per il rumore totale sovrapposto al segnale logico e danno dunque un'idea dell'immunità ai disturbi offerta da una particolare tecnologia.

Un'altra quantità comunemente usata è rappresentata dall'escursione logica o "logic swing", che risulta definita tramite

$$LS = V_{OH_MIN} - V_{OL_MAX}.$$

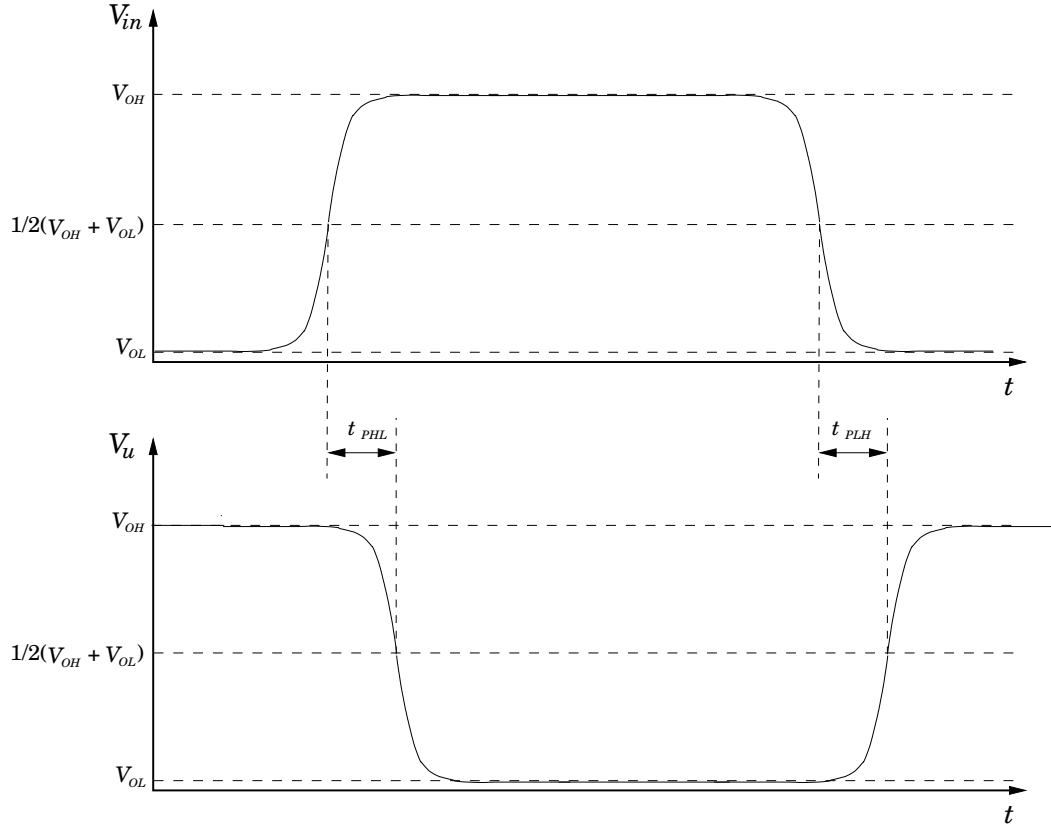
Possiamo anche definire una potenza dissipata, intendendo come tale la potenza che viene dissipata sulla porta in condizioni statiche. Poiché si assume che la porta si trovi il 50% delle volte nello stato alto e il restante 50% nello stato basso, la potenza media dissipata può essere ottenuta mediando le quantità relative ai due stati logici:

$$P_D = \frac{P_{DH} + P_{DL}}{2} = V_{DD} \frac{I_H + I_L}{2},$$

dove I_H e I_L sono, rispettivamente, le correnti assorbite nello stato logico alto e in quello basso dalla sorgente di alimentazione V_{DD} . È importante notare che in genere tale potenza dissipata in condizioni statiche è trascurabile rispetto a quella dissipata

durante i transitori di commutazione da uno stato all'altro, che esamineremo nel seguito.

In conseguenza della presenza di capacità sull'uscita della porta logica, che devono essere caricate per passare da uno stato logico all'altro, la commutazione dell'uscita avviene inevitabilmente con un certo ritardo rispetto a quella dell'ingresso, come illustrato nella figura seguente, per il caso di un inverter.

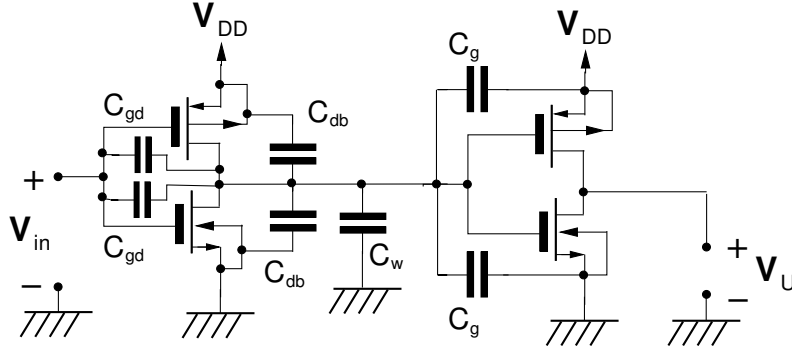


Se si prende come riferimento il valor medio tra il livello logico alto e quello basso in uscita, si può definire un ritardo tra l'istante in cui la tensione di ingresso attraversa il livello di riferimento e quello in cui esso è attraversato dalla tensione di uscita. In generale è possibile che tale ritardo sia diverso quando l'uscita passa dal livello alto a quello basso rispetto a quando l'uscita passa dal livello basso a quello alto: per tale motivo si definiscono due ritardi di propagazione: t_{PHL} per la transizione dal livello alto a quello basso e t_{PLH} per la transizione opposta. Si definisce anche un ritardo medio, corrispondente alla media dei due ritardi

$$t_P = \frac{t_{PHL} + t_{PLH}}{2}.$$

Esaminiamo ora quali sono i varî contributi alla capacità sull'uscita di una porta CMOS, che determina il valore del ritardo di propagazione. Consideriamo due inverter in cascata e vediamo quali sono le capacità che interessano l'uscita del primo e di conseguenza l'ingresso del secondo.

Esistono le capacità C_{db} tra drain e bulk, le quali sono connesse tra il punto a comune dei drain e la massa (per il transistor n) o l'alimentazione positiva (per il transistor p), essendo il bulk di tutti i transistori p in un circuito integrato connesso all'alimentazione positiva e quello di tutti i transistori n alla massa. Da un punto di vista del transitorio, questi due condensatori possono essere considerati in parallelo.



In parallelo a essi sono anche presenti la capacità C_w del collegamento tra l'uscita del primo inverter e l'ingresso del secondo e le due capacità di gate C_g viste sull'ingresso del secondo inverter. Un po' più complesso è il problema relativo al trattamento delle C_{gd} tra gate e drain dei transistori del primo inverter: considerando che, in corrispondenza della commutazione dell'inverter da uno stato all'altro, la tensione ai loro capi varia di $2V_{DD}$, possiamo considerare ragionevole sostituirli con dei condensatori equivalenti tra drain e massa di valore doppio, in modo che, in corrispondenza della variazione di tensione V_{DD} che si ha sull'uscita, la variazione di carica sugli stessi sia pari a quella sui C_{gd} . Si ottiene così una capacità totale equivalente in uscita al primo inverter pari a

$$C_{out} = 4C_{gd} + 2C_{db} + C_w + 2C_g.$$

Tale capacità C_{out} viene caricata, durante la transizione dal livello basso a quello alto, dal transistor p , che agisce come una resistenza di basso valore e, analogamente, viene scaricato dal transistor n durante il transitorio dal livello alto a quello basso. Si ha quindi, in ambedue i casi, un transitorio di tipo esponenziale. Se i due transistori presentano, nello stato "ON", resistenze pressoché uguali (avendo scelto le W/L nel rapporto inverso delle mobilità), i due transistori sono sostanzialmente simmetrici.

Vediamo come è possibile calcolare la potenza che viene dissipata dall'inverter durante la commutazione, in conseguenza del fenomeno di carica del condensatore di uscita. Nella transizione dal livello alto a quello basso si parte da un condensatore C_{out} carico a V_{DD} , che viene scaricato attraverso il transistor n . Quindi la dissipazione di energia corrisponde all'energia elettrostatica che era immagazzinata nel condensatore:

$$E_{HL} = \frac{1}{2}C_{out}V_{DD}^2.$$

Durante la transizione dal livello basso a quello alto il condensatore C_{out} , inizialmente scarico, viene caricato a V_{DD} . Il caricamento avviene attraverso il transistor p , che agisce come una resistenza di basso valore. L'energia assorbita dall'alimentazione V_{DD} nel transitorio, che immaginiamo duri da $t = 0$ a $t = T$, è data da

$$E_{PS} = \int_0^T V_{DD}i(t)dt = V_{DD}Q = V_{DD}CV_{DD} = CV_{DD}^2.$$

L'energia immagazzinata nel condensatore al termine del transitorio di carica risulterà

$$E_C = \frac{1}{2}C_{out}V_{DD}^2,$$

quindi l'energia dissipata nel transitorio stesso sarà pari alla differenza tra E_{PS} e E_C :

$$E_{LH} = E_{PS} - E_C = \frac{1}{2}C_{out}V_{DD}^2,$$

pertanto l'energia dissipata è la stessa nei due transistori. Se supponiamo che l'inverter venga commutato tra i due stati a una frequenza f la potenza dissipata risulterà $P_{\text{din}} = fC_{\text{out}}V_{DD}^2$. Sottolineiamo il fatto che durante la commutazione dell'inverter una certa quantità di energia viene dissipata anche in conseguenza del fatto che i transistori lavorano per un breve periodo con un prodotto tensione drain-source per corrente di drain non trascurabile. Non calcoliamo questo contributo alla dissipazione perché esso risulta in genere ordini di grandezza minore di quello associato al caricamento della capacità di uscita.

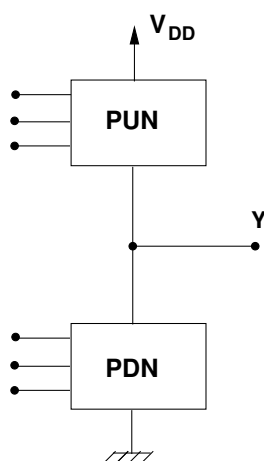
È possibile definire una cifra di merito consistente nel prodotto tra la potenza dissipata e il ritardo di propagazione (delay-power product), che tende a essere una costante per una data famiglia logica: $DP = P_{\text{din}}t_P$. Sostanzialmente il delay-power product ci dà un'indicazione della potenza dissipata durante ogni ciclo di clock, quindi nel caso dei CMOS risulta circa pari a $C_{\text{out}}V_{DD}^2$.

Si definisce fan-out il numero di ingressi che possono essere pilotati dall'uscita di una porta. Il fan-out è limitato, per la tecnologia CMOS, soprattutto dalla necessità di contenere il ritardo di propagazione, che aumenta all'aumentare della capacità in uscita, associata a un numero crescente di ingressi pilotati. Soprattutto in altre famiglie logiche (come la TTL che vedremo più avanti) il fan-out è limitato dal fatto che incrementando il numero di ingressi connessi a una uscita risulta incrementata V_{OL} e risulta diminuita V_{OH} , ottenendo come conseguenza un peggioramento dei margini di rumore.

Con fan-in si indica il numero massimo di ingressi che una porta logica può avere. Anche questo è limitato da considerazioni sulla velocità di commutazione e sui margini di rumore, come vedremo più in dettaglio in seguito.

15.4 Sintesi delle porte CMOS

Una porta logica CMOS consiste in una generalizzazione dell'inverter CMOS che abbiamo studiato finora. Così come l'inverter è costituito da un transistor PMOS e da un transistor NMOS che possono connettere l'uscita, rispettivamente, all'alimentazione o a massa, una porta CMOS è costituita da due reti che possono connettere l'uscita a massa (PDN, pull-down network) oppure alla tensione di alimentazione (PUN, pull-up network) a seconda del valore delle variabili di ingresso.

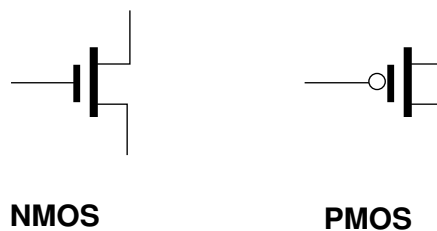


La PDN condurrà per tutte le combinazioni delle variabili di ingresso che richiedono un'uscita a livello basso, mentre la PUN condurrà per tutte quelle combinazioni che richiedono un'uscita al livello alto. Chiaramente, per quelle combinazioni per cui la

PDN conduce la PUN è un circuito aperto e per tutte le combinazioni per cui la PUN conduce la PDN è un circuito aperto.

Poiché la PDN è realizzata con una combinazione di transistori NMOS, questa verrà attivata per particolari scelte dei segnali di ingresso a livello alto. La PUN viene invece realizzata con una combinazione di transistori PMOS e quindi potrà essere attivata da opportune combinazioni degli ingressi a livello basso. Ciascuna delle due reti impiega combinazioni di transistori in serie per implementare funzioni di tipo AND e combinazioni di transistori in parallelo per implementare funzioni di tipo OR.

Prima di vedere alcuni esempi di PUN e di PDN, introduciamo dei simboli alternativi semplificati per i transistori PMOS e NMOS, che vengono di solito utilizzati negli schemi delle porte CMOS. Il transistore PMOS viene indicato con un cerchio sul gate che ricorda la negazione del valore logico.

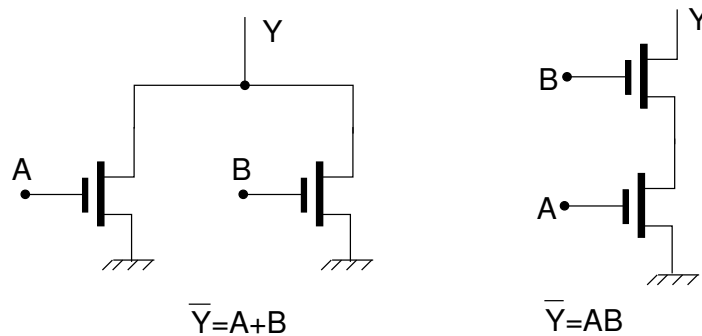


Una PDN costituita da due transistori NMOS in parallelo implementa la funzione logica

$$\bar{Y} = A + B,$$

mentre due transistori in serie forniscono la funzione logica

$$\bar{Y} = AB.$$



Se combiniamo in parallelo un transistore con la serie di altri due possiamo ottenere la funzione logica

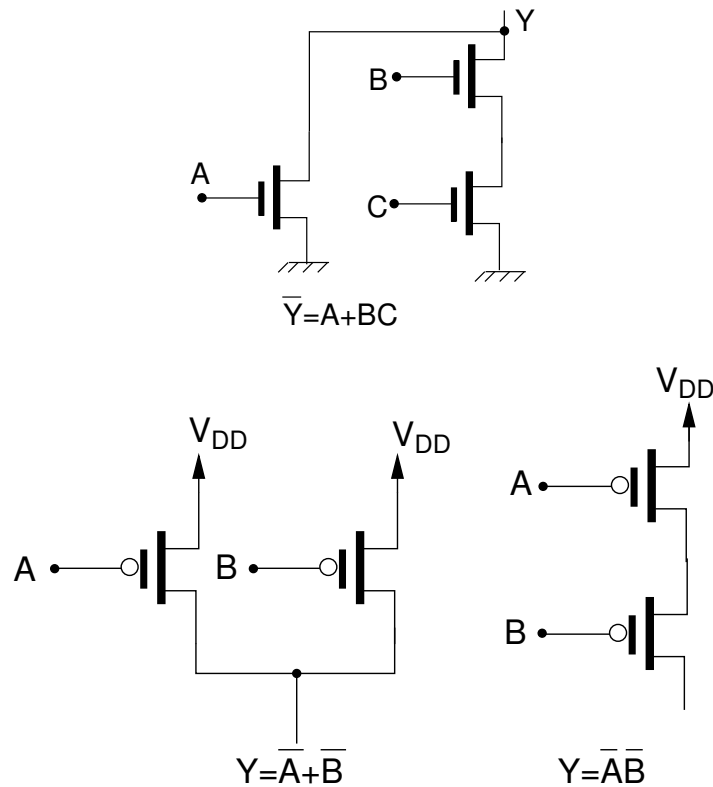
$$\bar{Y} = A + BC.$$

Per quanto riguarda invece le PUN, due transistori in parallelo danno una funzione logica

$$Y = \bar{A} + \bar{B},$$

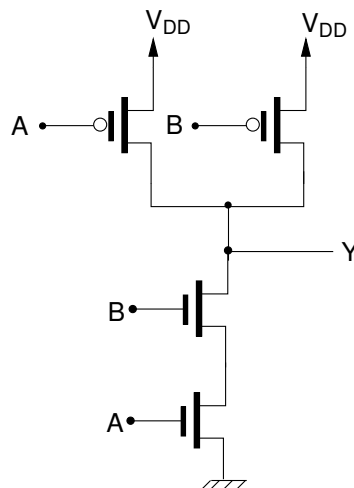
mentre due transistori in serie danno

$$Y = \bar{A} \bar{B}.$$



Combinazioni di più transistori in serie e in parallelo sono possibili anche in questo caso, allo scopo di ottenere funzioni logiche più complesse.

Riassumendo, la PDN può essere progettata a partire da un'espressione della variabile di uscita negata in funzione delle variabili di ingresso non negate. Viceversa, la PUN può essere progettata a partire da un'espressione della variabile di uscita non negata in funzione delle variabili di ingresso negate.



Vediamo come si fa a realizzare una porta NAND a due ingressi: la funzione logica del NAND è

$$Y = \overline{AB}.$$

Usando le leggi di De Morgan cerchiamo di esprimere la funzione di uscita negata in funzione delle variabili di ingresso non negate (per la PDN) e quella non negata in

funzione delle variabili di ingresso negate (per la PUN).

$$Y = \overline{A} + \overline{B},$$

$$\overline{Y} = AB.$$

Quindi la PDN deve essere formata da due NMOS in serie e la PUN da due PMOS in parallelo.

Prendiamo poi in considerazione la porta NOR a due ingressi, che ha una funzione logica

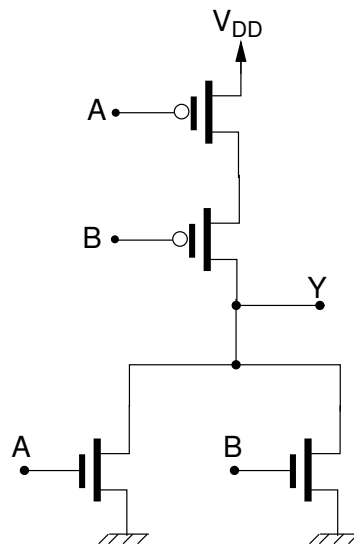
$$Y = \overline{A + B}.$$

Le funzioni per la PDN e per la PUN sono

$$\overline{Y} = A + B,$$

$$Y = \overline{A} \overline{B}.$$

Quindi la PDN deve essere formata da due NMOS in parallelo e la PUN da due PMOS in serie.



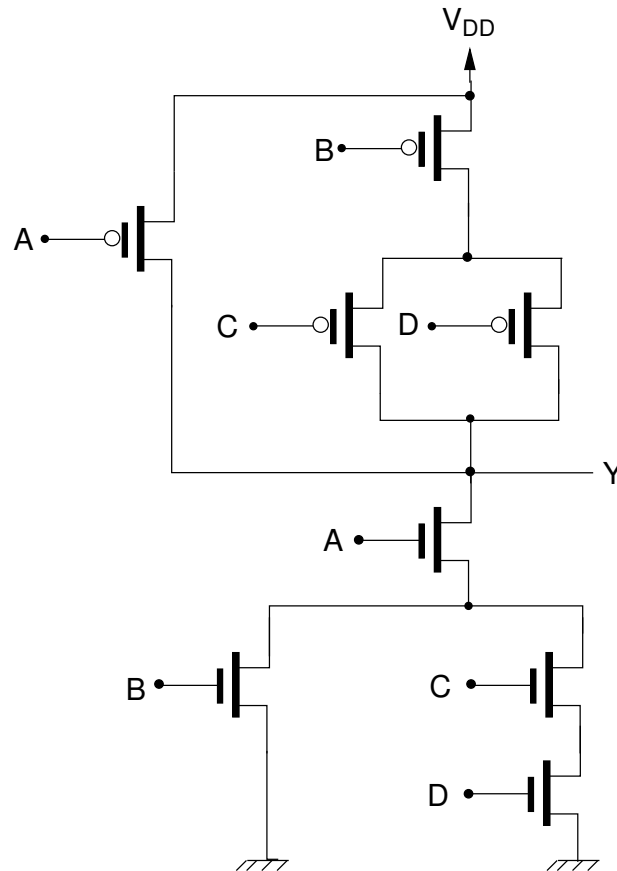
Con la stessa tecnica si possono sintetizzare anche porte più complesse. Per esempio, consideriamo la funzione logica $Y = \overline{A(B + CD)}$. Possiamo svilupparla usando le leggi di De Morgan, in modo da ottenere un'espressione dell'uscita non negata in funzione delle variabili negate

$$\begin{aligned} Y &= \overline{A(B + CD)} \\ &= \overline{A} + \overline{B + CD} \\ &= \overline{A} + \overline{B} \overline{CD} \\ &= \overline{A} + \overline{B}(\overline{C} + \overline{D}), \end{aligned}$$

mentre l'uscita negata in funzione delle variabili non negate si ottiene immediatamente negando ambo i membri

$$\overline{Y} = A(B + CD).$$

Da tali relazioni possono essere immediatamente sintetizzate la PUN e la PDN, ottenendo il circuito riportato di seguito.



Nelle porte viste finora la PDN e la PUN sono reti tra loro duali: se nell'una c'è un ramo in serie, nell'altra ce n'è uno corrispondente in parallelo e viceversa. Quindi l'una può essere ottenuta dall'altra con una procedura piuttosto semplice, senza la necessità di sintetizzarle indipendentemente. Si può partire dalla rete per la quale è più semplice ottenere l'espressione logica.

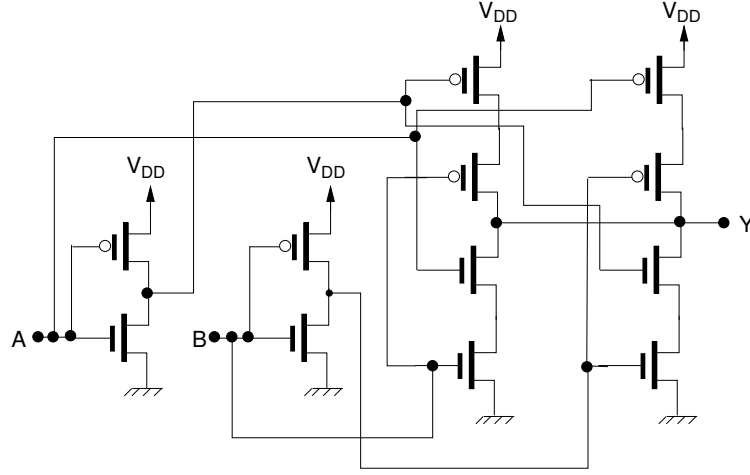
La procedura di sintesi vista finora non è però generale, infatti esistono molti casi in cui non è possibile ottenere, per esempio, un'espressione dell'uscita non negata in funzione delle sole variabili di ingresso negate. Un esempio abbastanza semplice di questo fatto è costituito dall'espressione logica per una porta OR esclusivo (XOR):

$$Y = A\bar{B} + \bar{A}B.$$

Possiamo realizzare la PUN utilizzando come ingressi della stessa non solo le variabili negate, ma anche quelle non negate. Per sintetizzare la PDN potremmo ricavare la rete duale della PUN o utilizzare ancora una volta le leggi di De Morgan:

$$\begin{aligned}\bar{Y} &= \overline{A\bar{B} + \bar{A}B} \\ &= \overline{A\bar{B}} \overline{\bar{A}B} \\ &= (\bar{A} + B)(A + \bar{B}) \\ &= \bar{A}A + \bar{A}\bar{B} + BA + B\bar{B} \\ &= \bar{A}\bar{B} + AB.\end{aligned}$$

Otteniamo quindi la rete seguente, nella quale abbiamo ricavato le variabili complementate tramite due inverter aggiuntivi:



Una volta determinato lo schema della porta che si intende progettare, resta soltanto da stabilire il valore di W/L per i vari transistori. Di solito questa scelta viene fatta in modo da garantire che la porta sia in grado nel caso più critico (worst case) di fornire la stessa corrente al condensatore equivalente in uscita, sia di carica sia di scarica, dell'inverter semplice.

Questo, in principio, garantirebbe un ritardo di propagazione per un gate generico minore o uguale di quello dell'inverter; in realtà la situazione non è esattamente questa, poiché la capacità totale effettivamente presente dipende da un numero di fattori, tra i quali è importante il fan-in, cioè il numero degli ingressi.

Indichiamo con p il rapporto W/L per il transistore PMOS dell'inverter e con n quello del transistore NMOS, sempre dell'inverter. Ricordiamo che la lunghezza L è sempre la stessa per i due transistori. Se, per esempio, scegliamo il valore 2 per n , $p = (\mu_n/\mu_p)n = 5$, se consideriamo il rapporto tra le mobilità pari a 2.5.

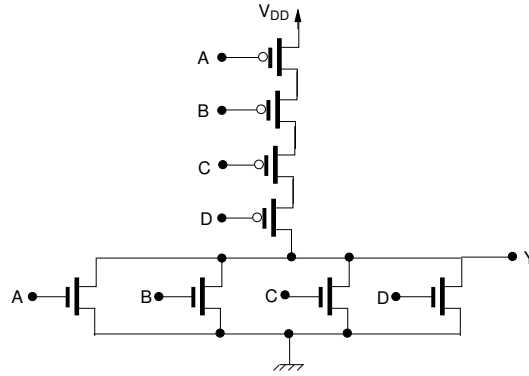
Per svolgere il calcolo del W/L relativo a ciascun transistore è necessario innanzitutto sapere qual è il W/L equivalente di una combinazione di transistori. Se i transistori sono connessi in parallelo, corrisponderanno a un transistore con una larghezza equivalente pari alla somma delle larghezze e quindi con un W/L pari alla somma dei W/L . Se invece i transistori si trovano in serie tra loro, la resistenza equivalente R_{eq} sarà pari alla somma delle varie resistenze di canale R_i , e quindi, poiché la resistenza è inversamente proporzionale a W/L ,

$$R_{eq} = \sum_i R_i = \sum_i \frac{\text{const}}{(W/L)_i} = \frac{\text{const}}{(W/L)_{eq}}.$$

Quindi

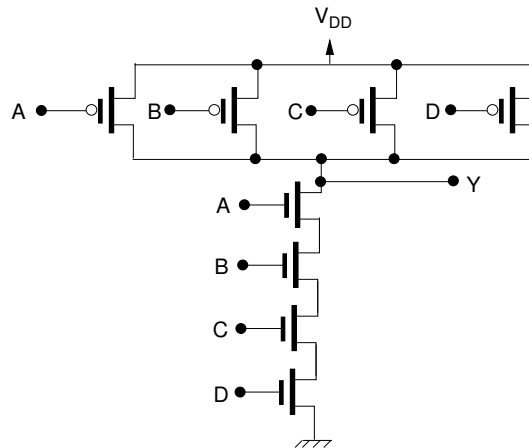
$$(W/L)_{eq} = \left[\sum_i \frac{1}{(W/L)_i} \right]^{-1}.$$

Consideriamo, per esempio, il NOR a quattro ingressi di cui viene di seguito riportato lo schema. Nella PDN il caso peggiore dal punto vista del pilotaggio della capacità di uscita si ha quando è alta una sola delle variabili di ingresso e quindi conduce solo un transistore. Ciascuno dei transistori della PDN dovrà quindi essere di dimensioni almeno pari a quelle del transistore NMOS di un inverter, pertanto



$(W/L)_n = n = 2$. Per quanto riguarda la PUN, vediamo che essa conduce solo se sono attivi tutti e quattro i transistori, i quali si trovano fra loro in serie, quindi $(W/L)_p = 4p = 20$.

Consideriamo poi il caso della porta NAND a quattro ingressi, riportato nello schema che segue. Per quanto riguarda la PUN, lo worst case si ha quando conduce un solo transistore, quindi $(W/L)_p = p = 5$. Per la PDN, invece, si ha passaggio di corrente solo se sono attivi tutti e quattro i MOS, quindi $(W/L)_n = 4n = 8$.



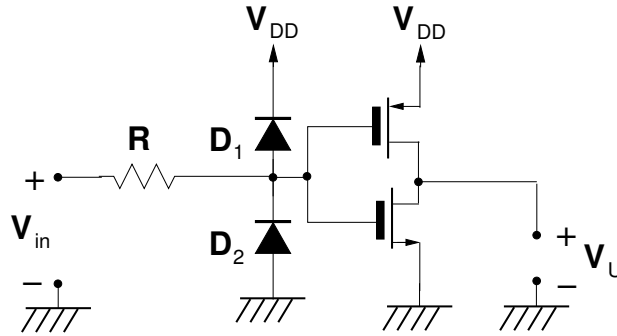
Possiamo ora stimare l'area delle porte NAND e NOR a quattro ingressi, ricordando che l'area del singolo transistore è pari a WL . Per il NOR abbiamo quindi: $A = 4nL^2 + 16pL^2 = 88L^2$. Per il NAND abbiamo invece: $A = 4pL^2 + 16nL^2 = 52L^2$. Ne consegue che i NAND in tecnologia CMOS occupano un'area nettamente inferiore rispetto ai NOR ed è per questo che nella sintesi di reti combinatorie con circuiti CMOS si preferisce utilizzare approcci basati sull'utilizzo di NAND piuttosto che di NOR.

Per concludere menzioniamo nuovamente il fatto che l'ottenimento della stessa capacità di fornire corrente dell'inverter semplice non garantisce che il tempo di ritardo risulti invariato rispetto all'inverter, perché la capacità totale aumenta, in conseguenza della maggiore complessità del circuito e dell'incrementato numero di transistori. L'aumento del fan-in porta infatti a un significativo incremento della capacità ed è per questo motivo che di solito non vengono realizzati NAND con più di quattro ingressi. Si preferisce infatti, nel caso siano necessari più di quattro ingressi, ricorrere a più livelli di logica, che rappresentano il miglior compromesso.

15.5 Protezione dalle scariche elettrostatiche

Data l'elevatissima resistenza e la piccolissima capacità presenti all'ingresso delle porte logiche CMOS, il contatto con oggetti come il corpo umano, che possono trovarsi a potenziali anche molto elevati rispetto a terra (per esempio in conseguenza di fenomeni di carica per sfregamento), sia pure con valori di carica elettrica immagazzinata molto ridotti, è in grado di danneggiare in modo irreversibile i dispositivi CMOS, tramite la perforazione del dielettrico tra gate e canale o tramite uno spostamento della carica intrappolata nell'ossido e una conseguente variazione della tensione di soglia. Infatti l'applicazione di una tensione di qualche decina di migliaia di volt tra gate e substrato dà luogo a un campo elettrico elevatissimo nello strato di ossido, che ha spessore nanometrico, portando quindi al superamento della rigidità dielettrica o alla migrazione di cariche.

È per questo motivo che i componenti CMOS vengono trasportati di solito in apposite buste realizzate in plastica conduttrice, in modo da impedire il formarsi di elevate differenze di potenziale tra terminali diversi. Un'ulteriore precauzione che viene messa in atto consiste nell'aggiunta di diodi di protezione sugli ingressi delle porte che possono venire a contatto con l'esterno e quindi con oggetti a elevati potenziali elettrostatici. Lo schema del circuito di protezione è riportato nella figura che segue.



I diodi D_1 e D_2 hanno la funzione di impedire che la tensione tra gate e massa e tra il gate e la sorgente di alimentazione V_{DD} superi valori di sicurezza. Infatti la tensione tra anodo e catodo di D_1 è data da $V_{D1} = V_G - V_{DD}$, mentre quella tra anodo e catodo di D_2 è data da semplicemente da $V_{D2} = -V_G$. Quindi, assumendo che la tensione ai capi dei diodi in polarizzazione diretta non superi V_γ , otteniamo

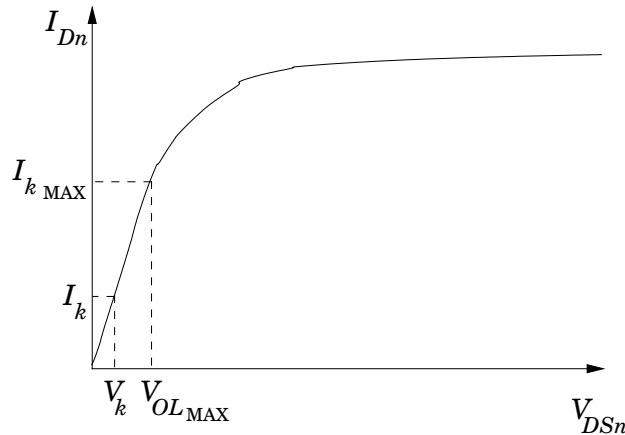
$$\begin{cases} V_G \leq V_{DD} + V_\gamma \\ V_G \geq -V_\gamma. \end{cases}$$

Questo ci garantisce che non si possono creare situazioni di rischio per l'integrità del dielettrico. La resistenza R ha lo scopo di limitare la corrente che attraversa i diodi nel caso si presenti una effettiva situazione di scarica elettrostatica, diluendo nel tempo tale fenomeno.

Questa forma di protezione degli ingressi presenta peraltro alcuni inconvenienti, in particolare legati al fatto che, anche in condizioni statiche, esisterà un assorbimento di corrente, poiché le correnti inverse che scorrono nei due diodi non saranno esattamente uguali. Anche se i diodi fossero identici, le loro caratteristiche difficilmente saranno tali da corrispondere perfettamente al comportamento ideale che prevederebbe una corrente inversa indipendente dalla tensione applicata. Si avranno quindi

correnti di ingresso, corrispondenti alla differenza tra le correnti inverse nei due diodi, dell'ordine dei nanoampere o addirittura microampere, invece dei picoampere tipici delle porte CMOS sprovviste di circuito di protezione.

Tali correnti di ingresso possono portare a significativi inconvenienti, in particolare a una non trascurabile riduzione del fan-out. Consideriamo infatti la situazione di un'uscita a livello basso che pilota una porta dotata di diodi di protezione all'ingresso: una corrente non trascurabile fluirà da tale ingresso nel transistor NMOS di pull-down della porta a monte, determinando uno spostamento del punto di lavoro di quest'ultimo rispetto a quello ideale.



La tensione di uscita passerà dunque dal valore ideale, pari a 0, a un valore V_k tanto più alto quanto maggiore è la corrente I_k che proviene dall'ingresso della porta a valle, dando quindi luogo a un degrado del livello logico basso. È chiaro che se le porte pilotate sono più di una le loro correnti si sommeranno e la tensione di uscita della porta pilotante sarà ulteriormente incrementata, fino a raggiungere valori superiori a $V_{OL\text{MAX}}$ e quindi non più consistenti con la rappresentazione dello stato logico 0.

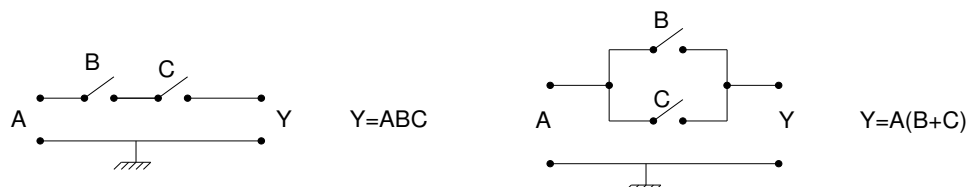
Nel caso di un'uscita a livello logico alto che pilota porte con circuito di protezione si verificherà un analogo inconveniente, a causa delle correnti che in questo caso fluiscono dal transistor PMOS di pull-up della porta a monte verso gli ingressi delle porte a valle, dando pertanto luogo a un degrado del livello logico alto.

È quindi necessario limitare il fan-out in maniera da mantenere i livelli logici in uscita entro gli intervalli previsti dalle specifiche della famiglia logica utilizzata. Questo problema di riduzione del fan-out è sentito in modo particolare nella realizzazione di circuiti logici tramite circuiti integrati a bassa scala di integrazione, per cui quasi tutte le porte vengono ad avere ingressi che necessitano di circuito di protezione perché possono venirsi a trovare in contatto con l'esterno. In circuiti a elevata scala di integrazione il problema è molto ridotto, perché soltanto poche porte sono a diretto contatto con il mondo esterno e abbisognano quindi di diodi di protezione.

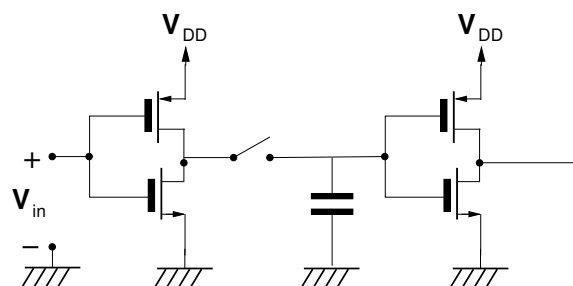
15.6 Logica basata sui pass transistor

L'implementazione vista finora delle porte logiche in tecnologia CMOS garantisce buone prestazioni, ma richiede spesso un numero piuttosto elevato di transistori. Esistono delle formulazioni alternative alla logica CMOS, che spesso vengono utilizzate in combinazione con il tipo di porte che abbiamo esaminato fino a questo punto. Tale logica alternativa è nota sotto il nome di "Pass Transistor Logic" perché sfrutta dei transistori MOS utilizzati come elementi di passo (interruttori) per la realizzazione di una logica a interruttori.

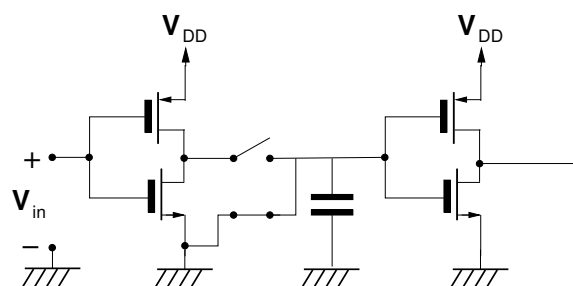
Il concetto di base di una logica a interruttori è mostrato in maniera molto semplice in figura, dove sono rappresentate porte in grado di implementare le funzioni logiche $Y = ABC$ e $Y = (B+C)A$, dove A corrisponde alla variabile logica applicata all'ingresso sotto forma di tensione mentre B e C corrispondono a variabili di controllo degli interruttori, che risultano chiusi quando queste sono al livello alto.



Una prima regola che deve essere sempre rispettata nella realizzazione di questo tipo di logica tramite componenti reali consiste nell'avere sempre ciascun nodo del circuito collegato a un altro nodo con tensione ben definita (0 o V_{DD}). Consideriamo infatti il circuito costituito da due inverter CMOS in cascata connessi tra loro tramite un interruttore, come indicato in figura, dove abbiamo anche rappresentato esplicitamente un condensatore corrispondente alla capacità di ingresso del secondo inverter.

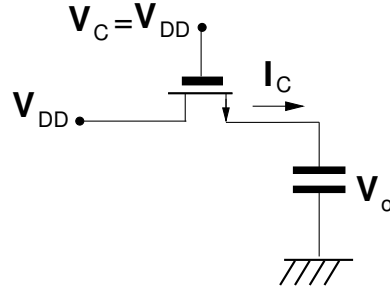


Se l'interruttore è inizialmente chiuso, e l'uscita del primo inverter è allo stato logico alto, il condensatore di ingresso del secondo inverter è carico alla tensione V_{DD} . Se l'interruttore viene aperto, la tensione all'ingresso del secondo inverter rimane per un certo tempo al livello logico alto e solo con un ritardo che può essere anche di alcuni secondi raggiunge il livello logico basso, grazie all'azione delle correnti di perdita. Questo problema può essere risolto utilizzando la regola prima citata e fornendo, quando l'interruttore è aperto, un collegamento diretto tra il nodo di ingresso del secondo inverter e la massa, tramite un secondo interruttore che risulta chiuso quando il primo è aperto e viceversa.



Gli interruttori necessari per l'implementazione di una logica come quella che stiamo descrivendo possono essere realizzati con minimo ingombro e con ridotte capacità parassite utilizzando dei transistori NMOS. Questa soluzione ha il vantaggio

della semplicità, ma presenta anche un comportamento lontano da quello ideale. Consideriamo infatti il caso in cui si voglia caricare, tramite un pass transistor NMOS, un condensatore inizialmente scarico, come nello schema rappresentato in figura.



All'inizio del transitorio di carica avremo $V_{GS} = V_{DD}$ e $V_{DS} = V_{DD}$, quindi il transistor opererà sulla caratteristica per $V_{GS} = V_{DD}$ e in condizione di saturazione. Non appena la tensione sul condensatore inizia a salire, sia la V_{GS} sia la V_{DS} iniziano a diminuire, quindi il punto di lavoro del transistor si sposta su caratteristiche di uscita più basse di quella iniziale e verso tensioni V_{DS} minori. Questo comportamento prosegue finché non si arriva alla condizione per cui $V_{GS} < V_T$, dove V_T è la tensione di soglia del transistor NMOS: a questo punto il transistor si interdice e la tensione sul condensatore non può salire oltre il valore $V_{DD} - V_T$ raggiunto.

Si noti che, nel caso che il circuito considerato sia realizzato in forma integrata, la tensione V_T non corrisponde alla tensione di soglia che si avrebbe se l'elettrodo di bulk fosse collegato a quello di source come nei MOS discreti, bensì a un valore diverso, più grande, a causa del cosiddetto "body effect". Tale effetto è conseguenza del fatto che in un circuito integrato il bulk di tutti i transistori a canale n viene collegato a massa e quello di tutti i transistori a canale p viene connesso a V_{DD} , in modo da garantire che le giunzioni formate tra le diffusioni di drain e source e il bulk siano sempre polarizzate inversamente. In questa condizione la tensione di soglia di un transistor MOS a canale n risulta incrementata rispetto a quella V_{T_0} che si avrebbe con source e bulk connessi tra loro e può essere espressa tramite la relazione

$$V_T = V_{T_0} + \gamma \left[\sqrt{2\phi_f + V_{SB}} - \sqrt{2\phi_f} \right],$$

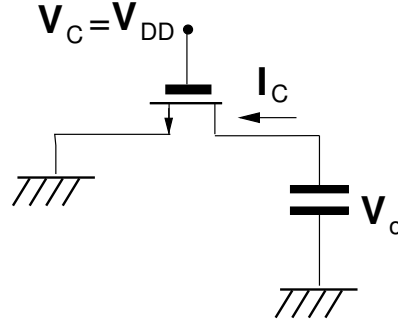
dove $2\phi_f$ è un parametro fisico del valore di circa 0.6 V, V_{SB} è la tensione tra source e bulk e γ è un parametro funzione del drogaggio del bulk N_A , della capacità dell'ossido C_{ox} e della costante dielettrica del semiconduttore ϵ_s tramite

$$\gamma = \frac{\sqrt{2qN_A\epsilon_s}}{C_{ox}}.$$

Un valore tipico per γ è di $0.5 \text{ V}^{0.5}$. Questa relazione, se si considerano i moduli delle varie grandezze e N_D al posto di N_A è valida anche per i transistori a canale p .

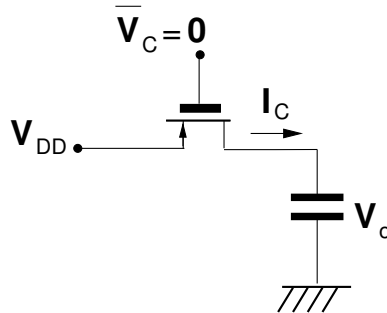
Raggiungiamo quindi la conclusione che un pass transistor realizzato con un transistor NMOS fornisce in uscita un livello logico alto degradato. Vediamo che cosa succede invece per il livello logico basso: supponiamo che il segnale di ingresso sia a livello logico basso, quindi a 0 V e che il condensatore sia inizialmente carico a V_{DD} . Lo schema è riportato nella figura: si noti che questa volta il source si trova a sinistra, invece che a destra; questo perché il transistor è intrinsecamente simmetrico

e l'attribuzione ai due elettrodi della funzione di drain o di source dipende esclusivamente dal segno della tensione applicata (supponiamo chiaramente di trovarci nel caso di un circuito integrato, in cui il bulk non è connesso a uno dei due elettrodi, ma a massa).



In questo caso la tensione V_{GS} è costante e pari a V_{DD} , quindi il transistor opera sempre sulla caratteristica di uscita corrispondente a $V_{GS} = V_{DD}$: inizialmente $V_{DS} = V_{DD}$, per cui il transistor è in saturazione e il condensatore viene scaricato a corrente costante, poi, quando V_{DS} scende al di sotto di $V_{DD} - V_{T0}$, il transistor passa in zona triodo, ma continua a scaricare il condensatore fino a tensione nulla. Dunque il livello logico zero prodotto in uscita da un pass gate realizzato con un transistor NMOS non è degradato.

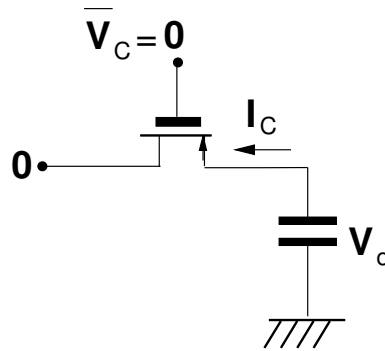
Prendiamo ora in considerazione il caso di un pass gate realizzato con un transistor PMOS. Tale pass gate conduce quando la tensione di gate, di solito indicata con \bar{V}_C , è a livello basso. Esaminiamo il processo di carica del condensatore (anche in questo caso la posizione del source è determinata dal verso della tensione V_{DS}).



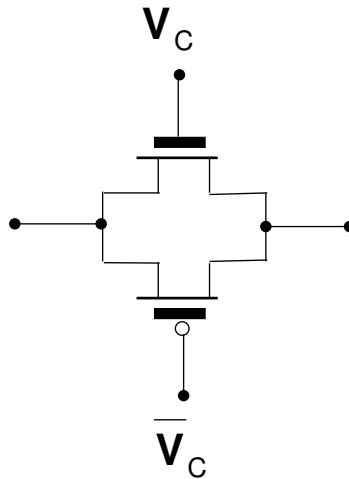
La tensione V_{GS} rimane costante e pari a $-V_{DD}$ durante tutto il transitorio di carica del condensatore, quindi il punto di lavoro del transistor si sposta, sempre sulla stessa caratteristica di uscita, passando dalla saturazione alla zona triodo quando la tensione V_o supera il modulo della tensione di soglia V_T . La carica del condensatore continua comunque fino al raggiungimento della tensione V_{DD} , senza degrado del livello logico alto.

Nel caso invece in cui si utilizzi un pass gate PMOS per scaricare un condensatore inizialmente carico a V_{DD} si presentano problemi analoghi a quelli visti per la carica con il transistor NMOS. Inizialmente $V_{GS} = -V_{DD}$, ma non appena il condensatore inizia a scaricarsi il modulo della V_{GS} inizia a decrescere, finché non viene raggiunto il valore di $|V_T|$, che risulta maggiore di $|V_{T0}|$ a causa del body effect, e il transistor si interdice, lasciando il condensatore carico a $-V_T$.

Possiamo quindi concludere osservando che un pass gate realizzato con un transistor NMOS consente di ottenere un “buon 0” ma un “cattivo 1”, mentre con un transistor



PMOS si ottiene il risultato opposto. Risulta quindi naturale pensare a un pass gate realizzato combinando un transistor PMOS e un transistor NMOS, in modo da sfruttare gli aspetti positivi di tutti e due. Tale pass gate CMOS deve essere pilotato con un segnale di comando attivo al livello alto, che viene connesso al gate dell'NMOS e dal segnale complementato, che controlla il gate del PMOS.

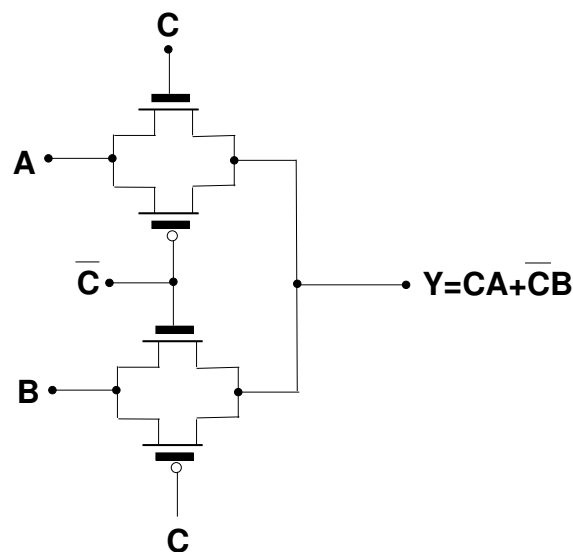


Il pass gate CMOS può essere utilizzato anche come interruttore analogico, con la variante di impiegare una tensione di comando pari a V_{DD} per il transistor NMOS e a $-V_{DD}$ per il transistor PMOS. In questo modo possono transitare tutti i segnali compresi nell'intervallo $[-V_{DD}, V_{DD}]$. Se invece applichiamo una tensione $-V_{DD}$ al gate dell'NMOS e V_{DD} a quello del PMOS, il gate impedisce il passaggio di qualunque segnale compreso nell'intervallo $[-V_{DD}, V_{DD}]$.

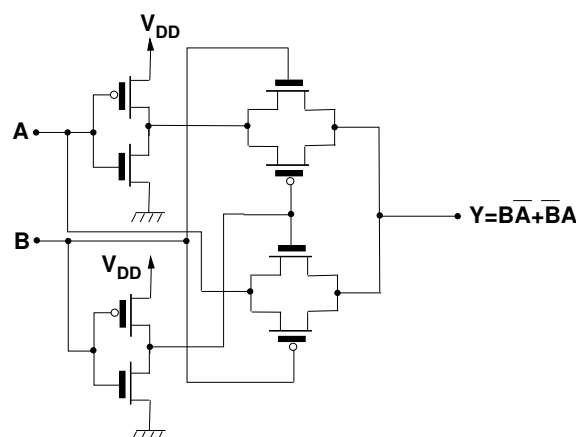
Il pass gate CMOS (nella versione per usi digitali) può essere impiegato come blocco costitutivo per la realizzazione di porte logiche, che in alcuni casi possono risultare più semplici delle corrispondenti porte ottenute con la tecnica vista per la sintesi standard CMOS. Consideriamo per esempio l'implementazione di un multiplexer 2-a-1, vale a dire di un circuito logico che connette alternativamente l'ingresso A o quello B all'uscita Y , a seconda che la variabile di controllo C sia, rispettivamente allo stato alto o a quello basso:

$$Y = CA + \bar{C}B.$$

La realizzazione di un tale circuito con pass gate CMOS è estremamente intuitiva: basta utilizzarne due con l'uscita a comune e controllati da segnali complementati.



Con l'aggiunta di un ulteriore inverter (oltre a quello per la complementazione della variabile di controllo) è possibile ottenere un OR esclusivo, ponendo sui due ingressi A e \bar{A} e impiegando B come variabile di controllo, come indicato in figura. Notiamo che in questo caso è stato possibile realizzare un OR esclusivo con soli 8 transistori invece dei 12 necessari per la realizzazione della versione presentata nel paragrafo 15.4.

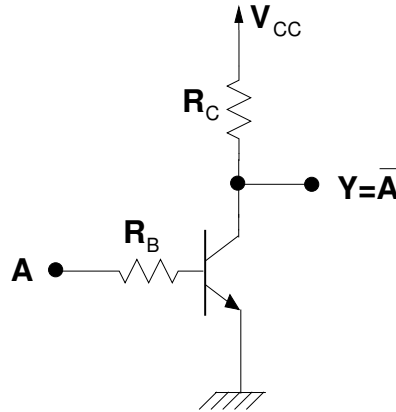


15.7 Famiglie logiche bipolari RTL e DTL

Finora abbiamo preso in esame il comportamento e la struttura dei circuiti logici realizzati con transistori MOS, che rappresentano la stragrande maggioranza di quelli che risultano attualmente impiegati nelle applicazioni. Tuttavia i primi circuiti logici furono realizzati per mezzo di transistori bipolari e sistemi basati su famiglie logiche bipolari sono ancora in uso, data la grandissima diffusione che questi ebbero fino ai primi anni '80.

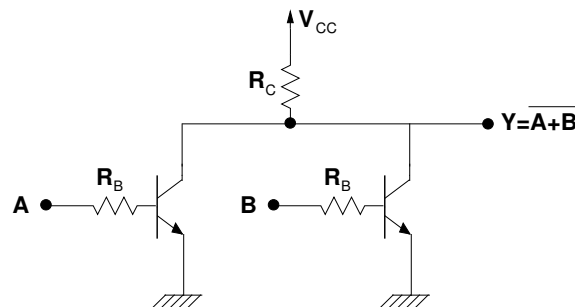
La maggior parte delle famiglie logiche basate su transistori BJT utilizza tali dispositivi come interruttori, quindi operanti in interdizione (vale a dire con i tre terminali disconnessi tra loro) o in saturazione (con tensione $V_{CE} = 0.2 \text{ V}$ e $V_{BE} = 0.7 - 0.8 \text{ V}$), anche se esistono logiche “non saturate” in cui i transistori vengono sempre fatti funzionare in zona attiva diretta.

La più semplice porta logica realizzabile con un BJT è l'inverter rappresentato di seguito, consistente in un BJT e due resistenze.



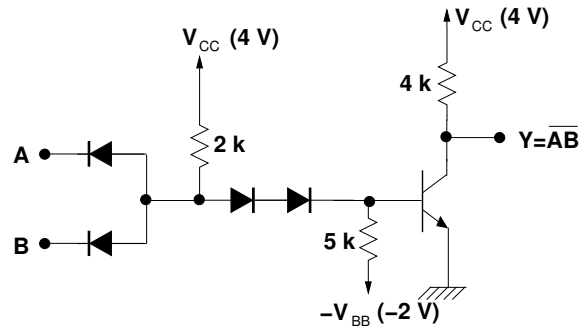
Se la tensione applicata all'ingresso è nulla o, comunque, inferiore a V_γ , il transistor è interdetto e il terminale di uscita si trova praticamente alla tensione di alimentazione, dato che non vi è caduta sulla resistenza di collettore. Se invece viene posta in ingresso una tensione corrispondente al livello logico alto, il transistor, purché R_B venga scelta di valore opportuno, si trova in saturazione e la tensione di uscita corrisponde a circa 0.2 V, quindi al livello logico basso. Nell'analisi dei circuiti logici a BJT verificheremo la condizione di saturazione andando a calcolare la corrente di collettore nell'ipotesi che $V_{CE} = V_{CE_{sat}}$: se tale corrente risulta inferiore a $I_B h_{FE}$, la condizione di saturazione risulta verificata (di solito si fa l'ipotesi semplificativa che il valore di h_{FE} sia costante e indipendente dalla corrente di collettore). Questa porta è molto semplice e corrisponde alle effettive realizzazioni in forma discreta dei primi circuiti logici a BJT, ma presenta evidenti svantaggi e limitazioni, come, per esempio, il fatto che è presente una significativa dissipazione di energia in condizioni statiche quando in uscita abbiamo il livello logico basso, a causa della corrente che scorre in R_C .

Più inverter elementari del tipo appena visto possono essere combinati secondo lo schema di seguito rappresentato, in modo da ottenere una porta NOR, che rappresenta l'elemento fondamentale della famiglia logica RTL (Resistor Transistor Logic).



Il fan-in può essere accresciuto incrementando il numero dei transistori connessi con i collettori a comune. Notiamo che la resistenza sui collettori deve fornire la corrente che pilota gli ingressi delle porte a valle, quindi un incremento nel fan-out porta a una riduzione del livello logico alto in uscita (V_{OH}) e risulta perciò chiaro che il fan-out massimo sarà abbastanza limitato, dato che la corrente di base necessaria per mandare un transistor BJT in saturazione non è trascurabile.

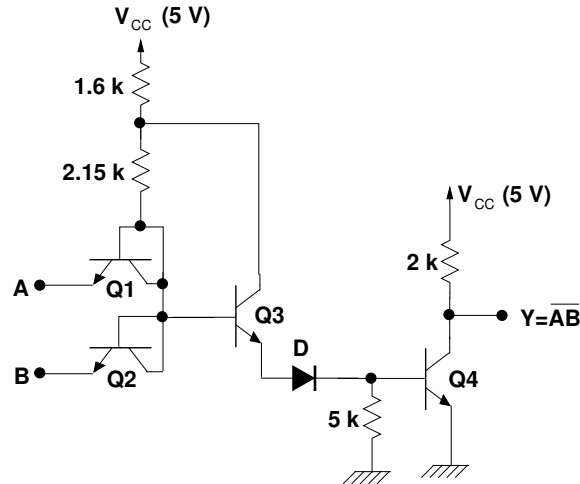
Un primo progresso rispetto alla logica RTL è rappresentato dalla logica DTL (Diode Transistor Logic) a componenti discreti, la cui porta di base, un NAND, è rappresentata nella figura seguente. Supponiamo che almeno uno dei due ingressi sia connesso a massa (livello logico basso): in tal caso il corrispondente diodo di ingresso si troverà in conduzione e la tensione sul nodo a comune tra i diodi di ingresso sarà pari a $V_\gamma = 0.7$ V.



I due diodi connessi alla base del transistore risulteranno pertanto in conduzione e quindi la base del transistore si troverà a $-V_\gamma$, determinandone l'interdizione. Dunque, consistentemente con la funzione logica che desideriamo implementare, se uno degli ingressi è a livello logico basso, l'uscita si trova a livello logico alto. Supponiamo ora di porre uno degli ingressi a livello logico alto e di innalzare gradualmente la tensione sull'altro ingresso: a un certo punto la tensione sulla base del transistore sarà tale da portarlo in conduzione. Quando tale tensione raggiungerà 0.7 V, il diodo di ingresso si interdirà, dato che la tensione sul punto a comune tra i due diodi di ingresso non potrà salire oltre $3V_\gamma = 2.1$ V. A questo punto tutta la corrente che attraversa la resistenza da 2 kΩ verrà deviata verso i diodi che collegano la base del transistore e quest'ultimo raggiungerà così la condizione di saturazione, presentando in uscita una tensione di 0.2 V, corrispondente allo 0 logico. Il motivo principale per cui si è inclusa una tensione di alimentazione negativa (che comporta non pochi problemi pratici) consiste nella necessità di estrarre la maggior corrente possibile dalla base del transistore quando lo si vuol far passare dalla condizione di saturazione a quella di interdizione. In saturazione la regione di base contiene un'elevata concentrazione di portatori minoritari, data la polarizzazione diretta di ambedue le giunzioni, e per uscire da questa condizione è necessario rimuovere una carica significativa: la transizione verso la condizione di interdizione sarà tanto più veloce quanto maggiore risulterà la corrente uscente dalla base durante il transitorio.

La porta DTL che abbiamo esaminato veniva anch'essa realizzata in forma discreta. Quando si passò alla realizzazione in forma integrata delle porte DTL furono apportate alcune significative modifiche, indicate nello schema che segue.

Osserviamo che i diodi di ingresso sono stati ottenuti, come si è soliti fare nei circuiti integrati, con transistori aventi base e collettore tra loro cortocircuitati. Osserviamo inoltre che il primo diodo sul ramo verso la base del transistore Q_4 è stato sostituito dalla giunzione base-emettitore del transistore Q_3 , allo scopo di ridurre la corrente di ingresso e, pertanto, di incrementare il fan-out massimo raggiungibile. Infatti in questo modo la corrente che scorre in un ingresso a livello basso a è limitata dalla serie delle due resistenze da 1.6 kΩ e 2.15 kΩ, che corrisponde quasi al doppio della precedente resistenza da 2 kΩ, mentre la corrente inviata sul diodo D per mandare in saturazione Q_4 quando tutti gli ingressi sono al livello alto è pari, se non superiore (vista la presenza di una resistenza da 1.6 kΩ e la ridotta caduta di tensione tra base e



collettore di Q_3 quando questo conduce; si noti che Q_3 non può andare in saturazione perché, in conseguenza della struttura del circuito, la tensione sul collettore è sempre più alta di quella sulla base), a quella ottenibile nel circuito DTL discreto. L'altra differenza che subito si nota è l'assenza di un'alimentazione negativa, che consente di semplificare notevolmente i circuiti di alimentazione, ma porta a un allungamento del tempo necessario a Q_4 per uscire dalla condizione di saturazione. Oltre alla non molto elevata velocità di recupero dalla saturazione, la logica DTL è affetta anche da un altro importante problema: il comportamento dello stadio di uscita su carico capacitivo è asimmetrico tra la transizione alto-basso e quella basso-alto. Infatti, un condensatore in uscita può essere scaricato molto rapidamente dalla forte corrente che attraversa il transistor Q_4 in condizione di saturazione, mentre, invece, il processo di carica è significativamente più lento, poiché deve avvenire, con Q_4 interdetto, attraverso la resistenza da $2\text{ k}\Omega$.

L'attività di progettazione volta a risolvere tali problemi ha portato allo sviluppo della famiglia logica TTL, che sarà l'oggetto del prossimo paragrafo.

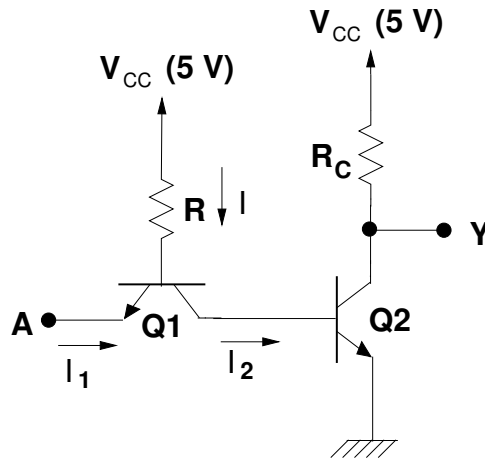
15.8 Famiglia logica TTL

La famiglia logica TTL (Transistor Transistor Logic) deriva il proprio nome dal fatto che i diodi di ingresso vengono sostituiti da transistori, come mostrato nella figura seguente, nella quale è rappresentato un tipico stadio di ingresso. Per l'analisi del funzionamento consideriamo dapprima presente in ingresso un valore di tensione V_{CC} , corrispondente a un 1 logico: in tal caso la giunzione base-emettitore di Q_1 è polarizzata inversamente, mentre quella base-collettore conduce e fornisce la corrente necessaria per mandare in saturazione il transistor Q_2 .

Il transistor Q_1 si trova quindi a operare in zona attiva inversa. La corrente I può essere calcolata immediatamente osservando che la tensione sulla base di Q_1 è pari alla somma delle cadute di tensione sulla giunzione base-collettore di Q_1 e sulla giunzione base-emettitore di Q_2 , quindi $2V_\gamma$:

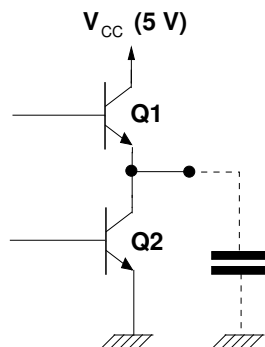
$$I = \frac{V_{CC} - 2V_\gamma}{R}.$$

Il transistor Q_1 è realizzato in maniera tale da avere un β_R estremamente basso (dell'ordine di 0.02), in modo che la corrente di ingresso, I_1 , pari a $\beta_R I$, sia molto piccola. La corrente I_2 è invece data da $I_2 = (\beta_R + 1)I$ (ricordiamo che il transistor



opera in zona attiva inversa) ed è sufficiente a portare in saturazione Q_2 . Se la tensione di ingresso viene poi portata al valore logico basso, corrispondente a una tensione di 0.2 V, la corrente I verrà deviata verso l'ingresso, polarizzando direttamente la giunzione base-emettitore del Q_1 e facendo quindi scendere la sua tensione di base a 0.9 V (gli 0.2 V di ingresso più la V_{BE}). Dato che il transistor Q_2 si trovava in saturazione, la sua tensione di base rimarrà circa a 0.7 V finché non sarà stato rimosso l'eccesso di cariche minoritarie. Quindi Q_1 si troverà a condurre, con una corrente di collettore $\beta_F I$, e porterà rapidamente Q_2 fuori dalla saturazione. Una volta che il transistor Q_2 si sarà interdetto, il transistor Q_1 andrà verso la saturazione (risultando anche la giunzione base-collettore in piena polarizzazione diretta), ma con una corrente di collettore trascurabile, per cui la V_{CE_1} sarà dell'ordine di 0.1 V e, conseguentemente, la tensione di collettore di Q_1 risulterà di 0.3 V.

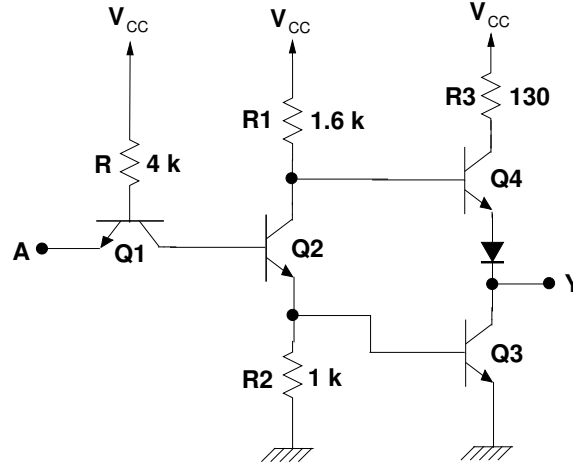
In questo modo viene quindi risolto uno dei problemi della logica DTL, quello della uscita lenta dalla condizione di saturazione. Vediamo ora come viene invece risolto il problema dell'asimmetria nel comportamento dello stadio di uscita. Abbiamo visto che lo stadio di uscita della logica DTL consente di ottenere una rapida scarica della capacità vista in uscita, ma non consente invece di caricarlo altrettanto rapidamente. Se invece avessimo in uscita uno stadio a collettore comune, questo sarebbe in grado di caricare rapidamente il condensatore, ma la scarica sarebbe invece più lenta perché dovrebbe avvenire attraverso la resistenza di emettitore. La soluzione adottata nella logica TTL consiste nel combinare i due circuiti, in modo da ottenere un comportamento veloce sia nel transitorio di scarica sia in quello di carica. Il risultato è rappresentato nella figura seguente e di solito viene indicato come circuito "totem pole", data la struttura con elementi sovrapposti (come nei totem).



I due transistori devono essere pilotati con segnali tra loro in opposizione di fase, per-

ché quando conduce uno deve essere interdetto l'altro: se conduce Q_1 , il condensatore viene caricato attraverso tale transistor e l'uscita va al livello alto; l'opposto accade se conduce il transistor Q_2 , attraverso il quale il condensatore viene scaricato.

Possiamo prendere a questo punto in esame un gate completo TTL, ancora con un solo ingresso per maggiore semplicità. Notiamo la presenza di un ulteriore transistor che svolge la funzione di “phase splitter”, fornendo due segnali in opposizione di fase per pilotare i due transistori dello stadio totem pole di uscita.



Notiamo anche la presenza di un diodo aggiuntivo nello stadio totem-pole e di una resistenza da 130 Ω , la cui funzione sarà discussa in seguito. Procediamo ora all'analisi del circuito quando in ingresso è presente una tensione corrispondente a un livello logico alto, supponiamo pari a 5 V. In tal caso, come abbiamo già visto, il transistor Q_1 opera in zona attiva inversa, con una tensione di base pari a $3V_\gamma = 2.1$ V (dato che dobbiamo includere, oltre al contributo della giunzione base-collettore di Q_1 e di quella base-emettitore di Q_3 , anche il contributo della giunzione base-emettitore del transistor Q_2). La corrente di base di Q_1 sarà dunque pari a

$$I = \frac{V_{CC} - V_{B1}}{R} = 0.725 \text{ mA}$$

e quindi la corrente assorbita dall'ingresso sarà data da

$$I_{IH} = \beta_R I = 14.5 \mu\text{A},$$

dove si è assunto $\beta_R = 0.02$. la corrente che esce dal collettore di Q_1 è invece data da $I_{C1} = (\beta_R + 1)I = 0.7395 \text{ mA}$, sufficiente a portare in saturazione Q_2 . La tensione sull'emettitore di Q_2 è pari a V_γ , data la condizione di polarizzazione diretta di Q_3 , per cui la corrente che scorre in R_2 risulta di 0.7 mA. Poiché Q_2 è in saturazione, la sua tensione di collettore è pari a quella di emettitore più 0.2 V, dunque $V_{C2} = 0.9$ V. La corrente I_{R1} che scorre in R_1 risulterà

$$I_{R1} = \frac{V_{CC} - V_{C2}}{R_1} = 2.56 \text{ mA}.$$

Poiché anche il transistor Q_3 si trova in saturazione, la tensione sul suo collettore, corrispondente a quella di uscita, sarà pari a 0.2 V. Quindi il transistor Q_4 non potrà condurre, dato che la differenza di potenziale tra la sua base e il collettore di

Q_3 è di soli 0.7 V, contro gli 1.4 V che sarebbero necessari per mandare in conduzione la giunzione V_{BE} del transistor Q_4 e il diodo. Pertanto, sia Q_4 sia il diodo sono interdetti e la corrente di base di Q_4 è praticamente nulla. Quindi la corrente di collettore di Q_2 è pari a I_{R1} . La corrente di emettitore di Q_2 si ottiene sommando la I_{B2} a I_{R1} , ottenendo 3.3 mA. Sottraendo la corrente che scorre in R_2 otteniamo la corrente di base del Q_3 , pari a 2.6 mA, sufficiente a mandare in saturazione il transistor Q_3 , come avevamo supposto.

Se non avessimo inserito il diodo in serie all'emettitore di Q_4 , tale transistor si sarebbe trovato sul punto di ingresso in conduzione, con la possibilità di determinare una forte corrente nella serie tra Q_3 e Q_4 .

Esaminiamo ora la situazione quando il terminale di ingresso è posto alla tensione di 0.2 V, che corrisponde allo zero logico. In tal caso la giunzione base-emettitore del transistor Q_1 è polarizzata direttamente, con una tensione sulla base pari a $V_{E1} + V_\gamma = 0.9$ V. La corrente I su R in questo caso risulterà

$$I = \frac{V_{CC} - V_{B1}}{R} = 1.025 \text{ mA}.$$

Una tensione di 0.9 V non è sufficiente a portare in conduzione la serie della giunzione base-collettore di Q_1 e di quella base-emettitore di Q_2 , quindi Q_2 è interdetto, mentre Q_1 si trova in saturazione, con una corrente di collettore praticamente trascurabile e quindi una caduta di tensione tra collettore ed emettitore di circa 0.1 V. Pertanto la base di Q_2 è a una tensione di 0.3 V. Tutta la corrente I va a costituire la corrente I_{IL} uscente dall'ingresso, che risulta quindi pari a 1.025 mA. Essendo Q_2 interdetto, la tensione ai capi di R_2 è nulla e anche Q_3 è interdetto. Il transistor Q_4 sarà in conduzione (dato che la base è collegata direttamente all'alimentazione V_{CC} tramite la resistenza R_1). A seconda del valore della corrente assorbita dal carico, Q_4 si troverà in zona attiva diretta o in saturazione (in zona attiva diretta la corrente di base di Q_4 sarà legata a quella di emettitore dalla relazione $I_{B4} = I_{E4}/(\beta_F + 1)$). La tensione di uscita dipenderà anch'essa dal carico e sarà al più (in caso di corrente nulla in uscita) pari a circa $V_{CC} - 2V_\gamma = 3.6$ V, assumendo nulla la caduta su R_1 e considerando quelle sulla giunzione base-emettitore di Q_4 e sul diodo. A questo punto possiamo anche comprendere la funzione della resistenza da 130 Ω posta in serie al collettore di Q_4 : serve per limitare la corrente di uscita in caso di cortocircuito accidentale del terminale di uscita con la massa e, soprattutto, a evitare eccessivi assorbimenti di corrente nel caso di passaggio dallo stato basso in uscita a quello alto. Infatti il transistor Q_4 entra in conduzione più rapidamente di quanto il transistor Q_3 non esca dalla saturazione, per cui, in assenza della resistenza da 130 Ω potrebbero aversi dei forti picchi di assorbimento durante le transizioni dell'uscita dal livello logico basso a quello alto, le quali potrebbero ripercuotersi, tramite le linee di alimentazione, sulle altre porte logiche del circuito, generando così un fenomeno di crosstalk. Nonostante la presenza della R_3 è sempre possibile avere dei fenomeni di crosstalk in conseguenza delle variazioni di assorbimento che si verificano in corrispondenza delle varie transizioni, quindi si ricorre sempre, come misura precauzionale, all'inserzione in parallelo ai terminali di alimentazione di ciascun integrato, di un condensatore, in genere da 0.1 μF , il quale può essere considerato come un elemento che abbassa l'impedenza vista tra i terminali di alimentazione alle alte frequenze o che fornisce le componenti impulsive della corrente richiesta dalle porte TTL.

Dall'analisi svolta notiamo che la corrente I_{IL} erogata dall'ingresso quando questo è posto al livello logico basso (1.025 mA) risulta molto maggiore della corrente I_{IH} assorbita dall'ingresso quando questo si trova al livello logico alto. Pertanto il fan-out delle porte TTL sarà limitato dalla I_{IL} .

La prima famiglia logica TTL è la serie 74, così denominata perché gli integrati facenti parte di tale famiglia vengono indicati con una sigla del tipo 74XX o 74YYY, dove XX e YYY rappresentano numeri, rispettivamente, di due o tre cifre. Per esempio, l'integrato capostipite della serie 74, il 7400, contiene quattro porte NAND a due ingressi. Per la serie 74 abbiamo i seguenti valori, per quanto riguarda le tensioni corrispondenti agli stati logici in ingresso e in uscita:

$$V_{OL_{MAX}} = 0.4 \text{ V}$$

$$V_{OH_{MIN}} = 2.4 \text{ V}$$

$$V_{IL_{MAX}} = 0.8 \text{ V}$$

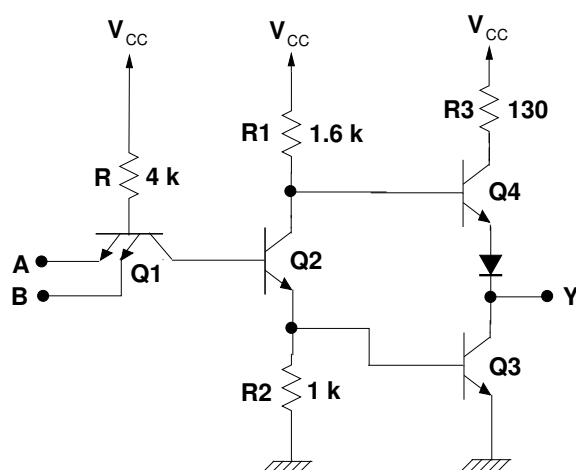
$$V_{IH_{MIN}} = 2 \text{ V}.$$

Possiamo quindi calcolare i margini di rumore, che risultano

$$NM_H = V_{OH_{MIN}} - V_{IH_{MIN}} = 0.4 \text{ V}$$

$$NM_L = V_{IL_{MAX}} - V_{OL_{MAX}} = 0.4 \text{ V}.$$

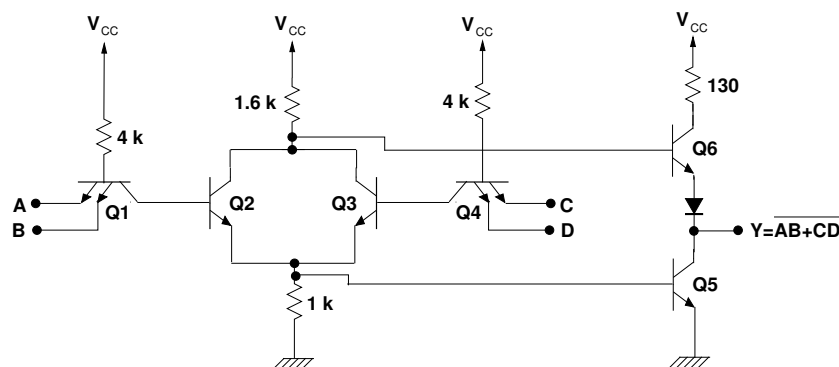
Vediamo ora come viene realizzata una porta NAND, la porta base della famiglia logica TTL. Sarebbe sufficiente aggiungere allo schema già visto un altro transistor di ingresso, con la base e il collettore collegati ai corrispondenti terminali di Q_1 e l'emettitore connesso al secondo ingresso. Considerato che i due transistor avrebbero comunque due elettrodi in comune, si ricorre a un unico componente di tipo particolare, un transistor con due emettitori distinti, ottenendo lo schema di seguito riportato.



È chiaramente possibile realizzare una porta NAND con un numero maggiore di ingressi, semplicemente aggiungendo ulteriori emettitori.

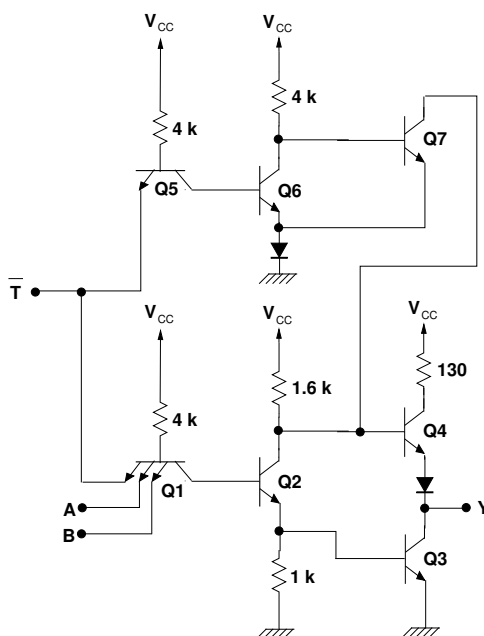
L'implementazione di porte logiche più complesse può essere fatta combinando porte elementari o, come di solito avviene se la realizzazione è effettuata a livello di singolo circuito integrato, tramite una combinazione semplificata di elementi base della tecnologia TTL. Per comprendere come questo sia possibile, vediamo un esempio di un AND-OR con uscita negata che implementa la funzione logica

$$Y = \overline{AB + CD}.$$



In questo caso la funzione logica AND viene svolta dai transistori d'ingresso a due emettitori Q_1 e Q_4 , mentre quella OR è ottenuta semplicemente ponendo in parallelo i due phase splitter. In questo modo abbiamo utilizzato un numero di componenti molto inferiore a quello che sarebbe stato necessario se avessimo combinato tra loro porte elementari TTL complete.

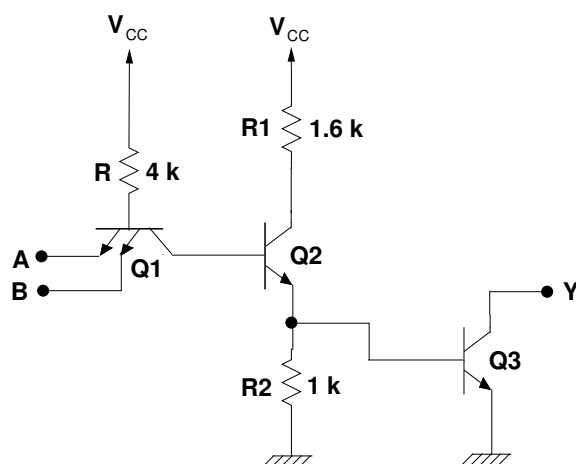
Un'altra importante variante delle porte TTL è rappresentata dall'uscita "tristate", che può presentare, oltre ai livelli logici 1 e 0, un terzo stato ad alta impedenza. Tale possibilità è molto utile, anzi essenziale, se si vogliono realizzare bus di comunicazione dati che sono condivisi da parti diverse di un circuito logico, ciascuna delle quali utilizza una porta per inviare dati. Di volta in volta una sola porta viene abilitata a "parlare" sul bus, mentre le uscite delle altre vengono portate in una condizione di alta impedenza, per evitare che interferiscano. La realizzazione di una porta tristate richiede l'aggiunta di alcuni componenti allo schema TTL standard, come indicato nella figura che segue.



L'ingresso che attiva la funzionalità tristate è indicato con \overline{T} essendo attivo al livello logico basso. Infatti se tale ingresso è a livello logico alto, nella parte inferiore del circuito non produce nessun effetto, rappresentando un ingresso allo stato alto di una porta NAND, e nella parte superiore del circuito determina il funzionamento in zona attiva inversa del transistor Q_5 , la conseguente saturazione del Q_6 e quindi

l'interdizione del Q_7 , dato che la tensione base-emettitore di quest'ultimo transistor risulta pari proprio alla V_{CE} del Q_6 . Essendo Q_7 interdetto, la parte superiore del circuito non ha alcun effetto sull'uscita. Se invece l'ingresso \overline{T} viene posto a livello basso, il transistor Q_1 si trova in saturazione, Q_2 è interdetto e lo stesso accade a Q_3 . Anche Q_5 risulta in saturazione, Q_6 interdetto e di conseguenza la corrente che scorre nella resistenza da $4\text{ k}\Omega$ sul collettore di Q_6 finisce tutta nella base di Q_7 portandolo in saturazione e quindi forzando la base del Q_4 a una tensione di 0.9 , insufficiente per far condurre il Q_4 stesso. Quindi, se $\overline{T} = 0$, ambedue i transistori di uscita della porta sono interdetti, dando luogo alla condizione di alta impedenza. Si noti che il diodo sull'emettitore di Q_6 svolge una funzione importante, poiché rende i livelli di commutazione dell'ingresso tristate uguali a quelli degli altri ingressi.

Porte logiche aventi circuiti di uscita di tipo totem-pole non consentono la connessione in parallelo delle uscite, dato che, non appena queste cercassero di imporre valori logici diversi, si avrebbe una tensione di uscita indeterminata e un forte passaggio di corrente, che porterebbe rapidamente alla distruzione dei dispositivi. Nel caso sia necessario avere la possibilità di connettere le uscite in parallelo, allo scopo, per esempio, di realizzare un AND cablato, si devono utilizzare porte TTL particolari, nelle quali la parte superiore del totem-pole di uscita è assente e il terminale di uscita è rappresentato direttamente dal collettore del transistor Q_3 .



Per un corretto funzionamento di una porta open collector è necessario porre sempre una resistenza di pull-up tra l'uscita e l'alimentazione positiva. Evidentemente, con la soluzione open collector si perdono i vantaggi dello stadio totem-pole e la transizione dal livello logico basso a quello alto risulta nettamente più lenta di quella dal livello logico alto a quello basso. Pertanto le porte open collector si usano soltanto in situazioni particolari.

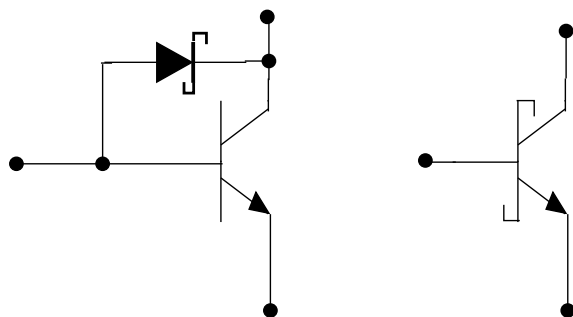
Nel caso in cui si voglia imporre un valore logico costantemente alto su un ingresso, è evidente dalla struttura circuitale delle porte TTL che sarebbe sufficiente lasciarlo non collegato. Tuttavia un terminale flottante può raccogliere disturbi elettromagnetici dall'ambiente circostante, che possono essere evitati connettendo tutti gli ingressi non utilizzati all'alimentazione attraverso una resistenza da $1\text{ k}\Omega$.

15.9 Famiglia logica TTL Schottky

Abbiamo visto che una delle limitazioni fondamentali alla velocità di commutazione delle porte TTL è rappresentata dal tempo che occorre per uscire dalla condizione di saturazione dei transistori, il quale rimane sempre uno degli inconvenienti più

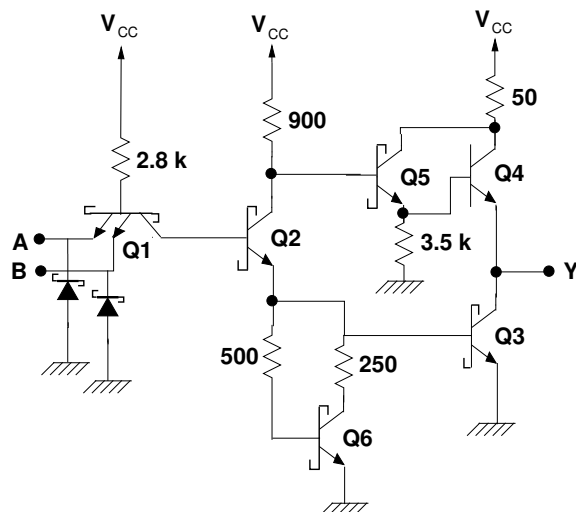
significativi, anche se si mettono in atto accorgimenti per una più rapida estrazione dell'eccesso di portatori minoritari immagazzinati in base in conseguenza della saturazione.

Una soluzione a questo problema è rappresentata dall'evitare che i transistori raggiungano mai la condizione di saturazione. Questo risultato può essere ottenuto sfruttando le proprietà dei diodi Schottky: un diodo Schottky è costituito da una giunzione tra metallo e semiconduttore, che ha un comportamento analogo a quello di una giunzione pn in cui il ruolo della regione n è svolto dal metallo. Con una opportuna scelta del metallo si possono ottenere in questo modo diodi con una tensione di cut-in di 0.5 V invece di quella tipica di 0.7 V delle giunzioni pn in silicio. Se si collega un diodo Schottky (il cui simbolo è analogo a quello di un diodo normale, con l'unica variante della rappresentazione del catodo sotto forma di una sorta di "S" stilizzata) tra collettore e base di un transistor, si ottiene il cosiddetto transistor Schottky, rappresentato di solito con la base a forma di "S".

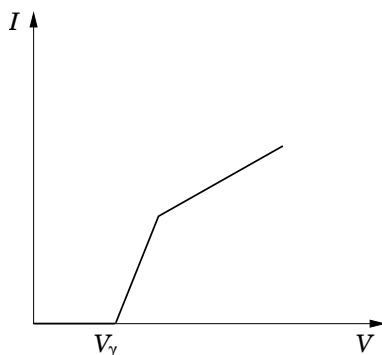


La presenza del diodo Schottky impedisce che la tensione diretta base-collettore possa salire oltre 0.5 V e quindi impedisce, di fatto, che questa possa mai essere polarizzata direttamente in modo da far passare una corrente significativa. In questo modo il transistor Schottky non entra mai in condizione di saturazione, dato che, una volta che il diodo Schottky sia entrato in conduzione, ogni ulteriore aumento della corrente esterna verso la base viene deviato nel collettore attraverso il diodo stesso. La tensione collettore-emettitore non scende al di sotto degli 0.3 V.

Una versione nettamente più veloce della logica TTL è stata realizzata sfruttando transistori Schottky. Di seguito riportiamo lo schema di una porta NAND in tecnologia TTL Schottky.



Notiamo che tutti i transistori, eccetto che per Q_4 , sono stati sostituiti con transistori Schottky. Non è necessario impiegare un transistor Schottky per Q_4 , dato che esso non può mai andare in saturazione, dato che la sua V_{CB} non si inverte mai, essendo pari alla V_{CE} del transistor Q_5 , che è stato aggiunto allo scopo di formare una sorta di transistor Darlington (si dice configurazione Darlington la combinazione di due transistori connessi con i collettori in comune e l'emettitore del primo collegato alla base del secondo; essi formano l'equivalente di un unico transistor ad alto guadagno, detto transistor Darlington) in combinazione con Q_4 , consentendo di fornire una corrente maggiore per il caricamento della capacità di uscita. Il diodo sull'emettitore del Q_4 non è più necessario perché ora la caduta di tensione richiesta è ottenuta tramite le due giunzioni base-emettitore di Q_5 e Q_4 . Sono stati aggiunti anche dei diodi "fissatori" sugli ingressi, che impediscono alle tensioni di ingresso di scendere al di sotto di -0.5 V, dato che tensioni negative potrebbero presentarsi in circuiti veloci a causa del comportamento da linee di trasmissione delle interconnessioni. Infine la resistenza sull'emettitore di Q_2 è stata sostituita con un circuito di pull-down attivo costituito dalle resistenze da 500 e $250\ \Omega$ e dal transistor Q_6 . Tale circuito si comporta da resistenza non lineare, con un valore pressoché infinito finché la tensione ai suoi capi è al di sotto di V_γ e il transistor è quindi interdetto, con un valore piuttosto basso quando il transistor è in zona attiva diretta (dell'ordine della decina di ohm) e, infine, con un valore pari al parallelo delle due resistenze quando il transistor raggiunge la condizione di tensione V_{CE} pari a 0.3 V, corrispondente alla "saturazione" Schottky.



Grazie a tale circuito i transistori Q_2 e Q_3 entrano in conduzione in maniera praticamente simultanea (dato che per tensioni ai suoi capi minori di V_γ il bipolo di pull-down attivo è un circuito aperto) e questo porta a ottenere una caratteristica di trasferimento molto più "squadrata" di quella delle porte TTL standard. In effetti il circuito di pull-down attivo contribuisce a rendere più veloce la commutazione di Q_3 sia dallo stato off a quello on sia viceversa. Infatti, finché Q_3 è interdetto, il bipolo di pull-down attivo è un circuito aperto e tutta la corrente di Q_2 finisce nella base di Q_3 , facilitandone l'attivazione. Quando invece il transistor Q_3 è attivo, il bipolo in questione conduce con una resistenza piuttosto bassa e contribuisce quindi, al momento della commutazione, a rimuovere carica dalla base di Q_3 , rendendo più rapido il processo.

Si nota che tutte le resistenze nel circuito ora esaminato (rappresentativo della porta base della famiglia 74S, dove la "S" sta per Schottky) sono circa la metà o minori di quelle presenti nelle porte TTL standard. Questo implica una dissipazione di energia assai maggiore di quella della logica TTL standard, in cambio di un notevole aumento di velocità. È stata anche sviluppata, in tempi successivi, una famiglia

logica TTL Schottky a bassa dissipazione di potenza, la 74LS (Low-Power Schottky), la quale è caratterizzata da un consumo molto basso. Un ulteriore miglioramento è rappresentato dalla famiglia 74ALS (Advanced Low-Power Schottky), che coniuga un'elevata velocità con una dissipazione di potenza estremamente ridotta.

15.10 Confronto tra le varie famiglie logiche

Le diverse famiglie logiche che abbiamo preso in considerazione si differenziano per velocità di commutazione e dissipazione di potenza. In particolare, i ritardi di propagazione vanno da 1.5 ns della TTL ALS ai 10 ns della TTL standard, mentre la dissipazione di potenza per porta varia da 1 mW per la TTL ALS a 20 mW per la TTL S. Come detto in precedenza, ormai la logica TTL non è più usata, eccetto che per casi particolari, perché sostituita da quella CMOS. Esiste una famiglia logica CMOS, la 74HC, la quale è compatibile pin-to-pin con i corrispondenti componenti TTL, vale a dire che un 74LSXX può essere per esempio sostituito con un 74HCXX senza modifica alcuna al circuito.

Il parametro decisivo nel confronto tra le prestazioni delle diverse famiglie logiche è rappresentato dal prodotto potenza-tempo di propagazione, che risulta molto più basso per la 74HC (data la molto minore dissipazione di potenza in tecnologia CMOS), come è possibile desumere dalla tabella comparativa.

	74	74S	74LS	74ALS	74HC
$P_D t_p$ (pJ)	100	60	20	4.8	0.25
P_D (mW)	10	20	2	1.2	0.025
t_p (nS)	10	3	10	4	10

Pertanto attualmente si preferisce sempre la famiglia 74HC, a meno che non si abbiano esigenze particolari di velocità e la dissipazione di potenza non rappresenti un problema.

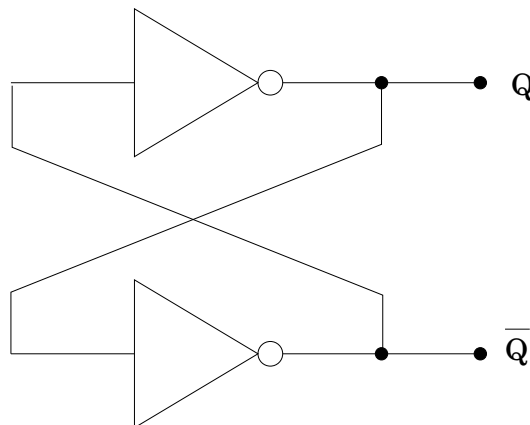
16. Logica sequenziale

16.1 Introduzione

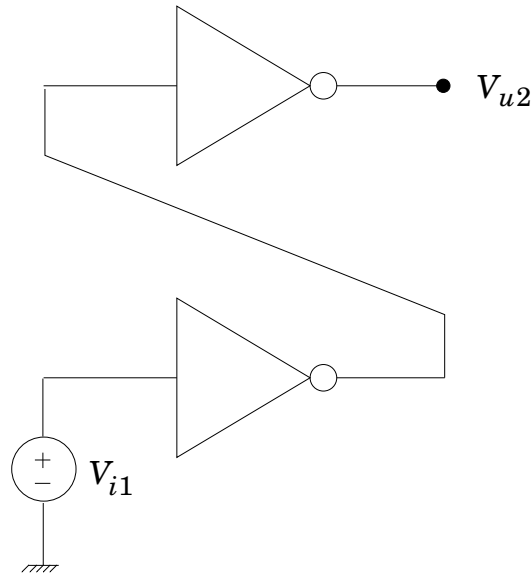
I circuiti logici sequenziali si differenziano da quelli semplicemente combinatori per l'esistenza di variabili di stato interne, i cui valori, insieme con quelli delle variabili di ingresso, determinano i dati in uscita. L'uscita di una rete sequenziale è quindi dipendente dalla storia precedente, della quale rimane traccia nelle variabili di stato, che rappresentano dunque elementi di memoria. La memorizzazione di un dato si può ottenere, rimanendo nell'ambito delle tecniche puramente elettroniche, in maniera "statica" tramite circuiti bistabili, cioè caratterizzati da due stati stabili e pertanto in grado di immagazzinare un bit di informazione, oppure in maniera "dinamica" sfruttando la carica di un condensatore, la tensione ai capi del quale può essere mantenuta quasi costante per un certo tempo, purché le correnti di perdita siano sufficientemente piccole. Tale capacità può anche essere costituita dalla capacità di ingresso dei transistori MOS, che consentono quindi la memorizzazione di un dato, sia pur per tempi molto brevi, in modo estremamente semplice.

16.1 Latch

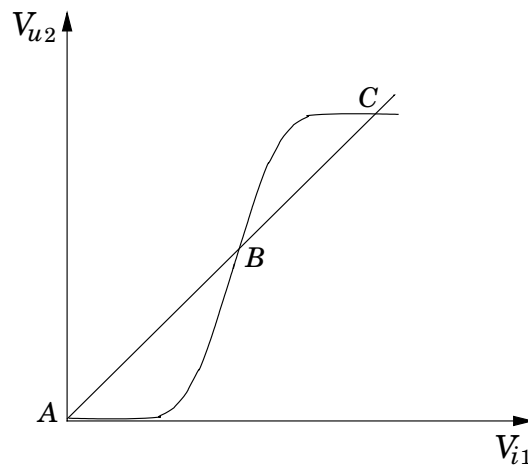
Un latch consiste in un circuito con due stati stabili, che è quindi in grado di memorizzare un bit di informazione. Il modo più semplice di realizzare un latch è rappresentato in figura e consiste nell'utilizzare due inverter in cascata, connettendo poi l'uscita all'ingresso.



Per capire il funzionamento del latch conviene interrompere il collegamento di reazione dall'uscita del secondo inverter (V_{u2}) all'ingresso del primo (V_{i1}), andando poi ad analizzare la caratteristica di trasferimento tra V_{i1} e V_{u2} . Dato che abbiamo due inverter in cascata, la caratteristica di trasferimento totale consiste nell'applicazione consecutiva, per due volte, della caratteristica di trasferimento di un singolo inverter. Nella normale condizione di funzionamento, il collegamento tra l'uscita del secondo latch e l'ingresso del primo impone che le tensioni in questi due punti siano uguali. Questa condizione può essere rappresentata, sul piano della caratteristica di trasferimento, con la retta $V_{u2} = V_{i1}$. Il circuito potrà trovarsi a funzionare in uno dei tre punti in cui tale retta interseca la caratteristica di trasferimento. Dimosteremo ora che il punto B è un punto di equilibrio metastabile, mentre i punti A e C sono di equilibrio stabile. Infatti in B il guadagno della cascata dei due inverter è molto maggiore dell'unità, quindi una piccola perturbazione che dà luogo a uno scostamento ΔV rispetto a B viene notevolmente amplificata e la reazione positiva fa sì che il



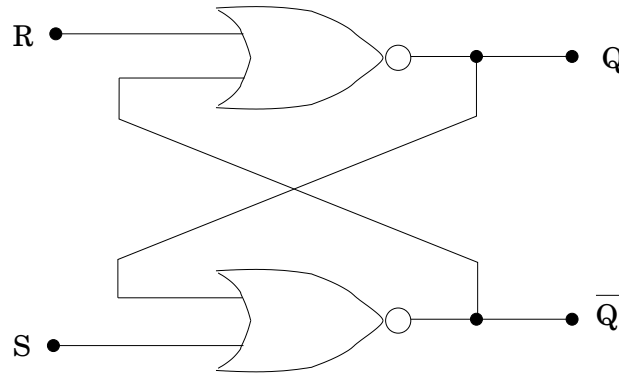
sistema evolva spontaneamente verso A (se $\Delta V < 0$) o verso C (se $\Delta V > 0$). Quando invece ci si trova in A o in C il guadagno della cascata dei due inverter è molto minore dell'unità, quindi una qualunque perturbazione viene attenuata e tende a scomparire, per cui si ha un equilibrio stabile.



Abbiamo pertanto a disposizione un elemento di memoria, in grado di immagazzinare un bit, ma dobbiamo trovare il modo di forzare tale elemento di memoria in uno dei due stati, agendo dall'esterno. L'insieme del latch e della parte di circuito necessaria per la memorizzazione di un dato proveniente dall'esterno si definisce flip-flop. Alcuni testi preferiscono invece utilizzare il termine latch in senso molto più generale, includendo anche i circuiti per l'introduzione del dato e riservano il nome flip-flop soltanto a quei circuiti il cui stato viene aggiornato in corrispondenza della transizione di un segnale di clock (quelli che noi chiameremo flip-flop edge-triggered).

16.2 Flip-flop set-reset

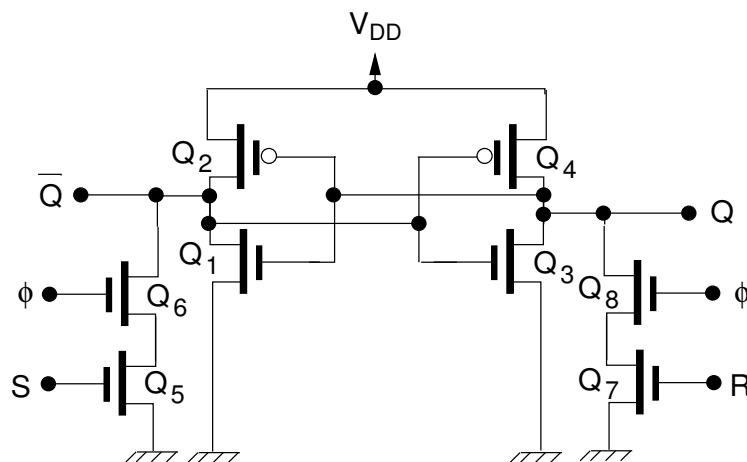
La realizzazione in termini di porte logiche di un flip-flop SR (set-reset) è rappresentata nella figura seguente: i due NOR formano un latch con l'aggiunta di due terminali che consentono di forzare lo stato del latch stesso.



Come risulta dalla tavola della verità, se ambedue gli ingressi sono a livello logico basso, l'uscita rimane nello stato in cui si trovava; se, invece, l'ingresso R si trova allo stato logico basso e quello S allo stato logico alto, l'uscita passa a 1. L'inverso accade se l'ingresso R si trova a 1, mentre quello S è a 0: l'uscita passa a 0. La condizione con tutti e due gli ingressi a 1 dà luogo a un comportamento indeterminato e dovrebbe essere evitata in condizioni normali.

Un analogo risultato si sarebbe ottenuto impiegando, invece delle porte NOR, delle porte NAND, con l'unica differenza che gli ingressi sarebbero stati attivi al livello basso, quindi si sarebbero avuti i terminali di ingresso \bar{S} e \bar{R} .

L'effettiva implementazione in tecnologia CMOS di un flip-flop SR può essere ottenuta semplicemente combinando porte NOR o NAND CMOS secondo gli schemi indicati, ma ciò porterebbe all'utilizzo di un numero di transistori sensibilmente maggiore di quello effettivamente necessario. Una realizzazione possibile, sfruttando un numero di transistori relativamente limitato è quella indicata nello schema che segue, nel quale sono stati inclusi anche degli ingressi di "enable" che consentono di abilitare o meno gli ingressi di set e di reset. Il cuore del flip-flop consiste nei due inverter, costituiti da Q_1 , Q_2 , Q_3 e Q_4 , che sono connessi tra loro nella configurazione già vista per il latch.



Se il valore della variabile ϕ è alto, gli ingressi, che agiscono sui transistori Q_5 e Q_7 , sono abilitati; in caso contrario il latch rimarrà nello stato in cui si trova, indipendentemente dal valore dei segnali applicati agli ingressi S e R . Consideriamo il caso in cui ϕ sia al valore alto, l'uscita Q si trovi al livello basso e sull'ingresso S venga applicato un livello logico alto: purché il rapporto W/L dei transistori Q_5 e Q_6 sia abbastanza grande (svolgeremo più avanti i calcoli per determinare in mo-

do preciso quanto deve valere tale rapporto), la tensione su \overline{Q} scende al di sotto di $V_{DD}/2$, che, nell'ipotesi di Q_3 e Q_4 simmetrici, rappresenta la soglia di commutazione dell'inverter. Da quel punto in poi, a causa della reazione positiva all'interno del latch, il circuito evolve spontaneamente, fino a che \overline{Q} non arriva al livello logico basso e Q a quello alto. Analogamente, se partiamo dalla condizione in cui Q è allo stato logico alto, l'applicazione di un 1 logico sull'ingresso R determina un abbassamento della tensione su Q al di sotto di $V_{DD}/2$, innescando quindi la commutazione del latch, che si conclude con l'uscita Q a livello logico basso. Il funzionamento finora descritto si verifica regolarmente se l'ingresso che determina la commutazione rimane al livello alto per un tempo sufficientemente lungo da provocare l'inizio del fenomeno di commutazione, che poi evolve in modo indipendente, grazie all'azione rigenerativa della reazione positiva.

Vediamo come si può procedere per determinare i rapporti W/L necessari per garantire il corretto funzionamento del flip-flop. Tali rapporti dovranno essere scelti, come già detto, in modo tale da assicurare che, quando viene applicato il segnale di set o quello di reset, la tensione all'ingresso dell'inverter corrispondente scenda al di sotto di $V_{DD}/2$. Prendiamo in esame il caso di una tensione di alimentazione V_{DD} di 5 V e di tensioni di soglia di 1 V per i transistori NMOS e di -1 V per i PMOS. Per semplicità svolgiamo i calcoli come se vi fosse un solo transistor NMOS equivalente al posto di Q_5 e di Q_6 , quindi dovremo studiare la serie di tale transistor e di Q_2 . Nella condizione limite per l'innescio della commutazione la tensione su \overline{Q} è $V_{DD}/2$, quindi

$$\begin{cases} V_{DS_n} = \frac{V_{DD}}{2} = 2.5 \text{ V} \\ V_{GS_n} - V_{T_n} = 4 \text{ V}, \end{cases}$$

per cui il transistor NMOS è in zona triodo. Lo stesso accade per il transistor PMOS:

$$\begin{cases} V_{DS_p} = -\frac{V_{DD}}{2} = -2.5 \text{ V} \\ V_{GS_p} - V_{T_p} = -4 \text{ V}, \end{cases}$$

pertanto $|V_{DS_p}| < |V_{GS_p} - V_{T_p}|$.

Quando abbiamo preso in esame i transistori MOS abbiamo visto soltanto l'espressione per la corrente di drain in condizione di saturazione, laddove risulta indipendente dalla tensione V_{DS} :

$$I_D = k(V_{GS} - V_T)^2(W/L),$$

dove $k = (1/2)\mu C_{ox}$ (μ è la mobilità e C_{ox} è la capacità tra gate e substrato per unità di area).

In zona triodo vale un'espressione un po' più complessa, nella quale compare anche V_{DS} :

$$I_D = \mu C_{ox} \frac{W}{L} \left[(V_{GS} - V_T)V_{DS} - \frac{1}{2}V_{DS}^2 \right],$$

che, in corrispondenza dell'inizio della zona di saturazione, quando $V_{DS} = V_{GS} - V_T$, diventa

$$I_D = \mu C_{ox} \frac{W}{L} \left[(V_{GS} - V_T)^2 - \frac{1}{2}(V_{GS} - V_T)^2 \right],$$

che coincide con l'espressione nota per la zona di saturazione.

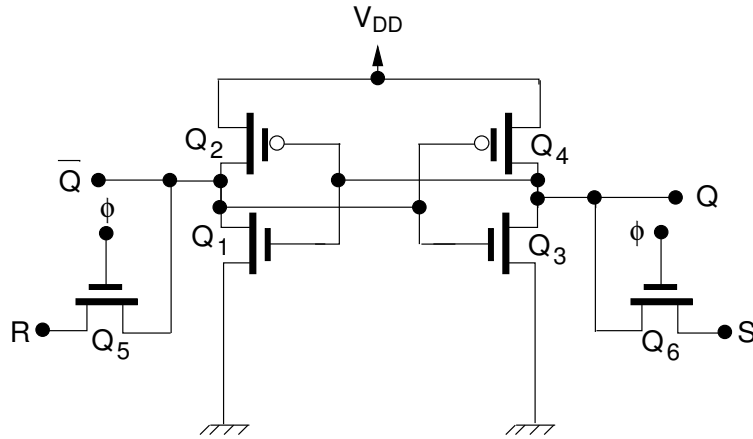
Supponiamo che sia fissato il rapporto W/L per il transistore PMOS, e sia pari a 5, come di solito avviene negli inverter. Nell'usuale ipotesi che la mobilità degli elettroni sia pari a 2.5 volte quella delle lacune, possiamo porre $\mu_n C_{ox} = 50 \mu\text{A}/\text{V}^2$ e $\mu_p C_{ox} = 20 \mu\text{A}/\text{V}^2$.

Poiché i due transistori sono in serie, possiamo imporre l'uguaglianza delle correnti di drain, ricordando che $V_{DS_n} = 2.5 \text{ V}$ e $V_{DS_p} = -2.5 \text{ V}$

$$\begin{aligned} \mu_n C_{ox} \left(\frac{W}{L} \right)_n \left[(V_{GS_n} - V_{T_n}) \frac{V_{DD}}{2} - \frac{1}{2} \left(\frac{V_{DD}}{2} \right)^2 \right] \\ = \mu_p C_{ox} \left(\frac{W}{L} \right)_p \left[(|V_{GS_p}| - |V_{T_p}|) \frac{V_{DD}}{2} - \frac{1}{2} \left(\frac{V_{DD}}{2} \right)^2 \right]. \end{aligned}$$

Da tale relazione, sostituendo i valori numerici, si ricava che $(W/L)_n = 2$. Quindi, per avere un corretto funzionamento del flip-flop è necessario utilizzare un transistore NMOS con W/L almeno pari a 2: di solito, per sicurezza, si usa un W/L almeno di 3. Inoltre, poiché in realtà, nel caso del flip-flop con enable, il transistore NMOS è formato da due transistori in serie, ciascuno di essi dovrà avere un rapporto W/L di 6.

Un'altra possibile realizzazione del flip-flop set-reset, che utilizza due transistori in meno, è indicata nella figura seguente: tale implementazione fa uso di transistori di passo per l'enable e utilizza direttamente i valori delle variabili di set e di reset per agire sullo stato del latch. Si noti che per questa realizzazione del flip-flop si ha una situazione di indeterminazione dell'uscita non solo quando sia l'ingresso di set sia quello di reset sono a 1, ma anche quando sono entrambi a 0.

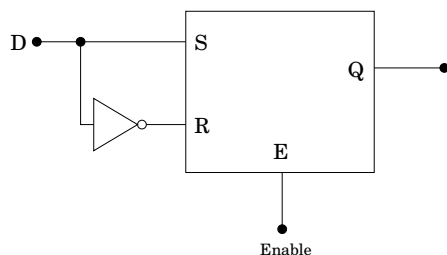


Spesso i pass transistor vengono in effetti realizzati con un gate completo CMOS (formato da un transistore NMOS insieme con un PMOS), per evitare i già discussi problemi di degrado dei livelli logici.

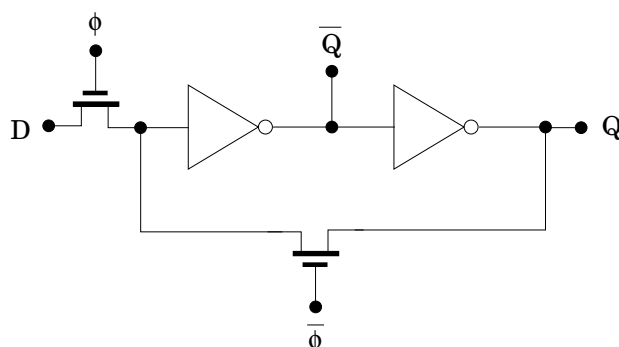
16.3 Flip-flop tipo D

I flip-flop di tipo D presentano in uscita il valore fornito all'ingresso finché il terminale di enable è allo stato alto. Se il terminale di enable viene portato allo stato basso, in uscita viene mantenuto il valore logico presente all'ingresso nell'istante immediatamente precedente la transizione della variabile di enable.

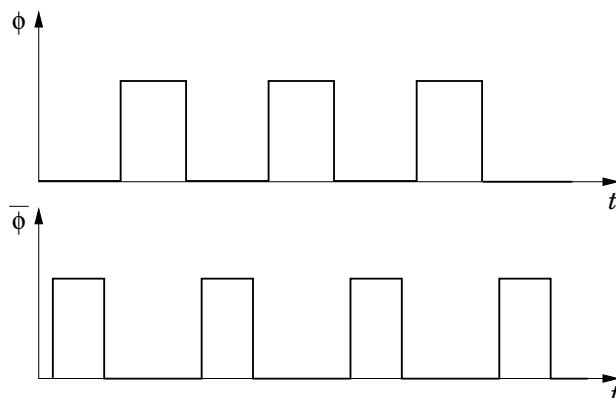
Da un punto di vista logico, perciò, un flip-flop di tipo D può essere ottenuto semplicemente utilizzando un flip-flop SR con enable unitamente a un inverter, come indicato in figura: se il terminale di enable è allo stato alto, l'uscita sarà una copia dell'ingresso, dato che per un valore logico alto in ingresso viene attivato il terminale di set e per un valore logico basso viene attivato quello di reset. Se il terminale di enable viene portato poi a livello basso, l'uscita del flip-flop rimane al valore memorizzato dal latch interno.



Esistono peraltro modi più semplici di realizzare il flip-flop di tipo D, impiegando l'effetto di memorizzazione dinamica che si ha sulla capacità di ingresso delle porte CMOS. Lo schema di principio è rappresentato nella figura seguente, dove $\bar{\phi}$ non è esattamente il complemento di ϕ , ma una fase “non overlapping”, vale a dire non sovrapposta all'altra.

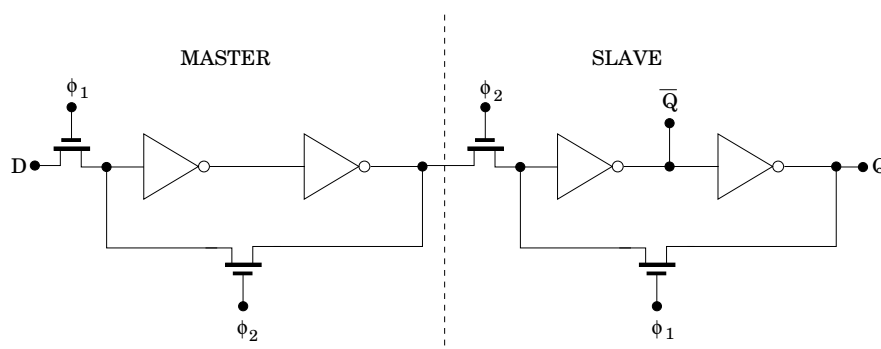


In un clock non sovrapposto le due fasi non sono mai contemporaneamente allo stato logico alto, come indicato nella figura seguente, nella quale gli intervalli con ambedue le fasi a livello logico basso sono stati esagerati per rendere più visibile il concetto; in realtà tale intervallo non è mai superiore a un decimo del periodo.



Esaminiamo il funzionamento del flip-flop D rappresentato: se la fase ϕ del clock è a livello alto, $\bar{\phi}$ è a livello basso e l'anello di reazione è interrotto. Il dato di ingresso attraversa due inverter in cascata per raggiungere l'uscita, dove compare inalterato. Quando invece ϕ passa al livello basso, si interrompe il collegamento con l'ingresso e si forma un latch, che mantiene il dato in uscita. Si noti che nell'intervallo in cui ambedue le fasi del clock sono a livello basso, la capacità di ingresso degli inverter mantiene il dato precedentemente presente in ingresso, finché questo non verrà memorizzato in modo stabile dal latch.

Sviluppando il circuito appena visto è possibile anche ottenere un flip-flop D edge-triggered, vale a dire che memorizza il valore logico presente in ingresso durante una specifica transizione del segnale di clock. Lo schema è quello riportato di seguito e consiste essenzialmente in due flip-flop di tipo D, uno definito master e l'altro slave.



I segnali ϕ_1 e ϕ_2 sono due fasi del clock non sovrapposte. Quando scende ϕ_1 e sale ϕ_2 il flip-flop master memorizza il valore della variabile D immediatamente precedente alla transizione. Essendo ϕ_2 a livello alto, il flip-flop slave risulta trasparente e riporta tale dato in uscita. Quando scende nuovamente ϕ_2 e risale ϕ_1 , il dato viene memorizzato dallo slave e il master ritorna a inseguire l'ingresso. In questo modo viene quindi “fotografato” il valore dell'ingresso in corrispondenza del fronte di discesa di ϕ_1 , trasferito in uscita in corrispondenza del fronte di salita di ϕ_2 e mantenuto fino al successivo ciclo di clock.

17. Memorie a semiconduttore

17.1 Introduzione

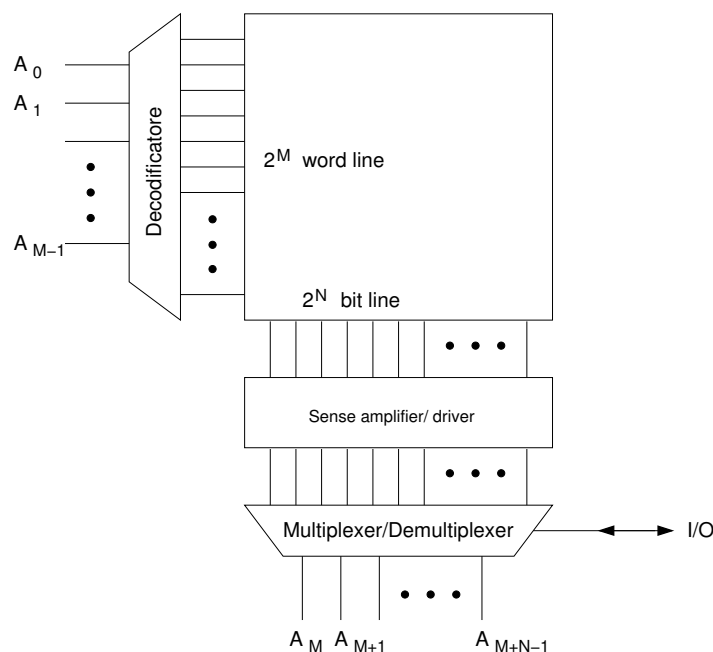
Le memorie rappresentano un componente fondamentale di qualunque sistema di trattamento dati digitale e possono essere realizzate con tecnologie molto diverse tra loro: basti pensare ai nastri magnetici, ai dischi ottici e alle memorie a semiconduttore. La maggior parte delle memorie che consentono un accesso non sequenziale ai dati immagazzinati (quindi con tempo sostanzialmente indipendente dalla posizione del dato), dette anche RAM (Random Access Memory), sono realizzate con tecnologia allo stato solido. Le RAM si distinguono tra “volatili”, nelle quali l'informazione permane soltanto in presenza della tensione di alimentazione e “non volatili”, che sono in grado di conservare i dati anche in assenza della tensione di alimentazione. Queste ultime vengono di solito indicate come ROM (Read Only Memory) se hanno un contenuto predeterminato in fabbrica o PROM (Programmable Read Only Memory) se il loro contenuto può essere modificato dall'utente. Tra le memorie volatili si fa un'ulteriore distinzione tra quelle “statiche” (SRAM), le quali mantengono i

dati indefinitamente fintanto che viene mantenuta l'alimentazione e quelle “dinamiche” (DRAM), nelle quali i dati devono subire un'operazione di “refresh” ogni pochi millisecondi, altrimenti vengono perduti.

L'accesso alle memorie può avvenire per gruppi di bit (byte, word, double-word) o un singolo bit per volta: negli esempi che tratteremo sarà sempre considerato il caso di accesso a un bit per volta.

17.2 Architettura delle memorie

Si potrebbe pensare di immagazzinare ogni bit in un latch e utilizzare poi, per selezionare il bit da leggere, un multiplexer con un numero di ingressi pari a quello dei bit. Una tale soluzione sarebbe però estremamente complessa, perché richiederebbe, dato l'elevato numero di bit contenuti nelle memorie attualmente in uso (siamo già alla DRAM da 1 Gbit), un multiplexer di dimensioni enormi, con un numero di componenti confrontabile a quello necessario per l'immagazzinamento dei dati o addirittura superiore. Si passa quindi da una struttura unidimensionale di questo tipo a una struttura bidimensionale, che corrisponde a organizzare la memoria in forma matriciale, dove ciascun bit si trova all'incrocio tra una riga (word line) e una colonna (bit line). Spesso la dimensione delle memorie è tale che il numero di bit è un quadrato perfetto, per cui è possibile strutturare la memoria nella forma di una matrice quadrata, di dimensione pari alla radice quadrata del numero di bit.



I primi M bit dell'indirizzo del dato che vogliamo andare a leggere o scrivere determinano la selezione, tramite un decodificatore, della word line che contiene il dato stesso. I rimanenti bit di indirizzo controllano un altro circuito di selezione, che opera da multiplexer, selezionando la bit line corrispondente al dato di interesse, in fase di lettura e da decodificatore, portando il dato da memorizzare alla bit line d'interesse, in fase di scrittura. Tra tale multiplexer/demultiplexer e le bit line vere e proprie è presente un ulteriore circuito che svolge la funzione di amplificatore di lettura e di driver per la scrittura.

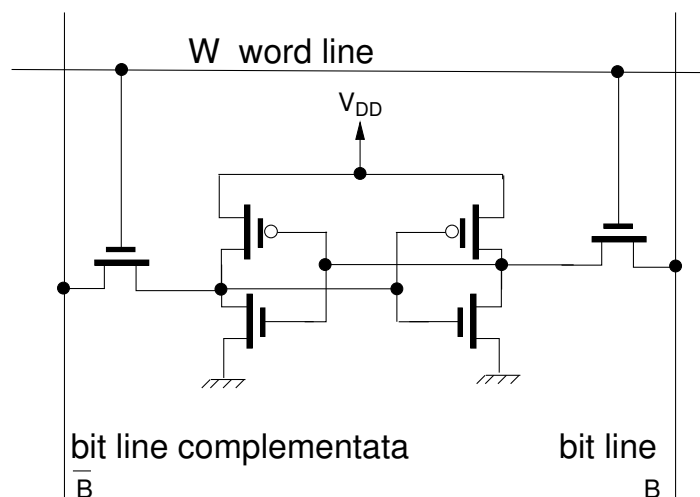
Per memorie particolarmente grandi si può ricorrere a una cosiddetta “organizzazione tridimensionale”, in cui la matrice viene suddivisa in più sottomatrici di

minori dimensioni (di solito in numero corrispondente a una potenza di 2), in modo da ridurre la lunghezza delle bit line e delle word line, allo scopo di contenerne la capacità entro limiti ragionevoli. In tal caso alcuni dei bit di indirizzo vengono utilizzati per selezionare la sottomatrice desiderata (è questo il motivo per cui il numero di sottomatrici è di solito una potenza di 2).

17.3 RAM statiche (SRAM)

Le memorie RAM statiche possono essere realizzate sia in tecnologia bipolare sia in tecnologia CMOS. Nel seguito esamineremo quest'ultima implementazione, che risulta di più diretta comprensione ed è utile anche per introdurre la struttura delle RAM dinamiche.

La cella base della RAM statica CMOS è rappresentata nella figura seguente e consiste sostanzialmente in un flip-flop SR con enable, i cui ingressi sono connessi alla bit line e alla bit line complementata.



I terminali di enable sono collegati alla word line, la quale è quindi in grado di attivare una fila di celle. Vediamo innanzitutto come viene effettuata la lettura: le capacità costituite dalle bit line, sia da quella complementata sia da quella non complementata, vengono precaricate a $V_{DD}/2$, dopodiché la word line viene portata a livello alto, in modo da connettere le uscite del flip-flop a B e a \bar{B} . Tale connessione permane per il tempo strettamente necessario per creare una tensione differenziale tra le B e \bar{B} dell'ordine di 100-200 mV, sufficiente per discriminare il valore logico letto, in modo da effettuare una lettura non distruttiva, vale a dire che non implica un'alterazione del dato memorizzato. Se il contenuto della cella è 1, la tensione sull'uscita dell'inverter di destra del flip-flop è al livello alto, viceversa nel caso che il contenuto della cella sia 0, si trova al livello alto l'uscita dell'inverter di sinistra. Quindi, se leggiamo un 1, la tensione della bit line B passerà da $V_{DD}/2$ a $V_{DD}/2 + \Delta V$, mentre quella della bit line complementata passerà da $V_{DD}/2$ a $V_{DD}/2 - \Delta V$. La differenza di tensione tra le due bit line ($2\Delta V$) viene letta da un apposito "sense amplifier", trasformata in un segnale elettrico a livello logico basso o alto e trasferita in uscita tramite il multiplexer.

Il sense amplifier opererà in modalità differenziale tra la bit line e la bit line complementata, in modo da consentire una buona reiezione dei disturbi a modo comune, come quelli causati da accoppiamenti capacitivi delle bit line con altre parti della memoria o con elementi circuitali esterni.

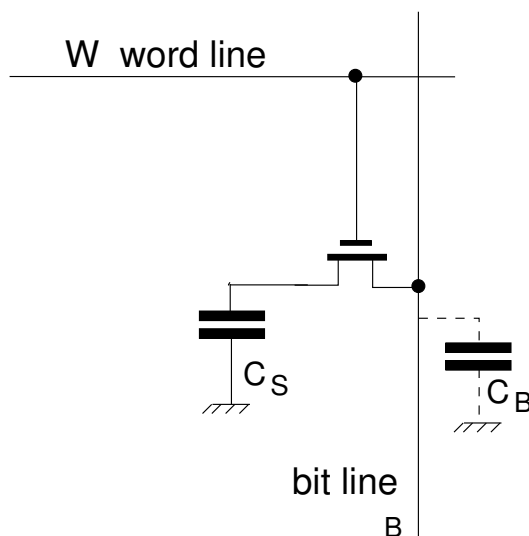
Vediamo ora la fase di scrittura: se desideriamo scrivere un 1, prepariamo B a V_{DD} e \bar{B} a zero (questa volta non precarichiamo semplicemente le capacità delle bit line, lasciandole poi flottanti, ma connettiamo effettivamente le bitline a delle sorgenti di tensione). Poi viene portata a 1 la word line, che determina la connessione tra le bit line e gli ingressi/uscite del flip-flop, che viene quindi forzato nello stato corrispondente a un 1 logico. Viceversa, se vogliamo scrivere uno zero, prepariamo B a 0 e \bar{B} a V_{DD} e procediamo poi nello stesso modo appena descritto.

Durante tali operazioni di scrittura la word line deve chiaramente rimanere a livello alto per un tempo sufficiente a garantire l'innesco della procedura di commutazione del latch contenuto nella cella di memoria.

Poiché i circuiti del sense amplifier, e quelli di precarica e di preparazione delle bit line sono sostanzialmente coincidenti con quelli usati per le memorie dinamiche, li esamineremo una volta trattate le DRAM.

17.4 RAM dinamiche (DRAM)

Le RAM dinamiche sono caratterizzate da una struttura della cella base molto più semplice di quella delle SRAM, con lo svantaggio però che il dato non viene mantenuto indefinitamente, ma deve essere rinfrescato ogni pochi millisecondi. L'informazione è immagazzinata nella tensione ai capi di un condensatore, che può essere connesso alla bit line tramite un transistor di passo NMOS, secondo lo schema della cella elementare rappresentato in figura.



La tensione sul condensatore diminuisce nel tempo a causa delle inevitabili perdite attraverso il condensatore stesso e, soprattutto, attraverso il transistor di passo, che, anche quando è interdetto, lascia passare una sia pur minima corrente. L'occupazione di area su silicio di una cella di memoria DRAM è estremamente ridotta, perché il condensatore viene realizzato con una tecnica a sviluppo verticale, scavando un solco, sulle cui pareti vengono poi realizzate le armature (trench capacitor).

Il condensatore può essere caricato alla tensione $V_{DD} - V_T$ o scaricato completamente tramite il transistor NMOS. Se si vuole memorizzare un 1 logico, si prepara la bit line a V_{DD} e poi si porta la word line al livello logico alto, per un tempo sufficiente a consentire la carica del condensatore. Se invece si vuole scrivere uno 0, si prepara la bit line a tensione nulla e poi si attiva la word line per un tempo sufficiente a garantire la scarica del condensatore.

Per la lettura, precarichiamo la capacità parassita della bit line (indicata con C_B in figura) a $V_{DD}/2$ e poi portiamo la word line a livello alto, connettendo quindi il condensatore di memoria alla bit line: poiché la capacità C_B della bit line è molto più grande di quella C_S della cella di memoria (qualche pF contro poche decine di fF), la tensione della bit line si sposterà di poco dal valore $V_{DD}/2$ e l'operazione di lettura sarà di tipo distruttivo, dato che la tensione sulla capacità di cella sarà equalizzata a quella della bit line. È quindi necessario, ogni volta che si effettua una lettura, riscrivere il dato appena letto nella corrispondente locazione di memoria.

Determiniamo il valore dello scostamento della tensione sulla bit line dovuto alla lettura di un 1 o di uno 0 su una cella DRAM. Nel caso venga letto un 1 logico, la tensione finale diventa $V_{DD}/2 + \Delta V$. Poiché la carica totale deve rimanere invariata, possiamo scrivere

$$C_S V_{CS} + C_B \frac{V_{DD}}{2} = \left(\frac{V_{DD}}{2} + \Delta V \right) (C_S + C_B).$$

Da questa equazione ricaviamo che

$$\Delta V = \frac{C_S}{C_B + C_S} \left(V_{CS} - \frac{V_{DD}}{2} \right) \simeq \frac{C_S}{C_B} \left(V_{CS} - \frac{V_{DD}}{2} \right),$$

dove l'ultimo passaggio è giustificato dal fatto che $C_S \ll C_B$. Ricordando che in presenza di un 1 logico il condensatore C_S è carico a $V_{DD} - V_T$ mentre in presenza di uno zero logico è scarico, otteniamo dunque

$$\begin{cases} \Delta V(1) \simeq \frac{C_S}{C_B} \left(\frac{V_{DD}}{2} - V_T \right) \\ \Delta V(0) \simeq - \frac{C_S}{C_B} \left(\frac{V_{DD}}{2} \right). \end{cases}$$

Per $C_B = 30C_S$, $V_{DD} = 5$ V, $V_T = 1.5$ V, otteniamo $\Delta V(0) = -83$ mV e $\Delta V(1) = 33$ mV. Tali valori risultano dello stesso ordine di grandezza dello sbilanciamento tra la bit line e la bit line complementata che compare nella fase di lettura delle SRAM. Pertanto i circuiti di lettura, in particolare il sense amplifier, saranno dello stesso tipo e li esamineremo una sola volta per ambedue i tipi di RAM.

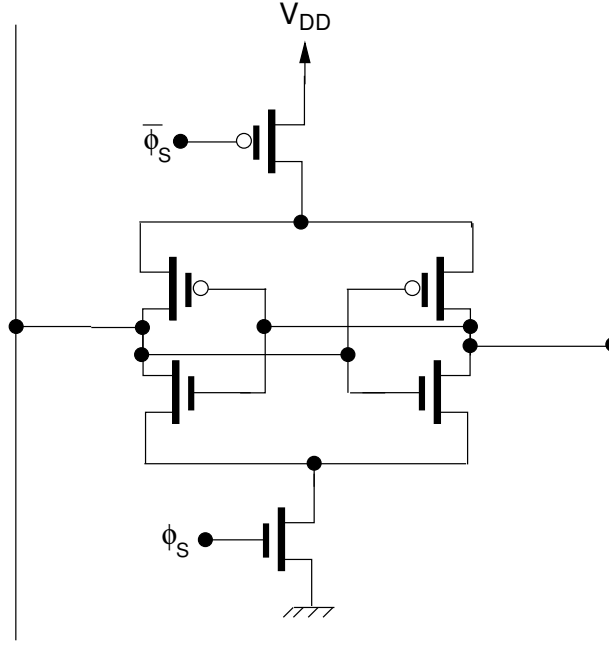
La procedura di rinfresco viene effettuata ponendo la bit line a V_{DD} se si è letto un 1 o a potenziale di massa se si è letto uno 0. A questo punto viene riattivata la word line e il condensatore di memoria può essere caricato alla tensione opportuna ($V_{DD} - V_T$) oppure 0. Dato che un'intera riga viene letta allo stesso tempo, tramite l'attivazione di una word line, tutta la riga può essere rinfrescata nello stesso istante.

Poiché è necessario effettuare un rinfresco ogni pochi millisecondi, un certo tempo viene impiegato a questo scopo (circa il 2-3% del tempo totale), durante il quale la memoria non è disponibile per altre operazioni.

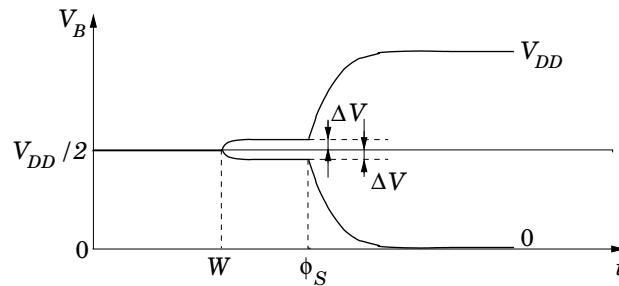
17.5 Procedure di lettura e scrittura

Per la lettura delle memorie DRAM e SRAM è necessario disporre di un amplificatore in grado di ottenere un segnale a livello logico alto o basso a seconda del segno della ΔV in ingresso. Tale amplificatore ha come uscita la tensione corrispondente a uno stato logico, quindi dovrà avere un comportamento non lineare.

L'implementazione di tale amplificatore è di solito ottenuta con un latch le cui uscite sono connesse alla bit line e alla bit line complementata. Tale latch non è alimentato in condizione di riposo e l'alimentazione viene fornita al momento opportuno tramite due pass transistor, uno a canale p per la V_{DD} e uno a canale n per la massa. La struttura descritta è rappresentata nella figura seguente, dove è stato indicato anche un segnale di abilitazione ϕ_S , che determina la connessione del latch alle alimentazioni



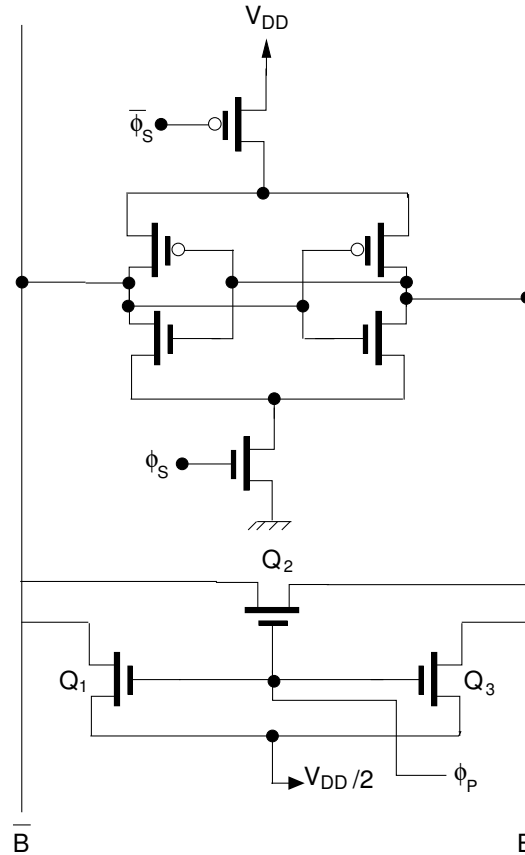
Per effettuare la lettura, una volta che sia stata portata allo stato alto la word line, si porta allo stato alto anche ϕ_S , in modo da alimentare il latch che costituisce il sense amplifier. Il latch inizia a funzionare partendo da una condizione di leggero sbilanciamento tra gli ingressi (intorno a $V_{DD}/2$) dovuto alla connessione alle bit line della cella di memoria: poiché intorno a $V_{DD}/2$ il guadagno degli inverter è molto elevato e il sistema è in reazione positiva, il latch evolve verso una delle due condizioni stabili, con V_B (tensione sulla bit line) che raggiunge V_{DD} se lo sbilanciamento iniziale rispetto a $V_{DD}/2$ era positivo e 0 nel caso di sbilanciamento iniziale negativo. Chiaramente la tensione sulla bit line complementata raggiunge il valore 0 nel primo caso e V_{DD} nel secondo. Tale evoluzione è rappresentata nella figura seguente, dove sono anche indicati gli istanti in cui vengono attivate la word line (W), che rimane alta solo per il tempo necessario a spostare di ΔV la tensione sulla bit line, e il terminale di enable del sense amplifier (ϕ_S).



Si ottiene quindi il risultato di avere sulla bit line e sulla bit line complementata i

valori logici corrispondenti al risultato della lettura “rigenerati”. Il valore logico sulla bit line può quindi essere trasferito in uscita tramite il multiplexer che segue i sense amplifier.

Insieme al sense amplifier viene realizzato anche il circuito necessario per pre-caricare le bit line al valore di tensione $V_{DD}/2$, realizzato con tre transistori NMOS per ciascuna coppia di bit line e bit line complementata, come indicato nella figura seguente, dove viene riportato il circuito complessivo, con l’inclusione anche del sense amplifier.



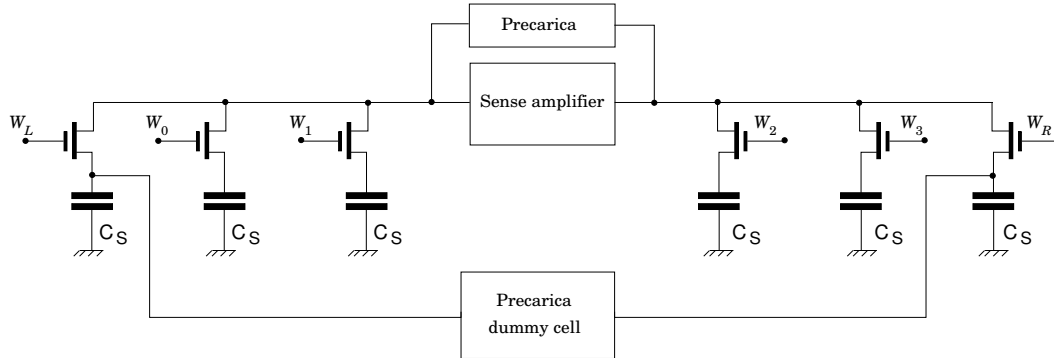
Quando ϕ_P va al livello alto, i transistori Q_1 e Q_3 connettono la bit line e la bit line complementata a una sorgente di tensione a $V_{DD}/2$, mentre il transistore Q_2 , entrando anch'esso in conduzione, garantisce che le tensioni su B e \bar{B} siano esattamente uguali.

Quindi un'operazione di lettura di una SRAM si svolge nel seguente modo: viene posto a livello alto ϕ_P per un tempo sufficiente a precaricare le bit line e le bit line complementate a $V_{DD}/2$, viene abilitata la word line del dato che vogliamo leggere per il tempo necessario a determinare una variazione ΔV sulla tensione delle bit line, viene posto a livello alto ϕ_S , in modo da attivare i sense amplifier, viene selezionato il dato desiderato tramite il multiplexer di uscita e inviato sulla linea di I/O.

Un'operazione di scrittura è più semplice, dato che consiste nel fornire il dato e il suo complemento rispettivamente sulla bit line e sulla bit line complementata, attraverso il multiplexer di uscita che opera in questo caso in senso inverso, come un demultiplexer, dopodiché è sufficiente attivare la word line e il dato viene immagazzinato nella cella di memoria.

Nel caso delle DRAM il circuito utilizzato è del tutto analogo, anche se, in principio, si avrebbe solo una bit line, senza la bit line complementata. Per ottenere

anche nelle DRAM una lettura di tipo differenziale, si suddivide ciascuna bit line in due parti e si pone il sense amplifier al centro con gli ingressi connessi alle due metà della bit line. Si aggiunge inoltre una “dummy cell” a ciascuna delle estremità “esterne” della bit line, come indicato in figura, per un caso semplificato di una DRAM 4×4 bit.



I condensatori delle dummy cell vengono precaricati esattamente a $V_{DD}/2$ e, al momento dell’attivazione della word line della cella di cui intendiamo leggere il contenuto, viene attivata la dummy cell di destra se la cella da leggere si trova nel tratto di sinistra della bit line e, viceversa, viene attivata la dummy cell di sinistra se la cella da leggere è nella porzione a destra della bit line. In questo modo agli ingressi del sense amplifier viene presentato un segnale differenziale, che consiste nella differenza tra la tensione determinata sulla bit line dall’operazione di lettura di un dato e $V_{DD}/2$, differenza che corrisponde proprio alla ΔV precedentemente calcolata.

Anche il refresh dei dati può essere fatto in maniera molto semplice: dopo un’operazione di lettura il sense amplifier presenta, sull’uscita corrispondente al tratto di bit line su cui si è fatta la lettura, una tensione derivante dalla rigenerazione del valore logico contenuto nella cella appena letta (V_{DD} se la cella contiene un 1 o 0 se la cella contiene uno 0). È quindi sufficiente riportare a livello alto la word line, per il tempo necessario per trasferire l’informazione sul condensatore. Ogni volta che si è effettuata una lettura è indispensabile effettuare un anche un refresh, perché l’operazione di lettura di una cella DRAM è di tipo distruttivo. Si noti che con questa operazione di refresh viene anche rinfrescato il contenuto di tutte le altre celle che si trovano sulla stessa word line.

17.6 Decoder e multiplexer per gli indirizzi

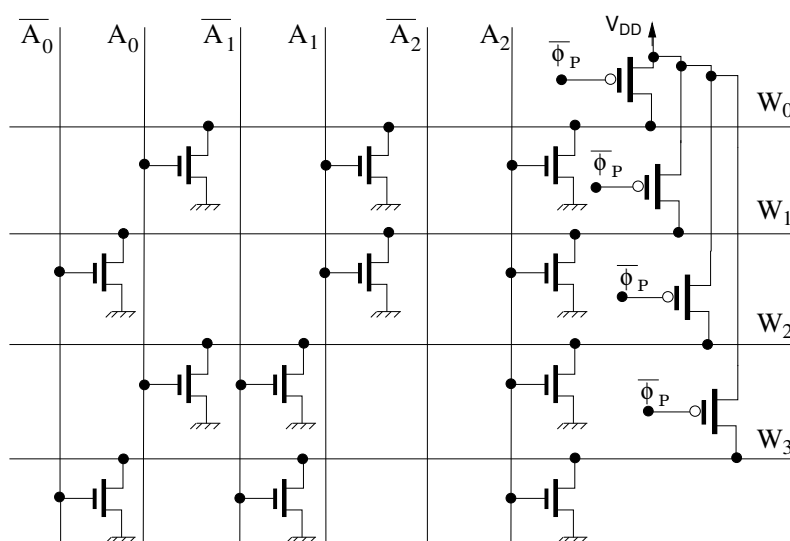
Ciascuna word line deve essere attivata soltanto se sul corrispondente blocco di indirizzi compare il dato associato. Questo risultato potrebbe essere ottenuto utilizzando una appropriata logica combinatoria a porte, ma risulterebbe piuttosto complicato. Sono pertanto stati progettati dei circuiti specifici per questo scopo, come quello che descriviamo nel seguito. L’idea nasce dall’osservazione che la variabile rappresentativa di ciascuna word line può essere espressa come prodotto delle variabili corrispondenti ai bit di ingresso e ai loro complementi, opportunamente scelti. Consideriamo, per semplicità un esempio con solo 3 bit di indirizzo per le word line: si hanno quindi $2^3 = 8$ word line. La prima word line (W_0) dovrà essere attivata sulla base dell’espressione

$$W_0 = \overline{A_0}\overline{A_1}\overline{A_2} = \overline{A_0 + A_1 + A_2}$$

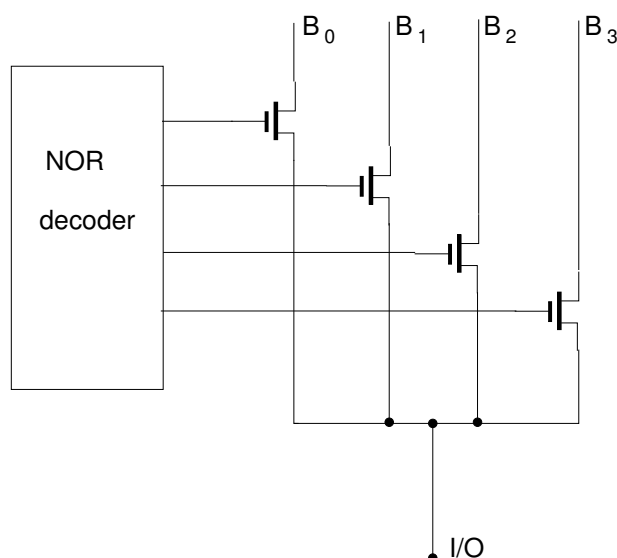
e la seconda dell'espressione

$$W_1 = A_0 \overline{A_1} \overline{A_2} = \overline{\overline{A_0} + A_1 + A_2},$$

dove abbiamo usato le leggi di De Morgan per trasformare il prodotto in somma e quindi l'operazione logica da AND a NOR. Il NOR risultante può essere realizzato in forma di “wired NOR” secondo lo schema che segue, dove il transistor PMOS viene posto in conduzione dal segnale $\overline{\phi_P}$ di precarica, in modo da precaricare tutte le word line, sfruttando la loro capacità, dopodiché tutte meno una, quella che vogliamo effettivamente attivare, vengono scaricate. In questo modo evitiamo significative dissipazioni di potenza. È chiaro che è necessario predisporre qualche altro elemento circuitale per evitare che la precarica di tutte le word line porti a una lettura contemporanea di tutte le celle di memoria, ma non ci addentriamo in questo aspetto.



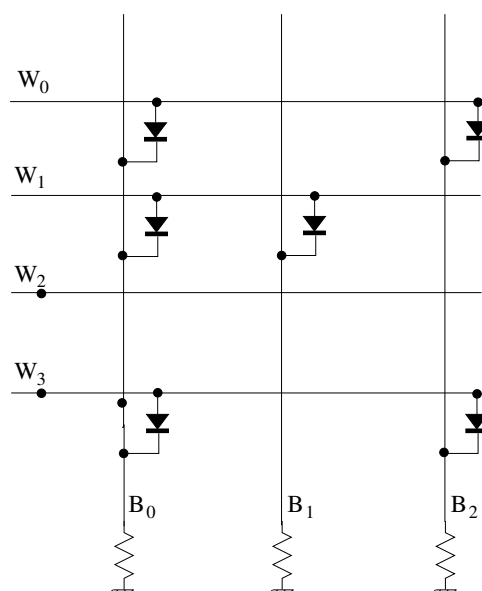
Per quanto riguarda il multiplexer-demultiplexer delle bit line possiamo utilizzare un circuito analogo a quello appena visto, che pilota dei transistori di passo, i quali a loro volta mettono in connessione il terminale di I/O con la bit line prescelta.



La generazione dei diversi impulsi di temporizzazione che abbiamo preso in considerazione viene fatta, con la sequenza corretta, da un'apposita circuiteria presente nel chip di memoria, per cui ben poco di tutto questo è visibile dall'esterno: dal punto di vista del circuito esterno è sufficiente che siano rispettate le temporizzazioni descritte sui data sheet per la presentazione degli indirizzi, dei dati e dei comandi di lettura/scrittura.

17.7 Le memorie ROM

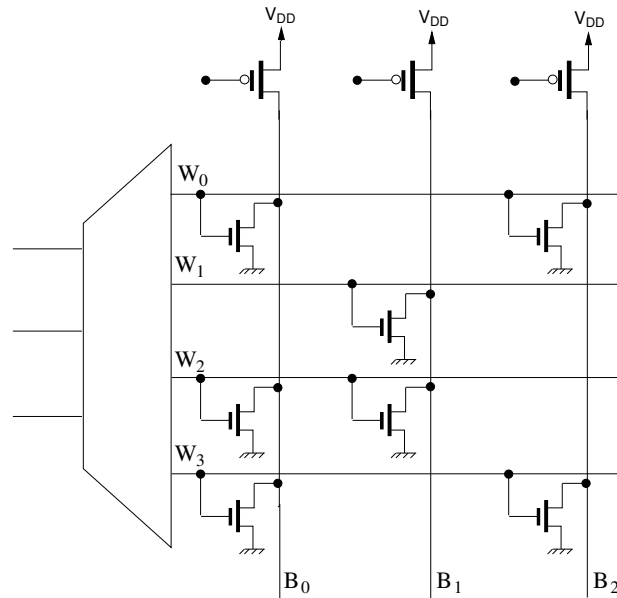
Come indicato dal nome (Read Only Memory), le memorie ROM sono memorie a sola lettura, la cui programmazione viene effettuata durante la fabbricazione. La memoria ROM concettualmente più semplice che si possa pensare consiste in una matrice di word line e bit line, tra le quali si pone un diodo nel caso in cui si voglia immagazzinare un 1 e non si pone alcunché per immagazzinare uno zero. Tale semplice layout è rappresentato nella figura seguente.



Se in corrispondenza di un incrocio tra word line e bit line è presente un diodo, quando la word line passa al livello alto, anche la bit line va a livello alto, mentre in assenza di diodo rimane a livello basso. Si noti che è importante la scelta di utilizzare dei diodi, quindi degli elementi unidirezionali, al posto, per esempio, di corto circuiti: infatti, se si mettessero dei corto circuiti, basterebbe avere due bit appartenenti alla stessa word line posti a livello logico 1 per cortocircuitare tra loro le corrispondenti bit line.

Questa implementazione per una ROM, impiegata talvolta per semplici realizzazioni a livello di circuiti discreti, non viene di solito utilizzata in circuiti integrati. A livello di ROM in forma integrata si preferiscono, per esempio, implementazioni basate su transistori MOS, come quella rappresentata di seguito.

I transistori PMOS hanno la funzione di elementi di pull-up per le bit line, ciascuna delle quali può essere portata a livello logico basso se in corrispondenza con l'intersezione della word line attivata è presente un transistoro NMOS. Il transistoro NMOS va realizzato con un rapporto W/L molto maggiore di quello del PMOS, in modo da poter forzare effettivamente la bit line a un valore di tensione basso. Si può ottenere lo stesso risultato in modo più semplice e con molto minore dissipazione di



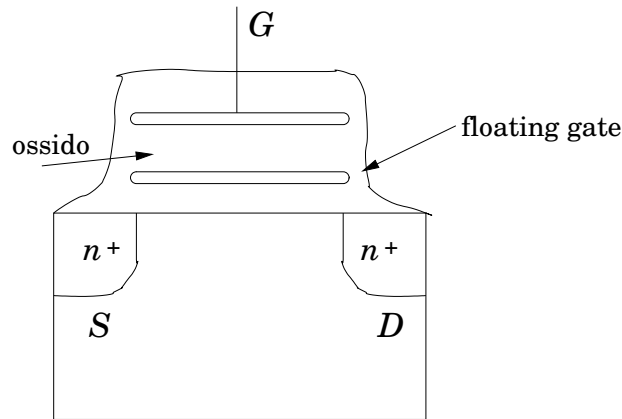
potenza impiegando una modalità di funzionamento dinamica della ROM, precaricando le bit line a V_{DD} tramite il transistor PMOS che viene poi interdetto prima che la word line relativa al bit che vogliamo leggere venga attivata. In tal modo, se è presente un transistor NMOS all'incrocio tra word line e bit line, quest'ultima viene scaricata e in uscita abbiamo uno zero logico. In assenza del transistor NMOS abbiamo invece un uno logico. In realtà si realizzano dei transistori in corrispondenza di ogni incrocio tra bit line e word line, ma vengono collegati alle word line solo i gate di quei transistori che si trovano in corrispondenza di una cella nella quale vogliamo immagazzinare uno zero. In questo modo è possibile differenziare il processo tra ROM contenenti una programmazione diversa soltanto a livello dell'ultima maschera, tramite la quale vengono definiti i collegamenti tra i gate e le word line. I costi di produzione risultano pertanto molto minori di quelli che si avrebbero nel caso si dovessero variare tutte le maschere che definiscono la posizione dei transistori NMOS.

17.8 Le memorie PROM, EPROM, EEPROM

Mentre il contenuto delle memorie ROM viene inserito all'atto della fabbricazione, quello delle memorie PROM (Programmable Read Only Memory) può essere definito dall'utente tramite una specifica procedura di programmazione. Le PROM propriamente dette consentono una sola programmazione, dopodiché il contenuto non può più essere variato, mentre EPROM (Erasable Programmable Read Only Memory) e EEPROM (Electrically Erasable Programmable Read Only Memory) consentono la cancellazione del contenuto e una successiva riprogrammazione. Tutte le memorie citate in questo paragrafo sono non volatili, nel senso che l'informazione in esse immagazzinata permane anche in assenza di tensione di alimentazione.

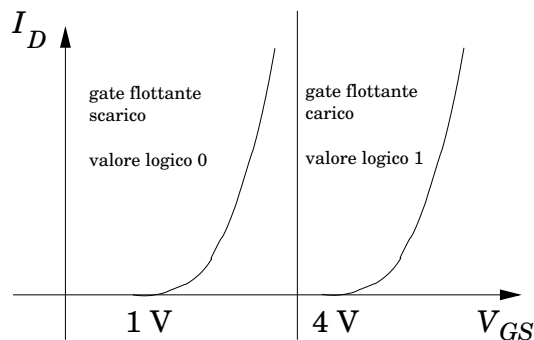
Una PROM può essere realizzata in modo molto semplice, inserendo in serie ai MOS di una ROM a MOS dei fusibili in polisilicio: tutti i gate dei transistori sono collegati alle word line e inizialmente tutti i fusibili sono intatti, quindi in tutti i bit della memoria è contenuto uno zero. Si fa poi passare una corrente abbastanza grande nei MOS che corrispondono alle celle nelle quali vogliamo immagazzinare un 1 logico: il fusibile in polisilicio fonde e il MOS corrispondente non può svolgere più alcuna funzione.

Le EPROM hanno anch'esse una struttura equivalente a una ROM a MOS, ma utilizzano dei transistori MOS molto particolari, con due gate sovrapposti, come illustrato nella figura seguente.



Il gate superiore è collegato al terminale esterno di gate e svolge una funzione analoga a quella del gate di un normale MOS. Il gate inferiore, invece, è localizzato all'interno dell'ossido ed è detto “floating gate”, proprio perché è isolato dal resto del dispositivo. Tali gate vengono realizzati in polisilicio, vale a dire silicio policristallino fatto crescere sull'ossido (mentre sul substrato è possibile far crescere silicio con il corretto ordinamento cristallino, sull'ossido, che è un materiale amorfo, possiamo ottenere soltanto silicio anch'esso amorfo, costituito da molti grani cristallini diversi). Il polisilicio, opportunamente drogato, ha un comportamento elettricamente simile a quello di un conduttore ed è più semplice da depositare di un metallo.

Se il gate flottante viene portato a un potenziale negativo accumulandovi degli elettroni, questo determina una traslazione verso destra della transcaratteristica del MOS: gli elettroni immagazzinati nel gate flottante respingeranno quelli che dovrebbero formare il canale e sarà quindi necessario applicare una tensione positiva più grande al gate superiore per ottenere la formazione del canale stesso. Un caricamento del gate flottante con elettroni porta dunque a un aumento della tensione di soglia V_T del transistor.



Vediamo ora come è possibile caricare il gate flottante. Si impone una tensione V_{DS} piuttosto elevata: questa determina la formazione nel canale dei cosiddetti “elettroni caldi” o “hot electron”, i quali hanno questa denominazione proprio perché, a causa dell'azione del forte campo elettrico, hanno un'energia nettamente superiore a quella del reticolo cristallino, l'interazione con il quale, tramite collisioni, non è sufficiente a creare un equilibrio termico. Proprio in virtù di questa elevata energia, gli elettroni

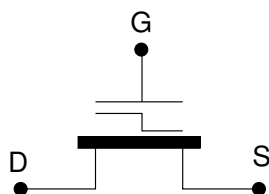
caldi sono in grado di abbandonare il canale e penetrare nell'ossido, una proprietà che li rende particolarmente dannosi nei normali transistori MOS, perché in essi porta a un danneggiamento dell'ossido e a una variazione della carica intrappolata nello stesso, che determinano una variazione nel tempo delle caratteristiche del dispositivo. Nei transistori MOS per le memorie EPROM, invece, questa proprietà viene sfruttata vantaggiosamente, perché applicando anche una tensione sufficientemente grande tra il gate superiore e il canale si riescono a trasferire elettroni caldi nel gate flottante per effetto "tunnel". L'effetto tunnel è una proprietà che nasce dalla natura ondulatoria delle particelle e che consente a una particella di superare, con una probabilità non nulla, una barriera di potenziale anche se questa ha altezza superiore all'energia posseduta dalla particella stessa. La probabilità di attraversamento dipende dalla differenza tra l'altezza della barriera e l'energia della particella (è tanto più piccola quanto più grande è questa differenza) e dallo spessore della barriera nella direzione di attraversamento (diminuisce al crescere dello spessore della barriera). La combinazione dell'accresciuta energia degli elettroni caldi con l'alterazione del profilo di potenziale in direzione verticale (in conseguenza della tensione applicata al gate) rendono la probabilità di tunnel abbastanza grande da caricare rapidamente il floating gate. Non appena si torna a valori normali della V_{DS} e della V_{GS} l'andamento della barriera di potenziale tra gate flottante e canale è tale da rendere la probabilità di ritorno degli elettroni verso il canale estremamente piccola, per cui questi rimangono intrappolati per tempi dell'ordine delle decine di anni. È possibile peraltro causare un rapido ritorno degli elettroni dal floating gate al canale fornendo loro l'energia necessaria per superare la barriera sotto forma di fotoni ultravioletti alla lunghezza d'onda di 253.7 nm. L'energia trasportata da un fotone (come fu per la prima volta postulato per spiegare l'effetto fotoelettrico) è infatti proporzionale alla sua frequenza ν tramite la costante h di Planck: $E = h\nu$. I fotoni della luce visibile, che hanno lunghezza d'onda superiore a circa 350 nm, possiedono quindi un'energia che non è sufficiente a portare l'elettrone al di sopra della barriera (ricordiamo che la frequenza ν è legata alla lunghezza d'onda λ dalla relazione $c = \lambda\nu$, dove c è la velocità della luce nel vuoto). Si noti che, come nell'effetto fotoelettrico, quello che conta è l'energia del singolo fotone e non l'intensità totale del fascio luminoso: anche utilizzando un laser non si riuscirebbe a causare il ritorno degli elettroni se la lunghezza d'onda non è abbastanza piccola.

Le tensioni utilizzate per la scrittura di un bit sono dell'ordine di 25 V per la V_{GS} e di 18 V per la V_{DS} . Prima di scrivere i dati in una EPROM la si cancella tramite esposizione a luce ultravioletta della corretta lunghezza d'onda: in questo modo tutte le celle hanno il floating gate scarico e contengono quindi uno 0 logico. Si procede poi alla scrittura di quelle celle nelle quali vogliamo che sia immagazzinato un 1.

Per la lettura, la word line alla quale è connesso il gate della cella alla quale siamo interessati viene portata a una tensione intermedia tra la V_{T_0} che si ha a gate flottante scarico e la V_{T_1} per gate flottante carico: in questo modo, se il gate flottante è scarico, il transistor conduce e la corrispondente bit line viene posta a livello logico basso; viceversa, se il gate flottante è carico, il transistor non conduce e la corrispondente bit line rimane al livello logico alto al quale era stata precaricata.

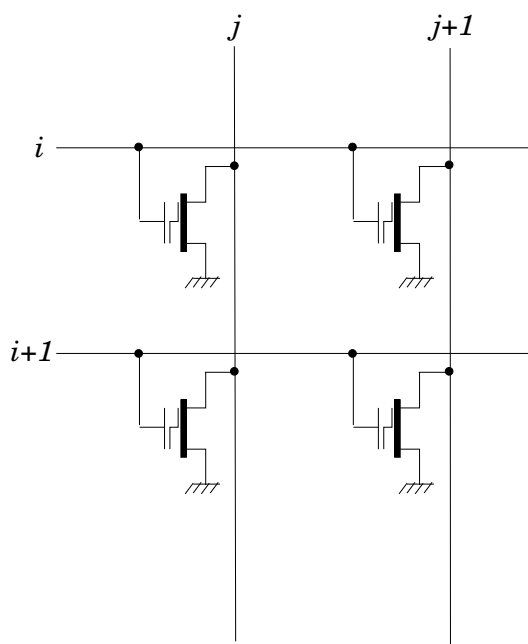
Il principale inconveniente delle memorie EPROM consiste nel fatto che per cancellarle e riprogrammarle è necessaria una esposizione a luce ultravioletta e quindi si tratta di un'operazione abbastanza complicata, che richiede attrezzature ausiliarie e la rimozione della memoria dall'apparecchiatura in cui è inserita. Un notevole progresso è rappresentato dalle memorie EEPROM (o E²PROM) (Electrically Erasable

Programmable Read Only Memory), le quali possono essere cancellate elettricamente, senza la necessità di esporle a luce ultravioletta. Anche le E²PROM si basano sulla struttura della ROM a MOS e impiegano un transistor con gate flottante, che però ha delle caratteristiche diverse da quello utilizzato nelle EPROM: lo strato di ossido che separa il gate flottante dal canale è estremamente sottile. Proprio per tale motivo questo tipo di transistor viene definito FLOTOX (Floating Gate Thin Oxide) MOS. Il simbolo circuitale del transistor FLOTOX è rappresentato nella figura seguente.



Se si pone il drain a un valore di tensione basso (0 V) e il gate viene invece portato a una tensione positiva relativamente elevata (18 V), la barriera tra canale e floating gate viene distorta in modo tale da consentire il passaggio di elettroni verso il floating gate stesso, che si carica. Se si inverte invece il segno della tensione tra gate e drain, si ha la scarica del floating gate, poiché gli elettroni riescono ad attraversare l'ossido in senso inverso. In presenza di tensioni tra gate e canale che rientrano nella gamma utilizzata per il normale funzionamento (lettura) della memoria, gli elettroni eventualmente intrappolati nel gate flottante possono rimanervi per un tempo estremamente lungo, superiore alla decina di anni. Resta da vedere come possiamo sfruttare queste interessanti proprietà del FLOTOX per realizzare una memoria E²PROM.

Potremmo pensare di utilizzare i MOS FLOTOX come i MOS di una ROM, ma nasce un problema: consideriamo per esempio una porzione di 2×2 bit di una ipotetica EEPROM realizzata in maniera analoga a una ROM.

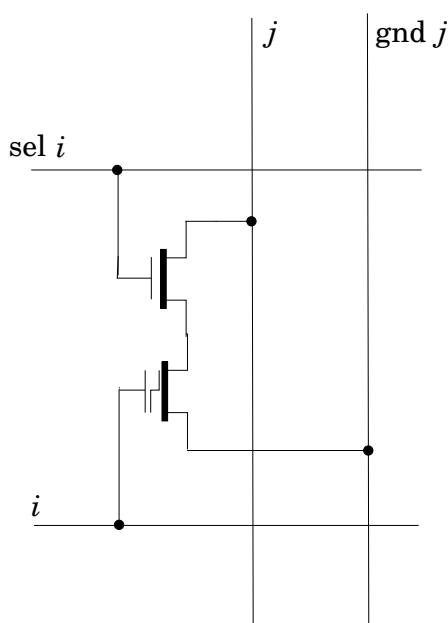


Supponiamo di voler programmare la cella (i, j) : dobbiamo dunque porre una tensione di 18 V sulla word line i e di 0 V sulla bit line j . Per evitare che venga scritta

anche la cella $(i, j + 1)$ dobbiamo porre la bit line $j + 1$ a 18 V e per evitare che venga la cella $(i + 1, j)$ dobbiamo mettere la word line $i + 1$ a 0 V. Quindi, essendo la word line $i + 1$ a 0 V e la bit line $j + 1$ a 18 V, si ha una cancellazione indesiderata del dato in $(i + 1, j + 1)$! Inoltre il transistore $(i, j + 1)$, avendo $V_{DS} = 18$ V e $V_{GS} = 18$ V, dissipa una notevole potenza. Questo semplice layout non è quindi utilizzabile in pratica per la realizzazione di una memoria EEPROM.

Si verificherebbero problemi anche durante l'operazione di cancellazione di una cella. Supponiamo di voler cancellare la cella (i, j) : poniamo dunque la word line i a 0 V e la bit line j a 18 V. Per non cancellare anche la cella $(i, j + 1)$ dovremo porre la bit line $j + 1$ a 0 e per non cancellare la cella $(i + 1, j)$ la word line $i + 1$ dovrà essere a 18 V. In questo modo, però la cella $(i + 1, j + 1)$ viene scritta!

La soluzione che effettivamente si utilizza si basa su una cella elementare più complessa, come quella rappresentata nella figura che segue.



In questo caso è stato aggiunto un secondo transistore, che svolge la funzione di enable (abilitazione) della cella. Inoltre risulta necessario raddoppiare il numero di word line (per l'aggiunta delle linee di selezione, indicate con “sel”) e quello di bit line (per l'aggiunta delle linee per la connessione a massa dei FLOTOX, indicate con “gnd”). L'architettura del chip risulta pertanto notevolmente più complessa, ma questa volta si ottiene il comportamento desiderato. Infatti per programmare la cella (i, j) poniamo la word line i a 18 V, la bit line j a 0, sel i a 18 V, in modo da far condurre il transistore di enable e lasciamo gnd j flottante (per esempio utilizzando una porta tristate). In questo modo il floating gate del transistore i, j viene caricato. Per evitare di cancellare la cella $i + 1, j + 1$ basta mettere sel $i + 1$ a 0, per cui il relativo transistore di enable risulta interdetto. Inoltre, rendendo gnd $j + 1$ flottante si evita una dissipazione di potenza indesiderata nel transistore $i, j + 1$.

Realizziamo una tabella con l'indicazione delle varie tensioni necessarie nelle fasi di scrittura, lettura e cancellazione di una EEPROM del tipo appena considerato. Rappresentiamo con W (write) l'operazione di scrittura (scrittura di un 1, corrispondente al caricamento del gate flottante), con R (read) quella di lettura e con E (erase) quella di cancellazione.

	sel i	i	j	gnd j	sel $i + 1$	$i + 1$	$j + 1$	gnd $j + 1$
W	18	18	0	F	0	0	18	F
E	18	0	18	F	0	0	0	F
R	5	2.5	$D_{i,j}$	0	0	0	$D_{i,j+1}$	0

Nella tabella si è indicata con F la condizione di terminale flottante e con $D_{k,m}$ il dato contenuto nella locazione $\{k, m\}$. Notiamo che tutti gli inconvenienti precedentemente incontrati sono risolti, infatti durante la scrittura della cella (i, j) la cella $(j + 1, i + 1)$ non viene più cancellata, essendo la linea sel $i + 1$ disabilitata; inoltre la cella $i, j + 1$ non è più attraversata da una forte corrente, dato che gnd $j + 1$ è flottante. Anche i problemi che si presentavano durante la cancellazione della cella (i, j) non sussistono più, come si può immediatamente verificare sulla base dei dati contenuti nella tabella.

Osserviamo che nella fase di lettura la word line viene posta alla tensione di 2.5 V: questa è stata scelta in modo da trovarsi in mezzo tra la tensione di soglia con floating gate scarico e quella con floating gate carico. In questo modo, se il floating gate è scarico il transistor conduce e in uscita abbiamo uno 0 logico; se il floating gate è invece carico il transistor non conduce e in uscita otteniamo un 1 logico.

La struttura interna di una EEPROM è, in base a quanto abbiamo visto finora, piuttosto complessa. Esistono peraltro EEPROM che si basano su una cella elementare con un solo transistor FLOTOX, per le quali la cancellazione può essere effettuata soltanto per tutte le celle allo stesso tempo, analogamente a quanto avviene per le EPROM, ma tramite un impulso elettrico. Tali memorie sono denominate FLASH EEPROM e richiedono, per il corretto funzionamento, una circuiteria di controllo piuttosto complessa che non esamineremo nel dettaglio. Esistono infine anche EEPROM con una struttura di cella intermedia tra quella a due transistori e quella della FLASH EEPROM: si tratta di memorie nelle quali la cella elementare è costituita da un singolo transistor che, grazie a un gate con una conformazione particolare, riesce a svolgere contemporaneamente funzioni analoghe a quelle del transistor di enable e il FLOTOX.

Ricordiamo infine che non è possibile scrivere e cancellare le EEPROM un numero indefinito di volte, poiché le caratteristiche ne risulterebbero progressivamente degradate. Il numero massimo di cicli di scrittura effettuabili su una EEPROM senza che si presentino problemi significativi varia in genere tra 10^4 e 10^5 . Le memorie EEPROM, a parte considerazioni di velocità e di costo, non possono essere quindi utilizzate al posto delle RAM nelle applicazioni, più tipiche, in cui il numero di cicli di scrittura è estremamente elevato (come nella memoria di lavoro di un microprocessore).

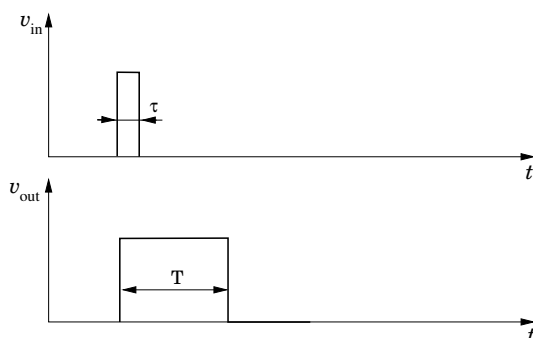
18. Circuiti monostabili, astabili e generazione di segnali di clock

18.1 Introduzione

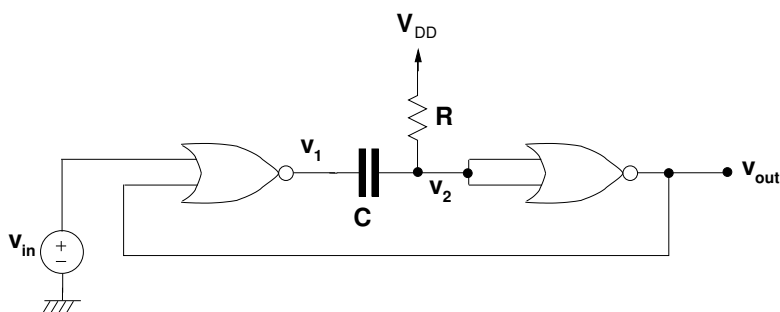
Prenderemo in esame vari circuiti basati su porte logiche e pochi componenti passivi esterni che consentono di produrre impulsi di durata prefissata, forme d'onda quadre e rettangolari, segnali di clock. Vedremo anche l'utilizzo del circuito integrato NE555 che è un interessante esempio di combinazione di funzioni digitali e analogiche e ha rilevante interesse applicativo.

18.2 Multivibratore monostabile

Il multivibratore monostabile, come suggerito dal nome, ha uno stato stabile dal quale può essere spostato tramite una perturbazione esterna e al quale poi ritorna una volta trascorso un intervallo di tempo determinato. Se consideriamo un monostabile per il quale la condizione stabile corrisponde al livello basso, avremo un comportamento come quello descritto nella figura seguente: l'uscita si trova normalmente al livello basso, finché in ingresso non si presenta un impulso di breve durata, il quale determina il passaggio dell'uscita al livello alto per un tempo T determinato dalle caratteristiche interne del monostabile e indipendente dalla durata τ dell'impulso di ingresso, purché questa sia minore di T e maggiore di un valore minimo necessario per causare la commutazione.



Una possibile realizzazione circuitale di un monostabile basato su porte logiche è riportata nello schema che segue, nel quale compaiono due porte NOR e due elementi passivi.



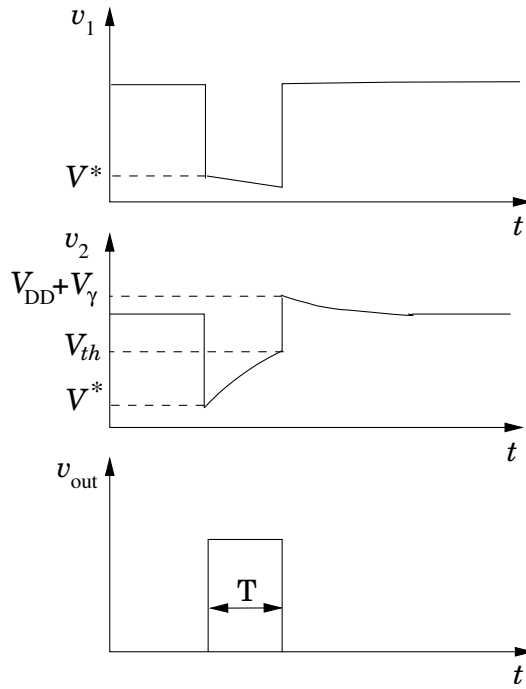
Per l'analisi del circuito facciamo l'ipotesi che siano presenti diodi di protezione agli ingressi delle porte, ma che non sia presente alcuna resistenza di protezione in serie o che tale resistenza sia di valore trascurabile. Cerchiamo innanzitutto di individuare

quale sia lo stato stabile del circuito: in condizioni di regime la resistenza R porterà l'ingresso del secondo NOR (usato semplicemente come inverter) al livello alto, caricando o scaricando opportunamente il condensatore C ; quindi nello stato stabile l'uscita sarà a livello basso. Di conseguenza, poiché in condizioni di riposo l'ingresso è anch'esso a livello basso, l'uscita v_1 del primo NOR sarà a livello alto e il condensatore C sarà scarico, avendo ambedue le armature a livello alto. Se in ingresso si presenta un impulso a livello alto, l'uscita del primo NOR commuta a livello basso, ma non raggiunge esattamente il valore nullo, poiché il transistor di pull-down all'uscita di tale NOR avrà una resistenza R_{ON} diversa da zero. Quindi la tensione v_1 , subito dopo la commutazione, sarà pari a

$$V^* = V_{DD}R_{ON}/(R + R_{ON})$$

e quindi la sua variazione, rispetto all'istante immediatamente precedente, risulterà

$$\Delta v_1 = \frac{R}{R + R_{ON}} V_{DD}.$$



La tensione v_2 subirà istantaneamente la stessa variazione, data l'inerzialità del condensatore C rispetto alla tensione, dopodiché inizierà a salire con legge esponenziale, a causa del transitorio di carica di C , tendendo a V_{DD} . Peraltro, quando v_2 raggiungerà la soglia di commutazione V_{th} dell'inverter rappresentato dal secondo NOR, l'uscita passerà di nuovo al livello basso, determinando la commutazione del primo NOR, la cui uscita v_1 non potrà però raggiungere istantaneamente il valore V_{DD} , poiché la variazione di tensione su v_1 deve essere uguale a quella su v_2 (data l'inerzialità del condensatore) e quest'ultima non può superare $V_{DD} + V_\gamma$, a causa della presenza all'ingresso della porta di diodi di protezione. Quindi sia la v_1 sia la v_2 possono variare istantaneamente solo di una quantità pari a $V_{DD} + V_\gamma - V_{th}$. È da notare che la v_1 raggiungerà poi il valore V_{DD} molto velocemente, dato che la costante di tempo

del transitorio relativo risulta pari al prodotto di C per la somma della R_{ON} (questa volta dovuta alla rete di pull-up della prima porta) e della resistenza del diodo (in conduzione) di protezione della seconda porta. Tali resistenze sono molto piccole, quindi la costante di tempo è trascurabile sulla scala del grafico rappresentato, per cui il ritorno di v_1 a V_{DD} è praticamente istantaneo. Il ritorno di v_2 a V_{DD} è invece più lento, poiché, non appena la tensione v_2 scende al di sotto di $V_{DD} + V_\gamma$, i diodi di protezione risultano interdetti e la costante di tempo diventa molto più lunga dato che è pari al prodotto di C per la somma di R e di R_{ON} . Il monostabile sarà pronto a ricevere un nuovo impulso di trigger in ingresso solo quando v_2 sarà tornata a V_{DD} (se desideriamo che il periodo per cui l'uscita rimane alta sia esattamente T): si parla quindi di un tempo di “recovery” che deve trascorrere dopo il termine dell'impulso di uscita prima che il monostabile sia pronto per operare nuovamente.

Procediamo ora a determinare l'effettiva durata T dell'impulso di uscita del monostabile. Si tratta di studiare il transitorio di carica del condensatore C . La tensione iniziale su v_2 risulta $V_i = V^*$ e quella finale a cui tenderebbe v_2 è $V_f = V_{DD}$. Dobbiamo determinare il valore dell'intervallo di tempo T trascorso il quale v_2 raggiunge il valore V_{th} di commutazione, che possiamo porre pari a $V_{DD}/2$. L'espressione del transitorio risulta:

$$\begin{aligned} v_2(t) &= V_f + (V_i - V_f)e^{-t/\tau} \\ &= V_{DD} + (V^* - V_{DD})e^{-t/\tau} \\ &= V_{DD} - V_{DD} \frac{R}{R + R_{ON}} e^{-t/\tau}, \end{aligned}$$

dove $\tau = C(R + R_{ON})$, dato che la resistenza vista da C è $R + R_{ON}$.

Poiché all'istante T viene raggiunta la tensione di soglia V_{th} , possiamo scrivere

$$V_{th} = V_{DD} - V_{DD} \frac{R}{R + R_{ON}} e^{-T/\tau},$$

che, con semplici passaggi, diventa

$$\frac{V_{DD}}{V_{DD} - V_{th}} \cdot \frac{R}{R + R_{ON}} = e^{T/\tau}$$

e quindi

$$T = (R + R_{ON})C \ln \left(\frac{V_{DD}}{V_{DD} - V_{th}} \cdot \frac{R}{R + R_{ON}} \right).$$

Osserviamo che la presenza dei diodi tagliatori non influenza il calcolo di T , dato che essi entrano in gioco solo quando l'uscita del monostabile torna al livello basso. Se assumiamo trascurabile il valore di R_{ON} , possiamo scrivere

$$T = RC \ln \left(\frac{V_{DD}}{V_{DD} - V_{th}} \right)$$

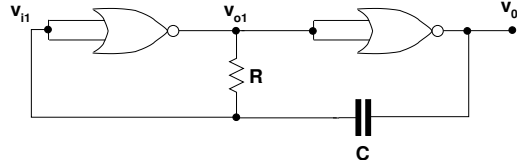
e, se assumiamo $V_{th} = V_{DD}/2$,

$$T = RC \ln 2.$$

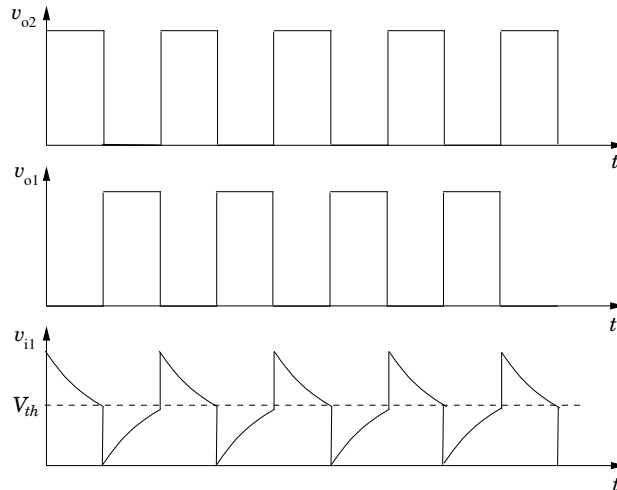
Il tempo di recovery che dobbiamo attendere prima di poter riutilizzare il monostabile, corrispondente al tempo necessario per riportare v_2 a V_{DD} , è pari ad alcune costanti di tempo RC (se R_{ON} è trascurabile rispetto a R).

18.3 Multivibratore astabile

È possibile, con l'utilizzo di porte logiche e di pochi altri componenti, ottenere anche dei multivibratori astabili, di cui forniamo un esempio nella figura seguente, dove sono indicati dei NOR, che svolgono peraltro la funzione di semplici NOT.



Analizziamo il comportamento del circuito facendo le ipotesi semplificative che la R_{ON} di uscita delle porte sia nulla e che i diodi di protezione sugli ingressi siano ideali, con $V_\gamma = 0$ e senza alcuna resistenza in serie. Supponiamo di partire da una condizione in cui l'ingresso del primo NOT (v_{i1}) è superiore alla soglia di commutazione V_{th} : in tal caso l'uscita v_{o1} del primo NOT è a livello basso e quella v_{o2} del secondo NOT è a livello alto. Il condensatore tende quindi a caricarsi (con polarità positiva a destra e negativa a sinistra), finché V_{i1} non raggiunge il valore V_{th} . A questo punto v_{o1} passa a V_{DD} e v_{o2} passa a 0.



In assenza dei diodi di protezione sugli ingressi, la tensione v_{i1} scenderebbe a un valore pari a $V_{th} - V_{DD}$, data l'inerzialità del condensatore alle variazioni di tensione (la tensione sul terminale a destra del condensatore subisce una variazione da V_{DD} a 0, quindi quello a sinistra dovrebbe subire un'identica variazione), tuttavia i diodi di protezione impediscono che v_{i1} diventi negativa e quindi si raggiunge semplicemente lo zero. Istantaneamente anche v_{o2} non può subire una variazione maggiore di quella di v_{i1} , ma raggiunge poi lo zero con una costante di tempo molto veloce (analoga a quella già vista per il transitorio di v_{i1} nel caso del monostabile) e trascurabile sulla scala del periodo del segnale di uscita. A questo punto il condensatore C comincia a caricarsi con polarità positiva a sinistra e negativa a destra, finché v_{i1} non raggiunge il valore V_{th} : ciò determina una nuova commutazione delle due porte, quindi v_{o1} passa a zero e v_{o2} risale a V_{DD} . In assenza dei diodi di protezione v_{i1} raggiungerebbe il valore $V_{th} + V_{DD}$, ma, proprio per l'azione di tali diodi, non può salire oltre V_{DD} . Da questo momento in poi inizia un nuovo ciclo, con una diminuzione della tensione v_{i1} fino al raggiungimento di V_{th} , ecc. Per completezza va detto che, essendo la resistenza per la limitazione di corrente posta in serie agli ingressi delle porte di valore abbastanza elevato, l'effettiva tensione raggiungibile ai terminali di ingresso delle porte logiche è

sostanzialmente la stessa che si avrebbe senza diodi di protezione: potremmo quindi trascurare completamente l'effetto di tali diodi, ottenendo risultati un po' diversi per i calcoli che seguono.

Procediamo ora alla determinazione del periodo della forma d'onda prodotta da questo circuito, definendo T_1 la porzione del periodo in cui l'uscita si trova a livello logico basso e T_2 la restante porzione. Nel transitorio relativo a T_1 la tensione v_{i1} varia secondo il transitorio

$$v_{i1} = V_{DD} \left(1 - e^{-t/\tau}\right),$$

dove τ è la costante di tempo RC . Poiché la commutazione avviene quando v_{i1} raggiunge V_{th} , abbiamo:

$$\frac{V_{DD} - V_{th}}{V_{DD}} = e^{-T_1/\tau}.$$

Pertanto

$$T_1 = RC \ln \frac{V_{DD}}{V_{DD} - V_{th}}.$$

Calcoliamo poi T_2 : poiché durante T_2 la tensione su v_{i1} scende da V_{DD} , tendendo a 0, possiamo scrivere

$$v_{i1} = V_{DD} e^{-t/\tau}.$$

Dato che la commutazione si verifica quando v_{i1} raggiunge V_{th} ,

$$V_{th} = V_{DD} e^{-T_2/\tau}.$$

Quindi

$$\frac{V_{th}}{V_{DD}} = e^{-T_2/\tau},$$

per cui

$$T_2 = RC \ln \frac{V_{DD}}{V_{th}}.$$

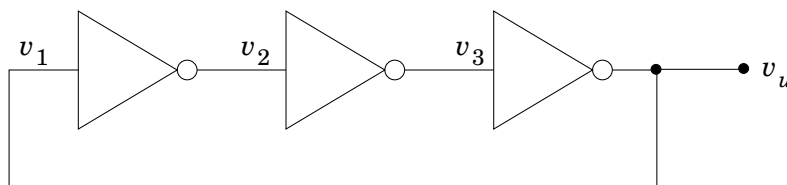
Il periodo totale sarà dato dalla somma di T_1 e T_2 :

$$T = T_1 + T_2 = RC \left[\ln \left(\frac{V_{DD}}{V_{DD} - V_{th}} \cdot \frac{V_{DD}}{V_{th}} \right) \right].$$

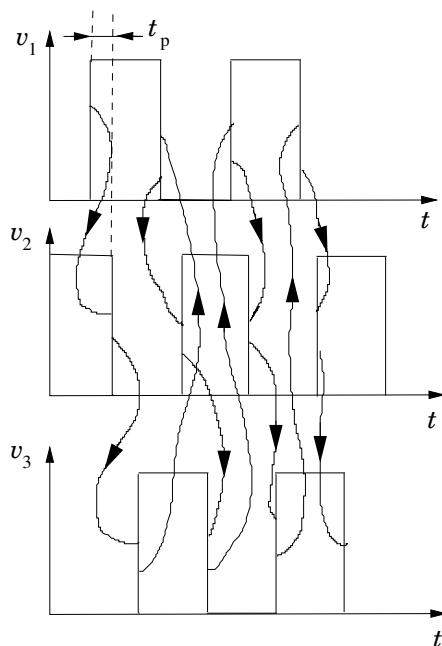
Nel caso in cui sia $V_{th} = V_{DD}/2$, T_1 e T_2 risultano tra loro uguali e la forma d'onda in uscita è quadra, con periodo $T = RC \ln 4$. Se si ripetesse lo stesso calcolo senza considerare l'azione dei diodi di protezione (assumendo la presenza di resistenze elevate in serie agli ingressi) otterremmo $T = 2RC \ln 3$.

18.4 Oscillatore ad anello

Se si pongono in cascata un numero dispari di circuiti NOT e poi si riporta l'uscita dell'ultimo all'ingresso del primo è possibile realizzare un particolare tipo di generatore di onda quadra detto oscillatore ad anello (ring oscillator). Esaminiamone il funzionamento prendendo in considerazione il caso di tre soli inverter in cascata (di solito se ne utilizzano almeno 5), in modo da rendere la trattazione abbastanza semplice.



Supponiamo di avere un fronte di salita su v_1 : questo si trasformerà, su v_2 , in un fronte di discesa ritardato di un tempo pari al ritardo di propagazione t_p ; dopo un ulteriore intervallo t_p avremo un fronte di salita su v_3 e, infine, dopo $3t_p$, un fronte di discesa arriverà su v_1 . Tale fronte di discesa si propagherà attraverso la catena di inverter, fino a ripresentarsi come fronte di salita su v_1 dopo un altro intervallo pari a $3t_p$. Ne consegue che il circuito in esame produce un'onda quadra con un periodo pari a $6t_p$ e quindi con frequenza $1/(6t_p)$. Se si ha un numero N di inverter (che deve essere dispari perché il comportamento descritto abbia luogo), otteniamo una frequenza pari a $1/(2Nt_p)$.



Quindi un oscillatore ad anello può rappresentare un modo molto semplice per misurare il ritardo di propagazione di un particolare inverter. La misura di una frequenza relativamente bassa (se si usa un numero sufficiente di inverter) è infatti un'operazione nettamente più semplice della misura di un ritardo molto piccolo.

18.5 Oscillatori quarzati

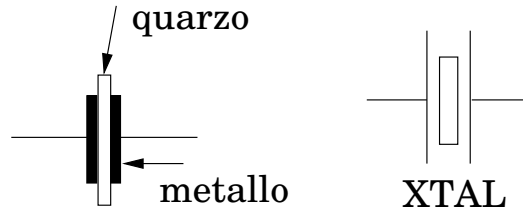
I generatori di forme d'onda che abbiamo visto finora sono adatti (eccetto quelli ad anello) a raggiungere frequenze non molto elevate (al più una decina di kHz) e la loro frequenza di lavoro dipende fortemente dal valore di elementi circuitali che presentano notevoli derive in funzione della temperatura. Quindi, se vogliamo ottenere un segnale di clock a frequenza elevata e con buone caratteristiche di stabilità in frequenza, dobbiamo ricorrere a soluzioni di tipo diverso. Un metodo piuttosto economico per raggiungere stabilità in frequenza dell'ordine della parte su milione consiste nell'utilizzo delle proprietà piezoelettriche di sottili lamine di quarzo.

I materiali piezoelettrici presentano la caratteristica di dar luogo a una differenza di potenziale elettrico se sollecitati meccanicamente (si pensi agli accendigas piezoelettrici) o di deformarsi se soggetti all'applicazione di un campo elettrico (si pensi agli altoparlanti piezoelettrici). Queste importanti proprietà consentono di far interagire un sistema risonante meccanico con un circuito elettrico, mantenendo l'elevato Q caratteristico del sistema meccanico. Il fattore di qualità Q rappresenta sostanzialmente il rapporto tra l'energia in gioco negli elementi reattivi e quella dissipata in

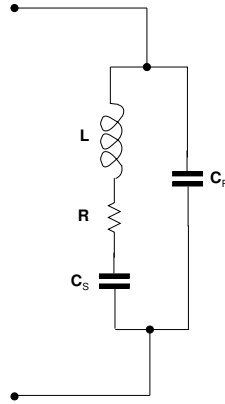
un ciclo negli elementi resistivi. Un sistema senza dissipazione sarebbe caratterizzato da un Q infinito, dato che l'energia dissipata risulterebbe nulla. Nel caso di un filtro passa banda il Q si definisce come il rapporto fra la frequenza centrale e la banda ($Q = f_0/B$), infatti tanto minore è la componente dissipativa, tanto più stretta è la curva di risposta del filtro.

Nei risonatori elettrici, a causa delle inevitabili resistenze parassite, non si riescono a raggiungere valori di Q superiori a qualche centinaio, mentre nei risonatori meccanici valori di molte migliaia sono possibili.

Il risonatore meccanico che si utilizza negli oscillatori di clock è costituito da una lamina di cristallo piezoelettrico (genericamente definito in campo elettronico semplicemente “quarzo”) sulle cui superfici sono stati depositati due elettrodi metallici, come nella figura che segue, dove è anche rappresentato il simbolo circuitale del quarzo.



Possiamo tracciare un circuito elettrico equivalente del quarzo, costituito dal parallelo di un circuito risonante RLC_s , rappresentativo della risonanza meccanica, e di un condensatore C_p , che rappresenta la capacità tra gli elettrodi.



Calcoliamo l'impedenza del circuito equivalente che abbiamo rappresentato: si tratta di valutare il parallelo dei due rami, trascurando la resistenza R , che risulta molto piccola, data la scarsa dissipazione. Abbiamo dunque:

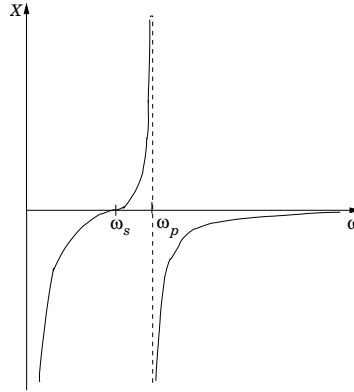
$$\begin{aligned} Z(j\omega) &= \frac{1}{j\omega C_p + \frac{1}{\frac{1}{j\omega C_s} + j\omega L}} = \frac{1 - \omega^2 LC_s}{(1 - \omega^2 LC_s)j\omega C_p + j\omega C_s} \\ &= \frac{1 - \omega^2 LC_s}{j\omega C_p \left(\frac{C_s + C_p}{C_p} - \omega^2 C_s L \right)} = \frac{\frac{1}{C_s L} - \omega^2}{j\omega C_p \left(\frac{C_s + C_p}{C_s C_p L} - \omega^2 \right)}. \end{aligned}$$

Se definiamo $\omega_s = 1/\sqrt{C_s L}$ e $\omega_p = \sqrt{(C_p + C_s)/(C_s C_p L)}$ (osserviamo che $\omega_p > \omega_s$, dato che $C_p C_s/(C_p + C_s)$ è minore di C_s), possiamo riscrivere la precedente equazione

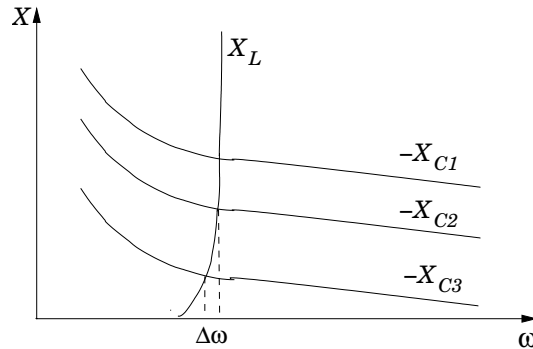
nella forma

$$Z(j\omega) = \frac{1}{j\omega C_p} \frac{\omega^2 - \omega_s^2}{\omega^2 - \omega_p^2}.$$

Notiamo che l'impedenza è puramente reattiva, dato che il primo fattore è immaginario e il secondo reale. Notiamo inoltre che la reattanza (la parte immaginaria dell'impedenza) risulta negativa per $\omega < \omega_s$, positiva per ω compresa nell'intervallo (molto piccolo) tra ω_s e ω_p e nuovamente negativa per $\omega > \omega_p$.



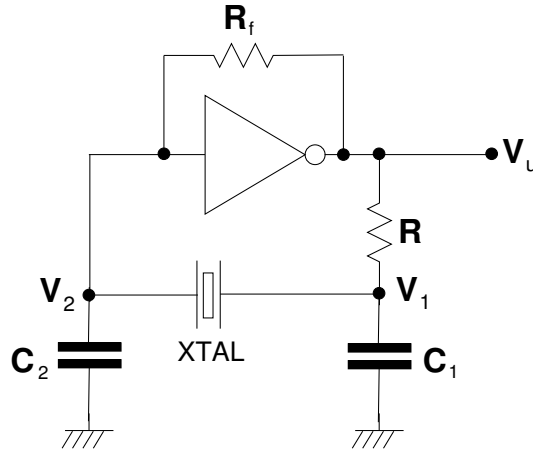
Notiamo inoltre che nell'intervallo in cui la reattanza è positiva (induttiva) ha anche un andamento molto ripido in funzione della frequenza. Questo è un aspetto importante, perché se utilizziamo il quarzo in questa regione in combinazione con una reattanza capacitiva esterna, si avrà una frequenza di risonanza che varia poco anche per grosse variazioni della componente capacitiva, come è facile comprendere dalla rappresentazione grafica contenuta nella figura seguente.



La risonanza si ottiene quando la reattanza totale risulta nulla, quindi quando $-X_C$ (reattanza capacitiva) è pari a X_L (reattanza induttiva del quarzo). Data la forte pendenza di X_L , anche ampie variazioni di X_C determinano uno spostamento $\Delta\omega$ molto piccolo della pulsazione alla quale si ha la risonanza.

Negli oscillatori il quarzo viene sempre utilizzato nella regione di frequenze nella quale ha un comportamento induttivo, sia perché in essa l'andamento in funzione della frequenza è più ripido sia perché è preferibile sostituire i componenti induttivi (i quali presentano perdite a causa della loro resistenza serie) piuttosto che quelli capacitivi, che hanno un comportamento più vicino a quello ideale.

Uno schema tipicamente impiegato è quello dell'oscillatore di Pierce, che utilizza come elemento attivo un inverter CMOS, secondo lo schema di seguito riportato.



La resistenza R_f è necessaria per mantenere il punto di lavoro esattamente nel centro della caratteristica, laddove il guadagno è molto elevato. Infatti il quarzo non consente il passaggio di una componente continua, quindi in assenza di R_f la componente continua subirebbe una deriva non controllabile. In conseguenza della presenza di R_f la componente continua della tensione di ingresso deve risultare pari a quella della tensione di uscita (non c'è caduta di tensione in continua su R_f , dato che non può fluire una corrente continua in C_2 , nel quarzo o nell'ingresso dell'inverter): se consideriamo il legame tra ingresso e uscita imposto dall'andamento della caratteristica di trasferimento dell'inverter, concludiamo che il valor medio all'ingresso e all'uscita sarà pari alla metà della tensione di alimentazione, quindi il punto di lavoro sarà, come desiderato, proprio al centro della caratteristica.

In base al criterio di Barkhausen, per ottenere il funzionamento di un oscillatore è necessario che il guadagno attraverso l'anello di reazione abbia fase nulla (in modo da riportare in ingresso un segnale esattamente in fase con quello che lo ha prodotto) e modulo unitario. In realtà la seconda condizione è indispensabile soltanto se si vuole ottenere una forma d'onda in uscita sinusoidale, altrimenti è sufficiente che il modulo sia maggiore dell'unità. A rigore si dovrebbe parlare di oscillatori solo nel caso di circuiti che producono una forma d'onda sinusoidale (mentre gli altri sarebbero più propriamente generatori di forma d'onda), ma è ormai entrato nell'uso comune definire oscillatori anche circuiti come quello che stiamo esaminando, la cui forma d'onda di uscita è quadra. Dato che in questo caso il guadagno di anello è maggiore dell'unità, l'ampiezza viene limitata dal fenomeno della saturazione dell'inverter. Supponiamo che il quarzo e C_2 siano in risonanza con C_1 , per cui, alla frequenza fondamentale di funzionamento dell'oscillatore, non scorre corrente in R (un circuito risonante parallelo presenta un'impedenza infinita alla frequenza di risonanza). In tal caso la tensione V_1 (sempre alla frequenza fondamentale) su C_1 è uguale a quella V_u presente in uscita. Quindi possiamo calcolare la tensione V_2 ai capi del condensatore C_2 considerando il partitore tra il quarzo (con impedenza jX) e C_2 :

$$V_2 = \frac{V_u \frac{1}{j\omega C_2}}{\frac{1}{j\omega C_2} + jX} = \frac{V_u}{1 - \omega C_2 X}.$$

Per ottenere una fase totale nulla, lo sfasamento tra V_u e V_2 dovrà risultare di 180° , dato che l'inverter introduce il rimanente sfasamento di 180° . Dovrà quindi risultare $1 - \omega C_2 X < 0$, che può anche scriversi $\omega C_2 X > 1$, quindi X deve essere maggiore di zero e dunque induttiva. Pertanto, il circuito che stiamo considerando è tale da far funzionare il quarzo nella zona induttiva.

Imponiamo ora che la serie del quarzo e di C_2 (sono in serie, dato che l'ingresso dell'inverter non assorbe corrente) sia effettivamente in risonanza con il condensatore C_1 :

$$j\omega C_1 + \frac{1}{jX + \frac{1}{j\omega C_2}} = 0$$

Con qualche passaggio otteniamo

$$j\omega C_2 = j\omega^2 C_1 C_2 X - j\omega C_1$$

e quindi

$$\omega = \frac{1}{X} \frac{C_2 + C_1}{C_1 C_2}.$$

Ricordiamo che X è anch'essa funzione di ω , per cui dobbiamo sviluppare ulteriormente i calcoli, per ottenere un'espressione esplicita della pulsazione di funzionamento. Sostituiamo in questa equazione l'espressione che abbiamo precedentemente ottenuto per $X(\omega)$, ottenendo

$$\omega = -\omega C_p \frac{\omega^2 - \omega_p^2}{\omega^2 - \omega_s^2} \frac{C_1 + C_2}{C_1 C_2}.$$

Sviluppando i calcoli ricaviamo

$$\omega^2 + \frac{C_1 C_2}{(C_1 + C_2) C_p} \omega^2 = \frac{C_1 C_2}{(C_1 + C_2) C_p} \omega_s^2 + \omega_p^2.$$

Poniamo

$$k = \frac{C_1 C_2}{(C_1 + C_2) C_p},$$

ottenendo

$$\omega^2(1 + k) = k\omega_s^2 + \omega_p^2,$$

da cui, dividendo per $1 + k$ e sommando e sottraendo un termine del tipo

$$\frac{\omega_p^2 k}{1 + k},$$

ricaviamo

$$\begin{aligned} \omega^2 &= \frac{k}{1 + k} \omega_s^2 + \frac{\omega_p^2}{1 + k} + \frac{\omega_p^2 k}{1 + k} - \frac{\omega_p^2 k}{1 + k} \\ &= \omega_p^2 - \frac{k}{1 + k} (\omega_p^2 - \omega_s^2). \end{aligned}$$

Osserviamo che il termine

$$\frac{k}{1 + k} (\omega_p^2 - \omega_s^2)$$

rappresenta una frazione dell'intervallo $[\omega_s^2, \omega_p^2]$, dato che $k > 0$ e quindi $k/(k+1) < 1$. Ne consegue che $\omega_s < \omega < \omega_p$, per cui il quarzo opera effettivamente nella regione in cui la sua reattanza è induttiva.

Possiamo ora ricavare l'espressione di $k/(1+k)$ dalla definizione di k precedentemente data:

$$\frac{k}{1+k} = \frac{C_1 C_2}{C_1 C_2 + C_p(C_1 + C_2)},$$

da cui otteniamo

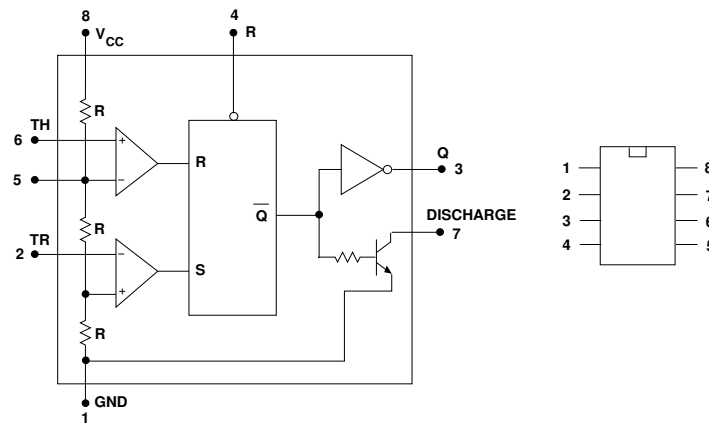
$$\omega^2 = \omega_p^2 + \frac{C_1 C_2}{C_1 C_2 + C_p(C_1 + C_2)} (\omega_s^2 - \omega_p^2).$$

È possibile sostituire uno dei condensatori con un condensatore variabile per ottenere una regolazione fine del valore della frequenza. Il campo di regolazione è comunque molto piccolo, perché anche una variazione piuttosto grande delle capacità sposta comunque di poco la frequenza di oscillazione (in genere meno di una parte per mille).

18.6 Applicazioni del circuito integrato NE555

Il circuito integrato NE555 contiene al suo interno dei comparatori di tensione e un flip-flop SR, oltre ad altri componenti come un transistor utile a scaricare capacità connesse esternamente. Lo NE555 ha un numero estremamente grande di applicazioni in circuiti analogico-digitali, in particolare nella realizzazione di timer, di multivibratori astabili, di formatori di impulsi.

Lo schema a blocchi interno è rappresentato nella figura seguente, dove è anche indicata la corrispondenza dei terminali con quelli del case plastico dual-in-line più comunemente utilizzato. Le tre resistenze R formano un partitore sul quale sono disponibili le tensioni $1/3 V_{CC}$ e $2/3 V_{CC}$.

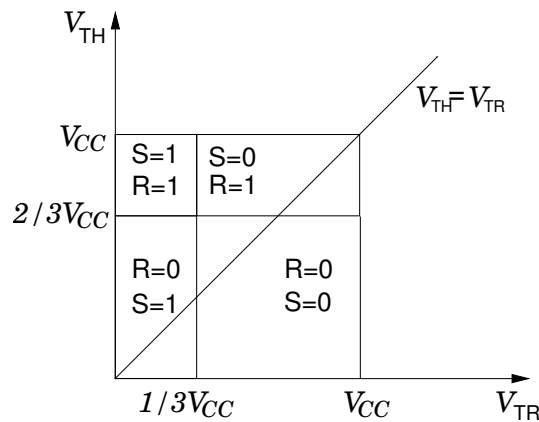


Se sull'ingresso TH (Threshold) è presente una tensione superiore a $2/3 V_{CC}$, viene attivato il terminale di reset del flip-flop; se invece è presente una tensione inferiore a $1/3 V_{CC}$ sull'ingresso TR (trigger), risulta attivato il terminale di set del flip-flop. Il flip-flop ha anche un altro terminale di reset, attivo allo stato basso, che è collegato al piedino 4. L'uscita negata del flip-flop viene portata al terminale di uscita esterno tramite un inverter in grado di erogare correnti fino a 200 mA e comanda anche un transistor open collector che viene portato in saturazione quando l'uscita esterna è allo stato basso.

L'attivazione del reset sul piedino 4 è prioritaria rispetto agli altri ingressi del flip-flop e tra gli altri due ingressi il set è prioritario rispetto al reset. La tensione di

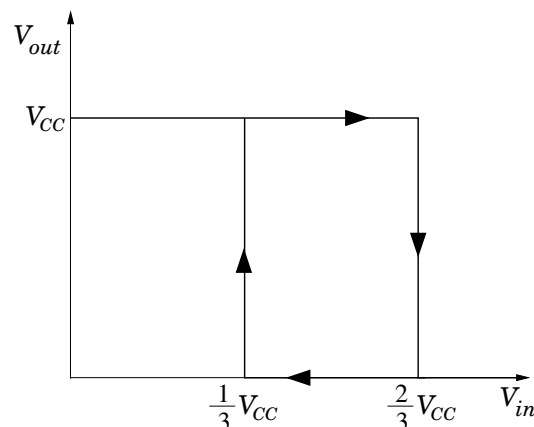
alimentazione V_{CC} può essere compresa tra 4.5 e 16 V e l'uscita, purché si scelga il corretto valore di V_{CC} , è compatibile con circuiti logici sia CMOS sia TTL.

Tracciamo un diagramma che rappresenti lo stato dell'uscita in funzione dello stato degli ingressi TH e TR.

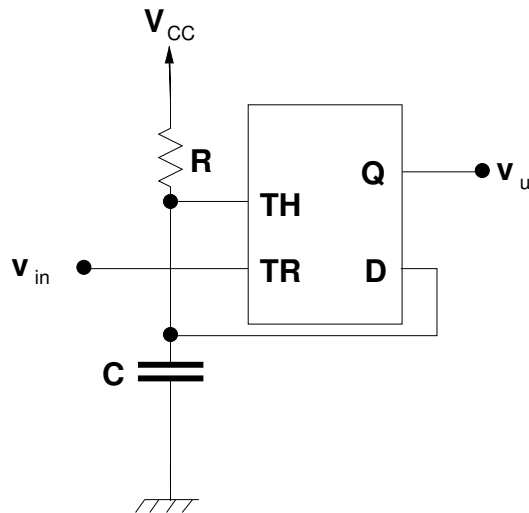


Per $V_{TR} < 1/3 V_{CC}$, il terminale di set è attivo e quindi l'uscita è sempre a livello alto, anche se $V_{TH} > 2/3 V_{CC}$, perché, come abbiamo già detto, il set ha priorità sul reset. Se invece V_{TR} supera $1/3 V_{CC}$, l'uscita rimarrà nello stato in cui si trovava precedentemente se $V_{TH} < 2/3 V_{CC}$ (dato che in tal caso né il set né il reset sono attivi) oppure sarà forzata al livello basso se $V_{TH} > 2/3 V_{CC}$.

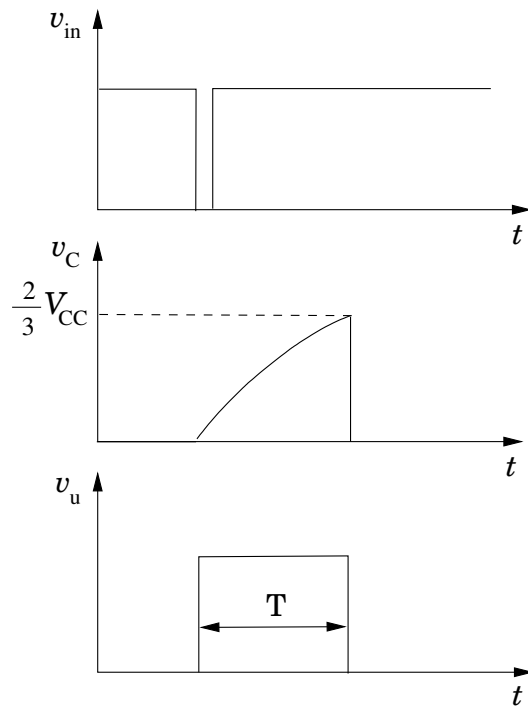
Prendiamo ora in esame il caso particolare consistente nel collegare insieme TH e TR: il comportamento è descritto dalla bisettrice del primo quadrante, indicata sul grafico precedente. Se partiamo da un valore basso di tensione di ingresso, ci troviamo nella regione ($S = 1, R = 0$), quindi l'uscita si trova al livello alto. Un ulteriore incremento della tensione di ingresso ci porta nella regione ($S = 0, R = 0$), per cui l'uscita rimane al livello alto, finché l'ingresso non supera la tensione di $2/3 V_{CC}$, e raggiungiamo la condizione ($S = 0, R = 1$), per cui l'uscita passa al livello basso. Tornando verso valori di tensione di ingresso bassi, attraversiamo di nuovo la regione ($S = 0, R = 0$), per cui l'uscita rimane al livello basso, finché non attraversiamo la soglia corrispondente a $1/3 V_{CC}$ e l'uscita torna al livello alto, dato che siamo nella regione ($S = 1, R = 0$). Abbiamo quindi ottenuto un trigger di Schmitt, la cui caratteristica di trasferimento è rappresentata nella figura seguente.



Vediamo ora come si può utilizzare lo NE555 per realizzare un timer, che sostanzialmente corrisponde a un monostabile (lo si definisce timer perché la durata dell'impulso di uscita può essere anche dell'ordine di qualche minuto).



A riposo, quando all'ingresso TR è presente una tensione superiore a $1/3 V_{CC}$, l'uscita si trova a livello basso, quindi il terminale di discharge mantiene il condensatore scarico e il terminale TH disattivato. La condizione descritta risulta pertanto consistente. Se applichiamo un impulso di valore inferiore a $1/3 V_{CC}$ sull'ingresso TR, l'uscita passa allo stato alto e il condensatore C comincia a caricarsi, come indicato nella figura seguente, dato che il terminale di discharge è ora flottante.



Quando la tensione sul condensatore raggiunge $2/3 V_{CC}$, avviene una nuova commutazione e l'uscita passa al livello basso, causando anche, tramite il terminale di discharge, la scarica del condensatore, riportando tutto nella condizione iniziale. Questo monostabile ha quindi il vantaggio, rispetto a quello a porte logiche visto in precedenza, di non richiedere alcun tempo di attesa prima che sia possibile fornire in ingresso un nuovo impulso, dopo che è terminato l'impulso di uscita.

Possiamo facilmente calcolare la durata dell'impulso di uscita: questa corrisponde al tempo impiegato dalla tensione sul condensatore per raggiungere $2/3 V_{CC}$:

$$\frac{2}{3}V_{CC} = V_{CC} \left(1 - e^{-T/RC}\right).$$

Quindi otteniamo

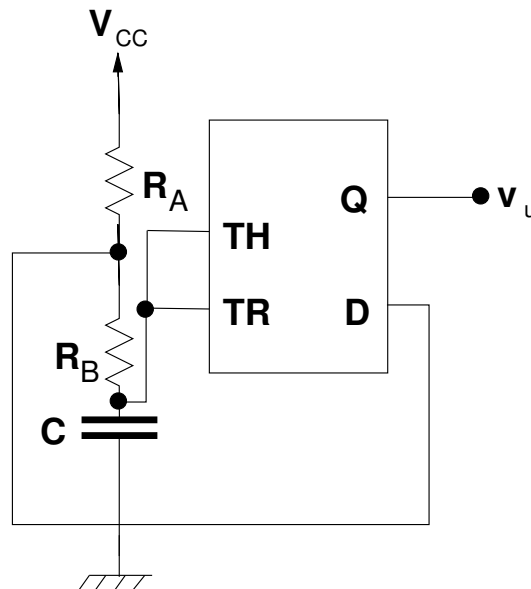
$$\frac{1}{3} = e^{-T/RC},$$

da cui

$$T = RC \ln 3.$$

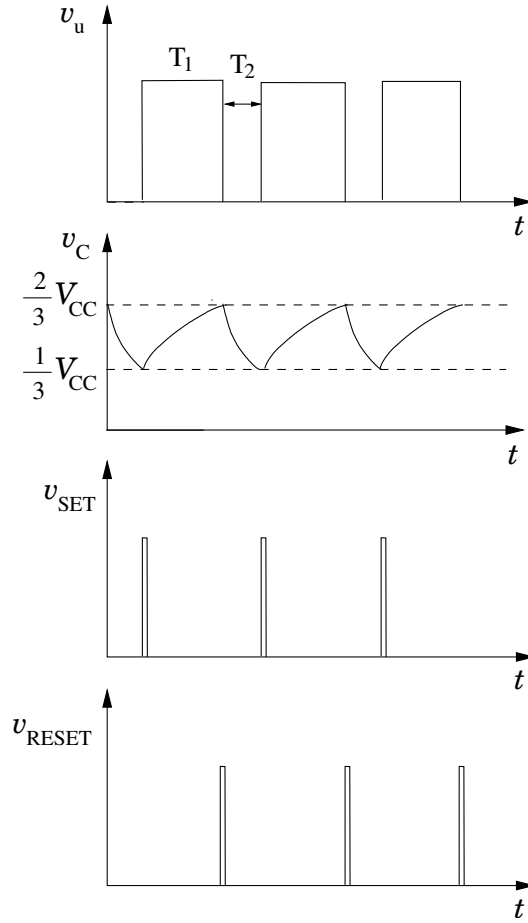
Possiamo ottenere tempi T anche piuttosto lunghi: se consideriamo di poter utilizzare resistenze R_C tra i 100Ω e qualche megaohm, con condensatori di qualche decina di microfarad si possono raggiungere tempi oltre il centinaio di secondi. Non si possono impiegare resistenze troppo piccole, altrimenti la corrente entrante nel terminale di discharge potrebbe raggiungere valori eccessivi né si possono utilizzare resistenze di valore troppo alto, perché le correnti parassite che fluiscono nell'integrato e nel condensatore (specialmente se si utilizzano condensatori elettrolitici) darebbero luogo a cadute di tensione eccessive.

Un'altra possibile applicazione dello NE555 consiste nella realizzazione di un multivibratore astabile, un possibile schema del quale è riportato di seguito.



Partiamo da una condizione nella quale l'uscita si trovi al livello basso e in cui quindi il terminale di discharge sia connesso a massa e determini la scarica di C tramite R_B : quando la tensione sul condensatore raggiunge $1/3 V_{CC}$, l'uscita passa al livello alto, D diventa flottante e il condensatore comincia a caricarsi tramite la serie di R_A e di R_B , fino a che la tensione ai suoi capi non raggiunge $2/3 V_{CC}$ e si verifica il passaggio dell'uscita al livello basso e l'inizio di una nuova scarica di C attraverso R_B . L'andamento delle tensioni è rappresentato nella figura che segue.

Calcoliamo il periodo della forma d'onda rettangolare ottenuta: il tempo T_1 per cui l'uscita rimane al livello alto corrisponde a quello di carica del condensatore da



$1/3 V_{CC}$ a $2/3 V_{CC}$ tramite $R_A + R_B$, quindi con costante di tempo $\tau_1 = C(R_A + R_B)$, per cui

$$\frac{2}{3}V_{CC} = V_{CC} + \left(\frac{1}{3}V_{CC} - V_{CC}\right)e^{-T_1/\tau_1}$$

e dunque

$$T_1 = C(R_A + R_B) \ln 2.$$

Il tempo T_2 corrisponde invece a quello di scarica, tramite R_B , da $2/3 V_{CC}$ a $1/3 V_{CC}$, con costante di tempo $\tau_2 = CR_B$:

$$\frac{1}{3}V_{CC} = \frac{2}{3}V_{CC}e^{-T_2/\tau_2},$$

pertanto

$$T_2 = CR_B \ln 2.$$

Il periodo della forma d'onda ottenuta risulta perciò

$$T = T_1 + T_2 = C(2R_B + R_A) \ln 2.$$

Notiamo che le tensioni sui terminali interni di set e di reset hanno una forma impulsiva, con durata di ciascun impulso molto breve. Il motivo della durata molto breve è semplice da capire: consideriamo per esempio il caso del terminale di reset. Quando la tensione sul condensatore raggiunge il valore $2/3 V_{CC}$, il terminale di reset viene portato a livello alto e avviene la commutazione dell'uscita al livello basso; a questo punto il terminale di discharge entra in azione, iniziando immediatamente a scaricare il condensatore, per cui la tensione scende subito al di sotto di $2/3 V_{CC}$ e il terminale di reset torna al livello logico basso.

19. Logica programmabile

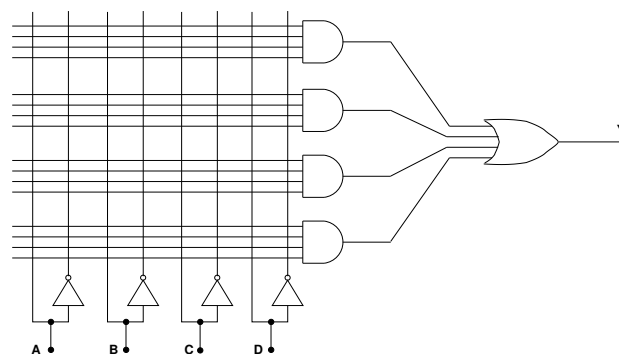
19.1 Introduzione

La realizzazione di circuiti logici di una certa complessità a partire da circuiti integrati SSI (Small Scale Integration), che contengono al più una decina di porte logiche, conduce a schede di dimensioni eccessive e a costi di realizzazione troppo elevati. In tali situazioni si cerca quindi di ricorrere a circuiti integrati di tipo diverso, contenenti un numero elevato di porte logiche: è chiaro che tali circuiti integrati devono essere concepiti in modo che sia possibile “programmarli” sul campo o, al limite, in sede di fabbricazione, in modo da ottenere esattamente le funzioni logiche desiderate. Un modo molto semplice di ottenere una rete combinatoria programmabile consiste nell’uso di una PROM: i terminali degli indirizzi corrispondono agli ingressi e il dato di uscita corrisponde al valore immagazzinato nella cella individuata da ciascuna combinazione delle variabili di ingresso. La funzione logica svolta da una rete combinatoria così realizzata è pertanto completamente programmabile. Il problema dell’implementazione di una rete combinatoria tramite una PROM sta nel fatto che le dimensioni di quest’ultima crescono piuttosto rapidamente con il numero delle variabili di ingresso, per cui una tale soluzione è valida solo se si ha un numero di ingressi piuttosto limitato.

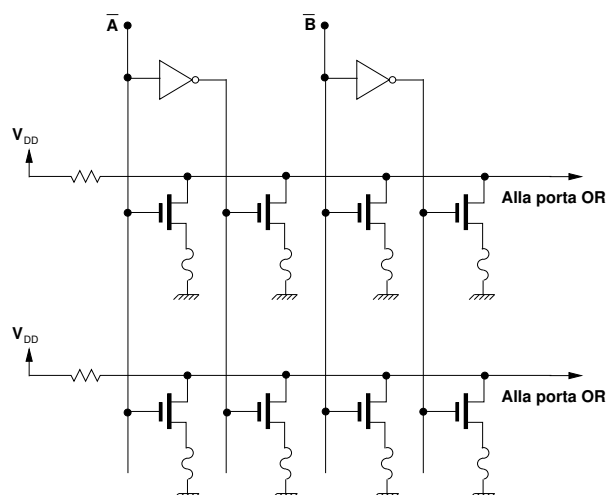
Sono stati pertanto studiati approcci di tipo diverso, che contengono anche elementi di logica sequenziale, e che verranno brevemente trattati in questo capitolo. L’utilizzo delle diverse soluzioni dipende sia dalla complessità del circuito che si intende realizzare sia dal numero di pezzi che si prevede di produrre, dato che il costo unitario può essere fortemente dipendente dal volume della produzione.

19.2 Programmable Array Logic (PAL)

L’approccio della Programmable Array Logic si basa sulla rappresentazione della funzione logica in termini di somma di prodotti, quindi su una struttura costituita da porte AND ai cui ingressi vengono presentate le variabili complementate o non complementate e le cui uscite sono connesse a una porta OR che svolge la somma dei prodotti. La struttura base di una PAL è riportata nella figura seguente.



La struttura a matrice che precede gli AND si definisce “piano AND” e rappresenta la parte più importante della PAL, quella sulla quale si svolge la programmazione, consistente nel creare le connessioni desiderate tra le linee verticali e quelle orizzontali. Vediamo quale può essere un’implementazione del piano AND di una PAL in termini circuitali. Lo schema riportato di seguito implementa la funzione logica AND se le



variabili di ingresso complementate sono scambiate, come indicato, con quelle non complementate.

Infatti la struttura è quella di un NOR, in cui, per esempio, abbiamo $Y = \overline{A + B + C} = \overline{A} \cdot \overline{B} \cdot \overline{C}$ e, quindi, negando ciascuna delle variabili di ingresso, si ottiene un AND. I componenti indicati sui source sono fusibili, che possono essere interrotti in fase di programmazione (eliminando quindi la corrispondente intersezione tra linee verticali e orizzontali). Una PAL riprogrammabile potrebbe essere ottenuta sostituendo i MOS indicati con dei MOS con floating gate come quelli delle EPROM, consentendo quindi la cancellazione tramite esposizione agli ultravioletti, o come quelli delle E²PROM, permettendo dunque la cancellazione elettrica.

Se il numero delle variabili di ingresso è molto elevato, la dimensione del piano AND cresce in modo eccessivo, anche se le intersezioni tra le linee verticali e quelle orizzontali sono in effetti in numero ridotto. Lo spazio sul chip non viene quindi sfruttato efficientemente da PAL di grosse dimensioni. Per tale motivo, per circuiti a molti ingressi, si ricorre a soluzioni di tipo diverso, che vengono discusse nel paragrafo seguente.

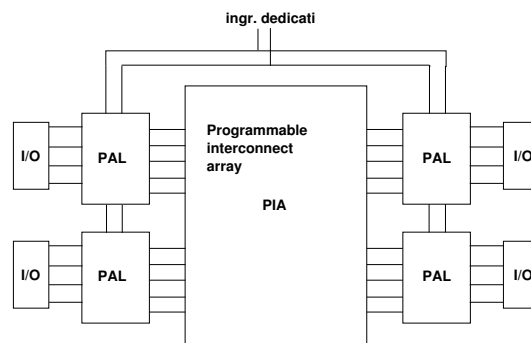
19.3 Programmable Logic Device (PLD)

Un PLD è costituito da un certo numero di PAL interconnesse tra loro da una matrice di connessioni bidirezionali detta PIA (Programmable Interconnect Array), che può essere programmata in modo da ottenere una connessione tra due qualsiasi punti del sistema. La PIA è una struttura molto complessa, dato che deve garantire qualunque interconnessione, mantenendo però sempre lo stesso ritardo. Spesso il suo progetto costituisce un segreto industriale gelosamente custodito.

Uno schema a blocchi di un PLD è riportato nella figura che segue, dove sono anche rappresentati i pad di I/O, che consentono l'interfacciamento con il mondo esterno e possono contenere della logica sequenziale.

Sono indicati anche dei collegamenti che dall'esterno raggiungono solo le PAL, per variabili che non necessitano di transitare dalla PIA, come, per esempio, dei segnali di reset globali. Notiamo che in questa struttura sono programmabili sia le PAL sia la PIA e, in parte, i pad di I/O. I PLD possono sostituire fino a un centinaio di integrati SSI (Small Scale Integration), mentre le PAL possono sostituirne intorno a una decina.

Per circuiti logici con complessità equivalente maggiore di qualche centinaio di integrati SSI è necessario ricorrere a soluzioni con progetto ad hoc, che consistono

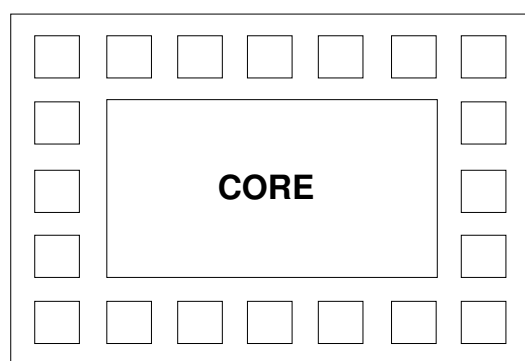


nei cosiddetti circuiti ASIC.

19.4 Application Specific Integrated Circuit (ASIC)

I circuiti ASIC vengono progettati specificamente per una determinata applicazione, quindi il layout può essere ottimizzato e il numero di componenti ridotto a quello effettivamente necessario. I vantaggi di un tale approccio sono evidenti: il consumo di potenza viene ridotto e la velocità risulta incrementata, grazie all'ottimizzazione del progetto; il circuito ha un ingombro minimo, essendo compreso tutto all'interno di un singolo chip; l'affidabilità risulta notevolmente incrementata, per l'assenza di saldature; infine l'effettiva struttura interna non è visibile, se non con mezzi molto sofisticati, e quindi è difficile che il progetto venga riprodotto da un concorrente. Gli svantaggi degli ASIC sono rappresentati dall'elevato costo di progetto, che richiede l'uso di tool CAD (Computer Aided Design) sofisticati e molto costosi, e dalla tecnologia richiesta per l'effettiva realizzazione del circuito integrato, che è disponibile solo presso poche silicon foundry nel mondo. Anche la fase di test è complessa, dato che, a meno di usare tecniche complesse e costose come la microscopia elettronica a contrasto di tensione, non è possibile avere accesso ai nodi interni del circuito. La preparazione di sequenze di dati di ingresso capaci di verificare al meglio la funzionalità di un circuito ASIC (creazione dei vettori di test) è una specialità a sé stante, che richiede notevoli sforzi anche a livello di ricerca. Un altro svantaggio è rappresentato dal "turn around time" (il tempo che trascorre dal momento in cui si inizia il progetto al momento in cui il chip è effettivamente disponibile) piuttosto lungo a causa delle interazioni, anche abbastanza complesse, che devono aversi tra progettisti e fonderia.

In un ASIC si fa una distinzione tra la parte che svolge tutte le funzioni logiche caratteristiche del particolare progetto (core) e i pad di I/O, che contengono i componenti necessari per interfacciare il core con l'esterno. La disposizione di tali parti sul chip è illustrata nella figura che segue, con i pad di I/O che formano una corona intorno al core.

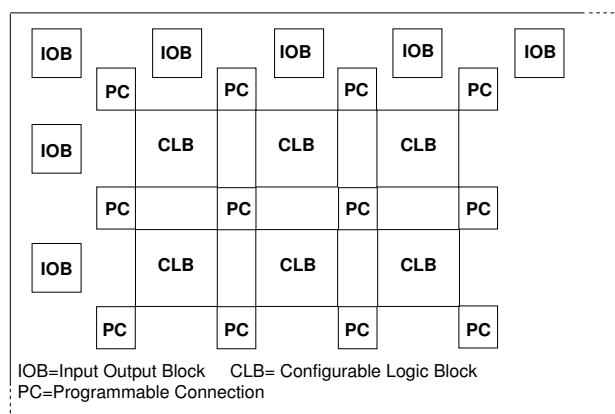


Gli ASIC si possono distinguere in due categorie principali: i Full-Custom e i Semi-Custom. Nel caso full-custom tutto il core deve essere progettato, mentre in quello del semi-custom gli elementi base del core sono già definiti dalla fonderia e il progetto riguarda sostanzialmente le connessioni tra tali elementi.

Ulteriori distinzioni si possono fare tra i vari tipi di ASIC semi-custom, che si distinguono nelle categorie principali dei "Gate Arrays", "Sea of Gates" e "Standard Cells". I "Gate Arrays" sono costituiti da file di transistori PMOS e NMOS i quali possono essere interconnessi tra loro in maniera selezionabile dal progettista. I "Sea of Gates" sono costituiti da un insieme di molti transistori distribuiti in modo meno regolare che nei "Gate Arrays". Infine l'approccio "Standard Cells" consiste in un core formato da celle preprogettate che svolgono funzioni logiche elementari e in questo caso il progettista deve solo definire la loro interconnessione.

19.5 Field Programmable Gate Arrays (FPGA)

Come abbiamo visto, la soluzione basata su circuiti ASIC diventa economicamente conveniente solo nel caso in cui siano previste produzioni di grande volume (in modo da poter ammortizzare i costi di progetto e di realizzazione delle maschere). Un approccio che consente di realizzare circuiti digitali anche complessi su un singolo chip senza dover ricorrere a tecnologie custom è rappresentato dalle FPGA (Field Programmable Gate Arrays), strutture contenenti celle elementari (che realizzano funzioni sia combinatorie sia sequenziali), la cui interconnessione viene definita tramite una procedura di programmazione. L'effettiva struttura interna delle FPGA dipende dal costruttore e può variare anche notevolmente dall'uno all'altro. Per esempio, le FPGA ACTEL hanno una struttura che ricorda quella "Standard Cells" degli ASIC e sono costituite da blocchi logici di tipo sequenziale o combinatorio allineati secondo una sequenza opportuna e tra i quali è possibile programmare le interconnessioni. Le FPGA Xilinx, molto utilizzate attualmente, hanno una struttura diversa, rappresentata schematicamente nella figura che segue, in cui è presente una corona esterna di pad di I/O programmabili detti IOB (Input Output Block).



All'interno della corona si ha una matrice di blocchi logici programmabili detti CLB (Configurable Logic Block) che possono essere collegati tra loro e con i pad di I/O tramite una rete di connessioni programmabile, ciascuna delle quali è controllata da un bit di configurazione. La complessità della struttura è notevole, così come quella della procedura di programmazione, che coinvolge sia i pad di I/O sia i CLB sia le connessioni programmabili.

Oltre che per la realizzazione di piccole serie, le FPGA si utilizzano anche per la realizzazione di prototipi di ASIC, data la facilità di programmazione e di riprogrammazione e la rapidità con cui il progetto può essere verificato, data l'immediata disponibilità sul mercato delle FPGA stesse.

Un esempio particolare di applicazione delle FPGA è rappresentato dal TERAMAC (sviluppato nell'ambito di una collaborazione tra Hewlett-Packard e University of California at Los Angeles), un calcolatore riconfigurabile basato su una struttura a matrice di FPGA scelte in maniera che alcune di esse siano sicuramente difettose e connesse con un cablaggio non affidabile al 100%. Un software appositamente realizzato va a esaminare la struttura hardware del TERAMAC, individuando i componenti non funzionanti e configura le FPGA in modo da realizzare un computer operante in modo corretto. Se durante il funzionamento intervengono altri guasti oppure vengono addirittura tolte delle schede, si può far girare di nuovo il software in questione e ottenere ancora un computer funzionante, anche se con prestazioni un po' inferiori. Lo scopo del TERAMAC è di tipo dimostrativo, per far vedere come sia possibile, purché si preveda un margine sufficiente di ridondanza, realizzare sistemi correttamente funzionanti anche a partire da componenti che siano in parte difettosi.