

M. Sampietro, G. Bertuccio, C. Fiorini

# FONDAMENTI DI ELETTRONICA

Tracce delle lezioni  
ad uso esclusivo degli studenti

## Lezione n. 1 : L'elettronica - Storia di un secolo

Per quasi 50 anni dall'inizio del secolo scorso fino a dopo la 2<sup>a</sup> guerra mondiale l'industria elettronica è stata dominata dalla tecnologia del *tubo elettronico a vuoto*.

La “**valvola**” come veniva gergalmente chiamata, permetteva di:

- generare un sufficiente flusso di elettroni, e quindi corrente elettrica;
- controllarne il suo valore mediante il potenziale applicato ad un sistema di griglie metalliche.



Le valvole tuttavia avevano grandi limitazioni, soprattutto se guardate con gli occhi di adesso; erano:

- fragili
- voluminose
- poco affidabili
- consumavano molto
- producevano molto calore

Fu solo nel 1947, con l'invenzione del **transistore bipolare**, che il “problema” delle valvole arrivò ad una soluzione.

La sola disponibilità delle valvole non ha comunque impedito, nella prima metà del secolo, lo sviluppo importante dell'elettronica. Anzi, enormi progressi scientifici sono stati compiuti in quel periodo e moltissime realizzazioni elettroniche che oggi consideriamo assolutamente normali videro la luce allora.

Non bisogna dimenticare che **Guglielmo Marconi** iniziò le comunicazioni radio ben prima del 1900 e che nei primi decenni del secolo la tecnologia dei sistemi di telecomunicazione si sviluppò al punto che le comunicazioni telefoniche, la radio e la televisione erano ormai oggetti di massa già ben prima della seconda guerra mondiale.

- 1895 Guglielmo Marconi
- 1900 Comunicazioni Telefoniche
- 1906 Radio
- 1930 Televisione

Certo, gli apparati risentivano delle limitazioni imposte dalla presenza delle valvole, quali consumi o affidabilità, e quanto ad ingombro ci si doveva accontentare. Nella foto è mostrata la *prima radio portatile* mai realizzata, donata da **Howard Armstrong** alla moglie durante il loro viaggio di nozze a Palm Beach nel dicembre del 1923. Il dono era frutto del suo ingegno. Armstrong, uno dei più grandi innovatori di quel periodo, inventò di tutto: le tecniche di rigenerazione del segnale che consentirono di sviluppare i trasmettitori radio, inventò i circuiti a supereterodina per selettivamente ricevere ed amplificare segnali radio debolissimi, inventò la radio a modulazione di frequenza (FM) che aprì la strada delle trasmissioni radio ad alta fedeltà e sviluppò un'infinità



di tecniche radar per il riconoscimento di oggetti.

Come accennato, fu solo nel **1947, con l'invenzione del transistore bipolare**, che il "problema" delle valvole arrivò ad una soluzione. Gli anni a cavallo della seconda guerra mondiale furono infatti anni importanti per il settore dei dispositivi elettronici.

**L'esigenza di sostituire gli interruttori meccanici delle centraline telefoniche con interruttori comandati più piccoli ed affidabili aveva fortemente sollecitato la ricerca scientifica sui materiali semiconduttori** e sulla possibilità che essi offrivano di fare circolare corrente in modo controllato all'interno di un minuscolo cristallo piuttosto che nel vuoto come nelle valvole. La ricerca culminò proprio con l'invenzione e la verifica del funzionamento del primo transistore a semiconduttore della storia, nel dicembre del 1947.

Il primo transistore era costituito da un cristallo di germanio drogato N (Base) e da due contatti in fili d'oro molto vicini tra di loro (Emettitore e Collettore), visibili nella foto a destra.

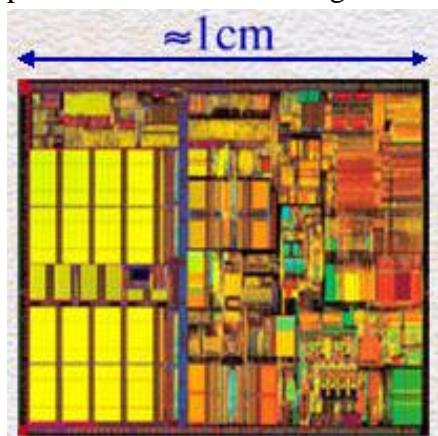
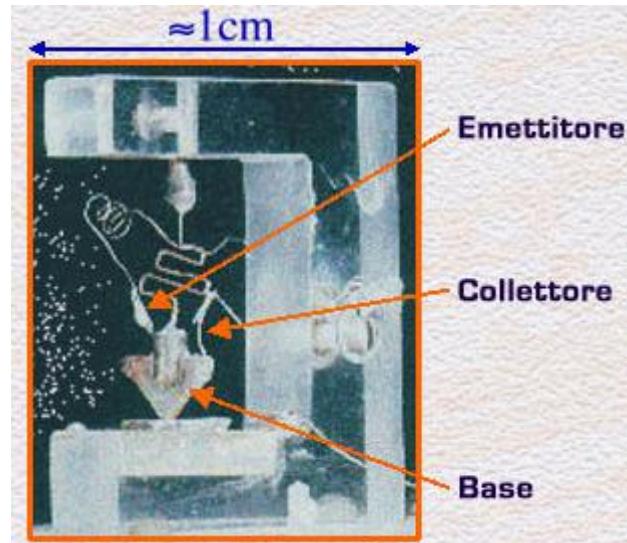
Il dispositivo dimostrava la possibilità di amplificare la potenza di un segnale elettrico a scapito della potenza continua fornitagli da una semplice batteria. A questo straordinario risultato, forse il più

importante di tutto il secolo nel campo dell'elettronica, giunsero con un lavoro congiunto durato anni i fisici **William Shockley, John Bardeen e Walter Brattain**. Piccolo, affidabile (non si rompeva come le valvole che dovevano essere continuamente sostituite), parco nei consumi di potenza elettrica e nello scaldarsi, il transistore da allora ha invaso ogni ambito in cui un circuito elettronico possa avere un senso! Da quel primo esemplare ad oggi, il transistore ha avuto uno sviluppo tecnologico enorme che lo ha reso sempre più efficiente nelle sue funzioni e sempre più piccolo nelle dimensioni: oggi, nello spazio di  $1\text{cm}^2$  più o meno occupato dal primo esemplare, si riescono a stipare più di un milione di transistori ! come rappresentato in figura a sinistra.

L'invenzione del transistore, così come molte altre fatte nei decenni precedenti e molte che seguiranno fino ai giorni nostri, è avvenuta in un laboratorio di ricerca particolarmente famoso che ha avuto il pregio ed il prestigio di raccogliere per quasi tutto il secolo gli scienziati più dotati nei campi della fisica, della chimica e delle scienze ingegneristiche. Questi laboratori, i "**Bell Laboratories**", sono entrati nella storia della scienza e della tecnologia come **fucina di Premi Nobel**: in essi hanno lavorato moltissimi premi nobel, tra cui appunto William Shockley, John Bardeen e Walter Brattain insigniti del Premio Nobel per la **Fisica nel 1956**. Nei laboratori Bell lavorò Shannon quando formulò la teoria dell'informazione, lì venne concepito il laser, il sistema operativo UNIX ed il linguaggio di programmazione C++.

- 1948 teoria dell'informazione di **C.E.Shannon**
- 1958 LASER
- 1969 sistema operativo **UNIX** (**K.Thompson e D.Ritchie**)
- 1983 linguaggio **C++** (**B.Stroustrup**)

Tra i tanti scienziati legati ai laboratori Bell, ha avuto un'importanza particolare ai nostri fini **Harold S. Black**, ingegnere 29enne quando entra nella nostra storia. L'oggetto della ricerca erano allora i sistemi per comunicazione telefonica su grande distanza, con l'obiettivo di arrivare ad apparati che permetessero un collegamento efficiente tra le due coste degli Stati Uniti e tra gli Stati Uniti e



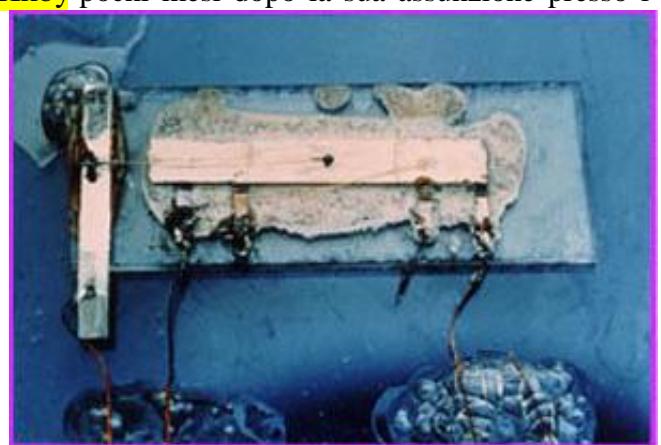
l'Europa. Le difficoltà che si dovevano affrontare erano soprattutto al non sapere come **progettare amplificatori sufficientemente stabili che non determinassero distorsioni eccessive dei segnali**. Le variazioni delle caratteristiche degli apparati, per effetto della temperatura o dell'invecchiamento, determinavano infatti un continuo cambiamento delle prestazioni degli amplificatori, in particolare del loro guadagno.

L'idea nuova che rivoluzionò e risolse il problema fu la **RETROAZIONE NEGATIVA** e venne ad Harold S. Black la mattina del 2 agosto 1927 mentre attraversava il fiume Hudson per recarsi al lavoro a Manhattan. Egli schizzò su di una pagina del New York Times il diagramma di un circuito reazionato negativamente e ne ricavò le proprietà fondamentali.



Quegli appunti riassumevano l'idea che sia la controllabilità dell'amplificazione che le distorsioni del segnale amplificato potevano essere estremamente migliorate se il segnale all'uscita del circuito veniva riportato in ingresso e sommato in controfase con il segnale applicato. Quando H. S. Black verificò sperimentalmente le caratteristiche del primo amplificatore reazionato negativamente della storia, misurò una **riduzione della distorsione di un fattore 100.000** su segnali di ingresso compresi tra 4 e 45 kHz ! Oggigiorno, praticamente tutti i circuiti elettronici di classe sono circuiti reazionati.

Ritornando al **transistore**, esso ovviamente **stimolò la creazione di circuiti sempre più complessi**, impieganti anche parecchie centinaia di dispositivi, non solo transistori ma anche diodi, capacità e resistenze. Il **PROBLEMA** divenne allora **come connettere tra loro questi elementi e come ridurre il costo che il lavoro manuale di saldatura aveva, oltre che come rendere questa fase di montaggio affidabile**. La **SOLUZIONE** fu proposta da **Jack Kilby** pochi mesi dopo la sua assunzione presso i laboratori di ricerca della Texas Instruments: mettere sullo stesso cristallo di semiconduttore dove già si riusciva a costruire il transistore anche gli altri componenti elettronici e le relative connessioni metalliche. Dimostrò che l'idea era praticabile, nel settembre del 1958 realizzando il primo circuito integrato della storia con un unico cristallo di germanio su cui erano presenti un transistore, un condensatore ed un resistore, opportunamente collegati tra loro, (vedi figura a destra). Il brevetto che ne seguì sancì la nascita dell'elettronica integrata come la conosciamo oggi e portò alla sua azienda grandissime soddisfazioni economiche !



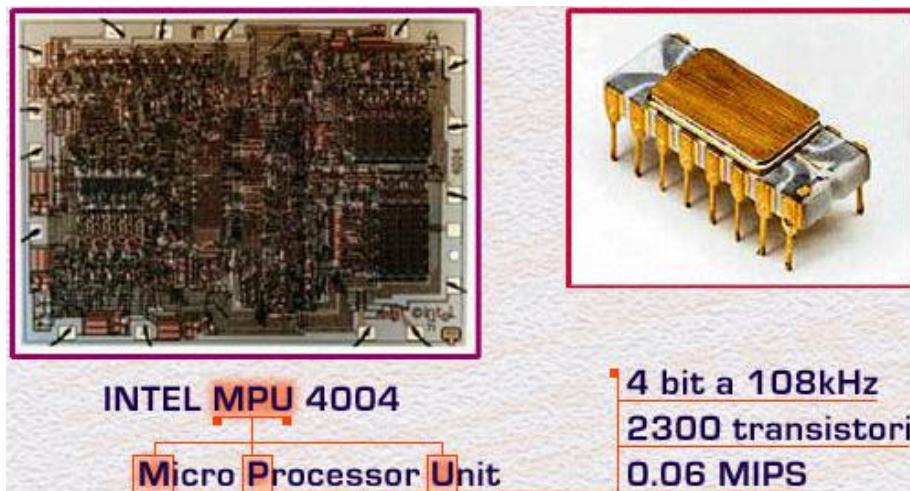
Si dovette comunque aspettare molti anni prima che i circuiti integrati arrivassero al grande pubblico. Il primo prodotto commerciale utilizzante circuiti integrati è stato una calcolatrice tascabile prodotta dalla Texas Instruments nel 1967 e progettata proprio da Kilby (oltre che da **Jerry Merryman** e **James Van Tassel**). Nelle pur minuscole dimensioni per allora di 10x15cm era comunque racchiusa una

potenza di calcolo che aveva uguali solo in grosse macchine non utilizzanti circuiti integrati. Jack Kilby è stato onorato del premio Nobel nel 2000.

Nel **1971** un'azienda giapponese che voleva produrre una nuova serie di calcolatori scientifici ad alte prestazioni, senza saperlo sollecitò un altro importante passo avanti nel campo dell'elettronica: il **microprocessore**. Alcuni ricercatori di una piccola azienda americana, la **INTEL**, decisero infatti, per soddisfare le specifiche imposte dall'azienda giapponese, di dare corpo all'idea di riunire nello stesso chip le funzioni di calcolo, di memorizzazione e di controllo della tastiera e del video che fino ad allora venivano svolte da circuiti distinti.

Riunendo queste funzioni non solo realizzarono il prodotto richiesto ma si trovarono tra le mani un oggetto di una flessibilità inaspettata e di una potenza di calcolo paragonabile perfino ai calcolatori di grandi dimensioni allora esistenti (cfr figura a destra).

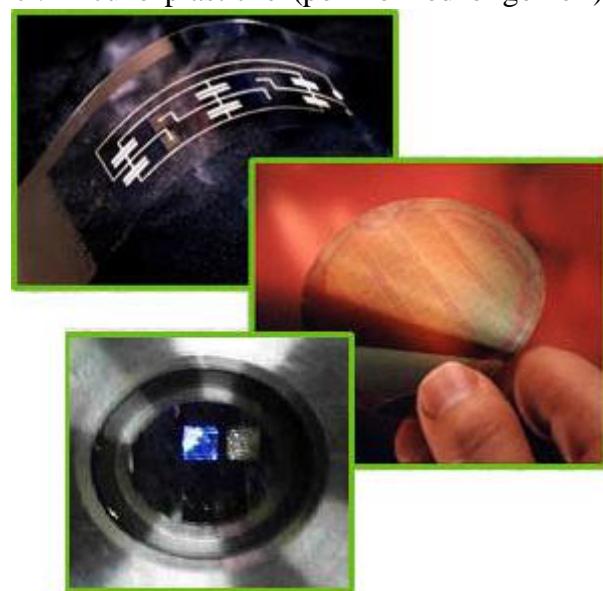
Il dispositivo venne chiamato **INTEL MPU 4004**, dove con MPU si volle sottolineare che si trattava di una Micro-Processor Unit. Funzionava su 4 bit a 108 kHz, conteneva 2300 transistori ed aveva una velocità stimabile in 0.06 MIPS. Fu il primo microprocessore prodotto industrialmente ed alla sua messa a punto partecipò anche un ingegnere italiano, **Federico Faggin**. Quella prima realizzazione ha aperto la strada ad un settore che ha avuto uno sviluppo enorme. Di sviluppo in sviluppo si è arrivati oggi a microprocessori che hanno quasi 50 milioni di transistori e possono essere fatti funzionare a frequenze ben superiori al GHz.



Altri materiali si stanno affacciando al mondo dell'elettronica. Dall'inizio degli anni novanta si è concretizzata la possibilità di usare materiali semiconduttori organici, (polimeri, oligomeri o molecole) per realizzare dispositivi e circuiti elettronici. Alcune plastiche (polimeri ed oligomeri) infatti hanno dimostrato di comportarsi come, ed in alcuni casi meglio, del tradizionale silicio. Oltre a transistori e, per ora, semplici circuiti, sono già stati ottenuti dispositivi emettitori di luce su qualunque colore del visibile. Le potenzialità di questi materiali plasticci sono molto elevate, dettate dalle doti caratteristiche di:

- flessibilità meccanica
- realizzabilità su grandissime superfici
- emissione di luce
- costi più bassi.

A riprova dell'importanza del settore e degli sviluppi scientifici avutisi in quest'ultimo decennio, nel 2000 il Premio Nobel per la chimica è stato assegnato a **A.J.Heeger** dell'Università americana di Santa Barbara che, tra i tanti, ha più approfondito le proprietà dei materiali semiconduttori organici per impieghi elettronici.



Durante tutto il secolo molte persone hanno contribuito, con la loro passione ed il loro ingegno, allo sviluppo delle conoscenze scientifiche. Per alcuni di essi c'è stato l'onore di essere insigniti di qualche importante premio a riconoscimento dei loro meriti e delle loro fatiche. Il più noto e, credo, il più ambito, è considerato il premio Nobel, assegnato ogni anno all'autore di una scoperta dimostratasi importante ed innovativa nei campi della fisica, della chimica e della medicina, per limitarci ai campi che hanno una attinenza con il nostro argomento.

Se non lo avete mai fatto, visitate il sito ufficiale del premio (<http://www.nobel.se>) e sfogliatelo alla ricerca di quelle scoperte, anche in campi in un primo momento diversi da quello dell'elettronica, che poi si sono rivelate decisive proprio per lo sviluppo della nostra materia.

## Lezione n. 2 : Elementi di base di elettronica

### Partizioni di tensione e di corrente

L'elettronica è quella disciplina che utilizza il moto dell'elettrone (**mezzo**) per trasmettere, memorizzare o acquisire dal mondo esterno informazione (**scopo**). L'elettronica ha quindi uno scopo preciso, l'**elaborazione dell'informazione**, realizzato con un mezzo ben definito, l'elettrone o più in generale la carica elettrica.

Anche l'**elettrotecnica**, come sapete, usa gli elettroni ma lo fa per uno scopo completamente diverso, quello di **trasferire energia**. La differenza tra le due finalità si manifesta anche negli ordini di grandezza delle tensioni o delle correnti in gioco: correnti di parecchi ampere e tensioni anche di migliaia di volt in elettrotecnica, correnti inferiori al milliampera e tensioni mai superiori a qualche decina di volt in elettronica, perché comunque sufficienti a trasmettere informazioni con precisione.

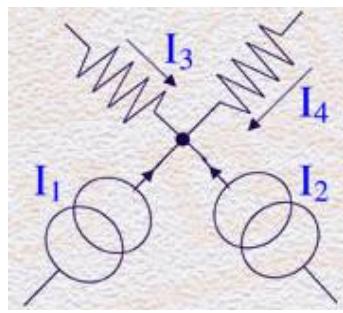
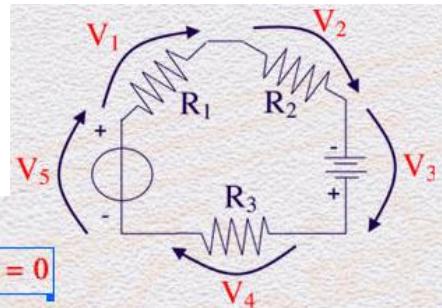
<b>Elettrotecnica:</b>	$1A < I < 1000A$	$10V < V < 10000V$
<b>Elettronica:</b>	$1fA < I < 10mA$	$0 < V < 20V$

Nonostante le differenze, alcuni strumenti tecnici di analisi dei circuiti elettronici sono identici. In questa lezione li richiamiamo alla memoria e li adatteremo al linguaggio dell'elettronica.

*L'analisi di un circuito elettronico a parametri concentrati si fonda sull'applicazione delle Leggi di Kirchhoff. La legge di Kirchhoff alle maglie dice che la somma algebrica delle tensioni elettriche misurate ordinatamente tra i nodi di una maglia di un circuito è nulla.*

**Legge delle maglie**

$$V_1 + V_2 + V_3 + V_4 + V_5 = 0$$



La legge di Kirchhoff ai nodi dice che la somma algebrica di tutte le correnti che fluiscono nei conduttori che afferiscono ad un nodo di un circuito è nulla.

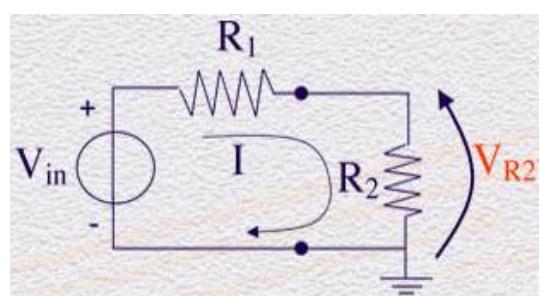
**Legge ai nodi**

$$I_1 + I_2 + I_3 + I_4 = 0$$

Esse sono di fondamentale importanza e sono redatte come **caso particolare delle più generali equazioni di Maxwell**: la prima esprime l'irrotationalità del campo elettrico, la seconda la solenoidalità della corrente nei conduttori.

Quale prima applicazione, si pensi al semplice circuito costituito da un generatore di tensione  $V_{in}$  e da due resistenze  $R_1$  e  $R_2$  collegati a formare una maglia chiusa, e si voglia calcolare la frazione di tensione  $V_{in}$  che si sviluppa ai capi di una delle due resistenze, ad esempio  $R_2$ .

Per la legge alle maglie, la tensione che cade ai capi della somma delle due resistenze è proprio  $V_{in}$ , cosicché la corrente circolante in esse sarà, in base alla Legge di Ohm:



**Corrente circolante**

$$I = \frac{V_{in}}{R_1 + R_2}$$

La corrente  $I$  scorre necessariamente in ognuno dei componenti presenti lungo la maglia e quindi anche in  $R_2$ , determinando una caduta di potenziale ai suoi capi pari a:

Ci si abitui a notare che se  $R_2$  è grande rispetto a  $R_1$  (ad esempio  $R_2=1k$  e  $R_1=50$ ) la frazione di tensione  $V_{in}$  che si ritrova ai capi di  $R_2$  è alta, ad esempio 95mV rispetto - ad esempio - ai 100mV di  $V_{in}$ .

Se invece  $R_2$  è piccola rispetto a  $R_1$  (ad esempio  $R_2=1k$  e  $R_1=12k$ ) allora la frazione di tensione ai capi di  $R_2$  è piccola, solo 7.7mV rispetto ai 100mV di  $V_{in}$  nel nostro esempio:

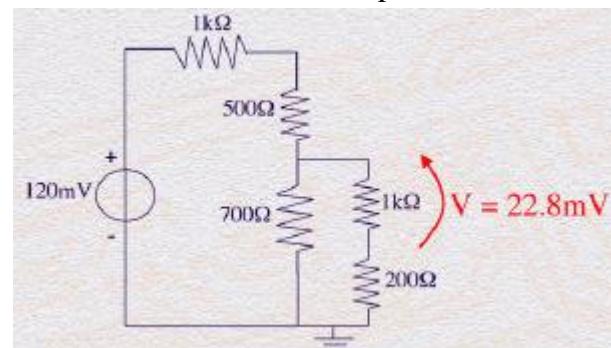
**Esempio 1:**  $R_1 = 50\Omega$ ,  $R_2 = 1k\Omega$ ,  $V_{in} = 100mV \rightarrow V_{R2} = 95mV$

**Esempio 2:**  $R_1 = 12k\Omega$ ,  $R_2 = 1k\Omega$ ,  $V_{in} = 100mV \rightarrow V_{R2} = 7.7mV$

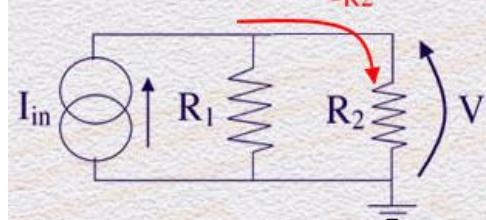
Poiché una tale configurazione circuitale definisce una partizione della iniziale tensione  $V_{in}$  disponibile, essa viene generalmente chiamata "*partitore di tensione*".

### Esempio:

Applicando le leggi di Kirchhoff e la legge di Ohm, verificare che la tensione ai capi della resistenza da 1k, nel circuito in basso a destra, è di 22.8mV.



Se il circuito contiene un generatore di corrente  $I_{in}$  in parallelo a dei componenti passivi, ad esempio due resistenze, si può voler calcolare la frazione di corrente  $I_{in}$  che circola in una delle due resistenze, ad esempio  $R_2$ . Poiché la corrente  $I_{in}$ , circolando sicuramente nel parallelo delle resistenze, produce una tensione ai loro capi pari a:



se ne deduce che la frazione che scorre in  $R_2$  sarà pari a:

$$I_{R2} = \frac{V}{R_2} = I_{in} \frac{R_1}{R_1 + R_2}$$

$$V = I_{in} \frac{R_1 \cdot R_2}{R_1 + R_2}$$

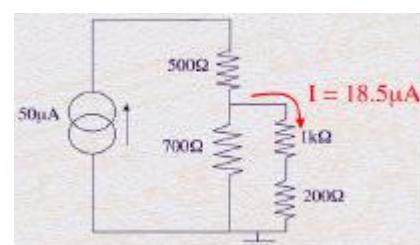
Anche in questa situazione di partizione di corrente, è utile abituarsi a notare che la corrente tende a ripartirsi maggiormente nella resistenza di minor valore.

**Esempio 1:**  $R_1 = 50\Omega$ ,  $R_2 = 1k\Omega$ ,  $I_{in} = 100mA \rightarrow I_{R2} = 4.8mA$

**Esempio 2:**  $R_1 = 12k\Omega$ ,  $R_2 = 1k\Omega$ ,  $I_{in} = 100mA \rightarrow I_{R2} = 92mA$

### Esempio:

A titolo di esercizio, verificare che la frazione di corrente che scorre nel ramo da 1.2k è di 18.5μA.



## Circuiti equivalenti

I circuiti elettronici sono spesso dei sistemi complessi di cui è importante valutare l'interazione con gli stadi elettronici successivi o con il carico, oltre naturalmente conoscere la sua funzionalità interna. I sistemi elettronici, inoltre, ricevono informazioni dal mondo esterno attraverso sensori, dispositivi che convertono la grandezza fisica di interesse (onda radio, temperatura, pressione, suono, etc...) in segnali elettrici. Benché le grandezze fisiche da convertire siano moltissime, e quindi siano moltissimi i sensori ed i trasduttori immaginabili ed effettivamente in commercio, tutti possono essere descritti, per quanto riguarda l'entità del segnale elettrico reso disponibile e la modalità con cui il segnale è fornito, da un semplice **circuito elettrico equivalente**.

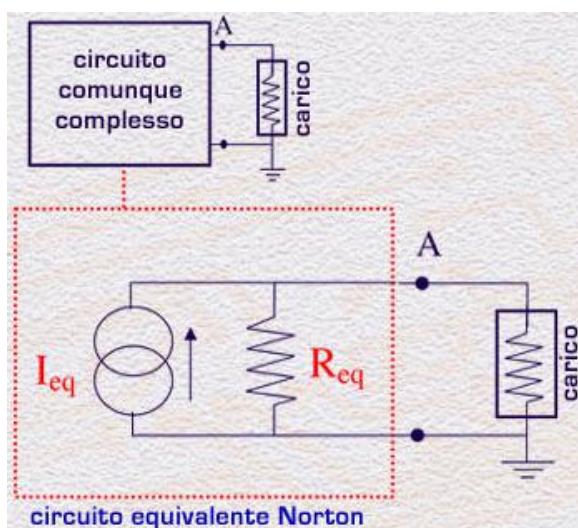
Con questo termine si intende appunto *ricondurre il circuito elettronico comunque complesso o il sensore ad un circuito elettrico che sintetizzi tutte e sole quelle caratteristiche che hanno interesse per la successiva elaborazione elettronica del segnale.*

Un possibile circuito elettronico equivalente è il **circuito equivalente Thevenin**, composto da un generatore equivalente ideale di tensione  $V_{eq}$  e da un'impedenza serie, per semplicità indicata come resistenza  $R_{eq}$ . Il generatore ideale di tensione di segnale fornisce la tensione a vuoto del circuito o del sensore, ottenuta rimuovendo il carico e valutando la tensione effettivamente presente nel punto A della figura corrispondente all'uscita del circuito o del sensore. Questo generatore rappresenta in termini elettrici l'informazione erogata dal circuito o dal sensore.

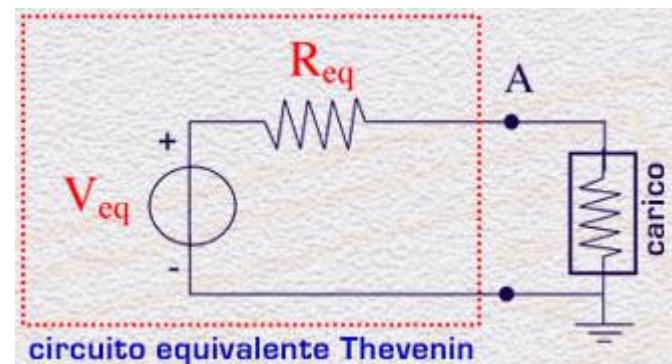
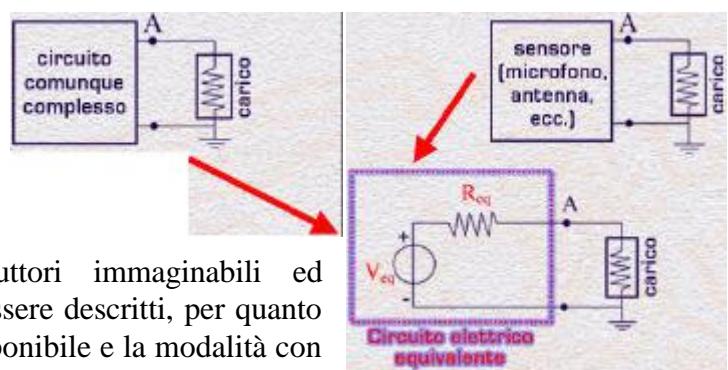
La resistenza equivalente sintetizza invece la resistenza con cui il sistema si presenta al successivo stadio elettronico. Essa è la resistenza che si misura ai terminali del circuito quando questo non fornisce segnale.

Se la **resistenza equivalente è nulla** siamo in presenza di un **generatore equivalente di tensione ideale**.

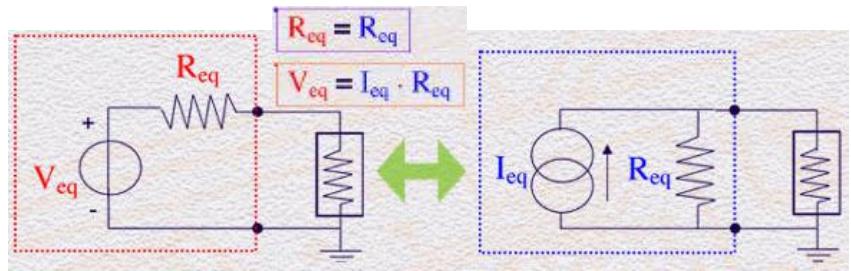
In alternativa, un circuito elettronico comunque complesso o un sensore può essere sintetizzato nel suo comportamento elettronico da un circuito elettrico equivalente composto da un generatore ideale di corrente  $I_{eq}$  e da un'impedenza in parallelo, per semplicità indicata come resistenza  $R_{eq}$ . Un tale circuito è chiamato circuito equivalente Norton. Il generatore ideale di corrente di segnale fornisce la corrente di cortocircuito del circuito o del sensore, ottenuta cortocircuitando l'uscita e valutando la corrente effettivamente circolante nel cortocircuito. Anche ora, il generatore equivalente Norton rappresenta in termini elettrici l'informazione erogata dal circuito o dal sensore. L'impedenza equivalente sintetizza invece la resistenza con cui il sistema si presenta al successivo stadio elettronico. Essa è l'impedenza che si misura ai terminali del circuito quando non fornisce segnale, come prima. Se la **resistenza equivalente è infinita** siamo in presenza di



un **generatore equivalente di corrente ideale**.



La scelta tra un circuito equivalente Thevenin o Norton dipende dalla comodità d'uso o dalla migliore rispondenza alla tipologia del segnale disponibile in uscita dal circuito od al fenomeno fisico di trasduzione del sensore. Da un circuito equivalente ci si può ricondurre all'altro notando che il valore della resistenza  $R_{eq}$  è sempre la stessa e che dalla tensione si passa alla corrente per semplice equivalenza dei due circuiti quando il carico è sconnesso.



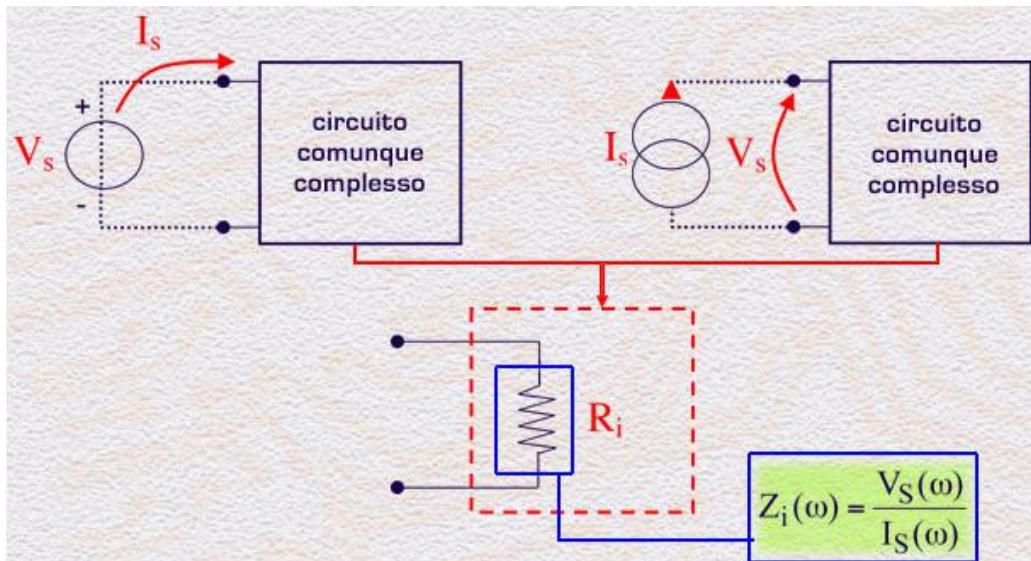
Prima di procedere è opportuno illustrare più precisamente come si calcola la resistenza equivalente mostrata tra due punti di un circuito. Si tratta di tradurre in termini operativi la definizione stessa di impedenza intesa come rapporto tra tensione e corrente.

Pertanto:

- si può applicare tra i due morsetti un segnale di tensione,  $V_s$ , e determinare la corrente  $I_s$  corrispondentemente assorbita dal circuito
- si può forzare un segnale di corrente  $I_s$  e determinare la tensione  $V_s$  che si sviluppa tra i morsetti.

$$R_i = \frac{V_s}{I_s}$$

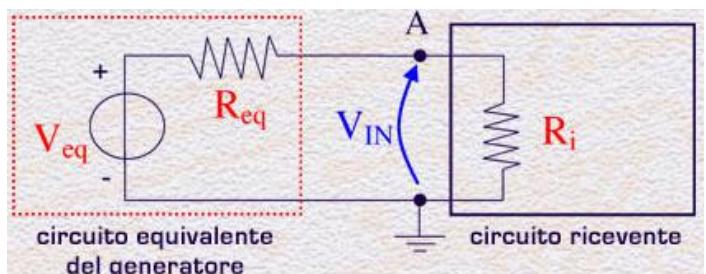
Con entrambe le procedure, il risultato viene evidentemente lo stesso. In generale, la procedura permette di ricavare l'impedenza equivalente in funzione della frequenza:



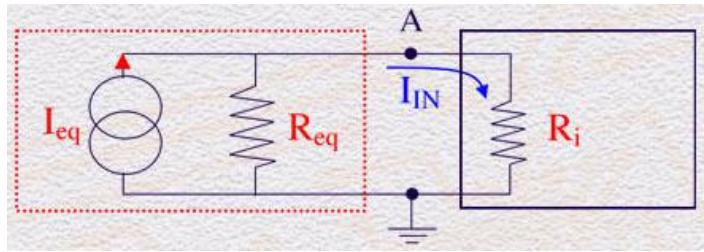
cioè rapporto tra i segnali sinusoidali di tensione e di corrente, al variare della pulsazione.

### Proprietà degli amplificatori

Dopo avere sintetizzato il circuito equivalente Thevenin del generatore di segnale e l'impedenza d'ingresso del circuito elettronico ricevente, il trasferimento di segnale tra i due è regolato dalla partizione tra  $R_{eq}$  ed  $R_i$ . Per avere il **massimo trasferimento del segnale di tensione** è bene avere  $R_{eq}$  piccola e/o  $R_i$  grande. Un circuito ricevente avente una **resistenza di ingresso infinita** consentirebbe al segnale  $V_{eq}$  di ripresentarsi esattamente uguale all'ingresso del circuito ricevente e quindi di massimizzare il trasferimento di segnale tra generatore e ricevitore. Per questo, il circuito ricevente sarebbe detto essere un "**lettore ideale di tensione**".

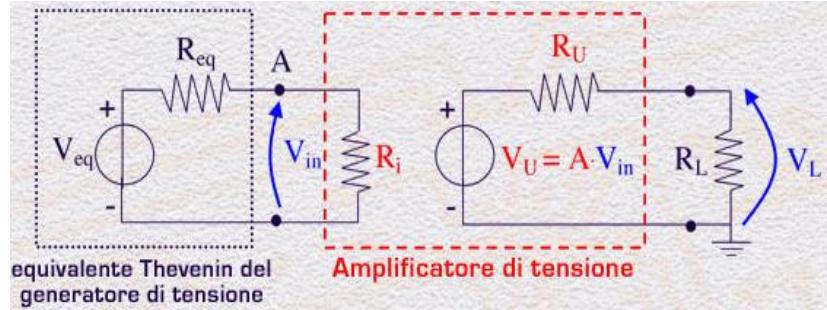


Analogo il discorso quando si debba trasferire un segnale di corrente, che supporremo disponibile in uscita da un circuito generatore sintetizzato dal suo equivalente Norton. Per avere il **massimo trasferimento del segnale di corrente** dal generatore al circuito ricevente è necessario che  $R_{eq}$  sia grande e/o  $R_i$  sia piccola. Al limite, un circuito ricevente avente una *resistenza di ingresso  $R_i$  nulla* assorbirebbe tutta la corrente di segnale  $I_{eq}$  indipendentemente dalla  $R_{eq}$ , massimizzando così il trasferimento di segnale tra l'uno e l'altro. Esso sarebbe indicato come un “*lettore ideale di corrente*”.



Il segnale fornito da un circuito o da un sensore viene normalmente elaborato da altri circuiti posti a valle. La più intuitiva delle elaborazioni possibili è quella di amplificare il segnale in modo da portarne il livello ad un valore maggiore. Questo compito è svolto dagli amplificatori.

Un amplificatore di tensione può essere rappresentato schematicamente dal circuito equivalente riportato nel riquadro tratteggiato centrale. Esso è caratterizzato dal valore del guadagno massimo di tensione  $A$  tra il morsetto di ingresso e di uscita, e dalle due impedenze  $Z_{in}$  tra l'ingresso e massa e  $Z_u$  tra l'uscita e massa.

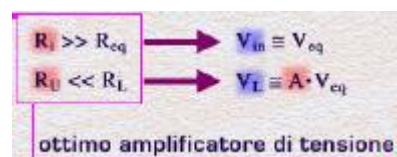


Nella figura sono riportate per semplicità le sole componenti resistive dell'impedenza ed è stato completato il sistema con il circuito equivalente Thevenin dello stadio a monte che fornisce il segnale di ingresso e con il circuito utilizzatore, rappresentato dal resistore di carico.

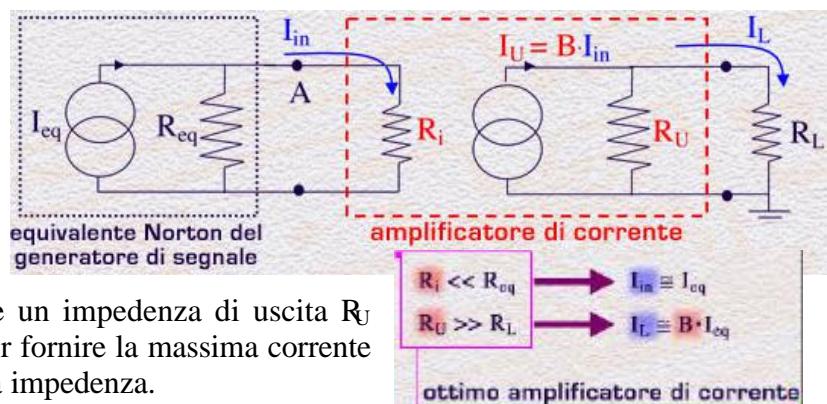
Affinché il segnale  $V_{eq}$  possa essere correttamente amplificato senza essere attenuato dalla partizione resistiva di ingresso, *l'impedenza di ingresso dell'amplificatore dovrebbe essere infinita o per lo meno ben maggiore della resistenza equivalente  $R_{eq}$*  del circuito elettronico che lo precede. Questa considerazione si esprime dicendo che il circuito d'ingresso dell'amplificatore di tensione deve essere un buon *lettore di tensione*, come già visto.

In generale *l'impedenza di ingresso di un circuito determina la frazione del segnale erogato dal generatore forzante effettivamente disponibile per comandare l'amplificatore*.

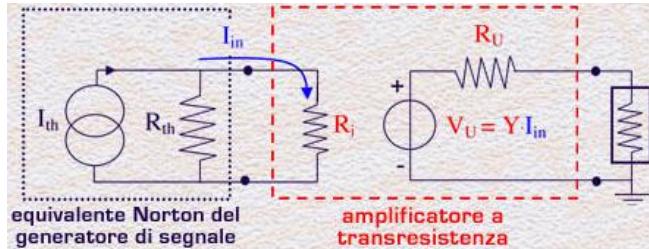
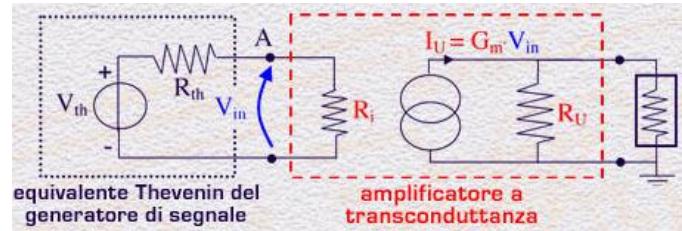
Un amplificatore di tensione deve anche erogare al carico il segnale amplificato. Quindi, affinché ai capi del carico si riesca a sviluppare la massima tensione disponibile, pari a  $V_L = A \cdot V_{in}$ , occorre che sia  $R_u \ll R_L$ . Ovvero, l'uscita di un amplificatore di tensione deve avere un'impedenza molto più bassa delle impedenze dei carichi a cui detto amplificatore è connesso.



Naturalmente vi sono anche **amplificatori di corrente**, che rilevano un segnale di corrente fornito dal generatore di segnale e lo amplificano rendendolo disponibile all'uscita. Per quanto visto fino ad ora, un buon amplificatore di corrente dovrà avere impedenza di ingresso bassa, al limite nulla, in modo da raccogliere la maggior parte possibile della corrente  $I_{eq}$  fornibile dallo stadio precedente; e deve avere un'impedenza di uscita  $R_u$  alta, al limite infinita, in modo da poter fornire la massima corrente possibile al carico qualunque sia la sua impedenza.



Sono previste anche altre combinazioni diverse di amplificatori. **Amplificatori a transconduttanza**, che rilevano un segnale di tensione fornito dal circuito a monte e lo convertono in un segnale di corrente da fornire ad un carico a valle.



**Amplificatori a transresistenza**, che rilevano un segnale di corrente fornito dal generatore di segnale a monte e lo convertono in un segnale di tensione con cui pilotare un carico a valle.

## Esercizio 1

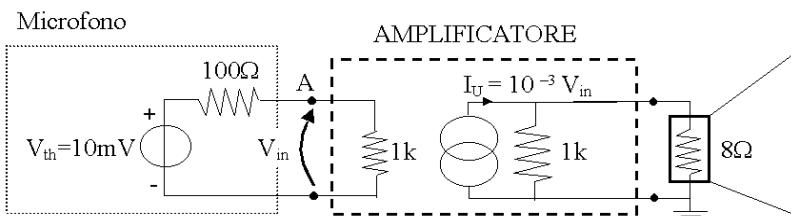
a- Perché il circuito equivalente Thevenin viene considerato "ideale" se  $R_{eq}=0 \Omega$  ?

b- Si consideri il caso di un microfono che sviluppa una tensione a vuoto ( $V_{eq}$ ) di 1mV e che si presenta con una  $R_{eq}=1k\Omega$  e lo si colleghi ad un carico (supponiamo che sia l'amplificatore del segnale audio) di  $R_c=2k\Omega$ . Calcolare quale frazione del segnale microfonico riesce effettivamente ad essere applicato all'ingresso dell'amplificatore (ossia ai capi del carico).

Un microfono con resistenza equivalente di  $100\Omega$  sarebbe stato meglio dal punto di vista del trasferimento del segnale? Provate a cercare quale è l'impedenza equivalente di un buon microfono commerciale e commentatene il valore.

## Esercizio 2

Calcolare, per il seguente sistema di amplificazione, quale tensione arriva all'altoparlante supposto che il microfono possa generare un segnale a vuoto di 10mV.



## Lezione n. 3 : Segnali elettronici e loro rappresentazione

### Segnali analogici e digitali

I segnali elettrici, intesi come variazioni nel tempo della tensione in un punto di un circuito o della corrente in un ramo di un circuito, trovano una rappresentazione naturale in grafici in cui sull'asse delle ascisse ci sia il tempo e sull'asse delle ordinate la grandezza elettrica, tensione o corrente.

Tale rappresentazione è sicuramente intuitiva ed evidenzia bene il valore della grandezza elettrica in ogni istante del suo evolversi.

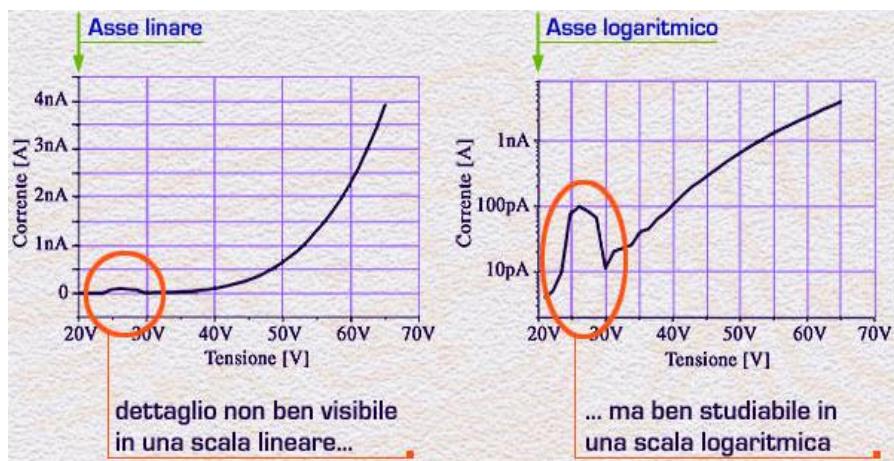
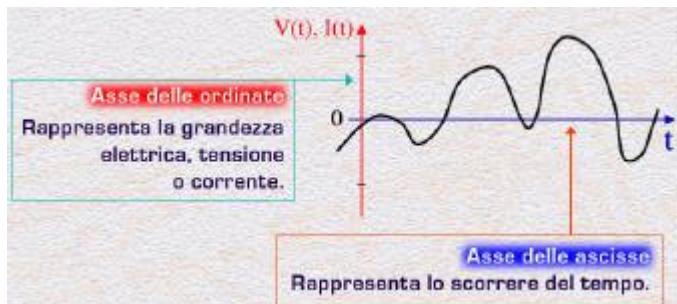
La rappresentazione nel tempo di un segnale elettrico in un circuito può effettivamente essere ottenuta grazie all'uso di uno strumento specifico, chiamato **oscilloscopio**. Questo strumento di misura è costituito da un tubo a raggi catodici, come lo schermo di un televisore, in cui la deflessione orizzontale del fascio è fatta a velocità costante e definisce lo scorrere del tempo, mentre la deflessione verticale è proporzionale al valore del segnale elettrico presente al suo ingresso. Lo strumento è corredata di tutte le regolazioni necessarie a modificare sia la velocità della deflessione orizzontale, così da raccogliere nell'immagine sullo schermo un intervallo di tempo più o meno esteso, sia l'ampiezza della deflessione verticale, in modo da rappresentare a pieno schermo anche segnali elettrici molto piccoli.

Proprio per la intuitività della visualizzazione nel tempo di un segnale, l'oscilloscopio è lo strumento che non manca mai in un laboratorio elettronico ed è sicuramente il più utilizzato.

La scala verticale dei grafici è normalmente una scala lineare. Tuttavia quando i segnali variano molto in ampiezza (si dice che la dinamica del segnale è elevata) una scala lineare non permette di vedere i dettagli del segnale quando è piccolo. Infatti, poiché il fondo scala dell'asse è scelto per poter includere il valore massimo del segnale, i segnali piccoli risultano tutti compresi in uno spazio piccolissimo nella parte bassa del grafico ed i loro dettagli sono indistinguibili.

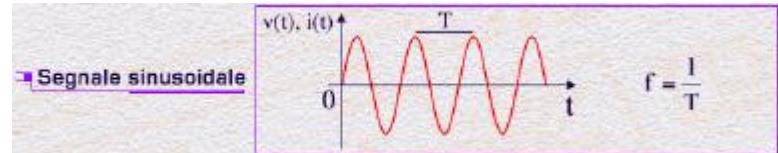
Quando si è in questi casi e quei dettagli sono importanti, l'inconveniente può essere superato dalla scelta di un asse verticale logaritmico. In questo modo si dà la stessa importanza (e cioè lo stesso spazio verticale sul foglio per rappresentarlo) ai segnali piccoli ed ai segnali grandi: ogni decade di valori di segnale ha lo stesso spazio rappresentativo in verticale sul foglio, sia che il segnale sia di pochi pA sia che sia di parecchi nA come nella figura.

I segnali variabili nel tempo in cui *il valore che assume la grandezza elettrica in ogni istante è l'elemento importante* sono detti **segnali analogici**. In essi quindi l'informazione è contenuta proprio nelle piccole o grandi variazioni del suo valore in ogni istante di tempo.

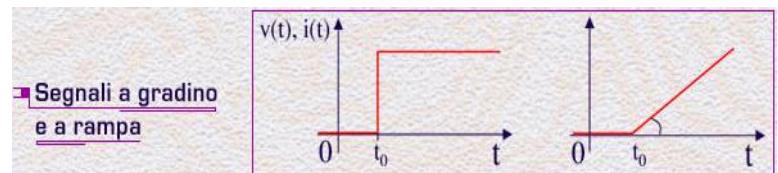


Alcuni esempi di segnali analogici a cui faremo riferimento durante il corso sono qui riportati:

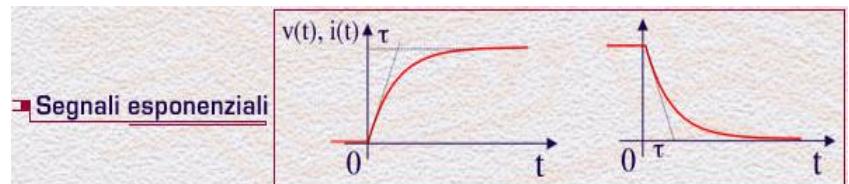
- il segnale sinusoidale, caratterizzato oltre che dalla sua ampiezza, dal suo periodo  $T$  o dalla sua frequenza  $1/T$ ;



- il segnale a gradino, in cui il valore della grandezza elettrica cambia istantaneamente per poi mantenersi sul nuovo livello ed il segnale a rampa, in cui l'aumento è costante;



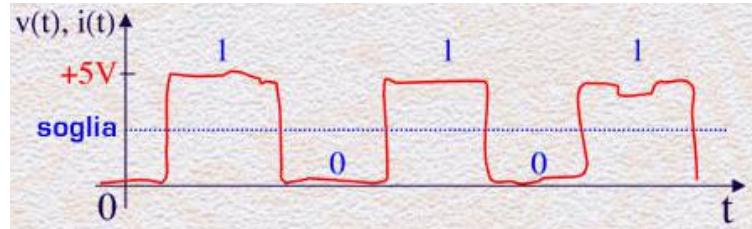
- il segnale esponenziale, crescente o decrescente, in cui l'ampiezza varia nel tempo con continuità secondo una legge esponenziale, caratterizzato dalla costante di tempo  $\tau$  dell'esponenziale.



I segnali variabili nel tempo in cui invece è importante solo che siano superiori o inferiori ad una data soglia, sono detti segnali digitali. In essi l'informazione è contenuta nello stato in cui si trova il segnale, sopra la soglia o sotto la soglia, e non nel valore dettagliato che assume la grandezza elettrica all'interno di ognuno di questi due stati. Ad ogni livello elettrico è associato il livello logico 0 od 1. Proprio per questa loro natura, i segnali digitali necessitano di convenzioni e standard accettati da tutti per riconoscere i due stati.

Per anni lo standard prevalente era quello di due stati, uno a 0V e l'altro a 5V. Ora, con l'esigenza di diminuire sempre più la potenza dissipata dai circuiti in modo da addensarne le funzioni, i livelli si stanno avvicinando. Tipici livelli dei circuiti integrati sono da 0V

e da 3V, ma valori anche minori da 0 e da 1.5V sarebbero desiderati ed in qualche caso sono stati anche realizzati.



**Standard TTL: 0V [0 logico] +5V [1 logico]**

**Nuovi standards: 0V/+3V oppure 0V/1.5V**

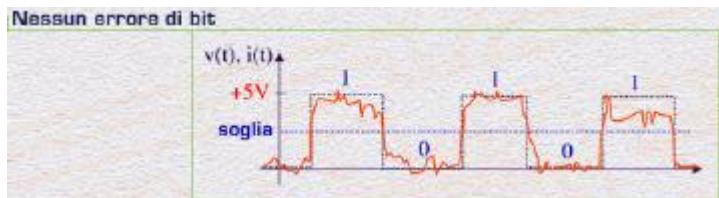
## Rumori e interferenze

Ai segnali elettrici, siano essi analogici o digitali, è sempre sovrapposto del rumore. *Il rumore è una fluttuazione casuale della tensione o della corrente attorno al valore teorico che dovrebbe avere.* Il rumore è una *conseguenza della casualità di molti dei processi fisici fondamentali in natura che sottostanno al funzionamento degli elementi* (resistori, transistori, ecc.) *presenti in un circuito.* Esso è quindi **sempre inevitabilmente minimizzato** con opportuni interventi di progetto ma mai interventi di progetto mai segnale reale analogico traccia non è una nitida linea, dovuto appunto alle gergo anche chiamata "barba" dalla radice quadrata del suo (indicato con il termine RMS: Root Mean Square) che corrisponde all'ampiezza della barba identificabile sull'oscilloscopio.

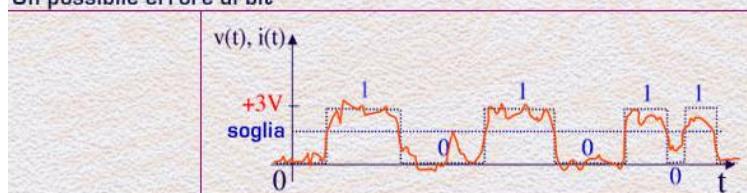


L'effetto della presenza del rumore sui segnali analogici è di variare bcalmente il valore rispetto al valore atteso e quindi di rendere l'informazione contenuta nella grandezza elettrica meno precisa. L'effetto è proprio quello che si percepisce nei segnali audio come "fruscio" di fondo sovrapposto al parlato od alla musica.

Nel caso di segnali digitali, l'effetto della presenza del rumore è molto diverso. Purché il rumore non raggiunga mai un'ampiezza superiore al valore della soglia, esso non causa alcuna alterazione dell'informazione contenuta nel segnale. Infatti, pur in presenza del rumore, i livelli logici 0 e 1 vengono ancora distinti con certezza.



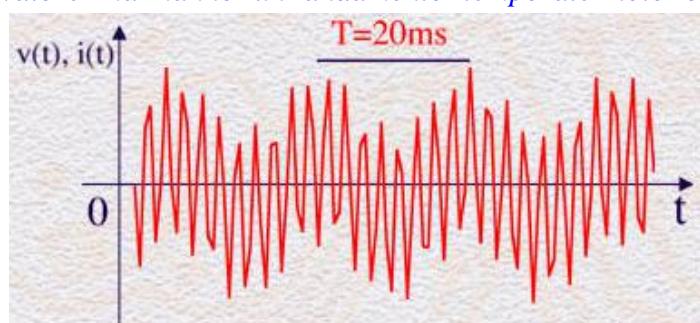
**Un possibile errore di bit**



Diverso è il discorso quando il livello della soglia è tale che c'è una probabilità non più trascurabile che il rumore faccia superare la soglia facendo erroneamente scattare il livello logico.

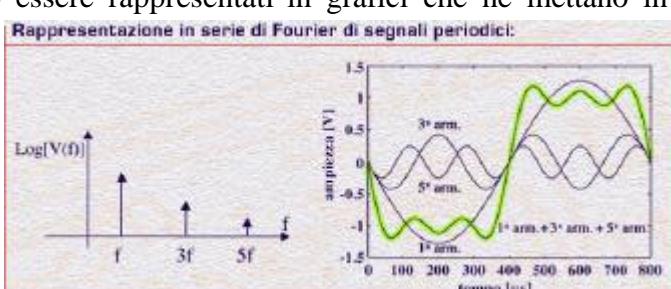
Poiché l'evoluzione dei circuiti digitali è quella di ridurre il più possibile l'ampiezza del segnale per dissipare meno potenza, anche il rumore deve essere ridotto pena l'aumento degli errori di bit. Nei più moderni circuiti si accetta un tasso di frequenza degli errori, chiamato **bit error rate**, a volte inferiore ad  $10^{-6}$ .

Oltre che per il rumore, il segnale elettrico di tensione o di corrente può discostarsi dal valore atteso a causa di disturbi o di interferenze. Con questi termini si intendono *segnali elettrici che si sovrappongono a quelli utili alterandone il valore ma hanno un andamento temporale noto o comunque riconoscibile*. Essi sono dovuti generalmente a cause esterne ai componenti del circuito in esame. Un tipico esempio di disturbo è quello *sinusoidale a 50Hz* presente nei sistemi elettronici alimentati dalla tensione di rete e dovuto alla non completa eliminazione della periodicità del segnale sinusoidale durante l'operazione di rettificazione e filtraggio per produrre la tensione continua. Nel grafico si vede quanto sia riconoscibile un disturbo a 50Hz su un segnale audio di 1kHz. Esso verrebbe riconosciuto dall'orecchio umano come un rumore basso di sottofondo al suono da 1kHz. **Disturbi ad alta frequenza** possono essere generati in un circuito per induzione dai segnali elettromagnetici presenti nell'ambiente, quali le onde radio o le portanti dei telefonini.

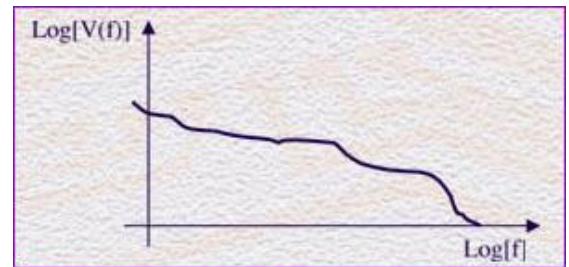


### Rappresentazione spettrale di un segnale

I segnali elettrici, oltre che nel tempo, possono essere rappresentati in grafici che ne mettano in evidenza le singole componenti spettrali. Tali grafici hanno quindi sull'asse delle ascisse le frequenze delle sinusoidi e su quello delle ordinate le loro ampiezze. Questo tipo di rappresentazione nasce dalla constatazione che ogni segnale elettronico nel tempo può essere scomposto nella somma di componenti sinusoidali a frequenze crescenti. Tale considerazione vale innanzitutto per segnali periodici, che possono essere espressi in modo univoco mediante la *serie di funzioni trigonometriche di frequenza f, 2f, 3f...ecc.*, di ampiezze e fasi opportune. Tale sviluppo è detto "sviluppo in serie di Fourier". Nella figura sono ad esempio visualizzate le prime 3 armoniche dello sviluppo in serie di un segnale digitale, con il relativo segnale somma. Appare evidente come le armoniche di ordine superiore siano necessarie per descrivere il fronte ripido delle transizioni a gradino.



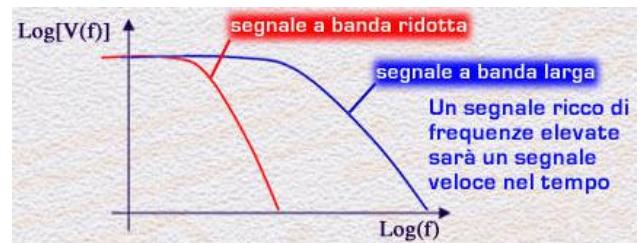
Sotto **opportune condizioni di regolarità**, non solo un segnale periodico, ma anche una generica funzione  $f(t)$  può essere descritta come sovrapposizione di componenti armoniche, detta **Trasformata di Fourier**. Nel caso dei circuiti elettronici si fa ampio uso di una ulteriore generalizzazione, detta **Trasformata di Laplace**, che permette di trattare tutti i segnali elettronici possibili. Essa fornisce un continuo di frequenze la cui composizione ricostruisce esattamente l'originario segnale nel tempo.



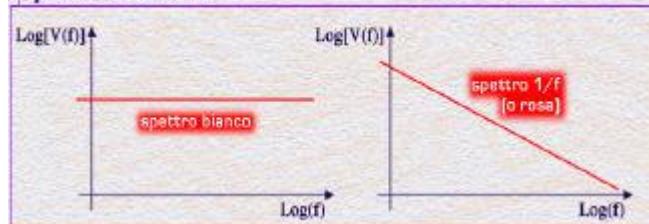
La rappresentazione spettrale di un segnale è complementare a quella nel tempo ed è spesso di grandissima utilità:

- visualizza chiaramente le componenti in frequenza che concorrono alla formazione del segnale originario nel tempo;
- fornisce pertanto tutte le indicazioni per un progetto ottimizzato del circuito che deve elaborarne il segnale, così ad esempio da attenuare alcune frequenze o amplificarne altre in base alle esigenze.

Per quanto riguarda la corrispondenza tra segnale nel tempo e suo spettro in frequenza, si ricordi che un segnale che nel tempo presenta variazioni veloci di ampiezza ha uno spettro ricco anche di frequenze elevate, proprio quelle necessarie per descrivere il fronte ripido delle transizioni. Segnali veloci sono così detti segnali a larga banda. Viceversa, un segnale con variazioni lente nel tempo avrà uno spettro limitato alle frequenze più basse.



#### Spettro del rumore



Anche il rumore ha uno spettro, molto spesso bianco (come la luce bianca con ugual intensità per ogni frequenza), cioè distribuito su tutte le frequenze in ugual modo, oppure uno spettro rosa, indicato anche come “spettro  $1/f$ ”, in cui c’è prevalenza di frequenze basse.

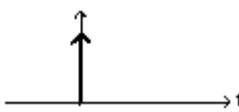
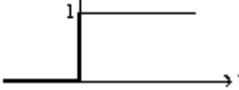
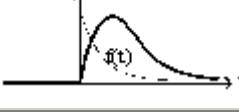
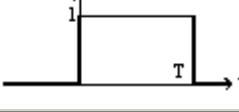
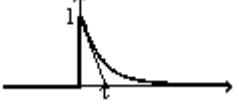
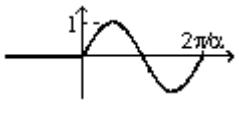
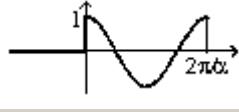
## Complementi

E’ opportuno riguardare e riprendere confidenza con le trasformate di Fourier e di Laplace.

Per comodità, qui di seguito è riportata una tabella riassuntiva delle proprietà della trasformata di Laplace ed una tabella della trasformata di Laplace di alcune funzioni elementari di frequente uso in elettronica.

PROPRIETA' [t]	FUNZIONE	TRASFORMATA	PROPRIETA' [s]
Linearità	$c_1.f_1(t)+c_2.f_2(t)$	$c_1.f_1(s)+c_2.f_2(s)$	Linearità
Derivazione	$\frac{df(t)}{dt}$	$s f(s)$	Moltiplicazione per s
Integrazione	$\int f(t)dt$	$\frac{1}{s} f(s)$	Divisione per s
Moltiplicazione per t	$t.f(t)$	$-\frac{df(s)}{ds}$	Derivazione

Traslazione nel tempo	$f(t-T)$	$e^{-sT}f(s)$	Moltiplicazione per un'esponenziale
Valore iniziale	$f(0_+)$	$\lim_{s \rightarrow \infty} sf(s)$	
Valore finale	$f()$	$\lim_{s \rightarrow 0} sf(s)$	
Convoluzione	$f(t)*g(t)$	$f(s) \cdot g(s)$	

NOME DELLA FUNZIONE	$f(t)$	RAPPRESENTAZIONE GRAFICA	$f(s)$
$\delta$ di Dirac	$\delta(t)$		1
Gradino unitario	$1(t)$		$\frac{1}{s}$
Rampa lineare	$t \cdot 1(t)$		$\frac{1}{s^2}$
Potenze n-esime	$t^n \cdot 1(t)$		$\frac{n!}{s^{n+1}}$
	$t^n \cdot f(t) \cdot 1(t)$		$(-1)^n \frac{d^n}{ds^n} f(s)$
Rettangolo	$1(t) - 1(t-T)$		$\frac{1}{s} (1 - e^{-sT})$
Discesa esponenziale	$e^{-t/\tau} 1(t)$		$\frac{1}{s + 1/\tau}$
Sinusode	$1(t) \cdot \sin(\alpha t)$		$\frac{\alpha}{s^2 + \alpha^2}$
Cosinusode	$1(t) \cdot \cos(\alpha t)$		$\frac{s}{s^2 + \alpha^2}$

### Esercizio 1

Provate a fare una ricerca sulle caratteristiche degli **oscilloscopi**:

Quale è la frequenza massima del segnale visualizzabile ?

Quale è la minima ampiezza visualizzabile ?

E la massima ?

Quanti segnali diversi si possono visualizzare contemporaneamente ?

Come si sincronizza un segnale ripetitivo in modo che appaia stabile sullo schermo ?

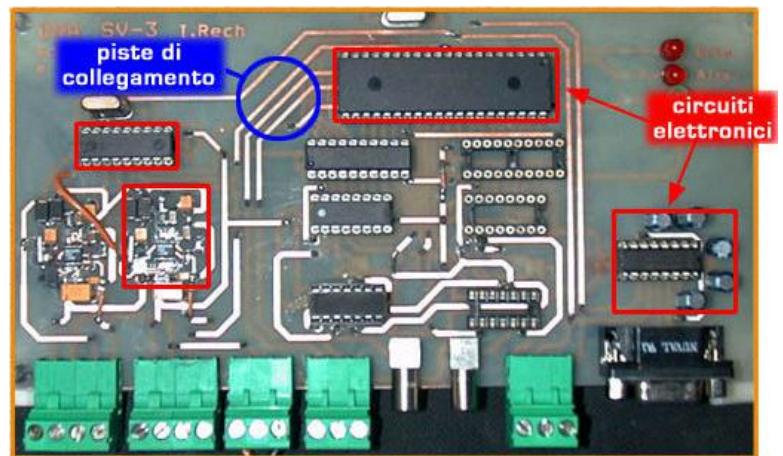
## Lezione n. 4 : Trasmissione dei segnali

### Circuito equivalente RC di una connessione e dinamica di risposta ad un gradino

I segnali elettrici generati o comunque presenti in un punto di un sistema elettronico si propagano nel circuito lungo le connessioni metalliche che collegano i singoli dispositivi e lungo i cavi di connessione tra circuiti ed apparati diversi.

Durante questo trasferimento, **il segnale non si mantiene inalterato**. Parte o tutte le sue caratteristiche di ampiezza, di componenti spettrali, di andamento temporale, di sfasamento o altro vengono infatti modificate in maniera più o meno accentuata.

Se prendiamo ad esempio una scheda elettronica come quella che siamo tutti abituati a vedere all'interno di un computer, notiamo che essa è fatta da tanti circuiti elettronici, ognuno che svolge un suo proprio compito, connessi tramite delle piccole piste metalliche serigrafeate sulla superficie della scheda.



Vediamo quale ripercussione abbia sul segnale questa connessione in apparenza innocua.

Se facessimo una sezione di una zona della scheda lungo una pista, troveremmo la situazione mostrata nella figura: due circuiti distinti connessi attraverso la pista che consente la trasmissione del segnale dall'uno all'altro.

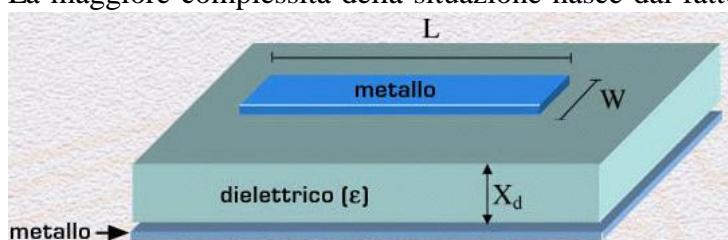
Supponiamo che il circuito di sinistra sia quello che genera il segnale da trasmettere disponibile alla sua uscita come segnale di tensione. Dal punto di vista elettronico, il circuito, comunque complesso sia al suo interno, può essere completamente descritto con il suo circuito equivalente Thevenin fatto dal generatore di tensione equivalente  $V_{in}$  e dalla sua resistenza equivalente di uscita  $R_{th}$ .

Anche il circuito di destra, che riceve il segnale, può essere completamente descritto dalla sua resistenza equivalente di ingresso  $R_i$ . Se la pista di collegamento fosse assimilabile ad un semplice filo metallico, il problema della trasmissione del segnale non sussisterebbe: il segnale prodotto in uscita dal circuito a sinistra sarebbe disponibile identico all'ingresso del circuito ricevente.

Se ad esempio il segnale  $V_{in}$  fosse

un segnale digitale con i livelli a 0V ed a +3V e le transizioni infinitamente veloci, si otterrebbe un segnale in arrivo  $V_U$  identico a quello in partenza, al più ridotto in ampiezza dalla partizione tra le due resistenze.

La maggiore complessità della situazione nasce dal fatto che la pista metallica è posta a contatto con la superficie dell'isolante della piastra, normalmente realizzata con resine plastiche, e che l'altra faccia della piastra è ad una tensione fissa, o perché anch'essa localmente metallizzata e tenuta volutamente a massa (come succede nella



maggior parte dei casi anche per motivi di schermatura dai disturbi) o perché indirettamente a massa per il segnale tramite piste contigue. La piastra costituisce perciò il dielettrico di un condensatore i cui piatti sono la pista metallica sopra e, nella soluzione più semplice, l'elettrodo di massa sotto.

La pista metallica si comporta dunque come una capacità di valore:

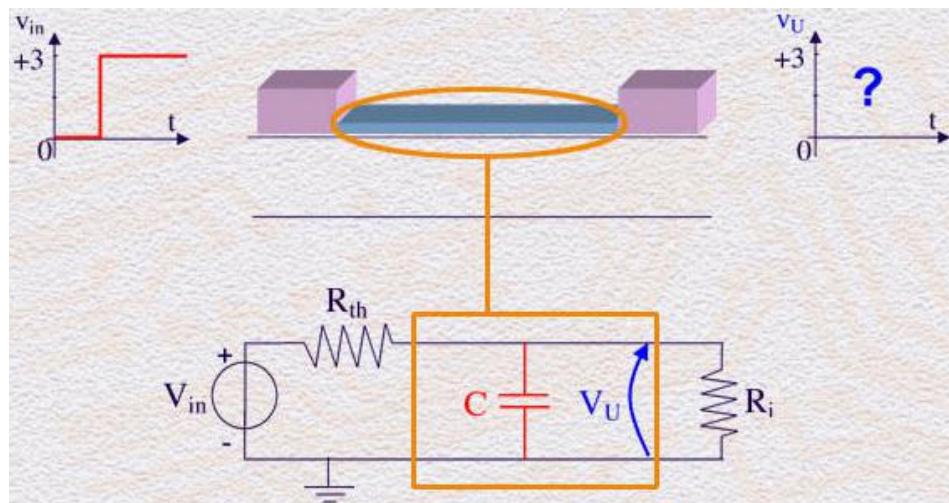
Essa è quindi proporzionale alla costante dielettrica  $\epsilon$  del materiale isolante con cui è realizzata la piastra ed all'area dei piatti del condensatore (nel nostro caso all'area della pista metallica di collegamento pari a  $WL$ ). La capacità è invece inversamente proporzionale alla distanza  $X_d$  tra le piste metalliche, cioè allo spessore della piastra.

Nel caso di una piastra di vetro con  $\epsilon = 10^{-12} \text{ F/cm}$  e  $X_d=1\text{mm}$ ,  $W=1\text{mm}$  ed  $L=1\text{cm}$  si otterebbe una capacità **C=1pF**.

La presenza di questa capacità modifica la rete elettrica che simula il collegamento. La capacità prodotta dalla pista di collegamento si pone infatti in parallelo all'ingresso del circuito di arrivo del segnale. Vogliamo vedere come la sua presenza influisce sul trasferimento di segnale attraverso la connessione. Supponiamo, per semplicità, che il circuito utilizzatore sia un lettore ideale di tensione, per cui  $R_i$  sia molto più grande di  $R_{th}$ , e quindi trascurabile.

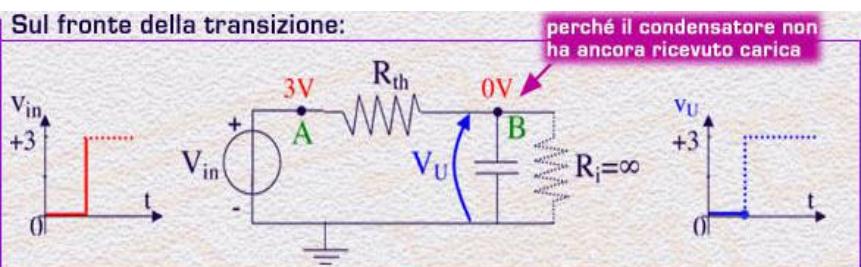
Supponiamo di applicare all'ingresso del circuito un segnale digitale con i livelli a 0V ed a +3V, e concentriamoci sulla transizione da 0 a 3V. Il segnale è quindi modellizzabile come un gradino, cioè una variazione istantanea, del valore della tensione all'ingresso.

$$C = \frac{\epsilon}{X_d} \cdot W \cdot L$$

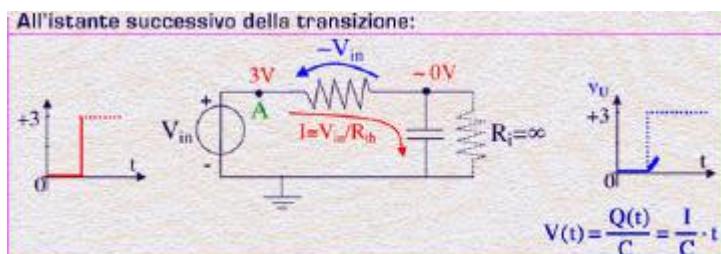


Sul fronte della transizione, il punto A necessariamente varia di  $V_{in}$  istantaneamente. Il punto B invece, in quello stesso istante non potrà variare la sua tensione: infatti il punto B coincide con uno dei due capi del condensatore, l'altro essendo fissato a massa.

Affinché B si alzi in tensione, deve essere depositata della carica positiva sul piatto superiore del condensatore. Ma la carica può arrivare solo dal percorso che attraversa la resistenza e sul fronte della transizione la carica accumulata è ancora nulla.



All'istante successivo della transizione:

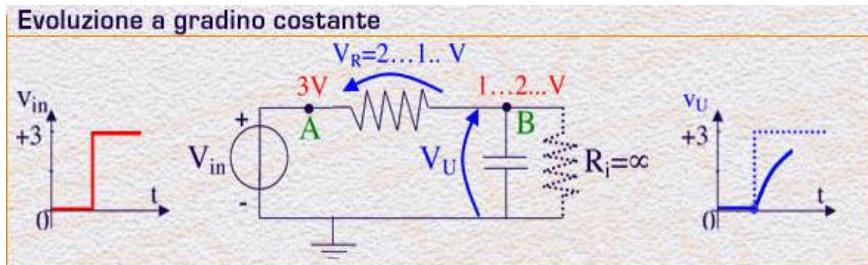


Solo all'istante successivo della transizione, e appunto con la tensione in B ancora a 0V, si sviluppa una tensione ai capi di  $R_{th}$  praticamente uguale a  $V_{in}$ . Pertanto in  $R_{th}$  inizierà a circolare una corrente pari a:  $I = V_{in}/R$ . Poiché la corrente è costituita da cariche positive che si muovono nel verso della freccia, essa va a caricare la capacità. La tensione in B comincerà quindi a salire di tensione secondo la legge:  $V(t) = Q(t)/C = I \cdot t/C$ , cioè con un inizio di rampa lineare.

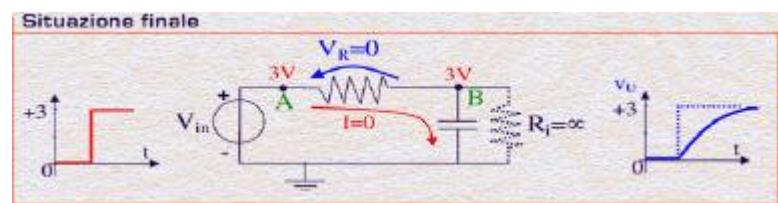
della freccia, essa va a caricare la capacità. La tensione in B comincerà quindi a salire di tensione secondo la legge:  $V(t) = Q(t)/C = I \cdot t/C$ , cioè con un inizio di rampa lineare.

Infatti nei primi istanti la corrente è praticamente costante perché la tensione in B rimane comunque piccola, e quindi la tensione ai capi di  $R_{th}$  varia poco: la capacità viene caricata con costanza e la tensione sale quasi linearmente.

Quanto più sale la tensione in B, tanto minore è la caduta di potenziale ai capi di  $R_{th}$  e quindi minore è la corrente che affluisce alla capacità. Conseguentemente, con il procedere del tempo, la capacità si carica sempre meno velocemente e la salita in tensione di B è sempre meno veloce.

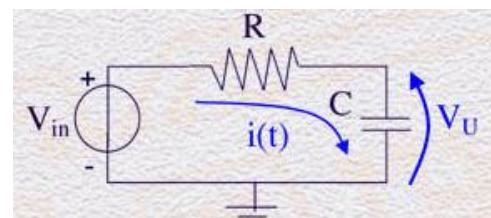


La situazione finale raggiunta dopo un certo tempo è quella in cui il punto B è salito in tensione proprio fino al valore di  $V_{in}$  per cui ai capi di  $R_{th}$  non c'è più differenza di potenziale e quindi non scorre più corrente che possa ulteriormente alimentare la capacità. Fintantoché  $V_{in}$  rimane fissa, anche  $V_u$  non cambia più.



Se guardiamo alla forma d'onda  $V_u(t)$  nel suo complesso, traiamo la conclusione che quando si applica un gradino di tensione  $V_{in}$  all'ingresso del semplice circuito resistenza-capacità, solo dopo un certo tempo la tensione ai capi del condensatore, cioè all'uscita del nostro circuito, finalmente raggiunge il valore di +3V. La trasmissione del segnale è quindi avvenuta con una certa lentezza e la forma d'onda in arrivo è diversa da quella in partenza. Per via della sua topologia, un tale circuito è generalmente chiamato "Rete RC".

La risposta precisa nel tempo della rete RC che abbiamo trovato in forma qualitativa può essere ricavata in modo rigoroso applicando la legge di Kirchhoff alla maglia, (cfr. figura a lato):  $V_u(t) = V_{in} - R \cdot i(t)$ .



**Legame tensione-corrente in un condensatore**

$$i(t) = C \frac{dV(t)}{dt}$$

Utilizzando il legame differenziale tra la corrente che attraversa il condensatore e la tensione ai suoi morsetti, si ottiene l'equazione differenziale:

$$V_u(t) = V_{in} - RC \frac{dV_u(t)}{dt}$$

che può essere più comodamente riscritta come a lato e che viene integrata con la **condizione iniziale**  $V_u(0)=0V$  (condensatore inizialmente scarico) e la cui soluzione è:

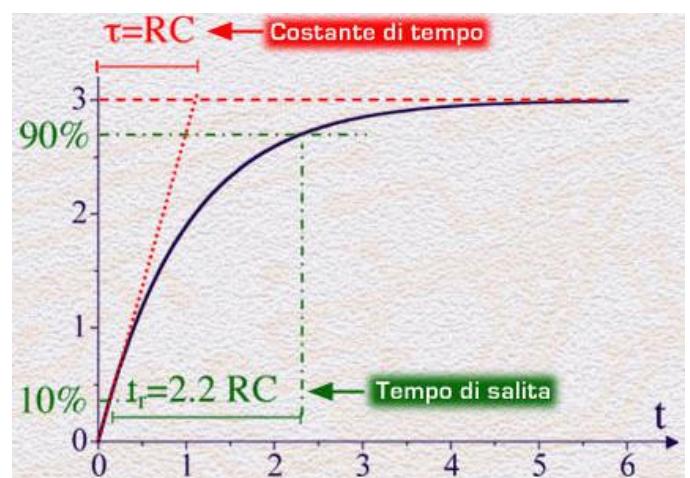
$$V_u(t) = V_{in} \cdot \left( 1 - e^{-\frac{t}{RC}} \right)$$

$$\frac{dV_u(t)}{dt} + \frac{1}{RC} V_u(t) = V_{in}$$

L'uscita quindi aumenta la sua tensione dal valore iniziale di 0V al valore finale pari a  $V_{in}$  seguendo un andamento esponenziale, come quello mostrato nella figura.

In esso vengono individuate due grandezze caratteristiche:

- la **costante di tempo**  $t = RC$  dell'esponenziale. Essa corrisponde all'intervallo di tempo che deve

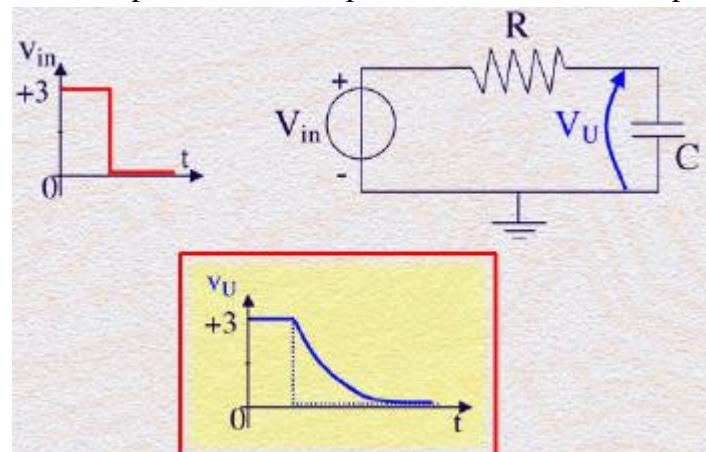


intercorrere affinché la tensione di uscita abbia raggiunto il 63% del valore asintotico. Si noti che la tangente nell'origine all'esponenziale  $V_u(t)$  interseca l'asintoto  $V_u = V_{in} = +3V$  proprio a  $t = RC$ .

- il **tempo di salita**  $t_r$  che corrisponde al tempo necessario affinché il segnale si porti dal 10% al 90% del valore asintotico; esso è pari a circa  $2.2 \cdot RC$ .

Ricordando quanto detto a proposito dell'aumento lineare della tensione  $V_u$  di uscita durante i primi istanti della transizione, questa rete RC è anche detta **integratrice approssimata** perché ad un segnale di ingresso costante quale è il gradino dopo il fronte risponde, ma solo per  $t < RC$  con una rampa lineare, che è proprio l'integrale del gradino.

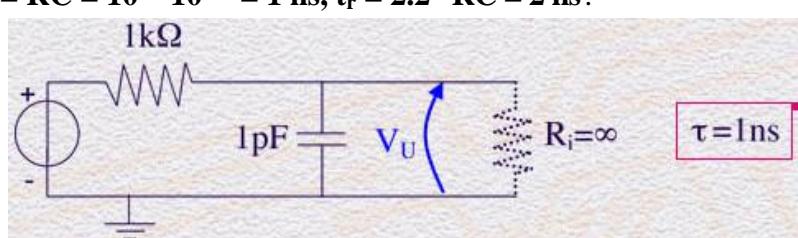
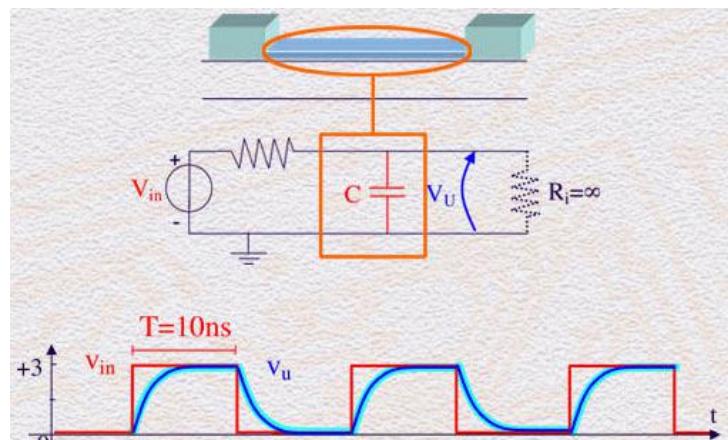
La stessa costante di tempo caratterizza anche la risposta nel tempo quando in ingresso è applicato il gradino in discesa. In questo caso la lentezza nella risposta è dovuta al tempo finito che ci vuole per scaricare il condensatore C di tutta la carica accumulata quando il livello stava a +3V. E' lasciato al lettore verificare per via analitica che si ottiene l'andamento mostrato in figura.



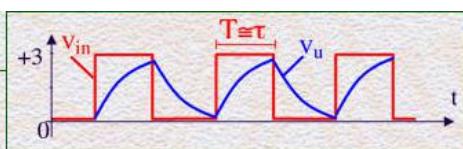
### Trasmissione di un segnale digitale

Se pensiamo di voler trasmettere un treno di impulsi digitali, ipotizzando che nel punto di inizio della trasmissione esso sia un segnale ideale tra 0V e +3V con transizioni infinitamente rapide indicato in rosso nel grafico, il segnale effettivamente letto dal circuito ricevente, in blu, è un segnale distorto con fronti di salita e di discesa più lenti. Fintanto che la frequenza del segnale di ingresso (intesa come numero di transizioni al secondo) è tale per cui il suo periodo è più lungo di alcune costanti di tempo della rete RC, il segnale di uscita benché distorto riesce a raggiungere il livello asintotico e quindi l'informazione digitale contenuta nei due livelli logici è mantenuta. Questo è il caso del grafico, in cui il segnale digitale è trasmesso alla velocità di circa 100Mbit/s (cioè un periodo di 10ns per bit) da un generatore con  $R_{th} = 1k$  lungo una pista da 1pF di capacità, per cui si ha:  $t = RC = 10^3 \cdot 10^{-12} = 1 \text{ ns}$ ,  $t_r = 2.2 \cdot RC = 2 \text{ ns}$ .

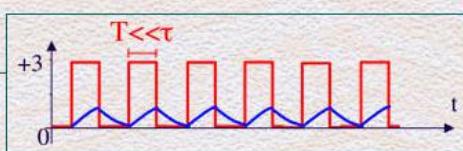
Se si aumentasse la frequenza di trasmissione, si correrebbe il rischio di non dare tempo al segnale di uscita di raggiungere non solo il valore asintotico ma neanche di superare la soglia di discriminazione tra i livelli. Il limite d'uso può essere considerato quello in cui il periodo di ogni bit si è ridotto fino ad arrivare ad essere equivalente alla costante di



**Se  $T \approx \tau$  ( $f_{max} \approx 700\text{Mbit/s}$ ):**



**Se  $T \ll \tau$  non si riesce più a trasmettere un segnale riconoscibile in arrivo**



tempo della rete. Frequenze maggiori hanno l'effetto, come si vede nel grafico, di non consentire mai all'uscita di superare una possibile soglia posta a metà dei livelli elettrici. Le reti RC pertanto, rallentando i fronti dei segnali, determinano una limitazione alla massima frequenza del segnale che può essere trasmesso.

Per potere trasmettere a velocità maggiori, si deve:

- **ridurre R**: per ridurre R bisogna progettare i circuiti al meglio in modo che la loro resistenza equivalente Thevenin sia piccola; questo non sempre è possibile perché possono intervenire altri vincoli progettuali che impediscono di scendere sotto un dato valore.
- **ridurre C** in modo da ridurre la costante di tempo di risposta del circuito: ridurre C vuole dire in primo luogo ridurre le dimensioni L (lunghezza) e W (larghezza) delle piste. La soluzione più efficace di questo problema è passare da un collegamento su scheda ad un microcollegamento realizzando tutti i circuiti in forma integrata sullo stesso chip di silicio.

In questo caso le dimensioni dei percorsi delle connessioni sono enormemente più piccole: tipici valori sono  $L=100\mu\text{m}$  e  $W=1\mu\text{m}$ , uno spessore dell'isolante  $X_d$  di soli  $0.1 \mu\text{m}$  (isolante normalmente di ossido di silicio  $\text{SiO}_2$  con una costante dielettrica di  $0.3 \cdot 10^{-12} \text{ F/cm}$ ). Tale realizzazione abbassa di colpo i valori delle capacità in gioco a valori di qualche decina di fF e consente la trasmissione di segnali digitali aventi bit/rate di parecchi Gbit/s.



Per ridurre ulteriormente il valore delle capacità dei collegamenti, ci sono aperte ancora due strade. La prima consiste naturalmente nella ulteriore riduzione delle dimensioni dei collegamenti. Il problema che nasce dall'adottare questa soluzione sulla tradizionale tecnologia al silicio con piste di alluminio è che *ad una riduzione delle piste di collegamento corrisponde un aumento della resistenza* della pista stessa. Questo fatto può risultare controproducente sul circuito, perché va ad aumentare le resistenze in gioco. La soluzione in avanzata fase di sperimentazione è quella di **usare il rame** come materiale

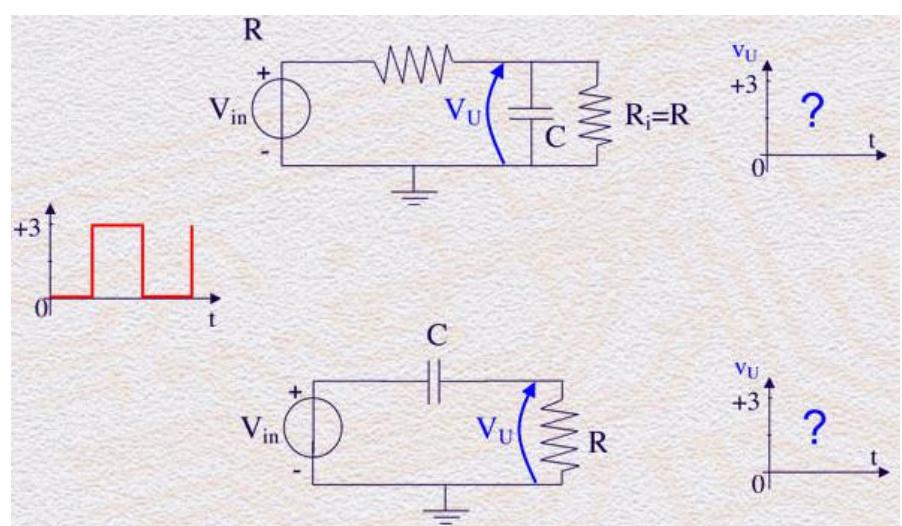


per la realizzazione delle piste, migliore conduttore della corrente rispetto all'alluminio. Per fare ciò si sono dovute superare le difficoltà legate alla incompatibilità del rame con il silicio attraverso una tecnologia nuova particolarmente raffinata.

Il secondo elemento su cui intervenire per ridurre la capacità del collegamento è la riduzione della costante dielettrica dell'isolante. A tal fine si è inventata una tecnica costruttiva raffinatissima per i microchip al silicio con cui si riesce, dopo avere realizzato le piste metalliche di connessione di rame, a togliere tutto il materiale di supporto in modo che le piste stesse vengano a trovarsi sospese in aria. Poiché **l'aria ha il valore di e minore possibile** esso è il risultato limite sotto cui non si potrà più andare.

Queste migliorie portano a realizzare circuiti con connessioni dalle capacità inferiori al fF che possono quindi gestire segnali anche di qualche centinaio di Gbit/s.

Da ultimo, sulla base delle considerazioni intuitive fatte nell'indagare il comportamento della rete RC, valutare con precisione la risposta nel tempo dei due circuiti qui riportati quando sollecitati da un segnale di ingresso digitale.



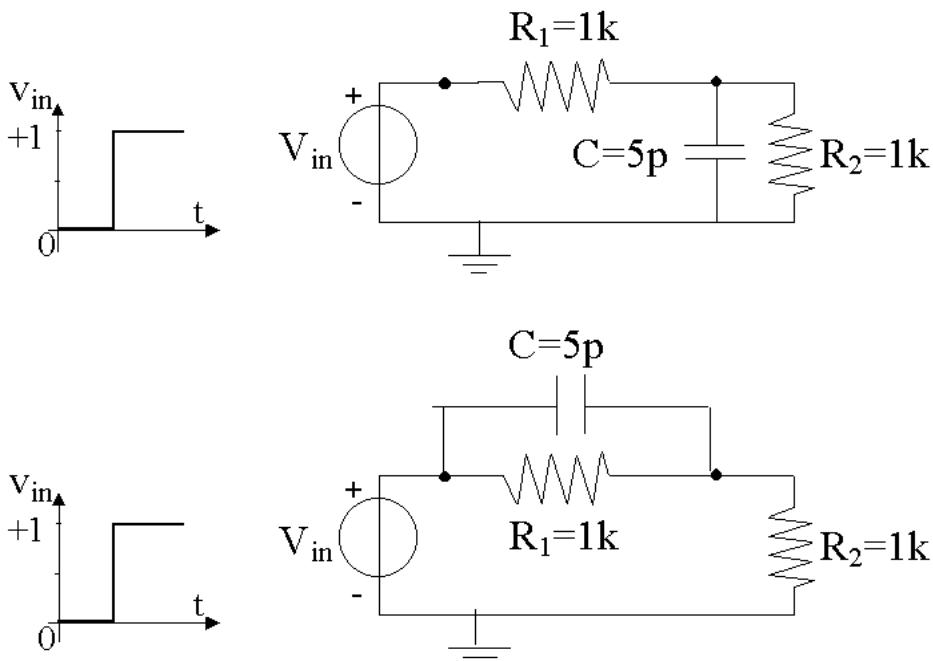
## Esercizio 1

Per diminuire la capacità di un collegamento tra due circuiti integrati posti su una piastra di un circuito stampato, quali delle seguenti cose fareste :

- diminuire lo spessore della piastra;
- diminuire la lunghezza del collegamento;
- cambiare il materiale con cui è realizzata la piastra (Quale scegliereste ? );
- diminuire la frequenza del segnale.

## Esercizio 2

Disegnare l'andamento nel tempo della risposta ad un gradino di tensione ampio 1V dei seguenti circuiti e calcolarne la costante di tempo ed il tempo di salita:



## Lezione n. 5 : I materiali semiconduttori per l'elettronica

### Il cristallo di silicio

Molti sono stati i fattori che hanno permesso lo sviluppo dell'elettronica come la conosciamo ai giorni nostri. Ma certamente il più importante è stato, ed è tuttora, l'esistenza di un particolare materiale naturale, il **SILICIO**, che presenta tutte quelle caratteristiche desiderate dagli scienziati che per primi, negli anni appena successivi alla seconda guerra mondiale, hanno pensato di costruire dispositivi elettronici a stato solido.

Vedremo in queste lezioni quali sono le proprietà più importanti del silicio e come queste caratteristiche abbiano permesso l'enorme sviluppo dell'elettronica integrata negli ultimi decenni.

Nelle prossime 2 lezioni, propedeutiche alla comprensione del funzionamento dei dispositivi elettronici utilizzati nei circuiti, ci proponiamo di capire sinteticamente:

- come è fatto un **cristallo** semiconduttore di Silicio
- che cosa significhi "materiale **semiconduttore**",
- quali **proprietà** manifestino questi materiali,
- in quali modi e sotto quali condizioni è possibile un passaggio di **corrente** al loro interno.

Come accennato avremo come protagonista il silicio che, nella sua forma cristallina, è il materiale semiconduttore ancora oggi di gran lunga più utilizzato per produrre dispositivi elettronici; e vedremo presto il perché di questa superiorità.

Cerchiamo di capire innanzitutto come sia fatto un cristallo di silicio e perché si dice che è un semiconduttore.

Il silicio è un elemento naturale che ha **numero atomico Z=14**, cioè ha un nucleo costituito da 14 protoni a cui, come noto, si attribuisce carica positiva, "attorniato" in un atomo neutro da 14 elettroni a cui si attribuisce carica negativa.

Aiuta in questo momento visualizzare i 14 elettroni come orbitanti attorno al nucleo e ad esso legati dalla attrazione coulombiana che le cariche positive del nucleo esercitano sulle cariche elettriche negative degli elettroni.

Gli elettroni stanno su "orbite" differenti, vale a dire subiscono forze attrattive differenti, maggiori per gli elettroni delle orbite più interne e minori per quelle più esterne.

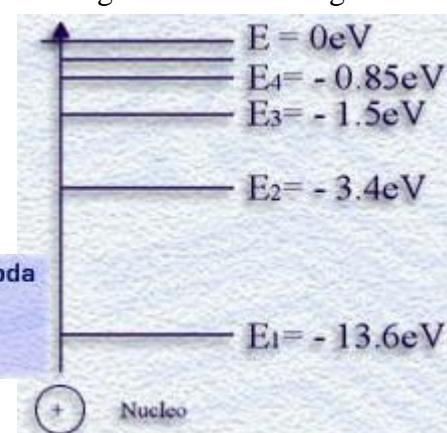
Per questo si dice che gli elettroni più interni hanno maggiore energia, intendendo che bisognerebbe spendere maggiore energia se si volesse liberarli dalla loro posizione nell'atomo e portarli ad essere elettroni liberi nello spazio.

Molti scienziati, nei primi decenni del XX secolo, hanno studiato in dettaglio le forze di legame tra elettroni e nuclei (la cosiddetta Fisica atomica quantistica) arrivando a concludere che gli elettroni possono assumere solamente ben determinate energie all'interno di un atomo.

Nella nostra immagine di elettroni "orbitanti" attorno al nucleo, questo equivale a dire che gli elettroni possono stare solo su alcune ben definite orbite o "**livelli energetici**". Le varie orbite possibili sono distanziate tra di loro di una certa quantità di energia, intendendo che un elettrone per passare da una all'altra deve ricevere o dare una ben definita quantità di energia. Su ogni livello energetico può stare un solo elettrone.

IA	IIA	IIIA	IVA	VA	VIA	VIIA	VIIIA	IB	IIB	IIIIB	IVB	VB	VIIB	VIB	VIIIB	VIII									
1 H Hydrogen 1.008	2 He Helium 4.003	3 Li Lithium 6.941	4 Be Boron 9.0128	5 B Boron 10.811	6 C Carbon 12.011	7 N Nitrogen 14.0074	8 O Oxygen 15.9994	9 F Fluorine 18.9984	10 Ne Neon 20.1797	11 Na Sodium 22.990	12 Mg Magnesium 24.306	13 Al Aluminum 26.9815	14 Si Silicon 28.0855	15 P Phosphorus 30.9738	16 S Sulfur 32.0656	17 Cl Chlorine 35.4527	18 Ar Argon 39.948								
11 Na Sodium 22.990	12 Mg Magnesium 24.306	13 Al Aluminum 26.9815	14 Si Silicon 28.0855	15 P Phosphorus 30.9738	16 S Sulfur 32.0656	17 Cl Chlorine 35.4527	18 Ar Argon 39.948	19 K Potassium 39.098	20 Ca Calcium 40.078	21 Sc Scandium 44.956	22 Ti Titanium 47.947	23 V Vanadium 50.9415	24 Cr Chromium 51.9963	25 Mn Manganese 54.9381	26 Fe Iron 55.847	27 Co Cobalt 58.9321	28 Ni Nickel 58.9354	29 Cu Copper 63.546	30 Zn Zinc 65.409	31 Ga Gallium 69.723	32 Ge Germanium 72.610	33 As Arsenic 74.9216	34 Se Selenium 78.960	35 Br Bromine 79.904	36 Kr Krypton 83.800
19 K Potassium 39.098	20 Ca Calcium 40.078	21 Sc Scandium 44.956	22 Ti Titanium 47.947	23 V Vanadium 50.9415	24 Cr Chromium 51.9963	25 Mn Manganese 54.9381	26 Fe Iron 55.847	27 Co Cobalt 58.9321	28 Ni Nickel 58.9354	29 Cu Copper 63.546	30 Zn Zinc 65.409	31 Ga Gallium 69.723	32 Ge Germanium 72.610	33 As Arsenic 74.9216	34 Se Selenium 78.960	35 Br Bromine 79.904	36 Kr Krypton 83.800								
37 Rb Rubidium 85.4678	38 Sr Strontium 87.62	39 Y Yttrium 88.9088	40 Zr Zirconium 91.224	41 Nb Niobium 92.9068	42 Mo Molybdenum 95.94	43 Tc Technetium 98.901	44 Ru Ruthenium 101.07	45 Rh Rhodium 102.905	46 Pd Palladium 106.42	47 Ag Silver 107.869	48 Cd Cadmium 114.811	49 In Indium 115.751	50 Sn Tin 118.710	51 Sb Antimony 121.767	52 Te Tellurium 127.661	53 I Iodine 126.9047	54 Xe Xenon 131.29								

L'unità di misura dell'energia comoda in elettronica è l'**elettron Volt, eV**:  
 $1 \text{ eV} = 1.6 \cdot 10^{-19} \text{ joules}$



Quando l'atomo è mantenuto a temperature basse, addirittura prossime alla temperatura dello zero assoluto ( $T=0$  K), NON c'è energia disponibile e l'elettrone occupa il livello energetico più vicino al nucleo.

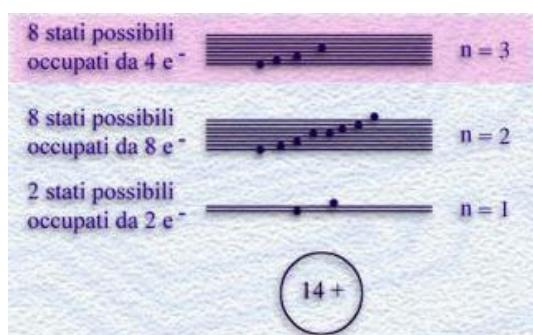
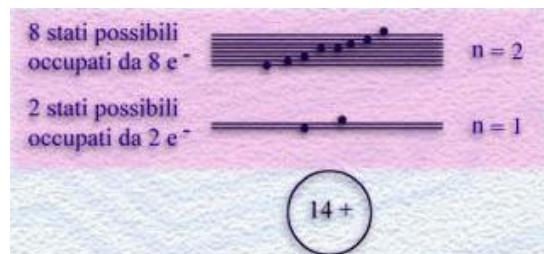
Solo fornendo energia all'elettrone almeno pari alla differenza con il livello successivo, esso può passare da un livello ad un altro.

Questo può essere fatto ad esempio scaldando l'atomo.

Ai fini del corso è utile notare che l'ordine di grandezza della distanza tra livelli energetici è, al più, di qualche *elettronVolt*.

Nel caso del **silicio** ( $Z=14$ ), l'analisi della disposizione energetica degli elettroni porta ad un risultato relativamente semplice da visualizzare nei suoi aspetti più importanti.

Dieci dei 14 elettroni occupano tutti i livelli energetici disponibili più vicini al nucleo. I legami con il nucleo sono per questi elettroni così forti che essi non vengono sostanzialmente perturbati dalle normali reazioni chimiche o dalle forze tra atomi contigui in strutture a più atomi.

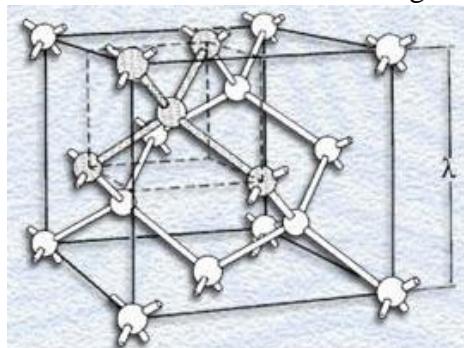


I rimanenti 4 elettroni del silicio, invece, occupano 4 "orbite" o livelli energetici più esterni. Essi sono debolmente attratti dal campo del loro proprio nucleo perché più lontani e perché la carica positiva di quest'ultimo è in parte schermata dagli altri 10 elettroni.

Questi 4 elettroni sono chiamati **elettroni di valenza** e partecipano attivamente ai legami con altri atomi.

## IL SILICIO : STRUTTURA CRISTALLINA

I 4 elettroni di valenza danno luogo a 4 **legami covalenti** con 4 altri atomi di silicio primi vicini. La disposizione nello spazio di questi legami è perfettamente simmetrica, con gli atomi disposti ai vertici di un **tetraedro**. Tetraedri contigui si legano a formare il **cristallo**, come mostrato nella figura.

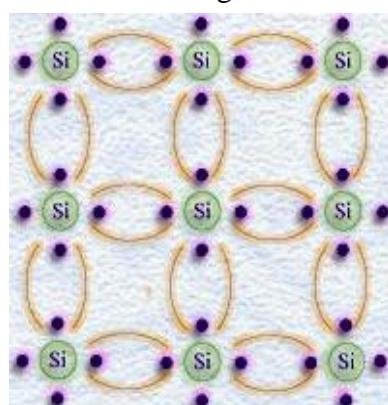


La struttura cristallina è molto compatta, con gli atomi distanti circa 5 Å uno dall'altro. Ne consegue che la densità di atomi di silicio nel cristallo è pari a  $5 \times 10^{22}$  atomi/cm<sup>3</sup>.

In molti casi una rappresentazione bidimensionale del cristallo è più comoda di una rappresentazione tridimensionale. Essa può consistere semplicemente nella **proiezione sul piano** della disposizione spaziale vera tridimensionale. Nel caso del silicio, essa dà luogo alla semplice struttura a scacchiera con quattro legami ad angolo retto.

Il disegno bidimensionale visualizza bene come ogni atomo di silicio spartisca i suoi 4 elettroni di valenza ognuno con uno dei 4 atomi primi vicini. Ogni cerchio rappresenta il nucleo ed i 10 elettroni più interni di ogni atomo, i pallini scuri rappresentano gli elettroni di valenza, ogni linea rappresenta il legame covalente con un ben definito atomo contiguo. Le linee che si dipartono da ogni cerchio sono otto perché ogni atomo rende disponibile i suoi 4 elettroni di valenza ma anche riceve un elettrone di valenza da ognuno degli atomi vicini.

Alla temperatura dello zero assoluto (cioè a -273°C), tutti i 4 elettroni di valenza del silicio sono impegnati a formare legami covalenti con i 4 atomi primi vicini. Non esistono elettroni liberi per condurre elettricità ed il cristallo si comporta da isolante.



## Elettroni e lacune

A temperature superiori, ed in particolare alla temperatura ambiente, (circa 27°C, ossia 300K), gli atomi acquisiscono energia termica ed alcuni legami covalenti si spezzano dando luogo a degli elettroni liberi.

L'entità dell'energia necessaria per rompere il legame covalente è di poco più di un elettronVolt nel caso del silicio, esattamente **1.1 eV**.

Dove prima stava l'elettrone si crea una mancanza dell'elettrone di legame, chiamato **lacuna**.

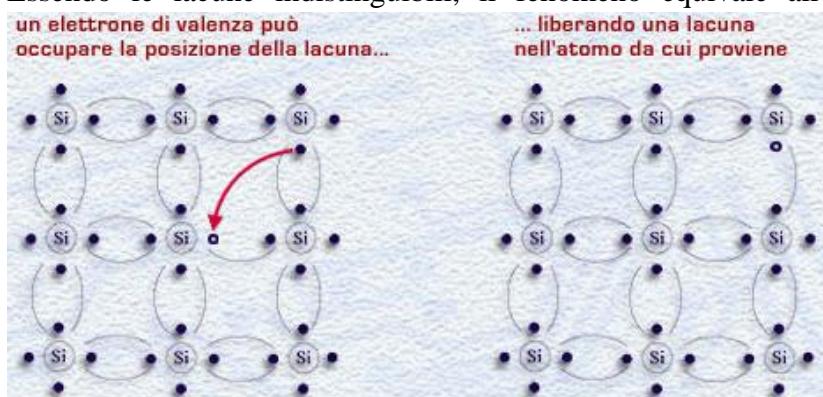
C'è stato quindi un processo di creazione contemporanea di un elettrone libero e di una lacuna nel cristallo di silicio, che comunque rimane neutro.

## CONDUZIONE DELL'ELETTRONE

Applicando un campo elettrico dall'esterno gli elettroni liberi acquistano una quantità di moto netta nella direzione dettata dal campo e danno luogo ad una corrente elettrica, che chiamiamo **corrente di elettroni liberi nel cristallo**.

## CONDUZIONE DELLA LACUNA

Un elettrone di valenza vicino al legame rotto può saltare nella posizione della lacuna, poiché la nuova posizione è per l'elettrone energeticamente identica alla sua attuale. Infatti, l'elettrone di valenza lascerebbe un legame covalente con un atomo di silicio per andare a saturare un legame covalente con un altro atomo di silicio. Questo spostamento fa sparire la lacuna nella posizione da cui era partito l'elettrone di conduzione ma fa nascere una lacuna attorno all'atomo da cui è partito l'elettrone di valenza. Essendo le lacune indistinguibili, il fenomeno equivale allo



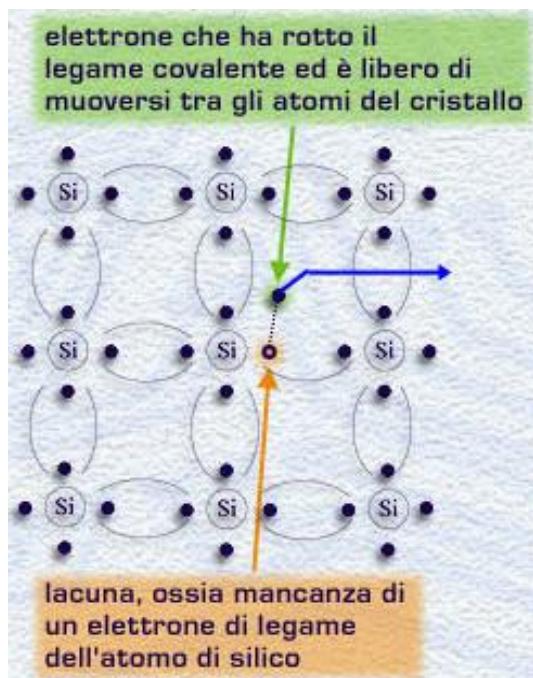
E' importante notare che ***l'elettrone libero che ha originato la lacuna non interviene nel processo di movimento di quest'ultima, e va per la sua strada, libero in maniera indipendente dalle lacune.***

In un semiconduttore come il silicio, quindi, la conduzione di carica è realizzata *contemporaneamente ed indipendentemente* da:

- **elettroni liberi** (o di conduzione) che si muovono "tra" gli atomi e quindi "tra" i legami covalenti che gli atomi di silicio stabiliscono tra di loro nel cristallo.
- **elettroni di valenza** che si muovono "lungo" i legami covalenti che tengono uniti gli atomi di silicio. Quest'ultimo meccanismo di trasporto è normalmente visualizzato dallo spostamento, in direzione opposta, delle rispettive lacune.

Quando si applica un campo elettrico, la corrente prodotta dal moto degli elettroni di valenza deve sommarsi a quella degli elettroni di conduzione, perché hanno lo stesso segno e sono sottoposti allo stesso campo. La corrente degli elettroni di valenza può anche essere espressa come moto delle corrispondenti *lacune*, purché queste siano immaginate avere *carica positiva* (in modulo *uguale a quella dell'elettrone*) perché muoventesi in senso opposto.

Gli elettroni liberi e le lacune sono anche chiamati **portatori** della corrente.



## CONCENTRAZIONE DI ELETTRONI DI CONDUZIONE E DI LACUNE IN UN SEMICONDUTTORE PURO

In un **cristallo puro di silicio**, poiché i portatori sono dovuti alla generazione di *coppie* elettrone-lacuna per eccitazione termica degli atomi di silicio, la densità di elettroni, **[n]**, espressa in [elettroni/cm<sup>3</sup> ], è quindi uguale alla densità di lacune, **[p]**, [lacune/cm<sup>3</sup> ]: **n = p**.

Il loro valore dipende dalla temperatura T, cioè dall'energia disponibile per ionizzare l'atomo di semiconduttore e dal valore della forza del legame.

Nel silicio alla temperatura ambiente (T=300K):

$$\mathbf{n = p = 1.45 \times 10^{10} \text{ [portatori/cm}^3\text{]}}$$

cioè 14 miliardi di elettroni e lacune.

Questo valore, benché grande in assoluto, è piccolo se confrontato con il numero di legami che si sarebbero potuti rompere, pari alla densità degli atomi di silicio che, come visto, è pari a  $5 \times 10^{22}$  atomi/cm<sup>3</sup> (50 mila miliardi di miliardi !).

## Materiali semiconduttori

A questo punto siamo in grado di catalogare più precisamente il cristallo di silicio come un **cristallo semiconduttore**.

Infatti, quando un materiale, alla *temperatura di interesse*, ha una modesta frazione di legami covalenti rotti e dà luogo sia ad elettroni che a lacune libere di muoversi e produrre corrente, esso viene chiamato **MATERIALE SEMICONDUTTORE**.

Esso si differenzia dagli isolanti o dai conduttori.

Un **ISOLANTE**, come il diamante o la ceramica (SiO<sub>2</sub>), non ha elettroni né lacune per la conduzione di corrente perché i legami tra gli atomi contigui sono così forti che nessuno è spezzato.

Un **CONDUTTORE**, come il rame o l'alluminio, è formato da atomi che perdono tutti un elettrone, fornendo quindi un numero di elettroni liberi per la conduzione pari al numero di atomi presenti e senza generare lacune.

La fortuna del silicio è anche legata alla facilità di realizzarvi un ossido dalle ottime proprietà isolanti, l'ossido di silicio (SiO<sub>2</sub>), e dall'essere compatibile con un ottimo conduttore come l'alluminio. L'avere a disposizione tutti e 3 questi materiali, un isolante, un semiconduttore ed un conduttore con tecnologie affini e compatibili, ha dato agli scienziati ed ai progettisti di circuiti tali gradi di libertà da permettere loro di realizzare dispositivi originali ed innovativi in grande quantità.

## ALTRI MATERIALI SEMICONDUTTORI

Il silicio non è l'unico materiale semiconduttore esistente. Anche altri elementi della IV colonna della tabella periodica possono manifestare proprietà da semiconduttore alla temperatura ambiente, oltre che alcuni composti formati dagli elementi a cavallo di questa.

In particolare, altri materiali semiconduttori utilizzati in elettronica sono:

- germanio
- i composti di 2 elementi delle colonne III-V : GaAs, InP...
- i composti di 3 (ternari) o di 4 (quaternari) elementi : GaAlAs, GaInAsP....
- i composti II-VI : CdTe ...

Eccetto che per l'arseniuro di gallio (GaAs) con il quale sono realizzati i transistori utilizzati nei circuiti operanti a frequenze molto elevate (>10 GHz), gli altri materiali trovano impiego in applicazioni specifiche e costose, nella maggior parte dei casi nell'ambito della optoelettronica.

La loro minore fortuna è data dalla maggiore difficoltà di fabbricazione e quindi dal maggior costo, dalla mancanza di isolanti compatibili tecnologicamente o da altre proprietà elettroniche meno interessanti.

IIIB	IVB	VB	VIB
5 <sup>3p<sub>1/2</sub></sup> <sub>1/2</sub> <b>B</b> Boron 10.811 (1s <sup>2</sup> ) <sup>2p</sup> 8.2980	6 <sup>3p<sub>1/2</sub></sup> <sub>1/2</sub> <b>C</b> Carbon 12.0107 (1s <sup>2</sup> ) <sup>2p</sup> 11.2603	7 <sup>3s<sub>1/2</sub></sup> <sub>1/2</sub> <b>N</b> Nitrogen 14.00674 (1s <sup>2</sup> ) <sup>2p</sup> 14.5341	8 <sup>3p<sub>1/2</sub></sup> <sub>1/2</sub> <b>O</b> Oxygen 16.9994 (1s <sup>2</sup> ) <sup>2p</sup> 13.8181
13 <sup>3p<sub>1/2</sub></sup> <sub>1/2</sub> <b>Al</b> Aluminum 26.98154 (1s <sup>2</sup> ) <sup>2p</sup> 8.9858	14 <sup>3p<sub>1/2</sub></sup> <sub>1/2</sub> <b>Si</b> Silicon 28.0855 (1s <sup>2</sup> ) <sup>2p</sup> 8.1517	15 <sup>3s<sub>1/2</sub></sup> <sub>1/2</sub> <b>P</b> Phosphorus 30.97376 (1s <sup>2</sup> ) <sup>2p</sup> 16.4867	16 <sup>3p<sub>1/2</sub></sup> <sub>1/2</sub> <b>S</b> Sulfur 32.066 (1s <sup>2</sup> ) <sup>2p</sup> 19.3030
30 <sup>3s<sub>1/2</sub></sup> <sub>1/2</sub> <b>Zn</b> Zinc 65.39 (1s <sup>2</sup> ) <sup>2p</sup> 8.9942	31 <sup>3p<sub>1/2</sub></sup> <sub>1/2</sub> <b>Ga</b> Gallium 69.729 (1s <sup>2</sup> ) <sup>2p</sup> 8.9993	32 <sup>3p<sub>1/2</sub></sup> <sub>1/2</sub> <b>Ge</b> Germanium 72.61 (1s <sup>2</sup> ) <sup>2p</sup> 7.8934	33 <sup>3s<sub>1/2</sub></sup> <sub>1/2</sub> <b>As</b> Arsenic 74.02160 (1s <sup>2</sup> ) <sup>2p</sup> 9.7865
48 <sup>3s<sub>1/2</sub></sup> <sub>1/2</sub> <b>Cd</b> Cadmium 112.411 (1s <sup>2</sup> ) <sup>2p</sup> 8.9958	49 <sup>3p<sub>1/2</sub></sup> <sub>1/2</sub> <b>In</b> Indium 114.818 (1s <sup>2</sup> ) <sup>2p</sup> 5.7984	50 <sup>3p<sub>1/2</sub></sup> <sub>1/2</sub> <b>Sn</b> Tin 116.715 (1s <sup>2</sup> ) <sup>2p</sup> 7.3430	51 <sup>3s<sub>1/2</sub></sup> <sub>1/2</sub> <b>Sb</b> Antimony 121.766 (1s <sup>2</sup> ) <sup>2p</sup> 8.0084
			52 <sup>3p<sub>1/2</sub></sup> <sub>1/2</sub> <b>Te</b> Tellurium 127.80 (1s <sup>2</sup> ) <sup>2p</sup> 9.9324

## **Complementi**

Provate ad andare nel seguente sito dove potete "prendere in mano" un cristallo di silicio e ruotarlo a vostro piacimento per osservarne la struttura da diverse angolazioni:

<http://jas2.eng.buffalo.edu/applets/education/solid/unitCell/home.html>

Verificate che esiste un punto di vista in cui il cristallo assume esattamente la disposizione bidimensionale a scacchiera che abbiamo introdotto nella lezione.

### **Esercizio 1**

Se si aumenta la temperatura di un cristallo di silicio puro, il numero di lacune per  $\text{cm}^3$  aumenta o diminuisce ? E quello degli elettroni ?

### **Esercizio 2**

Il diamante è stato introdotto in questa lezione come un isolante. In quale situazione manifesterebbe invece proprietà semiconduttrici ?

## Lezione n. 6 : I portatori e la corrente nei dispositivi semiconduttori

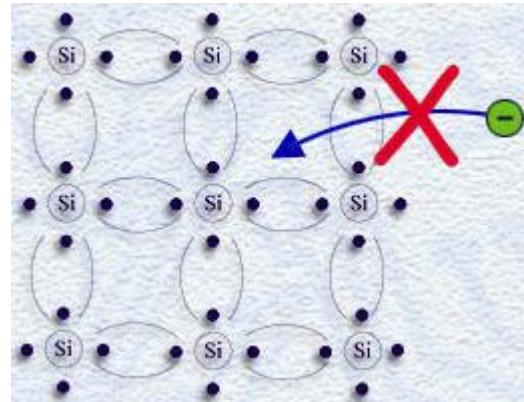
### Drogaggio di tipo N

#### **COME SI PUO' VARIARE LA CONCENTRAZIONE DI $n$ e $p$ ?**

Dopo avere visto nella precedente lezione che in un cristallo puro di silicio vi è una certa concentrazione di elettroni e di lacune libere di muoversi, sarebbe certamente interessante poterne variare a piacimento il numero così da poter pensare di modificare e controllare la corrispondente corrente elettrica circolante.

Il numero di portatori in un semiconduttore (elettroni e/o lacune) **NON** può essere aumentato semplicemente aggiungendone dall'esterno perché il cristallo si caricherebbe elettrostaticamente ed assumerebbe un potenziale tale da renderlo "intoccabile" a meno di scosse elettriche !

Neppure si può pensare di variare la concentrazione di elettroni o lacune variando la temperatura del pezzo di cristallo, perché decisamente poco pratico !

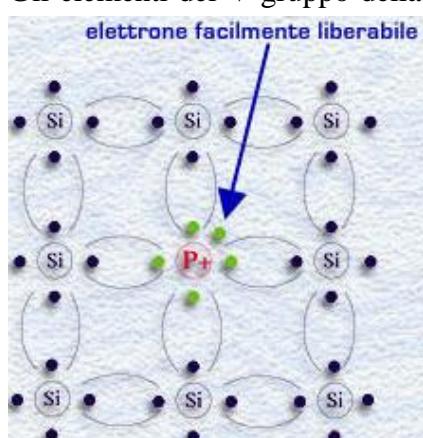


La soluzione al problema di come variare la concentrazione di  $n$  o di lacune  $p$  consiste nell'incorporare nel cristallo, in *sostituzione* ad alcuni atomi di silicio, atomi neutri di altri elementi.

Il processo tecnologico che realizza questa operazione è detto **drogaggio del semiconduttore**. Vediamo in particolare gli effetti di questa operazione.

### DROGANTI DEL V GRUPPO

Gli elementi del V gruppo della tabella periodica hanno 5 elettroni nei livelli più esterni. Quando uno



di questi atomi, normalmente il **fosforo**, prende il posto di un atomo di silicio, **un elettrone non partecipa al legame covalente** con i 4 atomi di silicio vicini.

Questo elettrone è quindi poco vincolato al cristallo, e l'energia per liberarlo dal suo atomo è molto minore di quella necessaria per rompere un legame covalente del silicio.

IIIB	IVB	VB	VIB
5 B Boron Y=0.11 1s <sup>2</sup> 2s <sup>2</sup> 2p <sup>1</sup> 8.2800	6 C Carbon Y=0.107 1s <sup>2</sup> 2s <sup>2</sup> 2p <sup>2</sup> 8.2800	7 N Nitrogen Y=0.074 1s <sup>2</sup> 2s <sup>2</sup> 2p <sup>3</sup> 10.2300	8 O Oxygen Y=0.064 1s <sup>2</sup> 2s <sup>2</sup> 2p <sup>4</sup> 10.2300
13 Al Aluminum Y=0.054 1s <sup>2</sup> 2s <sup>2</sup> 2p <sup>1</sup> 3s <sup>2</sup> 8.3800	14 Si Silicon Y=0.0508 1s <sup>2</sup> 2s <sup>2</sup> 2p <sup>2</sup> 3s <sup>2</sup> 8.3800	15 P Phosphorus Y=0.0376 1s <sup>2</sup> 2s <sup>2</sup> 2p <sup>4</sup> 3s <sup>2</sup> 10.4000	16 S Sulfur Y=0.0306 1s <sup>2</sup> 2s <sup>2</sup> 2p <sup>4</sup> 3s <sup>2</sup> 3p <sup>2</sup> 10.4000
31 Ga Gallium Y=0.029 1s <sup>2</sup> 2s <sup>2</sup> 2p <sup>1</sup> 3s <sup>2</sup> 3p <sup>1</sup> 8.3800	32 Ge Germanium Y=0.027 1s <sup>2</sup> 2s <sup>2</sup> 2p <sup>2</sup> 3s <sup>2</sup> 3p <sup>2</sup> 8.3800	33 As Arsenic Y=0.0216 1s <sup>2</sup> 2s <sup>2</sup> 2p <sup>4</sup> 3s <sup>2</sup> 3p <sup>3</sup> 10.4000	34 Se Selenium Y=0.016 1s <sup>2</sup> 2s <sup>2</sup> 2p <sup>4</sup> 3s <sup>2</sup> 3p <sup>4</sup> 10.4000

La sola agitazione termica presente nel cristallo di silicio a temperatura ambiente è sufficiente per ionizzare l'atomo di drogante (per il fosforo nel silicio, l'energia è di soli 44meV contro gli 1.1eV del silicio puro) e quindi rendere disponibile nel cristallo il suo elettrone, libero quindi di condurre corrente.

**Localmente**, associato all'atomo di drogante ionizzato, si ha una carica positiva, **globalmente** compensata, in qualche punto del cristallo, dal suo elettrone libero. Il cristallo, nel suo complesso, rimane quindi **neutro**.

Questo avviene, a temperatura ambiente ( $T=300K$ ), per tutti gli atomi di fosforo presenti, che quindi rendono disponibili nel cristallo un numero corrispondente di elettroni.

Gli **elementi del V gruppo della Tabella Periodica**, per la loro capacità di liberare un elettrone, sono detti atomi **donori**.

Il numero di donori per unità di volume [ $\text{cm}^3$ ] è detta **densità di donori**,  $N_D$  [atomi/ $\text{cm}^3$  ].

I processi tecnologici che realizzano il drogaggio del semiconduttore sono così sviluppati che è possibile controllare con ottima precisione la densità del drogante immesso e quindi anche la densità di elettroni liberi a disposizione.

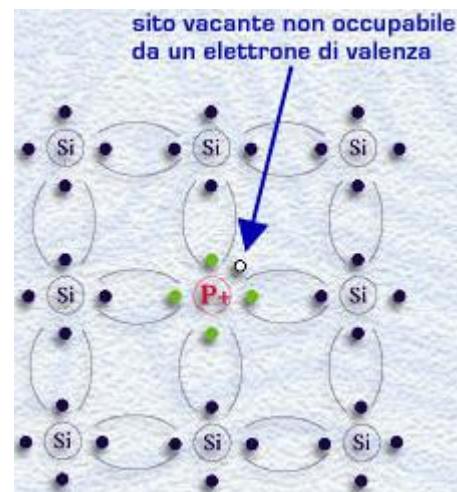
La ionizzazione dell'atomo di fosforo **NON** genera contemporaneamente anche una lacuna.

Infatti non c'è nessun elettrone vicino, tutti occupati in legami covalenti molto più forti, per il quale sia indifferente stare dove già è oppure andare ad occupare il legame lasciato libero dal quinto elettrone del donore. Quest'ultimo infatti, abbiamo detto, è di natura ben diversa da un legame covalente e molto più debole.

Con l'inserimento nel cristallo semiconduttore di atomi donori si può pertanto selettivamente aggiungere un numero voluto di elettroni liberi disponibili alla conduzione senza aggiungere un corrispondente numero di lacune.

In particolare, in un cristallo di silicio drogato con fosforo ci saranno:

- *elettroni generati dalla ionizzazione di tutti (a temperatura ambiente) gli atomi di drogante*
- *elettroni generati termicamente dalla rottura di alcuni legami covalenti del silicio*



Proprio per avere il controllo del numero di elettroni presenti, si fa in modo che i primi siano in numero prevalente.

Per cui basta drogare con  $N_D > 10^{12}$  atomi/cm<sup>3</sup> per avere  $n$  (*densità di elettroni liberi*) =  $N_D$  (*densità dei donori*).

Nello stesso cristallo ci saranno però anche delle lacune:

- *lacune generate termicamente dalla rottura di alcuni legami covalenti del silicio.*

Il loro numero è fissato da una relazione, che qui non dimostreremo, chiamata **legge di azione di massa**, che dice che in un semiconduttore all'equilibrio il prodotto tra la concentrazione degli elettroni e quella delle lacune è una costante secondo la relazione :

$$n \cdot p = 2 \cdot 10^{20} \text{ [portatori/cm}^3\text{]}$$

per cui quando si aumenta il numero di elettroni mediante l'aggiunta di atomi donori contemporaneamente diminuisce la concentrazione delle lacune. Nota l'una è immediatamente calcolabile l'altra.

Pertanto, il numero di lacune presenti in un cristallo di silicio drogato con  $N_D$  atomi di fosforo è pari a:

$$p = 2 \cdot 10^{20} / N_D \text{ [lacune/cm}^3\text{]}$$

un valore tanto più piccolo quanto più  $N_D$  è grande.

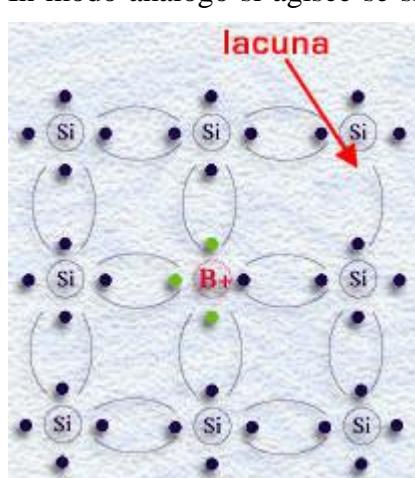
I semiconduttori drogati con elementi del V gruppo sono detti **semiconduttori di tipo n** perché i **portatori maggioritari** sono gli elettroni (indicati con n), ed i **portatori minoritari** sono le lacune (indicati con p).

### Drogaggio di tipo P

#### DROGANTI DEL III GRUPPO

In modo analogo si agisce se si vuole aumentare la concentrazione delle lacune. Gli elementi del III

gruppo della tabella periodica hanno solo 3 elettroni nei livelli più esterni. Quando uno di questi atomi, normalmente il **boro**, prende il posto di un atomo di silicio, i 3 elettroni di valenza saturano 3 dei 4 legami covalenti con gli atomi di silicio vicini. Il quarto legame non saturato esercita un grande richiamo sugli elettroni vicini impegnati nel legame covalente tra atomi di silicio. In termini energetici, per gli elettroni di legame è relativamente indifferente mantenere la loro posizione o cambiarla ed andare a colmare il legame mancante associato al drogante (in realtà c'è una piccola differenza tra la forza del legame



III	IV	V	VI
B Boro 12.0107 14.0119 16.0122 20.0125	C Carbonio 12.0107 14.0119 16.0122 20.0125	N Nitrizio 14.0107 15.0119 17.0122 21.0125	O Ossigeno 16.0107 17.0119 18.0122 20.0125
Al Alluminio 26.9815 28.9855 30.9876 32.9897	Si Silicio 28.9855 30.9876 32.9897 34.9917	P Fosforo 30.9876 32.9897 34.9917 36.9937	S Sulfuro 32.9897 34.9917 36.9937 38.9957
Ga Galio 69.7231 70.7251 71.7271 72.7291	Ge Germanio 69.7231 70.7251 71.7271 72.7291	As Antimonio 74.9216 75.9236 76.9256 77.9276	Se Selenio 78.9236 79.9256 80.9276 81.9296

non saturato dovuto alla presenza dell'atomo di drogante e quella del legame non saturato di un atomo di silicio privato dell'elettrone, l'atomo di silicio sarebbe carico positivamente ed eserciterebbe una azione maggiore sull'elettrone di quanto non possa fare il boro che invece è neutro). Nel caso del boro nel silicio, l'energia in più che bisogna fornire all'elettrone che è legato covalentemente al silicio affinché vada ad occupare il legame covalente associato al boro è, alla temperatura ambiente, ( $T=300K$ ), molto piccola e la sola energia termica presente nel cristallo è sufficiente per far fare questo salto all'elettrone verso l'atomo di boro. Come conseguenza di questo scambio, si è creata una **lacuna** nel semiconduttore, cioè un legame covalente non saturato associato ad un atomo di silicio.

**Localmente**, associato all'atomo di drogante, si ha una carica negativa, (e si parla di ionizzazione dell'atomo di drogante), **globalmente** compensata, in qualche punto del silicio, dalla lacuna creata. Il cristallo, nel suo complesso, rimane quindi sempre **neutro**.

Gli elementi del III gruppo, per la loro capacità di prendere un elettrone e liberare così una buca nel cristallo, sono detti **accettori**.

Il numero di accettori per unità di volume è detto **densità di accettori**,  $N_A$  [atomi/cm<sup>3</sup>].

Con l'inserimento nel semiconduttore di atomi accettori in sostituzione di alcuni atomi di silicio, si può pertanto selettivamente aggiungere un numero voluto di lacune libere disponibili alla conduzione senza aggiungere un corrispondente numero di elettroni.

In particolare, in un cristallo di silicio drogato con boro ci saranno:

- *lacune* generate dalla ionizzazione di tutti (a temperatura ambiente) gli atomi di drogante, (boro, alluminio, Gallio)
- *lacune* generate termicamente dalla rottura di alcuni legami covalenti del silicio.

Anche in questo caso, per avere il controllo del numero di lacune presenti, si fa in modo che le prime siano in numero prevalente drogando con  $N_A > 10^{12}$  [atomi/cm<sup>3</sup>]  $\Rightarrow p = N_A$  [lacune/cm<sup>3</sup>].

Nello stesso cristallo ci saranno però anche:

- *elettroni* generati termicamente dalla rottura di alcuni legami covalenti del silicio.

Il loro numero è fissato dalla **legge di azione di massa**,  $n \cdot p = 2 \cdot 10^{20}$  [portatori/cm<sup>3</sup>]<sup>2</sup>, ed è pari a:

$$n = 2 \cdot 10^{20} / N_A$$
 [elettroni/cm<sup>3</sup>]

un valore tanto più piccolo quanto più  $N_A$  è grande.

I semiconduttori di questo tipo sono chiamati **semiconduttori di tipo p** dove i **portatori maggioritari** sono le lacune, ed i **portatori minoritari** sono gli elettroni.

Lo sviluppo dell'elettronica si è basato, e si basa tutt'ora, sulla capacità di produrre silicio cristallino con una elevatissima purezza; ricordando che il numero di atomi di silicio per cm<sup>3</sup> è di circa  $5 \cdot 10^{22}$ , si arriva fino ad avere nel cristallo meno di  $10^{12}$  [atomi/cm<sup>3</sup>] di altri elementi, cioè un livello di impurezze di ~1 atomo su  $10^{10}$  atomi di silicio !!! Con nessun altro materiale si arriva a tanto.

In elettronica si tende a purificare il più possibile il silicio e poi a introdurre i droganti voluti. La **posizione** dei droganti nel reticolo in sostituzione degli atomi del silicio è **casuale**.

Le concentrazioni di drogante sono comunque piccole rispetto alla densità del silicio:

**densità di atomi di silicio:**  $5 \cdot 10^{22}$  atomi/cm<sup>3</sup>

**densità di atomi di drogante:**  $10^{12} < N_A, N_D < 10^{19}$  atomi/cm<sup>3</sup>

Riassumendo, drogando il silicio opportunamente riusciamo a controllare la concentrazione degli elettroni e delle lacune nel cristallo di silicio in modo che:

**Semiconduttore di tipo p**: le lacune sono **maggioritarie** e **pari alla densità del drogante** e gli elettroni sono **minoritari** e dettati dalla **legge di azione di massa**.

**Semiconduttore di tipo p** -  $N_A > 10^{12}$  droganti/cm<sup>3</sup>

$p = N_A$  lacune/cm<sup>3</sup>       $n = 2 \times 10^{20} / N_A$  elettroni/cm<sup>3</sup>

Lacune maggioritarie, elettroni minoritari.

**Semiconduttore di tipo n** gli elettroni sono **maggioritari** e **pari alla densità del drogante** e le lacune sono **minoritarie** e dettate dalla **legge di azione di massa**.

**Semiconduttore di tipo n -  $N_D > 10^{12}$  droganti/cm<sup>3</sup>**  
 **$n = N_D$  elettroni/cm<sup>3</sup>**     **$p = 2 \times 10^{20} / N_D$  lacune/cm<sup>3</sup>**  
**Elettroni maggioritari, lacune minoritarie.**

In entrambi i casi la densità di portatori maggioritari corrisponde praticamente alla densità di drogante, e la densità dei portatori minoritari segue la legge di azione di massa.

Nella pratica costruttiva dei dispositivi elettronici si è sempre in questa situazione, essendo i livelli di drogaggio normalmente utilizzati pari a  $10^{13} \div 10^{19}$  [atomi droganti/cm<sup>3</sup>].

Ad esempio, se si droga il silicio con boro (tipo **p**), con  $N_A = 5 \times 10^{15}$  [atomi/cm<sup>3</sup>], si avrà un materiale con densità di lacune pari a:  $p = 5 \times 10^{15}$  [lacune/cm<sup>3</sup>], e densità di elettroni pari a:  $n = 4 \times 10^4$  [elettroni/cm<sup>3</sup> ].

### Corrente di conduzione nei semiconduttori

#### COME SI PUO' FARE SCORRERE CORRENTE NEI SEMICONDUTTORI ?

Dopo avere visto che in un semiconduttore, ed in particolare nel silicio, è possibile controllare a piacimento la concentrazione ed il tipo dei portatori mediante la tecnica del drogaggio, vediamo ora come sia possibile controllare anche il loro moto all'interno del cristallo.

Gli elettroni liberi e le lacune in un cristallo di silicio all'equilibrio non sono ferme. Alla normale temperatura di utilizzo, c'è abbastanza energia da tenere costantemente in moto sia gli elettroni che le lacune in tutte le direzioni in maniera casuale.

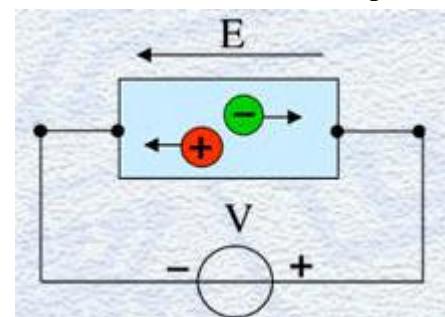
Nonostante che ogni carica che si muove produca una corrente elettrica, **il moto di tutte le cariche presenti in direzioni del tutto casuali produce correnti che si controbilanciano tutte**, per cui **nel cristallo non c'è una corrente netta circolante**.

Ora supponiamo di applicare un campo elettrico **E** al cristallo di silicio. Il moto termico casuale prima descritto continua a valere, ma ora ad esso si sovrappone l'effetto del campo elettrico. Concentriamoci quindi solo su quest'ultimo.

Le cariche mobili, elettroni e lacune, sentono una forza quando poste in un campo elettrico **E**, pari a :

$$\vec{F} = q \cdot \vec{E} = m \cdot \vec{a}, \text{ in cui } q \text{ è la carica dell'elettrone ed } E \text{ è l'intensità del campo elettrico, subendo una accelerazione pari a:}$$

$$\vec{a} = \frac{q}{m} \vec{E}, \text{ dove } m \text{ è un valore opportuno legato alla massa del portatore.}$$



Agli effetti del moto nel cristallo si considerano gli elettroni e le lacune come particelle libere.

Tuttavia, il potenziale periodico del reticolo di silicio e le vibrazioni del reticolo determinano delle forze deflettenti sul moto delle particelle. E' molto comodo rappresentare il moto delle particelle come soggette ad urti secchi contro gli atomi del cristallo, in corrispondenza dei quali la velocità si annulla e la nuova direzione è completamente casuale.

Se l'intervallo medio tra un urto e l'altro è **t**, la velocità media del portatore è dell'ordine di:

$$\vec{v} \approx \frac{q\tau}{m} \vec{E}$$

A causa quindi delle continue interazioni, la presenza di una forza sui portatori dà luogo ad una **velocità costante**, come in un moto viscoso.

Il fattore di proporzionalità tra velocità e campo elettrico è chiamato **mobilità**,  $m$  [cm<sup>2</sup>/Vs] del

portatore:  $\vec{v} \approx \frac{q\tau}{m} \vec{E}$         $\vec{v} \approx \mu \cdot \vec{E}$ .

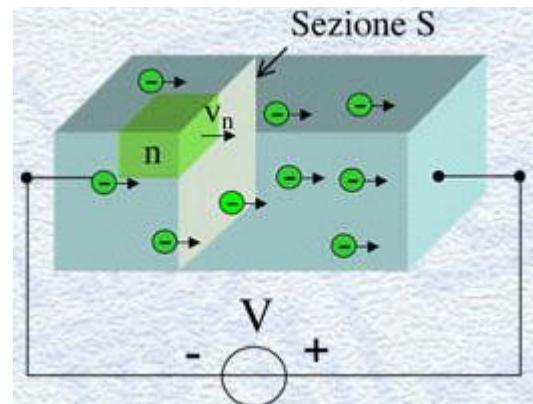
La mobilità può essere vista come un indice della facilità con cui i portatori si muovono nel cristallo. Ad esempio, un aumento del numero di collisioni, e quindi una diminuzione del tempo medio  $\tau$  tra di esse, diminuirebbe il valore della mobilità.

Data la loro diversa natura, gli elettroni di conduzione e gli elettroni di valenza (d'ora in avanti chiamate solo lacune) hanno in generale differenti mobilità, cioè differenti facilità a muoversi nel cristallo di silicio quando sottoposti a campi elettrici, pari a:

$$\mu_n \cong 1300 \frac{\text{cm}^2}{\text{Vs}} \quad \mu_p \cong 400 \frac{\text{cm}^2}{\text{Vs}}$$

Ciò che interessa per la comprensione dei dispositivi elettronici non è tanto il percorso di moto di un singolo portatore quanto il comportamento nel suo complesso di tutto il gran numero di portatori presenti nel cristallo soggetti ad un gran numero di interazioni. Questo può essere visualizzato come il comportamento di gruppo di tutti i portatori, aventi ognuno la velocità costante prima definita e muoventesi parallelamente (se lacune) o antiparallelamente (se elettroni) al campo elettrico esterno applicato.

In analogia ai flussi di fluidi viscosi, quindi, la corrente, definita come la quantità di carica trasportata attraverso una sezione S (normale al moto delle cariche stesse) nell'unità di tempo, può essere espressa da:



Essa è la somma di due componenti, una dovuta al moto degli elettroni, l'altra dovuta al moto delle lacune.

“n” e “p” sono infatti le concentrazioni (in un volume unitario) degli elettroni liberi e delle lacune, rispettivamente, e  $v_n$  e  $v_p$  sono le loro velocità.

I contributi delle lacune e degli elettroni si sommano perché lo stesso campo elettrico produce flussi in direzioni opposte di portatori con cariche opposte.

Tale corrente è chiamata **corrente di conduzione** (o di **deriva**) perché nasce dal trascinamento del portatore di corrente per effetto di un campo elettrico.

Nei semiconduttori drogati, uno dei due portatori è in netta prevalenza sull'altro (spesso di più di 10 ordini di grandezza !!!) per cui la corrente elettrica è praticamente dettata dal solo portatore maggioritario:

In un materiale di tipo *p* ( $p \gg n$ ), si ha:  $I \cong I_p = qp\mu_p E \cdot S$

In un materiale di tipo *n* ( $n \gg p$ ), si ha:  $I \cong I_n = qn\mu_n E \cdot S$

Ossia la corrente di deriva è portata essenzialmente dai maggioritari.

Si vede come la corrente sia proporzionale al campo elettrico applicato e sia tanto maggiore quanto più il materiale è drogato e quanto maggiore è la mobilità del portatore, oltre naturalmente a quanto maggiore è la sezione del dispositivo.

Proprio per separare le proprietà fisiche del materiale da quelle legate alle sue dimensioni, e quindi dare senso al confronto tra dispositivi diversi, si usa introdurre la **densità di corrente**,  $J$  [ $\text{A}/\text{cm}^2$ ]. Essa è definita come la corrente che attraversa una sezione di un  $\text{cm}^2$  di semiconduttore.

Dividendo quindi la corrente  $I$  per la superficie  $S$  espressa in  $\text{cm}^2$ , si ha l'espressione della densità di corrente per semiconduttori di tipo *n* e *p*:

$$\frac{I}{S} = J \left[ \frac{\text{A}}{\text{cm}^2} \right]$$

$$J_n = q \cdot n \cdot \mu_n \cdot E \quad J_p = q \cdot p \cdot \mu_p \cdot E$$

$\sigma_n$       **conducibilità**       $\sigma_p$

Il fattore di proporzionalità tra il campo elettrico  $E$  e la densità di corrente  $J$  è detto **conducibilità**,  $\sigma$ , del semiconduttore:

$$\sigma_n = qn\mu_n \quad o \quad \sigma_p = qp\mu_p$$

Nel silicio, gli elettroni hanno una mobilità maggiore, circa  $1300 \text{ cm}^2/\text{Vs}$ , di quella delle lacune, circa  $400 \text{ cm}^2/\text{Vs}$ , per cui i materiali drogati  $n$  offrono conducibilità più alte a pari droggaggio, e per questo vengono spesso preferiti nella realizzazione dei dispositivi elettronici.

**Nel silicio, la mobilità degli elettroni è maggiore di quella delle lacune**

$$\mu_n \approx 1300 \frac{\text{cm}^2}{\text{Vs}} \quad \mu_p \approx 400 \frac{\text{cm}^2}{\text{Vs}}$$

Il reciproco della conducibilità è chiamato **resistività**,  $r = 1/\sigma [\Omega \cdot \text{cm}]$ , del materiale:

$\rho \equiv \frac{1}{qn\mu_n} \quad \text{per materiale n},$	$\rho \equiv \frac{1}{qp\mu_p} \quad \text{per materiale p}$
---	--

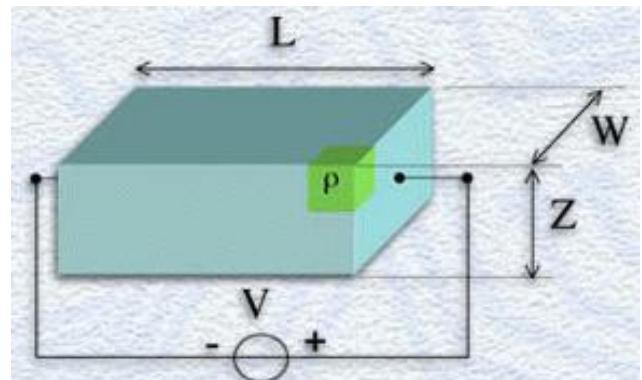
Essa corrisponde al valore di resistenza di un cubetto di semiconduttore di lato unitario, ed è quindi indipendente dalle dimensioni del dispositivo. È per questo motivo che viene introdotta ed utilizzata, e permettere il corretto confronto tra materiali diversi.

Dalla resistività si giunge alla resistenza di un dispositivo lungo  $L$  e di sezione  $S=W \times Z$  tramite la relazione:

$$R = \rho \frac{L}{W \cdot Z} \quad [\Omega]$$

Ricordando l'espressione della densità di corrente di deriva prima trovata in una forma indipendente dalle dimensioni del materiale:

$$J = \sigma E = \frac{1}{\rho} E$$



si giunge alla legge di Ohm, dipendente dalle dimensioni del dispositivo:

$$\text{Legge di Ohm: } I = \frac{V}{R} .$$

### Esercizio 1

- Dopo avere drogato un cristallo di silicio, tenuto alla temperatura ambiente, con  $5 \cdot 10^{16}$  atomi/cm<sup>3</sup> di fosforo, quanti elettroni liberi e quante lacune libere ci sono nel cristallo?
- Se lo stesso cristallo di silicio fosse tenuto a temperature bassissime, vicine a 0K, cosa succederebbe? Quanti elettroni e quante lacune ci sarebbero?
- E se lo stesso cristallo fosse scaldato a temperature molto alte (ad esempio di parecchie centinaia di gradi centigradi)?

### Esercizio 2

- Dopo avere drogato un cristallo di silicio, tenuto alla temperatura ambiente, con  $6 \cdot 10^{17}$  atomi/cm<sup>3</sup> di boro, quanti elettroni liberi e quante lacune libere ci sono nel cristallo?
- Se si drogasse contemporaneamente lo stesso cristallo di silicio sia con  $6 \cdot 10^{17}$  atomi/cm<sup>3</sup> di boro che con  $5 \cdot 10^{16}$  atomi/cm<sup>3</sup> di fosforo, quanti elettroni liberi e quante lacune libere ci sarebbero nel cristallo?

### Esercizio 3

- Dato un blocchetto di silicio drogato  $N_D = 10^{15}$  atomi/cm<sup>3</sup> di dimensioni  $W = 10\mu m$ ,  $Z = 0.1\mu m$  e  $L = 1\mu m$ , calcolare la corrente di elettroni e quella di lacune quando è applicata una tensione di 1V ai suoi capi.
- Se il blocchetto fosse stato drogato  $N_A = 10^{15}$  atomi/cm<sup>3</sup>, la corrente totale sarebbe stata uguale a prima o no? Perché?

## Lezione n. 7 : Il MOSFET come resistore variabile

### Struttura di un MOSFET

Nella figura vediamo la struttura di un transistore MOSFET (*Metal Oxide Semiconductor Field Effect Transistor*) a canale *n*. Il dispositivo è realizzato sulla superficie di un semiconduttore drogato *p* (il substrato rosso) ed è caratterizzato da due contatti drogati *n* (Source e Drain in verde) tra i quali è posto uno strato di dielettrico isolante - normalmente ossido di silicio - (in grigio), ricoperto da un contatto metallico detto di Gate (in blu).

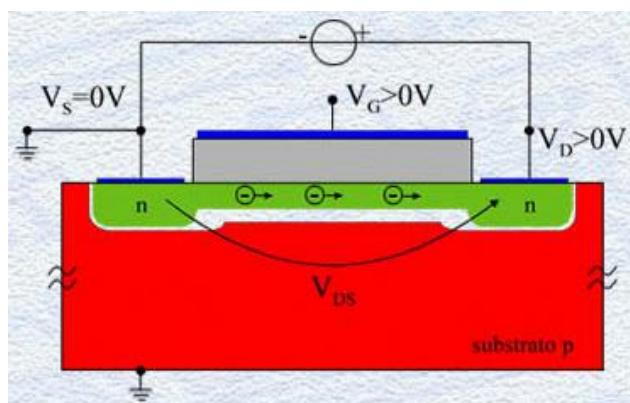
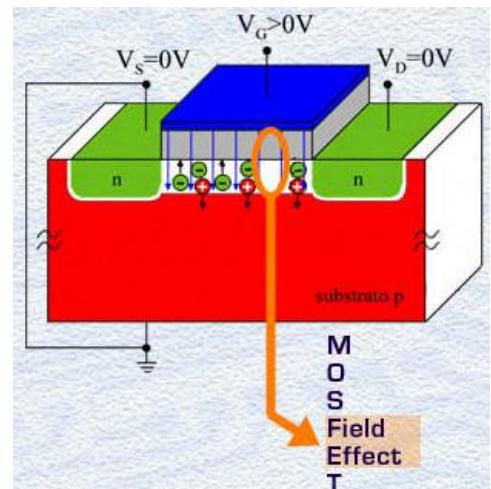
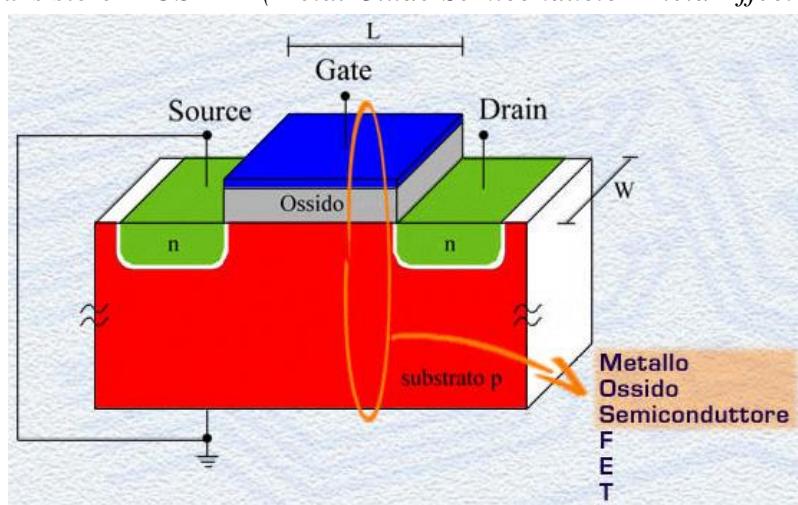
I due contatti *n* di Source e Drain risultano isolati elettricamente tra loro dal substrato *p* perché NON c'è una continuità di zona *n* (e quindi di elettroni) tra di essi. Se scorriamo in verticale il dispositivo lungo il suo asse centrale troviamo il Metallo di Gate, poi l'Ossido di silicio e poi il Semiconduttore, da cui il termine **MOS** nell'acronimo MOSFET.

Per semplicità supponiamo che il substrato sia tenuto allo stesso potenziale del contatto di Source, entrambi a massa. Chiamiamo **L** la lunghezza del Gate, cioè la distanza tra Source e Drain, e **W** la sua larghezza.

**Il funzionamento del dispositivo si fonda sulla possibilità di variare, agendo sulla tensione applicata al Gate, la concentrazione degli elettroni liberi presenti sotto l'ossido tra i due contatti *n*.**

Se infatti si **aumenta il potenziale del Gate** (dandogli tensioni positive), si crea un campo elettrico verticale tra Gate e substrato che **allontana le lacune verso il substrato e richiama elettroni liberi verso l'interfaccia tra semiconduttore e ossido**.

Proprio la presenza di questo campo elettrico giustifica la classificazione del dispositivo come "ad effetto di campo" contenuta nel termine **FE**, Field Effect, dell'acronimo MOSFET. L'arretramento delle lacune determina una zona svuotata di portatori maggioritari (appunto le lacune), nella regione sotto l'ossido che si estende tanto più dalla superficie verso l'interno quanto più è alta la tensione applicata al Gate. Contemporaneamente si ha un aumento della concentrazione degli elettroni sotto l'ossido, (attratti elettrostaticamente dalla tensione del Gate), che creano un cammino sempre più conduttivo (detto **canale**) tra Source e Drain, anch'essi pieni di elettroni.

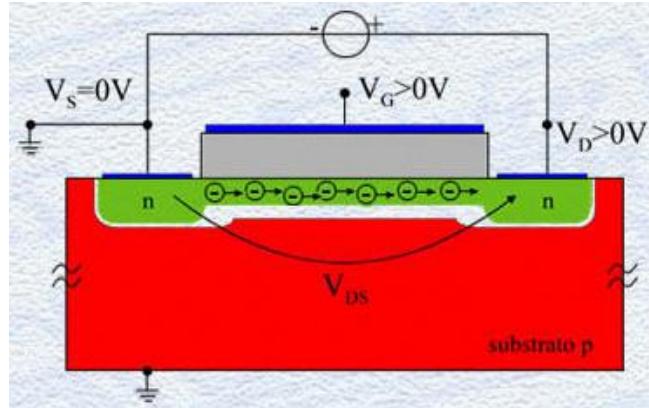


Quando ciò succede, i due terminali di Source e di Drain cessano di essere isolati elettricamente e se si applica un potenziale positivo  $V_D$  al Drain, tra i due contatti passa una corrente costituita dagli elettroni in moto dal Source al Drain. Questo fatto giustifica il nome dato ai due elettrodi, (Source: l'elettrodo sorgente degli elettroni; Drain: l'elettrodo di raccolta degli elettroni).

Aumentando ulteriormente la tensione positiva del Gate, la concentrazione di elettroni sotto l'ossido aumenta.

A pari tensione di Drain, anche la corrente aumenterà.

Si intuisce quindi come il **potenziale di Gate** possa agire come **variabile di controllo della corrente che fluisce tra i terminali di Source e Drain**.



### La tensione di soglia

E' opportuno evitare arbitrarietà nell'indicare la concentrazione di elettroni che dà inizio alla formazione di un canale conduttivo tra Source e Drain.

Pertanto si è convenuto di dire che:

**il canale inizia ad essere formato quando la concentrazione di elettroni sotto l'ossido giunge ad eguagliare la concentrazione che le lacune maggioritarie avevano lì prima di applicare tensione al Gate, cioè pari alla concentrazione di drogante nel substrato.**

Il valore della tensione  $V_{GS}$  (cioè  $V_G$  per  $V_S=0$ ), per cui si raggiunge questa condizione è detto **tensione di soglia**, indicata con  $V_T$ , e la relazione  $V_{GS} = V_T$  è nota come **condizione di inversione**.

Per  $V_{GS} < V_T$  i contatti di Source e Drain sono considerati in prima approssimazione ancora elettricamente isolati. Si parla di *inversione* per sottolineare che i portatori che garantiscono la conducibilità dello strato sotto l'ossido sono di *polarità opposta ai portatori maggioritari* del substrato di semiconduttore, appunto le lacune.

Si noti che tra il canale conduttivo di elettroni accumulati sotto l'ossido ed

zona priva di elettroni e di lacune

il substrato neutro si ha una zona priva sia di elettroni che di lacune, che assicura il completo isolamento elettrico tra i portatori mobili che fluiscono nel canale sotto l'ossido (elettroni) e quelli maggioritari (lacune) del substrato.

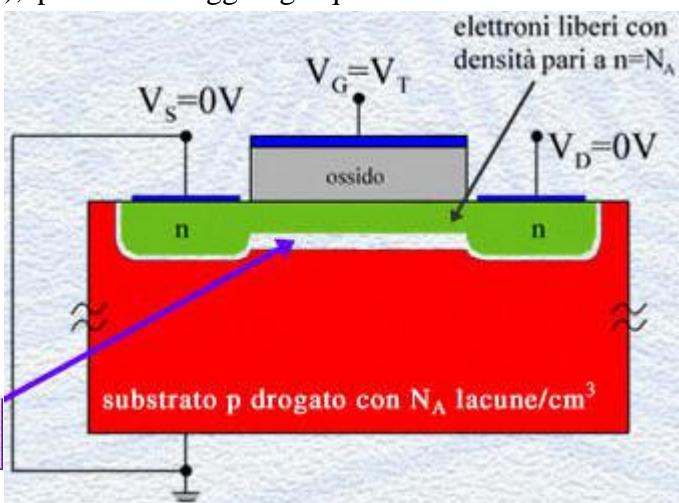
La tensione di soglia nel MOSFET cosiffatto ha, come visto, segno positivo dettato dalla necessità di indurre *elettroni* sotto l'ossido.

Per quanto detto sul suo significato, il valore della tensione di soglia  $V_T$  dipenderà da:

- spessore dell'ossido (più l'ossido è fino e maggiore è la carica indotta a pari tensione applicata)
- droggaggio del substrato (maggiore è il droggaggio, corrispondentemente maggiore dovrà essere la quantità di elettroni da indurre)
- eventuale presenza nell'ossido di Gate di ioni spuri, normalmente positivi (che indurrebbero loro stessi elettroni nel canale in aggiunta a quella indotti dalla tensione di Gate)

Per progettare dei circuiti, come vedremo meglio in seguito, è essenziale che il valore della tensione di soglia sia nota precisamente e sia uguale per tutti i dispositivi presenti nel circuito.

Questo ha imposto negli ultimi decenni alle ditte costruttrici di circuiti integrati uno sforzo tecnologico enorme per controllare precisamente i parametri appena citati, cioè lo spessore dell'ossido di silicio con una precisione migliore del nm, il droggaggio del substrato eventualmente modificandone localmente il valore proprio sotto l'ossido, e la carica spuria nell'ossido con una purificazione esasperata di tutti i processi produttivi.



Aumentando la tensione al Gate oltre il valore di soglia, la concentrazione di elettroni nel canale aumenta ulteriormente. E' importante analizzare questo aspetto in maggior dettaglio. Innanzitutto è utile notare che lo strato di cariche mobili nel canale e l'elettrodo conduttore di Gate formano le armature di un condensatore a facce piane e parallele, il cui dielettrico è l'ossido.

Ricordiamo che: - un condensatore costituito da due piatti conduttori di area  $A = L \times W$  affacciati tra loro e separati da un isolante spesso  $d$  e avente costante dielettrica  $\epsilon$ , ha capacità  $C$  pari a:  $C = (\epsilon / d) \times A$ .

Nel nostro caso possiamo quindi dire che la capacità di Gate, (detta anche capacità dell'ossido) è:  $C_{\text{gate}} = C_{\text{ox}} = \frac{\epsilon_{\text{ox}}}{d_{\text{ox}}} \text{Area}_{\text{gate}}$

Ricordiamo inoltre che in un condensatore di capacità  $C$ , la carica  $Q$  posta sulle sue armature determina una differenza di potenziale  $V$  tra esse pari a:  $V = Q/C$ .

Nel nostro caso, si può quindi dire, che la tensione tra Gate e canale, indicata anche come tensione tra Gate e Source ( $V_{GS}$ ), è pari alla carica che c'è nel canale di elettroni liberi ( $Q_n$ ), divisa per la capacità dell'ossido ( $C_{\text{ox}}$ ).

Un'analisi dettagliata del sistema **metallo-ossido-semiconduttore** mostrerebbe come, all'aumentare della **tensione di Gate oltre il valore di soglia**, effettivamente si formi il condensatore con i piatti separati dallo spessore dell'ossido  $d_{\text{ox}}$  e come la **tensione in eccesso** ( $V_{GS} - V_T$ ) rispetto alla minima necessaria per raggiungere la soglia vada a cadere quasi interamente ai capi dell'ossido. Per questo la concentrazione di carica nel canale (indicata con l'apice) cresce linearmente con la tensione secondo la relazione:

$$Q'_n = C'_{\text{ox}} (V_{GS} - V_T)$$

$Q'_n$  = elettroni liberi sotto l'ossido per ogni  $\text{cm}^2$  di area del Gate

$C'_{\text{ox}}$  = capacità del condensatore di Gate avente area di  $1\text{cm}^2$

## Il MOSFET come resistore

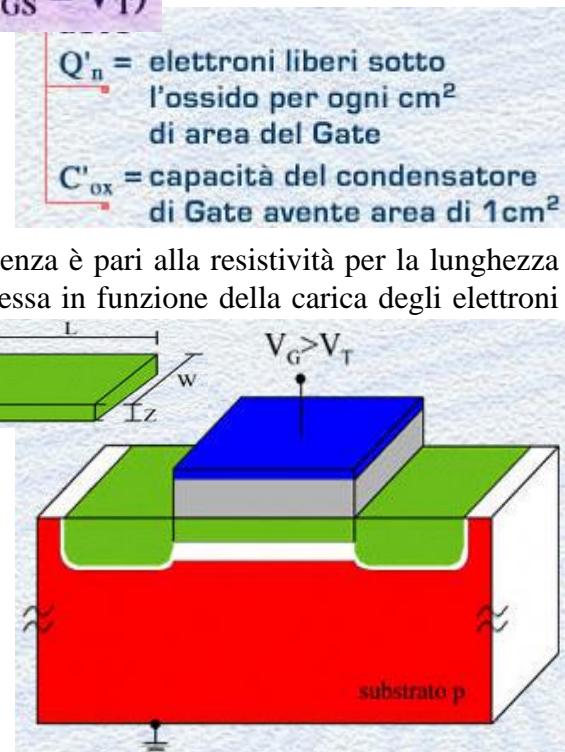
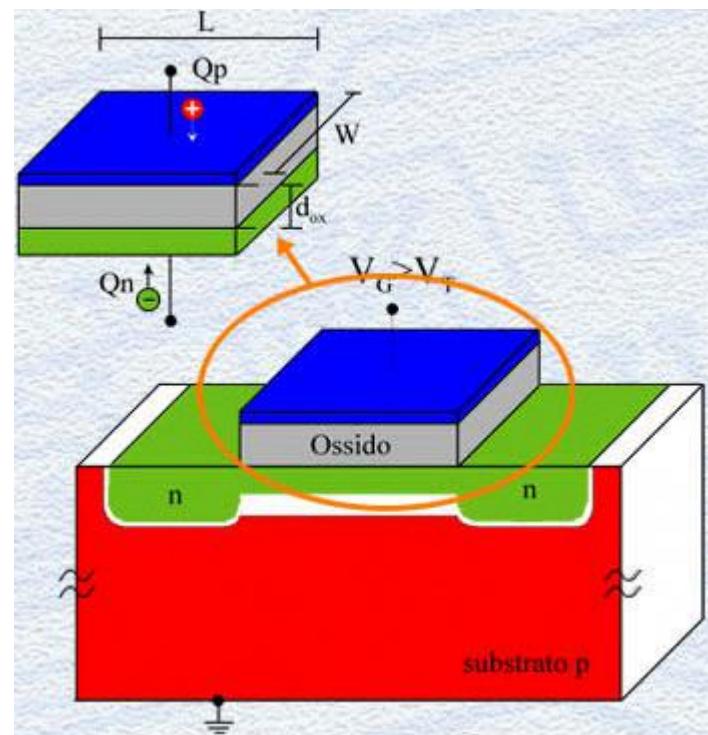
Ricordando le espressioni della resistenza di un blocchetto di semiconduttore e della sua resistività, viste

nella lezione precedente, si ha che la resistenza è pari alla resistività per la lunghezza diviso la sezione, mentre la resistività espressa in funzione della carica degli elettroni

quali portatori di corrente è:  $\rho = \frac{1}{q\mu_n n} [\Omega \cdot \text{cm}]$   
Si noti che moltiplicando gli elettroni, per unità di volume, al  $\text{cm}^3$  del canale, indicati con  $n$ , per lo spessore  $Z$  del canale stesso e per la carica dell'elettrone,  $q$ , si ottiene  $Q'_n$  prima definito. Pertanto la resistenza del canale tra Source e Drain è

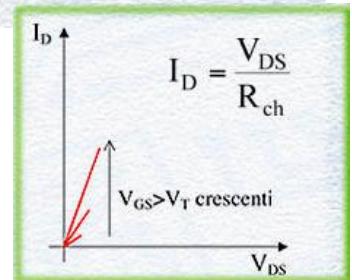
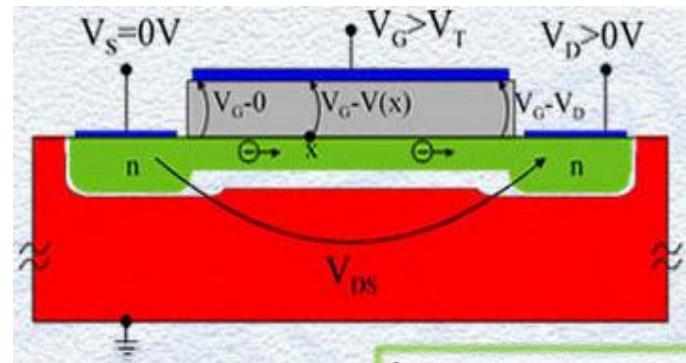
$$R_{\text{ch}} = \frac{1}{\mu_n \cdot Q'_n} \frac{L}{W}$$

$L$  e  $W$  sono rispettivamente la lunghezza lungo la direzione di moto dei portatori e la larghezza del canale, e  $\mu_n$  è la mobilità degli elettroni.

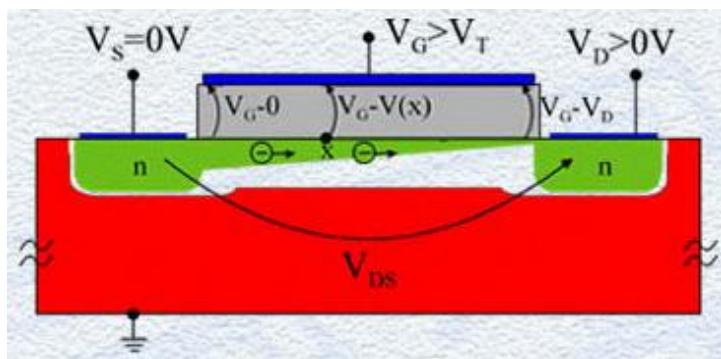


Al variare della tensione di Gate, come visto, la carica di canale può essere aumentata o diminuita e corrispondentemente quindi può essere **modulata la resistenza del canale**.

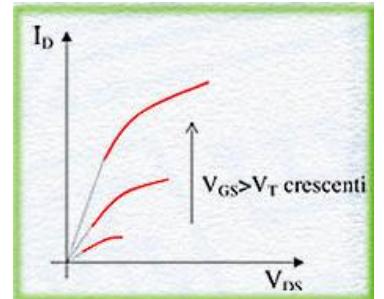
Basta applicare una piccola tensione positiva al Drain che nel canale fluisce una corrente che sarà tanto maggiore quanto più è piccola la resistenza del canale, cioè quanto più è alta la concentrazione di elettroni nel canale. Se mettiamo in un grafico la corrente circolante nel canale,  $I_{DS}$ , in funzione della tensione tra Drain e Source, otteniamo una delle curve caratteristiche mostrate nella figura. Esse sono sostanzialmente delle rette e differiscono in pendenza a seconda della tensione applicata al Gate: tensioni di Gate più elevate producono un canale meno resistivo e quindi permettono il passaggio di una corrente maggiore. Le curve rimangono sostanzialmente delle rette fintanto che la tensione  $V_{DS}$  rimane piccola, perché solo così la conducibilità del canale è praticamente costante lungo tutta la sua estensione. Per questo si dice anche che il dispositivo sta operando in **zona ohmica**, perché in queste condizioni esso si comporta a tutti gli effetti come un resistore.



Quando però si aumenta il potenziale del Drain, l'ulteriore aumento della corrente circolante segue



una legge meno che lineare. Infatti, la caduta di potenziale lungo il canale determinata dal passaggio della corrente fa diminuire la tensione ai capi dell'ossido man mano che ci si sposta verso il Drain. Quindi, la carica nel canale diminuisce procedendo dal Source verso il Drain e raggiunge il suo minimo



all'estremità del canale adiacente al Drain. Globalmente la carica nel canale sarà minore al crescere di  $V_{DS}$ . Questa diminuzione di carica fa sì che, aumentando la tensione  $V_{DS}$ , la resistenza media del canale aumenti e quindi che la corrente tenda ad aumentare sempre meno, come riportato nel grafico. La caduta di potenziale totale ai capi del canale è sempre pari a  $V_{DS}$ .

### Esercizio 1

- Cosa ci aspettiamo che succeda al MOSFET se diamo una tensione negativa al Gate ? Scorrerà corrente?
- Cosa succederebbe nel MOSFET se si mettesse il Drain a massa e si desse tensione positiva al Source? Quale conclusione traiamo?

### Esercizio 2

- Se la concentrazione di atomi di drogante Boro del substrato è di  $5 \cdot 10^{16}$  atomi/cm<sup>3</sup>, quale è la concentrazione di elettroni nel canale sotto l'ossido quando al Gate si applica la tensione di soglia ?
- Se la concentrazione del substrato fosse più alta (ad esempio  $8 \cdot 10^{17}$  atomi/cm<sup>3</sup>), la tensione di soglia sarebbe più alta o più bassa?

### Esercizio 3

- La resistenza di canale del MOSFET dipende o no dalla concentrazione del substrato su cui è costruito il dispositivo ?
- Come si deve costruire il MOSFET per avere bassa resistenza di canale, e quindi fare scorrere tanta corrente, ad una certa  $V_{DS}$  ?

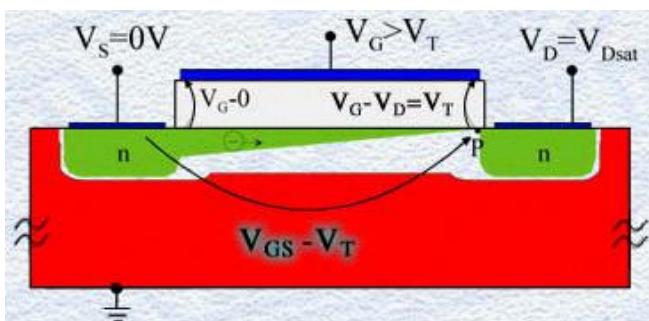
## Lezione n. 8 : Il MOSFET come transistore

### La corrente in saturazione

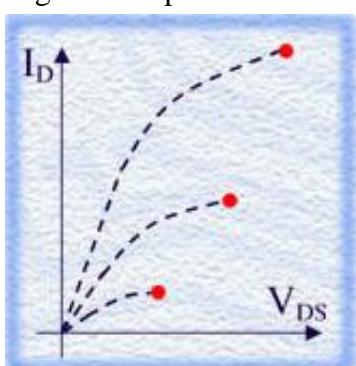
Nella lezione precedente abbiamo visto come sia possibile realizzare un dispositivo che svolga la funzione di resistenza variabile. La resistenza si manifesta tra i due morsetti di S e di D ed è regolata dalla tensione applicata al Gate. Benché interessante, questo dispositivo non è innovativo: la corrente tra i morsetti di S e di D dipende infatti sempre dalla tensione tra gli stessi due morsetti, come in un normale resistore.

Vediamo ora come invece lo stesso dispositivo riesca a realizzare una funzione elettronica ben più articolata ed innovativa. Riprendiamo quindi l'analisi del dispositivo dal punto dove eravamo arrivati.

Aumentando ulteriormente la tensione del Drain, la sua differenza con la tensione del Gate diminuisce e si arriva alla situazione in cui la tensione ai capi dell'ossido vicino al Drain non è più sufficiente per avere il canale conduttivo. Ciò accade esattamente quando la caduta di potenziale ai capi dell'ossido all'estremo vicino al Drain diventa pari a  $V_T$ , cioè quando  $V_G - V_D = V_T$ .



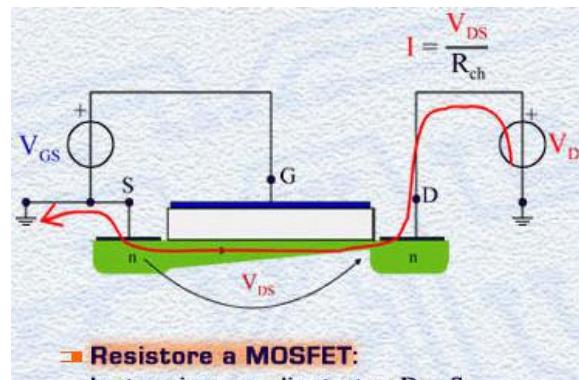
Per quanto visto, la corrente ora circolante nel canale tra i capi del canale, pari a  $(V_{GS} - V_T)$ , e la resistenza del canale rimasta tra il Source ed il punto P. Il calcolo di quest'ultima è complesso ed esula dagli scopi di



che avevamo chiamato  $R_{ch}$ , (R canale). Utilizzando la legge di Ohm ai suoi capi, come indicato nella figura, si ottiene l'espressione della corrente di Drain

come:

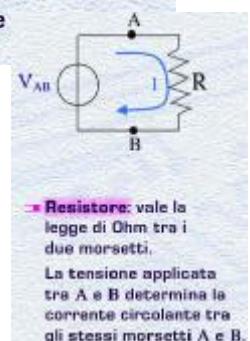
$$I_D = \frac{(V_{GS} - V_T)}{R_{sat}} = \frac{1}{2} \mu_n C_{ox} \frac{W}{L} (V_{GS} - V_T)^2$$



#### Resistore a MOSFET:

La tensione applicata tra D e S determina la corrente circolante tra gli stessi morsetti D e S.

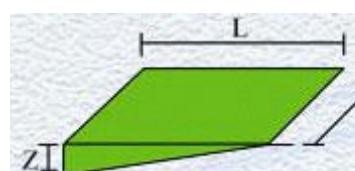
Al più il valore della resistenza può essere variato tramite la tensione al Gate.



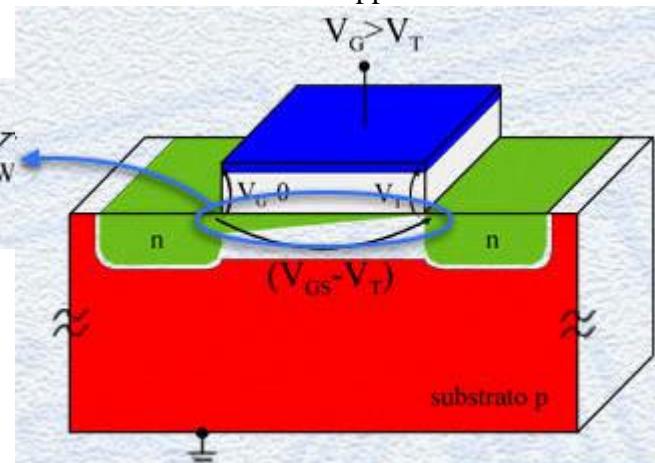
Resistore: vale la legge di Ohm tra i due morsetti.

La tensione applicata tra A e B determina la corrente circolante tra gli stessi morsetti A e B.

In questa situazione, detta condizione di **pinch-off** del canale (in italiano si direbbe di strozzamento del canale) la tensione ai capi del canale conduttivo è pari esattamente a  $(V_{GS} - V_T)$ . La tensione tra Drain e Source per cui si ha il *pinch-off* è detta **tensione di saturazione**, e vale  $V_{Dsat} = V_G - V_T$ .



questo corso, ma il suo risultato può essere ricostruito con la semplice considerazione che segue; il canale ora, per via della sua forma, ha sostanzialmente la metà dei portatori rispetto a quanto visto prima nel caso di zona ohmica. La sua resistenza, che chiamiamo  $R_{sat}$ , sarà pertanto doppia di quella allora calcolata,



$$R_{sat} = 2 \cdot R_{ch}$$

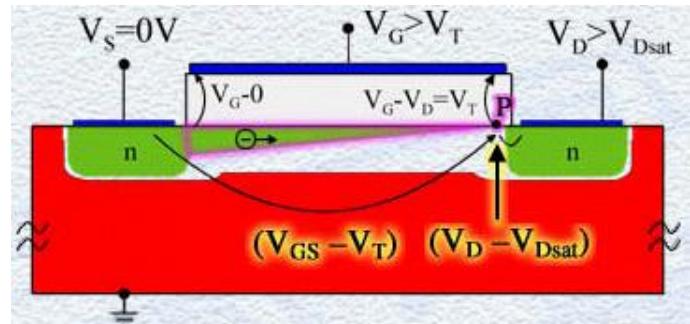
$$R_{ch} = \frac{1}{\mu_n \cdot Q_n} \frac{L}{W}$$

$$Q_n = C_{ox} (V_{GS} - V_T)$$

Questa espressione è proprio quella che si otterrebbe dai calcoli più raffinati che tengono conto della distribuzione di resistenza lungo il canale. Il valore di corrente trovato è indicato con il punto rosso sulle curve caratteristiche del dispositivo.

All'aumentare della tensione di Drain,  $V_D$ , oltre il valore di  $V_{Dsat}$  si verificherà che:

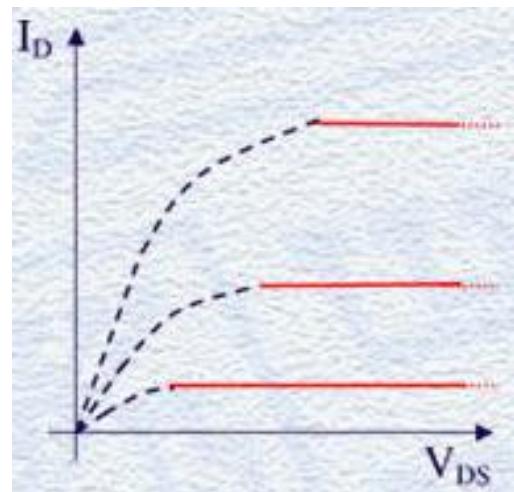
- Esisterà sempre il punto P, perché ci sarà sempre il punto del canale in corrispondenza del quale la concentrazione sarà inferiore a quella data alla soglia e la cui posizione al più arretra verso il Source
- Il canale avrà, praticamente, sempre la stessa forma e quindi la stessa resistenza
- La tensione ai capi del canale conduttivo, tra il Source ed il punto P, sarà sempre pari a  $(V_{GS} - V_T)$
- La corrente  $I_D$  rimane costante (saturata)



$$\rightarrow \text{sempre } I_D = \frac{(V_{GS} - V_T)}{R_{sat}} = \frac{1}{2} \mu_n C_{ox} \frac{W}{L} (V_{GS} - V_T)^2$$

All'aumentare della tensione  $V_{DS}$ , ai capi del canale conduttivo (tra Source ed il punto P) si

ha sempre la stessa caduta fissa di tensione, esattamente pari a  $(V_{GS} - V_T)$ . La tensione di Drain in eccesso rispetto a  $V_{Dsat}$  va a cadere tra il punto P ed il morsetto di Drain determinando un locale forte campo elettrico verso il Drain ed eventualmente un allargamento locale della regione svuotata. La corrente che fluisce nel transistore in condizioni di saturazione rimane quindi pressoché costante all'aumentare di  $V_{DS}$  ed è sempre data dal rapporto tra la tensione ai capi del canale  $(V_{GS} - V_T)$  e la resistenza di canale, quest'ultima pressoché costante. Quando la tensione al Drain è maggiore di  $V_{Dsat}$  si dice che il transistore opera in **saturazione** perché la corrente di Drain non varia più il suo valore (è saturata) nonostante la variazione della tensione di Drain. E' questa la zona di funzionamento di maggior interesse per le applicazioni circuitali analogiche lineari, come vedremo meglio nelle prossime lezioni. Si noti che il *pinch-off* del canale conduttivo non impedisce il passaggio della corrente. Infatti gli elettroni che partono dal Source, scorrono lungo il canale e giungono nel punto P, continuano ad essere accelerati verso il Drain dalla differenza di potenziale che c'è tra il punto P ed il Drain stesso.

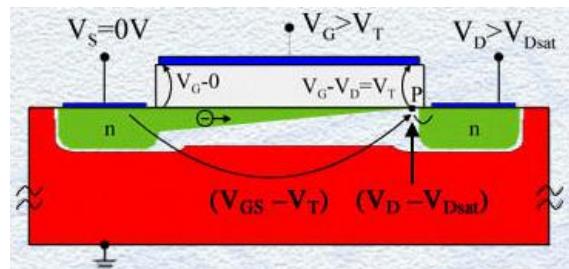


## Il MOSFET come transistore

Vediamo ora con più attenzione quale è l'effetto di una variazione della tensione applicata al Gate, mantenendoci nella condizione che  $V_{DS}$  sia sempre ad un valore superiore a  $V_{Dsat}$  in modo da essere sempre in zona di saturazione.

Un aumento della tensione di Gate determina due effetti primari:

- Il primo effetto è l'aumento della tensione ai capi del canale. Infatti, con il dispositivo in saturazione, la differenza di potenziale tra l'elettrodo di Gate ed il punto P rimane sempre pari a  $V_T$  per come è stato definito il punto P e quindi ogni variazione del potenziale sul Gate si trasmette rigidamente come una identica variazione del potenziale del punto P. In definitiva, attraverso un morsetto esterno (il Gate), si riesce a controllare il potenziale di un punto interno del dispositivo (il punto P) e conseguentemente il valore della tensione ai capi della resistenza del canale.
- Il secondo effetto dell'aumento del potenziale di Gate è l'aumento anche della carica laddove c'è canale, secondo la relazione:  $Q_{ch} = C_{ox} (V_{GS} - V_T)$ . Infatti, in ogni punto del canale, a cominciare dal Source, la tensione ai capi dell'ossido di Gate è maggiore di prima e quindi viene indotta più carica. Come conseguenza di ciò, diminuisce la resistenza del canale.



La combinazione di questi due effetti determina un aumento della corrente di Drain,  $I_D$  e ne giustifica l'aumento proporzionale al quadrato della tensione ( $V_{GS} - V_T$ ), secondo la relazione già trovata. Si usa indicare con  $k$  il parametro di proporzionalità tra tensione e corrente. Esso ha le dimensioni  $[A/V^2]$  ed ingloba le grandezze geometriche e fisiche del dispositivo.

Proprio per ottenere tanta corrente  $I_D$  con una minima tensione di comando  $V_{GS}$ , si cerca di avere il parametro  $k$  il più grande possibile, facendo lo spessore dell'ossido fine per avere  $C_{ox}$  grande e scegliendo una larghezza  $W$  di Gate grande ed una lunghezza di canale  $L$  la più corta possibile. Lo sviluppo della microelettronica in questi ultimi vent'anni è stata una continua rincorsa delle maggiori industrie del settore a tecnologie costruttive che rendessero possibile la costruzione di ossidi di silicio sempre più fini e di distanze tra le piste sempre più ravvicinate.

Da ultimo ricordiamo che il termine  $(V_{GS} - V_T)$  che indica, come più volte visto, la tensione ai capi dell'ossido in più rispetto al minimo valore necessario per iniziare a formare il canale, è spesso generalmente chiamato "tensione di overdrive".

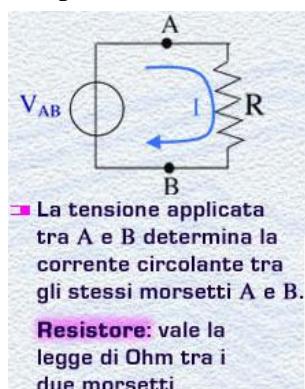
$$I_D = \frac{1}{2} \mu \cdot C_{ox} \frac{W}{L} (V_{GS} - V_T)^2 = k (V_{GS} - V_T)^2$$

$$k [A/V^2]$$

$$I_D = \frac{1}{2} \mu \cdot C_{ox} \frac{W}{L} (V_{GS} - V_T)^2$$

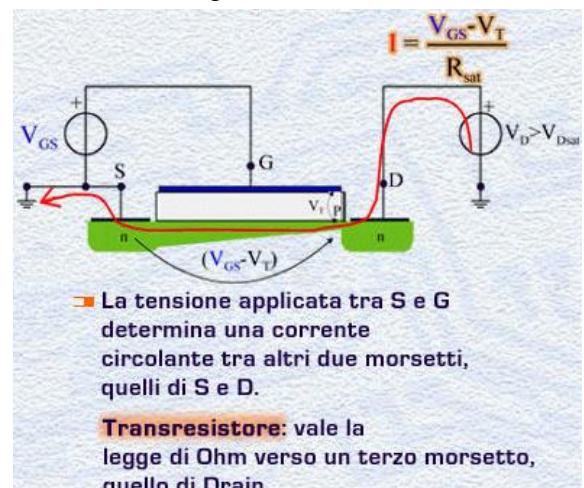
$V_{OD}$   
Tensione di overdrive

Riflettiamo un momento su quanto fin qui visto e confrontiamoci innanzitutto di nuovo con un semplice resistore.



In un resistore l'applicazione di una tensione tra i suoi morsetti produce una corrente (secondo la legge di Ohm) che scorre tra gli stessi due morsetti. Nel caso del MOSFET invece la situazione è estremamente più raffinata. L'applicazione di una tensione tra i morsetti di S e G definisce una differenza di potenziale ai capi della resistenza del canale "sommerso" sotto l'ossido. Questa produce una corrente

(sempre secondo la legge di Ohm) che però ora è disponibile ad un terzo morsetto, diverso dagli altri due, quello di Drain. Proprio per questo il dispositivo si chiama TRANS-RESISTORE o più semplicemente transistore.



Le particolarità del transistore, ed in particolare del MOSFET in saturazione, sono quindi le seguenti:

- la corrente che scorre nel canale è INDIPENDENTE dalla tensione  $V_D$ , del Drain stesso
- essa dipende solo dalla tensione applicata agli altri due morsetti, cioè S e G, per questo chiamati morsetti di comando del transistore, secondo la relazione:
- non c'è corrente continua attraverso il Gate. Pertanto il generatore di tensione di comando tra S e G non deve contemporaneamente anche fornire corrente. Il comando viene pertanto applicato spendendo una potenza stazionaria nulla ! Comando a  $P = V_{GS} \cdot I_G = 0$  !

$$I_D = \frac{1}{2} \mu \cdot C_{ox} \frac{W}{L} (V_{GS} - V_T)^2 = k (V_{GS} - V_T)^2$$

Tutto ciò è vero però solo se il MOSFET opera in zona di saturazione. Quando invece opera in zona Ohmica come visto nella lezione passata, la corrente circolante dipende anche dalla tensione di Drain, come avevamo visto, ed il dispositivo torna a comportarsi come un "semplice" resistore, al più con il valore di resistenza variabile tramite il Gate.

## Caratteristiche di un MOSFET

Per riassumere le relazioni tra le tensioni applicate ad un MOSFET e la corrente circolante tra S e D, riferiamoci ad un caso pratico di un dispositivo avente, per esempio,  $V_T=1V$  e  $k=3mA/V^2$ . Entrambi questi due dati sono forniti dal costruttore perché, come abbiamo visto, dipendono dalla geometria e dai parametri costruttivi dei vari elementi del dispositivo.

Le curve che rappresentano la corrente al Drain in funzione di  $V_{DS}$  sono dette **curve caratteristiche** del MOSFET e sono utili perché mostrano immediatamente la zona di funzionamento del dispositivo come transistore, quella in cui le curve sono orizzontali.

Si noti che il transistore è interdetto, cioè non porta corrente, fintanto che  $V_{GS}=1V$ , che infatti costituisce la sua tensione di soglia. Poiché per  $V_{GS}=0$  il transistore non conduce, in gergo si dice che esso è *normalmente spento (off)*. Solo applicando una tensione di Gate maggiore della tensione di soglia di 1V, si arricchisce il canale di portatori mobili (elettroni), la conducibilità del canale aumenta e nel dispositivo fluiscono correnti sempre più intense.

Questo comportamento classifica il dispositivo come un **MOSFET a canale n ad arricchimento** (in inglese *enhancement nMOSFET*).

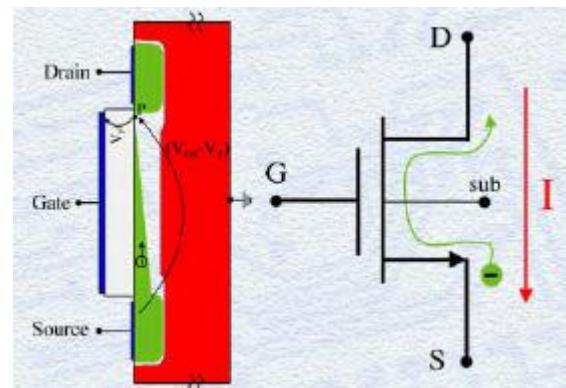
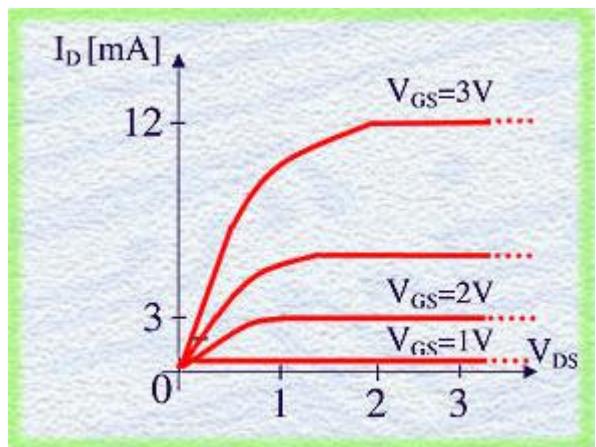
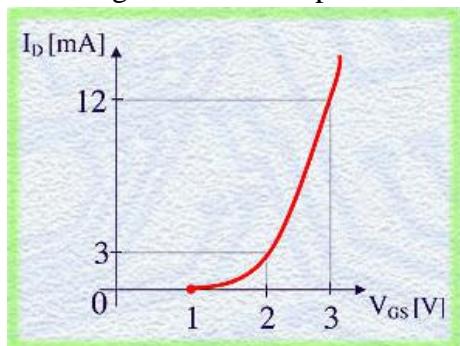
Le curve caratteristiche evidenziano anche come la zona di saturazione (vale a dire la parte piatta delle curve) inizi sempre più a destra quanto maggiore è il valore della tensione di Gate.

Nella figura sotto è riportato invece il grafico della relazione quadratica  $I_D = k(V_{GS} - V_T)^2$ ,

corrispondente ad un ramo di parabola con vertice in  $V_{GS}=V_T=1V$ . Essa è chiamata **curva transcaratteristica** del transistore. Anch'essa visualizza bene per quali valori di tensione di Gate il transistore porta corrente e quanta ne porta.

## I SIMBOLI CIRCUITALI

La figura riporta il simbolo circuitale del MOSFET ad arricchimento a canale n da usarsi quando si disegnano gli schemi circuituali. Il simbolo evidenzia il collegamento esistente tra S e D attraverso il canale. Il Source è individuato dalla presenza della freccia, il cui verso indica la direzione di scorrimento della corrente: in questo caso, dato che dal Source partono elettroni, la freccia è uscente dal dispositivo. Il simbolo evidenzia bene anche la presenza dell'ossido isolante di Gate che blocca il passaggio di corrente stazionaria attraverso di esso. Il terminale di substrato è spesso cortocircuitato con il Source e non è disponibile come contatto esterno. In questo caso, esso non appare nemmeno nel simbolo del dispositivo.



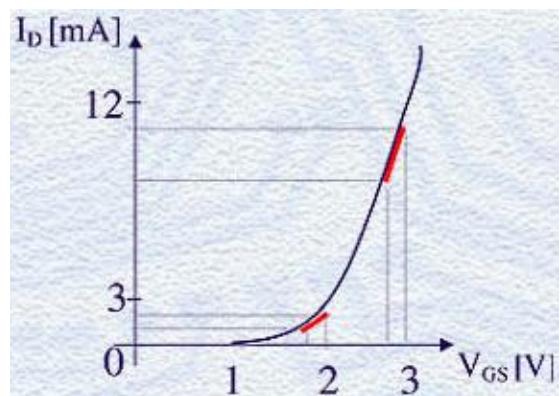
## LA TRANSCONDUTTANZA

L'efficacia del comando di tensione  $\delta V_{GS}$  nel variare la corrente erogata dal morsetto di Drain del transistore è espressa dalla transconduttanza, indicata con il termine  $g_m$ . Essa ci dice perciò di quanto varia la corrente di Drain per effetto di una variazione della tensione  $V_{GS}$ . Quando il dispositivo opera in zona di saturazione essa è data da due volte la corrente di Drain,  $I_D$ , diviso la tensione di overdrive.

$$g_m = \frac{\partial I_D}{\partial V_{GS}} = 2k(V_{GS} - V_T) = \frac{2I_D}{V_{ov}}$$

Maggiore è la transconduttanza, migliore è il transistore, in quanto maggiore è la variazione di corrente che si riesce ad inviare ad un eventuale carico collegato al Drain del MOSFET a pari sforzo di comando. Come si vede sia dalla formula che dal grafico, la transconduttanza è tanto maggiore quanto più grande è il valore della corrente stazionaria che già circola nel MOSFET. In altri termini, per avere un transistore con una alta transconduttanza quando si applicherà il segnale, bisogna dissipare molta potenza stazionaria.

La transconduttanza è sempre un numero positivo poiché ad un aumento della tensione di comando  $V_{GS} - V_T$  corrisponde sempre un aumento della corrente che fluisce nel dispositivo.



## Esercizio 1

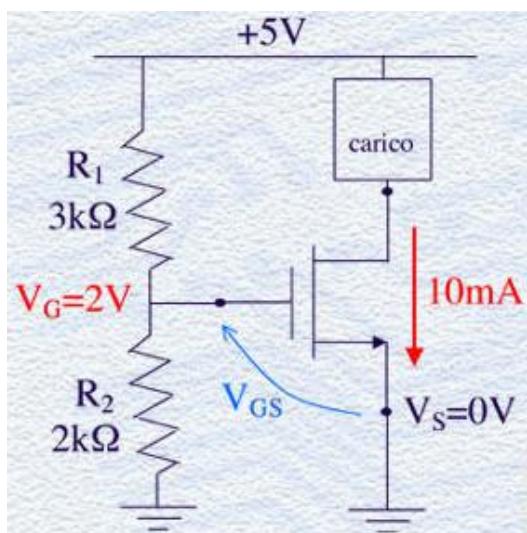
Provare a disegnare in un grafico con gli assi  $I_D$  e  $V_{DS}$  quotati, le curve caratteristiche del MOSFET in 3 situazioni diverse di tensione di Gate :  $V_{GS}=2V$ ,  $V_{GS}=3V$  e  $V_{GS}=4V$ . Si assuma che il MOSFET abbia  $k=3mA/V^2$  e  $V_T=1V$ .

## Lezione n. 9 : Principi d'uso del MOSFET e sue caratteristiche

### Generatori di corrente a MOSFET

La prima funzione che si può pensare di fare svolgere ad un transistore è quella di **generatore di corrente**. Infatti ricordando la relazione caratteristica:  $I_D = k(V_{GS} - V_T)^2$  basterà fissare la tensione tra Gate e Source ad un valore opportuno per ottenere al Drain il voluto valore di corrente.

Il circuito nella figura sotto è un esempio di come realizzare questa funzione. Il transistore è collegato con il Source a massa (quindi a 0V) ed il Gate ha la tensione definita dalla partizione delle resistenze  $R_1$  e  $R_2$ , poste tra la tensione di alimentazione di +5V e massa.



Il circuito nella figura sotto è un esempio di come realizzare questa funzione. Il transistore è collegato con il Source a massa (quindi a 0V) ed il Gate ha la tensione definita dalla partizione delle resistenze  $R_1$  e  $R_2$ , poste tra la tensione di alimentazione di +5V e massa.

Nel caso in esame il Gate viene a trovarsi a  $V_G=2V$ . Si noti che la tensione del Gate è unicamente fissata dal partitore resistivo in quanto il Gate non assorbe corrente.

Poiché né  $V_G$  né  $V_S$  cambiano nel tempo, al Drain sarà disponibile una **corrente costante nel tempo**.

Se ad esempio si usa un MOSFET che ha una  $V_T=1V$  ed un  $k=10\text{mA/V}^2$  (entrambi i dati sono forniti dal costruttore del dispositivo), la corrente sarà:  $I_D=10[\text{mA/V}^2](2 - 1)^2 [\text{V}^2] = 10\text{mA}$ . Questa corrente può essere fatta scorrere in un qualsiasi circuito o apparato che ne abbia bisogno, chiamato nel linguaggio dei progettisti "carico".

L'unica precauzione che bisogna prendere affinché il circuito funzioni effettivamente da generatore di corrente è che la tensione del morsetto di Drain del transistore non diventi mai più piccola di  $V_D=+1V$ . Se ciò succedesse, il transistore non funzionerebbe più correttamente nella zona di saturazione ma in quella

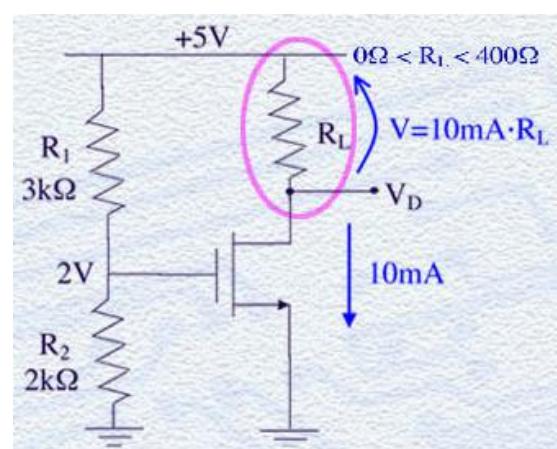
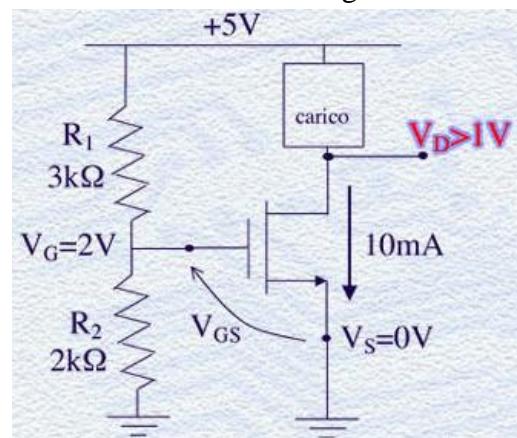
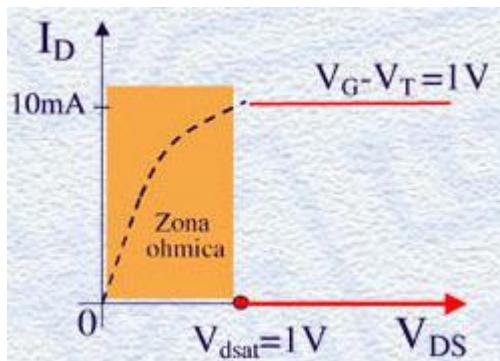
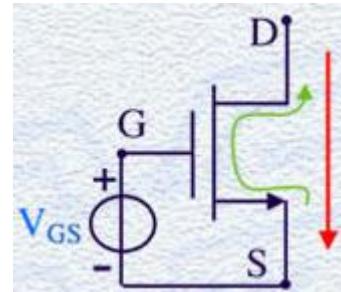
"Ohmica", come abbiamo visto nelle lezioni precedenti, e quindi la corrente elettrica non sarebbe più semplicemente fissata dalla relazione quadratica

$I_D = k(V_{GS} - V_T)^2$  ma sarebbe minore e dipenderebbe fortemente dalla stessa tensione di Drain.

Il più semplice carico che possiamo immaginarcì di collegare all'uscita del generatore di corrente è una resistenza posta tra il morsetto di Drain del MOSFET e la tensione di alimentazione di 5V.

Sia che il carico sia una resistenza da  $10\Omega$  o da  $100\Omega$ , sempre in essa scorrerà la corrente di 10mA. Nei due casi varierà solo la tensione che si forma al morsetto di Drain del MOSFET. Per quanto detto prima sul mantenere il transistore in saturazione, **la resistenza di carico non dovrà avere valori maggiori di 400 W**. Se ciò succedesse, infatti, la tensione al Drain scenderebbe sotto il valore limite di 1V.

E' utile abituarsi a sintetizzare un insieme di elementi circuituali che svolgono una ben definita funzione nel corrispondente "*elemento equivalente*" che svolga la stessa identica funzione. Nel nostro caso, la parte di



circuito tratteggiata può essere concettualmente sintetizzata nel simbolo del generatore di corrente. Questa procedura aiuta perché, una volta capiti i dettagli di una parte del circuito, semplifica il circuito complessivo e permette di concentrarsi sulla sua funzione.

Naturalmente, se volessimo generare una corrente costante di valore diverso dai 10 mA di prima, dovremo semplicemente cambiare il valore della tensione al Gate del MOSFET modificando corrispondentemente il valore delle resistenze del partitore resistivo. Per fare ciò in modo regolabile si può pensare di sostituire una delle due resistenze del partitore con un resistore variabile.

Questa semplice applicazione del MOSFET come generatore di corrente ci ha permesso di rivedere applicati in modo diretto i **principi del transistore**, che qui riassumiamo:

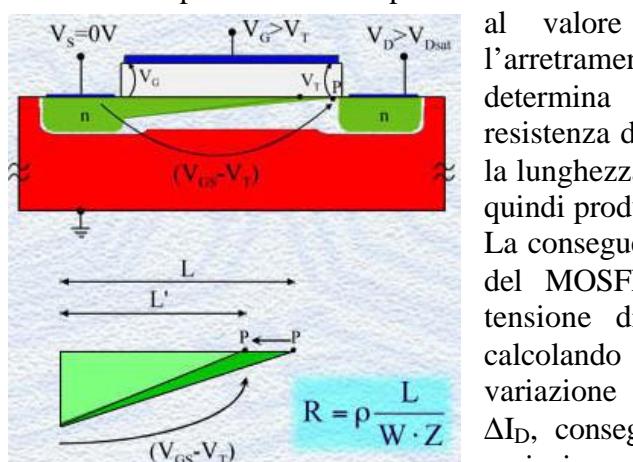
- avere a disposizione al morsetto di Drain una corrente elettrica **indipendente** dalla tensione a cui lo stesso Drain si trova;
- fissare questa corrente **unicamente** fissando la tensione tra il Gate ed il Source;
- operare una amplificazione di potenza tra la maglia di ingresso del transistore ( $P_{in} = I_B * V_G = 0$ ) e quella di uscita ( $P_{out} = I_D * V_D$ ).

Prima di analizzare modi d'uso più complessi del transistore quando si è in presenza di segnali variabili nel tempo, conviene approfondire alcuni aspetti del MOSFET non ancora visti.

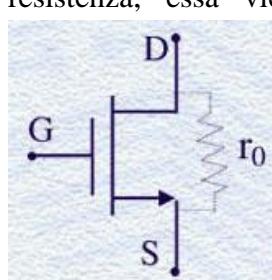
### Ulteriori proprietà dei MOSFET

Innanzitutto, purtroppo, le curve caratteristiche di un MOSFET reale prodotto da un'industria, non sono perfettamente orizzontali nella zona di saturazione ma hanno una più o meno accentuata pendenza.

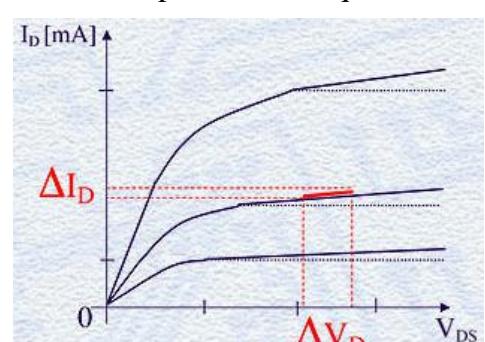
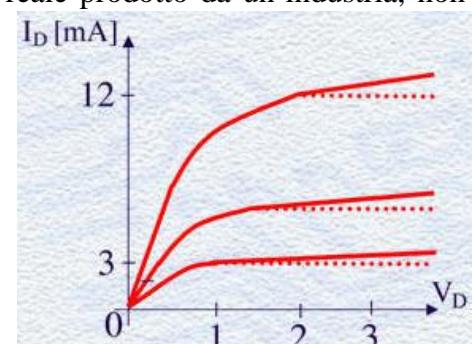
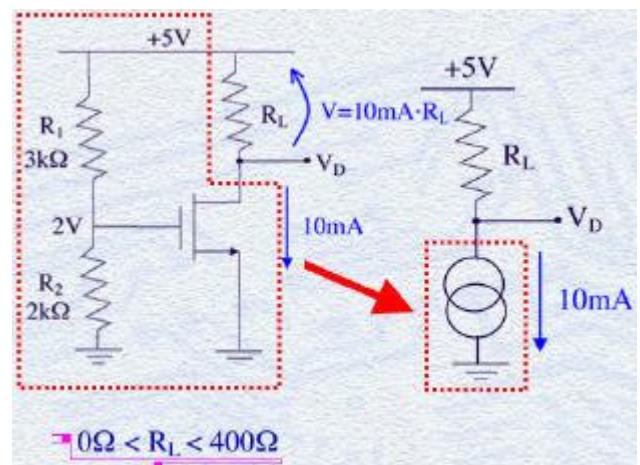
Questo fatto è causato dall'arretramento del punto P di pinch-off del canale quando la tensione di Drain aumenta. Poiché la differenza di potenziale tra il punto P ed il Source è sempre fissa al valore di  $(V_{GS} - V_T)$ , l'arretramento di P determina una minore resistenza del canale perché la lunghezza L diminuisce e quindi produce un aumento della corrente circolante.



di Drain,  $\Delta V_D$ . Poiché il rapporto di queste due quantità è una resistenza, essa viene chiamata **resistenza di Drain** del



MOSFET o anche resistenza di uscita del MOSFET, e la indicheremo con  $r_0$ . Spesso può essere utile immaginare questa resistenza come nascosta all'interno del transistore e posta tra S e D. Un transistore ideale dovrebbe avere resistenza di uscita infinita. Un transistore reale ha una resistenza di uscita finita, sperabilmente alta. Poiché i transistori più moderni sono sempre più corti (hanno cioè lunghezze L di canale sempre più piccole, fino a meno di un

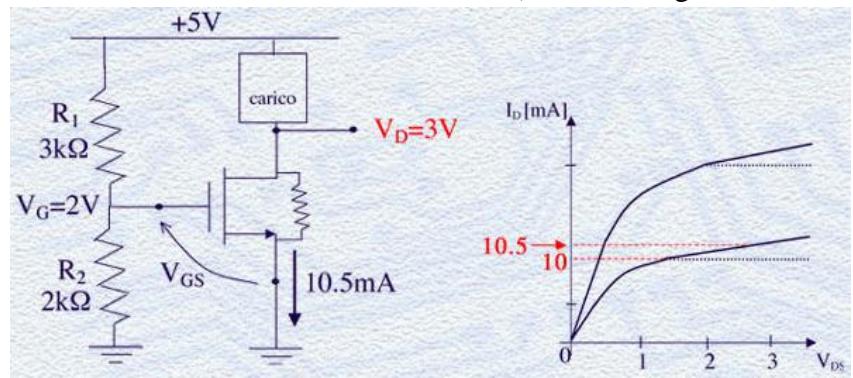


decimo di micron), anche piccoli spostamenti del punto P sono percentualmente rilevanti e danno luogo a variazioni percentuali di resistenza significative. Pertanto i più moderni transistori tendono ad avere resistenze di uscita modeste, a volte inferiori a  $10 \text{ k}\Omega$ .

Valori tipici di  $r_0$  sono:  $5 \text{ k}\text{W} < r_0 < 100 \text{ k}\text{W}$ .

La presenza di una  $r_0$  finita (pendenza finita sulle curve caratteristiche), ha conseguenze sulle prestazioni dei circuiti che impiegano transistori. Riprendendo l'esempio del generatore di corrente, la corrente da esso fornita non sarà più completamente indipendente dalla tensione ai capi del carico.

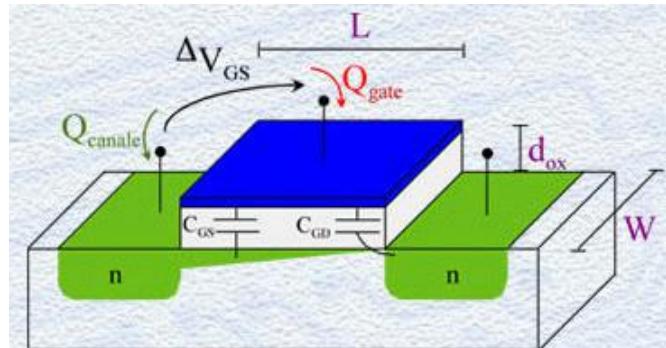
Quando, ad esempio  $V_D=3\text{V}$ , la corrente portata sarebbe di  $10.5\text{mA}$  invece dei nominali  $10\text{mA}$  che si avrebbero se  $r_0$  fosse infinita.



Un altro aspetto importante del MOSFET è legato alla capacità di Gate verso il canale.

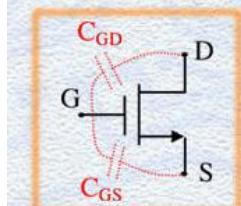
Infatti, per come è fatto il MOSFET, il Gate ed il canale costituiscono i piatti di un condensatore il cui dielettrico è l'ossido di silicio. Proprio per questo, ogni volta che si varia la tensione  $V_{GS}$ , cioè appunto la tensione ai capi del condensatore, si dovrà contemporaneamente anche depositare una corrispondente quantità di carica sui suoi piatti; carica di un segno sul Gate e di segno opposto nel canale.

Se per esempio si aumenta  $V_{GS}$ , della carica positiva deve essere posta sul metallo di Gate e corrispondentemente deve essere aumentata



$$C_{\text{gate}} = C_{GS} + C_{GD}$$

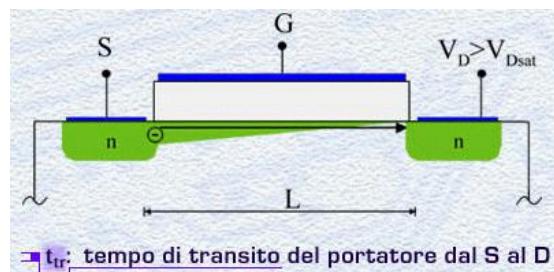
$$C_{\text{gate}} = \frac{\epsilon_{\text{ox}}}{d_{\text{ox}}} W \cdot L$$



della stessa quantità la carica negativa di elettroni nel canale. Questo fatto ha delle conseguenze pratiche sul funzionamento dei circuiti. Infatti, la capacità di Gate del MOSFET forma con le resistenze presenti nel circuito e ad essa collegate, una rete RC che impone i suoi tempi di risposta quando sollecitata da un segnale all'ingresso. In altri termini, ogni volta che si applica un segnale al Gate, bisogna mettere in conto che la sua tensione varierà più lentamente del previsto perché bisognerà attendere che tutta la carica necessaria, portata da una corrente di valore finito, sia arrivata sul condensatore. Vedremo durante il corso esempi dell'effetto di questa situazione. È comodo immaginare le due capacità fisiche presenti nel dispositivo, cioè  $C_{GS}$  e  $C_{GD}$ , come appunto nascoste nel dispositivo stesso ed indicarle tratteggiate nel simbolo.

Un altro aspetto importante di un MOSFET è legato al fatto che la corrente di uscita al Drain non varia istantaneamente nel momento in cui si varia la tensione  $V_{GS}$  di comando all'ingresso, ma varia con un certo ritardo.

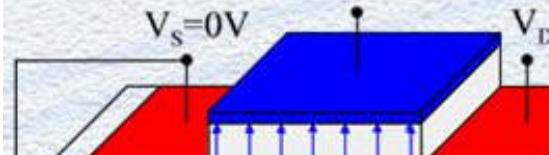
Infatti quando variamo  $V_{GS}$  dobbiamo immaginarcì che un ulteriore elettrone lasci il Source ed inizi il suo cammino verso il Drain; questo elettrone porta con sé l'informazione che il Gate ha cambiato valore, ma l'informazione arriva in uscita al dispositivo, cioè al Drain del MOSFET, solo quando l'elettrone effettivamente arriva al Drain. Questo ritardo nella trasmissione dell'informazione è legato al tempo di transito dell'elettrone nel canale: più il canale è lungo e più tempo impiegherà l'elettrone a percorrerlo. Questo tempo di transito è un parametro importante nel definire il MOSFET, e poiché dipende dalla geometria del dispositivo e da altri parametri fisici viene fornito dal costruttore.

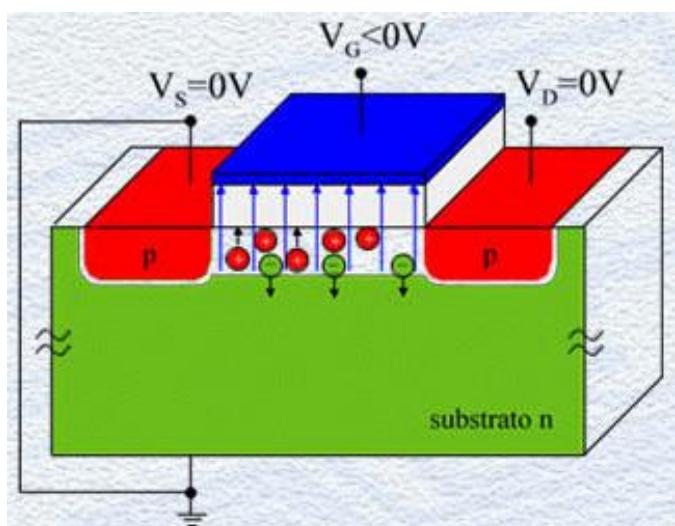


In verità invece che essere fornito direttamente come tempo di transito viene fornito come il suo reciproco cioè la frequenza:  $f_T = \frac{1}{2p \cdot t_{tr}}$ ; essa è la frequenza massima con cui si può pensare di variare la tensione di Gate oltre la quale non si nota più alcuna corrispondente modulazione della corrente di uscita. La frequenza  $f_T$  è chiamata **frequenza di taglio** del MOSFET od anche **frequenza di transizione**. Per quanto detto, transistori con  $f_T$  elevate permettono di realizzare circuiti in cui siano presenti segnali a frequenze elevate. Lo sviluppo delle tecnologie di fabbricazione dei MOSFET è sempre andato nella direzione di diminuire la lunghezza L del canale e realizzare circuiti per applicazioni sempre a più alte frequenze, come le telecomunicazioni od i computer ad altissima frequenza di clock. Valori tipici della lunghezza di canale L sono:  $0.1 \mu\text{m} < L < 3 \mu\text{m}$  ed i corrispondenti valori di  $f_T$  sono:  $1\text{GHz} < f_T < 50 \approx 100\text{GHz}$ .

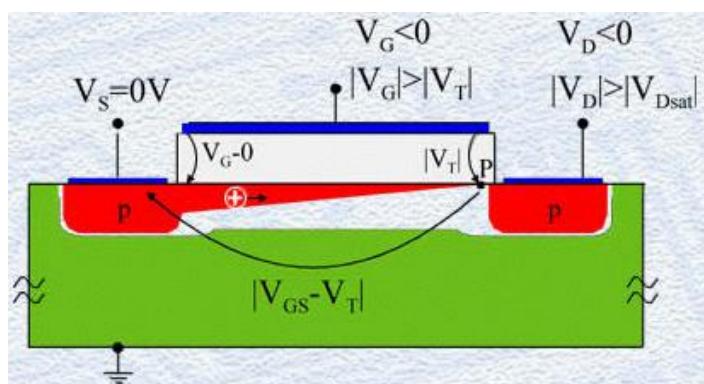
**pMOSFET e CMOS**

Il transistore MOSFET può anche essere costruito in modo che siano le lacune a muoversi nel canale invece che gli elettroni. Per fare ciò basta invertire i drogaggi delle varie zone di semiconduttore rispetto al MOSFET a canale n. Pertanto un MOSFET ad arricchimento a canale p è fabbricato su di un substrato di tipo n in cui le regioni di Source e di Drain sono ottenute drogando tipo p. Il funzionamento è analogo a quello visto per i MOSFET a canale n, ma ora la tensione da applicare al Gate è negativa rispetto al Source in modo da indurre carica positiva (le lacune) sotto l'ossido e realizzare così il canale conduttivo con portatori dello stesso tipo di quelli già presenti nei contatti di Source e di Drain, in questo caso appunto le lacune. La *tensione di soglia* sarà pertanto *negativa*.





Inoltre dovrà essere applicata una tensione al Drain negativa rispetto al Source in modo da fare



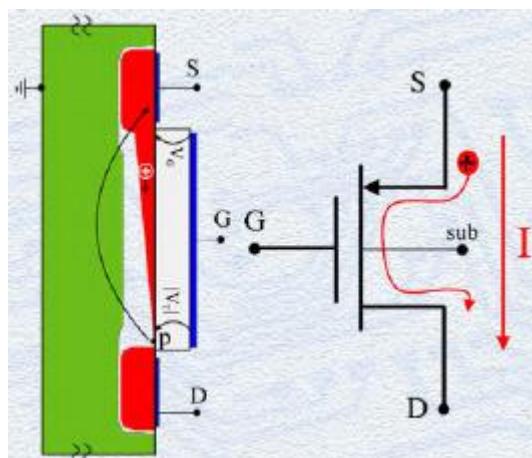
negativa rispetto al Source in modo da fare scorrere le lacune dal Source al Drain, come i nomi dei due elettrodi lasciano intendere. La corrente elettrica, per definizione equiversa al moto delle cariche positive, circolerà quindi tra Source e Drain. Quando operante in saturazione, la corrente nel MOSFET è sempre data dalla solita relazione:

$$I_D = \frac{(V_{GS} - V_T)}{R_{sat}} = \frac{1}{2} \mu_n C_{ox} \frac{W}{L} (V_{GS} - V_T)^2$$

cioè  $I_D \equiv k(V_{GS} - V_T)^2$ .

La figura mostra anche il simbolo circuitale usato per rappresentare un MOSFET a canale p ad arricchimento. Il verso della freccia indica il verso della corrente ed è posta sull'elettrodo di Source.

Le curve caratteristiche e la transcaratteristica hanno lo stesso andamento visto per nMOSFET a patto di considerare le tensioni e le correnti in modulo.



Per abituarvi ad utilizzare i pMOSFET, provate a calcolare la corrente prodotta da questo generatore di corrente che utilizza un MOSFET con tensione di soglia  $V_T = -0.8V$  e  $k$  pari a  $2 \text{ mA/V}^2$ . Notate come nel disegnare i circuiti sia opportuno, per motivi di chiarezza e d'ordine, mettere le alimentazioni maggiori in alto (nel nostro caso +3V) e quelle di minor valore in basso (nel nostro caso la massa) e conseguentemente fare scorrere le correnti stazionarie dall'alto verso il basso.

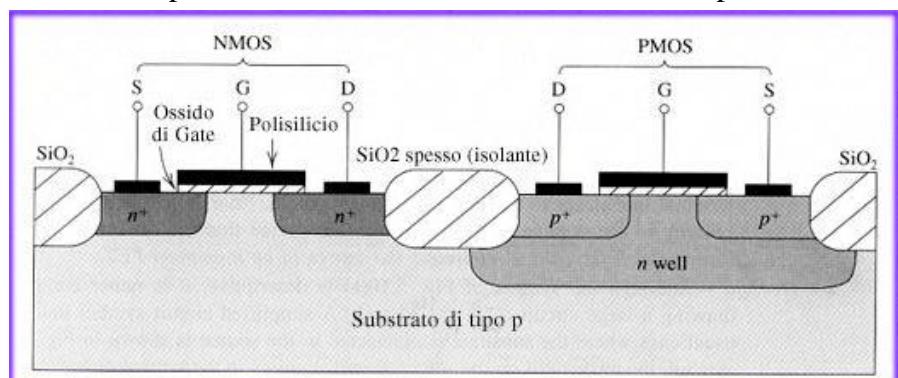
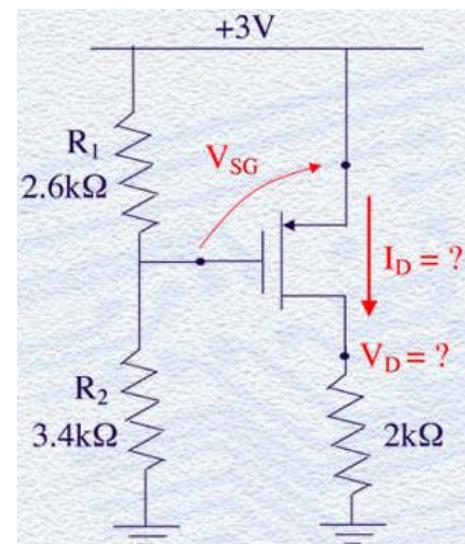
Sempre in questo circuito, calcolate quale è il valore massimo che può avere la resistenza  $R_L$  oltre cui il transistore esce dalla saturazione.

### **CONFRONTO TRA NMOSFET E PMOSFET**

Ritornando ai MOSFET a canale n, essi sono molto più utilizzati nei moderni circuiti integrati grazie alla maggiore mobilità degli elettroni rispetto alle lacune, che consente di realizzare circuiti che portano più corrente a pari geometria, ed alla minore tensione di soglia per creare il canale, che consente di realizzare circuiti con una minore tensione di alimentazione.

Tuttavia i MOSFET a canale p hanno un ruolo fondamentale nella realizzazione dei circuiti integrati digitali perché, in unione con gli nMOSFET, permettono di costruire circuiti elettronici più raffinati e potenti che non se ci fossero solo transistori dello stesso tipo. La tecnologia che permette di avere sullo stesso chip entrambi i tipi di MOSFET è chiamata tecnologia COMPLEMENTARE o, più brevemente, tecnologia CMOS. Nella figura a destra si può osservare una sezione di un *circuito integrato in cui sono presenti entrambi i tipi di MOSFET*.

Dato che 1 substrato deve essere lo stesso, ed è normalmente scelto di tipo p, laddove si vuole realizzare il pMOSFET si droga il substrato con atomi di fosforo in modo da realizzare una base n, chiamata in inglese *Well*, in cui costruire il transistore p. L'isolamento tra i due transistori è assicurato da uno spesso strato di ossido di silicio.



### **Esercizio 1**

Utilizzando lo schema circuitale introdotto nel modulo 1, in cui ora la tensione di alimentazione sia di 3.3V e  $R_1=15\text{k}\Omega$ ,  $R_2=18\text{k}\Omega$  ed il MOSFET abbia  $V_T=0.8V$  e  $k=1\text{mA/V}^2$ , calcolare:

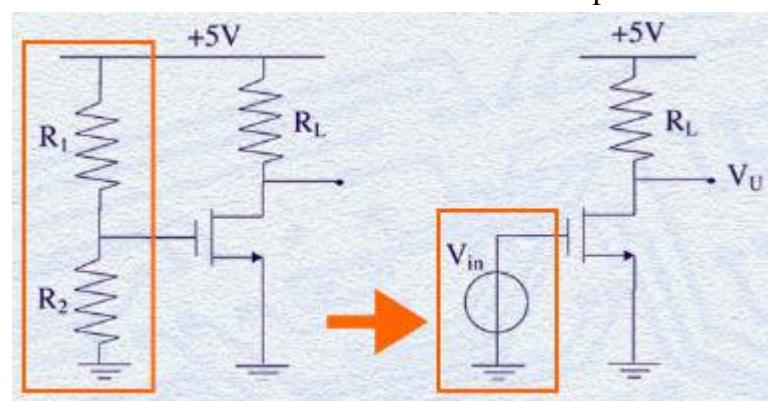
- la corrente prodotta dal generatore
- la massima resistenza di carico che può essere collegata.

## Lezione n. 10 : L'invertitore a MOSFET

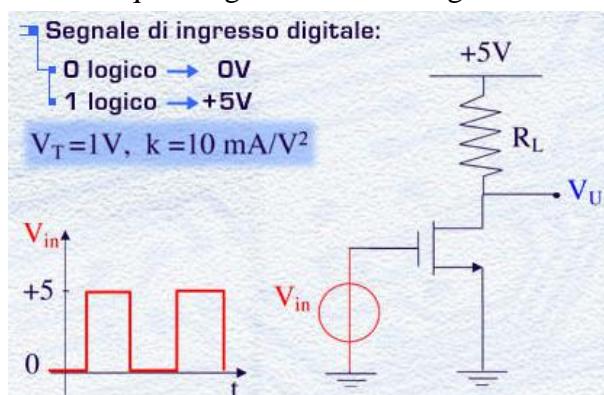
### Funzionamento di un circuito invertitore

Riprendiamo il circuito del generatore di corrente che abbiamo introdotto nella lezione precedente. In questa lezione vediamo come si comporta il MOSFET quando gli viene applicato al Gate un segnale di tensione variabile nel tempo.

Innanzitutto bisogna rendere possibile lo spostamento in tensione del Gate. Per fare ciò, si può ad esempio sostituire il partitore resistivo che avevamo introdotto proprio per fissare la tensione di Gate con un semplice generatore di tensione ideale direttamente collegato tra Gate e massa.

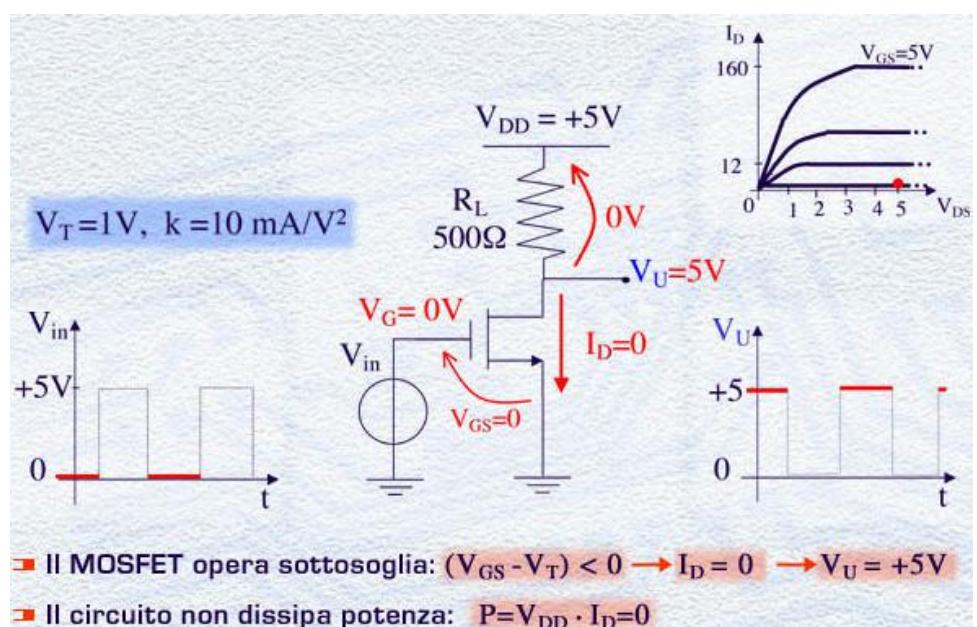


Tramite questo generatore immaginiamo di applicare al Gate del circuito un segnale digitale di tensione, il cui 0 logico sia dato dal livello di tensione di 0V e l'1 logico sia dato dal livello di tensione di +5V.



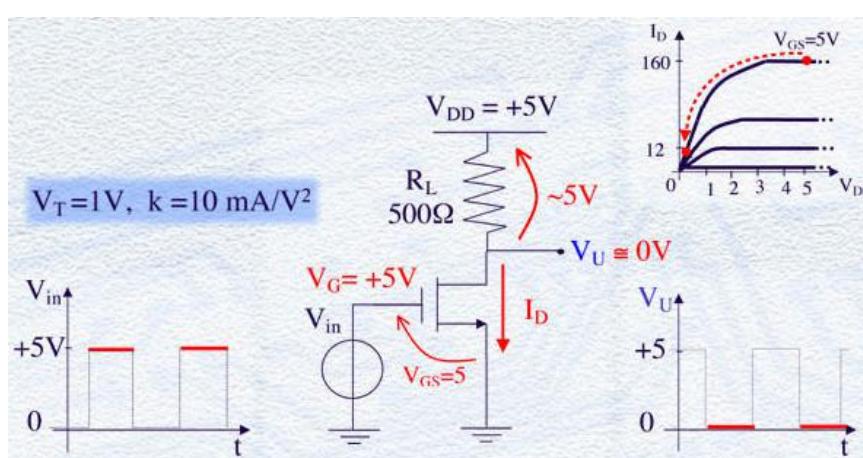
Vogliamo studiare come reagisce il circuito a questo segnale digitale applicato all'ingresso, vale a dire come varia nel tempo la tensione  $V_U$  all'uscita del circuito presa sul Drain del MOSFET. Per fare ciò analizziamo separatamente le due situazioni di quando il segnale d'ingresso è basso (0V) e di quando è alto (+5V). Immaginiamo di avere scelto anche ora un MOSFET con  $V_T=1V$  e  $k=10 \text{ mA/V}^2$ .

Quando  $V_{in}=0V$ , anche la tensione di Gate è a 0V ( $V_G=0V$ ) essendo il morsetto di Gate direttamente collegato al generatore di tensione di segnale. In questo caso la tensione  $V_{GS}$  di comando del MOSFET è uguale a 0V ed è inferiore al valore minimo dettato dalla soglia  $V_T=1V$  perché il MOSFET non scorre corrente tra S e D. Pertanto  $I_D=0V$ . Il punto di lavoro del MOSFET è mostrato nel grafico delle curve caratteristiche. Poiché non scorre corrente in  $R_L$ , non vi sarà neppure caduta di tensione ai suoi capi. Il valore della tensione di uscita è quindi  $V_U=+5V$ . Se la situazione di ingresso non cambia, cioè se  $V_{in}$  rimane a 0V, anche l'uscita non cambia e rimane a 5V. Per tutto questo tempo la potenza dissipata dal circuito è praticamente zero, perché la corrente circolante nel transistore è praticamente zero.



Quando  $V_{in}$  scatta a +5V, la tensione  $V_{GS}$  scatta anch'essa a +5V per cui il MOSFET si mette a funzionare sopra soglia con una tensione di overdrive ( $V_{GS}-V_T=4V$ ). Quindi il MOSFET è nella condizione di portare molta corrente, al più data dal valore:  $I_D = k(V_{GS}-V_T)^2 = 10(5-1)^2 = 160 \text{ mA}$ . Ma tale corrente, circolando anche attraverso la resistenza di carico  $R_L$ , produce una grande caduta di tensione ai capi di  $R_L$  con conseguente discesa del valore della tensione  $V_U$  dal valore precedente di +5V ad un valore prossimo a 0V, che è il valore minimo a cui un punto del circuito può scendere. A questo punto, in cui  $V_{DS}$  si è ridotta a poche centinaia di mV, il MOSFET è uscito dalla saturazione perché il Drain è sotto la tensione di Gate di più della soglia. Pertanto il MOSFET si è messo a portare una corrente ben inferiore ai 160 mA inizialmente previsti. La corrente portata dal MOSFET sarà ora fissata dalla legge di Ohm sulla resistenza  $R_L$  e circa pari a:  $I_{Dmax} \approx \frac{5V - 0}{R_L} = 10\text{mA}$ , nel nostro caso in cui  $R_L=500\Omega$ .

Il grafico delle curve caratteristiche mostra quanto è successo al MOSFET. Poiché la tensione al Gate non cambia mai, bisogna stare sempre sulla curva caratteristica contrassegnata con  $V_{GS}=5V$ . Il transistore passa dal punto in saturazione al punto in zona ohmica man mano che la tensione  $V_{DS}$  diminuisce dagli iniziali 5V al valore quasi 0 finale. Se la situazione di ingresso non cambia, cioè se  $V_{in}$  rimane a 5V, anche l'uscita non cambia e rimane a 0V. Per tutto questo tempo però il circuito dissipava potenza elettrica pari al prodotto della corrente circolante nel Drain (10mA) con la



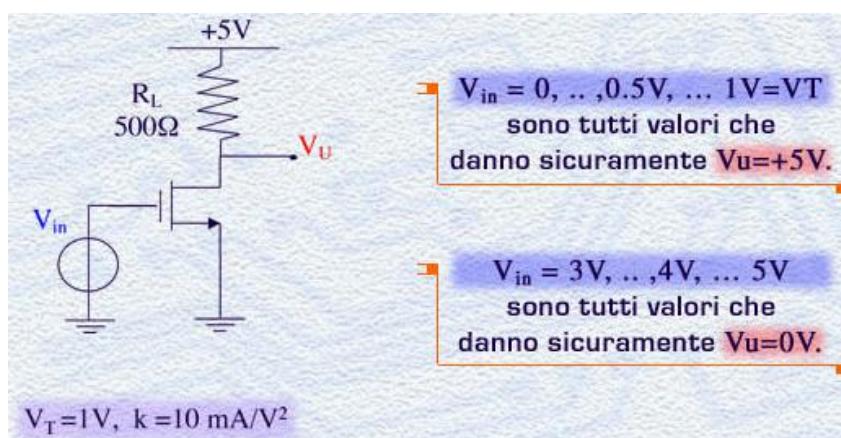
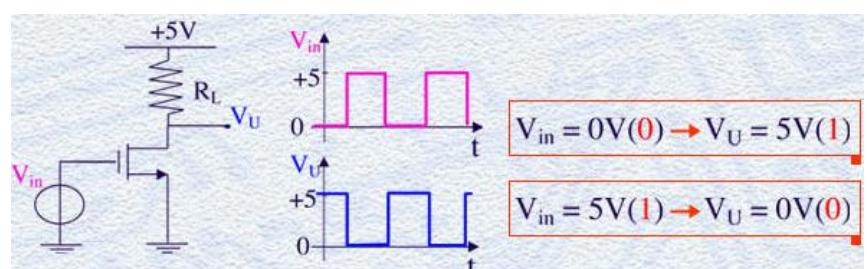
tensione di alimentazione (5V):  $P = V_{DD} * I_D = 5V * 10\text{mA} = 50\text{mW}$ .

Il circuito, per quanto visto, si sta comportando in modo da "invertire" il valore di tensione presentatagli al suo ingresso.

Avendo associato a 0V lo zero logico ed a 5V l'uno logico, un tale circuito svolge la funzione di **"INVERTITORE LOGICO"** e costituisce, in questa forma semplice o in altre più raffinate, uno dei mattoni fondamentali di

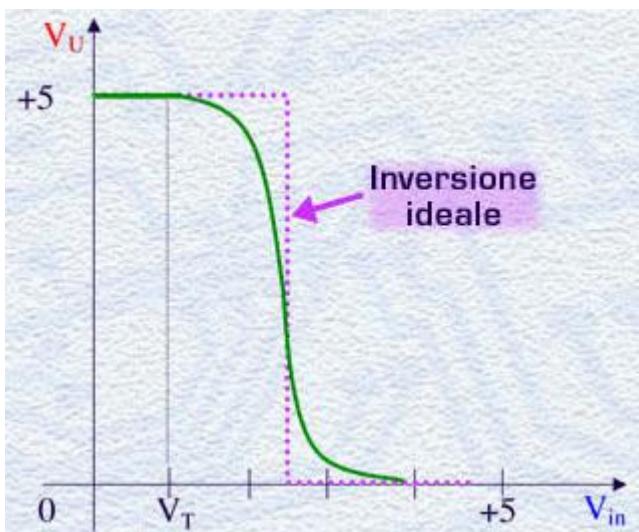
cui è composto un circuito digitale complesso. *Affinché l'invertitore svolga il suo compito NON è necessario che il segnale al Gate sia di 0V o di 5V.*

Ad esempio, anche un segnale di ingresso di 0.5V impone all'uscita di stare alta a +5V; infatti 0.5V è inferiore alla soglia del MOSFET e non gli fa portare comunque corrente.



Analogalemente, anche un segnale di ingresso di 3V impone all'uscita di stare bassa a 0V; infatti con 3V l'overdrive del MOSFET è di 2V e la corrente che il transistore potrebbe portare sarebbe di 40 mA, sufficiente ad abbassare  $V_U$  a zero volt, dato il valore di  $R_L=500\Omega$ .

Proprio per visualizzare con precisione il comportamento dell'invertitore anche per tutti i valori di tensione compresi tra 0V e +5V, la relazione tra  $V_{in}$  e  $V_U$  del circuito viene spesso mostrata in un unico grafico riassuntivo sul cui asse delle ascisse c'è  $V_{in}$  ed in quello delle ordinate c'è  $V_U$ .

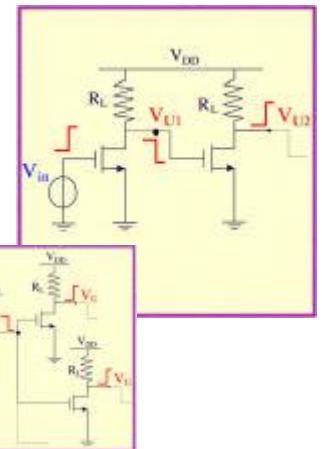


Il grafico mette bene in evidenza l'intervallo di valori di  $V_{in}$  per cui l'uscita è alta e quello per cui l'uscita è bassa, oltre ai valori di  $V_{in}$  in corrispondenza dei quali l'uscita assume valori intermedi tra 0 e +5V. Idealmente si vorrebbe questa parte centrale la più ripida possibile in modo che l'uscita stia sempre o al livello alto o a quello basso. In realtà la curva è più dolce, dipendendo dalle caratteristiche del circuito, in particolare dal  $k$  del MOSFET; *maggior è k più ripida è la curva perché il MOSFET porterebbe più corrente e quindi l'uscita scenderebbe prima a 0V*, e dal valore di  $R_L$ : *maggior è R\_L più ripida è la curva perché, a pari corrente circolante, la caduta di tensione ai suoi capi sarebbe maggiore e quindi l'uscita scenderebbe prima a 0V*.

Il circuito invertitore appena visto si presta bene ad essere collegato in serie ad altri circuiti uguali, a formare circuiti più complessi.

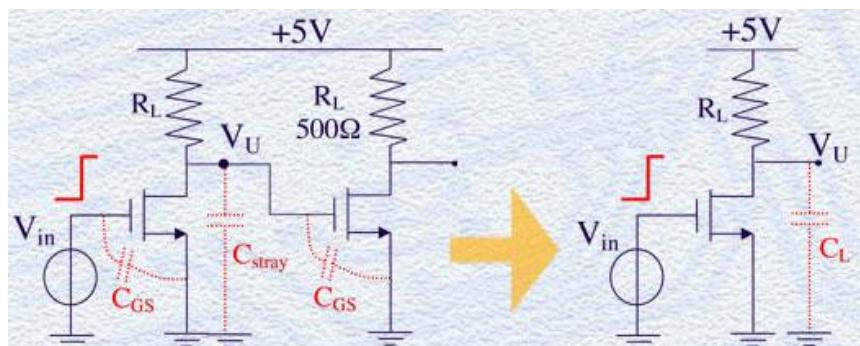
Come si vede nella figura più grande, l'uscita del primo invertitore pilota l'inversione del secondo e così via. Ma si può pensare ad un collegamento in parallelo di più invertitori, in cui un primo circuito pilota contemporaneamente il Gate di parecchi altri invertitori.

In tutti questi casi subentra l'aspetto nuovo di valutare se e quanto la presenza del nuovo circuito a valle del primo condiziona il suo comportamento.



### Dinamica dell'inversione

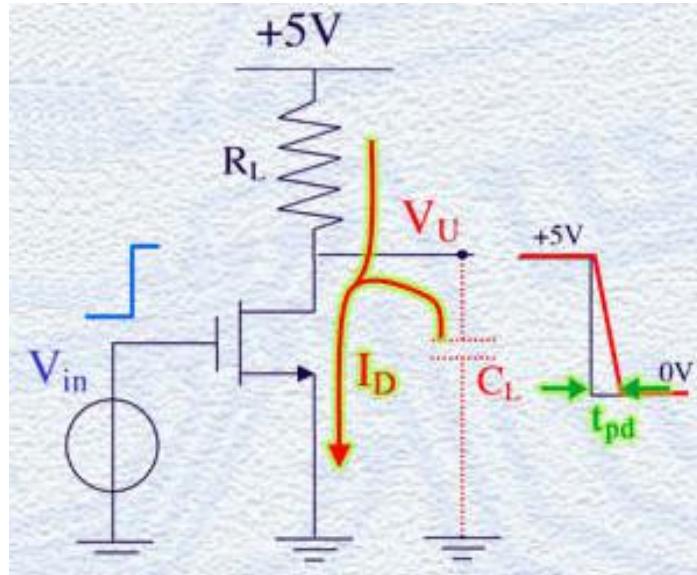
Il problema nasce dal fatto che tutte le volte che abbiamo a che fare con delle capacità collegate tra un punto che si muove in tensione e massa, la variazione di tensione avviene solo se contemporaneamente si fornisce al condensatore la carica corrispondente, data dalla relazione  $Q=C \cdot V$ . Poiché tale carica è fornita dalle correnti circolanti, bisognerà aspettare un certo tempo prima che sui piatti del condensatore si sia effettivamente depositata tutta la carica necessaria. Ciò rende le varie commutazioni dell'invertitore non più istantanee ma lente. Maggiori sono le capacità da caricare e maggiore sarà il tempo di commutazione del circuito. Pertanto, come regola generale e prima ancora di entrare nei dettagli del problema, un buon progetto presuppone di minimizzare le capacità scegliendo **MOSFET delle dimensioni più piccole possibili** e realizzando i **collegamenti tra i MOSFET più corti possibili**, perché anche le piste di collegamento, come visto nelle prime lezioni del corso, generano delle capacità verso massa.



Per semplicità di trattazione, si concentrano tutte le capacità che afferiscono al nodo  $V_U$  in un'unica capacità  $C_L$  che le somma tutte.

Vediamo quindi come si comporta l'invertitore durante la commutazione dell'uscita da alto a basso, corrispondente alla commutazione dell'ingresso da basso a alto. Poiché già sappiamo molto sull'inversione, concentriamoci sull'effetto introdotto dalla capacità  $C_L$ . Prima della commutazione,  $V_U$  è a +5V. Questo vuole dire che la capacità  $C_L$  ha molta carica sui suoi piatti, esattamente pari a  $Q = C_L * 5$  [coulomb]. Questa è la carica che bisogna togliere se si vuole che l'uscita  $V_U$  scenda fino a 0V.

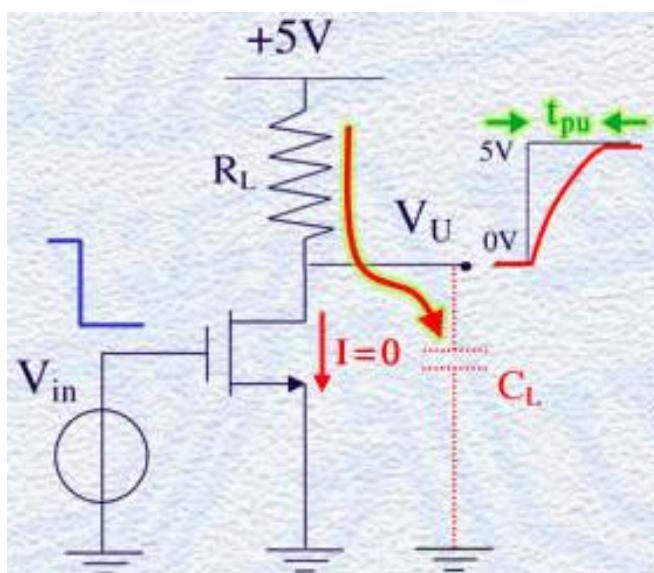
Quando  $V_{in}$  commuta da basso ad alto, il MOSFET si accende e tende a portare molta corrente dal Drain verso il Source. Questa corrente, oltre a provenire dalla resistenza  $R_L$ , proviene anche da  $C_L$  prelevandogli carica. Poiché la corrente  $I_D$  è molto elevata (avevamo visto che inizialmente era di 160mA) la carica del condensatore può venire prelevata molto velocemente. Grazie a questo intervento attivo del MOSFET, la



commutazione in giù del nodo di uscita tende ad essere veloce. Il tempo di commutazione dell'invertitore sarebbe ancora minore se la corrente iniziale del MOSFET fosse maggiore dei 160mA. In generale, quindi, possiamo dire che il tempo di commutazione dell'invertitore nella fase di discesa dell'uscita, detta anche di **pull-down**, è breve e limitato dalla corrente del transistore.

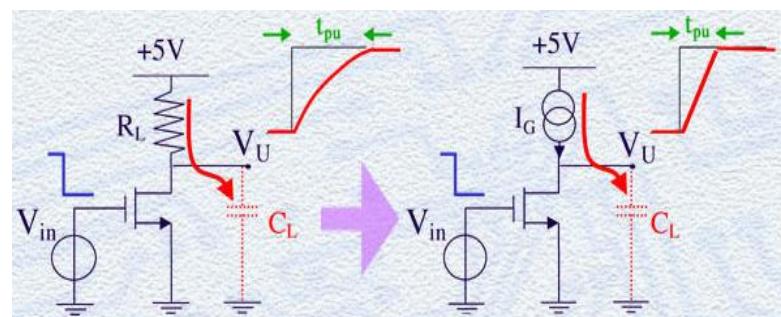
Quando invece la commutazione dell'uscita avviene dal basso verso l'alto, detta anche fase di **pull-up** dell'invertitore, la situazione è radicalmente diversa.

L'uscita prima della commutazione, è a 0V, per cui ai capi del condensatore la tensione è nulla e sui suoi piatti non c'è carica. Quando  $V_{in}$  commuta da alto a basso, il MOSFET viene spento e conseguentemente  $I_D=0$ . Poiché inizialmente  $V_U$  era a 0V, nella resistenza  $R_L$  passa comunque una corrente che, non potendo più scorrere attraverso il transistore, va a caricare la capacità  $C_L$ . La tensione  $V_U$  inizia quindi a salire in valore. Così facendo però si riduce la caduta di tensione ai capi di  $R_L$  e quindi la corrente proveniente dal resistore diminuisce. La capacità  $C_L$  viene così caricata da una corrente via via minore quanto più lei stessa si carica. La salita di  $V_U$  è quindi sempre meno veloce con più passa il tempo. Poiché la corrente è solo dettata da  $R_L$  e  $C_L$ , l'andamento di  $V_U$  è il classico andamento esponenziale tipico di una rete RC. Riassumendo, possiamo dire che il tempo di commutazione in fase di **pull-up** dell'uscita è **lungo** e pari a circa 3



volte la costante di tempo  $R_L \cdot C_L$ :  $t_{pu} @ 3 \cdot R_L \cdot C_L$ .

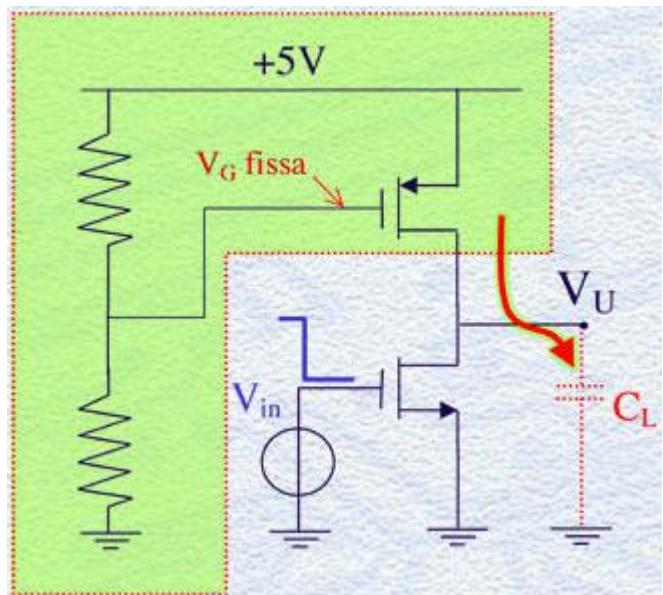
In base a quanto appena detto circa la corrente di carica della capacità che diminuisce al crescere della tensione  $V_U$ , sorge naturale proporre la seguente semplice modifica al circuito: sostituire la resistenza  $R_L$  con un generatore di corrente. In questo modo, la corrente  $I_G$  da esso portata rimane costante per tutta la durata della salita di  $V_U$ , rendendo la commutazione più veloce, tanto più veloce quanto maggiore è il valore di  $I_G$ .



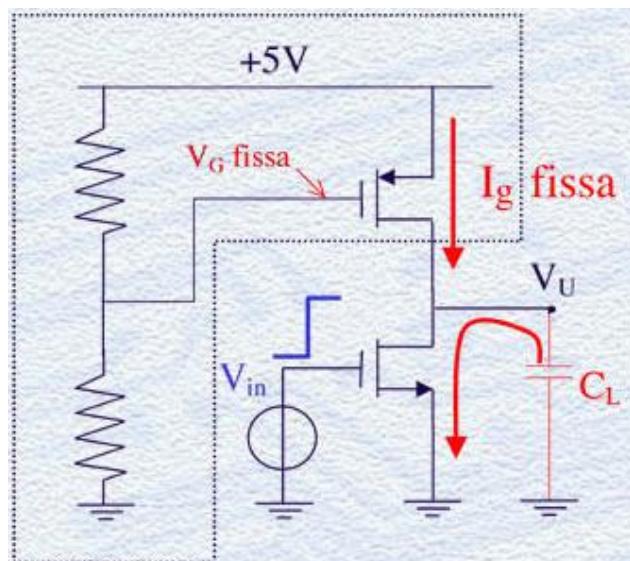
Come abbiamo visto nella lezione scorsa, il generatore di corrente può essere facilmente realizzato con un semplice pMOSFET polarizzato a tensione di Gate fissa ad un valore tale da far portare al transistore la corrente  $I_G$  desiderata. Quando  $V_{in}$  transisce da alto a basso, la corrente di carica della capacità è proprio quella fornita dal pMOSFET e tale rimane fin quando l'uscita ha quasi raggiunto i +5V.

Quando  $V_U$  si avvicina ai +5V, la tensione  $V_{DS}$  ai capi del pMOSFET si è ridotta così tanto da far uscire il pMOSFET dalla saturazione e farlo entrare in zona ohmica. Esso quindi inizia a portare sempre meno corrente fino ad annullarsi completamente quando  $V_U$  arriva a +5V. Questo comportamento del generatore di corrente quando l'ingresso è basso è vantaggioso ai fini del funzionamento complessivo dell'invertitore perché esso *non fornisce più corrente quando questa non serve più!*

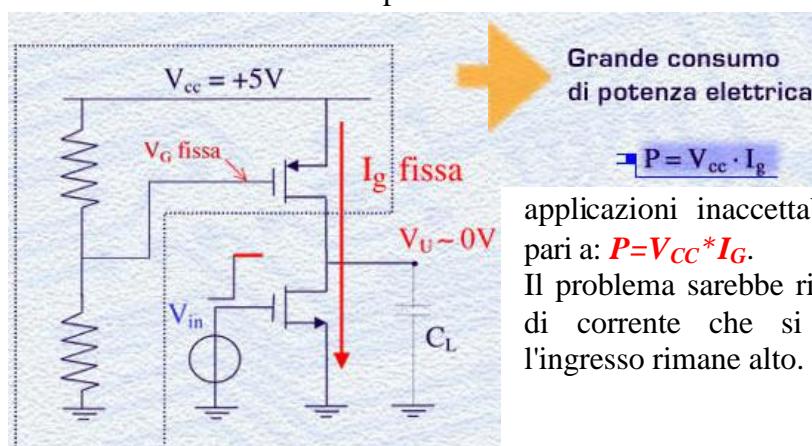
Quando invece il segnale di ingresso è alto, e quindi l'uscita tende ad essere bassa, il pMOSFET è



sempre correttamente in saturazione e porta sempre la stessa corrente  $I_G$  fissa. Poiché l'nMOSFET tende a portare invece molta più corrente di  $I_G$ , la corrente in più viene inizialmente utilizzata per svuotare la capacità  $C_L$  e rendere così veloce la transizione verso il basso dell'uscita. Quando  $V_U$  è scesa vicino a 0V, l'nMOSFET esce dalla saturazione ed inizia a portare sempre meno corrente. Alla fine della transizione, l'nMOSFET sarà costretto a portare esattamente la stessa corrente  $I_G$  che gli arriva dall'alto dal generatore di corrente a pMOSFET.



Questa situazione finale a cui siamo giunti è però svantaggiosa. Infatti, per tutto il tempo in cui  $V_{in}$  rimane alto a +5V e corrispondentemente l'uscita rimane bassa a circa 0V, scorre una corrente continua  $I_G$  attraverso i due transistori, che parte dall'alimentazione positiva ed arriva a massa. Questo fatto produce una elevata, e nella maggioranza delle applicazioni inaccettabile, dissipazione di potenza elettrica pari a:  $P = V_{CC} \cdot I_G$ .



**Grande consumo di potenza elettrica:**

$$P = V_{CC} \cdot I_G$$

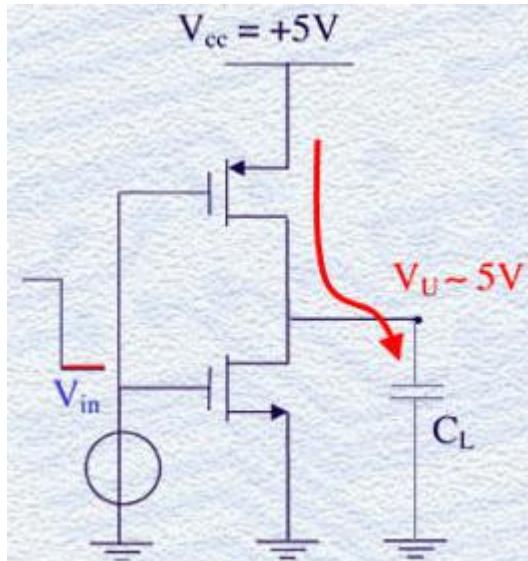
Il problema sarebbe risolto se si disponesse di un generatore di corrente che si spegnesse automaticamente quando l'ingresso rimane alto.

## L'invertitore CMOS

Questo è proprio l'effetto che si ottiene se si collega direttamente il segnale d'ingresso  $V_{in}$  anche al Gate del pMOSFET, eliminando il partitore resistivo. Così facendo, infatti, il **pMOSFET**, quando  $V_{in}$  è alto, viene ad avere una tensione  $V_{GS}$  pari a 0V e quindi risulta **spento**, come si voleva succedesse. Per quanto riguarda invece il funzionamento dell'**nMOSFET** esso risulta inizialmente **acceso** e quindi svuota la capacità  $C_L$  facendo scendere  $V_U$ , poi entra in zona ohmica e si spegne; sostanzialmente, per l'nMOSFET, non cambia nulla rispetto a quanto visto prima.

E' per certi aspetti clamoroso come una modifica così piccola possa dare luogo ad un circuito così semplice, fatto unicamente di due transistori, ma che risolve esattamente il problema che avevamo riscontrato: ora infatti non scorrendo più corrente stazionaria tra  $V_{CC}$  e massa, si è riusciti a ridurre a zero la sua dissipazione di potenza !

Per essere certi di potere adottare definitivamente questo circuito, vediamo come si comporta nel caso di segnale di ingresso  $V_{in}$  basso a 0V. Ora è il transistore nMOS che è spento, mentre il pMOS è sicuramente acceso con il massimo della tensione  $V_{GS}$  possibile, pari sostanzialmente a tutta l'alimentazione.



Questo fa sì che il pMOSFET porti, almeno inizialmente, moltissima corrente e quindi carichi  $C_L$  molto più velocemente di quanto non succedesse prima con il pMOSFET collegato con  $V_G$  fissa. Quindi  $V_U$  sale più velocemente di prima. Quando  $V_U$  arriva alto, come prima il pMOSFET entra in zona ohmica ed è costretto a portare a regime una corrente nulla. Ne consegue una potenza dissipata di nuovo nulla.

Concludendo, anche con  $V_{in}$  basso, l'invertitore così trovato si comporta bene se non addirittura meglio della configurazione precedente.

L'invertitore cosiffatto è chiamato **INVERTITORE A MOSFET COMPLEMENTARI** o più sinteticamente

**INVERTITORE CMOS**. Il nome deriva dalla presenza di un MOSFET a canale *n* e del suo complementare a canale *p*. Per quanto sinteticamente visto, esso è un piccolo gioiello di circuito ed è ormai utilizzato nella stragrande maggioranza dei circuiti digitali. Quando si legge sulle riviste tecniche che un certo processore è realizzato con qualche miliardo di transistori, state pur certi che una buona parte di essi va a formare invertitori CMOS.

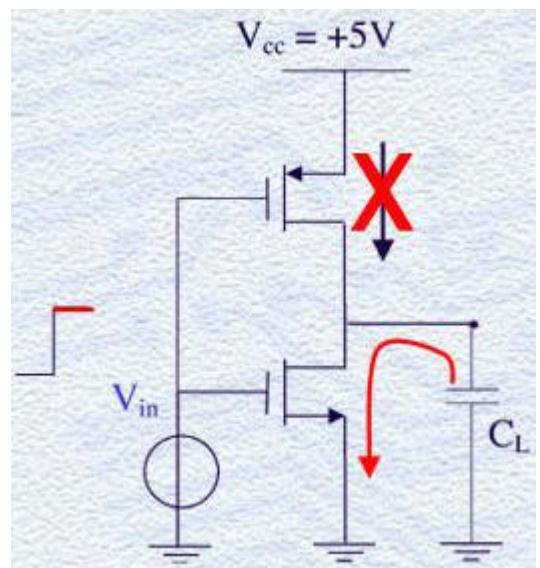
### Esercizio 1

Sapreste progettare un invertitore utilizzando un *pMOSFET*? Provate a farlo immaginando di alimentarlo con una tensione di 3V, utilizzando un *pMOSFET* con  $|V_T|=0.8V$  e  $k=500\mu A/V^2$ .

Quale valore minimo deve avere la resistenza  $R_L$  affinché l'inversione avvenga correttamente in corrispondenza di un segnale digitale con livelli 0V e 3V ?

### Esercizio 2

Provate a confrontare tra di loro i tempi di commutazione,  $t_{pu}$  e  $t_{pd}$ , calcolandoli per il circuito visto nel modulo 2, avente  $R_L=500\Omega$  e  $C_L=0.1pF$ .



## Lezione n. 11 : L'amplificatore di segnali analogici

### La polarizzazione di un circuito

In questa lezione vediamo come reagisce un MOSFET all'applicazione sul Gate di un segnale variabile nel tempo che indichiamo con  $v_{in}$ . Poiché questo segnale potrebbe anche essere piccolo, se lo applicassimo direttamente al Gate del MOSFET esso potrebbe non produrre alcun segnale di corrente al Drain perché sarebbe inferiore alla soglia minima di funzionamento del transistore. Pertanto dovremo pensare di sovrapporre a questo segnale analogico variabile un segnale di tensione costante, che chiamiamo  $V_{pol}$ , che attivi il MOSFET indipendentemente dalla presenza di  $v_{in}$ .

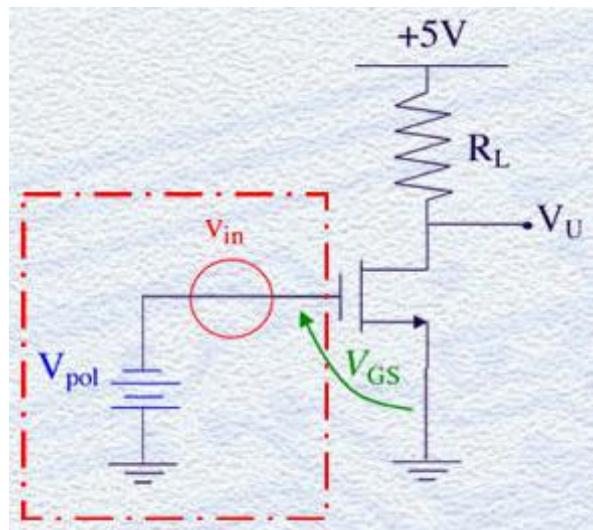
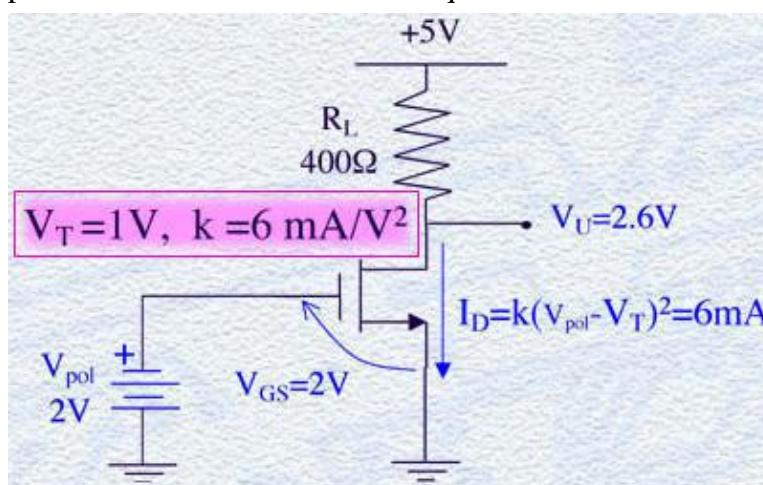
Così è stato fatto collegando in serie i due generatori cosicché i loro valori di tensione si sommino per andare a formare la  $V_{GS}$  totale effettivamente applicata tra Gate e Source del MOSFET.

La novità della presenza del generatore di tensione fissa che attivi il MOSFET indipendentemente dal segnale  $v_{in}$  merita qualche approfondimento.

La sua funzione è quella di definire una tensione tra Gate e Source del MOSFET superiore alla soglia affinché il transistore porti una corrente stazionaria prefissata. Questa funzione viene chiamata "**polarizzazione**" del transistore, o più in generale del circuito, e corrisponde a scegliere le correnti e le tensioni stazionarie che verranno poi cambiate dal segnale applicato. Bisogna pertanto mettersi nella condizione che:

- nel MOSFET circoli corrente,
- che il MOSFET lavori in saturazione,
- che la tensione  $V_U$  sia tale da potersi spostare quando gli si applica il segnale  $v_{in}$ .

Queste considerazioni fanno intuire come non esista una unica soluzione di polarizzazione ma ce ne possano essere tante tutte comunque soddisfacenti. Vedremo in seguito alcune delle conseguenze di



una scelta di polarizzazione piuttosto che un'altra da cui potremo dedurre dei criteri guida per operare al meglio. Se prendiamo il circuito della figura, utilizzante un MOSFET con  $V_T=1V$  e  $k=6\text{mA/V}^2$ , una possibile scelta di polarizzazione è quella di fissare  $V_{pol}$  a 2V, cosicché la tensione di overdrive del MOSFET sia di 1V. Questo produce una corrente di Drain di 6mA e posiziona l'uscita  $V_U$  alla tensione di 2.6V rispetto a massa. E' immediato verificare che il MOSFET è effettivamente in saturazione: infatti la sua tensione di

Drain (2.6V) è maggiore di quella di Gate (2V). Il MOSFET rimane in saturazione fintanto che il suo Drain resta sopra il valore di +1V. Come già accennato, anche altri valori di  $V_{pol}$  sarebbero stati possibili, ad esempio  $V_{pol}=1.5V$  o  $V_{pol}=2.2V$ . Vedremo più avanti quali criteri aggiuntivi guidano nella scelta della polarizzazione più appropriata.

### Funzionamento di un amplificatore invertente

A questo punto sovrapponiamo alla tensione stazionaria prima impostata il segnale  $v_{in}$  variabile nel tempo. Innanzitutto vediamo che, essendo le due tensioni applicate in serie, esse si sommano e definiscono una tensione totale tra Gate e Source del MOSFET pari a  $V_{GS}=V_{pol}+v_{in}$ .

Il nostro fine è quello di calcolare come si modificano la corrente di Drain, e conseguentemente la tensione di uscita  $V_U$ , per effetto del segnale  $v_{in}$  applicato.

Per rendere più chiara la rappresentazione, in blu ed a caratteri maiuscoli sono indicati i valori costanti di polarizzazione, in rosso ed a caratteri minuscoli sono indicati i segnali variabili ed in verde ed a caratteri maiuscoli corsivi sono indicate le composizioni dei segnali con la polarizzazione.

La corrente totale di collettore sarà

$$\text{data dalla relazione: } I_D = k(V_{pol} + v_{in} - V_T)^2 \text{ e se si sviluppa il quadrato si ottiene: } I_D = k((V_{pol} - V_T) + v_{in})^2 \\ = k(V_{pol} - V_T)^2 + 2k(V_{pol} - V_T) * v_{in} + k v_{in}^2.$$

La formula mette in luce tre addendi:

- il primo  $k(V_{pol} - V_T)^2$  è esattamente il valore di corrente di polarizzazione calcolato in assenza del segnale;
- il secondo addendo  $2k(V_{pol} - V_T) * v_{in}$  è funzione lineare del segnale di tensione  $v_{in}$  in cui il fattore di proporzionalità dipende ancora dalla scelta della polarizzazione attraverso il termine  $(V_{pol} - V_T)$ ;
- il terzo addendo  $k v_{in}^2$  è funzione quadratica del segnale  $v_{in}$  secondo il fattore  $k$  di costruzione del MOSFET.

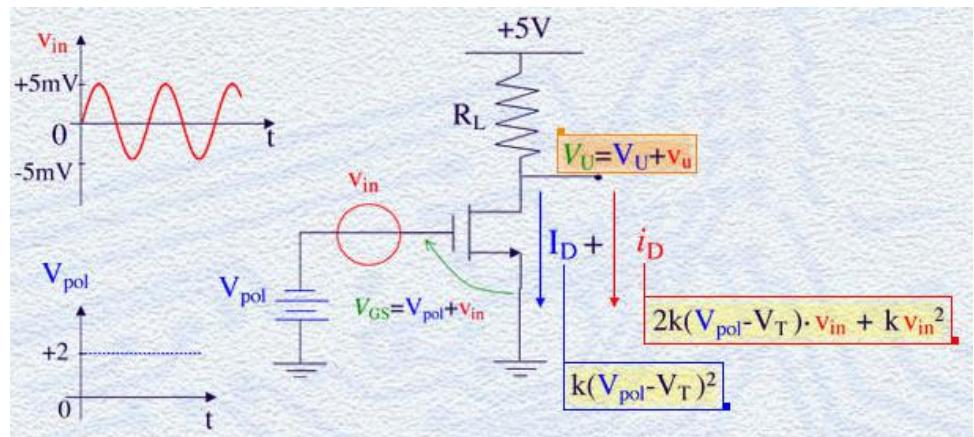
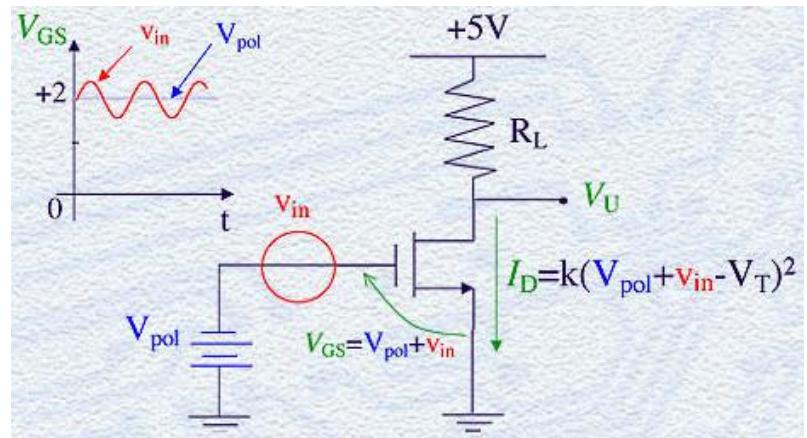
La prima interessante considerazione che si può fare è che la corrente totale al Drain è pari alla SOMMA della corrente di polarizzazione con la corrente generata dal solo segnale d'ingresso  $v_{in}$  ossia  $I_D = I_D + i_D = I_D + 2k(V_{pol} - V_T) * v_{in} + k v_{in}^2$

Grazie al fatto che le due correnti si sommano, nel progetto e nell'analisi dei circuiti esse vengono calcolate separatamente, prima la polarizzazione e poi il segnale. Questa è una comodità enorme, sia dal punto di vista della semplificazione dei calcoli sia da quello della facilità di comprensione del funzionamento del circuito. Si noti che, poiché le correnti di Drain si sommano, anche la tensione di uscita totale  $V_U$  sarà necessariamente data dalla somma della tensione costante di uscita fornita dalla

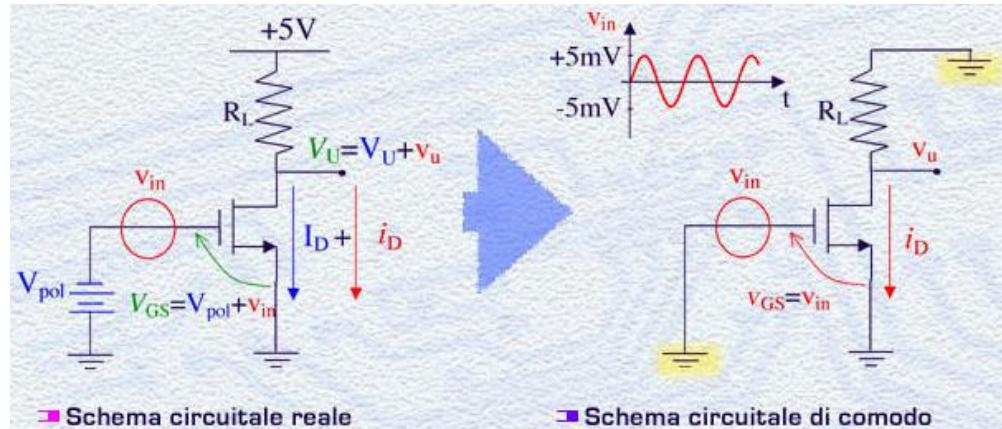
polarizzazione più la variazione dovuta alla sola variazione di corrente innescata dal segnale  $v_{in}$ .

In maniera sintetica, possiamo dire che il segnale si sovrappone ai valori stazionari definiti dalla polarizzazione e ne costituisce una variazione, in genere piccola.

Dato che abbiamo già visto come si analizza il circuito per la polarizzazione e come se ne calcola la corrente di Drain, vediamo ora come sia possibile rappresentare il circuito in modo da permettere il calcolo del solo effetto dovuto al segnale di tensione di ingresso  $v_{in}$ .



Poiché il segnale produce solo delle variazioni ad una situazione preesistente fissata dalla polarizzazione, ci si deve concentrare solo su di esse. Tutti i valori fissi di tensione, qualunque siano i loro effettivi valori, saranno per noi dei punti fissi di tensione tutti equivalenti e pertanto li possiamo tutti indicare con lo stesso simbolo che, per comodità, prendiamo quello di massa. Nella figura questo è stato fatto per il generatore costante  $V_{pol}$  e per l'alimentazione costante di +5V. Per quanto riguarda le correnti stazionarie di polarizzazione alle quali le correnti di segnale si sommano, non si dovrà fare altro che non indicarle esplicitamente. Il circuito che rimane è quello che formalmente mi permette di calcolare tutte le variazioni di tensione e di corrente nel circuito.



Naturalmente, *non bisogna mai dimenticare che questa rappresentazione circuitale è una rappresentazione di comodo che nasconde una polarizzazione sempre presente. Pertanto se pensiamo, ad esempio, di non dare segnale (cioè  $v_{in}=0$ ) non si giunga alla conclusione che il MOSFET ha  $V_{GS}=0$  e quindi non porta alcuna corrente. La conclusione corretta è che il MOSFET non porta alcuna corrente in più di quella che circola già per la polarizzazione e data da  $V_{GS}=V_{pol}$ .*

A questo punto possiamo concentrarci sul solo segnale di corrente di Drain prodotto dal solo segnale di tensione di ingresso  $v_{in}$ .

L'espressione della corrente avevamo visto essere:  $i_D = 2k(V_{pol} - V_T) * v_{in} + k * v_{in}^2$

Il fattore di proporzionalità,  $2k(V_{pol} - V_T)$ , tra  $v_{in}$  e  $i_D$  del primo addendo è chiamato "**La transconduttanza del MOSFET**" ed è indicata con  $g_m$ . Il termine equivale infatti a quello che si otterebbe derivando rispetto a  $V_{GS}$  la relazione caratteristica della corrente di Drain. La transconduttanza quindi rappresenta la variazione della corrente di Drain che si riesce ad ottenere con una piccola variazione della tensione Gate-Source. Proprio per questo, la transconduttanza del MOSFET è uno dei suoi parametri più importanti e dà proprio l'idea di quanto si sia efficaci nel produrre corrente al Drain con il segnale  $v_{in}$  applicato tra Gate e Source.

L'espressione della transconduttanza comprende il termine ( $V_{pol} - V_T$ ) che costituisce l'overdrive del MOSFET e dipende da come si era polarizzato il circuito. Questo aspetto può essere messo ulteriormente in luce ricavando ( $V_{pol} - V_T$ ) dalla relazione caratteristica del MOSFET e sostituendola nella  $g_m$ : si otterebbe che  $g_m$  dipende dalla radice quadrata della corrente di polarizzazione circolante nel MOSFET. Pertanto maggiore è la corrente di polarizzazione, maggiore è la transconduttanza del MOSFET quando gli si applica un segnale, (ma è anche maggiore la potenza stazionaria dissipata dal transistore).

Ecco che la transconduttanza costituisce un primo legame tra polarizzazione del circuito e suo comportamento su segnale, di cui quindi tenere conto quando si polarizza il circuito.

Il secondo addendo presente nella espressione della corrente di segnale dipende dal quadrato della tensione  $v_{in}$  secondo il fattore  $k$ . Si noti che se  $g_m v_{in} \gg k v_{in}^2$  cioè se  $v_{in} \ll (g_m/k) = 2(V_{pol} - V_T)$  questo secondo addendo può essere anche trascurato, e che la corrente dipende solo LINEARMENTE dal segnale di ingresso  $v_{in}$  secondo la semplicissima relazione:  $i_D = g_m v_{in}$ .

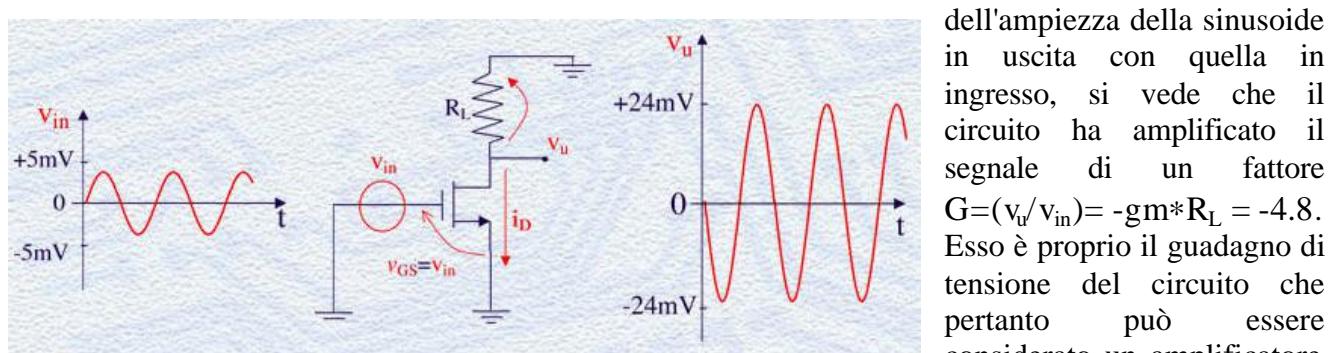
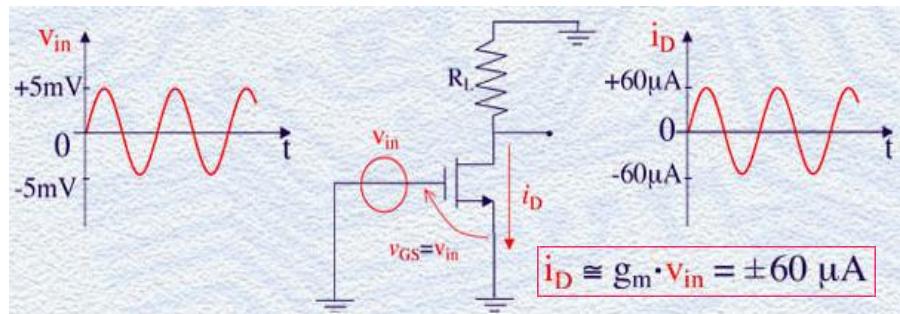
$$g_m = \frac{\partial I_D}{\partial V_{GS}} = \frac{\partial k(V_{GS} - V_T)^2}{\partial V_{GS}} = 2 \cdot k(V_{GS} - V_T)$$

$$g_m = 2k(V_{GS} - V_T) = 2\sqrt{k \cdot I_D}$$

La situazione in cui  $v_{in} \ll 2(V_{pol} - V_T)$  per cui vale la linearizzazione citata, è detta “*condizione di piccolo segnale*” applicato al circuito. Rientrano in questa categoria segnale  $v_{in}$  fino a qualche centinaia di mV pensando a tensioni di overdrive attorno ad un volt. Riprendiamo quindi il circuito in esame. Poiché  $v_{in}$  è di  $\pm 5\text{mV}$  e la tensione di overdrive di 1V, siamo in condizioni di piccolo segnale. La transconduttanza, ricavata in base alla polarizzazione scelta prima, vale  $g_m = 12\text{mA/V}$ , pertanto la corrente di segnale sarà una corrente sinusoidale in fase con la sinusoide di tensione applicata all'ingresso ed avente una ampiezza picco-picco di  $120\text{\mu A}$ .

La corrente di segnale  $i_D$  trovata scorre in  $R_L$  e determina ai suoi capi una variazione di tensione pari a  $v_u = -i_D R_L = -g_m R_L v_{in}$  in cui il segno meno deriva dal fatto che quando  $i_D$  aumenta nel verso della freccia, la caduta di tensione in  $R_L$  aumenta e, poiché il capo sopra della resistenza è fissato a massa, il capo sotto (cioè  $v_u$ ) deve necessariamente scendere in tensione.

La variazione della tensione di uscita sarà quindi anch'essa sinusoidale, di ampiezza di picco pari a  $v_u = -12\text{mA/V} * 400\Omega * 5\text{mV} = -24\text{mV}$  e sfasata di  $180^\circ$  rispetto all'ingresso. Confrontando il valore



di tensione. Il segno meno, derivante dallo sfasamento, ci dice anche che l'amplificatore è *invertente*. Come si vede il guadagno dipende dalla resistenza  $R_L$  e dalla  $g_m$  del MOSFET: esso cioè dipende da come si era disegnato il circuito (la scelta di  $R_L$ ) e da come lo si era polarizzato (la corrente stazionaria circolante, e quindi le tensioni di alimentazione).

L'arte del progettare circuiti elettronici sta proprio nel calibrare la scelta dei componenti e la determinazione delle condizioni di polarizzazione per avere il guadagno su segnale desiderato senza avere troppi altri svantaggi collaterali.

Tra questi c'è ad esempio:

- la potenza dissipata (se si aumenta la corrente di polarizzazione per avere transconduttanza più elevata, contemporaneamente si dissiperà molta più potenza elettrica)
- la tensione di alimentazione troppo alta (se aumento  $R_L$  per avere un guadagno maggiore, dovrò corrispondentemente aumentare la tensione di alimentazione dai 5V previsti ai 10 o 15V per fare in modo che possa scorrere la stessa corrente di polarizzazione, e ciò non è sempre accettabile, si pensi agli apparecchi portatili alimentati da delle pile).

### Accoppiamenti DC e AC

Normalmente la polarizzazione non viene imposta al circuito tramite un generatore specifico per questa funzione, ma si fa in modo di sfruttare i generatori fissi già necessariamente presenti nel circuito. Un tipico esempio è l'uso delle tensioni di alimentazione per fissare la tensione di polarizzazione del Gate del MOSFET. La figura che segue, mostra come un semplice partitore di tensione possa svolgere egregiamente questa funzione. Scegliendo ad esempio  $R_1=60\text{k}\Omega$  e  $R_2=40\text{k}\Omega$  si riottiene  $V_{pol}=2\text{V}$  come desiderato, sul Gate del MOSFET.

In questa nuova situazione bisogna trovare il modo di aggiungere alla tensione di polarizzazione il segnale  $v_{in}$ . La soluzione di *collegare direttamente il generatore di tensione di segnale  $v_{in}$  al Gate del MOSFET è sbagliata!*

Infatti, nei momenti in cui non c'è segnale, cioè nel caso in cui  $v_{in}=0V$ , il generatore di segnale imporrebbe al Gate la sua tensione di 0V invece che lasciare quel punto al valore di  $V_{pol}=2V$  a cui dovrebbe stare. Il generatore di segnale entrerebbe quindi in conflitto con la polarizzazione invece di convivere con essa aggiungendosi alla polarizzazione stessa.

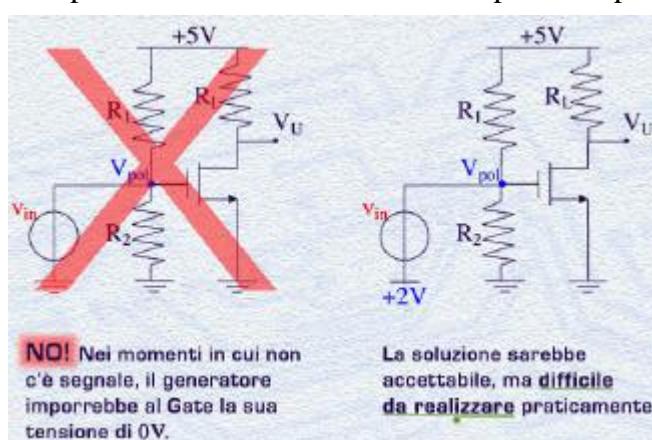
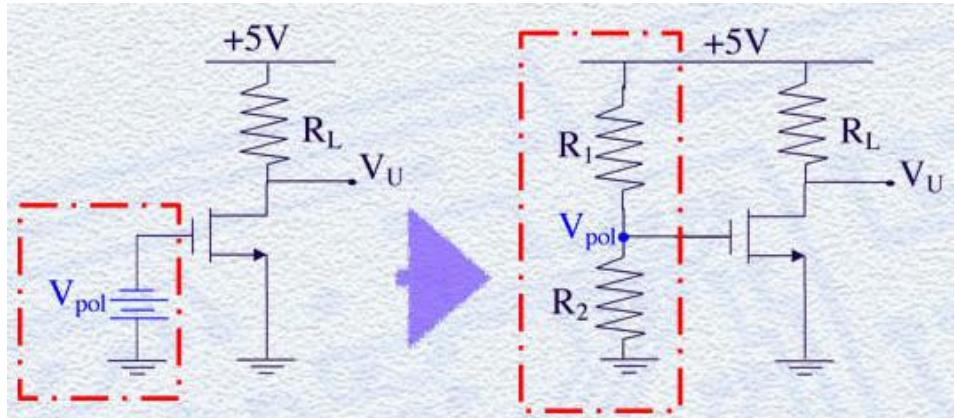
Una prima soluzione corretta a cui si potrebbe pensare è quella di riferire il generatore di segnale non più a massa ma proprio al valore di 2V di  $V_{pol}$ .

In questo modo, quando  $v_{in}=0V$  il generatore applicherebbe al Gate la tensione di 2V che è proprio quella lì imposta dal partitore di polarizzazione. In questo modo non ci sarebbe conflitto, e quindi la soluzione sarebbe accettabile, ma rimarrebbe la difficoltà pratica di realizzare ciò.

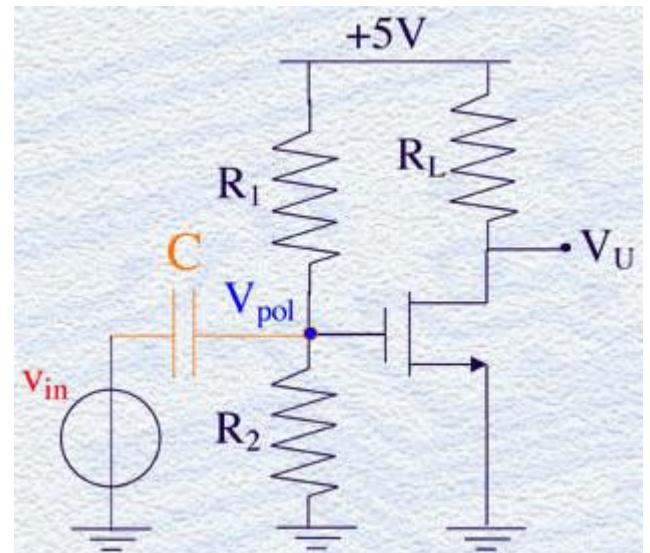
Una soluzione spesso più semplice da realizzare è quella di collegare il generatore di segnale al Gate del MOSFET tramite un condensatore. La capacità C rende infatti indipendenti i valori di tensione ai suoi capi in polarizzazione, che possono quindi assumere i valori che il circuito impone loro. Nel nostro caso il capo a sinistra può stare a 0V (corrispondente a  $v_{in}=0V$ ) ed il capo a destra può stare a 2V (corrispondente a  $V_{pol}=2V$ ).

Ma il condensatore ha anche la proprietà di trasmettere il segnale variabile  $v_{in}$  rendendo possibile la variazione del Gate del MOSFET proprio di  $v_{in}$ . Poiché il condensatore forma, con le resistenze di polarizzazione, una rete RC, la bontà del trasferimento del segnale dipende dalla sua frequenza rispetto alla costante di tempo RC: maggiore è la capacità C e migliore è il trasferimento del segnale anche per segnali a bassa frequenza.

Il collegamento del circuito al segnale  $v_{in}$  tramite una capacità è chiamato in gergo "*collegamento in AC*" per sottolineare che solo il segnale passa e non la polarizzazione, distinguendolo dal collegamento diretto del generatore al Gate del MOSFET visto prima chiamato "*collegamento in DC*" per sottolineare che anche la continua può venire trasmessa.



La soluzione sarebbe accettabile, ma difficile da realizzare praticamente.



### Comportamento di un amplificatore con grandi segnali all'ingresso

Rimuoviamo ora l'ipotesi fatta che il segnale di ingresso  $v_{in}$  al circuito sia piccolo. In tal caso la corrente di segnale prodotta dal MOSFET è data dalla relazione completa:

$$i_D = +2k(V_{pol} - V_T) * v_{in} + k * v_{in}^2$$

e non è più possibile trascurarne il secondo addendo.

La relazione tra la  $v_{in}$  d'ingresso e la corrente d'uscita **non è più lineare**. Questo fatto in generale disturba, perché vuole dire che l'amplificazione del circuito non è più un numero fisso ma dipende dall'ampiezza del segnale stesso fornito in ingresso.

Ma questa considerazione non è la sola a suggerire di evitare una tale situazione circuitale.

Ben più importante è la seguente.

Si consideri di avere un segnale di ingresso sinusoidale di frequenza  $f$ :  $v_{in} = A_0 \sin(2\pi f t)$

La corrispondente corrente di segnale nel Drain è pari a:  $i_D = g_m A_0 \sin(2\pi f t) + k A_0^2 \sin^2(2\pi f t)$

Sviluppando il quadrato della funzione seno, si ottiene  $i_D = g_m A_0 \sin(2\pi f t) + k A_0^2/2 - k A_0^2 \cos(2 \cdot 2\pi f t)/2$

Come si vede, la corrente di segnale, a causa del termine quadratico, dipende in maniera più articolata dal segnale d'ingresso di quanto non lo fosse prima.

In particolare si nota che sorge un termine cosinusoidale avente FREQUENZA DOPPIA rispetto a quella del segnale di ingresso. Questo fatto è a dir poco imbarazzante. Immaginate di usare il circuito per amplificare il suono di uno strumento musicale: benché lo strumento generi una certa armonica ad una definita frequenza, il suo segnale amplificato contiene anche l'armonica alla frequenza doppia!

Proprio per questo, il fenomeno, spurio e non voluto, prende il nome di "*distorsione armonica*". Esso nasce dalle non-linearietà del dispositivo amplificante ed è esaltato quando i segnali di ingresso sono "*grandi*".

Nella maggior parte delle applicazioni, la distorsione armonica del circuito deve essere evitata o almeno ridotta. Le tecniche sono le più varie:

- riduzione dei segnali applicati ai capi del MOSFET,
- inserimento nel circuito di filtri che taglino le armoniche superiori,
- altro.

I corsi più avanzati di progettazione dei circuiti elettronici analizzano in dettaglio tutti questi aspetti, in modo da fornire le conoscenze e le tecniche per un progetto consapevole anche di circuiti molto complessi. Tuttavia non si pensi che le distorsioni armoniche siano sempre da evitare. Ci sono applicazioni anzi che le generano volutamente per sfruttarle in modo nobile. Nel semplice caso che abbiamo visto noi, la distorsione viene utilissima se si volesse proprio generare una sinusode ad alta frequenza avendo a disposizione solo un segnale sinusoidale a frequenza più bassa della metà. In questo caso, se facessimo seguire al circuito disegnato un filtro passa alto che lasci passare l'armonica a frequenza  $2f$  e tagliasse quella a frequenza  $f$ , avremmo realizzato un circuito generatore di tensione sinusoidale di frequenza doppia dell'ingresso.

### Esercizio 1

Calcolare la polarizzazione del circuito (a) che utilizza un *n*MOSFET con  $V_T=0.8V$  e  $k=10mA/V^2$ . Verificare che il MOSFET funzioni correttamente in saturazione e trovare quale sarebbe il valore di  $V_{pol}$  massima oltre cui il MOSFET non sarebbe più polarizzato correttamente. Rispondere alle stesse domande riferite al circuito (b) il cui pMOSFET ha  $|V_T|=1.2V$  e  $|k|=1mA/V^2$ .

### Esercizio 2

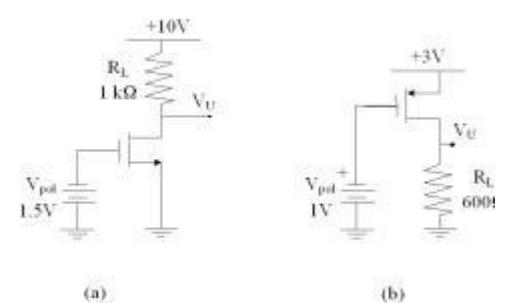
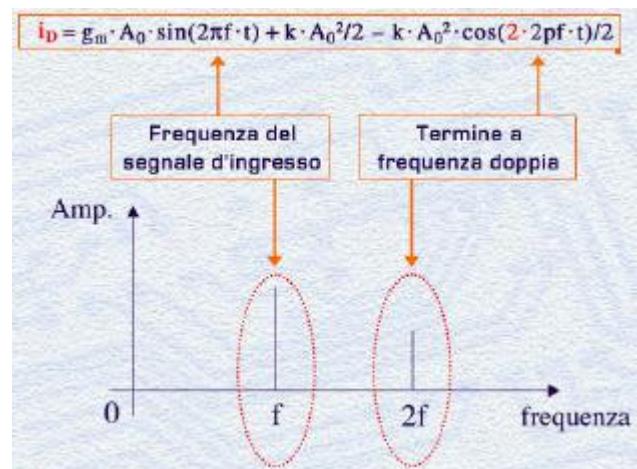
Riprendere i due circuiti dell'esercizio 1 di cui si era calcolata la polarizzazione e calcolare nell'ordine e per ognuno di essi:

- la transconduttanza  $g_m$ ;
- il guadagno  $G$  di tensione tra ingresso ed uscita;
- l'ampiezza del segnale di uscita quando all'ingresso è applicata una sinusode di ampiezza  $\pm 12mV$ .

Da ultimo, provate a fare i progettisti! Modificando  $V_{pol}$  e  $R_L$ , riuscite a raddoppiare il guadagno degli amplificatori?

### Esercizio 3

Se scegliessimo, nel circuito appena visto, di usare una capacità di disaccoppiamento AC una  $C=100nF$ , da quale frequenza in poi si riuscirebbe a trasmettere il segnale da  $v_{in}$  al MOSFET?

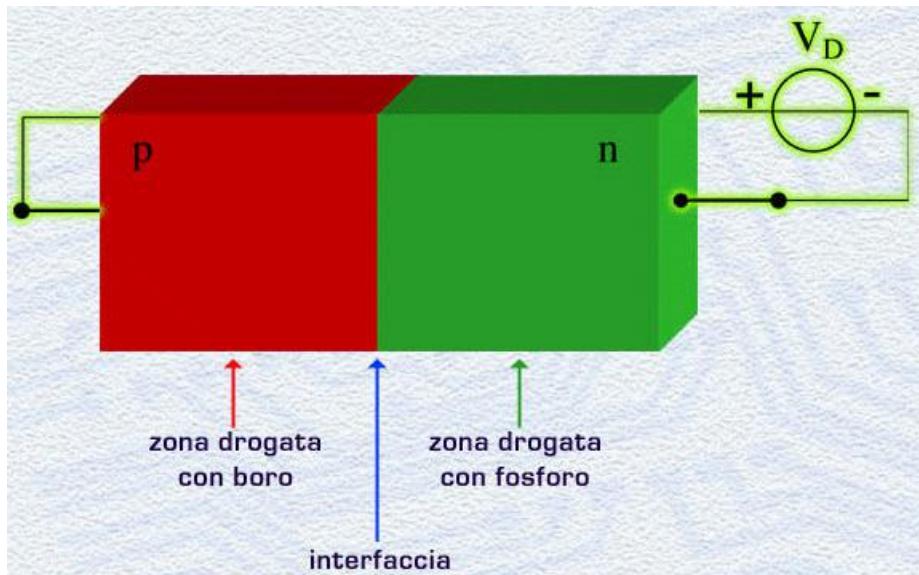


## Lezione n. 12 : Il diodo a giunzione

### La giunzione P-N

Il diodo a giunzione è un dispositivo elettronico molto importante per l'elettronica ed anche il primo ad essere stato prodotto con un cristallo semiconduttore negli anni quaranta.

Esso è oggi realizzato partendo da un cristallo di silicio drogato, per una parte, con atomi di boro e per un'altra parte subito contigua, con atomi di fosforo. Per semplicità supponiamo che la transizione da silicio drogato *p* a silicio drogato *n* avvenga bruscamente. Una tale realizzazione viene anche chiamata "**Giunzione P-N**". La zona di transizione è chiamata interfaccia della giunzione. Il dispositivo ha inoltre un collegamento elettrico ad ognuna delle due zone della giunzione, tramite cui applicare una tensione ed attraverso cui vedere scorrere una corrente.



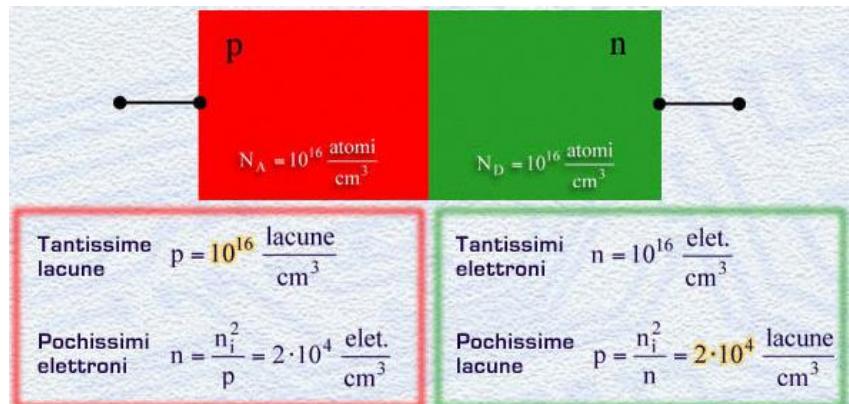
Per capire il suo funzionamento dobbiamo concentrarci, come abbiamo fatto per il MOSFET, sul movimento dei portatori di carica, elettroni e lacune, nel cristallo di silicio prima e dopo l'applicazione di una tensione applicata dall'esterno.

Incominciamo a vedere cosa succede quando alla "giunzione *p-n*" non viene applicata alcuna tensione. Il dispositivo è cioè all'equilibrio con  $V_D=0V$ .

La situazione in cui da una parte dell'interfaccia ci sono tantissime lacune, perché lì abbiamo drogato *p*, e dall'altra ci sono tantissimi elettroni perché là abbiamo drogato *n*, non può mantenersi nel tempo.

Si noti che lo squilibrio di portatori dello stesso tipo è elevatissimo. Se pensiamo ad esempio di drogare sia la zona *p* che la zona *n* con  $10^{16}$  atomi di drogante (boro a sinistra e fosforo a destra) a sinistra avremo  $10^{16}$  lacune/cm<sup>3</sup> mentre a destra ne avremo solo  $2 \cdot 10^4$ . La differenza è di ben 12 ordini di grandezza (pari ad un milione di milioni), una quantità enorme! La stessa cosa vale per gli elettroni. Questa disuniformità di concentrazione nello spazio non si mantiene nel tempo. Infatti la naturale tendenza di un sistema costituito da elementi distribuiti disuniformemente nello spazio è quella di portarsi ad una situazione di omogeneità, a meno che non intervengano altri fattori. Di questa tendenza, assolutamente generale nel mondo fisico, abbiamo tutti conoscenza quando pensiamo al calore emesso da un calorifero che tende a distribuirsi in tutta una stanza portandola ad una temperatura più o meno uniforme.

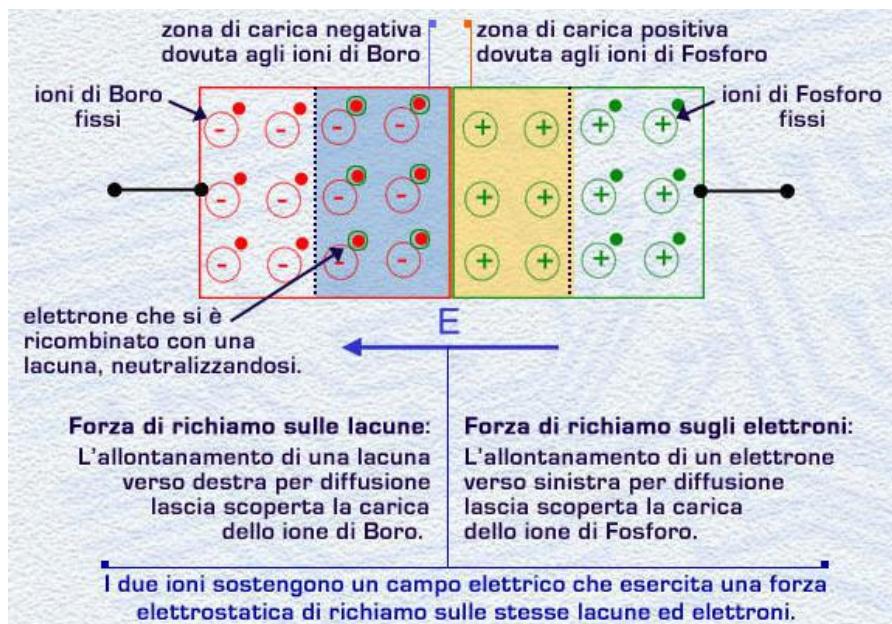
Nel nostro caso la presenza di molte lacune da una parte della giunzione e di pochissime dall'altra, fa sì che si instauri un moto di lacune che va dalla zona dove sono in maggior numero a sinistra verso la zona dove sono in minor numero a destra.



Con una terminologia più fisica si dice che una differenza (detta anche gradiente) di concentrazione di lacune tra due punti attiva un flusso di lacune dalla zona dove sono in maggior numero verso la zona dove sono in minor numero. Il flusso è tanto più intenso quanto maggiore è la differenza di concentrazione iniziale tra i due punti. Il moto sorretto da questo meccanismo è detto "moto di diffusione" delle lacune, in contrapposizione con il "moto

di deriva" dei portatori definito nella 6<sup>a</sup> lezione che è quello in cui il campo elettrico che trascina le cariche, su cui si basa ad esempio il funzionamento del MOSFET. Quanto visto per le lacune avviene ovviamente anche per gli elettronni, che tendono a muoversi per diffusione da destra verso sinistra in maniera indipendente dalla presenza delle lacune.

Nella giunzione *p-n* la situazione è complicata dal fatto che quando l'elettrone lascia il suo atomo donore per trasferirsi a sinistra della giunzione e lì ricombinarsi con l'atomo accettore, nel suo punto di partenza rimane l'atomo donore con un elettrone in meno, cioè rimane lo ione positivo di fosforo. A sinistra della giunzione, gli atomi di boro che hanno preso un elettrone sono diventati ioni negativi. I due ioni sostengono un campo elettrico che esercita una forza elettrostatica di richiamo sulle stesse lacune ed elettroni.

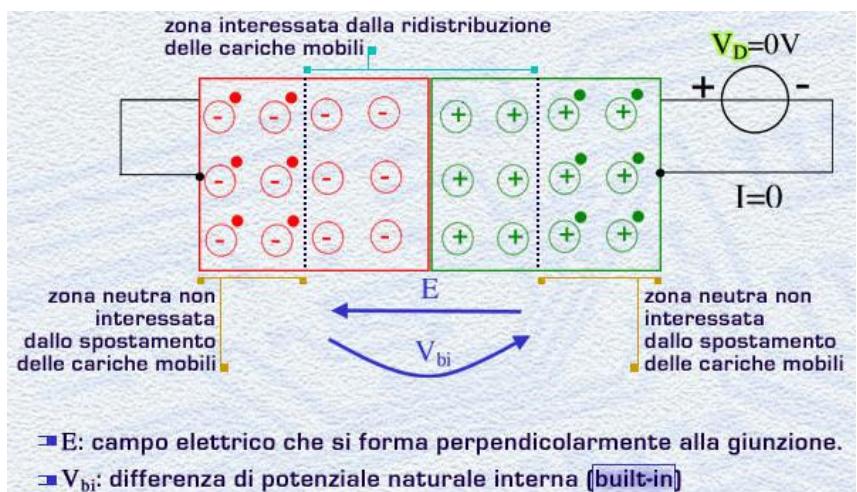


elettroni carichi negativamente. Si crea quindi una situazione di tendenze opposte sugli elettroni: vorrebbero spostarsi a sinistra per diffusione ma vengono richiamati indietro dal campo elettrico attivato dal loro stesso spostamento.

La stessa cosa può essere vista avvenire per le lacune, per le quali la differenza di concentrazione tenderebbe a creare un flusso di diffusione verso destra, contrastato dal campo elettrico di richiamo verso sinistra degli ioni fissi nel cristallo. La situazione finale effettivamente presente è proprio quella che bilancia esattamente questi due fenomeni.

Ad equilibrio raggiunto, si vede che solo una piccola regione a cavallo della giunzione è stata interessata dal movimento delle cariche.

Come si può vedere, a sinistra della giunzione ci sono cariche negative fisse (gli atomi di boro più gli elettroni) ed a destra ci sono cariche positive fisse (gli atomi di fosforo senza gli elettroni). Questa distribuzione di cariche produce un campo elettrico perpendicolare alla giunzione che a sua volta definisce una differenza di



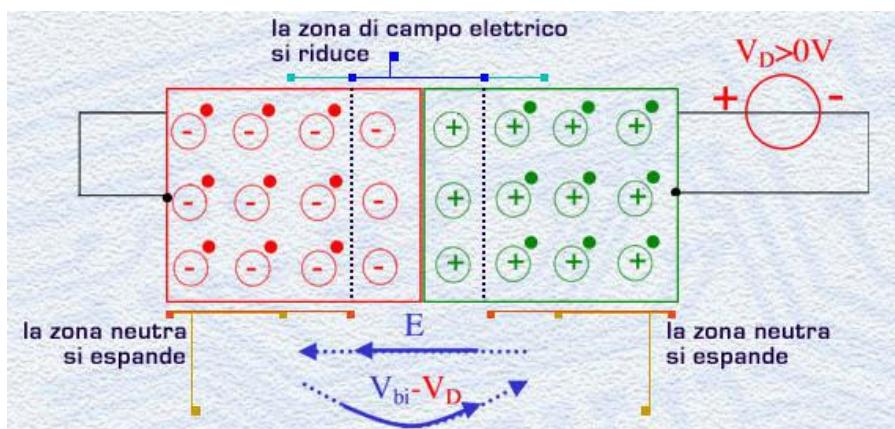
potenziale tra la zona a carica positiva e la zona a carica negativa del diodo. Questa tensione naturale che viene a crearsi all'interno del diodo a cavallo della giunzione quando è all'equilibrio è anche chiamata, con terminologia inglese, tensione di *built-in*, ed indicata con  $V_{bi}$ . Il valore della tensione di built-in dipende da quanto sono drogati le due regioni *p* ed *n* ed è normalmente compreso tra 0.8 e 0.9V.

La corrente circolante in questa situazione è nulla grazie alla situazione di equilibrio.

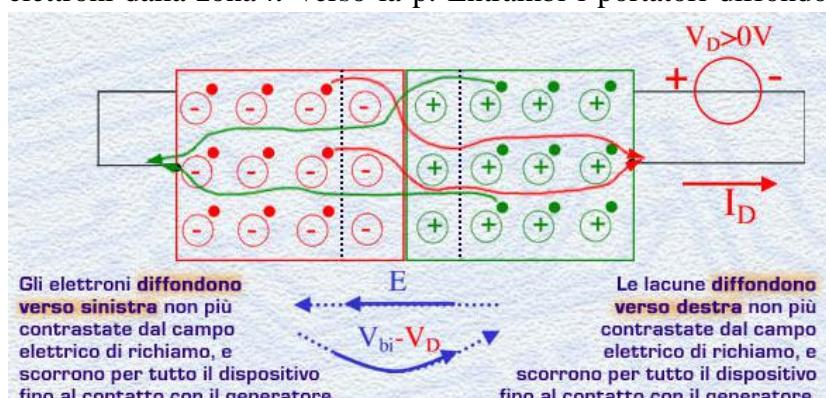
Il principio di funzionamento del diodo si basa tutto sulla possibilità di modificare, tramite la tensione esterna applicata  $V_D$ , la tensione naturale interna creatasi all'equilibrio, in modo da rompere l'equilibrio tra deriva e diffusione prima raggiunto.

## Diodo in diretta

Quando  $V_D$  è applicata con il segno tale da diminuire la tensione naturale, si dice che si sta applicando una tensione diretta al diodo. Per fare ciò, il polo positivo del generatore di tensione esterno è applicato alla zona *p* ed il polo negativo alla zona *n*. In questo caso, poiché si diminuisce la tensione a cavallo della giunzione, viene anche a diminuire il campo elettrico di richiamo per i portatori e quindi diminuisce il numero di ioni presenti. L'equilibrio tra la tendenza alla diffusione in una direzione ed il richiamo del campo elettrico nell'altra è quindi rotto a favore di un maggior flusso di portatori per diffusione, lacune dalla zona *p* verso la *n* ed elettroni dalla zona *n* verso la *p*. Entrambi i portatori diffondono liberamente per tutto il dispositivo



fino ai rispettivi contatti con i fili del collegamento elettrico con il generatore di tensione esterno. Entrambi questi flussi, fatti di cariche di segno opposto muoventesi in direzioni opposte, si sommano nel definire la **corrente totale** circolante  $I_D$ . Fintanto che  $V_D$  rimane applicata e fissa, nuovi portatori partiranno dalle zone neutre e diffonderanno verso il



contatto opposto, alimentando così una corrente continua.

La corrente  $I_D$  è tanto maggiore quanto più la tensione naturale è ridotta rispetto al valore iniziale. La relazione tra  $V_D$  e  $I_D$ , andando a svolgere i conti precisi del processo di diffusione, è in effetti di tipo

$$I_D = I_0 \left( e^{\frac{V_D}{kT/q}} - 1 \right)$$

esponenziale, cioè:  $I_D = I_0 \left( e^{\frac{V_D}{kT/q}} - 1 \right)$ . L'espressione contiene il fattore di proporzionalità  $I_0$  che riassume le caratteristiche costruttive del diodo (geometria, livelli di droggaggio, costanti fisiche) ed una tensione di riferimento  $kT/q$  che, nel caso di funzionamento a temperatura ambiente ( $T=300K$ ), vale circa 25mV.

Per quanto riguarda la geometria,  $I_0$  è direttamente proporzionale alla superficie della interfaccia del diodo: maggiore è questa dimensione, maggiore è la corrente portata a parità di tensione applicata.

Inoltre si noti che basta che  $V_D$  sia maggiore di 100mV che il termine **-1** diventa assolutamente

$$I_D = I_0 \left( e^{\frac{V_D}{kT/q}} - 1 \right) \approx I_0 \cdot e^{\frac{V_D}{kT/q}}$$

trascurabile:  $I_D = I_0 \left( e^{\frac{V_D}{kT/q}} - 1 \right) \approx I_0 \cdot e^{\frac{V_D}{kT/q}}$ . La formula ci dice quindi che quando variamo la tensione  $V_D$  in diretta, la corrente  $I_D$  circolante nel diodo aumenta enormemente come un

esponenziale. Il grafico che mostra il legame corrente-tensione raffigura la curva caratteristica del diodo. Giustamente la relazione ci dice innanzitutto che se non diamo tensione ( $V_D=0V$ ), la corrente circolante si riduce a zero: quindi la curva passa per l'origine degli assi.

Per come abbiamo visto essere il fenomeno fisico, sappiamo anche che al massimo potremo applicare al diodo una tensione diretta pari alla tensione naturale presente all'equilibrio. In tal caso infatti la tensione di *built-in* sarà praticamente azzerata e quindi il flusso di diffusione si esprimerebbe senza più alcun limite. Questo si manifesta a valori di tensione diretta di circa 0.8V poiché è questa la tensione  $V_{bi}$  normalmente presente in un diodo. A questi valori di tensione non si deve mai arrivare però perché si rischierebbe di bruciare il diodo: infatti, essendo la corrente elevatissima, la potenza termica generata (pari a  $I_D \cdot V_D$ ) sarebbe grandissima ed alzerebbe la temperatura del cristallo fino ad un danneggiamento irrecuperabile.

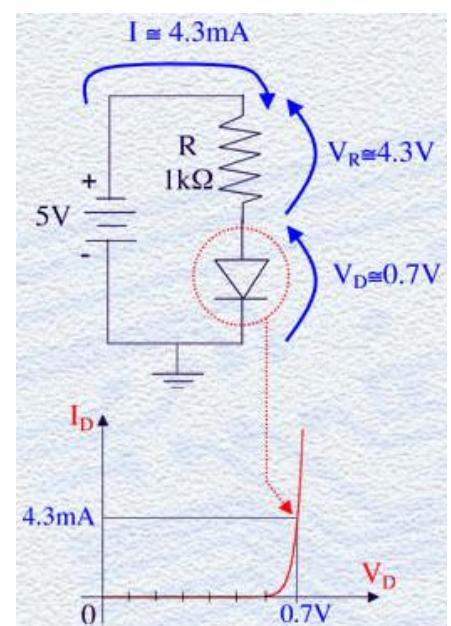
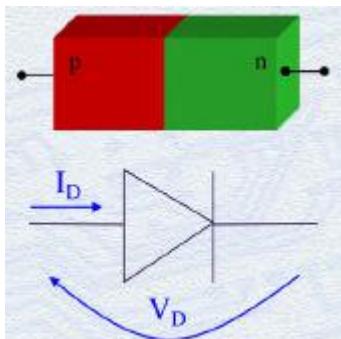
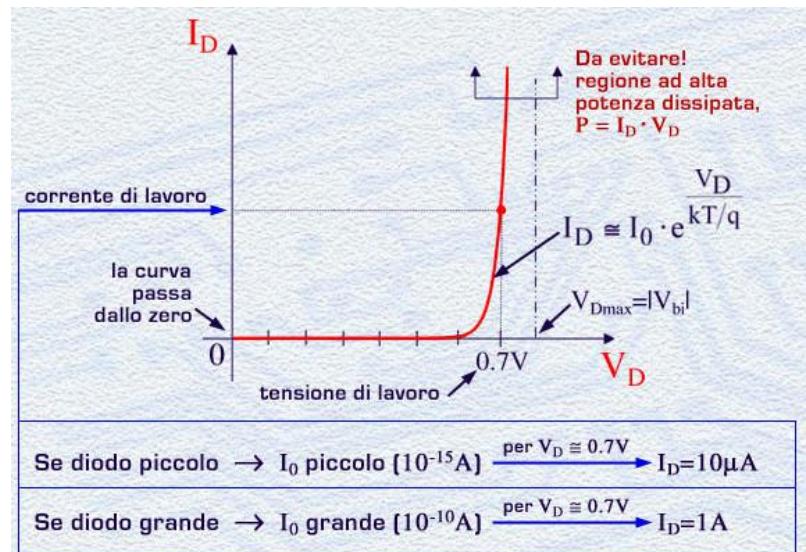
I diodi sono costruiti avendo dimensioni diverse a seconda delle applicazioni in modo che tutti portino la corrente per cui sono stati previsti con una tensione ai capi sempre:  $V_D \approx 0.7V$ .

Quando il diodo è usato in un circuito di un apparecchio portatile, che quindi ha piccole correnti circolanti in modo da dissipare complessivamente poca potenza per far durare le batterie a lungo, esso sarà di piccolissime dimensioni e la piccola corrente  $I_D$  sarà ottenuta con una tensione di circa 0.7V. Nella situazione opposta all'altro estremo, quando invece il diodo è usato in un circuito di grande potenza, come ad esempio nelle motrici dei treni, esso sarà di grandi dimensioni in modo da portare la elevata corrente richiesta sempre con una tensione ai suoi capi di circa 0.7V, oltre che da disperdere il calore generato dalla alta potenza elettrica dissipata.

Nella figura è riportato il simbolo circuitale del diodo: il triangolo rappresenta la zona *p* e la barra la zona *n*. La corrente diretta circola con il verso della punta del triangolo quando la tensione è applicata direttamente come nella figura.

Il fatto che la relazione tra  $I$  e  $V$  sia esponenziale e quindi che la corrente varia tantissimo a fronte di piccolissime variazioni di tensione, spesso giustifica la visione semplificata del diodo come di un dispositivo che fino a quando la sua tensione non ha raggiunto il valore di 0.7V non porta quasi corrente e solo attorno a 0.7V porta tutta la corrente necessaria.

Questa considerazione è in pratica sfruttata quasi sempre quando si deve trovare la corrente circolante in un circuito contenente anche diodi. Si consideri ad esempio il semplicissimo circuito costituito da una resistenza con in serie un diodo, alimentati da una tensione fissa, di cui si voglia calcolare la corrente circolante nella maglia. Se abbiamo scelto il diodo della giusta dimensione, quando in esso circolerà corrente la tensione ai suoi capi sarà di circa 0.7V. Ne risulterà che la tensione ai capi della resistenza sarà di 4.3V, tale quindi da fare circolare una corrente di 4.3mA. Quando si realizzerà questo circuito, bisognerà effettivamente scegliere un diodo che porti correnti di quell'ordine di grandezza. I fabbricanti indicano questo valore con il termine di *corrente nominale* del diodo.



## Diodo in inversa

Ritorniamo ai principi di funzionamento del diodo e vediamo ora cosa succede quando si applica una tensione esterna con il segno tale da aumentare (e non diminuire come fatto prima) la tensione naturale presente ai capi della giunzione quando è all'equilibrio.

Per fare ciò, lasciando invariato il generatore di tensione esterno, bisognerà applicare una tensione  $V_D$  negativa ai capi del diodo. Tale tensione è detta "tensione inversa" applicata al diodo.

Aumentando la tensione interna, aumenta anche il campo elettrico di richiamo sulle lacune e sugli elettroni. Per sostenere questo campo elettrico maggiore, un maggior numero di atomi di drogante si ionizzano, allargando spazialmente la zona a cavallo della giunzione interessata al fenomeno.

L'equilibrio tra la tendenza dei portatori alla diffusione in una direzione per effetto della differenza di concentrazione e la tendenza alla deriva nella direzione opposta per effetto del campo elettrico di richiamo, viene rotto a favore di quest'ultimo.

Se ne deduce che la corrente elettrica deve sicuramente cambiare il verso di percorrenza.

A differenza di prima però, la rottura dell'equilibrio non provoca un aumento della corrente circolante. Infatti, tolte inizialmente le lacune e gli elettroni della zona subito contigua alla giunzione, il campo

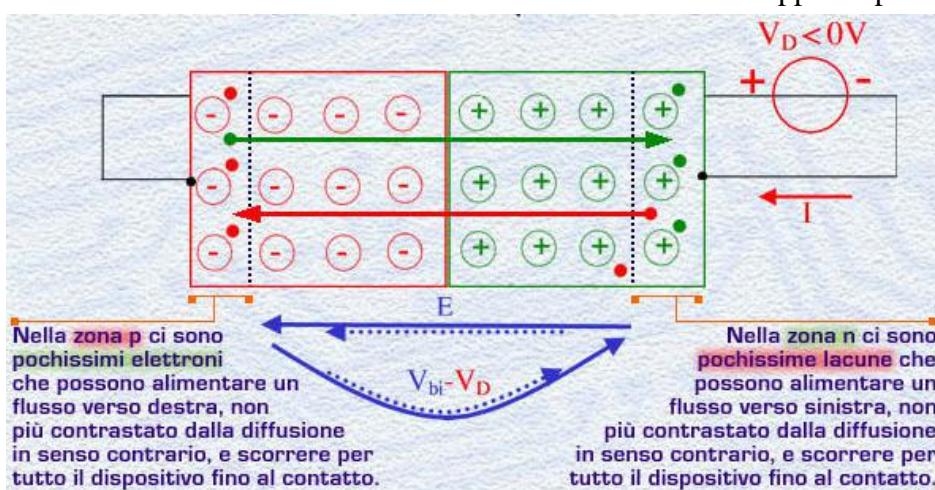
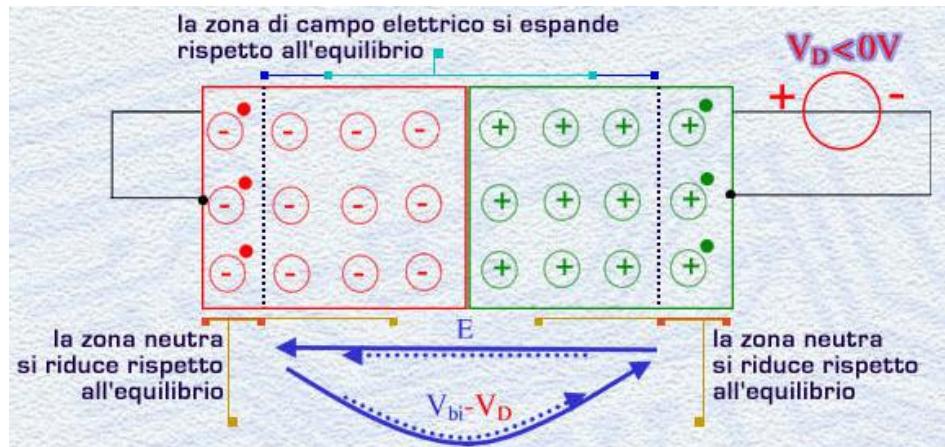
elettrico di richiamo, ora prevalente, non ha praticamente portatori da richiamare e mettere in moto con continuità. A destra, nella zona drogata n, infatti ci sono pochissime lacune minoritarie che possano sentire il campo elettrico e muoversi verso sinistra fino all'altro contatto; così pure a sinistra, nella zona drogata p, ci sono solo i pochissimi elettroni minoritari che possono essere messi in moto verso destra.

La corrente elettrica circolante quando è applicata una tensione esterna inversa è quindi piccolissima, ed è in pratica indipendente dal valore della tensione inversa applicata. Il suo valore può venire ricavato analizzando di nuovo in modo preciso i fenomeni che regolano il moto dei portatori minoritari lungo il dispositivo. Troveremmo così il valore:  $I_D @ -I_0$

dove  $I_0$  è lo stesso termine che compariva nella relazione trovata prima con la polarizzazione diretta. Questo è il valore di corrente che avremmo trovato utilizzando la formula esponenziale:

$$I_D = I_0 \left( e^{\frac{V_D}{kT/q}} - 1 \right) \cong -I_0$$

e calcolandola per  $V_D$  negativi. Il segno meno indica che il verso della corrente è opposto a quello trovato quando in diretta. Possiamo concludere che la stessa formula valga sia in diretta che in inversa.



elettrico di richiamo, ora prevalente, non ha praticamente portatori da richiamare e mettere in moto con continuità. A destra, nella zona drogata n, infatti ci sono pochissime lacune minoritarie che possano sentire il campo elettrico e muoversi verso sinistra fino all'altro contatto; così pure a sinistra, nella zona drogata p, ci sono solo i pochissimi elettroni minoritari che possono essere messi in moto verso destra.

La corrente elettrica circolante quando è applicata una tensione esterna inversa è quindi piccolissima, ed è in pratica indipendente dal valore della tensione inversa applicata. Il suo valore può venire ricavato analizzando di nuovo in modo preciso i fenomeni che regolano il moto dei portatori minoritari lungo il dispositivo. Troveremmo così il valore:  $I_D @ -I_0$

dove  $I_0$  è lo stesso termine che compariva nella relazione trovata prima con la polarizzazione diretta. Questo è il valore di corrente che avremmo trovato utilizzando la formula esponenziale:

$$I_D = I_0 \left( e^{\frac{V_D}{kT/q}} - 1 \right) \cong -I_0$$

e calcolandola per  $V_D$  negativi. Il segno meno indica che il verso della corrente è opposto a quello trovato quando in diretta. Possiamo concludere che la stessa formula valga sia in diretta che in inversa.

La curva caratteristica del diodo valida sia per tensioni dirette che per tensioni inverse è riportata nella figura. Per la corrente inversa, è stato indicato oltre al termine  $I_0$  prima trovato anche un termine aggiuntivo chiamato  $I_{GEN}$ , che spesso è presente nei diodi reali e che tiene conto delle lacune e degli elettroni generati per effetto termico nella zona di campo elettrico a cavallo della giunzione. Benché la corrente inversa rimanga piccolissima anche per ampie escursioni della tensione inversa applicata, tuttavia non bisogna mai superare il valore massimo di tensione inversa, indicato con  $V_{max}$ , (tensione di rottura o *breakdown*), in corrispondenza della quale il campo elettrico nella giunzione raggiunge un valore così elevato che si innescano scariche elettriche che portano al danneggiamento del dispositivo.

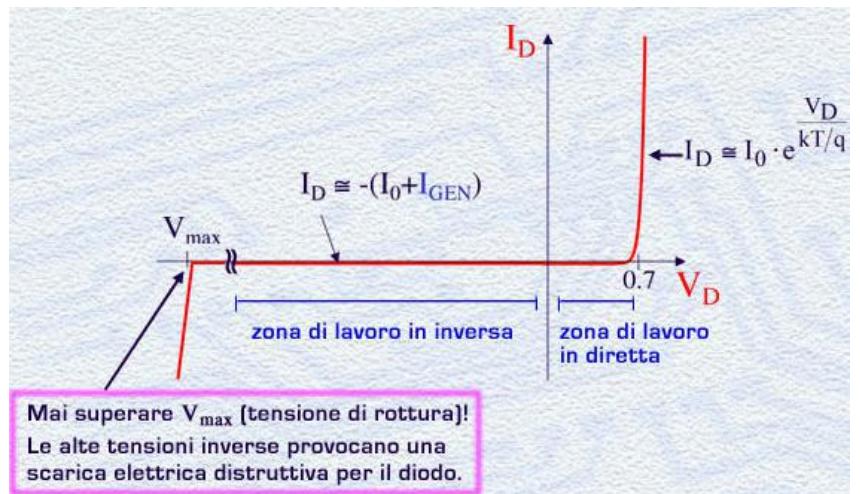
Se un diodo è inserito in un circuito e polarizzato in inversa, esso non permette il passaggio di corrente nella maglia. Si prenda ad esempio lo stesso semplicissimo circuito visto per la diretta e si giri il diodo. Il diodo sarà sicuramente polarizzato con la tensione inversa: infatti, pur non conoscendola ancora, la tensione ai suoi capi potrà assumere solo valori compresi tra 0V e 5V in inversa. Poiché in inversa, attraverso di esso non scorrerà corrente. Quindi la caduta di potenziale ai capi della resistenza sarà nulla. Pertanto, la tensione di 5V fornita dal generatore di tensione di alimentazione cadrà praticamente tutta ai capi del diodo, con il verso indicato dalla freccia nella figura che appunto mostra che è in inversa.

Se il circuito fosse stato alimentato con la tensione di 10V invece che di 5V la tensione ai capi del diodo sarebbe stata di circa 10V. Infatti, come detto commentando la curva caratteristica del diodo, la corrente è piccolissima per qualunque tensione inversa presente ai suoi capi.

L'unica precauzione pratica che bisognerà prendere quando si realizzerà il circuito sarà quella di scegliere un diodo che abbia una *tensione di rottura ben maggiore in modulo della tensione di lavoro* trovata.

## Esercizio 1

Calcolare la corrente circolante nella maglia del seguente circuito:



della quale il campo elettrico nella giunzione raggiunge un valore così elevato che si innescano scariche elettriche che portano al danneggiamento del dispositivo.

Se un diodo è inserito in un circuito e polarizzato in inversa, esso non permette il passaggio di corrente nella maglia. Si prenda ad esempio lo stesso semplicissimo circuito visto per la diretta e si giri il diodo. Il diodo sarà sicuramente polarizzato con la tensione inversa: infatti, pur non conoscendola ancora, la tensione ai suoi capi potrà assumere solo valori compresi tra 0V e 5V in inversa. Poiché in inversa, attraverso di esso non scorrerà corrente. Quindi la caduta di potenziale ai capi della resistenza sarà nulla. Pertanto, la tensione di 5V fornita dal generatore di tensione di alimentazione cadrà praticamente tutta ai capi del diodo, con il verso indicato dalla freccia nella figura che appunto mostra che è in inversa.

Se il circuito fosse stato alimentato con la tensione di 10V invece che di 5V la tensione ai capi del diodo sarebbe stata di circa 10V. Infatti, come detto commentando la curva caratteristica del diodo, la corrente è piccolissima per qualunque tensione inversa presente ai suoi capi.

L'unica precauzione pratica che bisognerà prendere quando si realizzerà il circuito sarà quella di scegliere un diodo che abbia una *tensione di rottura ben maggiore in modulo della tensione di lavoro* trovata.

## Lezione n. 13 : Circuiti con diodi

### L'alimentatore stabilizzato di tensione

Per familiarizzarci con l'uso del diodo nei circuiti elettronici, analizziamo il funzionamento di uno dei circuiti probabilmente più utilizzato al mondo e presente in ogni apparecchio alimentato dalla tensione di rete: l'alimentatore stabilizzato di tensione.

Esso ha il compito di convertire la tensione alternata di rete da 220V RMS nella tensione costante utilizzata nei circuiti elettronici. Questa tensione costante può variare da apparecchio ad apparecchio in base alla sua funzione ed in genere si colloca tra i 3V dei più moderni circuiti digitali ed i 25V di alcuni circuiti analogici.

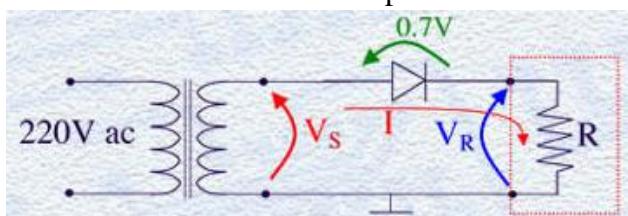
Un buon alimentatore deve poter fornire la tensione ad un valore il più possibile costante nel tempo, indipendentemente dalla quantità di corrente prelevata dal circuito utilizzatore e dalle fluttuazioni della tensione di rete. Per questo è detto "stabilizzato".

Quando si parla di tensione di rete da 220V si intende il valore RMS dell'onda sinusoidale oscillante alla frequenza di 50Hz e quindi con periodo di 20ms. L'ampiezza di picco dell'oscillazione è invece di  $220 \cdot \sqrt{2} = 311\text{V}$ .

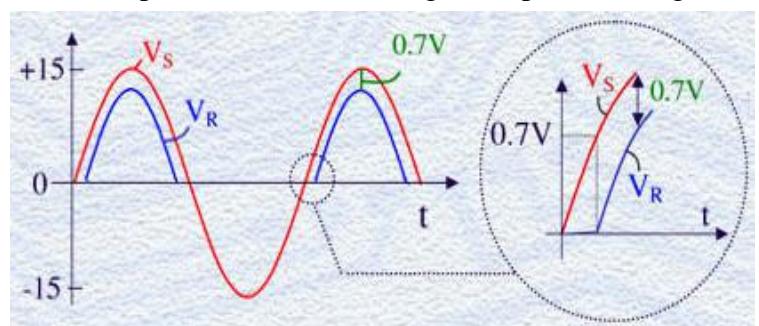
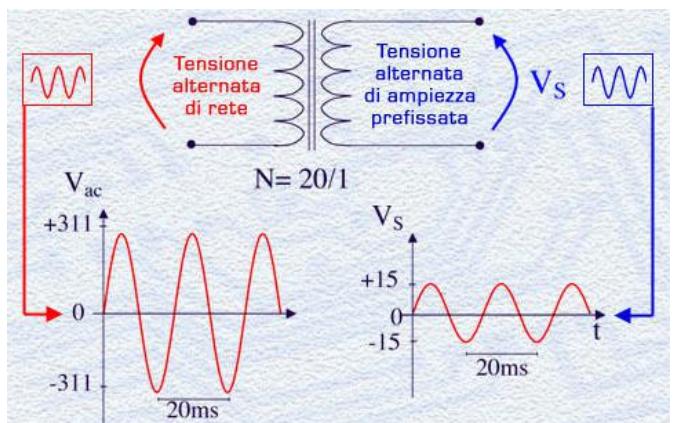
La diminuzione dell'ampiezza della sinusoide dal valore di rete di 311V di picco ad un valore prossimo a quello costante che si vorrà ottenere, viene realizzata con un trasformatore. Esso riduce l'ampiezza di una sinusoide in base al rapporto N tra le spire del primario e le spire del secondario. Pertanto ai capi del secondario abbiamo ancora una tensione sinusoidale,  $V_S$ , ma di ampiezza ridotta.

Se ad esempio si sceglie un rapporto di spire di 20/1, la sinusoide del secondario ha una ampiezza di picco di circa 15V.

La prima operazione da compiere sulla tensione alternata  $V_S$ , ora a disposizione sul secondario, è di trasformarla in una tensione avente sempre lo stesso segno. Per fare questo utilizziamo proprio la proprietà del diodo di condurre corrente solo quando è polarizzato direttamente e di non condurre quando in inversa. Un circuito che può compiere questa operazione è detto circuito RETTIFICATORE. Un esempio è mostrato nella

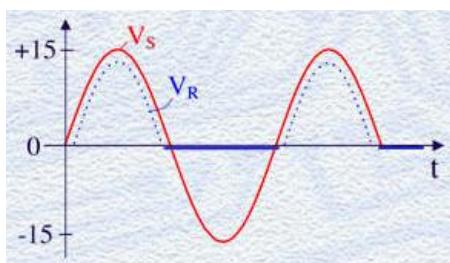


semionda di  $V_S$  è trasmessa ai capi di  $R$  ridotta in ampiezza di 0.7V. Se si guarda più in dettaglio il fenomeno, si vede che quando l'onda  $V_S$  è vicina allo 0V, precisamente tra 0V e 0.7V,  $V_R$  rimane comunque quasi a 0V perché il diodo non è ancora riuscito a polarizzarsi direttamente. Solo quando  $V_S$  raggiunge 0.7V, il diodo inizia ad essere correttamente polarizzato in diretta ed a mantenere la tensione di 0.7V ai suoi capi, e quindi  $V_R$  a salire in tensione.



Durante la semionda negativa di  $V_S$  il diodo è sicuramente polarizzato in inversa e quindi attraverso di esso non potrà scorrere corrente. Pertanto la caduta di potenziale ai capi di  $R$  è nulla e conseguentemente il segnale di uscita rimane a zero.

Si noti che in questa fase, tutta la tensione  $V_S$



cade in inversa ai capi del diodo che deve quindi essere scelto con una **tensione di rottura maggiore di  $V_S$** , nel nostro caso 15V. Grazie al circuito rettificatore, si è ottenuto un segnale di uscita sempre positivo, benché ancora ben lungi dall'essere il segnale costante desiderato!

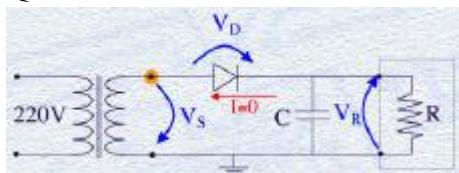
Come spesso si fa quando si vuole ridurre l'escursione di tensione di un punto di un circuito, la semionda in uscita può essere "filtrata" tramite l'aggiunta di una capacità  $C$  in parallelo al carico  $R$ .

Rivediamo il funzionamento del circuito, come prima, ma ora con l'aggiunta della capacità.

Durante la salita della semionda, in cui il diodo è in diretta, la corrente, oltre ad alimentare il circuito a valle, carica la capacità. La tensione ai capi della capacità e della resistenza segue quindi, come prima, la tensione  $V_S$ , ridotta di 0.7V.

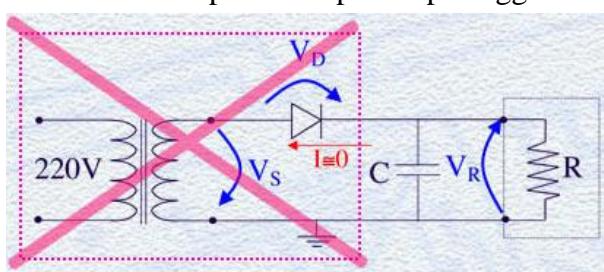
Quando  $V_S$  ha raggiunto il massimo, anche l'uscita ha raggiunto il massimo della tensione e quindi il condensatore ha accumulato il massimo della carica.

Quando la semionda inizia la ridiscesa, il diodo viene ad avere il capo a sinistra che diminuisce di



tensione mentre il capo a destra, dove è collegato il condensatore, che tende a mantenere la stessa tensione di prima grazie alla proprietà di una capacità di memorizzare una tensione ai suoi capi fintanto che non si riesce a variare la carica sui suoi piatti. Il diodo quindi si ritrova immediatamente una tensione inferiore a 0.7V e non permette più il passaggio di corrente attraverso di esso. Il circuito a sinistra del condensatore viene così ad essere disattivato e rimane solo la capacità  $C$  ad alimentare il carico  $R$ .

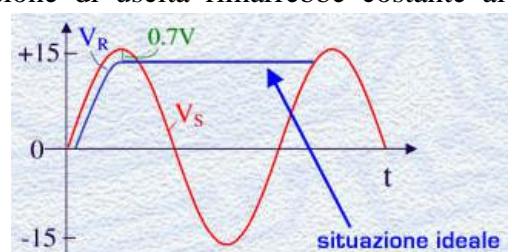
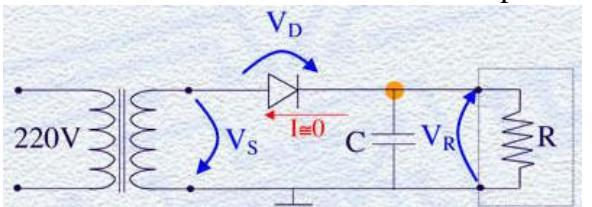
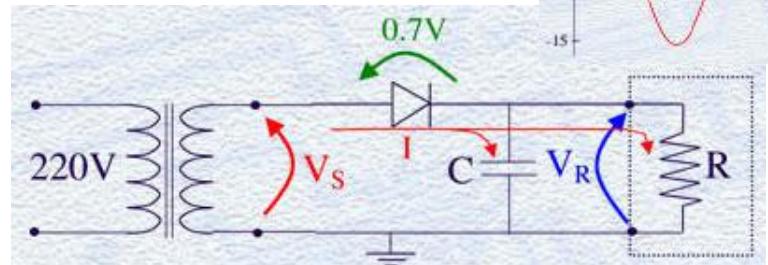
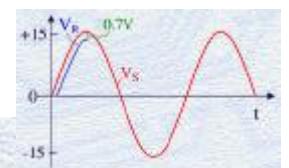
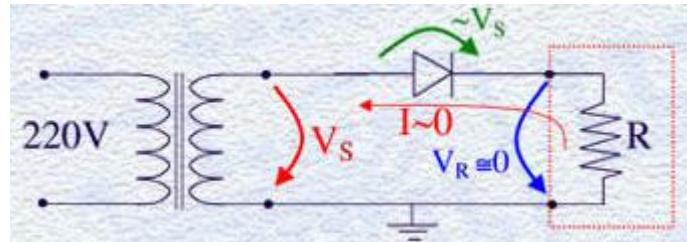
Se la carica del condensatore si mantenesse sui suoi piatti, la tensione di uscita rimarrebbe costante al valore di picco prima ottenuto ed avremmo



raggiunto l'obiettivo di produrre una tensione continua partendo da una tensione alternata.

Il diodo si manterebbe in inversa anche per tutta la semionda negativa. Si noti come l'aggiunta della capacità rimetta in discussione il criterio di scelta del diodo che ora dovrà poter sopportare una **tensione inversa ai suoi capi doppia di prima**, (circa 30V in questo caso).

In realtà, dato che l'alimentatore deve alimentare un circuito a valle che assorbirà corrente, da noi sintetizzato con la resistenza di carico  $R$ , la carica sul condensatore non rimane fissa ma diminuisce nel tempo.



Infatti la tensione ai capi del condensatore impone una corrente in R pari a  $V_R/R$  nello stesso verso di prima che scarica il condensatore stesso. La tensione ai capi del condensatore, e quindi ai capi del carico R, diminuisce pertanto nel tempo. Poiché il circuito è ora una semplice rete RC, la diminuzione avviene con la solita legge esponenziale (cfr figura a destra), dove  $V_{max}=14.3V$  nel nostro esempio. La scarica continua fino all'istante  $t_1$  in cui la successiva semionda ha un valore maggiore di  $V_R$  di 0.7V così da riaccendere il diodo e da

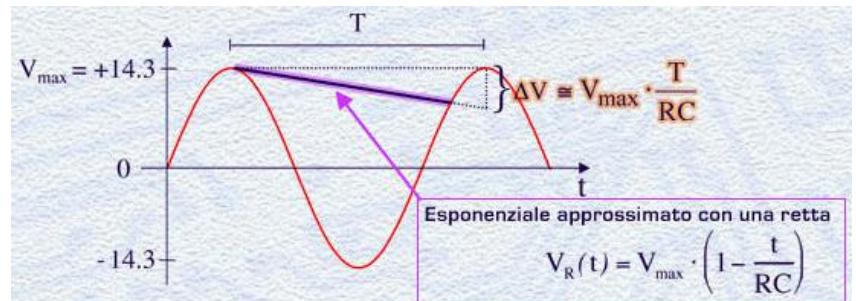
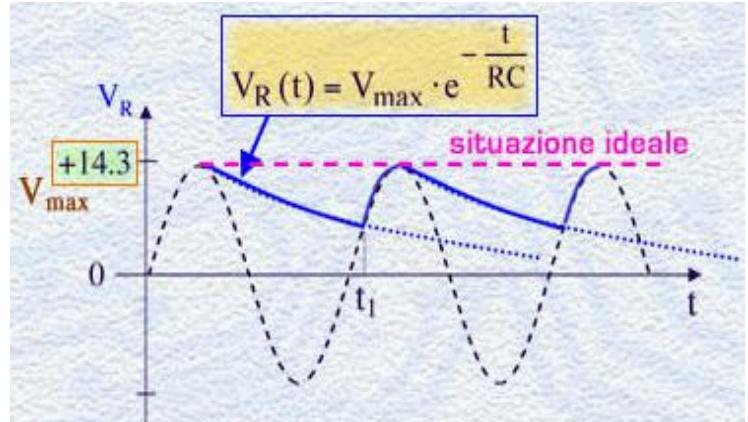
ricaricare la capacità con la sua corrente diretta fino al nuovo valore di picco della sinusoide.

L'andamento della tensione  $V_R$  che va ad alimentare il carico R è ancora purtroppo un po' diversa dall'andamento costante desiderato.

Per fare in modo che la diminuzione sia piccola, bisogna che la corrente assorbita dal carico sia piccola (cioè che il circuito a valle consumi poco o, nella nostra schematizzazione, che R sia grande) e che la quantità di carica immagazzinata nel condensatore sia grande (cioè che C sia grande).

I fenomeni descritti producono una ondulazione residua, indicata nella figura con il termine .V, della tensione di alimentazione di uscita, e spesso citata con il termine inglese di *ripple*. Per calcolarne il valore supponiamo di esserci posti comunque nella condizione di minimizzarne il valore avendo scelto un condensatore sufficientemente grande per cui la sua costante di tempo di scarica  $\tau = RC$  sia molto maggiore dei 20ms del periodo T dell'onda sinusoidale. In questo caso la discesa esponenziale può essere approssimata da una retta, (cfr figura a destra).

Una piccola sovrastima dell'ondulazione residua la si può immediatamente avere sulla base del triangolo indicato nella figura, come. Se ad esempio il carico è  $R=10k\Omega$  e la capacità  $C=100\mu F$  e  $V_{max}=14.3V$ , si otterebbe una ondulazione residua di circa 0.3V, pari al 2% del valore medio della tensione raddrizzata.



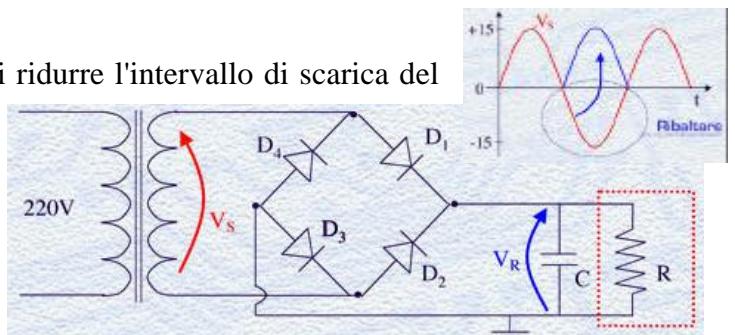
$$R=10k\Omega, C=100\mu F \text{ e } V_{max}=14.3V \rightarrow \Delta V \approx 0.3V (2\%)$$

Una tale ondulazione residua si trasmette in una pari instabilità delle tensioni di alimentazione dei vari circuiti elettronici ad esso collegati, e potenzialmente può trasferirsi anche nei segnali da essi amplificati. Ci sono applicazioni in cui un tale effetto è assolutamente da evitarsi, come ad esempio tutte le applicazioni audio ad altissima qualità.

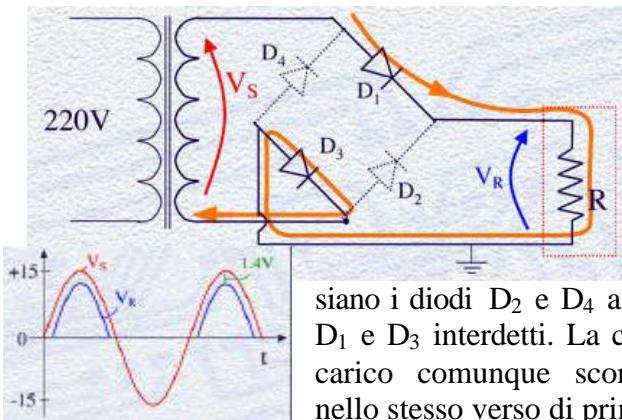
Se si volesse ridurre il ripple al valore dello **0.1%** del valore  $V_{max}$ , si scoprirebbe che bisognerebbe usare un condensatore da **2000mF!** A questo valore molto elevato di capacità corrisponde un condensatore di dimensioni particolarmente grosse e costoso. Se vi capita di aprire un apparecchio ad alta fedeltà, i cui circuiti assorbono molta potenza dalla rete per poterla trasferire agli altoparlanti, i grossi cilindri a fianco del trasformatore sono proprio i condensatori di filtro della tensione di alimentazione.

## Raddrizzatore a ponte di diodi

Un modo per ridurre il **ripple** è anche quello di ridurre l'intervallo di scarica del condensatore, prima coincidente con il periodo  $T$  della sinusoide di rete. Il modo più diretto per fare ciò è quello di sfruttare anche la semionda negativa della sinusoide, ribaltandola. Un circuito che realizza ciò è quello riportato nella figura a lato e chiamato "raddrizzatore ad onda intera" o anche "**raddrizzatore a ponte di diodi**". Durante la semionda positiva, solo i diodi  $D_1$  e  $D_3$  sono in diretta mentre  $D_2$  e  $D_4$  sono in inversa. La corrente quindi scorre nel carico  $R$  nel verso della freccia.



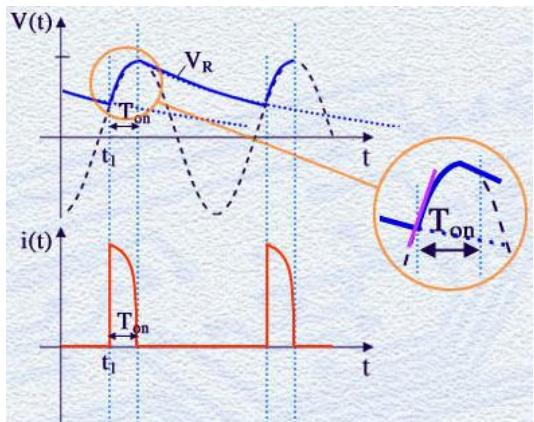
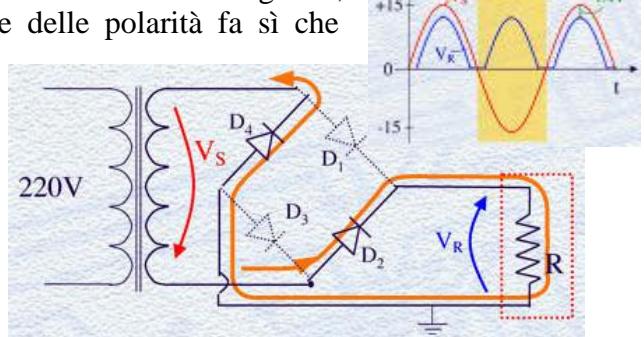
Siamo i diodi  $D_2$  e  $D_4$  a condurre e  $D_1$  e  $D_3$  interdetti. La corrente nel carico comunque scorre sempre nello stesso verso di prima e quindi alimenta il circuito in modo corretto. A differenza del caso a singola semionda, ora con il raddrizzatore a ponte di diodi si sfruttano tutte le semionde, sia quelle positive che quelle negative, per l'erogazione di potenza al carico.



Aggiungendo a valle del ponte il condensatore di filtro, si ottengono gli stessi effetti visti prima con il raddrizzatore a semionda, con la differenza che ora si è dimezzato il periodo in cui il condensatore da solo deve fornire potenza al carico, e pertanto si riduce di un fattore 2 il dimensionamento della capacità.

Meglio di così, avendo a disposizione all'ingresso la tensione di rete, non si può fare.

Dopo avere analizzato in dettaglio l'andamento della tensione, vediamo come varia nel tempo la corrente assorbita. Nei momenti in cui il segnale di tensione  $V_R$  è in salita ed il diodo è polarizzato in diretta, indicata con  $T_{on}$  nella figura, la capacità viene caricata esattamente della quantità di carica che essa ha dovuto fornire al circuito a valle per farlo funzionare. Questa carica arriva sul condensatore passando attraverso il diodo raddrizzatore in forma di corrente. Questa corrente è tutt'altro che costante nell'intervallo  $T_{on}$ . Ricordando che la corrente in un condensatore è legata alla derivata della tensione nel tempo secondo la relazione:  $i(t_0) = C \frac{dv(t)}{dt} \Big|_{t=t_0}$  se ne deduce che nel



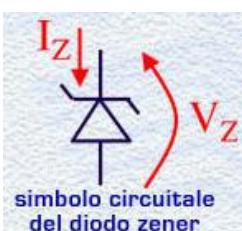
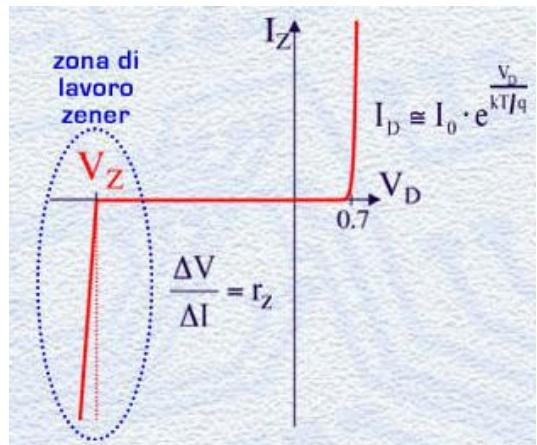
momento  $t_1$  di attacco delle risalita della tensione (in cui la variazione è massima) si avrà la corrente massima. Essa è minore negli altri istanti dell'intervallo  $T_{on}$  perché la derivata diminuisce.

La corrente del picco può essere molto alta anche dell'ordine di qualche A ed il diodo dovrà essere scelto in modo che possa portarla senza distruggersi. Si noti come questo **picco** sia **tanto maggiore quanto maggiore** è la **capacità di filtro** scelta: minore è il **ripple** che si vuole ottenere e maggiore è il picco di corrente che si deve gestire.

## Il diodo Zener

Per ridurre le ondulazioni residue, non potendo aumentare ulteriormente C per i motivi visti, si realizza una cella di stabilizzazione utilizzante un nuovo dispositivo elettronico chiamato "**diodo zener**".

Il diodo zener è un diodo in cui la tensione inversa limite, che ora indichiamo con  $V_Z$ , è ad un ben preciso valore ed è realizzato in modo che pur operando in quella regione di funzionamento esso non si danneggia. La sua curva caratteristica è del tutto simile a quella di un diodo normale, con la corrente inversa che aumenta rapidamente quando è raggiunto il valore di  $V_Z$ . Esso viene **usato solo in polarizzazione inversa** proprio sfruttando la presenza di  $V_Z$ : in questa regione per qualunque valore della corrente circolante, la tensione ai suoi capi è praticamente stabile a  $V_Z$ . In verità la curva caratteristica dello zener non è perfettamente verticale in corrispondenza di  $V_Z$  ma ha una pendenza finita e pertanto quando cambia la corrente inversa circolante al suo interno, cambia anche un poco la tensione effettiva ai suoi capi. I costruttori degli zener forniscono la pendenza della curva in termini di resistenza di zener proprio per fare con precisione questo conto. Tipici valori di questo dato sono: **5W < r<sub>Z</sub> < 100W**.

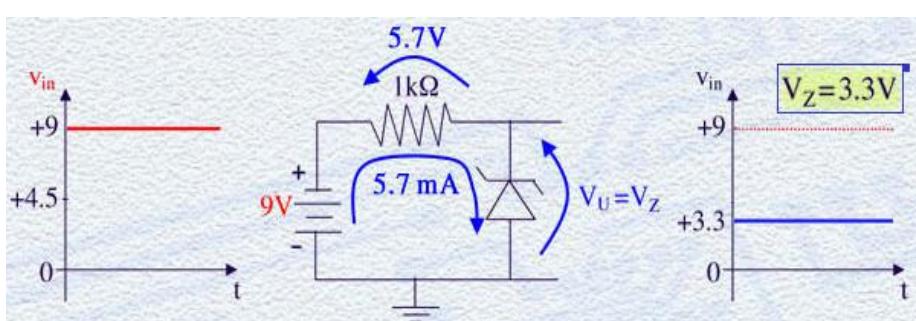


Il suo simbolo circuitale, simile a quello del diodo, riflette il suo uso particolare: quando passa la corrente inversa nel verso della freccia, il diodo sviluppa ai suoi capi la tensione  $V_Z$  nel verso della freccia.

Lo zener svolge quindi la funzione propria di un generatore di tensione e come tale è normalmente utilizzato.

Per prendere confidenza con il suo funzionamento, si pensi al semplice circuito fatto da un generatore di tensione, da una resistenza R e dal diodo zener collegato come in figura. Supponiamo di scegliere uno zener con tensione  $V_Z=3.3V$ . Se il generatore di tensione produce una tensione costante, ad esempio di +9V, la maglia è tale che sicuramente lo zener è polarizzato in inversa. Più in inversa di 3.3V però non può andare, perché se lo facesse passerebbe una

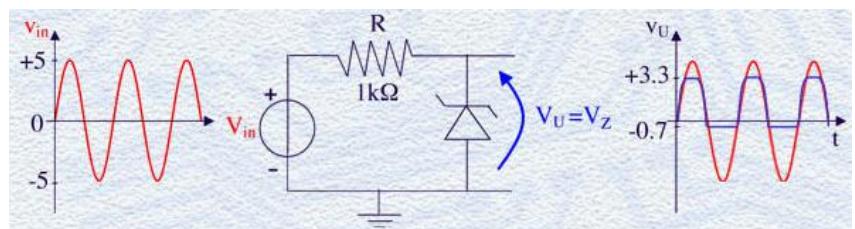
corrente enorme che scorrendo nella resistenza R ridurrebbe automaticamente la tensione ai suoi capi. Se il diodo zener tiene 3.3V in inversa ai suoi capi, allora ai capi del resistore ci sono 5.7V nel verso della freccia. Questo produce una corrente di 5.7mA che può passare in



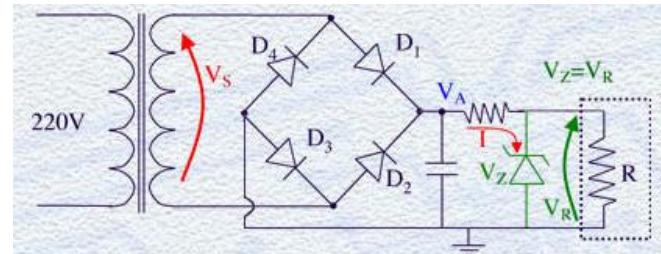
inversa attraverso lo zener e sostenere i 3.3V ai suoi capi inizialmente ipotizzati. Il risultato di questo circuito è che l'iniziale tensione a disposizione di 9V è stata ridotta al valore di 3.3V del diodo zener ed è disponibile per essere applicata ad un carico da collegarsi a valle.

In prima approssimazione, se la tensione di ingresso viene modificata, ad esempio passando dai 9V ai 10V, nulla cambia nella tensione di uscita che rimane a 3.3V.

Come altro esempio, si pensi di applicare un segnale di tensione sinusoidale all'ingresso di ampiezza di picco di 5V. Ci si convinca che l'uscita di tensione ai capi del diodo zener è la curva indicata in blu nel grafico, compresa tra +3.3V e -0.7V.



Lo zener può essere utilmente utilizzato in un alimentatore in continua proprio come elemento di stabilizzazione della tensione di uscita. Infatti aggiunto a valle del ponte di diodi di rettificazione e della capacità di filtraggio, esso costituisce una cella di stabilizzazione ulteriore della tensione  $V_R$  applicata al circuito di carico. Grazie al diodo zener infatti, le ondulazioni residue della tensione raddrizzata presente in  $V_A$



vengono ulteriormente ridotte, se non proprio completamente eliminate a causa della pendenza non infinita della curva caratteristica dello zener. Bisognerà ovviamente avere scelto uno zener con una tensione di zener pari alla tensione con cui vogliamo alimentare il circuito ed avere dimensionato la parte prima di circuito in modo che  $V_A$  sia sempre maggiore di  $V_Z$  in modo che lo zener sia sempre percorso da una corrente inversa che lo tenga acceso.

## Esercizio 1

Calcolare la corrente massima che circola nei diodi di un raddrizzatore a ponte in cui  $C=1000\mu F$ ,  $R=10k\Omega$  e  $V_{max}=14V$ . Sarete stupiti del suo alto valore!

Se la capacità di filtro fosse stata di  $C=10mF$  per ridurre ulteriormente il *ripple*, di quanto sarebbe aumentata  $I$  ?

# Lezione n. 14 : Introduzione ai circuiti digitali

## INTRODUZIONE

Con questa lezione ha inizio lo studio dei CIRCUITI ELETTRONICI DIGITALI, ovvero di quei circuiti che **elaborano i segnali digitali**, che sono segnali che contengono informazioni codificate in codice binario. Inizieremo analizzando i circuiti più semplici, ma anche fondamentali, che sono le porte logiche.

Prima di iniziare a studiare le porte logiche vediamo brevemente come è costituito in generale un Sistema Elettronico Digitale, di cui le porte logiche sono dei componenti particolari.

## I SISTEMI ELETTRONICI DIGITALI

Un sistema elettronico digitale che riceva in ingresso dei dati, li elabori e fornisca in uscita altri dati è costituito, generalmente, da una grande varietà di circuiti, semplici o complessi, che interagiscono tra di loro più o meno direttamente.

Tra questi circuiti se ne possono individuare alcuni che svolgono le seguenti funzioni elementari: Operazioni Logiche Booleane, come l'operazione di negazione (NOT), di prodotto logico



(AND), di somma logica (OR) di due o più variabili. Le operazioni aritmetiche, come somme, sottrazioni, elaborazioni particolari di segnali come la codifica, la decodifica, il multiplexing, la memorizzazione di dati; inizieremo a studiare i circuiti che svolgono le operazioni logiche su variabili booleane.

## OPERAZIONI LOGICHE BOOLEANE

Le operazioni logiche vengono svolte su variabili, dette booleane, che possono assumere, in un dato istante, uno tra due soli valori possibili. I due valori possibili vengono indicati, in modo equivalente con: Vero o Falso oppure con i numeri 1 o 0 oppure con Alto o Basso. Le variabili Booleane possono essere rappresentabili da grandezze elettriche quali una tensione o una corrente perché, come vedremo, è semplice far assumere ad una tensione in un nodo di un circuito un valore



basso (ad esempio 0 volt) oppure un valore alto (ad esempio +3V).

I circuiti che operano su grandezze elettriche, rappresentanti variabili booleane, svolgendo operazioni logiche, si chiamano **PORTE LOGICHE** o **Logic Gates**.

## LE PORTE LOGICHE

Le porte logiche oltre a svolgere operazioni booleane costituiscono anche i **blocchi elementari** con cui si sintetizzano funzioni digitali più complesse, quali la codifica, la decodifica, il multiplexing ... solo per fare alcuni esempi. Le porte logiche vanno anche considerate come **circuiti fondamentali** perché l'analisi di molti circuiti complessi è riconducibile all'analisi di porte logiche; ad esempio la determinazione dei tempi d'accesso di memorie ROM o di RAM statiche (SRAM) è riconducibile alla determinazione di tempi caratteristici di porte logiche.

## LE FAMIGLIE LOGICHE

Storicamente le porte logiche sono state realizzate utilizzando due tecnologie principali:

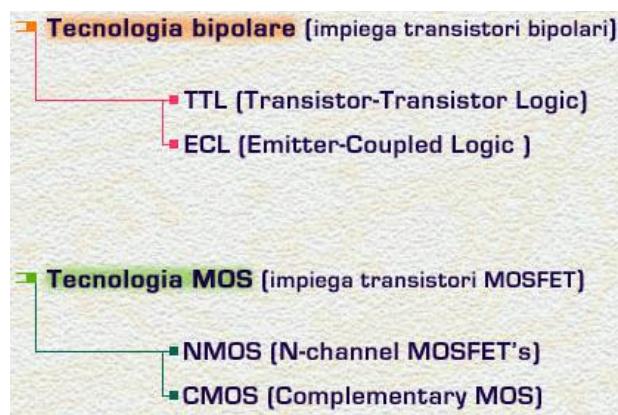
- la **tecnologia bipolare**, che impiega particolari transistori detti bipolari, non inclusi nel programma di questo corso;
- la **tecnologia MOS**, che impiega i transistori MOSFET.

In ciascuna di queste tecnologie le porte logiche vengono realizzate impiegando differenti topologie circuitali. Porte logiche appartenenti ad una particolare tecnologia e tipologia costituiscono una **FAMIGLIA LOGICA**. Ogni Famiglia Logica contiene tutte le porte necessarie a svolgere le operazioni logiche.

Esiste la Famiglia logica TTL e la Famiglia logica ECL per quanto riguarda la tecnologia bipolare.

In tecnologia MOS le principali famiglie logiche sono la N-MOS, nella quale si impiegano come transistor, solo MOSFET a canale N, e la famiglia logica CMOS, nella quale si impiegano sia transistor MOSFET a canale N che a canale P. In

questo corso studieremo solo le porte logiche della famiglia CMOS, perché la maggior parte dei sistemi digitali dei nostri giorni sono realizzati con tecnologia CMOS.



### COME SI ANALIZZA UNA PORTA LOGICA

Prima di studiare com'è fatta una porta logica da un punto di vista circuitale vediamo come in generale si analizza una porta logica.

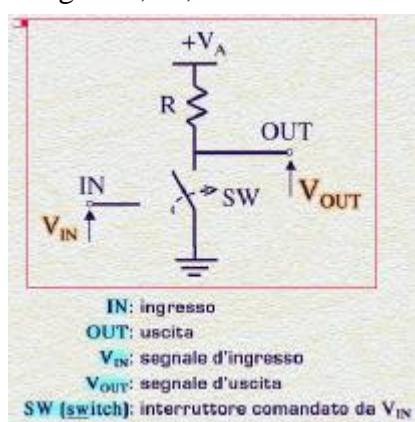
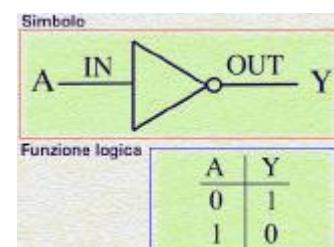
Una porta logica si analizza determinandone:

- La **funzione logica**,  $Y = f(A, B, C \dots)$ , ovvero la relazione booleana che lega la variabile d'uscita Y alle variabili d'ingresso A, B, C ...
- La **caratteristica ingresso – uscita**,  $V_{OUT} = g(V_{IN})$ , ovvero la funzione che lega la variabile elettrica tensione d'uscita,  $V_{OUT}$ , alla tensione d'ingresso,  $V_{IN}$ .
- L'analisi della caratteristica ingresso – uscita permetterà di determinare alcuni parametri importanti della porta che sono:
  - i **livelli logici** ovvero i valori di tensione corrispondenti al bit 1 e al bit 0,
  - la **soglia di commutazione**
  - i **margini di rumore** che studieremo successivamente.
- Inoltre di una porta logica occorre conoscere il **consumo di potenza elettrica**.
- Infine è importante conoscere anche la **velocità di risposta** attraverso parametri quali:
  - il **tempo di propagazione  $t_P$** ,
  - la **massima frequenza di operazione**,
  - il massimo tasso di ripetizione di bit che la porta può elaborare o detto **BIT RATE massimo**.

### L'INVERTITORE LOGICO

Iniziamo a studiare la più semplice porta logica: l'invertitore logico o NOT GATE.

In figura a lato vediamo il simbolo e la funzione logica: quando la variabile d'ingresso, A, assume il valore logico basso (0) la variabile d'uscita, Y, assume il valore logico alto (1); analogamente quando A è 1, Y assume il valore 0.



### CIRCUITO DI UN INVERTITORE

Vediamo una semplice implementazione circuitale di un invertitore logico. Il circuito è costituito da un resistore R e da un interruttore elettronico indicato con SW (come 'SWITCH'). Il resistore è collegato da una parte alla tensione di alimentazione  $V_A$ , dall'altra al morsetto d'uscita OUT.  $V_{IN}$  e  $V_{OUT}$  sono le tensioni d'ingresso e d'uscita rispettivamente, riferite al potenziale di massa. L'interruttore, collegato da una parte a massa, è comandato in apertura e chiusura dalla tensione d'ingresso  $V_{IN}$ . Vediamo il principio di funzionamento e le caratteristiche di questo circuito.

## PRINCIPIO DI FUNZIONAMENTO

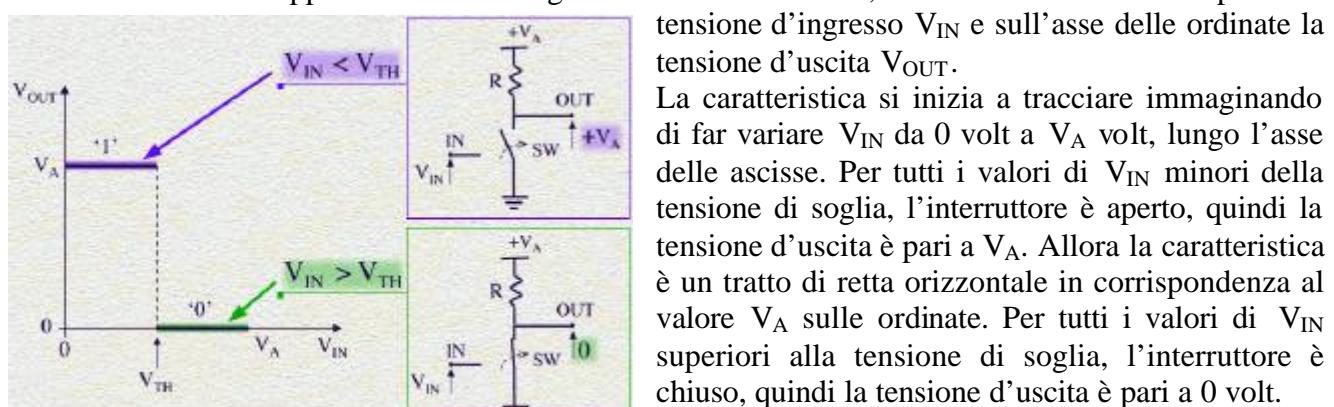
Indichiamo con  $V_{TH}$  un particolare valore di tensione chiamato “TENSIONE DI SOGLIA”.  $V_{TH}$  è un valore compreso tra 0 e la tensione di alimentazione  $V_A$ . Se la tensione d’ingresso  $V_{IN}$  è minore della tensione di soglia allora l’interruttore rimane in posizione “APERTO”. Nel resistore R non passa corrente, quindi la tensione d’uscita è uguale alla tensione di alimentazione  $V_A$ .

Se la tensione d’ingresso è maggiore di  $V_{TH}$  allora l’interruttore si porta in posizione “CHIUSO”. Il nodo d’uscita risulta quindi cortocircuitato a massa e la tensione d’uscita è perciò uguale a 0. Vediamo quindi che questo circuito svolge la funzione di inversione logica del bit in ingresso. Infatti se la tensione d’ingresso è 0 volt (corrispondente a bit 0) la tensione d’uscita è  $V_A$  volt, corrispondente al bit 1. Viceversa, quando la tensione d’ingresso è  $V_A$  (bit 1) la tensione d’uscita è 0. I livelli logici sono quindi 0 per il bit 0 e  $V_A$  per il bit 1.

## LA CARATTERISTICA INGRESSO – USCITA

Vediamo ora come ricavare la caratteristica ingresso – uscita dell’invertitore.

La caratteristica si rappresenta in un diagramma cartesiano dove, sull’asse delle ascisse è posta la tensione d’ingresso  $V_{IN}$  e sull’asse delle ordinate la tensione d’uscita  $V_{OUT}$ .



La caratteristica si inizia a tracciare immaginando di far variare  $V_{IN}$  da 0 volt a  $V_A$  volt, lungo l’asse delle ascisse. Per tutti i valori di  $V_{IN}$  minori della tensione di soglia, l’interruttore è aperto, quindi la tensione d’uscita è pari a  $V_A$ . Allora la caratteristica è un tratto di retta orizzontale in corrispondenza al valore  $V_A$  sulle ordinate. Per tutti i valori di  $V_{IN}$  superiori alla tensione di soglia, l’interruttore è chiuso, quindi la tensione d’uscita è pari a 0 volt.

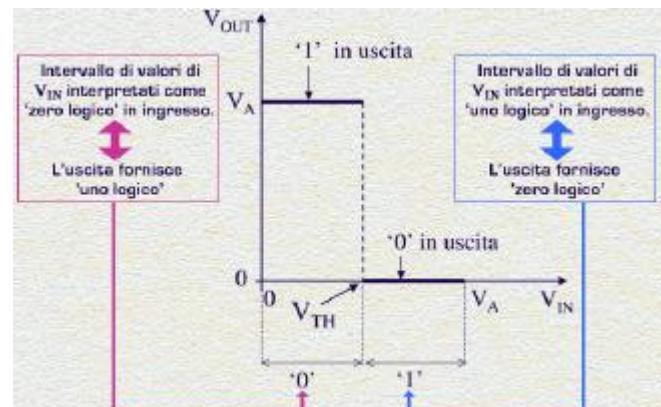
In questo caso la caratteristica è un tratto di retta orizzontale sovrapposto all’asse delle ascisse. In corrispondenza del valore  $V_{IN} = V_{TH}$  si ha perciò una discontinuità nella caratteristica ingresso – uscita.

## ANALISI DELLA CARATTERISTICA IN-OUT

Discutiamo ora più in dettaglio la caratteristica ingresso – uscita dell’invertitore.

Concentriamo l’attenzione sulla tensione d’ingresso: osserviamo che non solo il valore  $V_{IN}$  uguale a zero volt rappresenta il bit 0, ma tutti i valori compresi nell’intervallo tra 0 volt e la tensione di soglia sono “interpretati” dal circuito come 0 logico in ingresso. Infatti il circuito fornisce in uscita un uno logico, indipendentemente dal particolare valore di  $V_{IN}$  interno all’intervallo indicato. Un’analoga osservazione può essere fatta per l’intervallo di  $V_{IN}$  compresi tra la tensione di soglia e  $V_A$ , che vengono tutti interpretati come ‘1’ logico in ingresso.

Queste osservazioni ci permettono ora di fare una considerazione di carattere generale riguardo all’elaborazione di segnali digitali con porte logiche.



## ELABORAZIONE DI UN SEGNALE CON UN INVERTITORE IDEALE

In figura è rappresentato un segnale digitale costituito da una sequenza di bit distribuiti lungo l'asse dei tempi. Nel caso ideale i bit sono rappresentati da due soli livelli di tensione: 0 volt e  $V_A$  volt. Un invertitore che riceva in ingresso questo segnale fornirà in uscita la sequenza complementare rappresentata in figura, con livelli logici ancora pari a 0 e  $V_A$  volt.

Nella realtà un segnale come quello mostrato può venire alterato da disturbi generati da campi elettro – magnetici che interessano il mezzo in cui si propaga il segnale. Tali disturbi vengono solitamente indicati con il termine di rumore. Nella figura in alto a destra vediamo una rappresentazione semplificata del fenomeno: lo stesso segnale di sinistra ma con una alterazione significativa dei livelli di tensione corrispondenti ai bit '0' e '1'. Tuttavia la caratteristica a soglia dell'invertitore permette, in molti casi, una corretta interpretazione della sequenza anche in presenza di rumore significativamente alto. Infatti, per l'invertitore, è sufficiente che la tensione in ingresso sia minore della tensione di soglia affinché il bit venga letto come 0 oppure che la tensione d'ingresso sia superiore alla soglia affinché il bit venga letto come 1. Tale condizione è verificata per tutti i bit della sequenza rumorosa, quindi l'invertitore fornirà in uscita la sequenza corretta.

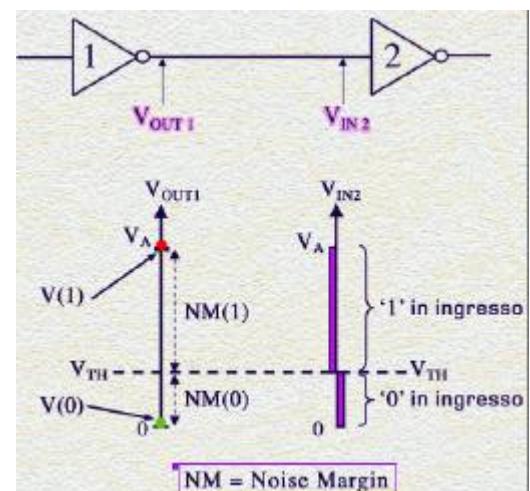
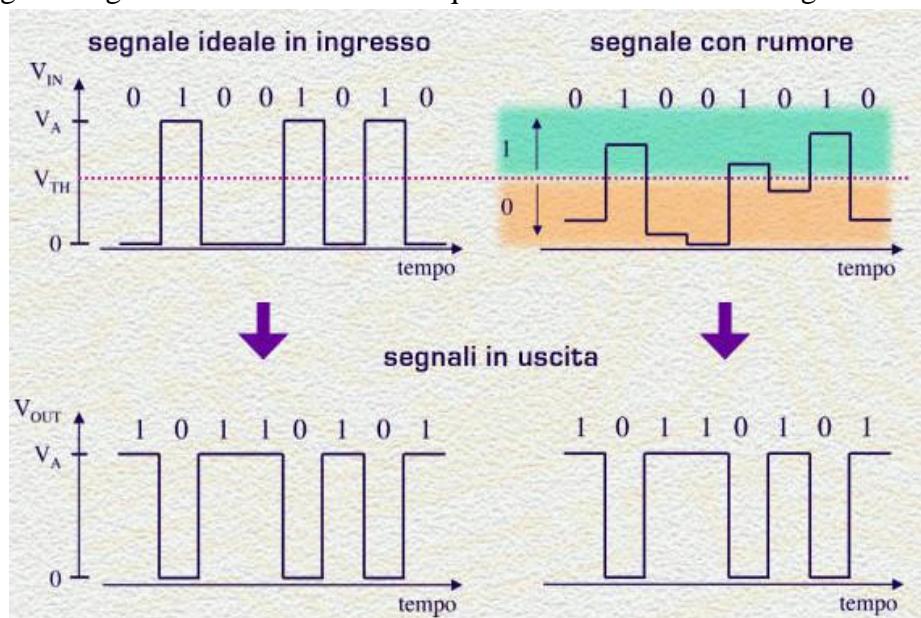
Notiamo inoltre che l'invertitore non solo ha interpretato correttamente i bit in ingresso ma, alla sua uscita, ha ripristinato i livelli logici a 0 volt e  $V_A$  volt. Osserviamo quindi una caratteristica fondamentale dei sistemi digitali che è ***'l'alto grado di immunità ai disturbi.***

Infine occorre osservare che esiste un limite superiore all'ampiezza del rumore affinché l'informazione binaria non venga modificata con l'introduzione di errori nella sequenza. Infatti, la lettura corretta si ha solo nel caso in cui l'ampiezza del rumore sia tale che il livello alto (bit 1) rimanga comunque sopra la tensione di soglia ed il livello basso (bit 0) sotto la tensione di soglia. Questa condizione si traduce nel concetto di **MARGINE DI RUMORE** che ora introdurremo.

### I MARGINI DI RUMORE

Consideriamo due invertitori identici in cascata: l'uscita dell'invertitore 1 è collegata all'ingresso dell'invertitore 2. Indichiamo con  $V_{OUT1}$  la tensione d'uscita dell'invertitore 1 e con  $V_{IN2}$  la tensione d'ingresso dell'invertitore 2. Ovviamente le due tensioni coincidono in valore ma è opportuno indicarle con due nomi distinti. Rappresentiamo su un asse verticale i valori che  $V_{OUT1}$  può assumere e su un altro asse parallelo i valori che  $V_{IN2}$  può assumere. I due invertitori sono identici, quindi hanno la **stessa tensione di soglia**.

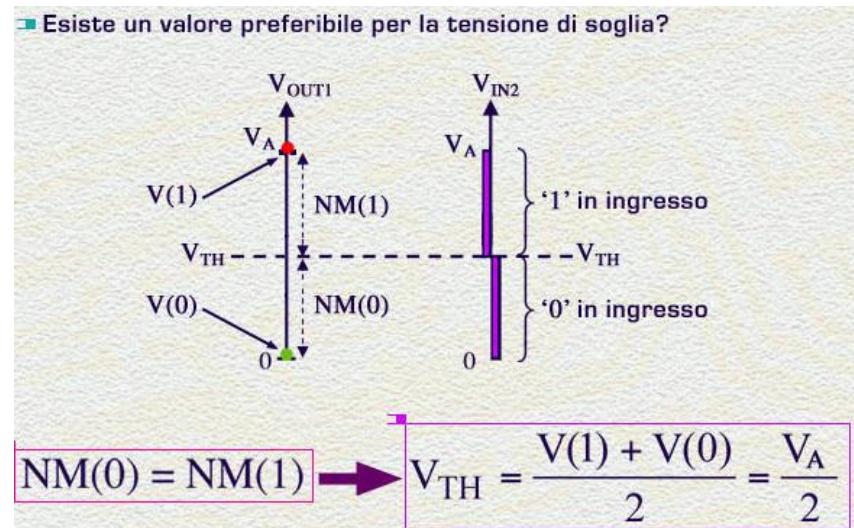
Sull'asse di  $V_{IN2}$  indichiamo la fascia di valori sotto la tensione di soglia che sono interpretati come bit 0 in ingresso. Analogamente indichiamo la fascia di valori sopra la soglia che sono interpretati come bit 1 in ingresso. Supponiamo ora che la porta 1 abbia in uscita 0 volt, valore nominale del livello logico basso. Supponiamo che in uscita alla porta 1 si presenti un disturbo che modifica il livello di  $V_{OUT1}$ .



Affinché la porta 2 non risenta di tale variazione è necessario che  $V_{OUT1}$  non superi la tensione di soglia ovvero che il disturbo sia di ampiezza inferiore alla quantità indicata in figura come **NM(0)**, che sta per **NOISE MARGIN ZERO** ovvero margine di rumore associato al livello 0. Analogamente vale per il livello logico alto. In tale caso si parla di **MARGINE DI RUMORE** associato al livello 1, in figura indicato con **NM(1)**.

### SCELTA OTTIMALE DI $V_{TH}$

In base a quanto visto riguardo ai margini di rumore di un invertitore possiamo fare una importante osservazione riguardo alla tensione di soglia di un invertitore. Possiamo chiederci: supponendo di poter scegliere il valore della tensione di soglia, quale valore scegliere? Ovvero: esiste un valore preferibile per la tensione di soglia? La risposta è sì! Esiste un valore preferibile. Come abbiamo visto la tensione di soglia determina i valori dei due margini di rumore. Dato che le quantità di bit 0 e bit 1 elaborati da un sistema sono in media uguali e dato che i disturbi possono in uguale modo abbassare o alzare i livelli di tensione, è preferibile avere entrambi i margini di rumore più alti possibile e non è utile che un livello logico abbia un margine di rumore superiore a quello dell'altro. La condizione preferibile è quindi avere uguali i due margini di rumore.



Questa condizione si traduce nella scelta della **tensione di soglia che dev'essere posta a metà tra il livello logico alto  $V(1)$  e il livello logico basso  $V(0)$** .

Nel caso in esame in cui  $V(0)$  è 0 volt e  $V(1)$  è pari a  $V_A$ , la **soglia ottimale si trova a  $V_A/2$** .

## Lezione n. 15 : Porte logiche: generalità, 1

### INVERTITORE REALE: CARATTERISTICA INGRESSO – USCITA

Abbiamo visto che la caratteristica ingresso – uscita di un invertitore ideale presenta due soli valori per la tensione d'uscita ed una discontinuità in corrispondenza della tensione di soglia in ingresso. Nella realtà un circuito elettronico potrà solo approssimare tale caratteristica ideale. La caratteristica di un invertitore reale può essere rappresentata qualitativamente come mostrato in figura. Le differenze con la caratteristica ideale che, nel caso più generale possono presentarsi sono:

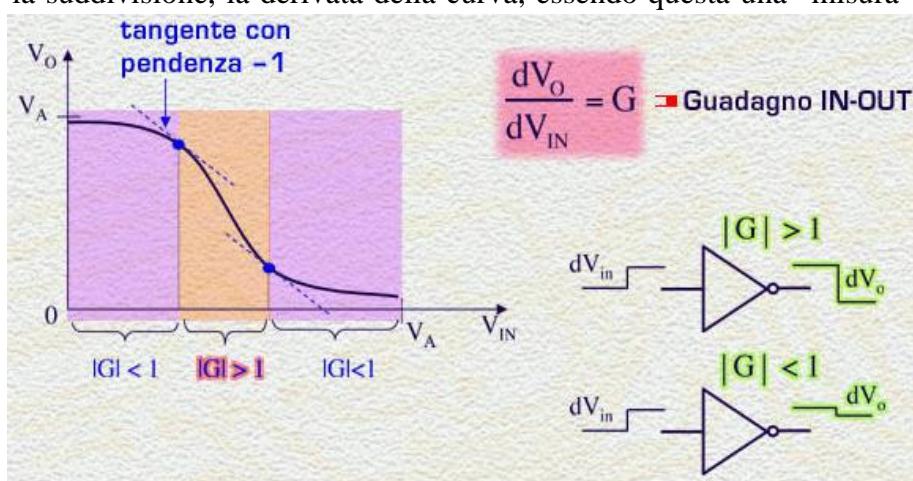
- Il livello logico alto è inferiore alla tensione di alimentazione  $V_A$
- Il livello logico basso è superiore a 0 volt
- La transizione tra i due livelli è graduale con pendenza finita
- I tratti di caratteristica in prossimità dei livelli alto e basso non sono orizzontali ma presentano una pendenza non nulla.

Questa caratteristica rappresenta il caso più generale nel senso che alcune delle differenze evidenziate potrebbero non sussistere. Ad esempio, vedremo che esistono invertitori reali con livello logico alto pari a  $V_A$  e livello basso pari a 0 volt. Comunque, tutti gli invertitori reali presentano una transizione graduale tra i due livelli logici e questa costituisce la differenza principale tra il caso reale ed il caso ideale. Questa differenza richiede alcune considerazioni.

### LA CARATTERISTICA DELL'INVERTITORE REALE

Guardando la caratteristica ingresso – uscita di un invertitore reale, possiamo identificare tre regioni in cui suddividerla in base alla pendenza; le due regioni a bassa pendenza in prossimità dei livelli logici basso ed alto e la regione di transizione tra i due livelli, caratterizzata da una elevata pendenza. Faremo ora una divisione precisa tra queste regioni, che ci sarà utile successivamente, prendendo come parametro, per effettuare la suddivisione, la derivata della curva, essendo questa una “misura” della pendenza. La derivata

della caratteristica viene anche chiamata **guadagno ingresso – uscita** del circuito su piccoli segnali. Infatti se consideriamo il circuito operante in un preciso punto della caratteristica, la derivata in quel punto ci dà informazioni su quanto una piccola variazione  $dV_{IN}$  della tensione d'ingresso, variazione che chiameremo



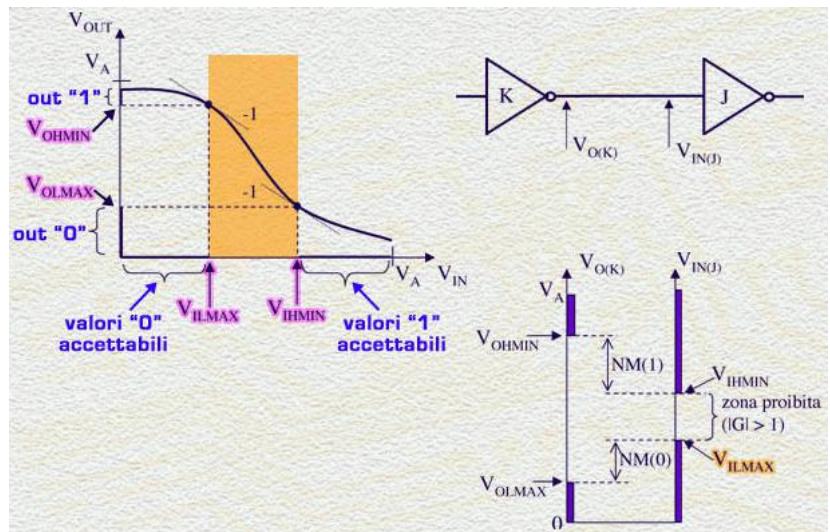
per ora segnale, si ripercuote come variazione  $dV_{OUT}$  in uscita. Se la derivata è, in modulo, maggiore di 1, la variazione della tensione d'uscita risulterà maggiore rispetto a quella in ingresso e parleremo di guadagno effettivo, cioè in modulo maggiore di 1, sul segnale. Se viceversa la derivata è, in modulo, minore di 1 si verificherà una attenuazione del segnale passando dall'ingresso all'uscita del circuito. La stessa cosa avviene qualora la variazione  $dV_{IN}$  fosse dovuta ad un disturbo o rumore. Quindi il parametro guadagno, largamente utilizzato per i circuiti analogici come ad esempio gli amplificatori, risulta utile anche nell'analisi di un circuito digitale come l'invertitore perché permette di dire come reagirà il circuito ad una variazione indesiderata, disturbo o rumore, della tensione d'ingresso. Ovvero, se tale disturbo verrà amplificato in uscita, se il guadagno è in modulo  $> 1$  o attenuato nel caso opposto. Risulta quindi immediato suddividere la caratteristica ingresso – uscita prendendo come punti di suddivisione quelli in cui il guadagno vale  $-1$ , ovvero quei due punti in cui

la retta tangente alla curva è parallela alla bisettrice del 2° e 4° quadrante. Ora, affinché non si verifichi una amplificazione dei disturbi eventualmente presenti all'ingresso dell'invertitore, è necessario che la porta **operi sempre con  $V_{IN}$  interno alle due regioni con modulo del guadagno minore di 1**. La regione di transizione in cui si ha  $|G| > 1$  viene **proibita** in quanto operare all'interno di questa regione potrebbe portare ad errori nell'elaborazione dei bit, a causa dell'amplificazione di eventuali disturbi.

### MARGINI DI RUMORE DELL'INVERTITORE REALE

L'aver identificato la regione proibita a guadagno in modulo  $> 1$  ci permette di definire i margini di rumore dell'invertitore reale. Infatti i due punti a tangente  $-1$  definiscono due valori per  $V_{IN}$  e due valori per  $V_{OUT}$ . Vediamoli:

$V_{ILMAX}$  rappresenta il massimo valore che è concesso alla tensione d'ingresso nello stato logico basso.  $V_{IHMIN}$  rappresenta il minimo valore concesso alla tensione di ingresso nello stato logico alto. Quindi questi due valori delimitano le regioni di tensioni accettabili in ingresso come bit '0' e bit '1' rispettivamente. Ora, se la tensione d'ingresso di un invertitore è interna alle due regioni ammesse, la sua tensione d'uscita sarà interna ai due intervalli mostrati in figura, in cui:



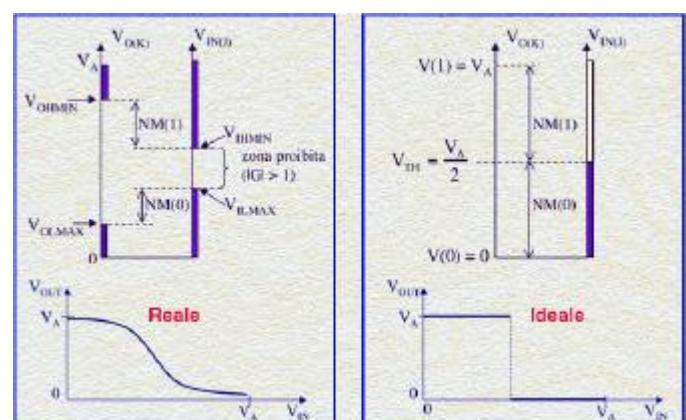
$V_{OLMAX}$  rappresenta il massimo valore in uscita per il livello logico basso e  $V_{OHMIN}$  rappresenta il minimo valore in uscita per il livello logico alto. Riprendendo quindi il concetto di margine di rumore visto per l'invertitore ideale, consideriamo due invertitori K e J in cascata. Nel grafico sotto gli invertitori, sui due assi, sono indicati i valori possibili per la tensione d'uscita di K e i valori accettabili per la tensione d'ingresso di J. In analogia con quanto visto per l'invertitore ideale i margini di rumore per lo stato basso e per lo stato alto sono quelli indicati in figura. Vediamo ad esempio il margine di rumore del livello logico basso  $NM(0)$ . Supponiamo quindi che la prima porta fornisca in uscita il bit '0' ovvero una tensione non superiore a  $V_{OLMAX}$ . Supponiamo ora che si presenti un disturbo che alza la tensione  $V_o(K)$ . Affinché la seconda porta operi nella regione consentita è necessario che  $V_{O(K)}$  non superi 1 valore  $V_{ILMAX}$ . La situazione più critica si verifica quando  $V_{O(K)}$  è pari a  $V_{OLMAX}$  perché in questo caso l'ampiezza massima consentita al rumore è la più bassa e appunto pari a  $NM(0)$ . Analoghe considerazioni valgono per il margine di rumore 1.

### CONFRONTO INVERTITORE REALE – IDEALE

I **margini di rumore devono essere entrambi massimizzati per ottenere una porta in grado di operare, senza errori, anche in presenza di rumore**. Vediamo allora che i margini di rumore assumono il loro valore massimo quando:

- La zona proibita si riduce in estensione fino ad annullarsi.
- $V_{OLMAX}$  si riduce a 0 volt e  $V_{OHMIN}$  sale a  $V_A$  volt
- La tensione di soglia si pone a  $V_A/2$ .

Osserviamo che a questa situazione corrisponde proprio la caratteristica ingresso – uscita dell'invertitore ideale. Comprendiamo quindi il perché, alla caratteristica a gradino mostrata, si attribuisce il termine di ideale. È quella a cui corrispondono i più alti margini di rumore a parità



di tensione di alimentazione. La caratteristica ideale rappresenta quindi ciò con cui si confrontano e a cui devono tendere le caratteristiche di invertitori reali.

### LA TENSIONE DI SOGLIA $V_{TH}$

Osserviamo la caratteristica dell'invertitore ideale: vediamo che esiste un preciso valore della tensione di ingresso, chiamato tensione di soglia, a cui avviene la transizione alto – basso in uscita. Il termine “soglia” indica proprio il fatto che il superamento di questo valore da parte della tensione d’ingresso provoca un cambiamento “netto” di stato logico in uscita. Osserviamo ora la caratteristica reale: in questa non esiste un preciso valore ma piuttosto un intervallo di valori entro cui avviene la transizione tra i livelli logici. Ci chiediamo quindi: si può ancora identificare una tensione di soglia oppure non ha più senso parlarne? La risposta è che possiamo identificare la tensione di soglia anche per un invertitore reale. Vediamo come e che valore ha.

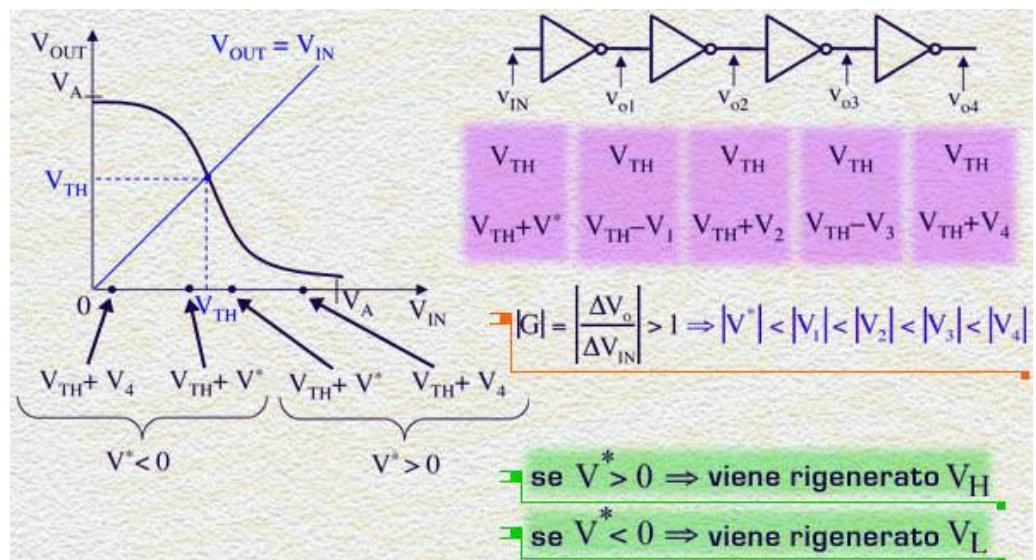


### TENSIONE DI SOGLIA DELL'INVERTITORE REALE

Mostreremo ora che si può identificare come tensione di soglia dell'invertitore reale quel valore della tensione d’ingresso a cui corrisponde un ugual valore della tensione d’uscita. La tensione di soglia si può quindi ottenere geometricamente intersecando la caratteristica con la retta  $V_{OUT} = V_{IN}$ .

Vediamo perché questo particolare valore può essere visto come soglia dell'invertitore. Consideriamo a questo proposito una catena di un numero pari di invertitori in cascata. Se l’ingresso  $V_{IN}$  è a livello logico alto anche l’uscita della catena sarà alta, se l’ingresso a basso anche l’uscita sarà bassa. Quindi l’uscita della catena riproduce il livello logico presente in ingresso.

Consideriamo ora una tensione d’ingresso  $V_{IN}$  pari alla tensione di soglia  $V_{TH}$ , dato che questa è la tensione per cui l’uscita uguaglia l’ingresso, tutti gli invertitori presenteranno una tensione d’uscita pari a  $V_{TH}$ .



Consideriamo ora una  $V_{IN}$  leggermente superiore a  $V_{TH}$ , pari a  $V_{TH} + V^*$  con  $V^* > 0$ . È immediato verificare graficamente che l’uscita del primo invertitore sarà minore di  $V_{TH}$ , l’uscita del secondo maggiore e così via.

Considerando ora che nell’intorno della tensione  $V_{TH}$  il guadagno dell’invertitore è, in modulo, maggiore di 1, si deduce che, in modulo,  $V_1$  sarà maggiore di  $V^*$ ,  $V_2$  di  $V_1$  e così via. Ovvero, lo scostamento dal valore  $V_{TH}$  aumenta man mano che si procede lungo la catena.

Quindi, se l’ingresso è anche leggermente superiore a  $V_{TH}$ , l’uscita sarà prossima al livello logico alto. Si è così verificata una “rigenerazione” del livello logico alto  $V_H$  a partire da un valore leggermente superiore a  $V_{TH}$ .

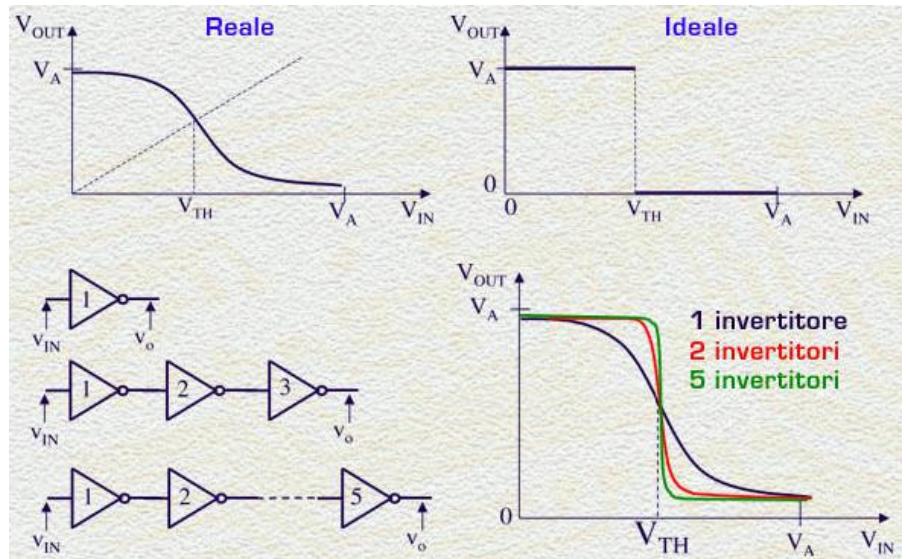
Analogamente si può discutere il caso di  $V^*$  negativo: l’ingresso  $V_{IN}$  sarà minore di  $V_{TH}$  e lo scostamento dalla tensione di soglia  $V_{TH}$  sempre crescente, porterà l’uscita  $V_{O4}$  verso il livello logico basso. Viene quindi “rigenerato” il livello basso  $V_L$  a partire da una tensione leggermente inferiore a  $V_{TH}$ . Possiamo perciò concludere che il valore  $V_{TH}$ , come l’abbiamo definito, costituisce effettivamente una soglia tra i livelli basso e alto. Infatti i valori superiori a  $V_{TH}$  sono correttamente interpretati come ‘1’ logico in ingresso ed i valori inferiori a  $V_{TH}$  come ‘0’ logico.

## TENSIONE DI SOGLIA

Mostriamo ora come esista un legame tra la tensione di soglia di un invertitore ideale e quella di un invertitore reale. Consideriamo tre circuiti: il primo costituito da un solo invertitore, il secondo da tre invertitori in cascata e il terzo da cinque invertitori. L'operazione svolta da ognuno di questi tre circuiti è sempre quella di inversione logica.

Se osserviamo ora le caratteristiche di questi tre circuiti, noteremo che all'aumentare del numero di invertitori la caratteristica ingresso – uscita tende verso quella dell'invertitore ideale, con transizione sempre più ripida in corrispondenza di  $V_{TH}$ , che rimane invariata.

L'aumento di ripidità della transizione è spiegabile con l'aumento di guadagno del circuito all'aumentare del



numero di stadi. La tensione di soglia  $V_{TH}$  rimane la stessa perché è il valore per il quale la tensione d'uscita di ogni invertitore uguaglia quella al suo ingresso, relazione questa che si mantiene anche ponendo più invertitori in cascata. Questa analisi ci offre un'ulteriore spiegazione di quanto precedentemente visto sulla rigenerazione dei livelli logici con il passare di un segnale, appena sopra o appena sotto la soglia, attraverso più stadi di elaborazione, qui rappresentati, per semplicità, da invertitori.

## Lezione n. 16 : Porte logiche: generalità, 2

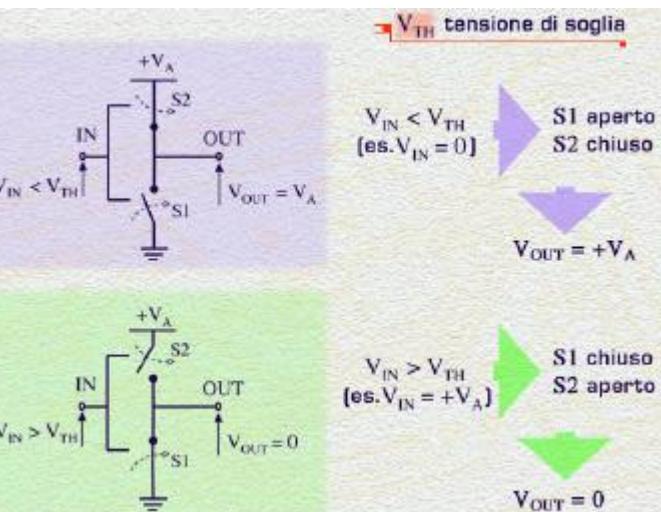
### L'INVERTITORE A DUE INTERRUTTORI

In figura vediamo un'altra possibile implementazione di un invertitore logico. È costituita da due interruttori S1 e S2 che sono comandati dalla stessa tensione d'ingresso  $V_{IN}$ . Quando  $V_{IN}$  è minore della tensione di soglia  $V_{TH}$ , S1 è in posizione "aperto" mentre S2 è in posizione "chiuso", quindi la tensione d'uscita vale  $V_A$ , corrispondente all'1 logico.

Quando  $V_{IN}$  è maggiore della tensione di soglia, S1 è chiuso ed S2 è aperto, quindi l'uscita è pari a 0 volt. Vediamo così che anche questo circuito svolge la funzione di inversione logica.

### LA CARATTERISTICA INGRESSO – USCITA.

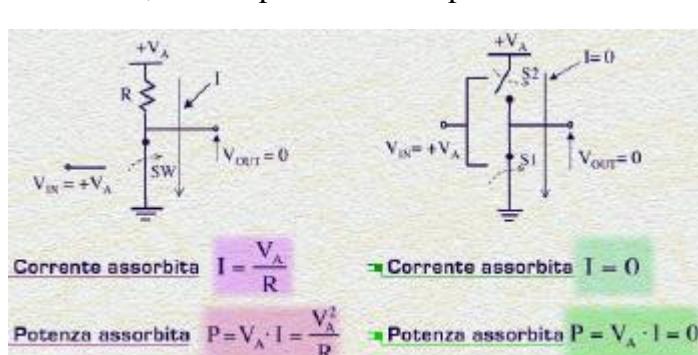
La caratteristica ingresso – uscita dell'invertitore a due interruttori è identica a quella vista per l'invertitore ad un interruttore ed una resistenza.



Per tensioni d'ingresso minori di  $V_{TH}$  l'uscita è alta, pari a  $V_A$ ; per tensioni  $V_{IN} > V_{TH}$  l'uscita è bassa, pari a 0. L'invertitore a due interruttori presenta però un grande vantaggio rispetto all'invertitore ad un solo interruttore: quando l'uscita è bassa l'invertitore con la resistenza assorbe dall'alimentazione una corrente pari a  $I = V_A/R$ . Questo comporta un consumo di potenza uguale a:  $P = V_A \cdot I = V_A^2 / R$ . L'invertitore a due interruttori invece non assorbe alcuna corrente,

quindi il consumo di potenza è nullo. Si può facilmente vedere che, con l'uscita alta, entrambi gli invertitori non assorbono corrente.

L'invertitore a due interruttori presenta quindi il vantaggio di non assorbire corrente in regime stazionario, ovvero per tutto il tempo in cui l'uscita ha un valore costante (0 o 1 logico).



diversamente alla tensione  $V_{IN}$ : infatti per una certa  $V_{IN}$ , se l'interruttore S1 si chiude S2 si deve aprire. Comunque il vantaggio di non consumare potenza in condizioni stazionarie è notevole e giustifica pienamente lo sforzo per la tecnologia più avanzata. Il tema del consumo di potenza elettrica dei circuiti elettronici ed in particolare di quelli digitali richiede un'ulteriore analisi che ora svolgeremo.

### L'INVERTITORE A DUE INTERRUTTORI

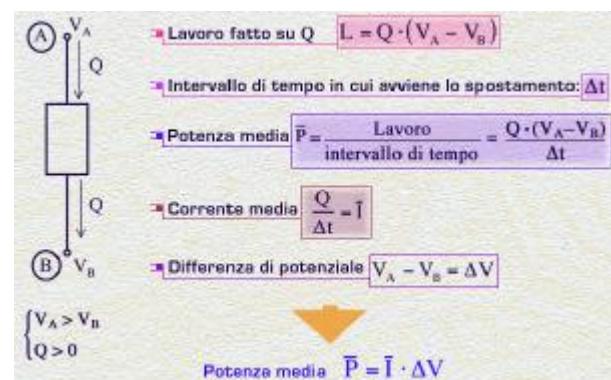
Facciamo una premessa che servirà come richiamo del concetto di potenza elettrica e ci aiuterà a comprendere meglio successive considerazioni. Consideriamo a tale proposito un generico bipolo posto tra due nodi A e B, a potenziale  $V_A$  e  $V_B$  rispettivamente e con  $V_A > V_B$ . Supponiamo che una

Questo vantaggio deriva dal fatto che in ognuno dei due stati logici c'è un interruttore aperto (S1 o S2). Questo vantaggio ha comunque un costo che è la necessità di realizzare due interruttori invece che uno solo considerando che un interruttore è tecnologicamente più complesso di un resistore. Inoltre l'interruttore S2, che sostituisce il resistore, è diverso dall'interruttore S1 perché deve rispondere

quantità di carica  $Q$  positiva parta dal nodo A e raggiunga il nodo B. Il lavoro fatto dal campo elettrico su tale carica è pari a  $Q \cdot (V_A - V_B)$ . Se tale spostamento di carica avviene in un tempo  $\Delta t$  allora la potenza media impiegata è uguale a:

$$\bar{P} = \frac{Q \cdot (V_A - V_B)}{\Delta t}$$
 e, dato che  $Q/\Delta t$  è la corrente media ( $\bar{I}$ )

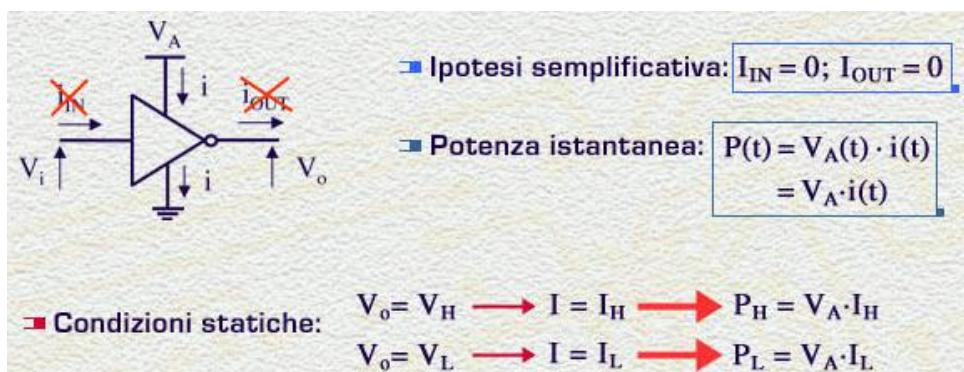
( $\bar{I}$ ) nel bipolo durante  $\Delta t$  otteniamo la nota equazione:  $\bar{P} = \bar{I} \cdot \Delta V$  (potenza media = corrente media  $\times$  differenza di potenziale).



### CONSUMO DI POTENZA STATICO.

Vediamo ora come calcolare il consumo di potenza di una porta logica, per semplicità un invertitore. Supponiamo che la corrente di ingresso  $I_{IN}$  e la corrente d'uscita  $I_{OUT}$  siano nulle o trascurabili rispetto alla corrente  $i$  che scorre tra l'alimentazione e massa. L'ipotesi di trascurare  $I_{IN}$  e  $I_{OUT}$  è verificata solo per alcuni circuiti e, generalmente, in condizioni particolari. Riteniamo comunque valida questa ipotesi per semplificare l'analisi del problema in questa fase introduttiva.

La potenza istantanea fornita dall'alimentatore all'istante di tempo  $t$  è  $P(t) = V_A(t) \cdot i(t)$ . Se supponiamo che  $V_A$  non varia:  $P(t) = V_A \cdot i(t)$ . Consideriamo inoltre una condizione statica, ovvero una condizione in cui le tensioni d'ingresso e d'uscita non variano. In generale la corrente  $i(t)$  è diversa, a seconda che l'uscita sia allo stato basso o alto. Indichiamo quindi con  $I_H$  la corrente  $i(t)$  quando l'uscita è allo stato alto. La potenza assorbita è quindi  $P_H = V_A \cdot I_H$ .



Analogamente quando l'uscita è allo stato basso  $P_L = V_A \cdot I_L$ . In generale  $P_H$  è diverso da  $P_L$ .

Per calcolare la potenza media consumata in un dato intervallo di tempo  $\Delta t$  è necessario conoscere per quanto tempo l'uscita si è trovata nello stato alto, tempo che indichiamo con  $\Delta t_H$  e per quanto tempo nello stato basso, che indicheremo con  $\Delta t_L$ .

La potenza statica media è data dall'espressione che vediamo.

$$\bar{P}_{\text{STATICA}} = \frac{(V_A I_H \Delta t_H + V_A I_L \Delta t_L)}{\Delta t_H + \Delta t_L}$$

energia totale

intervallo di tempo

Al numeratore è scritta l'energia totale impiegata, suddivisa in due addendi: il primo relativo al tempo in cui l'uscita è stata a livello alto il secondo relativo allo stato basso.

L'espressione si semplifica se si suppone che bit 0 e bit 1 siano equiprobabili e quindi che i tempi  $\Delta t_H$  e  $\Delta t_L$  siano in media uguali. Sotto questa ipotesi l'espressione della potenza statica media si riduce a:

■ bit 0 e 1 equiprobabili  $\rightarrow \Delta t_H = \Delta t_L$

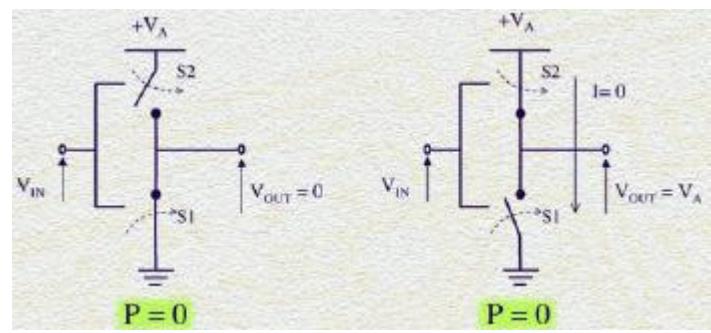
$$\rightarrow \bar{P}_{\text{STATICA}} = \frac{V_A \cdot (I_H + I_L)}{2}$$

### VALUTAZIONE DEL CONSUMO DI POTENZA STATICO

Consideriamo ora l'invertitore a due interruttori. Abbiamo già visto che in condizioni statiche questo circuito non assorbe corrente dall'alimentazione, in quanto in entrambi gli stati uno dei due interruttori è sempre aperto. Concludiamo quindi che il consumo di potenza statico di questo invertitore è nullo.

Questo non significa però che la porta logica non consumi potenza elettrica altrimenti sarebbe come dire che la porta svolga una funzione (quella di inversione del bit in ingresso) senza impiego di energia. Ciò non è possibile ed infatti mostreremo come il circuito consumi potenza proprio per commutare il suo stato: da 0 a 1 o viceversa. Il consumo di potenza si verifica quindi durante le commutazioni e non in condizioni statiche. Parleremo quindi di **CONSUMO DI POTENZA DINAMICO**.

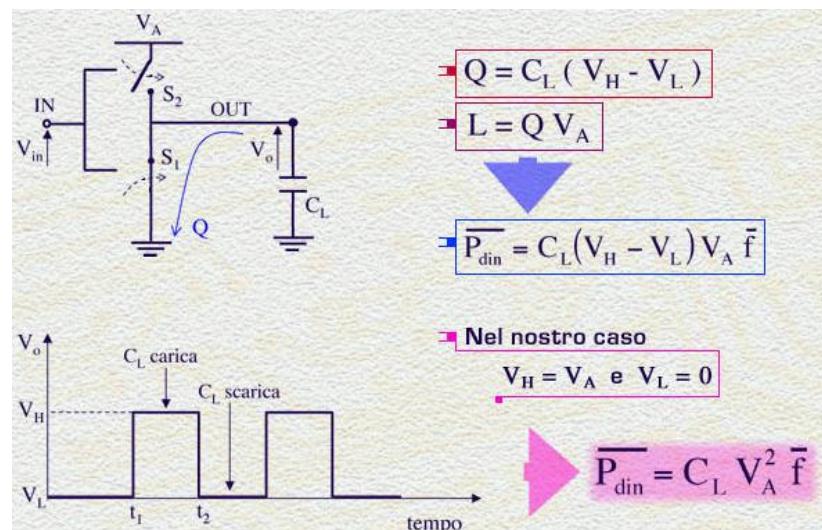
Vediamo come si calcola.



### CONSUMO DI POTENZA DINAMICO.

Consideriamo quindi l'invertitore a due interruttori e supponiamo che all'ingresso sia presente una sequenza di bit in modo che la tensione  $V_O$  in uscita, vista in funzione del tempo, sia quella rappresentata in figura. Inizialmente  $V_O$  è a 0 volt, all'istante  $t_1$  la porta commuta stato e la tensione sale a  $V_H$ . Occorre ora aver presente che, per cambiare il potenziale di un nodo di un circuito, è necessario aggiungere o togliere una certa quantità di carica elettrica. Ad esempio, se al nodo arriva una carica positiva, il potenziale del nodo aumenta. Viceversa se viene tolta una carica positiva il potenziale si abbassa. Nei circuiti elettronici ciò viene modellizzato tramite una capacità posta tra il nodo in questione e massa, nel nostro caso indicata con  $C_L$ . Perché la tensione d'uscita vari è necessario mettere o togliere carica dalle armature del condensatore. Inizialmente la tensione d'uscita è nulla, quindi il condensatore è scarico. Al tempo  $t_1$ , l'interruttore  $S_2$  si chiude e la tensione d'uscita sale a  $V_H$ . La tensione ai capi del condensatore  $C_L$  è cambiata da  $V_L$  a  $V_H$ , ciò significa che il condensatore, inizialmente scarico, ha ricevuto una quantità di carica positiva pari a  $C_L \cdot (V_H - V_L)$ . Tale carica è stata portata, nell'istante  $t_1$ , dalla linea di alimentazione  $V_A$ , sull'armatura del condensatore attraverso una corrente, come indicato in figura.

All'istante successivo  $t_2$ ,  $S_2$  si apre,  $S_1$  si chiude e la tensione del condensatore ritorna a  $V_L$ . Ciò significa che la carica  $Q$  lascia l'armatura del condensatore attraverso la corrente indicata, che fluisce verso massa. Se consideriamo un ciclo completo di carica e scarica di  $C_L$  osserviamo che, durante tale ciclo, una quantità di carica  $Q = C_L \cdot (V_H - V_L)$  è stata trasferita dal nodo di alimentazione, a potenziale  $V_A$  al nodo di massa a potenziale 0, con una sosta momentanea della carica, diciamo così, su  $C_L$ . L'alimentatore ha quindi compiuto un lavoro pari al prodotto tra la carica spostata e la differenza di potenziale, pari a  $V_A$ . Quindi  $L = Q \cdot V_A$ .



Per calcolare ora la potenza consumata occorre valutare il lavoro compiuto nell'unità di tempo. Se la porta opera ad una frequenza media pari ad  $f$  soprassegnato ( $\bar{f}$ ), significa che, in un secondo, sono avvenuti in media  $\bar{f}$  cicli di carica/scarica di  $C_L$ . Il lavoro medio compiuto nell'unità di tempo, pari per definizione alla potenza impiegata, è perciò uguale a:  $L = \bar{P}_{din} = C_L \cdot (V_H - V_L) \cdot V_A \cdot \bar{f}$ ; questa equazione esprime il consumo di potenza dinamico dell'invertitore. Nel nostro caso dato che  $V_H = V_A$  e  $V_L = 0$ , l'espressione si riduce a  $\bar{P}_{din} = C_L \cdot V_A^2 \cdot \bar{f}$ .

Discutiamo ora l'equazione che abbiamo appena dedotto. Facciamo alcune semplici osservazioni e considerazioni. Osserviamo che:

- la potenza cresce linearmente con il valore della capacità  $C_L$ , che rappresenta, in maniera equivalente, la somma di tutte le capacità connesse al nodo d'uscita.
  - Per ridurre il consumo di potenza è quindi importante minimizzare tutte queste capacità riducendo, ad esempio, le dimensioni di tutti i conduttori connessi al nodo d'uscita. Si può a tale proposito osservare che il continuo progresso tecnologico verso la riduzione delle dimensioni di dispositivi e circuiti ha come effetto diretto proprio la riduzione delle capacità ai nodi dei circuiti e quindi, a parità degli altri parametri, porta ad un minor consumo di potenza.
- la potenza dinamica con il quadrato della tensione di alimentazione. È quindi evidente il grande vantaggio di realizzare dispositivi e circuiti in grado di operare a basse tensioni.
  - Ad esempio se fosse possibile dimezzare la tensione di alimentazione, si otterrebbe un abbattimento del consumo di potenza di un fattore 4!
  - Ci possiamo allora domandare quali condizioni limitino la riduzione della tensione di alimentazione di una porta logica al di sotto di un certo valore minimo.
    - Una prima risposta è che la riduzione di  $V_A$  comporta una riduzione della tensione dei margini di rumore, e questo è uno svantaggio.
    - Inoltre vedremo successivamente che al diminuire della tensione di alimentazione, diminuisce anche la massima frequenza a cui la porta può operare, e anche questo è un limite di cui tenere conto.
- la potenza dinamica cresce linearmente con la frequenza media a cui opera il circuito. Questo ci dice che:
  - la scelta della frequenza a cui far funzionare un sistema digitale non può prescindere dal considerare anche la potenza necessaria. Questa condizione è particolarmente importante per tutti i sistemi portatili alimentati a batteria.

## VELOCITÀ DI RISPOSTA DELLE PORTE

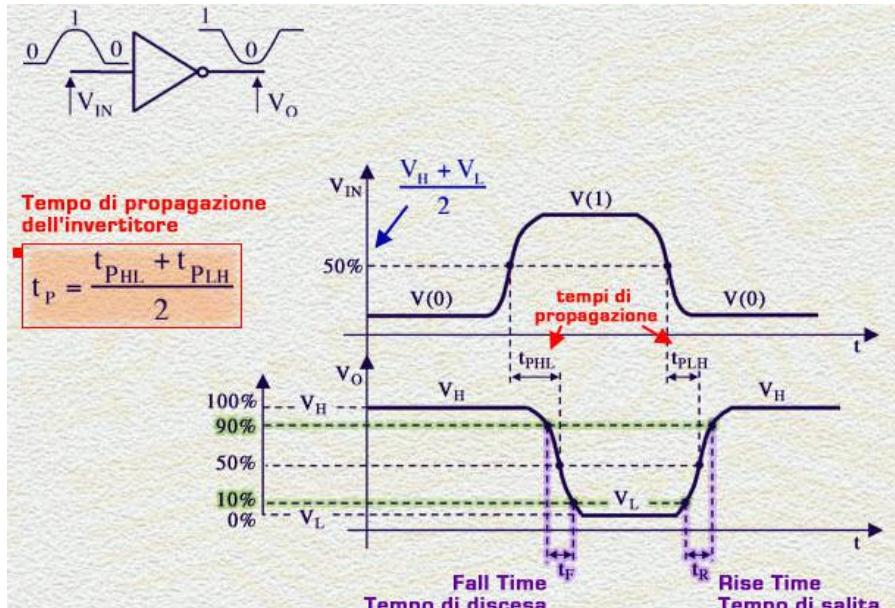
Consideriamo un invertitore con in ingresso il segnale mostrato in figura. Il segnale d'ingresso transisce dal livello basso al livello alto e, in seguito, ritorna a livello basso.

Osserviamo ora il segnale d'uscita dell'invertitore. I livelli logici sono ovviamente invertiti, ma notiamo un effetto importante: il segnale d'uscita risulta in ritardo rispetto a quello d'ingresso. La misura di tale ritardo ci

fornisce un'informazione sulla velocità con cui l'invertitore risponde al segnale d'ingresso. Quantifichiamo questo ritardo: per far ciò consideriamo dapprima la transizione in uscita da alto a basso. Individuiamo sui fronti di transizione i punti in cui i segnali d'ingresso e d'uscita si trovano al 50% della dinamica logica, ovvero gli istanti di tempo in cui le tensioni in ingresso e in uscita valgono  $(V_H + V_L)/2$ .

L'intervallo di tempo che intercorre tra questi due istanti è definito "tempo di propagazione della transizione alto – basso". In modo analogo viene definito il tempo di propagazione della transizione basso – alto. Questi due tempi, in generale, sono diversi, ma in casi particolari possono anche essere uguali. Comunque, si definisce 'tempo di propagazione dell'invertitore' la media aritmetica dei tempi di propagazione relativi alle due transizioni.

Concentrando ora la nostra attenzione unicamente al segnale d'uscita possiamo definire altri due tempi che caratterizzano la velocità della transizione tra i due livelli logici. A tale proposito

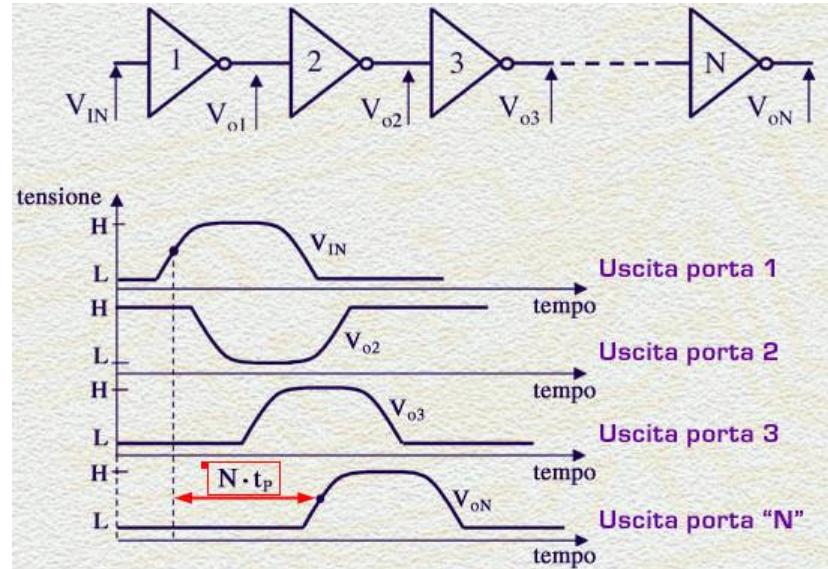


individuiamo i punti che si trovano al 10% e al 90% della dinamica logica. Questi punti definiscono due intervalli di tempo, uno relativo alla transizione di discesa, da alto a basso, l'altro relativo alla transizione di salita. Questi due tempi si definiscono tempo di discesa o fall-time e tempo di salita o rise-time. In una delle prossime lezioni impareremo a valutare tali tempi a partire dal circuito della porta logica. Ora ci limiteremo a capire l'influenza che questi tempi hanno sulle prestazioni di un sistema digitale in termini di velocità di elaborazione.

## TEMPO DI PROPAGAZIONE

Iniziamo a comprendere meglio il significato del tempo di propagazione e del suo effetto su di un sistema digitale. Consideriamo un sistema costituito da  $N$  invertitori posti in cascata. Questo caso dev'essere visto come una rappresentazione semplificata del percorso di un bit che viene elaborato attraverso più porte logiche generiche. Indichiamo con  $V_{IN}$  il segnale in ingresso al primo invertitore e con  $V_{oN}$  quella d'uscita dell' $n$ -esimo invertitore.

Si può dimostrare che il ritardo col quale si presenta il bit in uscita all' $N$ -esimo invertitore rispetto a quello d'ingresso è pari al numero  $N$  di porte traversate per il tempo di propagazione.



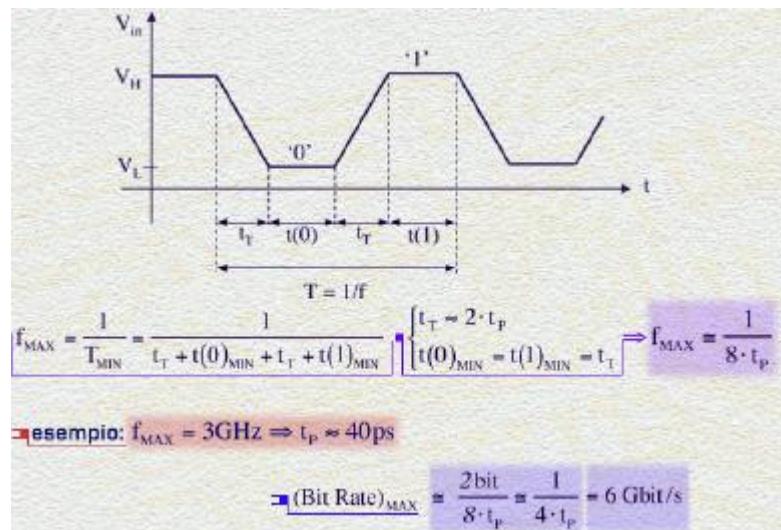
## MASSIMA FREQUENZA E BIT RATE

Vediamo ora come, conoscendo il tempo di propagazione di una porta logica possiamo determinare la massima frequenza a cui la porta stessa può operare e di conseguenza anche il tasso di bit, bit rate in inglese, che può sostenere. Consideriamo un invertitore il cui segnale d'uscita sia quello rappresentato in figura, ovvero una sequenza di bit 1-0-1-0. Il segnale è periodico di periodo  $T$  e frequenza  $f$ . Il periodo  $T$  è composto da quattro intervalli di tempo: 2 relativi alle transizioni 1 - 0 e 0 - 1, indicati con  $t_T$  e 2 intervalli che rappresentano la durata del bit '0' e del bit '1', indicati con  $t(0)$  e  $t(1)$ .

La massima frequenza è determinata dal minimo valore possibile del periodo  $T$ . Riteniamo ragionevole che la durata minima dei bit '0' o '1' debba essere pari al tempo di transizione  $t_T$ , in modo che la porta, che riceva in ingresso tali bit, possa completare la sua eventuale transizione. Possiamo inoltre supporre che, in prima approssimazione, il tempo di transizione  $t_T$  sia pari a 2 volte il tempo di propagazione, ricordando che in un tempo pari a  $t_p$  il segnale sicuramente conclude metà dell'intera transizione.

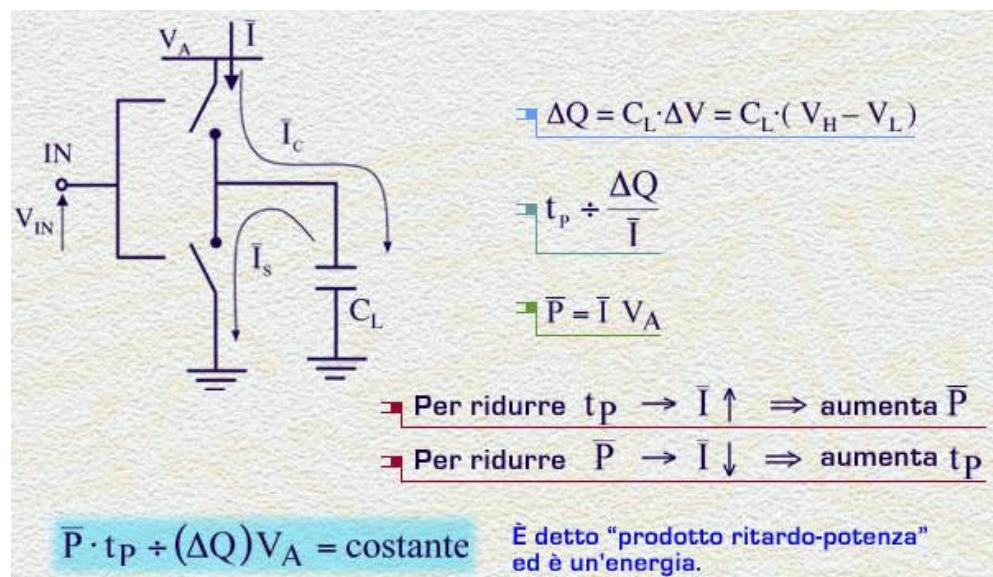
Quindi si ricava che il periodo minimo è, in prima approssimazione, uguale a 8 volte  $t_p$  e la frequenza massima sarà perciò circa uguale a  $1/(8t_p)$ . Ad esempio una frequenza massima pari a 3 GHz richiede che la porta abbia un tempo di propagazione di circa 40 ps.

Per quanto riguarda il massimo tasso di bit processabili, consideriamo che nel periodo  $T$  vengano elaborati 2 bit, quindi il massimo bit rate è pari a  $1/(4t_p)$ . Nell'esempio fatto, con 40 ps di tempo di propagazione, il massimo bit rate è pari a circa 6 Gbit per secondo.



## PRODOTTO RITARDO×POTENZA

Vediamo infine una interessante relazione tra la velocità di risposta di una porta e il suo consumo di potenza. Consideriamo sempre un invertitore. Indichiamo rispettivamente con  $\bar{I}_c$  e  $\bar{I}_s$  le correnti medie di carica e scarica della capacità  $C_L$  al nodo d'uscita durante le commutazioni. Queste due correnti possono essere uguali o diverse, indichiamo con  $\bar{I}$  il valore medio della corrente che scorre tra  $V_A$  e massa. Se ora indichiamo con  $\Delta Q$  la quantità di carica che dev'essere fornita o tolta dalla capacità  $C_L$  per far commutare l'uscita da basso ad alto o viceversa, possiamo dire che il tempo di propagazione è proporzionale al rapporto tra  $\Delta Q$  e la corrente media  $\bar{I}$ . D'altra parte la potenza media assorbita dalla porta è proporzionale proprio alla corrente media.



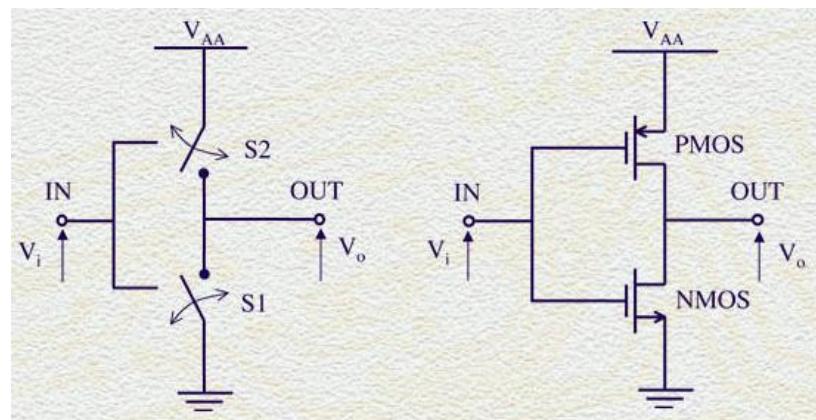
Vediamo quindi che, **per ridurre il tempo di propagazione, occorre aumentare la corrente**, ma ciò porta ad un aumento di potenza assorbita. Viceversa **per ridurre l'assorbimento di potenza occorre ridurre la corrente con il conseguente aumento del tempo di propagazione**. Quindi, supposto di poter scegliere il valore della corrente  $\bar{I}$ , tale scelta verrà fatta considerando sia l'assorbimento di potenza che la risultante velocità di risposta della porta. Inoltre, dato che il tempo di propagazione è inversamente proporzionale ad  $I$  mentre la potenza è proporzionale ad  $I$ , il prodotto tra la potenza e tempo di propagazione  $t_p$ , risulterà indipendente dalla corrente e pari ad una costante proporzionale al prodotto tra  $\Delta Q$  e  $V_A$ . Tale prodotto viene comunemente chiamato "**prodotto ritardo – potenza**" ed è un'energia. Nello specifico è una quantità legata strettamente all'energia impiegata dalla porta nell'effettuare una singola commutazione tra i due livelli logici. La **porta logica ottimale è quella per cui tale energia è minima, ovvero quella porta che possiede un basso tempo di propagazione e allo stesso tempo un basso assorbimento di potenza**. Vediamo che ciò è possibile solo con una riduzione della capacità  $C_L$  o della dinamica logica, detta anche *swing logico*, o della tensione di alimentazione.

## Lezione n. 17 : L'invertitore logico CMOS, 1 – analisi statica

### Introduzione

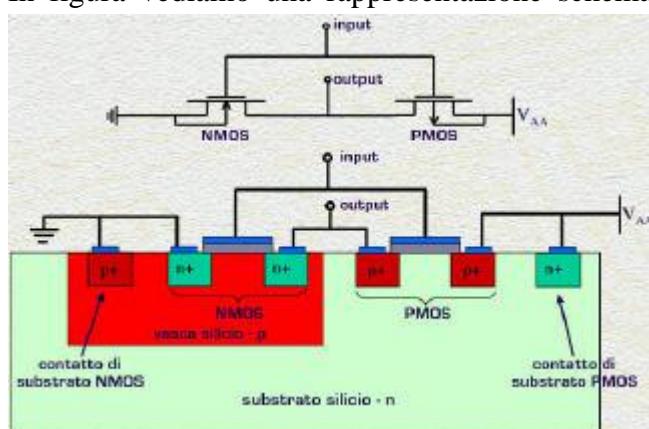
Nelle precedenti lezioni abbiamo visto l'invertitore a due interruttori, come è mostrato in figura. Questo invertitore può essere realizzato impiegando due MOSFET per svolgere la funzione di interruttori elettronici comandati dalla tensione  $V_i$  presente al morsetto IN.

In figura è mostrato il circuito, che è costituito da un transistore MOSFET a canale N, che svolge la funzione dell'interruttore S1 e un MOSFET a canale P per svolgere la funzione di S2. Gli elettrodi di GATE sono collegati insieme e costituiscono l'ingresso dell'invertitore mentre gli elettrodi di DRAIN costituiscono l'uscita dell'invertitore.



Il terminale di massa, a cui è connesso il SOURCE dell'nMOS è il terminale comune ad ingresso ed uscita al quale sono riferite le tensioni  $V_{IN}$ ,  $V_{OUT}$  e  $V_{AA}$ . Questo invertitore si chiama CMOS; che sta per COMPLEMENTARY MOS, dal nome della tecnologia microelettronica che permette la realizzazione, sullo stesso chip, di transistori MOS a canale N e a canale P.

In figura vediamo una rappresentazione schematica di un invertitore realizzato con la tecnologia CMOS.



Il MOSFET a canale N è realizzato all'interno di una vasca di silicio p che costituisce il substrato del transistor.

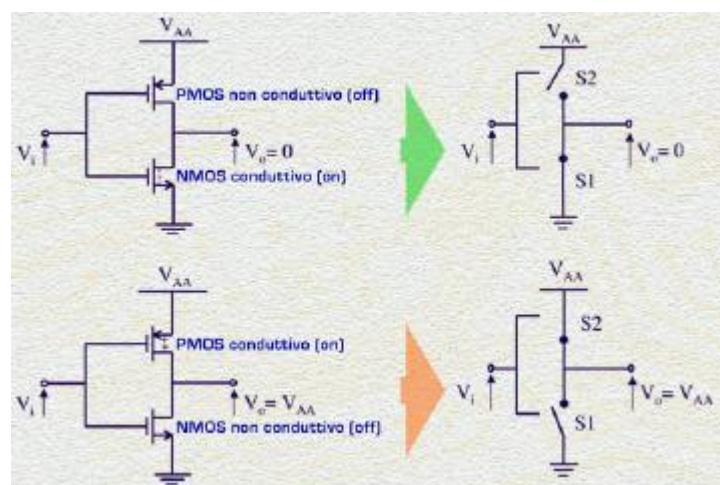
Il MOSFET a canale P è invece realizzato direttamente sul substrato di silicio N. Notiamo inoltre i contatti di substrato dell'nMOS e del pMOS cortocircuitati con il source dei rispettivi transistor in modo che le giunzioni P-N che source e drain formano con il substrato non conducano corrente.

Deduciamo per via sintetica il principio di funzionamento dell'invertitore CMOS.

Osserviamo che se per qualche valore della tensione d'ingresso  $V_i$ , l'nMOS presenta il canale conduttivo ed il pMOS invece no, allora il potenziale in uscita sarà quello di massa perché l'nMOS si comporterà equivalentemente ad un interruttore chiuso e il pMOS ad un interruttore aperto.

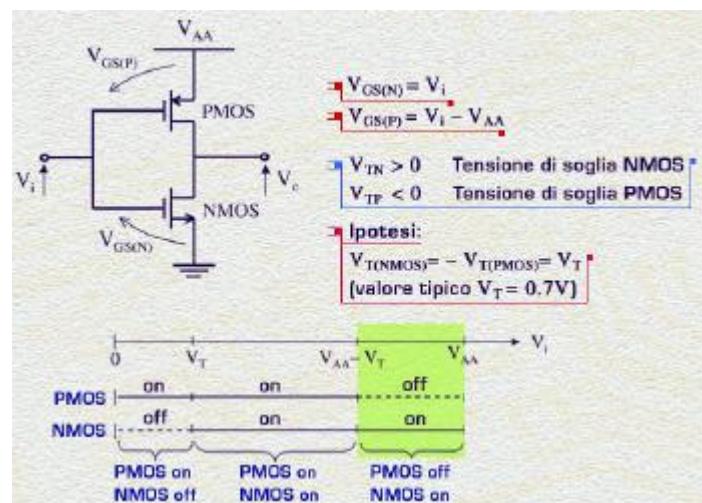
Analogamente se per qualche valore della tensione d'ingresso il pMOS è conduttivo e l'nMOS no, allora la tensione d'uscita sarà quella dell'alimentazione,  $V_{AA}$ , equivalentemente al caso dell'interruttore S2 chiuso ed S1 aperto.

Ora chiediamoci per quali valori della tensione d'ingresso si verificano il primo ed il secondo caso, ovvero determiniamo per quali valori di  $V_i$  l'nMOS e il pMOS sono conduttori e per quali non lo sono.



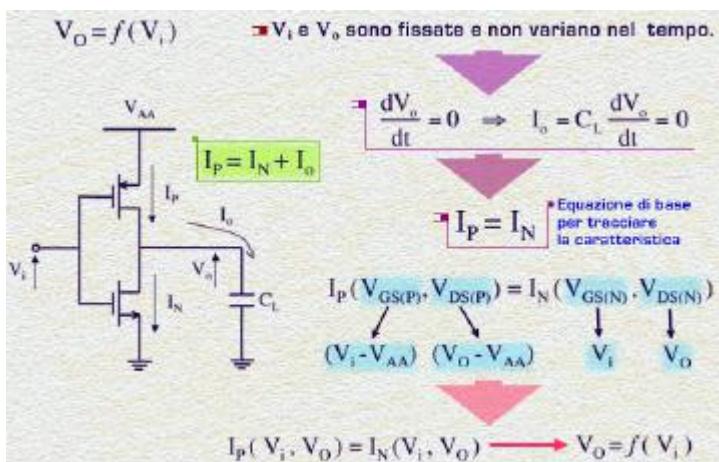
Iniziamo con l'osservare che la tensione gate – source dell'nMOS, che comanda la formazione o meno del canale conduttivo del transistor è proprio la tensione d'ingresso  $V_i$ . Analogamente osserviamo che la tensione gate – source del pMOS è pari a  $V_i - V_{AA}$ . Per determinare se un

MOSFET ha canale oppure no occorre verificare se la sua tensione gate – source è maggiore o minore della tensione di soglia del transistor. Facciamo quindi una premessa sulle tensioni di soglia dei due MOSFET. Come sappiamo la tensione di soglia di un nMOS  $V_{TN}$  è positiva mentre quella di un pMOS  $V_{TP}$  è negativa. Supporremo che tali due tensioni siano uguali in modulo e indicheremo tale valore con  $V_T$ . L'uguaglianza in modulo delle tensioni di soglia è comunemente verificata nelle tecnologie CMOS. Detto questo possiamo quindi rappresentare sull'asse delle tensioni d'ingresso  $V_i$  gli intervalli in cui i MOSFET sono accesi. L'nMOS sarà conduttivo per tutti i valori di  $V_i$  superiori alla tensione di soglia  $V_T$  e non conduttivo al di sotto di  $V_T$ . Il pMOS invece sarà conduttivo per tutti i valori di  $V_i$  inferiori a  $V_{AA} - V_T$ , infatti per tutti questi valori il gate del pMOS è a potenziale più basso rispetto al suo source di una quantità superiore alla tensione di soglia, condizione necessaria per formare il canale di lacune. Cominciamo ad osservare tre intervalli distinti di  $V_i$ . Il primo tra 0 e  $V_T$  in cui l'nMOS è spento ed il pMOS acceso, il secondo tra  $V_T$  e  $V_{AA} - V_T$  in cui entrambi i MOSFET sono accesi, il terzo, tra  $V_{AA} - V_T$  e  $V_{AA}$  in cui l'nMOS è acceso ed il pMOS spento.



### La caratteristica statica

Vediamo ora come determinare i valori della tensione d'uscita  $V_o$  al variare di  $V_i$  ovvero determiniamo la caratteristica ingresso – uscita dell'invertitore CMOS. La caratteristica ingresso –



uscita di una porta logica è data dalla funzione che lega la tensione d'uscita a quella d'ingresso in continua, ovvero in condizione statica, condizione nella quale sia la tensione d'ingresso che la tensione d'uscita non variano nel tempo. Sperimentalmente una caratteristica si ottiene per punti, fissando  $V_i$  ad un determinato valore, misurando quindi  $V_o$  dopo aver aspettato un tempo sufficiente affinché la stessa tensione d'uscita si sia stabilizzata al suo valore di equilibrio. Quindi si cambia valore a  $V_i$  e si acquisisce

il nuovo valore di  $V_o$ . Matematicamente, la caratteristica si ricava ricercando un'equazione che lega  $V_o$  a  $V_i$ . Nel nostro caso, possiamo osservare che il valore di  $V_i$  fissa le tensioni gate – source sia del pMOS che dell'nMOS. Queste tensioni a loro volta contribuiscono a fissare le correnti  $I_P$  e  $I_N$  nei transistori. Il bilancio di correnti al nodo d'uscita fornisce un'equazione  $I_P = I_N + I_o$  che permette di determinare la caratteristica, come vedremo tra poco. Dedurremo la caratteristica assumendo che il carico connesso all'uscita sia capacitivo, come lo è nella maggioranza dei casi. Infatti all'interno di un sistema digitale CMOS una porta logica comanda, con la sua tensione d'uscita, altre porte o circuiti il cui ingresso è costituito da gate di MOSFET, che dal punto di vista impedenziale sono appunto capacità. Se il carico d'uscita è capacitivo allora, in regime statico, in cui la tensione d'uscita non varia, si avrà anche corrente d'uscita nulla, e l'equazione di bilancio al nodo d'uscita si semplifica in  $I_P = I_N$ , ovvero nell'uguaglianza delle correnti dei due MOSFET. Ora, sia  $I_N$  che  $I_P$  sono funzioni note delle tensioni gate – source e gate – drain dei rispettivi MOSFET's, queste a loro volta sono tutte riconducibili alle tensioni  $V_i$  e  $V_o$  del circuito come mostrato. Quindi l'equazione  $I_N = I_P$  stabilisce proprio una relazione funzionale tra la tensione d'uscita e quella d'ingresso. Esplicitando quest'equazione otterremo la caratteristica ingresso – uscita dell'invertitore.

Iniziamo a tracciare la caratteristica ingresso – uscita relativa agli intervalli in cui uno dei due transistori è spento, ovvero non conduttivo. Consideriamo dapprima l'intervallo di  $V_i$  compreso tra 0 e la tensione di soglia  $V_T$ . In quest'intervallo l'nMOS è spento quindi la sua corrente  $I_N$  è nulla. Dato che consideriamo una condizione statica è  $I_P = I_N$  quindi anche il pMOS, pur essendo conduttivo, ha corrente nulla. Da ciò si deduce che la tensione drain – source del pMOS dev'essere nulla altrimenti, se non lo fosse, si avrebbe una corrente nel pMOS. Pertanto, dato che la  $V_{DS}$  del pMOS è 0, la tensione d'uscita è uguale alla tensione  $V_{AA}$  di alimentazione.

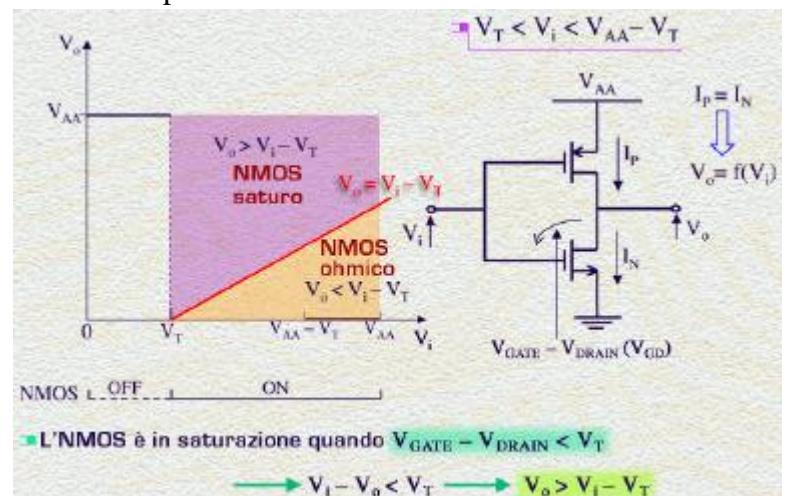
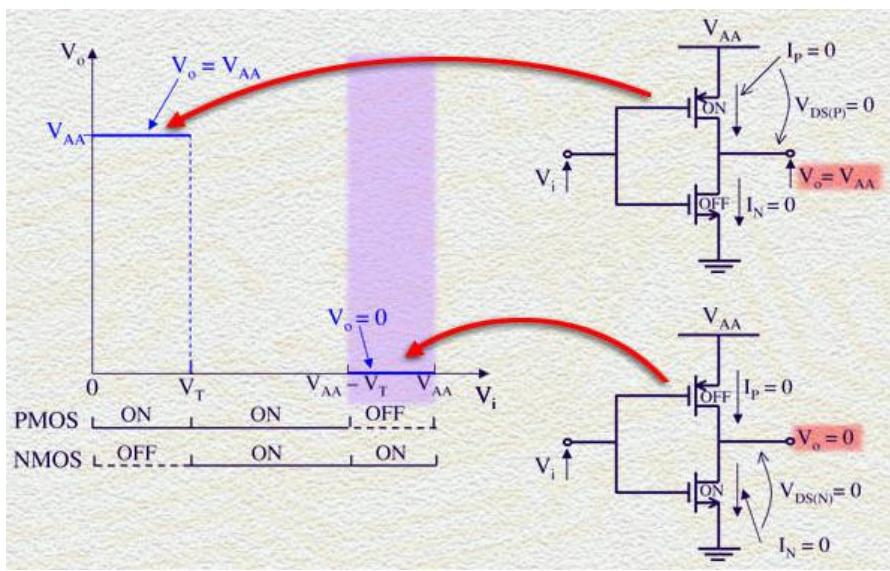
Il valore  $V_{AA}$  in uscita si mantiene per tutti i valori di  $V_i$  per cui l'nMOS è spento. Perciò, per  $V_i$  compreso tra 0 e  $V_T$ , la caratteristica ingresso – uscita è il tratto di retta orizzontale  $V_O = V_{AA}$ . Analogamente quando  $V_i$  è compresa tra  $V_{AA} - V_T$  e  $V_{AA}$  il pMOS è non – conduttivo, quindi  $I_P$  e  $I_N$  sono entrambi nulle. L'nMOS è conduttivo e, non portando corrente, ha una tensione drain – source necessariamente nulla. Perciò anche la tensione al nodo d'uscita è nulla per tutti i valori di  $V_i$  per cui l'nMOS è spento. La caratteristica tra  $V_{AA} - V_T$  e  $V_{AA}$  è quindi un tratto orizzontale coincidente con l'asse delle ascisse.

Perciò, per  $V_i$  compreso tra 0 e  $V_T$ , la caratteristica ingresso – uscita è il tratto di retta orizzontale  $V_O = V_{AA}$ . Analogamente quando  $V_i$  è compresa tra  $V_{AA} - V_T$  e  $V_{AA}$  il pMOS è non – conduttivo, quindi  $I_P$  e  $I_N$  sono entrambi nulle. L'nMOS è conduttivo e, non portando corrente, ha una tensione drain – source necessariamente nulla. Perciò anche la tensione al nodo d'uscita è nulla per tutti i valori di  $V_i$  per cui l'nMOS è spento. La caratteristica tra  $V_{AA} - V_T$  e  $V_{AA}$  è quindi un tratto orizzontale coincidente con l'asse delle ascisse.

Vediamo ora come ricavare la caratteristica ingresso – uscita quando la tensione d'ingresso è compresa tra  $V_T$  e  $V_{AA} - V_T$ . In questo intervallo entrambi i MOSFET sono conduttori. Ciò significa che, a differenza dei due precedenti casi, ora esiste un percorso conduttivo tra l'alimentazione e massa e di conseguenza una corrente scorre nei due MOSFET. Per poter esplicitare l'equazione  $I_P = I_N$ , e ricavare quindi la funzione  $V_O = f(V_i)$ , occorre prima determinare le condizioni di polarizzazione dei due MOSFET, ovvero sapere se ciascun MOSFET è polarizzato in zona ohmica o in saturazione.

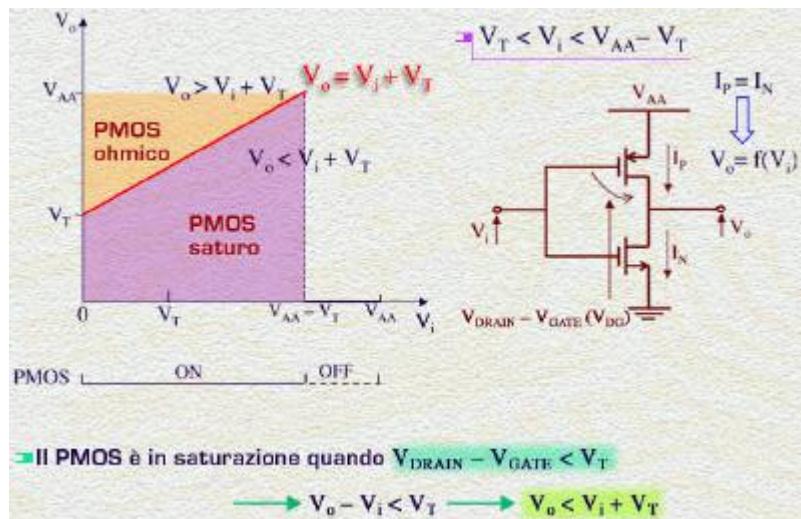
Vediamo quindi come dedurre quest'informazione. Consideriamo dapprima l'nMOS. Sappiamo che un MOSFET a canale N, se è conduttivo, risulta polarizzato in zona di saturazione quando in prossimità del drain il canale si interrompe. Questo si verifica se la differenza di potenziale tra gate e drain è minore della tensione di soglia. Dato che la tensione di gate coincide con la tensione d'ingresso e la tensione di drain con la tensione d'uscita, la condizione è  $V_i - V_o < V_T$  che scritta in termini di  $V_O$  è  $V_O > V_i - V_T$ . Questa disequazione è soddisfatta in tutti i punti che si trovano al di sopra della retta  $V_O = V_i - V_T$ , ovvero nella regione di piano indicata in figura. In questa regione l'nMOS opera in saturazione. Al contrario la regione al di sotto della retta rappresenta i valori di  $V_i$  e  $V_O$  per cui l'nMOS opera in regione ohmica.

Questo risultato può essere verificato qualitativamente per via intuitiva: infatti fissata la tensione di gate di un nMOS, la regione ohmica è caratterizzata da tensioni drain – source più basse rispetto a quelle che si hanno in regione di saturazione, quindi la regione sotto la retta, che è quella in cui il MOSFET ha tensione drain – source, pari a  $V_O$  più bassa, deve rappresentare la regione ohmica.

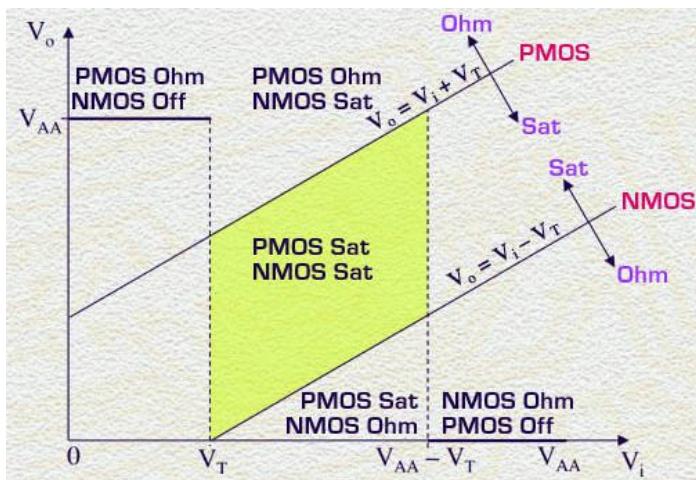


Un analogo discorso vale per il pMOS. Il pMOS è polarizzato in regione di saturazione quando la differenza di potenziale tra drain e gate è minore del modulo della tensione di soglia.

Infatti affinché non  $V_i$  sia canale di lacune in prossimità del drain è necessario che il potenziale di gate non sia più basso di quello di drain di una quantità superiore a  $V_T$ . Dalla condizione di saturazione ricaviamo la disequazione  $V_o < V_i + V_T$ , che è soddisfatta in tutti i punti che stanno sotto la retta  $V_o = V_i + V_T$  rappresentata in figura. Al di sopra della retta il pMOS opera in regione ohmica.



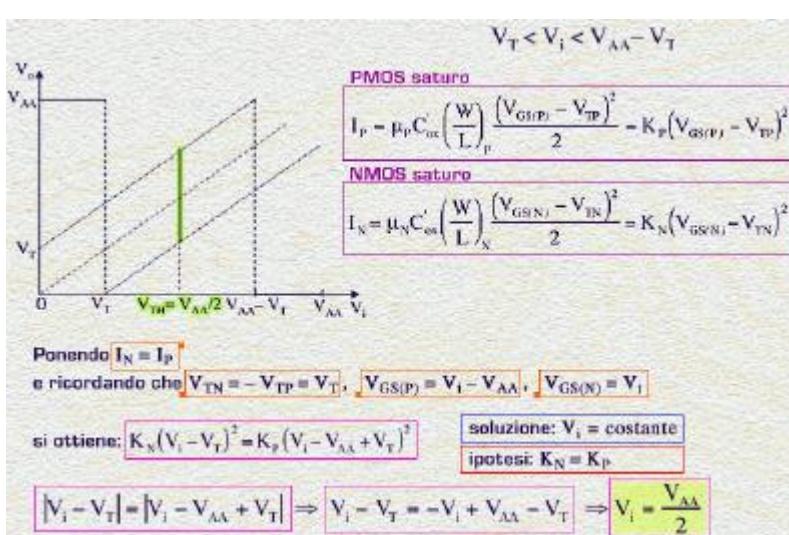
Riassumiamo ora in un grafico tutte le regioni di polarizzazione dei MOSFET entro il piano  $V_o - V_i$ .



Con  $V_i$  compreso tra 0 e  $V_T$  l'nMOS è spento e con  $V_i$  compreso tra  $V_{AA} - V_T$  e  $V_{AA}$  il pMOS è spento. Tracciamo quindi la retta  $V_o = V_i + V_T$  al di sopra della quale il pMOS è ohmico e sotto la quale saturo. Tracciamo infine la retta  $V_o = V_i - V_T$  al di sopra della quale l'nMOS è saturo e sotto la quale è ohmico. Consideriamo l'intervallo di valori di  $V_i$  compresi tra  $V_T$  e  $V_{AA} - V_T$ . Essendo ora a conoscenza dello stato di polarizzazione dei MOSFET, possiamo ricavare la caratteristica dell'invertitore quando la tensione d'ingresso è interna all'intervallo tra  $V_T$  e  $V_{AA} - V_T$ .

Ricaviamo dapprima la caratteristica nella regione in cui entrambi i MOSFET sono saturi. Esplicitiamo quindi l'equazione  $I_p = I_N$  con le espressioni delle correnti dei MOS in regione di saturazione. Semplifichiamo la notazione indicando con  $K_p$  e  $K_n$  i prodotti  $\mu C_{ox}' \frac{W}{L}$  dei rispettivi MOSFET. Ricordando ora che  $V_{TN} = -V_{TP} = V_T$  e che le tensioni gate – source dei transistor sono  $V_i - V_{AA}$  per il pMOS e  $V_i$  per l'nMOS, otteniamo l'equazione delle correnti riportata nel riquadro rosa in figura. Notiamo che in questa equazione non compare la tensione d'uscita  $V_o$  e che la soluzione sarà quindi nella forma  $V_i = \text{costante}$ , che nel piano  $V_o - V_i$  è rappresentata da una retta verticale. Determiniamo il valore della costante risolvendo l'equazione nell'ipotesi semplificativa di  $K_p = K_n$ . Tale ipotesi è comunque giustificata dal fatto che, a livello di progetto dell'invertitore, la scelta dei rapporti di forma  $W/L$  dei due transistor viene fatta affinché la condizione  $K_p = K_n$  sia soddisfatta.

regione in cui entrambi i MOSFET sono saturi. Esplicitiamo quindi l'equazione  $I_p = I_N$  con le espressioni delle correnti dei MOS in regione di saturazione. Semplifichiamo la notazione indicando con  $K_p$  e  $K_n$  i prodotti  $\mu C_{ox}' \frac{W}{L}$  dei rispettivi MOSFET. Ricordando ora che  $V_{TN} = -V_{TP} = V_T$  e che le tensioni gate – source dei transistor sono  $V_i - V_{AA}$  per il pMOS e  $V_i$  per l'nMOS, otteniamo l'equazione delle correnti riportata nel riquadro rosa in figura. Notiamo che in questa equazione non compare la tensione d'uscita  $V_o$  e che la soluzione sarà quindi nella forma  $V_i = \text{costante}$ , che nel piano  $V_o - V_i$  è rappresentata da una retta verticale. Determiniamo il valore della costante risolvendo l'equazione nell'ipotesi semplificativa di  $K_p = K_n$ . Tale ipotesi è comunque giustificata dal fatto che, a livello di progetto dell'invertitore, la scelta dei rapporti di forma  $W/L$  dei due transistor viene fatta affinché la condizione  $K_p = K_n$  sia soddisfatta.



Comprenderemo successivamente le ragioni che sostengono questa scelta. Semplifichiamo quindi  $K_n$

e  $K_p$ , estraiamo la radice di entrambi i membri ricordando che  $\sqrt{x^2} = |x|$ . Notiamo ora che a secondo membro, dentro il modulo, il termine è negativo essendo  $V_i < V_{AA} - V_T$ , quindi togliendo il modulo occorre cambiare segno. La soluzione dell'equazione è  $V_i = V_{AA}/2$ , che corrisponde ad un tratto di retta verticale delimitato entro la regione in cui entrambi i MOSFET sono saturi. Notiamo che la particolare posizione del tratto verticale, a  $V_{AA}/2$ , cioè metà della logica dinamica, deriva dall'aver scelto  $K_N = K_p$ . Se  $K_N$  fosse stato maggiore di  $K_p$  la retta si sarebbe trovata in corrispondenza di tensioni d'ingresso minori ovvero spostate a sinistra o viceversa se  $K_N$  fosse  $< K_p$ . Notiamo inoltre che tale valore di  $V_i = V_{AA}/2$  è proprio la tensione di soglia  $V_{TH}$  della porta logica in quanto la bisettrice del quadrante taglia la caratteristica proprio in questa regione. Si capisce ora che uno dei motivi della scelta di  $K_N = K_p$  è proprio quello di avere una soglia logica a metà dinamica, massimizzando così i margini di rumore.

Ricaviamo ora la caratteristica nelle due regioni in cui un transistore è ohmico e l'altro saturo.

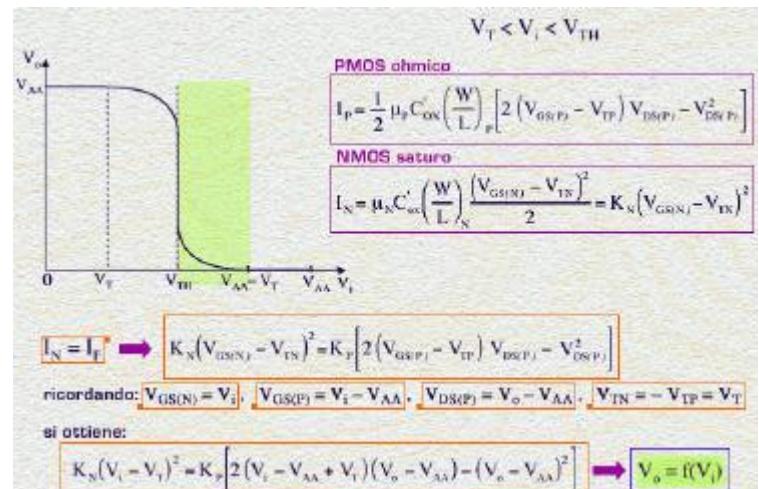
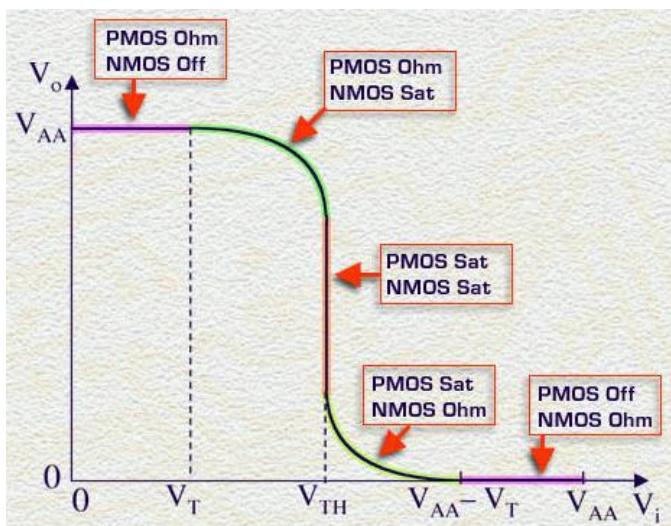
Consideriamo ad esempio la regione con  $V_i$  compreso tra  $V_T$  e  $V_{TH}$  in cui il pMOS è ohmico e l'nMOS è saturo.

Scriviamo le equazioni per le correnti dei due MOSFET's. La caratteristica si ottiene dall'equazione che uguaglia le correnti  $I_N = I_P$ .

Explicitando le tensioni gate – source e gate – drain in termini delle tensioni  $V_i$  e  $V_o$ , deduciamo un'equazione che lega proprio, in modo implicito, la tensione d'uscita a quella d'ingresso.

In modo analogo si ricava il tratto di caratteristica tra  $V_{TH}$  e  $V_{AA} - V_T$ .

Vediamo quindi in un grafico riassuntivo tutta la caratteristica ingresso – uscita dell'invertitore CMOS suddivisa nei suoi cinque tratti, ciascuno caratterizzato da una diversa regione di polarizzazione della coppia MOSFET's.



## Esercizio 1

Si determini la tensione di soglia di un invertitore logico CMOS nel caso generale in cui  $K_N \neq K_p$  e  $V_{TN} \neq -V_{TP}$  (si ricordi che  $V_{TN} > 0$  mentre  $V_{TP} < 0$ ). Si discuta il risultato.

## Soluzione

Determiniamo la tensione di soglia,  $V_{TH}$ , di un invertitore CMOS nel caso generale in cui  $K_N \neq K_P$ , considerando inoltre  $V_{TN} \neq -V_{TP}$  (ricordiamo che  $V_{TN}$  è maggiore di zero, mentre  $V_{TP}$  è minore di zero). Per calcolare il valore della tensione di soglia uguagliamo le correnti del PMOS e del NMOS nella condizione di saturazione, da qui ricaveremo il valore della tensione di ingresso che corrisponde alla tensione di soglia.

Quindi si ha:

$$K_N \cdot (V_{GS(N)} - V_{TN})^2 = K_P \cdot (V_{GS(P)} - V_{TP})^2$$

ricordando che

$$V_{GS(N)} = V_i = V_{TH} \quad V_{GS(P)} = V_i - V_{AA} = V_{TH} - V_{AA}$$

si ha

$$K_N \cdot (V_{TH} - V_{TN})^2 = K_P \cdot (V_{TH} - V_{AA} - V_{TP})^2$$

da qui

$$\begin{aligned} \sqrt{K_N} \cdot |V_{TH} - V_{TN}| &= \sqrt{K_P} \cdot |V_{TH} - V_{AA} - V_{TP}| \\ \sqrt{K_N} \cdot (V_{TH} - V_{TN}) &= \sqrt{K_P} \cdot (V_{AA} + V_{TP} - V_{TH}) \end{aligned}$$

raccogliendo opportunamente i termini si ricava il valore di  $V_{TH}$

$$V_{TH} = \frac{\sqrt{K_P} \cdot V_{AA} + \sqrt{K_N} \cdot V_{TN} + \sqrt{K_P} \cdot V_{TP}}{\sqrt{K_N} + \sqrt{K_P}}$$

ricordando che  $K = \frac{1}{2} \cdot \mu C_{ox} \cdot \left( \frac{W}{L} \right)$  e raccogliendo  $\sqrt{K_P}$  possiamo ridurre l'espressione nella

forma

$$V_{TH} = \frac{V_{AA} + V_{TP} + \sqrt{\frac{\mu_N \cdot (W/L)_N}{\mu_P \cdot (W/L)_P} \cdot V_{TN}}}{1 + \sqrt{\frac{\mu_N \cdot (W/L)_N}{\mu_P \cdot (W/L)_P}}}$$

se consideriamo che nella maggior parte dei casi  $V_{TN} = -V_{TP} = V_T$  si ha

$$V_{TH} = \frac{V_{AA} + V_T \left( \sqrt{\frac{\mu_N \cdot (W/L)_N}{\mu_P \cdot (W/L)_P}} - 1 \right)}{1 + \sqrt{\frac{\mu_N \cdot (W/L)_N}{\mu_P \cdot (W/L)_P}}}$$

A questo punto se si impone che  $K_N = K_P$ , cioè che  $\frac{(W/L)_P}{(W/L)_N} = \frac{\mu_N}{\mu_P}$ , si ottiene che la tensione di

soglia è pari a

$$V_{TH} = \frac{V_{AA}}{2}$$

## Lezione n. 18 : L'invertitore logico CMOS, 2 – analisi dinamica

### Stima del tempo di propagazione

Valuteremo ora i tempi di propagazione di un invertitore CMOS. Risolveremo il problema con alcune ipotesi semplificative. La prima ipotesi è che il segnale d'ingresso presenti transizioni istantanee tra i due livelli logici, ovvero abbia tempi di salita e di discesa trascurabili rispetto ai tempi di transizione del segnale d'uscita.

Questa è un'ipotesi semplificativa, vera solo in casi particolari. Infatti, in generale il segnale d'ingresso di una porta proviene da porte analoghe ed è quindi caratterizzato da tempi di salita e discesa confrontabili con quelli della porta in esame.

Indichiamo con  $C_L$  la capacità tra il nodo d'uscita e massa. I tempi di propagazione dell'invertitore sono strettamente legati ai tempi necessari alla carica o alla scarica della capacità  $C_L$ .

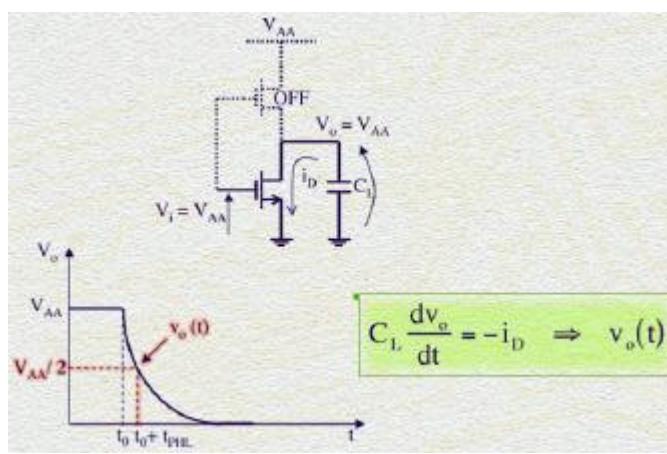
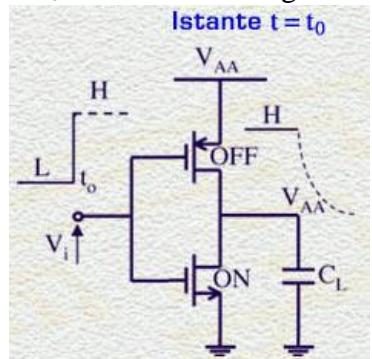
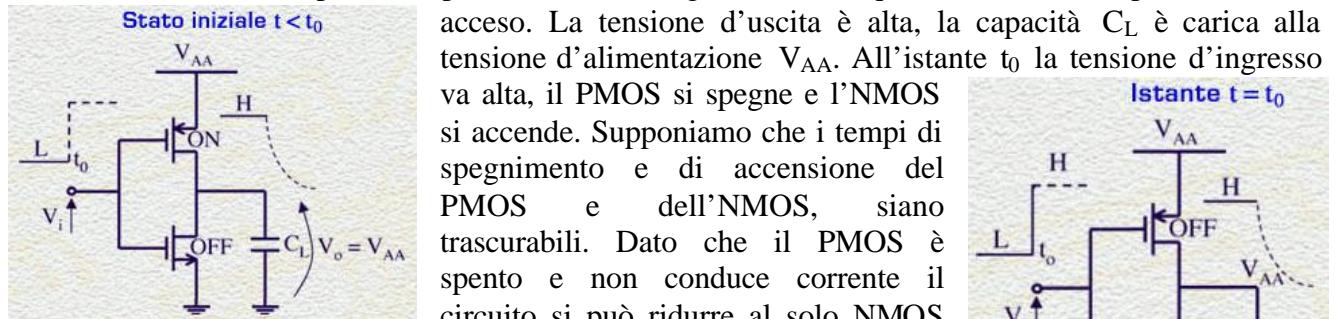
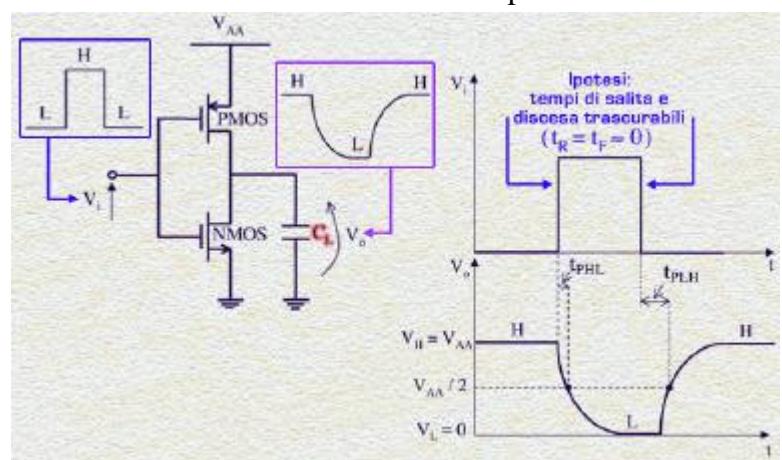
Valutiamo dapprima il tempo  $t_{PHL}$ , ovvero il tempo di propagazione da alto a basso.

Consideriamo quindi la transizione in ingresso da basso ad alto indicando con  $t_0$  l'istante in cui avviene.

La situazione statica di partenza, per  $t < t_0$ , è con ingresso basso, quindi con NMOS spento e PMOS acceso. La tensione d'uscita è alta, la capacità  $C_L$  è carica alla tensione d'alimentazione  $V_{AA}$ . All'istante  $t_0$  la tensione d'ingresso va alta, il PMOS si spegne e l'NMOS si accende. Supponiamo che i tempi di spegnimento e di accensione del PMOS e dell'NMOS, siano trascurabili. Dato che il PMOS è spento e non conduce corrente il circuito si può ridurre al solo NMOS con la capacità  $C_L$ .

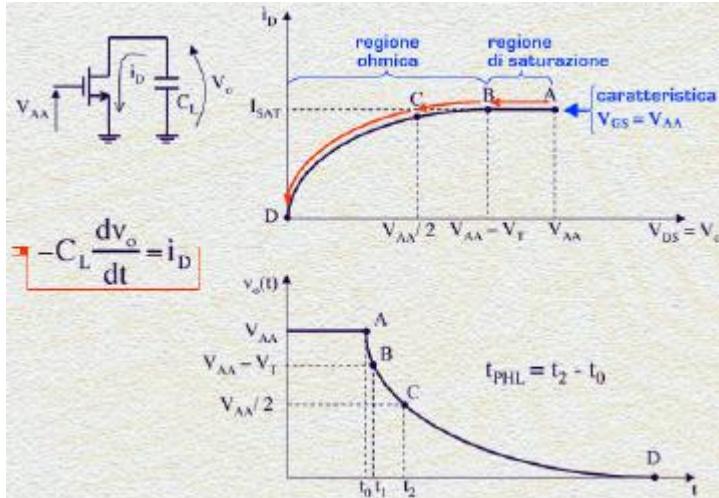
La differenza di potenziale tra drain e source è pari alla tensione d'uscita  $V_o$  che, all'istante  $t = t_0$ , è uguale a  $V_{AA}$ . Dato che l'NMOS è acceso, ovvero presenta un canale conduttivo tra drain e source, vi sarà una corrente  $i_D$  di canale che scorre dal drain verso il source, sottraendo cariche positive al nodo d'uscita ovvero alla capacità  $C_L$  che quindi inizierà a scaricarsi. La scarica della capacità si concluderà solo all'annullarsi della corrente  $i_D$  che avverrà quando la tensione  $V_o$  tra drain e source dell'NMOS, si annullerà. Per determinare il tempo di propagazione sarebbe sufficiente conoscere l'andamento nel tempo della tensione  $V_o$ , ovvero la funzione  $V_o(t)$ . Quindi si può ricavare l'istante in cui  $V_o(t)$  assume il valore  $V_{AA}/2$ , da cui si ottiene il tempo di propagazione. La funzione  $V_o(t)$  si deduce dall'equazione che lega la tensione ai capi del condensatore  $C_L$  con la corrente  $i_D$  che scorre nel condensatore stesso, ovvero  $C_L(dV_o/dt) = -i_D$ .

Si noti il segno meno a secondo membro dovuto al verso della corrente  $i_D$ : infatti se  $i_D$  è positiva il condensatore si scarica, la tensione  $V_o$  decresce nel tempo e quindi la derivata di  $V_o$  è negativa.



Si noti il segno meno a secondo membro dovuto al verso della corrente  $i_D$ : infatti se  $i_D$  è positiva il condensatore si scarica, la tensione  $V_o$  decresce nel tempo e quindi la derivata di  $V_o$  è negativa.

Per risolvere quest'equazione è necessario conoscere la corrente  $i_D$  che scorre nel MOSFET. La corrente del MOSFET dipende dalla tensione gate – source e dalla tensione gate – drain. Notiamo che la tensione gate – source è costante e pari a  $V_{AA}$ , la tensione drain – source invece varia passando dal valore iniziale  $V_{AA}$ , quando  $C_L$  è carica, al valore 0 quando, a transitorio esaurito,  $C_L$  è stata completamente scaricata. La corrente  $i_D$  del MOSFET può quindi essere rappresentata nel piano corrente di drain – tensione drain-source disegnando la caratteristica corrispondente alla tensione di gate  $V_{GS} = V_{AA}$ . In ogni istante di tempo del transitorio, la relazione tra  $i_D$  e la tensione d'uscita  $V_o$  è fissata dalla caratteristica del MOSFET.



quella ohmica. Da questo istante in poi il MOSFET opera in regione ohmica e la corrente  $i_D$  non è più costante ma decresce man mano che  $V_o$  si riduce. Al tempo  $t_2$  la tensione d'uscita ha raggiunto il valore  $V_{AA}/2$  corrispondente al punto di lavoro C sulla caratteristica. Il transitorio continua con il punto che si muove verso l'origine degli assi dove corrente e tensione si annullano ed il transitorio della tensione d'uscita ha termine.

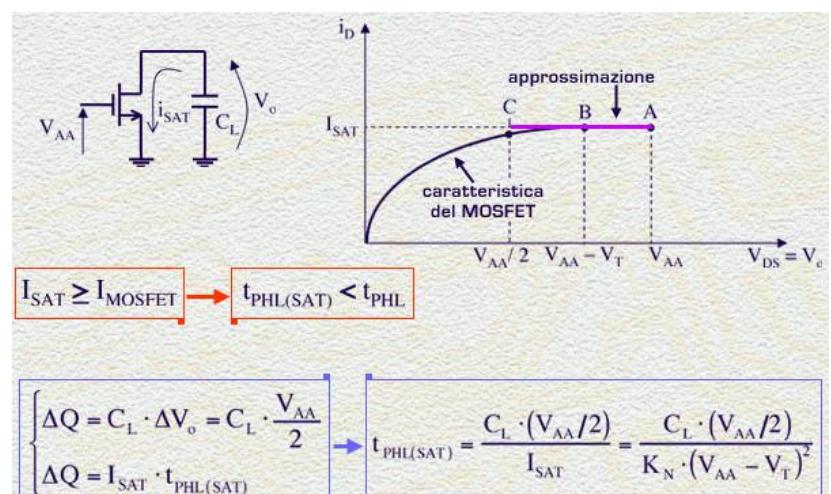
Il tempo di propagazione  $t_{PHL}$  è il tempo  $t_2 - t_0$  necessario alla tensione d'uscita per passare da  $V_{AA}$  a  $V_{AA}/2$ , ovvero, sulla caratteristica del MOSFET il tempo per passare dal punto A al punto C. Questo tempo può essere stimato velocemente con l'ipotesi semplificativa che ora vediamo, successivamente affronteremo il calcolo esatto.

L'ipotesi semplificativa consiste nell'approssimare il tratto di caratteristica del MOSFET tra i punti A e C con un tratto orizzontale corrispondente alla corrente di saturazione del MOSFET. Così facendo si sovrasta la corrente di scarica della capacità nel tratto B – C, ottenendo perciò una stima per difetto del tempo di propagazione.

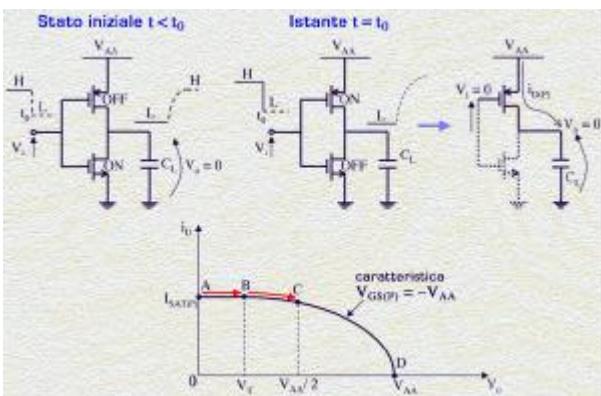
Indicheremo questa stima con  $t_{PHL(SAT)}$ , il pedice SAT ci ricorda che la stima è fatta supponendo che la corrente del MOSFET sia sempre pari al valore di saturazione  $I_{SAT}$ . Il calcolo di  $t_{PHL(SAT)}$  è immediato, si tratta infatti della scarica di una capacità a corrente costante. La tensione ai capi di  $C_L$  varia da  $V_{AA}$  a  $V_{AA}/2$  quindi la carica tolta dalla capacità è pari a  $C_L \cdot (V_{AA}/2)$ . Tale carica viene tolta attraverso una corrente costante  $I_{SAT}$  in un tempo

pari a  $t_{PHL(SAT)}$ . Uguagliando queste due cariche si ricava la stima del tempo di propagazione. Sostituendo a  $I_{SAT}$  l'espressione  $K_N (V_{AA} - V_T)^2$  otteniamo l'equazione finale.

Un discorso del tutto analogo può essere fatto anche per il transitorio d'uscita da basso ad alto. In tal caso è il MOSFET N che si spegne e il MOSFET P che si accende. Il problema si riconduce quindi alla carica della capacità  $C_L$  con la corrente fornita dal PMOS. Anche in questo caso possiamo rappresentare l'andamento della corrente di carica sul piano corrente di drain – tensione d'uscita. Notiamo che il PMOS opera sulla caratteristica a  $V_{GATE-SOURCE}$  costante pari a  $(-V_{AA})$ . Nel piano  $i_D -$



Vo tale caratteristica si sviluppa come mostrato in figura. Infatti quando la tensione d'uscita è pari a  $V_{AA}$ , la tensione drain – source del MOSFET è nulla quindi il punto  $V_0 = V_{AA}$  è l'origine della caratteristica. Per  $V_0 = 0$  Volt la  $V_{DRAIN-SOURCE}$  del MOSFET è pari a  $V_{AA}$ . A  $V_0$  pari alla tensione di soglia del MOSFET si ha il passaggio tra regione di saturazione e regione ohmica. Come per il transitorio precedente, il punto di lavoro del circuito si muove sulla caratteristica del PMOS partendo dal punto A, passando per B e C e raggiungendo il punto finale D



in cui la tensione d'uscita è pari a quella dell'alimentazione, la corrente del MOSFET si annulla e quindi la carica della capacità ha termine.

La stima del tempo di propagazione da basso ad alto può essere velocemente fatta come in precedenza, supponendo costante e pari a  $I_{SAT}$  la corrente del PMOS. Otteniamo quindi un'espressione per  $t_{PLH(SAT)}$  identica a quella trovata per il transitorio precedente, l'unica

differenza sta nel fattore  $K_p$  del MOSFET P al posto del  $K_n$  che era relativo al MOSFET N.

Consideriamo quindi le due espressioni trovate la prima per il tempo di propagazione da alto a basso,

$$\left\{ \begin{array}{l} \text{transitorio L} \rightarrow H \\ (\text{carica di } C_L \text{ con PMOS}) \end{array} \right. \quad t_{PLH(SAT)} = \frac{C_L \cdot (V_{AA}/2)}{K_p \cdot (V_{AA} - V_T)^2}$$

$$\left\{ \begin{array}{l} \text{transitorio H} \rightarrow L \\ (\text{scarica di } C_L \text{ con NMOS}) \end{array} \right. \quad t_{PHL(SAT)} = \frac{C_L \cdot (V_{AA}/2)}{K_n \cdot (V_{AA} - V_T)^2}$$

dei tempi di propagazione reali, valutati considerando la caratteristica corrente tensione reale del MOSFET. Vedremo comunque che l'espressione esatta del tempo di propagazione differisce solo per

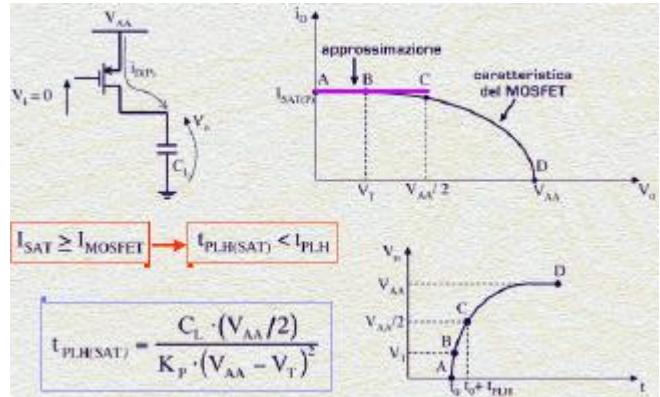
un fattore moltiplicativo, circa costante. Iniziamo quindi a discutere queste due espressioni. Esse differiscono solo per il fattore  $K$  dei MOSFETS, quindi, se vogliamo rendere i due tempi di propagazione uguali è

sufficiente porre uguali  $K_n$  e  $K_p$ . Ricordando le espressioni di  $K_n$  e  $K_p$ , l'uguaglianza tra questi si traduce in una relazione tra i rapporti di forma  $W/L$  del PMOS e dell'NMOS. Nello specifico è sufficiente che il PMOS abbia un  $W/L$  circa 3 volte maggiore di quello dell'NMOS per rendere uguali i due tempi.

**Questa relazione ci dice che, a parità di lunghezza di gate L, il PMOS dev'essere circa 3 volte più largo rispetto all'NMOS, per avere lo stesso K e quindi condurre la stessa corrente a parità di tensione gate-source in modulo. Il fatto che il PMOS debba essere 3 volte più largo per portare la stessa corrente è dovuta alla 3 volte più bassa mobilità delle lacune rispetto a quella degli elettroni.**

Analizziamo ora l'espressione del tempo di propagazione. Innanzitutto osserviamo che  $t_p$  è dato dal prodotto di due fattori di cui il primo,  $C_L/K$ , non dipende dalle tensioni mentre il secondo fattore dipende solo dalla tensione di alimentazione del circuito e dalla tensione di soglia del MOSFET.

Iniziamo ad analizzare l'espressione concentrandoci sul primo fattore ( $C_L/K$ ). Questo primo fattore ci dice che il tempo di propagazione è proporzionale alla capacità  $C_L$  associata al nodo d'uscita



relativo quindi alla scarica della capacità  $C_L$  da parte dell'NMOS; la seconda espressione per il tempo di propagazione da basso ad alto, relativo alla carica di  $C_L$  da parte del PMOS. Queste due equazioni forniscono solo delle **stime in difetto**

$$K_n = K_p = K \rightarrow t_{P(SAT)} = \frac{C_L \cdot (V_{AA}/2)}{K \cdot (V_{AA} - V_T)^2}$$

$$\left\{ \begin{array}{l} K_n = \frac{1}{2} \mu_n C_{ox} \left( \frac{W}{L} \right)_n \\ K_p = \frac{1}{2} \mu_p C_{ox} \left( \frac{W}{L} \right)_p \end{array} \right. \quad \left\{ \begin{array}{l} K_n = K_p \\ \left( \frac{W}{L} \right)_p = \left( \frac{\mu_n}{\mu_p} \right) \cdot \left( \frac{W}{L} \right)_n \Rightarrow \left( \frac{W}{L} \right)_p \approx 3 \cdot \left( \frac{W}{L} \right)_n \end{array} \right.$$

dell'invertitore. Ciò è comprensibile pensando che il tempo necessario a caricare o scaricare  $C_L$  è

$$t_{P(SAT)} = \left( \frac{C_L}{K} \right) \left[ \frac{V_{AA}/2}{(V_{AA} - V_T)^2} \right]$$

t<sub>P</sub>  $\propto$  C<sub>L</sub> perché t<sub>scarica</sub>  $\propto$  Q  $\propto$  C<sub>L</sub>  
 il tempo di scarica è proporzionale alla carica  
 la carica è proporzionale a C<sub>L</sub>  
 t<sub>P</sub>  $\propto$  1/K perché t<sub>scarica</sub>  $\propto$  1/I<sub>media (MOSFET)</sub>  
 e I<sub>SAT</sub>  $\propto$  K I<sub>OHM</sub>  $\propto$  K

proporzionale al fattore K. Possiamo comprendere questa dipendenza ricordando che la corrente del MOSFET che scarica/carica la capacità è proporzionale a K. Più è alto K maggiore sarà la corrente media e quindi minore sarà il tempo di carica o scarica di C<sub>L</sub>. Ricordiamo a tale proposito

l'espressione di K:  $K = \frac{1}{2} \mu C_{ox} \left( \frac{W}{L} \right)$  quindi il tempo di propagazione, a parità di C<sub>L</sub>, si riduce al

crescere del prodotto  $\mu C_{ox}$  caratteristico della tecnologia con la quale il MOSFET è realizzato, e si riduce anche al crescere del rapporto di forma, W/L, del MOSFET stesso. Infine possiamo osservare la dipendenza di t<sub>P</sub> dalla tensione di alimentazione del circuito. Notiamo che il tempo di propagazione è, in prima approssimazione, inversamente proporzionale alla tensione di alimentazione. Questa dipendenza è particolarmente importante per la progettazione di sistemi portatili a bassa tensione.

$$K = \frac{1}{2} \mu C_{ox} \left( \frac{W}{L} \right) \rightarrow \begin{cases} \mu C_{ox} \uparrow \rightarrow t_P \downarrow & \text{dipende dalla tecnologia del MOSFET} \\ (W/L) \uparrow \rightarrow t_P \downarrow & \text{dipende dalla geometria del MOSFET} \end{cases}$$

$t_P \propto 1/V_{AA}$   $\rightarrow V_{AA} \downarrow \rightarrow t_P \uparrow$  inversamente proporzionale alla tensione di alimentazione. Questa dipendenza è particolarmente importante per la progettazione di sistemi portatili a bassa tensione.

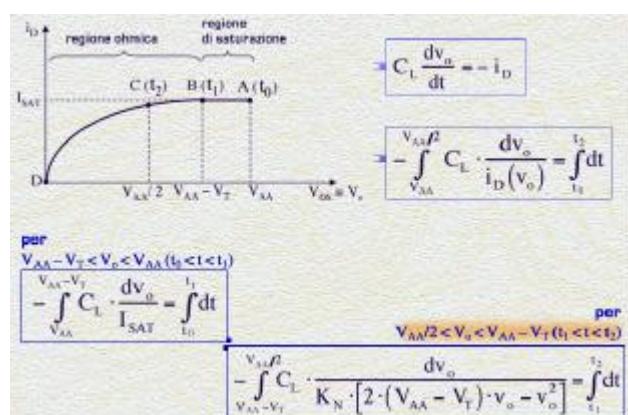
### Analisi completa

Vedremo ora come calcolare esattamente il tempo di propagazione. Consideriamo la transizione da alto a basso in cui la capacità d'uscita viene scaricata dall'NMOS. Sulla caratteristica del MOSFET indichiamo i punti A, B e C con i corrispondenti istanti di tempo t<sub>0</sub>, t<sub>1</sub>, t<sub>2</sub> della transizione.

Il tempo di propagazione alto – basso t<sub>PHL</sub> pari a t<sub>2</sub> – t<sub>0</sub> si può determinare risolvendo l'equazione

differenziale  $C_L \frac{dv_o}{dt} = -i_D$  essendo i<sub>D</sub> una

funzione nota della variabile v<sub>o</sub>, che è uguale alla tensione drain – source del MOSFET, è possibile risolvere l'equazione separando le variabili ed integrando. La corrente i<sub>D</sub> assume due espressioni diverse a seconda che il MOSFET operi in saturazione o in ohmica quindi è necessario suddividere l'integrazione nei due intervalli di tempo e di tensione v<sub>o</sub> corrispondenti. Il primo intervallo vede v<sub>o</sub> variare da V<sub>AA</sub> a V<sub>AA</sub> – V<sub>T</sub> e la



$V_{AA} - V_T < v_o < V_{AA}$  ( $t_0 < t < t_1$ ) N-MOSFET in regione di saturazione

$$-\int_{V_{AA}}^{V_{AA}-V_T} C_L \cdot \frac{dv_o}{I_{SAT}} = \int_{t_0}^{t_1} dt$$

quindi  $t_1 - t_0 = \frac{C_L \cdot V_T}{I_{SAT}}$  con  $\begin{cases} I_{SAT} = K_N \cdot (V_{GS(N)} - V_T)^2 \\ V_{GS(N)} = V_{AA} \end{cases}$

$$t_1 - t_0 = \frac{C_L \cdot V_T}{K_N (V_{AA} - V_T)^2}$$

corrente i<sub>D</sub> è costante pari a I<sub>SAT</sub>, nel secondo intervallo v<sub>o</sub> varia da V<sub>AA</sub> – V<sub>T</sub> a V<sub>AA</sub>/2 e la corrente i<sub>D</sub> assume l'espressione riportata, relativa al MOSFET operante in regione ohmica. Il primo integrale si risolve immediatamente giungendo al risultato mostrato a lato, che è del tutto simile a quello trovato nella trattazione semplificata, tranne che per la variazione della tensione d'uscita che ora è pari a V<sub>T</sub>.

Il secondo integrale, relativa alla parte di transitorio durante la quale il MOSFET è in regione ohmica, ha per soluzione quella mostrata nella figura seguente. Il tempo di propagazione si

$$V_{AA}/2 < V_o < V_{AA} - V_T \quad (t_1 < t < t_2)$$

N-MOSFET  
in regione  
ohmica

$$-\frac{C_L}{K_N} \int_{V_{AA}-V_T}^{V_{AA}/2} \frac{dv_o}{[2 \cdot (V_{AA} - V_T) \cdot v_o - v_o^2]} = t_2 - t_1$$

$$t_2 - t_1 = \left( \frac{C_L}{K_N} \right) \cdot \left[ \frac{1}{2 \cdot (V_{AA} - V_T)} \cdot \ln \left( \frac{3V_{AA} - 4V_T}{V_{AA}} \right) \right]$$

molto diversa da quella trovata nella trattazione semplificata. Consideriamo il fattore in parentesi quadra. Notiamo che questo dipende solo dalla tensione di alimentazione del circuito e dalla tensione di soglia del MOSFET. Per analizzare un'espressione di questo tipo e trarne delle informazioni utili è sufficiente inserirla in un calcolatore e tracciare alcuni grafici scegliendo ad esempio  $V_{AA}$  come variabile e  $V_T$  come parametro.

Un'altra possibilità, che è quella che seguiremo, è di elaborare l'equazione matematicamente cercando di esprimere in una forma più compatta e semplice e che soprattutto, ci mostri un legame con la soluzione approssimata. Raccogliendo alcuni termini e introducendo il parametro  $a$  definito come rapporto tra la tensione di alimentazione e la tensione di soglia, otteniamo l'equazione mostrata. Possiamo osservare che il primo fattore, in parentesi graffe, è nient'altro che il valore che il tempo di propagazione assumerebbe se il MOSFET operasse in zona di saturazione per tutto il tempo della transizione da  $V_{AA}$  a  $V_{AA}/2$ . Questo è proprio il tempo  $t_{P(SAT)}$  che avevamo ricavato nella trattazione semplificata. Il secondo fattore  $b$  può essere quindi interpretato come fattore correttivo che tiene conto

$$t_p = \left\{ \frac{C_L \cdot (V_{AA}/2)}{K \cdot (V_{AA} - V_T)^2} \right\} \cdot \beta = t_{P(SAT)} \cdot \beta$$

**tempo di propagazione  
nell'approssimazione di  
MOSFET saturo**

**fattore correttivo**

$$\beta = \frac{2}{\alpha} + \frac{(\alpha - 1)}{\alpha} \cdot \ln \left( \frac{3\alpha - 4}{\alpha} \right)$$

$$1 \leq \beta \leq 1.1$$

$$t_p \approx t_{P(SAT)}$$

del fatto che il MOSFET opera, per un certo periodo, anche in regione ohmica. Il fattore  $b$  dipende solo da  $a$ , rapporto tra  $V_{AA}$  e  $V_T$ .  $a$  può variare da un valore minimo di 2, quando la tensione di

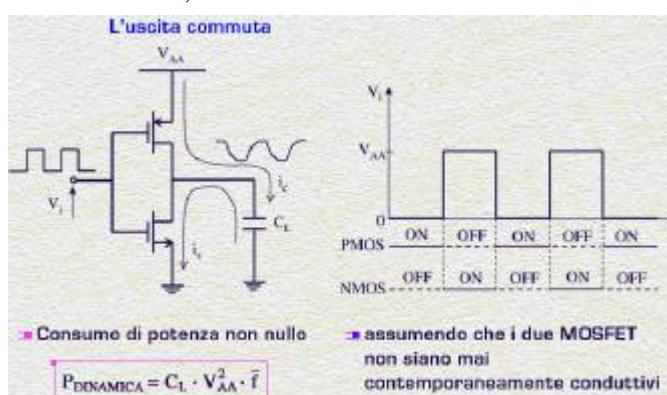
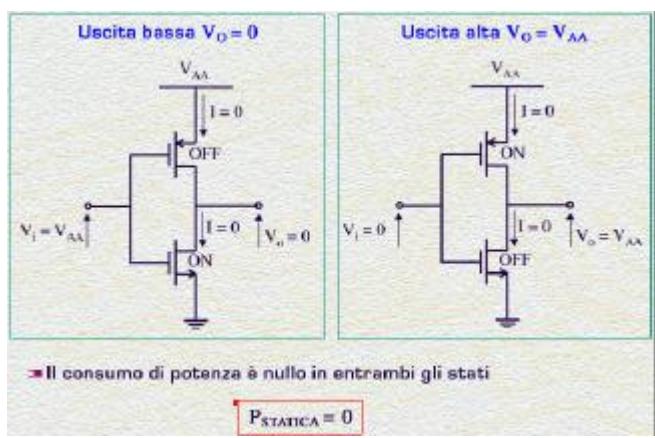
alimentazione è due volte quella di soglia, ad un valore massimo inferiore a 10. Per  $a = 2$  il fattore  $b$  è uguale a 1 perché durante i transistori di propagazione i MOSFET sono sempre in zona di saturazione. Il fattore  $b$  è una funzione monotona crescente di  $a$  tendendo al valore asintotico di  $\ln 3$ , pari a 1.1. Si verifica quindi che  $b$  è compreso al massimo tra 1 e 1.1, quindi il tempo di propagazione è ottimamente approssimato, con un errore inferiore al 10% da  $t_{P(SAT)}$ , valutato tramite l'analisi semplificata. Infine possiamo valutare il prodotto ritardo – potenza dell'invertitore CMOS. Scriviamo quindi l'espressione del tempo di propagazione, quindi della potenza dinamica impiegata da un invertitore che opera alla frequenza media di commutazione  $f$ . Eseguiamo poi il prodotto. Un invertitore è tanto più prossimo all'ideale quanto minore è il suo tempo di propagazione e tanto più basso è il suo consumo di potenza. Quindi un invertitore, o una porta logica in generale, è tanto

Tempo di propagazione	$t_p = \frac{C_L \cdot (V_{AA}/2)}{K \cdot (V_{AA} - V_T)^2}$
Potenza dinamica	$P = C_L \cdot V_{AA}^2 \cdot f$
Prodotto ritardo x potenza	$t_p \cdot P = \frac{C_L^2 \cdot V_{AA}^3 \cdot f}{2K \cdot (V_{AA} - V_T)^2}$
$\begin{cases} t_p P + 1/K \\ t_p P + C_L^2 \\ t_p P + V_{AA} \end{cases}$	

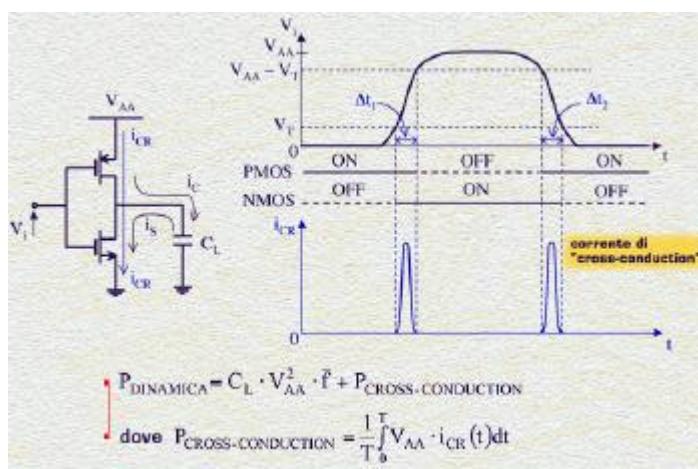
migliore quanto più basso è il suo prodotto ritardo × potenza. Nel caso dell'invertitore CMOS osserviamo che tale prodotto diminuisce al crescere del fattore K dei MOSFET's, decresce inoltre quadraticamente con la capacità  $C_L$  e si riduce circa linearmente con la tensione di alimentazione. Quindi una riduzione della capacità, della tensione di alimentazione ed un aumento del K dei MOSFET portano ad una riduzione del prodotto ritardo – potenza.

Analizzeremo ora un problema connesso alla commutazione di un invertitore CMOS. In una precedente lezione avevamo mostrato che un invertitore analogo a quello CMOS e costituito da due interruttori ha un consumo di potenza nullo quando l'uscita NON commuta, ovvero in condizioni statiche, perché quando l'NMOS è conduttivo (uscita bassa) il PMOS non lo è e quando è il PMOS ad esser conduttivo (uscita alta) l'NMOS è spento. Quindi la corrente nei due MOSFET è in condizioni statiche nulla di conseguenza anche il consumo di potenza.

Diversamente, se l'invertitore commuta ad una frequenza media  $f$  (soprassegnato) allora si ha un consumo di potenza pari a  $C_L V_{AA}^2 f$  originato dalle correnti di carica e scarica della capacità equivalente  $C_L$  del nodo d'uscita. Avevamo chiamato questo consumo di potenza, dinamico, per evidenziare che si ha in condizioni dinamiche e non statiche.



istantaneamente da basso ad alto e viceversa ed i MOSFET rispondono istantaneamente al segnale d'ingresso. Cosicché si vede che non c'è nessun istante in cui entrambi i MOSFET siano accesi. Consideriamo ora una situazione più realistica in cui l'ingresso non commuti istantaneamente tra i due livelli logici, ma che la transizione avvenga in un intervallo di tempo finito. Questa situazione è



intervallo  $\Delta t_1$ , nella transizione da basso a alto in cui si verifica la stessa conduzione di corrente. Questa corrente, indicata con  $i_{CR}$  in figura, viene detta di cross – conduction e comporta un consumo di potenza aggiuntivo oltre a quello già valutato e dovuto alle correnti  $i_C$  e  $i_S$  di carica e scarica della capacità equivalente d'uscita. La potenza consumata a seguito della conduzione contemporanea dei due MOSFET negli intervalli  $\Delta t_1$  e  $\Delta t_2$  può essere calcolata conoscendo l'andamento della corrente  $i_{CR}$  nel tempo. Si può comunque mostrare che la potenza di cross – conduction risulta inferiore rispetto a quella derivata dalla carica e scarica della capacità  $C_L$ .

L'espressione del consumo dinamico di potenza è stata ricavata supponendo che i due MOSFET non siano mai contemporaneamente conduttori durante le transizioni. Questo è vero perché si è assunto che il segnale d'ingresso commuti

realistica pensando che il segnale d'ingresso proviene da circuiti con propri tempi di commutazione. Osserviamo la transizione della tensione d'ingresso da basso ad alto: quando  $V_i$  supera la tensione di soglia  $V_T$  dell'NMOS, questo si accende mentre il PMOS si spegne quando viene superato il valore  $V_{AA} - V_T$ . Notiamo quindi che esiste un intervallo di tempo  $\Delta t_1$  durante il quale entrambi i MOSFET sono conduttori. Ciò comporta un passaggio diretto di corrente dall'alimentazione  $V_{AA}$  verso massa. Analogamente possiamo osservare un

<pagina bianca per data-sheet>

## Lezione n. 19 : Porte logiche a più ingressi

### Porte NAND e NOR

Finora abbiamo visto e studiato le caratteristiche statiche e dinamiche dell'invertitore logico, ora analizzeremo porte logiche a più ingressi. Come esempio introduttivo consideriamo il circuito mostrato in figura, avente due ingressi logici A e B. Determiniamo la funzione logica del circuito. Per  $A = 0$  e  $B = 0$  i due MOSFET sono entrambi non conduttori, quindi nel resistore R non scorre corrente e la tensione d'uscita sarà uguale alla tensione di alimentazione  $V_{AA}$ . L'uscita logica è quindi 1. Se A oppure B oppure sia A che B sono alti allora uno o tutte due i MOSFET saranno conduttori, in R scorrerà corrente e l'uscita sarà bassa. La tavola della verità è quella mostrata da cui si riconosce che la porta realizza la funzione NOR.

Analogamente si può mostrare che il circuito mostrato in figura con i due MOSFET in serie realizza la funzione NAND, in quanto l'uscita è pari a 0 se e solo se entrambi i MOSFET sono conduttori, ovvero quando  $A = B = 1$ .

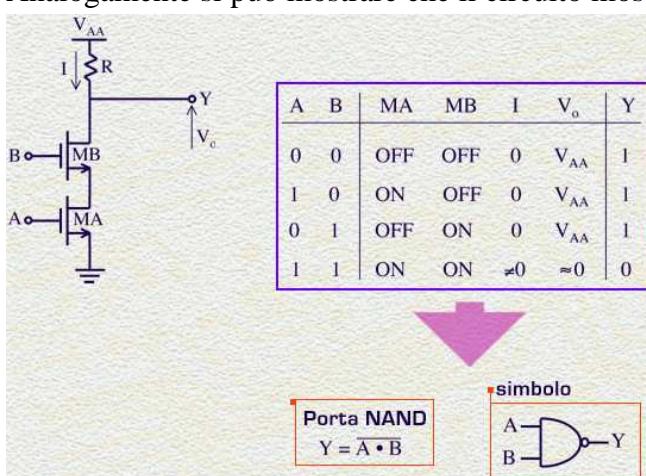
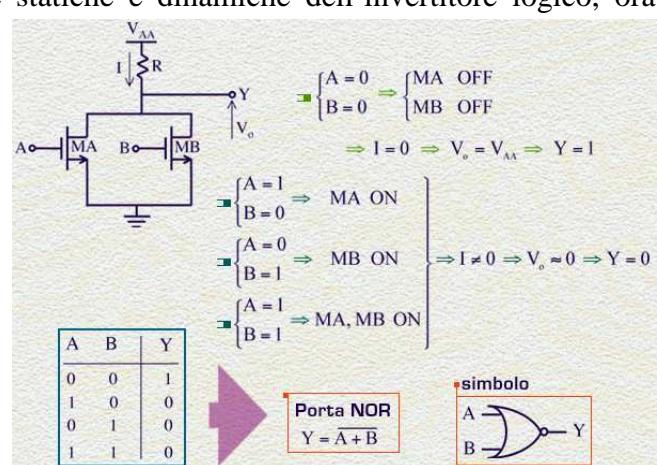
**Entrambi i circuiti possono essere realizzati con un numero di ingressi in linea teorica arbitrario, ponendo in parallelo o in serie un MOSFET per ogni ingresso.** I circuiti con la resistenza presentano tuttavia l'inconveniente di **consumare potenza in condizioni statiche** ogni volta che l'uscita è bassa.

Per annullare il consumo statico di potenza è necessario realizzare queste porte impiegando configurazioni a MOSFET complementari, così come abbiamo visto per l'invertitore CMOS.

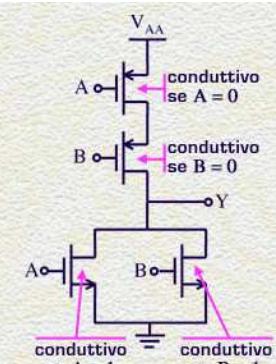
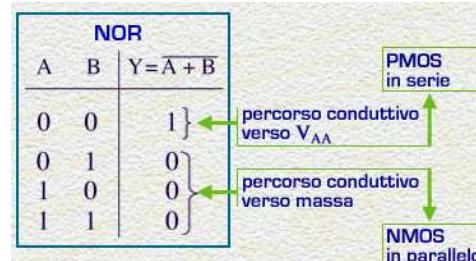
Vediamo come sintetizzare una porta CMOS NOR a due ingressi. A tale proposito ricordiamo, riferendoci all'invertitore, che un MOSFET P conduce quando al gate è presente la tensione 0 volt corrispondente al livello logico 0.

Consideriamo quindi la tavola della verità di una porta NOR a due ingressi, osserviamo che l'uscita è pari all'uno logico, ovvero alla tensione di alimentazione, solo quando entrambi gli ingressi sono a livello 0. Ciò significa che vi saranno due PMOS posti in serie che collegano l'uscita con l'alimentazione. Analogamente, lo zero logico in uscita si ottiene quando

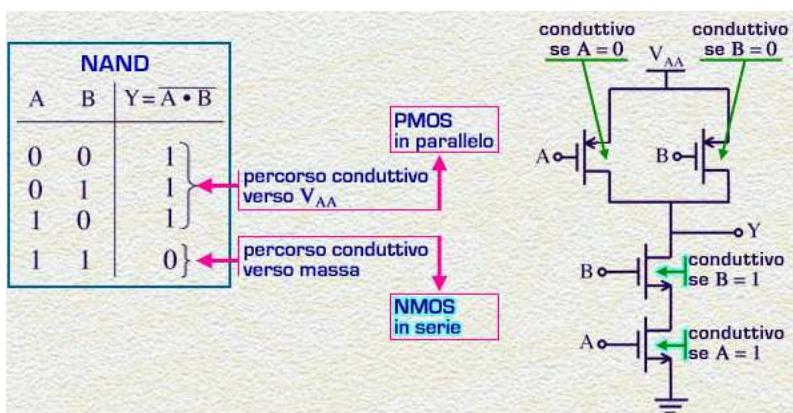
vi è un percorso conduttivo tra l'uscita e massa, ottenibile tramite NMOS. I due NMOS saranno in parallelo in modo che quando almeno uno dei due gate è a livello alto, quindi quando A oppure B sono 1, l'uscita sarà collegata verso massa tramite il canale conduttivo del MOSFET o dei MOSFET con ingresso alto. Riguardo al consumo di potenza in condizioni statiche possiamo osservare che, dato che l'accensione di un NMOS comporta il contemporaneo spegnimento del PMOS avente lo stesso ingresso, non si avrà mai, in condizioni statiche, un percorso conduttivo tra l'alimentazione e massa, quindi la potenza statica sarà nulla.



NOR		
A	B	$Y = \bar{A} + \bar{B}$
0	0	1
0	1	0
1	0	0
1	1	0



Con la stessa metodologia impiegata per la sintesi della porta NOR possiamo ricavare il circuito della porta NAND in tecnologia CMOS.



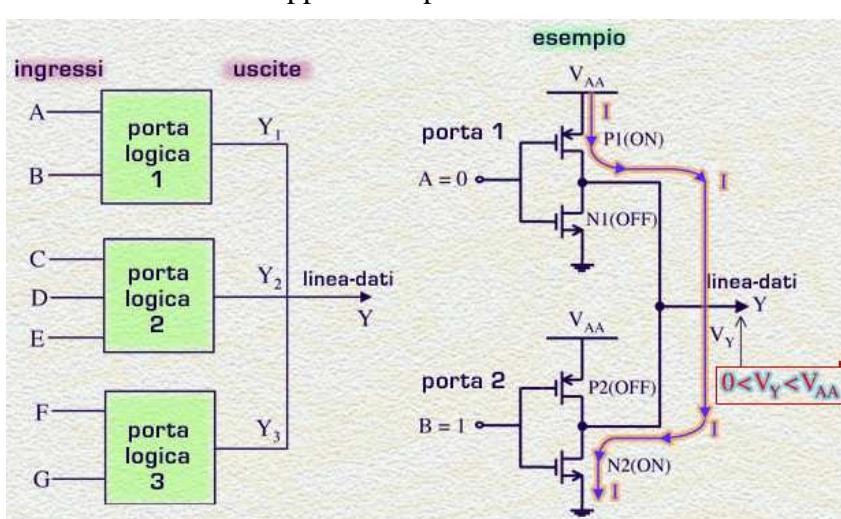
Osservando la tavola della verità della NAND vediamo che l'uscita è alta quando almeno uno degli ingressi è basso, quindi i due PMOS sono in parallelo.

Diversamente gli NMOS sono in serie perché l'uscita dev'essere collegata a massa solo se entrambi gli ingressi sono alti, ovvero quando i due NMOS sono entrambi conduttori.

### Invertitore tri-state

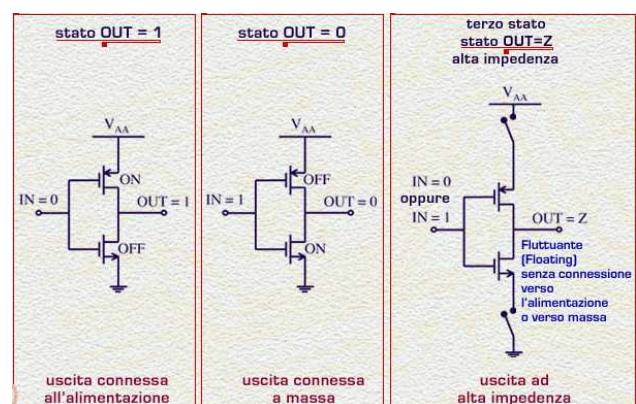
Consideriamo ora un problema legato alla coesistenza di più porte logiche entro un sistema digitale. Consideriamo tre porte logiche le cui uscite devono condividere un'unica linea-dati che può essere l'uscita del sistema oppure semplicemente la connessione con altre parti del sistema. Si può subito

intuire che una connessione delle uscite siffatta può generare un conflitto nel caso in cui una porta logica dovesse fornire in uscita un 1 logico ed un'altra porta fornire lo 0 logico. Quale bit sarebbe presente sulla linea dati? Consideriamo l'esempio mostrato di due invertitori con le uscite connesse ad un'unica linea Y. Supponiamo che l'ingresso dell'invertitore 1 sia  $A=0$  e l'ingresso dell'invertitore 2 sia  $B=1$ . Nel caso le uscite  $Y_1$  e  $Y_2$  fossero scollegate si avrebbe  $Y_1=1$



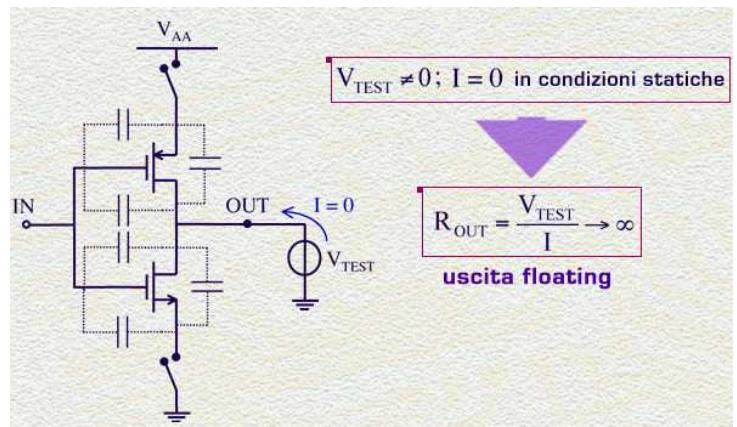
e  $Y_2=0$ . Nel caso di uscite collegate insieme osserviamo che, essendo conduttori il MOSFET P1 della prima porta e il MOSFET N2 della seconda, è presente un cammino conduttivo dall'alimentazione verso massa. Si verifica quindi un passaggio di corrente in condizioni statiche. L'uscita Y si porrà ad un valore di tensione intermedio tra la tensione d'alimentazione  $V_{AA}$  e il potenziale di massa, a seconda di quale MOSFET, P1 o N2 è più conduttivo. L'unica condizione di compatibilità di una connessione di questo tipo si ha quando tutte due gli invertitori forniscono lo stesso bit in uscita.

Vediamo ora come risolvere questo conflitto pur mantenendo un'unica linea-dati su cui far convergere le uscite di diverse porte. A tal fine notiamo che il conflitto sorge perché le porte logiche che abbiamo finora esaminato, hanno sempre l'uscita corrispondente ad uno stato logico, o zero o uno. Questo accade perché la topologia circuitale implica che l'uscita sia sempre collegata, tramite un MOSFET, o verso massa o verso l'alimentazione. Ma allora, per risolvere il conflitto, è sufficiente modificare il circuito dell'invertitore in modo che sia possibile interrompere le sue connessioni verso massa e verso l'alimentazione. Ciò può essere fatto inserendo due interruttori tra i SOURCE dei MOSFET e le linee di alimentazione e massa rispettivamente. Nel caso in cui questi interruttori sono entrambi aperti, il nodo d'uscita non risulta più connesso ad alcun potenziale di riferimento qualunque sia lo stato, conduttivo o meno dei MOSFET. Si dice che l'uscita è fluttuante o **floating**, nel senso che



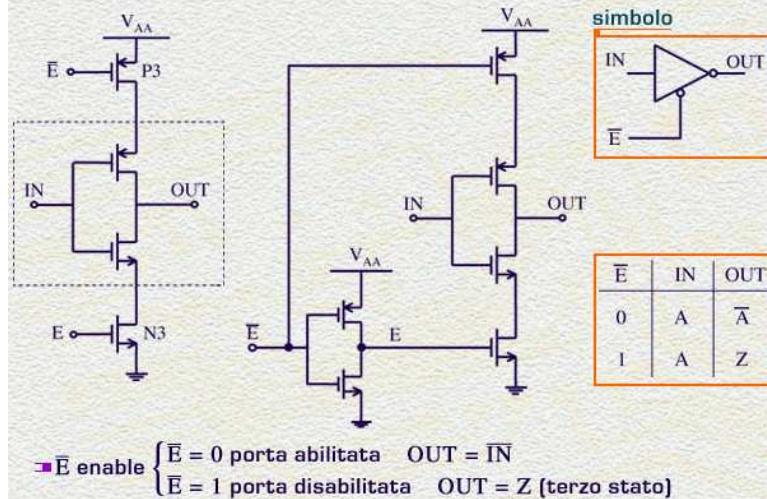
il suo potenziale può assumere qualsiasi valore, in linea teorica, dipendente solo dalla carica elettrostatica presente sul nodo.

Il nodo d'uscita risulta connesso agli altri nodi del circuito solo attraverso delle capacità, quali le capacità tra gate e drain e tra drain e source dei MOSFET. La porta logica si trova quindi in un **terzo stato** caratterizzato dall'avere l'uscita ad alta impedenza. Infatti pensando di applicare in uscita un generatore di tensione continua  $V_{TEST}$ , una volta caricate tutte le capacità associate al nodo d'uscita a transitorio esaurito, la corrente fornita dal generatore è nulla, indicando quindi che l'impedenza d'uscita del circuito, in continua, è infinita.



Vediamo ora l'implementazione circuitale di un invertitore con uscita a tre-stati. La connessione dell'invertitore dalla massa e dall'alimentazione può essere realizzata con due MOSFET's: un MOSFET N per la connessione verso massa ed un MOSFET P per la connessione verso l'alimentazione. Questi due MOSFET devono essere contemporaneamente chiusi o aperti, quindi devono venire comandati sul gate da segnali complementari: E ed E negato. Quando E è alto ( $V_{AA}$  volt) sia l'NMOS N3 che il PMOS P3 sono conduttori, la porta risulta quindi connessa ad alimentazione e massa ed è perciò abilitata ad eseguire la sua funzione logica. Quando E=0 la porta è disabilitata e l'uscita si trova nello stato ad alta impedenza.

In figura vediamo il circuito completo, il simbolo logico e la tavola della verità.

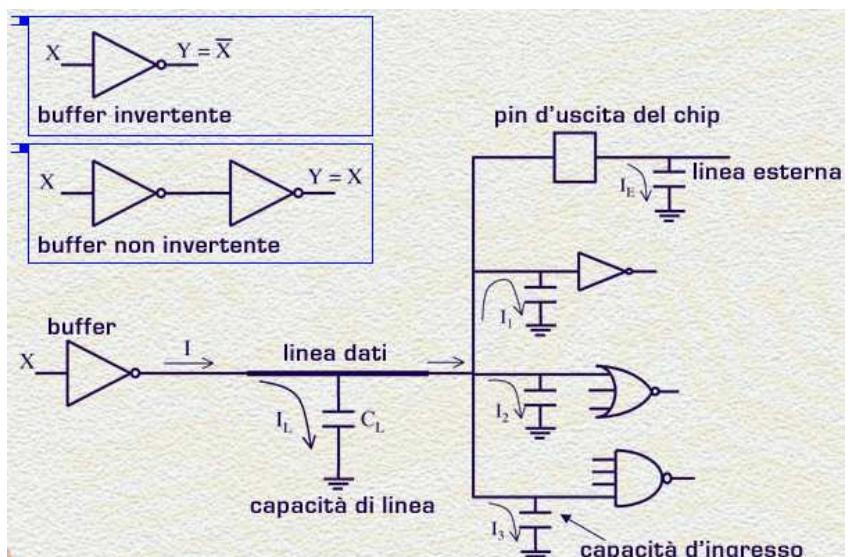


Il simbolo logico mostra l'input IN e l'enable  $\bar{E}$ . La tavola della verità mostra i valori di  $\bar{E}$ , IN e OUT.

La connessione dell'invertitore dalla massa e dall'alimentazione può essere realizzata con due MOSFET's: un MOSFET N per la connessione verso massa ed un MOSFET P per la connessione verso l'alimentazione. Questi due MOSFET devono essere contemporaneamente chiusi o aperti, quindi devono venire comandati sul gate da segnali complementari: E ed E negato. Quando E è alto ( $V_{AA}$  volt) sia l'NMOS N3 che il PMOS P3 sono conduttori, la porta risulta quindi connessa ad alimentazione e massa ed è perciò abilitata ad eseguire la sua funzione logica. Quando E=0 la porta è disabilitata e l'uscita si trova nello stato ad alta impedenza.

## Buffer e line-driver

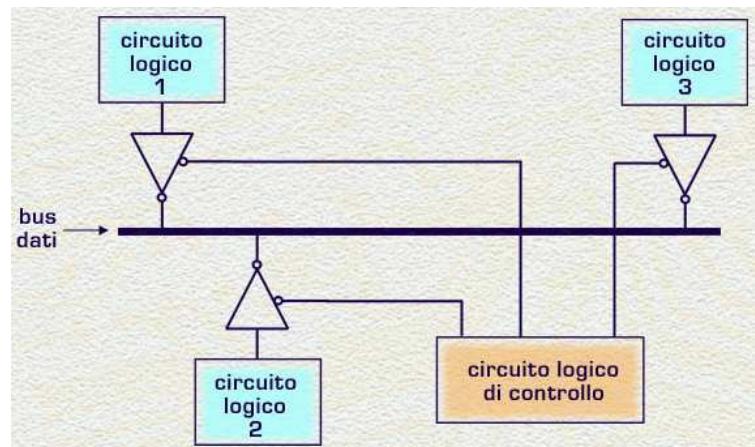
La configurazione tri-state è impiegata soprattutto in porte logiche chiamate buffer o linee-driver, ovvero circuiti costituiti da uno o più invertitori in cascata e progettati appositamente per pilotare bus-dati o linee d'uscita di un chip. Tali linee sono in genere caratterizzate dall'avere capacità verso massa di valore relativamente alto, a causa della loro lunghezza o dimensione fisica o delle numerose connessioni della linea stessa con circuiti che costituiscono carichi capacitivi. I buffer o linee driver sono quindi progettati con transistori d'uscita in grado di erogare forti correnti e quindi caricare o scaricare capacità anche elevate in tempi brevi.



Vediamo alcuni esempi di applicazioni di buffer tri-state.

Vediamo alcuni esempi di applicazione di buffer tri-state. In figura è rappresentato un bus-dati su cui convergono le uscite di 3 circuiti logici, attraverso buffer invertenti tri-state. L'ingresso ENABLE dei buffer è pilotato da un circuito logico di controllo che abilita un solo buffer alla volta, mantenendo tutti gli altri in condizioni di alta impedenza d'uscita. In tal modo il bus può ricevere, in intervalli di tempo diversi, le uscite di più circuiti logici, senza che si generino conflitti.

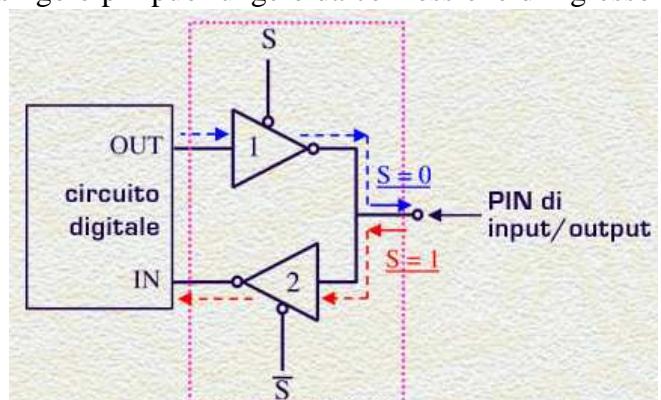
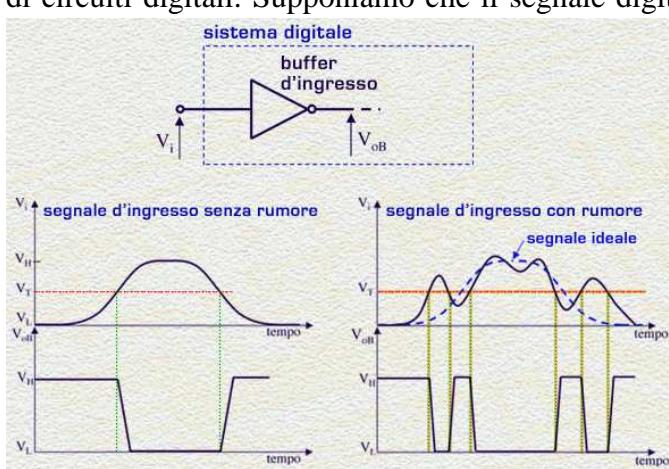
Un altro impiego della configurazione tri-state è nella realizzazione di buffer bidirezionali, ovvero circuiti in grado di regolare la trasmissione di bit in due sensi opposti. In figura vediamo l'implementazione di questo circuito, costituito da due invertitori tri-state i cui ingressi di abilitazione sono comandati da segnali logici complementari  $S$  e  $S$  negato.



realizzando un pin per ognuno di questi. Pertanto un singolo pin può fungere da connessione d'ingresso o connessione d'uscita in tempi diversi mediante l'impiego di due buffer tri-state, come mostrato in figura. Se  $S=0$  sarà abilitato il buffer 1 e disabilitato il 2. Al pin saranno quindi presenti i bit in uscita dal circuito. Viceversa quando  $S$  viene posto alto, il buffer 1 è disabilitato, il 2 è abilitato e l'ingresso del circuito può ricevere dati dal pin.

Il pin si comporta da ingresso o da uscita del circuito a seconda del segnale  $S$  che comunemente viene generato dallo stesso circuito digitale.

Accenneremo ora ad un particolare tipo di buffer comunemente impiegato nelle interfacce d'ingresso di circuiti digitali. Supponiamo che il segnale digitale in ingresso  $V_i$  abbia tempi di transizioni lenti rispetto a quelli propri del buffer. Nel caso ideale, mostrato in figura, in cui il segnale d'ingresso non è soggetto a rumore, il buffer presenterà transizioni in corrispondenza dell'attraversamento della sua soglia logica  $V_T$ . Si può tuttavia verificare comunemente che il segnale d'ingresso sia affetto da rumore, dovuto ad interferenze, come mostrato in figura. In tal caso la soglia viene attraversata più volte ed il buffer avendo tempi di transizione brevi risponde con delle commutazioni che tuttavia sono indesiderate in quanto sono causate dal rumore sovrapposto al segnale.



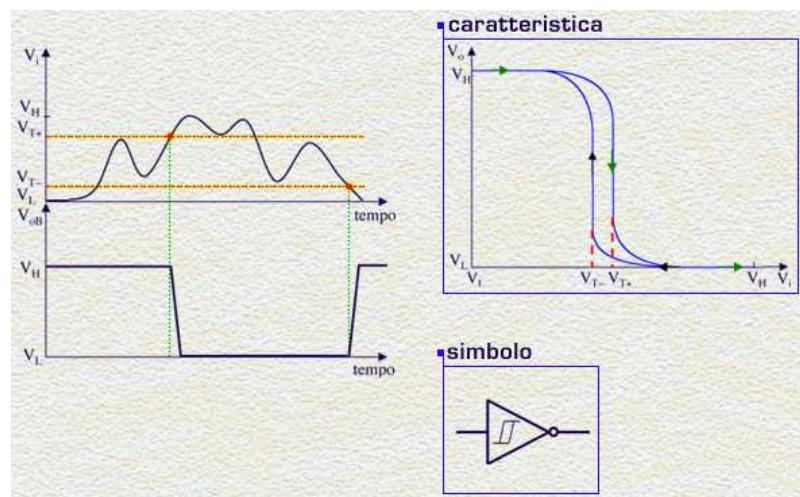
realizzando un pin per ognuno di questi. Pertanto un singolo pin può fungere da connessione d'ingresso o connessione d'uscita in tempi diversi mediante l'impiego di due buffer tri-state, come mostrato in figura. Se  $S=0$  sarà abilitato il buffer 1 e disabilitato il 2. Al pin saranno quindi presenti i bit in uscita dal circuito. Viceversa quando  $S$  viene posto alto, il buffer 1 è disabilitato, il 2 è abilitato e l'ingresso del circuito può ricevere dati dal pin.

Il pin si comporta da ingresso o da uscita del circuito a seconda del segnale  $S$  che comunemente viene generato dallo stesso circuito digitale.

Accenneremo ora ad un particolare tipo di buffer comunemente impiegato nelle interfacce d'ingresso di circuiti digitali. Supponiamo che il segnale digitale in ingresso  $V_i$  abbia tempi di transizioni lenti rispetto a quelli propri del buffer. Nel caso ideale, mostrato in figura, in cui il segnale d'ingresso non è soggetto a rumore, il buffer presenterà transizioni in corrispondenza dell'attraversamento della sua soglia logica  $V_T$ . Si può tuttavia verificare comunemente che il segnale d'ingresso sia affetto da rumore, dovuto ad interferenze, come mostrato in figura. In tal caso la soglia viene attraversata più volte ed il buffer avendo tempi di transizione brevi risponde con delle commutazioni che tuttavia sono indesiderate in quanto sono causate dal rumore sovrapposto al segnale.

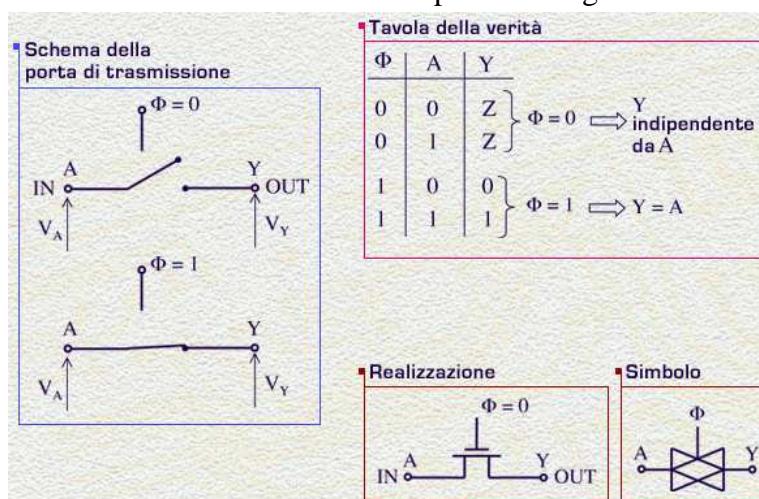
Per conferire al buffer un certo grado di immunità a tali disturbi, sono stati ideati dei buffer aventi due soglie di commutazione invece che una sola. Una soglia superiore  $V_{T+}$  in corrispondenza della quale il buffer commuta quando il segnale d'ingresso cresce ed una soglia inferiore  $V_{T-}$  a cui il buffer commuta quando il segnale decresce. In tal caso la risposta del buffer al segnale affetto da rumore non presenta più commutazioni multiple fintanto che l'ampiezza del rumore è minore della differenza tra le due tensioni di soglia. Un invertitore di questo tipo presenta una caratteristica ingresso – uscita come mostrato in figura.

Per  $V_i$  crescenti da  $V_L$  a  $V_H$  la curva seguita è quella con la tensione a  $V_{T+}$  mentre per  $V_i$  decrescenti dal valore alto  $V_H$  al valore basso  $V_L$  la curva caratteristica transisce a  $V_{T-}$ . Un invertitore con questa caratteristica è chiamato "invertitore con isteresi" o anche "Schmitt-trigger".



## Porte di trasmissione

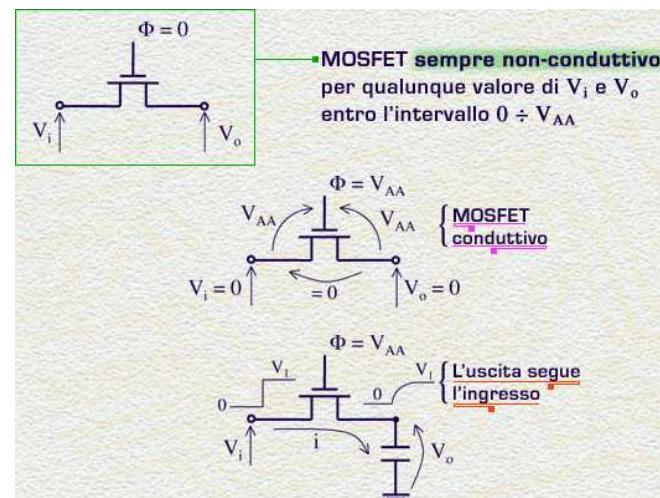
Studieremo ora un altro tipo di porta logica: la porta di trasmissione. La porta può essere schematizzata da un interruttore posto tra ingresso e uscita e comandato in chiusura e apertura da un segnale digitale  $\Phi$ .



Se  $\Phi = 0$  l'interruttore è aperto, la variabile Y risulta pertanto di valore indipendente dall'ingresso A. Se  $\Phi = 1$  l'interruttore è chiuso e  $Y = A$ . Vediamo quindi la tavola della verità. Una porta di trasmissione può essere realizzata con un MOSFET comandato sul gate dal segnale  $\Phi$ .

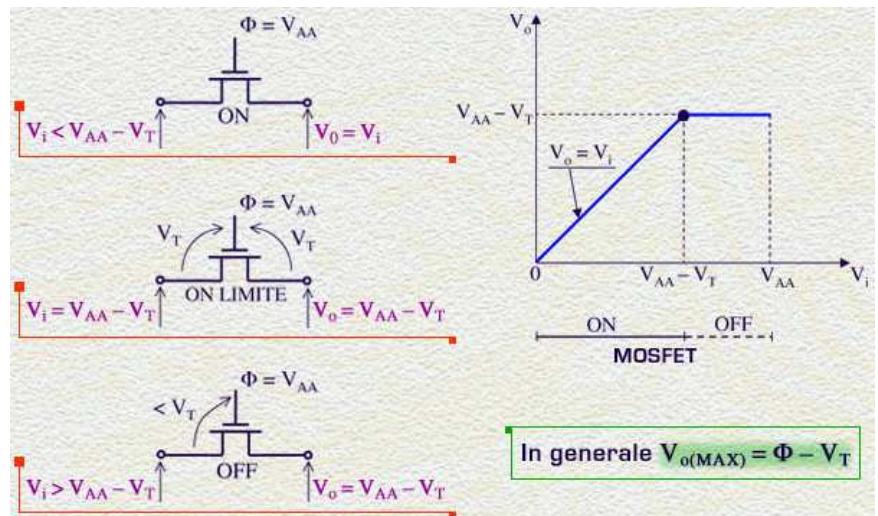
Tracciamo ora la caratteristica statica ingresso – uscita di questa porta. Supponiamo che inizialmente sia  $\Phi = 0$ . In tal caso qualsiasi siano i valori delle tensioni  $V_i$  e  $V_o$ , entro l'intervallo tra il

valore logico basso (0 Volt) e la tensione di alimentazione  $V_{AA}$ , il MOSFET risulta non conduttivo. Infatti perché il MOSFET sia conduttivo è necessario che il gate sia posto ad un potenziale rispetto ad almeno uno degli altri due elettrodi, ingresso o uscita, di una quantità almeno pari alla tensione di soglia del MOSFET stesso. Quando  $\Phi = 0$  questa condizione non è mai verificata e quindi il MOSFET è equivalente ad un interruttore aperto. Supponiamo ora che  $\Phi$  sia pari alla tensione di alimentazione  $V_{AA}$  e che sia  $V_i$  che  $V_o$  siano inizialmente 0 Volt. Dato che  $V_{AA}$  è superiore alla tensione di soglia il MOSFET è conduttivo ed il suo punto di lavoro è sulla caratteristica  $V_{GS} = V_{AA}$  nell'origine della caratteristica stessa essendo nulla la differenza di potenziale tra i due elettrodi di canale del MOSFET.



Aumentiamo ora la tensione  $V_i$  all'elettrodo d'ingresso. Dato che  $V_o$  è inizialmente a 0 Volt si avrà una differenza di potenziale tra ingresso e uscita che genererà una corrente nel MOSFET diretta dall'ingresso verso l'uscita. La capacità del nodo d'uscita inizierà quindi a caricarsi e  $V_o$  a crescere. Quando  $V_o$  raggiunge il valore di  $V_i$  la corrente si annulla ed è raggiunta una condizione di equilibrio, in cui  $V_o = V_i$ .

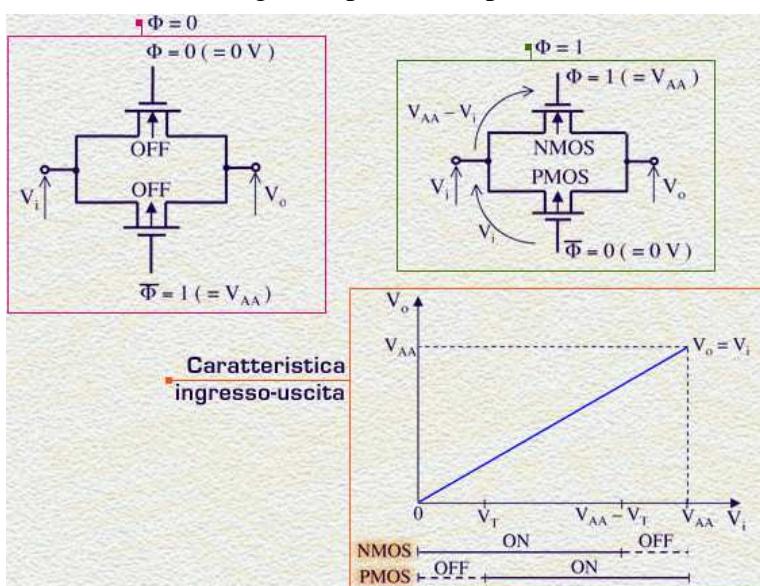
Quindi, fintanto che il MOSFET è conduttivo il potenziale del nodo d'uscita a seguire il potenziale del nodo d'ingresso. In condizioni statiche, ovvero a transitori esauriti,  $V_o = V_i$ . La caratteristica statica ingresso – uscita è quindi la retta  $V_o = V_i$ .



Occorre però osservare che quando  $V_i$ , crescendo, raggiunge il valore  $V_{AA} - V_T$  e tale valore sarà assunto anche da  $V_o$  che segue  $V_i$ , si verifica che la tensione tra gate e gli elettrodi di canale è pari a  $V_T$ , quindi il MOSFET è al confine tra conduzione e interdizione. Se ora la tensione d'ingresso sale sopra  $V_{AA} - V_T$  il MOSFET diviene non – conduttivo, non vi sarà corrente nel canale e quindi la capacità del nodo d'uscita rimarrà carica a  $V_{AA} - V_T$ , che è il valore raggiunto quando il MOSFET conduceva. La caratteristica per  $V_i > (V_{AA} - V_T)$  è quindi un tratto orizzontale. La caratteristica vista ci dice che, se i livelli logici in ingresso alla porta sono 0 e  $V_{AA}$  quelli in uscita saranno 0 e  $V_{AA} - V_T$ , ovvero la porta trasmette immutato il livello logico basso mentre quello alto viene trasmesso ridotto in ampiezza di  $V_T$  Volt.

Generalizzando possiamo dire che il *valore massimo di tensione che il nodo d'uscita può raggiungere è pari alla tensione  $\Phi$  imposta al gate meno la tensione di soglia*.

Per trasmettere sia il livello basso che quello alto inalterati si impiega la porta di trasmissione in configurazione CMOS. Si noti che il gate del PMOS è comandato con il segnale  $\Phi$  (negato) complementare. Con  $\Phi = 0$  entrambi i MOSFET sono non – conduttori per qualsiasi valore di  $V_i$  e  $V_o$ . Per l'NMOS abbiamo già visto perché. Per quanto riguarda il PMOS,  $\Phi = 0$  significa  $\Phi$  negato = 1, ovvero che la tensione al gate è pari alla tensione di alimentazione  $V_{AA}$ . Il PMOS non può quindi condurre avendo il gate al potenziale più alto.



Poniamo ora  $\Phi = 1$  e tracciamo la caratteristica statica ingresso – uscita. Come già visto l'NMOS risulta conduttivo per  $V_i$  compreso tra 0 e  $V_{AA} - V_T$ . Per quanto riguarda il PMOS possiamo dire che è conduttivo per tutti i valori di  $V_i$  superiori alla tensione di soglia  $V_T$ . Infatti per tutti questi valori il gate del PMOS è ad un potenziale inferiore di almeno  $V_T$  rispetto agli elettrodi di canale. Ne consegue che per ogni valore di  $V_i$  compreso tra 0 e  $V_{AA}$  c'è sempre almeno un MOSFET conduttivo e perciò la caratteristica statica sarà la retta  $V_o = V_i$  in tutto l'intervallo. Entrambi i valori logici

vengono quindi trasmessi dall'ingresso all'uscita senza venire alterati.

<pagina bianca per esercizio >

<pagina bianca per data-sheet>

## Lezione n. 20 : Circuiti logici

### Introduzione

Studieremo ora un circuito chiamato decodificatore. Questo circuito costituisce parte integrante delle memorie digitali a semiconduttore oltre che avere ulteriori applicazioni. Non procederemo definendo il decodificatore in modo diretto ma piuttosto esamineremo un problema legato alla realizzazione di una memoria digitale e da questo arriveremo a capire la necessità di questo nuovo circuito.

Così facendo perseguiremo tre finalità:

- Avremo un esempio di metodologia progettuale ingegneristica che, partendo da un problema legato ad un'applicazione, arriva ad ideare una soluzione.
- Sintetizzeremo un circuito semplice, che impiega le porte logiche finora studiate, a partire da un'applicazione specifica.
- Inizieremo a porre alcune basi per lo studio delle memorie che affronteremo sistematicamente in una prossima lezione.

Iniziamo quindi con una breve introduzione sulla memorizzazione di bit.

Memorizzare un bit significa, in pratica, **mantenere nel tempo un livello di tensione, alto o basso**.

Vi sono diversi circuiti che possono svolgere questa funzione, ne vedremo ora uno tra i più semplici. La funzione di mantenere nel tempo una tensione può essere svolta da un **condensatore**. Infatti, consideriamo un condensatore carico alla tensione  $V_H$  corrispondente al livello logico alto. Se il condensatore è isolato ovvero se non entra e non esce carica dalle sue armature e se non vi è conduzione parassita di corrente tra le due armature, allora la tensione ai capi del condensatore viene mantenuta indefinitamente nel tempo. Analogamente, un condensatore scarico, quindi con tensione nulla ai suoi capi, corrispondente al livello logico basso, rimane tale indefinitamente.

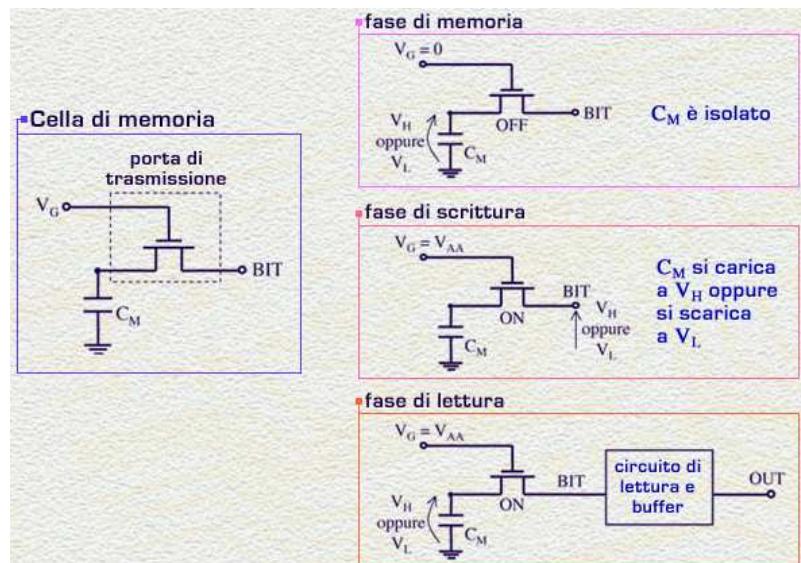
I due stati: di condensatore carico o scarico realizzano quindi la memorizzazione del bit 1 e del bit 0 rispettivamente. Scrivere il bit 1 significa quindi caricare il condensatore a tensione alta e scrivere il bit 0 significa scaricarlo a tensione 0. Leggere il bit memorizzato significa verificare se il condensatore è carico o scarico.

Entrambe le operazioni di scrittura e lettura richiedono l'accesso al condensatore, che può essere realizzato tramite una porta di trasmissione come rappresentato in figura. In fase di memoria la tensione  $V_G$  al gate del MOSFET è tenuta bassa (0 volt), il MOSFET è non – conduttivo ed è equivalente ad un interruttore aperto.

Il condensatore risulta quindi isolato e mantiene il suo stato di polarizzazione. In fase di scrittura o di lettura della cella di memoria, la tensione al gate del MOSFET viene portata alta, il transistor è conduttivo ed il condensatore è accessibile elettricamente dall'elettrodo indicato con BIT. In fase di scrittura, all'elettrodo BIT verrà imposta una tensione alta o bassa a seconda che si voglia caricare o scaricare il condensatore di memoria e quindi scrivere il bit 1 o il bit 0 nella cella.

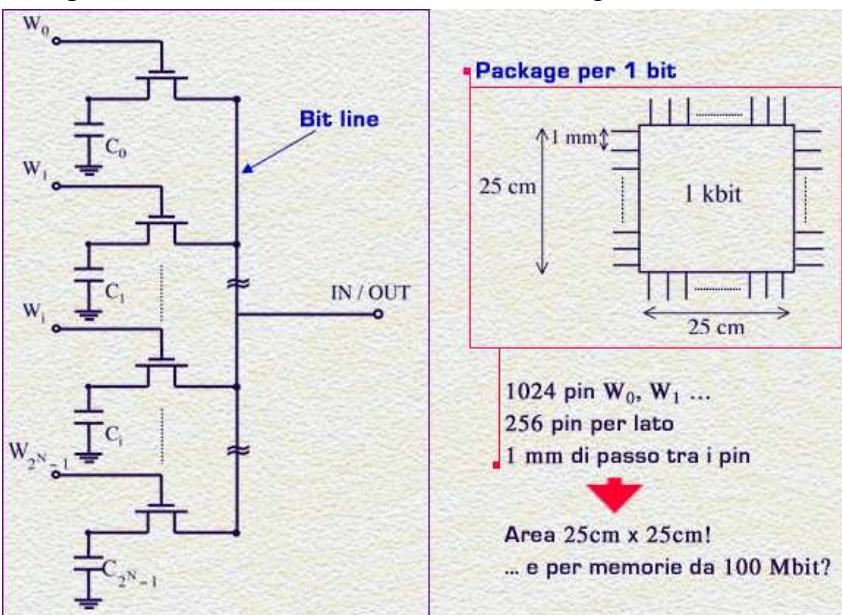
In fase di lettura, all'elettrodo BIT, è connesso un circuito capace di "leggere" la tensione dell'elettrodo, tensione che sarà alta o bassa a seconda che il condensatore sia carico o scarico. Il livello logico corrispondente alla tensione letta viene quindi reso disponibile all'uscita tramite un buffer in grado di pilotare altri circuiti. In sintesi abbiamo visto che una cella in grado di memorizzare un bit può essere realizzata da un condensatore ed una porta di trasmissione.

Se vogliamo quindi realizzare una memoria in grado di contenere  $2^N$  bit possiamo impiegare  $2^N$  celle come quella appena esaminata. Supponendo per ora che sia sufficiente, per la nostra applicazione,



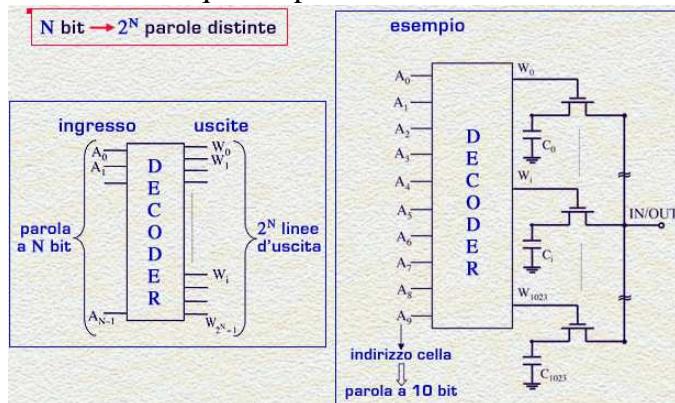
poter scrivere o leggere un solo bit alla volta, possiamo collegare insieme tutti gli elettrodi BIT in un'unica linea detta BIT LINE, così da convergere su di un unico elettrodo di ingresso/uscita per introdurre o leggere i bit dalle celle.

La cella  $i$ -esima viene resa accessibile per scrittura o lettura ponendo alta la tensione all'elettrodo  $W$  indice  $i$  che comanda il gate della  $i$ -esima porta di trasmissione e mantenendo tutti gli altri elettrodi  $W$  a tensione bassa. Pensando di racchiudere tale memoria in un contenitore, comunemente denominato con il termine inglese "package", possiamo subito verificare come non sia possibile, anche per memorie di piccola capacità, rendere disponibile all'esterno del package un numero di elettrodi, detti anche "pin", pari al numero degli elettrodi  $W$  di selezione delle celle. Infatti consideriamo ad esempio una memoria da 1 kbit, ovvero con 1024 celle. Supponendo la distanza tra due pin pari a 1 mm ed i 1024 pin distribuiti su quattro lati del package, risulterebbe un package di 25 cm di lato per una memoria di solo 1 kbit. Chiediamoci quindi quale superficie occuperebbe una memoria da 100 Mbit pur pensando ad una disposizione dei pin non limitata ai lati ma che sia distribuita su tutta la superficie inferiore del package.



## Decoder

Per risolvere questo problema ricordiamo che con  $N$  bit è possibile formare  $2^N$  parole distinte.

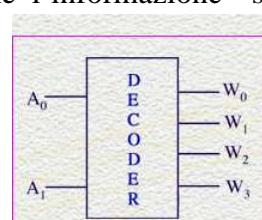


d'ingresso per la memoria da 1 kbit si ridurrebbe drasticamente da 1024 a 10. Il circuito che realizza la funzione vista si chiama "decodificatore" perché l'informazione "seleziona la cella  $i$ -esima" contenuta in codice nella parola d'ingresso, viene de – codificata, ovvero interpretata ed attuata.

Vedremo ora come sia possibile realizzare un decodificatore impiegando le porte logiche studiate. Per semplicità analizzeremo il decodificatore più piccolo, ovvero con parola d'ingresso a 2 bit e quindi quattro uscite. Per il decoder è possibile costruire la tavola della verità riportata in cui vediamo che ogni linea  $W$  va alta solo in corrispondenza di una particolare parola in ingresso. Possiamo quindi scrivere la relazione funzionale logica tra ogni variabile d'uscita  $W$  e le variabili d'ingresso  $A_1$   $A_0$ . Vediamo così che  $W_0$  è 1 solo se

Nell'esempio della memoria a 1024 celle, ogni cella può quindi essere identificata da una particolare parola a 10 bit. Generalizzando possiamo dire che sono sufficienti solo  $N$  bit per identificare una particolare cella su  $2^N$  celle. Allora possiamo pensare di realizzare un circuito che abbia solo  $N$  linee d'ingresso a cui vengono applicati gli  $N$  bit di una parola e, a seconda della parola in ingresso, una e una sola delle  $2^N$  linee d'uscita venga portata alta.

Con un circuito siffatto il numero di pin



$$\begin{aligned} W_0 &= \overline{A}_1 \cdot \overline{A}_0 = \overline{A}_1 + \overline{A}_0 \\ W_1 &= \overline{A}_1 \cdot A_0 = \overline{A}_1 + \overline{A}_0 \\ W_2 &= A_1 \cdot \overline{A}_0 = \overline{A}_1 + A_0 \\ W_3 &= A_1 \cdot A_0 = \overline{A}_1 + \overline{A}_0 \end{aligned}$$

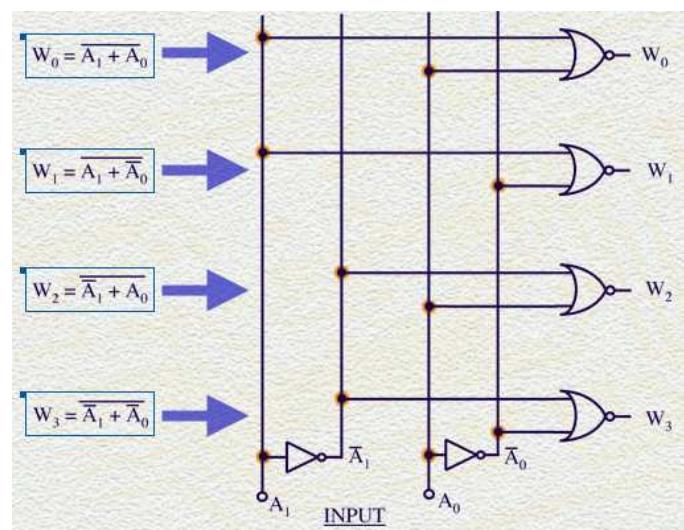
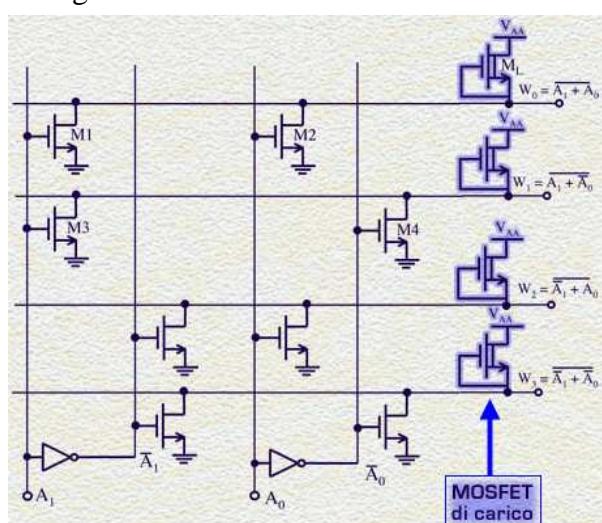
parola d'ingresso	uscite				
$A_1$	$A_0$	$W_0$	$W_1$	$W_2$	$W_3$
0	0	1	0	0	0
0	1	0	1	0	0
1	0	0	0	1	0
1	1	0	0	0	1

Il circuito può essere realizzato con porte logiche AND o NOR una volta che si abbiano a disposizione gli ingressi  $A_0$  e  $A_1$ .

$A_0$  e  $A_1$  valgono entrambi 0, quindi  $W_0$  è il prodotto logico  $A_1$  negato per  $A_0$  negato, e così via tutte le uscite W. Per le leggi di De Morgan possiamo anche tradurre i prodotti in somme logiche come mostrato. Queste equazioni ci dicono che il circuito può essere realizzato con porte logiche AND oppure con porte logiche NOR con ingressi  $A_0$ ,  $A_0$  negato,  $A_1$  e  $A_1$  negato.

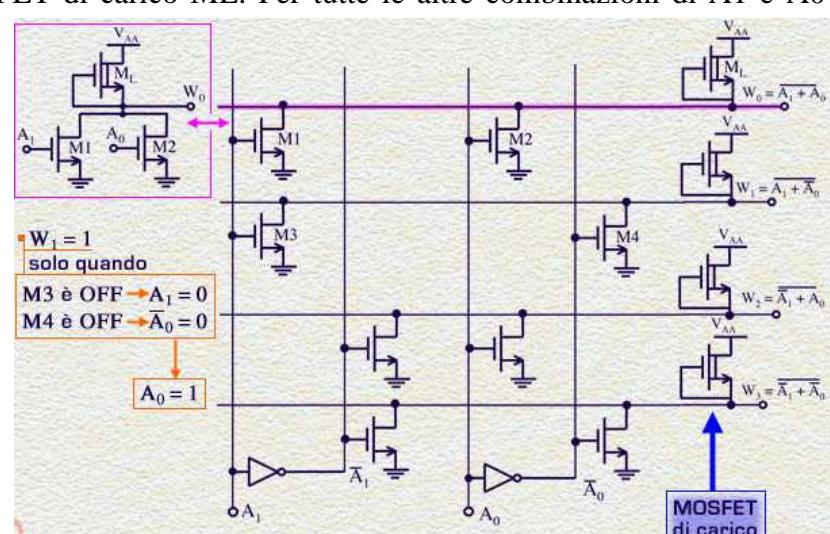
In figura vediamo lo schema del decodificatore a 2 ingressi e 4 uscite realizzato con porte NOR. Osserviamo la struttura a matrice dove sulle righe vi sono le piste di ingresso delle porte NOR e sulle colonne le piste dei bit d'ingresso  $A_0$ ,  $A_0$  negato,  $A_1$  e  $A_1$  negato. Le connessioni tra righe e colonne sono fatte solo ove contrassegnato con il punto. Ad esempio gli ingressi della NOR corrispondente all'uscita  $W_0$  sono connessi alle colonne di  $A_1$  e  $A_0$  così da realizzare la funzione  $(A_1 + A_0)$  negato.

In figura vediamo il circuito del decodificatore 2-4 realizzato con tecnologia impiegante solo transistori MOSFET a canale N. Analizziamo il circuito: osserviamo innanzitutto che ogni riga  $W_0$ ,  $W_1$ ,  $W_2$  e  $W_3$  sono connesse all'alimentazione  $V_{AA}$  tramite MOSFET's con gate e source cortocircuitati. Questi MOSFET, il cui simbolo differisce dagli altri, per la linea che collega source e drain, sono particolari transistor a canale N detti "a svuotamento". Sono MOSFET realizzati in modo che esista il canale anche per tensione gate – source nulla. Quindi sono sempre conduttori e collegano le linee W verso l'alimentazione in maniera, se non equivalente, simile a quella che farebbero dei resistori al loro posto. Per questo motivo possiamo denominarli MOSFET di carico. Supponendo non – conduttori tutti gli altri MOSFET allora tutte le righe  $W_0$ ,  $W_1$ ,  $W_2$  e  $W_3$  saranno tenute alla tensione di alimentazione  $V_{AA}$  dai rispettivi MOSFET di carico.



Analizziamo ora il **funzionamento del circuito concentrando la nostra attenzione sulla sola riga  $W_0$** .

Se  $A_1 = 0$  e  $A_0 = 0$  allora i due MOSFET M1 e M2 saranno entrambi non – conduttori e la linea  $W_0$  sarà tenuta a tensione  $V_{AA}$  dal MOSFET di carico ML. Per tutte le altre combinazioni di  $A_1$  e  $A_0$  almeno uno dei due MOSFET M1 o M2 sarà conduttivo e collegherà la linea  $W_0$  verso massa. Dato che i MOSFET M1 e M2 vengono realizzati con condutività di canale molto più grande del MOSFET di carico ML, la tensione di  $W_0$  sarà prossima agli 0 Volt del riferimento di massa. Quindi l'uscita  $W_0$  risulta alta solo in corrispondenza della parola 00 in ingresso. I transistor M1 e M2 insieme ad ML costituiscono una porta NOR come mostrato in figura. Analogamente, l'uscita  $W_1$  è alta solo se i transistor M3 e M4 sono entrambi non – conduttori, condizione che si verifica solo se  $A_1 = 0$  e  $A_0 = 1$ . E così via per le altre uscite.

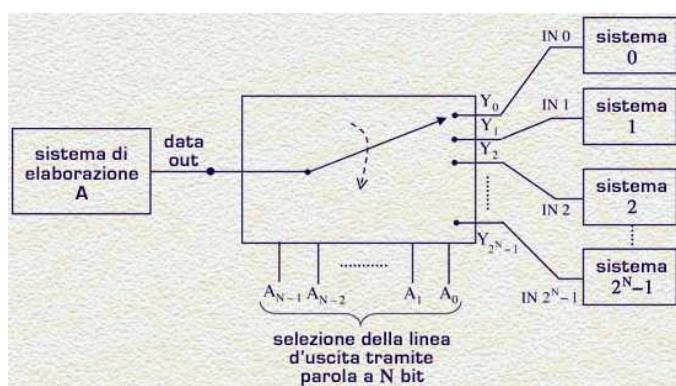


La costruzione della matrice di MOSFET è molto semplice perché rispecchia la variabile d'uscita. Ad esempio l'uscita  $W_0$  è la NOR di  $A_1$  e  $A_0$  ed i MOSFET M1 e M2 della riga sono posti proprio in corrispondenza delle colonne  $A_1$  e  $A_0$ . L'uscita  $W_1$  è la NOR di  $A_1$  e  $A_0$  negato ed i MOSFET M3 ed M4 sono sulle colonne di  $A_1$  e  $A_0$  negato e così via.

Il decodificatore può essere impiegato anche in altre applicazioni oltre a quella vista relativa alle memorie. In generale, ogni qualvolta occorra far corrispondere una particolare operazione ad una parola a  $N$  bit si può impiegare un decodificatore. Vedremo ora due circuiti logici che impiegano il decodificatore come loro parte integrante.

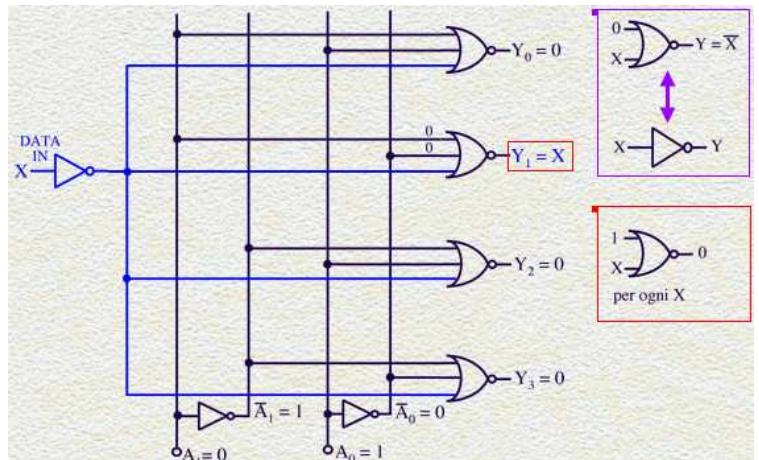
## Demultiplexer

Supponiamo che sia necessario inviare i bit in uscita da un sistema di elaborazione A ad uno solo dei sistemi numerati con  $0, 1, 2 \dots 2^N - 1$ .



Deve inoltre esser possibile scegliere a quale sistema inviare i dati. Occorre perciò realizzare un circuito che, ricevendo al suo ingresso i dati del sistema A possa indirizzarli su una e una sola delle sue  $2^N$  uscite connesse ai sistemi riceventi. Questo circuito viene chiamato DEMULTIPLEXER e può essere rappresentato schematicamente come un interruttore con un numero di posizioni pari al numero delle sue linee d'uscita. La selezione della particolare uscita viene effettuata attraverso la decodifica della parola di selezione a  $N$  bit,  $A_0, A_1, A_2 \dots A_{N-1}$  che identifica una tra le  $2^N$  linee. Possiamo quindi intuire come all'interno del DEMULTIPLEXER vi debba essere il circuito di decodifica della parola di indirizzo. Per semplicità costruiremo lo schema di un demultiplexer a 4 uscite, quindi con parola di indirizzo a 2 bit  $A_0, A_1$ .

Consideriamo quindi come base di partenza per la realizzazione di un demultiplexer lo schema del decodificatore. Per ciascuna parola  $A_1 A_0$  il decoder tiene a livello logico alto solo la linea selezionata mentre tutte le altre sono tenute a livello basso. Ad esempio con l'indirizzo  $A_1 = 0$   $A_0 = 1$  risulta alta solo la linea d'uscita  $Y_1$ , avendo la corrispondente porta NOR entrambi i suoi ingressi  $A_1$  e  $A_0$  negati bassi.



Aggiungendo allora un altro ingresso a tale porta NOR è possibile trasmettere il relativo bit notando che una NOR con un ingresso basso si comporta come un invertitore rispetto all'altro ingresso. Quindi affinché il dato d'ingresso X appaia in uscita occorre aggiungere un invertitore sulla linea d'ingresso. Analogamente possiamo aggiungere un ingresso a tutte le porte NOR ed applicare X negato contemporaneamente a tutte.

Solamente l'uscita selezionata però trasmetterà la variabile X perché tutte le altre linee rimarranno vincolate a livello 0 indipendentemente da X. Infatti tutte le altre porte NOR, per dare uscita 0, devono avere almeno un ingresso alto ed una NOR con un ingresso alto fornisce un'uscita bassa indipendentemente dai valori presenti agli altri ingressi.

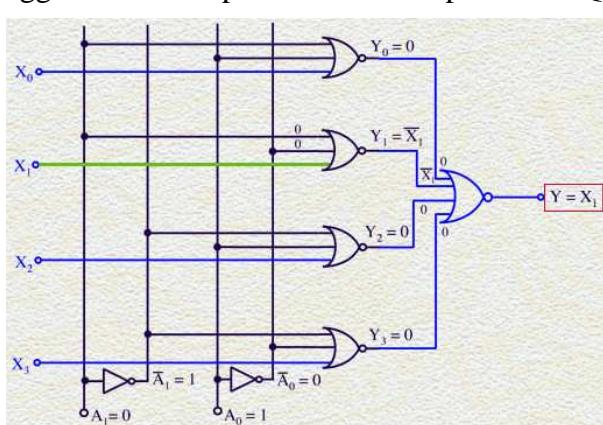
## Multiplexer

Vediamo adesso un altro circuito basato sul decodificatore. Questo circuito, che si chiama MULTIPLEXER, è rappresentato schematicamente in figura. Il MULTIPLEXER svolge la funzione di collegare uno ed uno solo degli ingressi X all'uscita Y. La selezione del particolare ingresso viene effettuata attraverso la decodifica della parola di selezione a N bit, che identifica quindi uno tra i  $2^N$  ingressi.

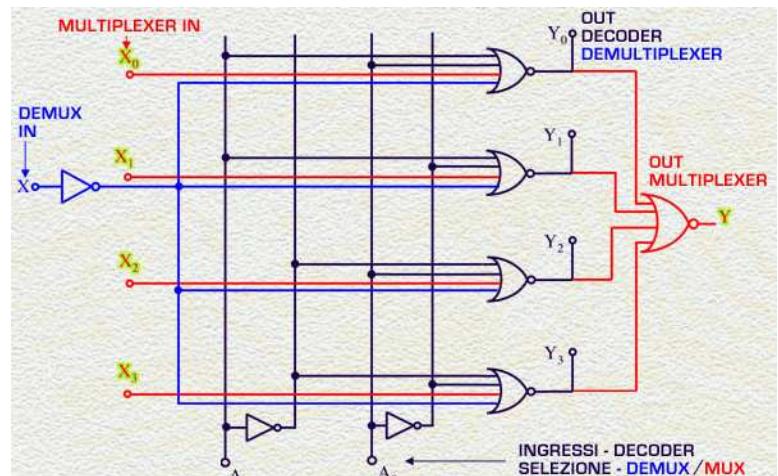
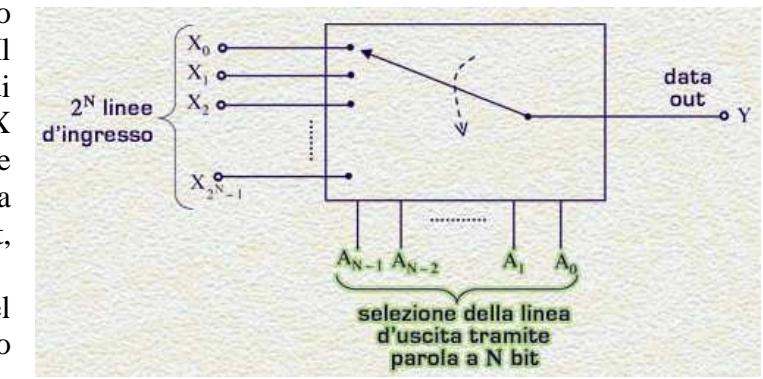
Vediamo come realizzare lo schema del multiplexer a 4 ingressi a partire da quello del decodificatore.

Come in precedenza visto la linea selezionata del decoder può trasmettere il bit inserito ad un ingresso aggiuntivo della porta NOR corrispondente. Quindi possiamo introdurre un ingresso per ogni NOR. A questo punto dato che il MULTIPLEXER deve avere un'unica uscita possiamo collegare tutte le uscite del decodificatore ad un'unica porta NOR a 4 ingressi. È semplice verificare che l'ingresso trasmesso in uscita sarà solo quello corrispondente alla linea selezionata.

Infatti tutte le altre linee rimarranno fisse a livello 0. Facciamo un esempio: supponiamo di selezionare la linea 1 tramite il codice  $A_1 = 0$ ,  $A_0 = 1$ , allora  $Y_0$ ,  $Y_2$  e  $Y_3$  sono a 0 mentre  $Y_1$  sarà uguale a  $X_1$  negato. La porta NOR d'uscita si comporta da invertitore dell'ingresso  $Y_1$ , quindi in uscita sarà presente il bit  $X_1$ .

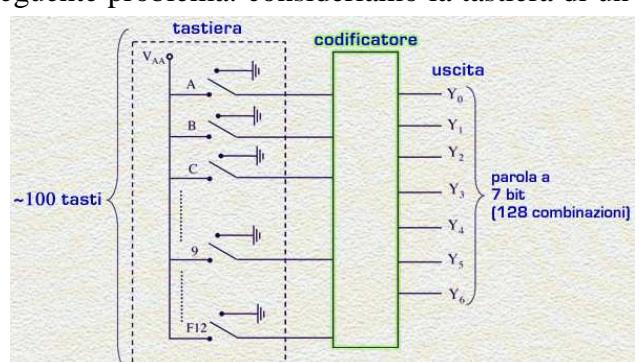


Abbiamo quindi visto che demultiplexer e multiplexer posseggono una struttura comune costituita dallo schema del decodificatore. Quindi, se necessario, possiamo realizzare un unico circuito che implementi tutte e tre le funzioni. In figura è mostrato lo schema del circuito che può operare da decodificatore con ingressi  $A_1$   $A_0$  ed uscite  $Y_0$ ,  $Y_1$ ,  $Y_2$   $Y_3$ ; può operare da demultiplexer con ingresso  $X$  ed uscite  $Y_0$ ,  $Y_1$ ,  $Y_2$   $Y_3$  oppure può essere impiegato come multiplexer utilizzando gli ingressi  $X_0$ ,  $X_1$ ,  $X_2$   $X_3$  e l'uscita  $Y$ .



## Encoder

Studieremo adesso un ultimo circuito; partiamo dal seguente problema: consideriamo la tastiera di un computer con circa 100 tasti tra lettere, numeri e funzioni particolari. Quando viene premuto un tasto la corrispondente linea va alta e l'informazione dev'essere inviata dalla tastiera all'unità di elaborazione. Non è tuttavia realizzabile un collegamento con un numero di conduttori pari al numero di tasti. Si introduce quindi un circuito che traduce in codice binario l'informazione che un particolare tasto è stato premuto e che quindi una particolare linea al suo ingresso è allo stato alto.

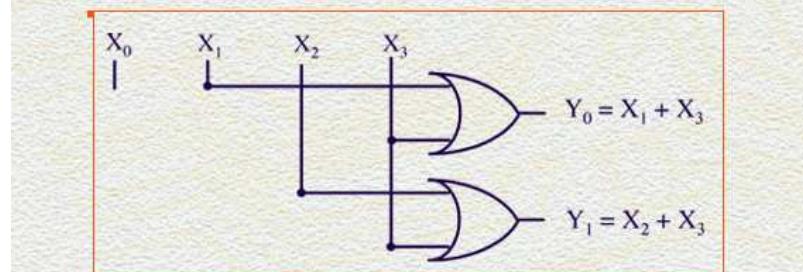
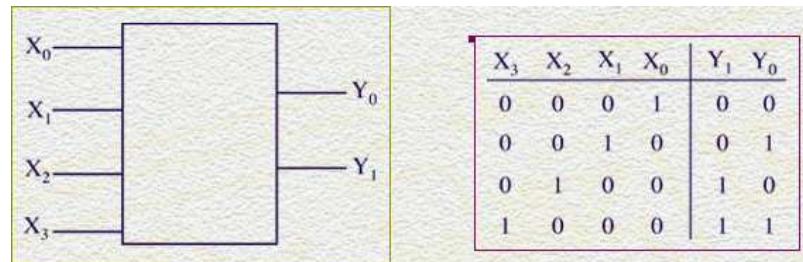


Dato che una parola a  $N$  bit può codificare fino a  $2^N$  informazioni. Nel caso di una tastiera sono sufficienti 7 bit per trasmettere fino a 128 informazioni. La riduzione nel numero di conduttori richiesti è quindi notevole. Dato che il circuito traduce in codice un'informazione viene denominato CODIFICATORE o ENCODER.

Vediamo come ricavare lo schema di un ENCODER a 4 ingressi. La parola d'uscita sarà quindi di 2 bit. Consideriamo la tavola della verità che rappresenta la funzione logica svolta dal circuito. Supponiamo che uno e uno solo degli ingressi possa essere nello stato alto. Osserviamo quindi che quando  $X_0$  è 1 la parola d'uscita è 00,

quando  $X_1$  è 1 l'uscita è 01 e così via. Scriviamo ora le funzioni logiche che legano le uscite  $Y_0$  e  $Y_1$  agli ingressi. Per quanto riguarda  $Y_0$  osserviamo che è 1 quando  $X_1$  è 1 oppure quando  $X_3$  è 1, quindi  $Y_0$  è dato dalla somma logica  $X_1 + X_3$ .

Analogamente  $Y_1$  è 1 quando  $X_2$  o  $X_3$  sono 1, quindi  $Y_1 = X_2 + X_3$ . Il circuito dell'ENCODER è pertanto realizzabile con porte OR come mostrato in figura. L'ingresso  $X_0$  non è collegato a nulla dato che quando  $X_0$  è 1 la parola d'uscita deve essere 00.



<pagina bianca per data-sheet>

<pagina bianca per data-sheet>

# Lezione n. 21 : Le memorie a semiconduttore

## Introduzione

Iniziamo con il vedere quali sono i parametri principali che caratterizzano una generica memoria.

- **volatilità** o meno dei dati. Una memoria viene detta volatile se i dati in essa contenuti svaniscono quando viene a mancare la tensione di alimentazione. Viceversa è detta non – volatile se i dati sono mantenuti anche in assenza di alimentazione.
- **riscrivibilità** o meno dei dati, ovvero la possibilità di cancellare i dati scritti e riscriverne di nuovi, non tutte le memorie lo permettono: ci sono memorie in cui i dati possono essere scritti una sola volta.
- **la modalità con cui i dati memorizzati possono venire cancellati dalla memoria:** alcune memorie consentono una cancellazione:
  - del singolo bit,
  - per gruppi di bit
  - per l'intera memoria.

Altri parametri caratterizzanti una memoria sono:

- **la velocità di scrittura e di lettura,**
- **la capacità di memoria** (ovvero il numero di bit memorizzabili),
- **il consumo di potenza** che può essere distinto nelle fasi di:
  - scrittura,
  - lettura
  - in stand-by o riposo.

Un ulteriore parametro è:

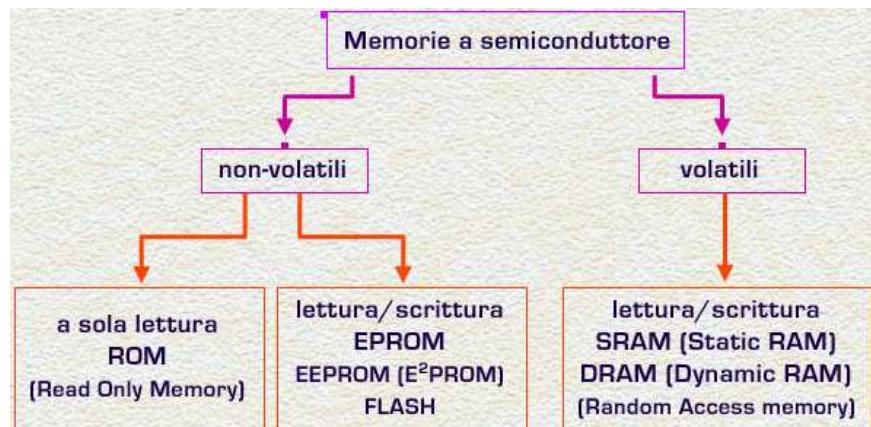
- **la vulnerabilità** o meno dei dati. Alcune memorie a differenza di altre possono operare anche in presenza di radiazioni  $\alpha$ ,  $\gamma$  o flussi di particelle senza che questo provochi perdita o alterazione dei dati. Questo aspetto è particolarmente importante per i sistemi che operano nello spazio, come sui satelliti, dove i circuiti possono essere sottoposti a flussi non trascurabili di radiazioni.

Un ultimo parametro a cui le industrie produttrici di memorie guardano con particolare attenzione è il:

- **costo per bit.**

Facciamo ora una panoramica sulle memorie che studieremo. Distinguendo nei due gruppi di memorie non – volatili e memorie volatili.

Nel primo gruppo vedremo le memorie a sola lettura o ROM (Read Only Memory) e le memorie derivate da queste ma che consentono una riscrittura dei dati: sono le memorie EPROM, E<sup>2</sup>PROM e FLASH. Nel gruppo delle memorie volatili studieremo le RAM statiche e le RAM dinamiche.



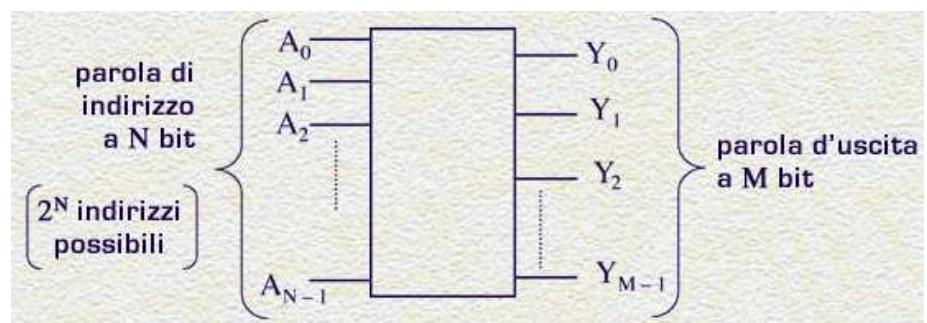
## La memoria ROM

In questa lezione studieremo le memorie non-volatili iniziando dalla memoria a sola lettura chiamata ROM (Read Only Memory). Questa memoria si chiama così perché i bit in essa contenuti vengono inseriti in fase di fabbricazione della memoria stessa e non possono più essere cancellati e venire riscritti nuovi dati, ma possono essere solamente letti.

Memorie di questo tipo svolgono quindi la funzione di rendere disponibili informazioni destinate a rimanere immutate nel tempo. Una memoria ROM può essere vista come nella figura che segue. La parola d'ingresso a N bit  $A_0 A_1 \dots A_{N-1}$  si chiama indirizzo e rappresenta il numero d'ordine dell'informazione a cui si vuole accedere. Applicato un certo indirizzo agli ingressi  $A_0 \dots A_{N-1}$  sulle

uscite  $Y_0, Y_1 \dots Y_{M-1}$  sarà disponibile la parola a  $M$  bit che codifica l'informazione memorizzata in corrispondenza di quel particolare indirizzo.

Chiediamoci: **quante parole di  $M$  bit può contenere la memoria?** La risposta è che, al massimo, può contenere un numero di parole pari al **numero di indirizzi possibili che è  $2^N$** . Il numero

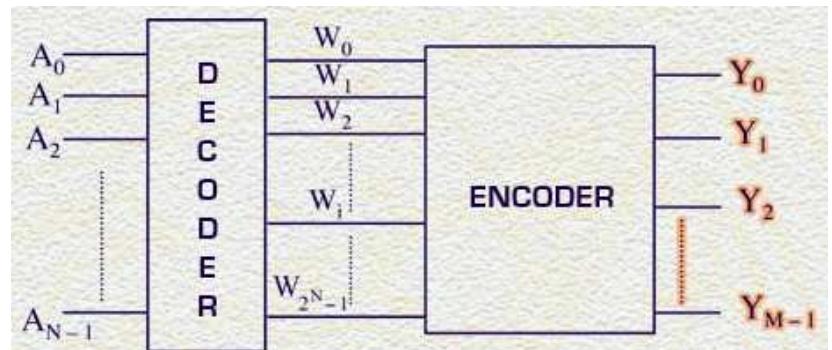


complessivo di **bit contenuti nella memoria è quindi  $2^N \times M$** . La stessa memoria ROM può essere anche vista come un circuito in cui è memorizzata la relazione funzionale tra la variabile d'ingresso, costituita dalla parola  $A_0 A_1 \dots A_{N-1}$  e la variabile d'uscita, la parola  $Y_0 Y_1 \dots Y_{M-1}$ .

Vediamo ora come è fatta una memoria ROM.

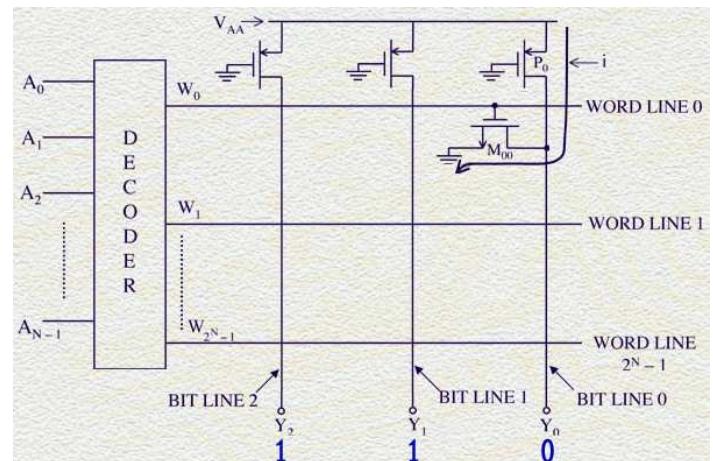
La ROM è costituita semplicemente da due circuiti che abbiamo analizzato in una precedente lezione: un decoder ed un encoder.

Il decoder riceve in ingresso la parola di indirizzo  $A_0 A_1 \dots A_{N-1}$  e porta alta una e una sola delle sue  $2^N$  linee d'uscita, indicate con  $W$ . Le linee  $W$  costituiscono gli ingressi dell'encoder, che fornisce in uscita la parola



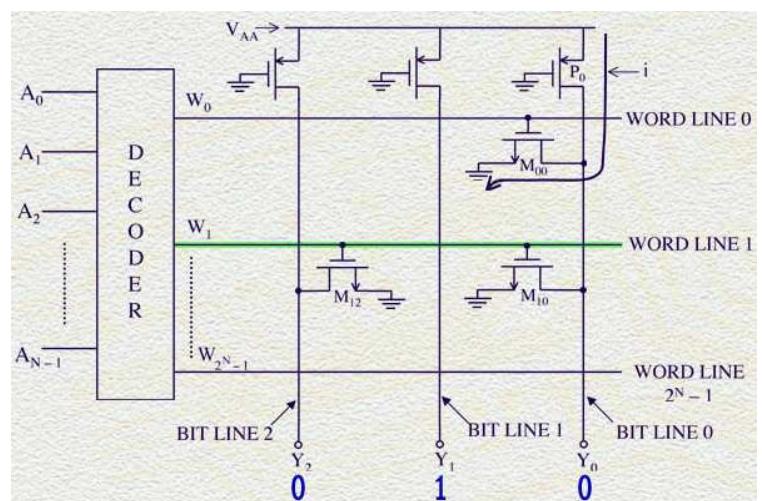
$Y_0 Y_1 \dots Y_{M-1}$  corrispondente alla linea  $W$  che si trova a livello alto. Scendiamo ora più nel dettaglio: il circuito del decoder è stato già presentato ed analizzato in una precedente lezione, per quanto riguarda l'encoder abbiamo già visto il suo schema con porte logiche. Tuttavia dato che l'encoder costituisce il circuito in cui i bit sono memorizzati lo analizzeremo più precisamente.

In figura è rappresentata una memoria ROM in cui è mostrato in dettaglio il circuito dell'encoder con sole 3 uscite, per semplicità. Come si vede la struttura dell'encoder è a matrice: le  $2^N$  uscite del decoder costituiscono le linee orizzontali chiamate WORD LINES. Le linee verticali, in numero pari ai bit d'uscita, sono chiamate appunto BIT LINES e sono connesse alla tensione di alimentazione tramite dei PMOS con il gate ad una tensione (0 V) inferiore rispetto a quella di source che è a  $V_{AA}$  volt, e dato che la tensione di alimentazione  $V_{AA}$  è maggiore del modulo della tensione di soglia del MOSFET (pari a circa 0.7V), ne risulta che i PMOS sono conduttori e tengono ogni BIT LINE alla tensione  $V_{AA}$  di alimentazione. Supponiamo ora di inviare la parola di indirizzi  $A_0, A_1 \dots A_{N-1}$  costituita da tutti zeri così da portare alta la word line  $W_0$ . Supponiamo di volere



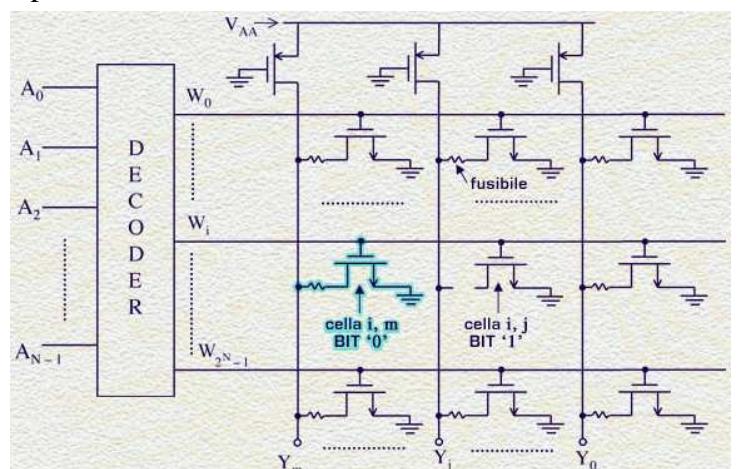
che, in tali condizioni, l'uscita sia  $Y_2=1 Y_1=1 Y_0=0$ . Le bit line  $Y_2$  e  $Y_1$  sono già a livello 1, tenute dai rispettivi PMOS. Per quanto riguarda  $Y_0$ , affinché sia portata a 0, è sufficiente porre un NMOS  $M_{00}$  come in figura, collegato con il gate alla WORD LINE  $W_0$  con il drain alla bit line e con il source a massa. Quando la WORD LINE  $W_0$  va alta l'NMOS  $M_{00}$  diviene conduttivo e connette la bit – line  $Y_0$  a massa. La tensione della bit line  $Y_0$  si porta al livello logico basso prossimo a 0 volt perché l'NMOS  $M_{00}$  viene realizzato molto più conduttivo del PMOS  $P_0$ . Occorre osservare che il fatto che sia  $P_0$  che  $M_{00}$  siano conduttori comporta un passaggio di corrente dall'alimentazione verso massa in condizioni statiche, e perciò un consumo di potenza. Sintetizzando quanto visto possiamo dire che l'indirizzo costituito da tutti zeri ha portato alta la word line  $W_0$ , la quale ha acceso il MOSFET  $M_{00}$ . L'uscita della ROM è perciò la parola 110.

Analogamente se vogliamo che attivando la WORD LINE 1 la parola d'uscita sia  $Y_2 = 0$   $Y_1 = 1$  e  $Y_0 = 0$  allora porremo i MOSFET  $M_{12}$  e  $M_{10}$  nella matrice in modo da portare a livello basso le bit line  $Y_2$  e  $Y_0$ . E così via ... l'uscita corrispondente all'attivazione di  $W_{2^N-1}$  sarà 101 avendo posto il MOSFET all'incrocio tra la WORD LINE  $W_{2^N-1}$  e la BIT LINE  $Y_1$ , (cfr figura sotto). Si comprende quindi anche il significato del nome WORD LINE dato alle linee orizzontali della matrice: infatti ognuna di queste, quando viene selezionata dal decoder e portata a livello alto, mette in stato di conduzione i MOSFET's ad essa connessi generando così la parola (WORD) in uscita. Comprendiamo inoltre, osservando il circuito dell'encoder, come



ogni incrocio riga – colonna della matrice, rappresenti una cella di memoria per un bit. Il bit contenuto nella cella sarà '0' se in quell'incrocio è presente un MOSFET oppure sarà '1' se non c'è alcun MOSFET. Da ciò capiamo come, una volta posti i MOSFET in particolari incroci della matrice, la relazione funzionale ingresso – uscita rimanga definitivamente fissata. In altri termini, il contenuto informativo della memoria non può più essere cambiato ma solo letto, da cui il nome di memoria a sola lettura: ROM.

Il fatto che il contenuto di una ROM debba essere fissato a livello di fabbricazione e poi non possa più essere modificato comporta un forte impegno a livello di produzione ed una **stringente limitazione a livello dell'utente**. A livello di fabbricazione infatti è necessario realizzare un particolare encoder per ogni particolare contenuto della ROM, a livello dell'utente significa non aver possibilità di effettuare eventuali modifiche correttive o aggiornamenti periodici del contenuto della memoria. Si rivelerebbe pertanto utile una memoria avente la stessa struttura circuitale della ROM ma il cui contenuto informativo non sia fissato a livello di fabbricazione ma possa essere inserito successivamente ed eventualmente anche modificato periodicamente se fosse necessario. Per realizzare una tale ROM si può pensare di fabbricarla con un MOSFET ad ogni incrocio dell'encoder, e poi, successivamente alla fabbricazione, pensare ad un modo per scollegare o disattivare solo alcuni di questi MOSFET e precisamente quelli delle celle in cui si vuole sia memorizzato il bit '1'. Nacquero così le ROM programmabili o PROM che impiegavano connessioni tra i transistor e le bit line che agivano come fusibili e potevano essere quindi interrotte dal passaggio di una corrente elevata che il programmatore poteva impostare solo ai transistor delle celle scelte, essendo possibile raggiungere ogni transistor della matrice da una particolare combinazione riga – colonna. Ad esempio la **cella i, j** in figura memorizza il BIT '1' non potendo il transistor connettere la bit line j a massa. Mentre la **cella i, m** memorizza il BIT '0'. Quindi, queste memorie vengono



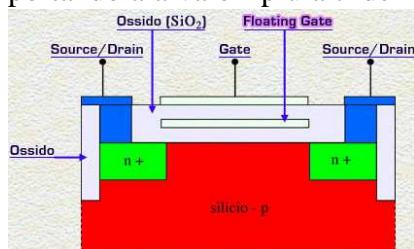
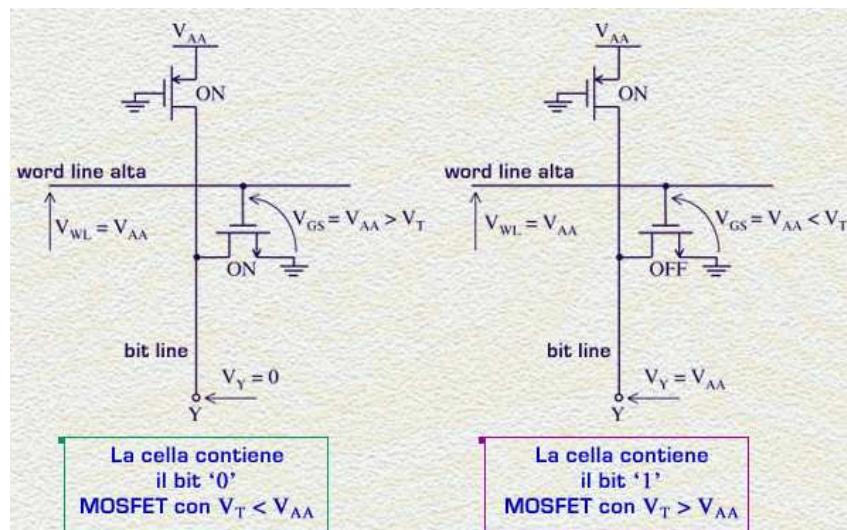
fabbricate tutte uguali, e questo è un vantaggio a livello produttivo, e poi vengono programmate sul campo dall'utente. Per questo vengono anche chiamate "field programmable ROM".

Tuttavia tali memorie impongono ancora il limite di non permettere la modifica delle informazioni memorizzate in quanto le connessioni interrotte nella fase di programmazione non possono più essere ripristinate.

In tal senso occorre considerare il vantaggio derivante dalla possibilità di riprogrammare la ROM sarebbe notevole in quanto consentirebbe di effettuare periodici aggiornamenti del contenuto della memoria senza doverla sostituire con un'altra. Per rendere possibile questo è necessario riuscire a disabilitare un MOSFET senza dover interrompere fisicamente il collegamento con la bit-line. Chiediamoci quindi: come può un MOSFET pur rimanendo collegato a word-line e bit-line a comportarsi come se non ci fosse? La risposta a questa domanda è semplice: la presenza del MOSFET è effettiva nella memorizzazione del bit 0 perché quando il suo gate viene portato alto dalla word line il MOSFET diventa conduttivo.

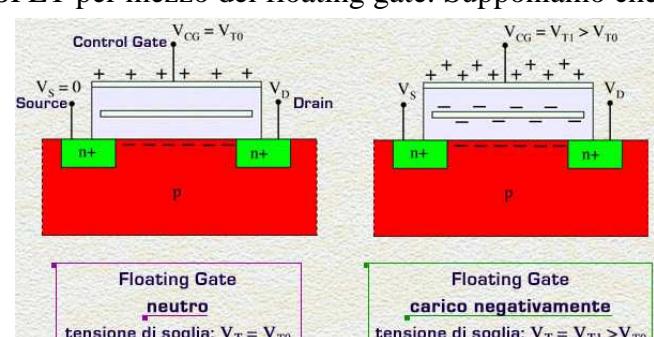
Se ciò non avvenisse, ovvero se il MOSFET continuasse a non condurre anche con la word line alta, il bit memorizzato sarebbe 1 e non 0.

Per far sì che il MOSFET non conduca anche quando la word line è alta, ovvero alla tensione di alimentazione  $V_{AA}$ , è sufficiente che la tensione di soglia del MOSFET sia più alta di  $V_{AA}$ . Quindi per disabilitare un MOSFET sarebbe sufficiente poter modificare la sua tensione di soglia portandola a valori più alti del livello logico alto. Questa idea ha condotto all'invenzione di un nuovo tipo di MOSFET avente in più rispetto al MOSFET che conosciamo un elettrodo conduttivo interposto tra il gate e silicio. Questo elettrodo è circondato da ossido di silicio e non ha connessioni con altri elettrodi. Risulta quindi essere un conduttore fluttuante in potenziale e per questo viene chiamato "**floating gate**". Vediamo ora come, attraverso il floating gate, sia possibile modificare la tensione di soglia del MOSFET per mezzo del floating gate. Supponiamo che



il floating gate non vi sia alcuna carica elettrica netta, ovvero che l'elettrodo sia elettricamente neutro. In tal caso il MOSFET funziona come se il floating gate non ci fosse. In questa condizione la tensione di soglia è  $V_T = V_{T0}$ , valore minore della tensione  $V_{AA}$  corrispondente al valore logico alto. Quindi applicando  $V_{AA}$  al gate di controllo, il MOSFET diviene conduttivo. Supponiamo ora di essere riusciti a portare un certo numero di elettroni sul floating gate, questi sarà quindi carico negativamente. In tal caso la tensione di soglia del MOSFET non è più  $V_{T0}$  ma un valore  $V_{T1}$  più alto, ovvero la presenza di cariche negative sul floating gate alza la tensione di soglia del transistor. Intuitivamente questo può essere compreso ricordando che la tensione di soglia per un MOSFET a canale N corrisponde alla tensione minima da applicare tra gate e source per formare il canale di elettroni. Per alzare la tensione al gate sopra il valore di soglia, inducendo così elettroni nel canale, è necessario porre cariche positive sull'elettrodo di gate. La carica negativa sul floating gate contrasta l'accumulo di elettroni all'interfaccia ossido-silicio in quanto tende a richiamare nel silicio cariche positive! Per generare il canale di elettroni tra source e drain è quindi necessario porre sul gate di controllo un numero di cariche positive maggiore, rispetto al caso di floating gate neutro. Ciò

sul floating gate non vi sia alcuna carica elettrica netta, ovvero che l'elettrodo sia elettricamente neutro. In tal caso il MOSFET funziona come se il floating gate non ci fosse. In questa condizione la tensione di soglia è  $V_T = V_{T0}$ , valore minore della tensione  $V_{AA}$  corrispondente al valore logico alto. Quindi applicando  $V_{AA}$  al gate di controllo, il MOSFET diviene conduttivo. Supponiamo ora di essere riusciti a portare un certo numero di elettroni sul floating gate, questi sarà quindi carico negativamente. In tal caso la tensione di soglia del MOSFET non è più  $V_{T0}$  ma un valore  $V_{T1}$  più alto, ovvero la presenza di cariche negative sul floating gate alza la tensione di soglia del transistor. Intuitivamente questo può essere compreso ricordando che la tensione di soglia per un MOSFET a canale N corrisponde alla tensione minima da applicare tra gate e source per formare il canale di elettroni. Per alzare la tensione al gate sopra il valore di soglia, inducendo così elettroni nel canale, è necessario porre cariche positive sull'elettrodo di gate. La carica negativa sul floating gate contrasta l'accumulo di elettroni all'interfaccia ossido-silicio in quanto tende a richiamare nel silicio cariche positive! Per generare il canale di elettroni tra source e drain è quindi necessario porre sul gate di controllo un numero di cariche positive maggiore, rispetto al caso di floating gate neutro. Ciò



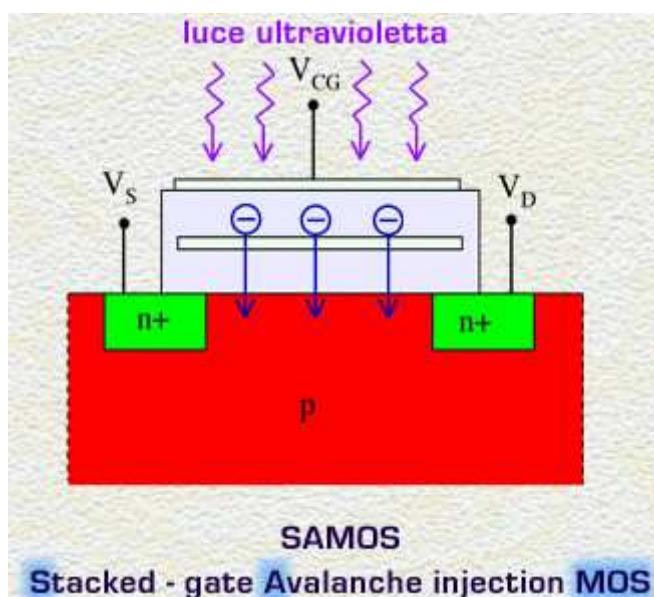
significa che occorre una tensione di gate maggiore per formare il canale, che è lo stesso che dire che la tensione di soglia  $V_{T1}$  è maggiore di  $V_{T0}$ .

Ora vediamo come sia possibile far arrivare elettroni sul floating gate, visto che questo elettrodo non è accessibile dall'esterno essendo circondato da ossido che, come sappiamo, è non-conduttore.

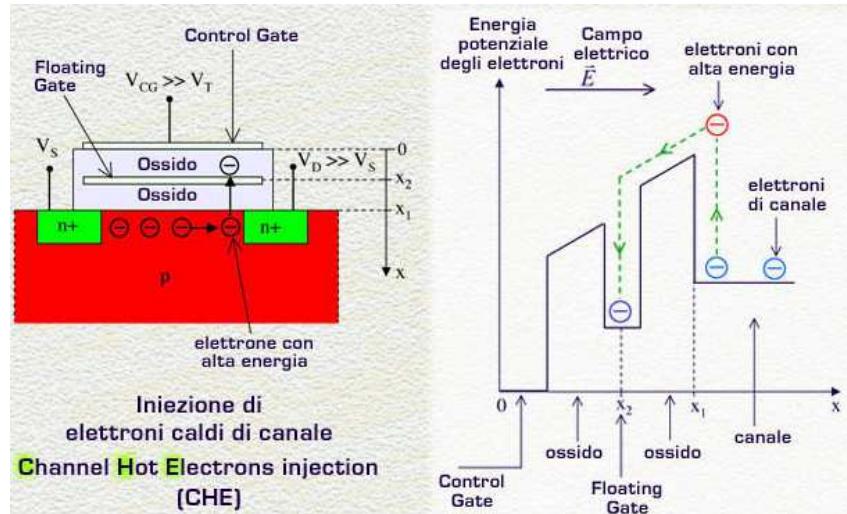
Gli elettroni vengono fatti arrivare sul floating gate dal canale, fornendo loro sufficiente energia per superare lo strato di ossido che li separa dal floating gate. L'ossido infatti viene visto dagli elettroni di canale come una barriera di potenziale molto alta che solo elettroni con elevata energia possono superare. Il primo passo consiste quindi nel fornire elevata energia agli elettroni presenti nel canale. Ciò viene fatto imponendo tra drain e source una elevata tensione così da generare un forte campo elettrico in prossimità del drain.

Gli elettroni di canale prossimi all'elettrodo di drain acquistano alta energia cinetica tale da provocare una ionizzazione a valanga simile al fenomeno di rottura a valanga visto per le giunzioni. Gli elettroni generati dalla ionizzazione hanno energia sufficientemente alta da poter superare la barriera dell'ossido, e contemporaneamente vengono spinti verso il floating gate dal campo elettrico generato dal potenziale positivo imposto al gate di controllo. Man mano che il floating gate si carica di elettroni il fenomeno di iniezione si attenua fino ad annullarsi perché la carica negativa sul floating gate tende a spegnere il MOSFET e ad annullare il campo elettrico che sospinge gli elettroni energetici dal canale verso il gate. Il fenomeno di INIEZIONE è quindi auto limitante.

Gli elettroni che possiedono un'energia elevata, tale da permettere loro il superamento della barriera di potenziale dell'ossido, vengono chiamati elettroni caldi o hot electrons per questo il fenomeno di iniezione descritto è denominato ‘iniezione di elettroni caldi dal canale’ o “Channel Hot Electrons injection”, che indicheremo in seguito con l'acronimo **CHE**.



modalità normale di operazione del MOSFET gli elettroni iniettati non possono lasciare il floating gate a causa dell'ossido circostante, il MOSFET rimane perciò indefinitamente con tensione di soglia alta. Per ripristinare le condizioni di soglia bassa occorre togliere gli elettroni dal floating gate. Ciò viene fatto illuminando il dispositivo con luce ultravioletta, che genera una corrente fotoelettrica riportando gli elettroni nel substrato.



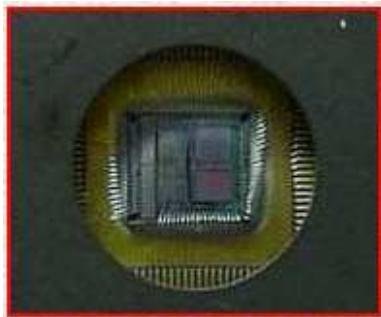
Questo dispositivo è stato proposto nel 1967 e poi realizzato nel 1976. Per la sua modalità di funzionamento è stato denominato **SAMOS**, acronimo di Stacked-gate Avalanche injection MOS ed è una variante di un dispositivo simile proposto due anni prima ed avente solo il floating gate senza il control gate, e chiamato, questo dispositivo, **FAMOS** Floating-gate Avalanche injection MOS. Riprendiamo ora l'analisi del funzionamento del dispositivo. Una volta che gli elettroni sono stati iniettati nel floating gate vengono ripristinate le normali tensioni di funzionamento al control gate ed al drain. In

modo normale di operazione del MOSFET gli elettroni iniettati non possono lasciare il floating gate a causa dell'ossido circostante, il MOSFET rimane perciò indefinitamente con tensione di soglia alta. Per ripristinare le condizioni di soglia bassa occorre togliere gli elettroni dal floating gate. Ciò viene fatto illuminando il dispositivo con luce ultravioletta, che genera una corrente fotoelettrica riportando gli elettroni nel substrato.

## EPROM ed E<sup>2</sup>PROM

Le memorie che impiegano il SAMOS sono chiamate EPROM che sta per ERASABLE Programmable ROM. La loro struttura è del tutto simile a quella delle ROM con la differenza di utilizzare i MOSFET con floating-gate su ogni incrocio word-line/bit-line.

Le EPROM vengono programmate elettricamente mediante il processo di iniezione di elettroni nel floating-gate e possono in seguito essere cancellate, prima di successive programmazioni, esponendole a luce ultravioletta. Per permettere la cancellazione, le EPROM sono contenute in package aventi una finestra trasparente sopra il chip come mostrato in figura.

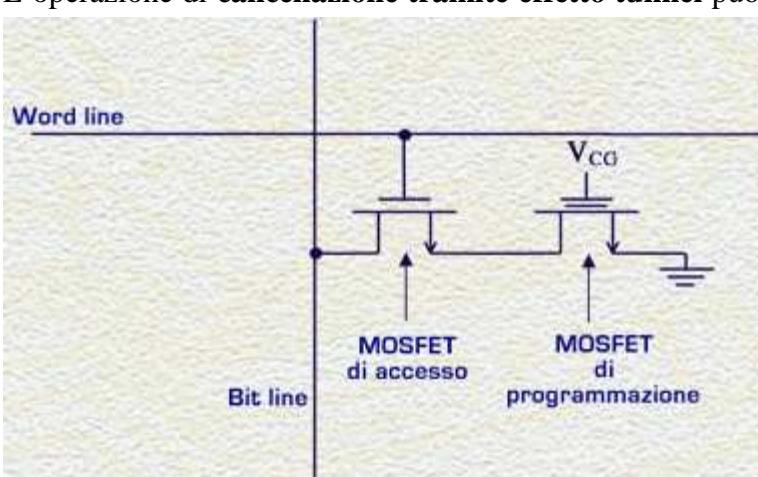
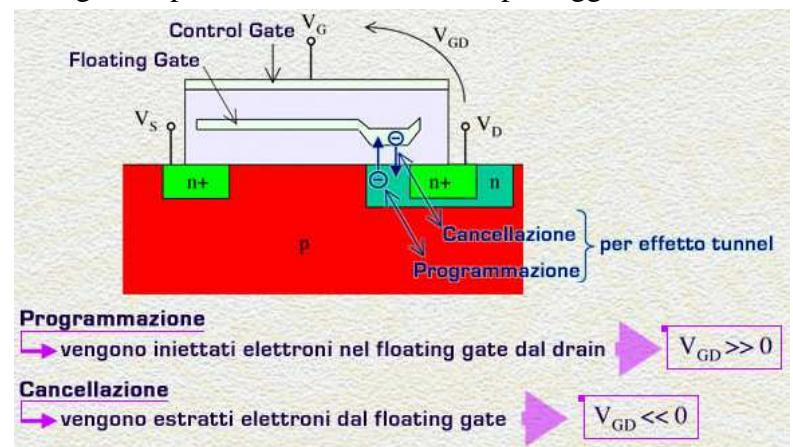
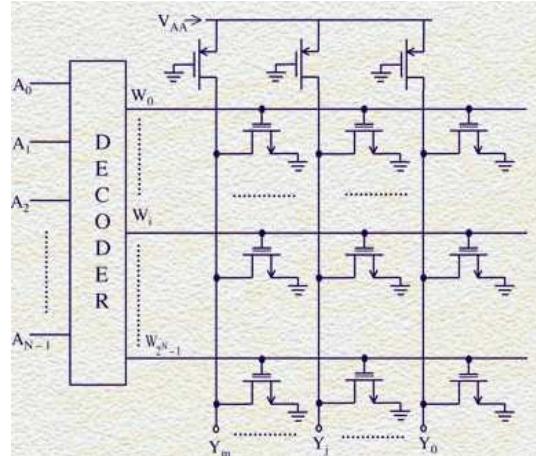


L'intera memoria viene esposta e quindi cancellata in blocco. Un'ulteriore evoluzione delle ROM ha riguardato la possibilità di poter cancellare elettricamente le celle invece che usare l'esposizione a ultravioletti. Ciò è stato possibile realizzando un MOSFET avente il gate floating con una parte sovrapposta al drain e separata da questo da un ossido estremamente sottile, con spessore inferiore a 10 nm. Imponendo una tensione positiva dell'ordine di 10 V al control gate rispetto al drain si causa un passaggio di elettroni dal drain verso il floating gate attraverso l'ossido sottile. Tale conduzione avviene per un fenomeno fisico chiamato **effetto tunnel**, che consiste nel passaggio degli elettroni attraverso la barriera di potenziale dell'ossido.

separata da questo da un ossido estremamente sottile, con spessore inferiore a 10 nm. Imponendo una tensione positiva dell'ordine di 10 V al control gate rispetto al drain si causa un passaggio di elettroni dal drain verso il floating gate attraverso l'ossido sottile. Tale conduzione avviene per un fenomeno fisico chiamato **effetto tunnel**, che consiste nel passaggio degli elettroni attraverso la barriera di potenziale dell'ossido.

Il **processo è reversibile** per cui, imponendo una tensione positiva al drain rispetto al control gate, gli elettroni lasciano il floating gate. È così possibile caricare e scaricare il floating gate elettricamente. Per questo le memorie che impiegano questo dispositivo si chiamano EEPROM o E<sup>2</sup>PROM, acronimo di Electrically Erasable Programmable ROM.

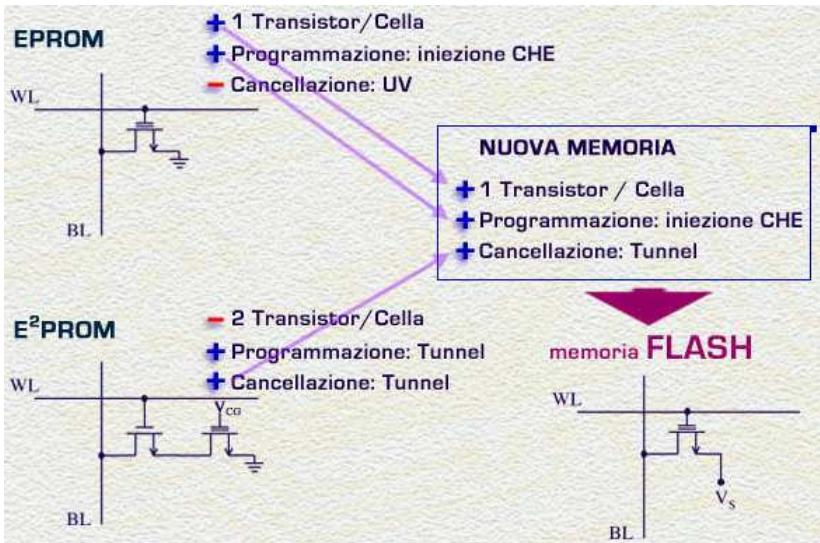
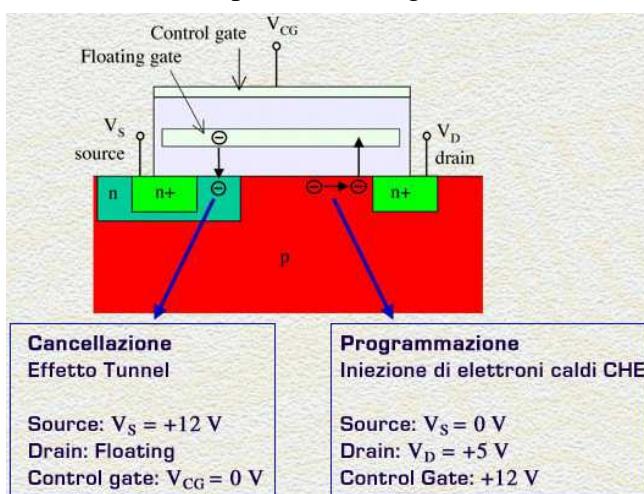
L'operazione di cancellazione tramite effetto tunnel può provocare una “**sovrascarica**” del floating gate, ovvero può caricare il floating gate positivamente. Ciò porterebbe la tensione di soglia a valori negativi ed il MOSFET sarebbe perciò sempre conduttivo anche per tensione nulla al control gate, ciò bloccherebbe la bit line permanentemente a massa. Per evitare che questo si verifichi la cella di una E<sup>2</sup>PROM è formata da due transistor: un MOSFET di accesso, comandato dalla word-line ed in serie a questa il MOSFET a doppio gate che si comporta quindi da interruttore aperto chiuso a



seconda che il suo floating gate sia carico di elettroni o meno.

## La memoria FLASH

Riassumiamo ora quanto abbiamo visto sulle memorie EPROM e E<sup>2</sup>PROM:  
 la EPROM possiede un solo transistor per cella a differenza della E<sup>2</sup>PROM la cui cella è costituita da due transistor. La cella EPROM si programma mediante iniezione di elettroni dal canale e si cancella esponendola a luce ultravioletta. Diversamente la cella E<sup>2</sup>PROM viene sia programmata che cancellata impiegando l'effetto tunnel tra drain e floating–gate. Il vantaggio della E<sup>2</sup>PROM di poter essere cancellata elettricamente è quindi accompagnato dallo svantaggio di avere due transistor per cella con la conseguente **riduzione di un fattore 2** della capacità di memoria per unità di area del chip. D'altra parte il vantaggio della EPROM di permettere densità doppia di celle ha per controparte il forte svantaggio di un metodo di cancellazione scomodo e che richiede tempi lunghi, dell'ordine di grandezza di decine di minuti. Costituirebbe quindi un passo in avanti una memoria programmabile e cancellabile elettricamente, come la E<sup>2</sup>PROM, ma avente un solo transistor per cella come la EPROM. La ricerca in questa direzione ha portato all'ideazione e alla realizzazione di una nuova memoria proprio con queste caratteristiche: tale memoria è chiamata FLASH per ragioni che vedremo tra poco. La cella della flash è costituita da un unico transistor, impiega lo stesso metodo di programmazione della EPROM ovvero l'iniezione di elettroni dalla regione di drain e permette la cancellazione per effetto tunnel come la E<sup>2</sup>PROM con la differenza che nella E<sup>2</sup>PROM il tunnel avviene verso il drain mentre nella flash avviene verso il source. Per questo nello schema della cella il source non è collegato a massa ma ad un potenziale V<sub>S</sub> che può essere cambiato come vedremo tra poco, esaminando il dispositivo. In figura vediamo una sezione del MOSFET a doppio gate di una memoria flash.



Come già accennato la programmazione viene fatta attraverso iniezione di elettroni ad alta energia, detti anche elettroni "caldi", dalla porzione di canale prossima al drain. Per far questo il MOSFET viene portato in forte conduzione imponendo delle tensioni relativamente alte al drain ed al control gate rispetto al source. Per la cancellazione il control–gate viene posto a 0 volt ed il source ad una tensione positiva relativamente alta in modo da favorire il tunnel degli elettroni dal floating gate verso il source. Si noti la asimmetria del MOSFET avendo il source una doppia regione N realizzata al fine di evitare fenomeni di rottura a valanga della giunzione source–substrato a causa della elevata tensione imposta al source in fase di cancellazione. Non è tuttavia possibile cancellare una sola cella alla volta in quanto questo richiederebbe l'accesso ad ogni source di ciascun MOSFET. Le celle di memoria flash sono perciò suddivise in più gruppi aventi ciascuno un elettrodo comune connesso al source di tutti i transistor del gruppo. In tal modo la cancellazione avviene contemporaneamente per tutte le celle appartenenti al gruppo selezionato. Per questo motivo e per il fatto che la cancellazione avviene in tempi rapidi le memorie sono state chiamate flash.

valanga della giunzione source–substrato a causa della elevata tensione imposta al source in fase di cancellazione. Non è tuttavia possibile cancellare una sola cella alla volta in quanto questo richiederebbe l'accesso ad ogni source di ciascun MOSFET. Le celle di memoria flash sono perciò suddivise in più gruppi aventi ciascuno un elettrodo comune connesso al source di tutti i transistor del gruppo. In tal modo la cancellazione avviene contemporaneamente per tutte le celle appartenenti al gruppo selezionato. Per questo motivo e per il fatto che la cancellazione avviene in tempi rapidi le memorie sono state chiamate flash.

<pagina bianca per data-sheet>

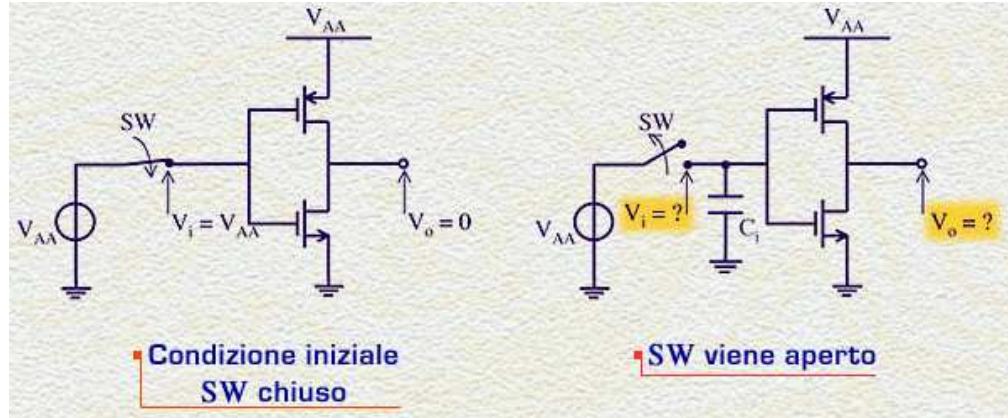
<pagina bianca per data-sheet>

## Lezione n. 22 : I circuiti bistabili

### Introduzione

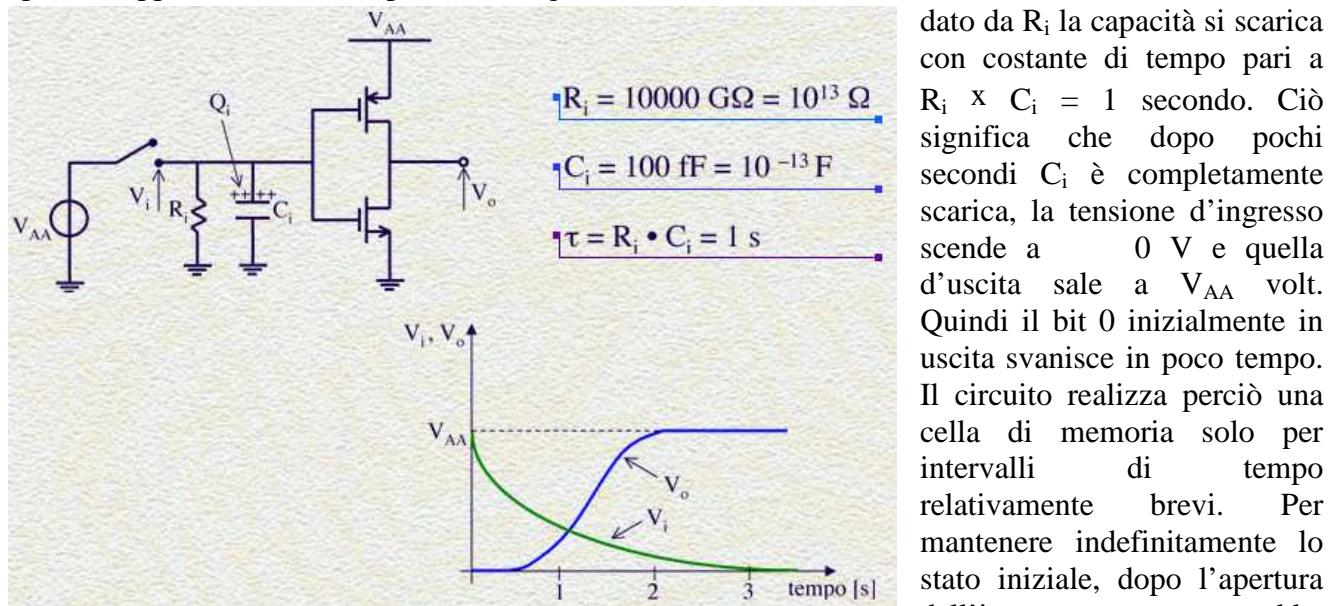
In questa lezione studieremo un particolare circuito in grado di memorizzare un bit. Le implementazioni di questo circuito si trovano in numerosi sistemi digitali, tra cui i registri, i contatori e alcuni tipi di memorie. Consideriamo l'invertitore in figura, supponiamo che inizialmente sia imposto un '1' logico in ingresso attraverso la chiusura dell'interruttore SW ed il generatore di tensione  $V_{AA}$ .

L'uscita sarà quindi a livello basso. Ora apriamo l'interruttore e domandiamoci: cosa succede al livello logico fornito in uscita dalla porta? Rimane 0 o cambia? Per rispondere a questa domanda dobbiamo sapere che



cosa succede alla tensione del nodo d'ingresso  $V_i$  dopo che l'interruttore viene aperto. Se la tensione  $V_i$  si mantiene a livelli prossimi a  $V_{AA}$ , allora l'uscita rimarrà a 0, altrimenti no. Per comprendere cosa accade, ricordiamo che l'impedenza d'ingresso di un invertitore può essere rappresentata da una capacità  $C_i$  verso massa. Quando l'interruttore viene chiuso il generatore carica la capacità  $C_i$  alla tensione  $V_{AA}$ . **Una volta aperto l'interruttore, la capacità  $C_i$  mantiene o meno il valore di tensione a seconda che la carica  $Q_i$  lascia o meno l'armatura del condensatore.** In un circuito idealizzato, come quello rappresentato in figura, la carica presente al nodo d'ingresso dell'invertitore non ha cammini conduttori verso massa e quindi l'ingresso rimane indefinitamente a tensione  $V_{AA}$  e conseguentemente l'uscita rimane a 0 V.

Tuttavia nella realtà esistono sempre cammini conduttori verso massa che provocano la scarica di  $C_i$ . Supponiamo ad esempio che vi sia un cammino verso massa con resistenza pari a  $10.000 \text{ G}\Omega$  ovvero  $10^{13} \Omega$ . In molti circuiti una tale resistenza sarebbe considerata del tutto equivalente ad un circuito aperto. Supponiamo che la capacità  $C_i$  valga  $100 \text{ fF}$  ovvero  $10^{-13} \text{ F}$ . Per via del cammino conduttivo dato da  $R_i$  la capacità si scarica con costante di tempo pari a  $R_i \times C_i = 1$  secondo. Ciò significa che dopo pochi secondi  $C_i$  è completamente scarica, la tensione d'ingresso scende a 0 V e quella d'uscita sale a  $V_{AA}$  volt.



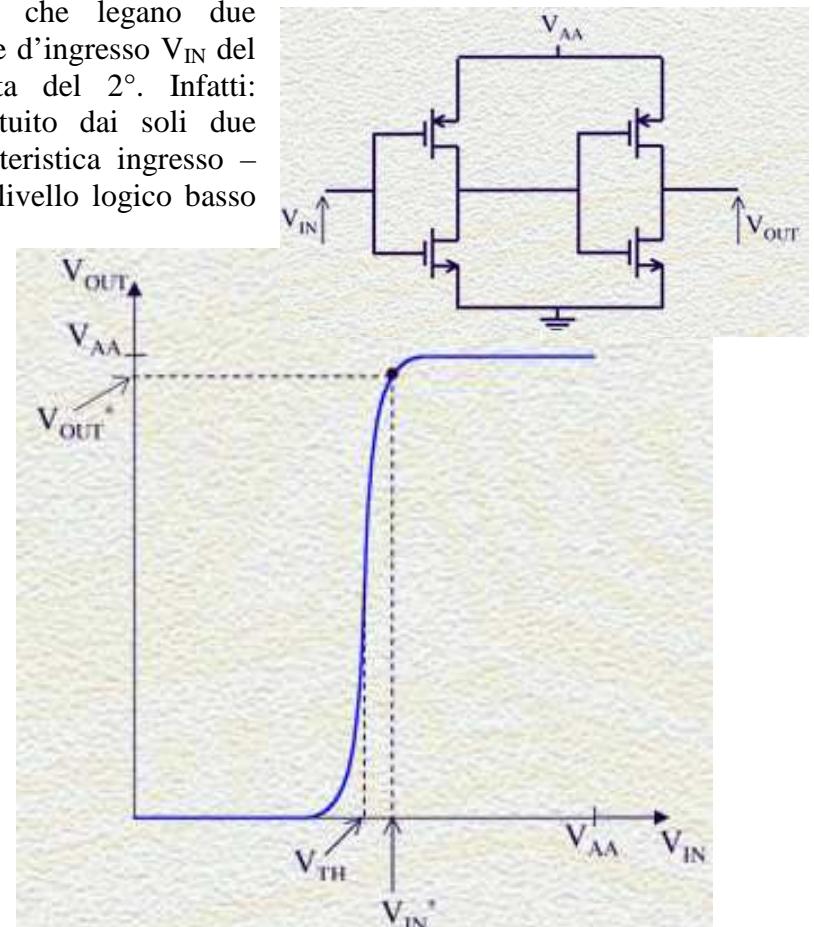
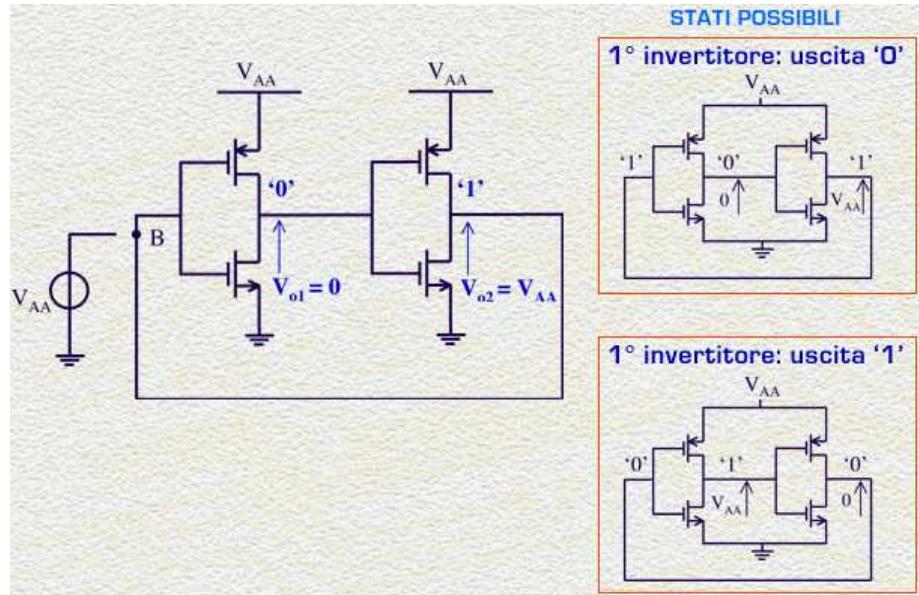
sufficiente che il circuito stesso possa generare il bit 1 fornendolo permanentemente all'ingresso, senza più bisogno del generatore.

Questo può essere realizzato aggiungendo un altro invertitore in cascata e riportando la sua uscita, '1' logico, all'ingresso del primo.

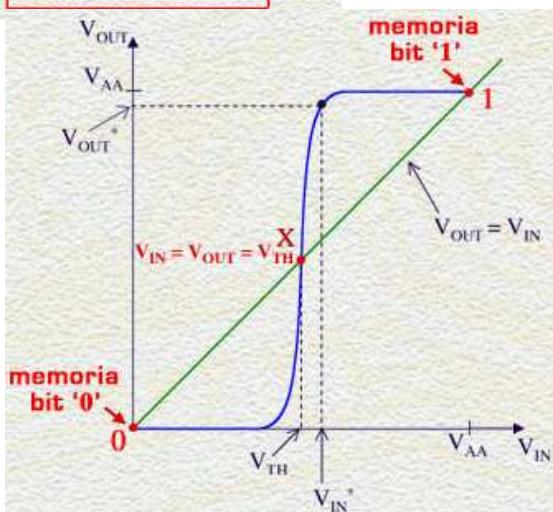
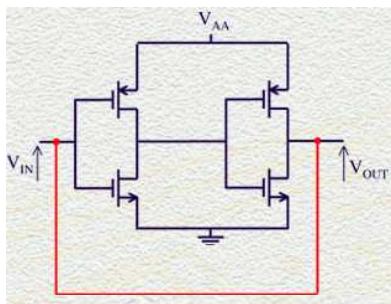
Esaminiamo il circuito mostrato. Supponiamo che inizialmente l'interruttore sia in posizione A, l'uscita del primo invertitore è 0 V e l'uscita del 2° è  $V_{AA}$  volt. Ora commutiamo la posizione dell'interruttore su B. Il primo invertitore si ritrova in ingresso ancora  $V_{AA}$  e quindi continuerà a fornire in uscita 0 volt. Il circuito non può accorgersi che ora la tensione  $V_{AA}$  al suo ingresso è generata da lui stesso attraverso il secondo invertitore. Pertanto il circuito così connesso si manterrà indefinitamente nello stato imposto inizialmente, comportandosi da cella di memoria. Il bit memorizzato può essere quello in uscita al 1° invertitore o quello complementare in uscita al secondo. Analogamente si può osservare che il circuito

può mantenersi indefinitamente anche nello stato complementare in cui l'uscita del primo invertitore è alta e l'uscita del secondo è bassa. Si noti che questa cella di memoria è volatile in quanto, se viene interrotta l'alimentazione al circuito, le capacità dei nodi tendono a scaricarsi a causa di correnti parassite e l'informazione quindi viene persa.

Il circuito di memoria visto è quindi costituito da due invertitori in cascata con un collegamento detto di **"retroazione"** dall'uscita del 2° invertitore all'ingresso del primo. Il fatto che questo circuito abbia degli stati di equilibrio stabili, corrispondenti alla memorizzazione del bit 0 e del bit 1, può anche essere visto considerando le relazioni che legano due particolari variabili del circuito: la tensione d'ingresso  $V_{IN}$  del 1° invertitore e  $V_{OUT}$  tensione d'uscita del 2°. Infatti: consideriamo dapprima il circuito costituito dai soli due invertitori, senza la retroazione. La caratteristica ingresso – uscita è quella mostrata: quando  $V_{IN}$  è a livello logico basso anche  $V_{OUT}$  è bassa e quando  $V_{IN}$  è alta anche  $V_{OUT}$  è alta. Infatti l'uscita logica di due invertitori in cascata è uguale all'ingresso. Osserviamo inoltre, nella caratteristica  $V_{OUT} - V_{IN}$ , la transizione tra i due livelli logici che avviene nell'intorno della tensione di soglia  $V_{TH}$  degli invertitori stessi. Imposto un certo valore  $V_{IN}^*$  all'ingresso del circuito, questo modificherà il suo stato fino a portare la sua tensione d'uscita al valore  $V_{OUT}^*$  fissato dalla caratteristica. In altre parole possiamo dire che, in condizioni statiche, quindi **NON di transitorio**, il circuito forza una relazione ben precisa tra i valori di  $V_{IN}$  ed i valori di  $V_{OUT}$ . Questa relazione è appunto la caratteristica rappresentata. In condizioni statiche quindi il punto di lavoro del circuito non può trovarsi al di fuori della curva caratteristica, bensì deve essere necessariamente sulla caratteristica.



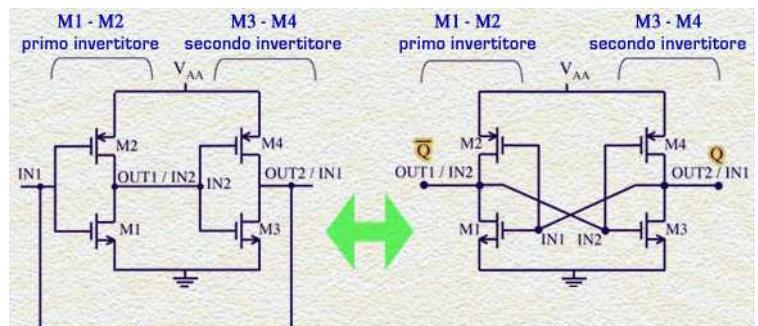
Ora aggiungiamo al circuito la connessione tra l'uscita del secondo invertitore e l'ingresso del primo.



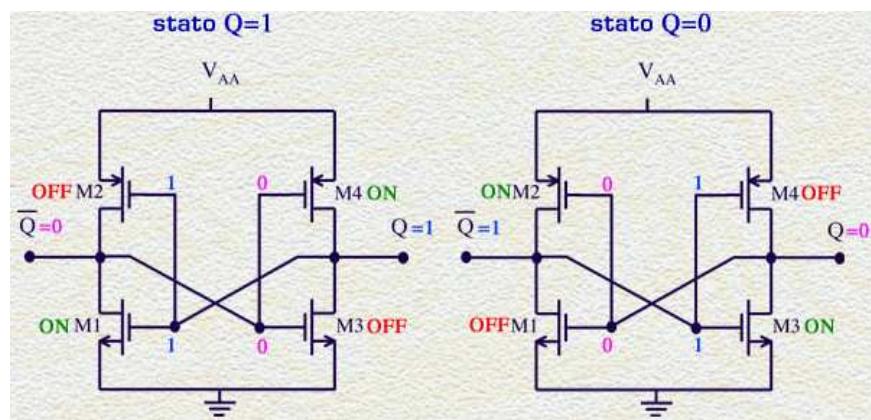
Questa connessione impone la relazione  $V_{OUT} = V_{IN}$  che sul piano è rappresentata dalla retta bisettrice. Il nuovo circuito, costituito dai due invertitori e dalla connessione ingresso–uscita, potrà avere come possibili valori di  $V_{IN}$  e  $V_{OUT}$ , in condizioni statiche, solo quelli che soddisfano entrambe le relazioni: quella del doppio invertitore e la  $V_{OUT} = V_{IN}$  della connessione. Gli unici punti che soddisfano entrambe le relazioni sono i punti di intersezione tra le due caratteristiche. Osserviamo che esistono 3 intersezioni: il circuito tende spontaneamente a portarsi in uno di questi 3 punti e lì può permanere indefinitamente finché non vi siano interventi che forzatamente lo inducono a cambiare stato. I tre punti vengono pertanto chiamati **punti di equilibrio**. Due di questi li avevamo già individuati; e sono il punto in cui  $V_{IN} = V_{OUT} = 0$  e il punto  $V_{IN} = V_{OUT} = V_{AA}$ . Questi due punti rappresentano gli stati di memoria del bit ‘0’ e del bit ‘1’. Indichiamoli quindi con 0 e 1. Il terzo punto di equilibrio, indicato con X, corrisponde a  $V_{IN}=V_{OUT}=V_{TH}$ . Vi è una differenza sostanziale tra i due punti 0 e 1 rispetto al punto X. Il **punto 0 ed il punto 1** rappresentano **stati di equilibrio stabile** del circuito mentre **il punto X** rappresenta invece un **punto di equilibrio instabile**. Vediamo cosa significa: assumiamo che il circuito si trovi nel punto di lavoro ‘1’ e che, a causa

di un’interferenza, o rumore, di durata istantanea, il valore di  $V_{IN}$  o quello di  $V_{OUT}$  cambino portando quindi il circuito ad operare in un punto del piano diverso, sebbene vicino al punto ‘1’. Ciò che si verifica è che il circuito tende spontaneamente a riportarsi nel punto ‘1’ iniziale. La stessa cosa si potrebbe dire se lo stato di partenza fosse quello del punto ‘0’. Se invece lo stato iniziale fosse il punto X allora anche un minimo scostamento da X provocherebbe l’allontanamento progressivo del punto di lavoro da X. Il circuito si porterebbe nel punto ‘0’ o nel punto ‘1’ a seconda che lo scostamento da X sia avvenuto con un avvicinamento verso lo ‘0’ o verso l’‘1’.

Per questo motivo il punto X viene detto di equilibrio instabile, mentre i due punti ‘0’ e ‘1’ di equilibrio stabile. In pratica quando si dà tensione di alimentazione ad un circuito di questo tipo, esso si porta in uno dei due punti stabili. Il punto X difficilmente viene raggiunto e mantenuto nel tempo perché qualsiasi scostamento da esso, dovuto anche a rumore di bassa intensità, comporterebbe il definitivo allontanamento dal punto stesso. Il circuito visto ha quindi due stati di equilibrio stabile che possono essere mantenuti indefinitamente nel tempo, finché permane l’alimentazione. Il circuito viene comunemente chiamato ‘Latch’, traducibile in italiano con ‘saliscendi’, in analogia all’antico sistema di chiusura delle imposte con un gancio che può stare in due posizioni stabili. Un altro nome, dato a questo circuito è ‘bistabile’, in virtù del fatto di possedere due stati stabili. Il bistabile viene comunemente rappresentato, per ragioni di comodità che saranno chiare in seguito, con gli ingressi dei due invertitori affacciati tra loro, come rappresentato in figura. Il terminale indicato con **OUT1/IN2** è connesso con l’uscita del primo invertitore e l’ingresso del secondo, mentre il terminale **OUT2/IN1** è connesso con l’uscita del 2° invertitore e l’ingresso del primo. Nei due punti di equilibrio stabile questi due terminali assumono valori logici complementari quindi verranno di seguito indicati con le variabili  $\bar{Q}$  e Q. Infatti  $\bar{Q}$  è l’uscita del 1° invertitore il cui ingresso è Q. Analogamente Q è l’uscita del 2° invertitore il cui ingresso è  $\bar{Q}$ . Q rappresenta quindi il *bit memorizzato dal circuito, insieme al suo complemento  $\bar{Q}$* .

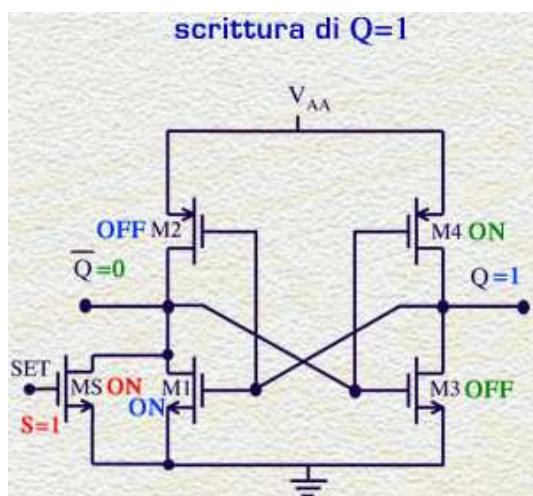


Esaminiamo ora lo stato dei transistor nei due stati di equilibrio stabile del circuito. Nello stato corrispondente a  $Q = 1$ , M3 è spento ed M4 è acceso al fine di tenere l'uscita Q alta. L'uscita Q alta tiene acceso M1 e spento M2 per cui l'uscita  $\bar{Q}$  sarà bassa. Nello stato  $Q = 0$ , M3 è acceso e M4 è spento coerentemente con Q basso, il quale a sua volta tiene spento M1 e acceso M2 stabilendo quindi l'uscita  $\bar{Q}$  alta. Vediamo ora come si deve agire sul circuito per portarlo in uno di questi due stati, in altre parole come fare per scrivere il bit  $Q = 1$  o il bit  $Q = 0$  nella cella.



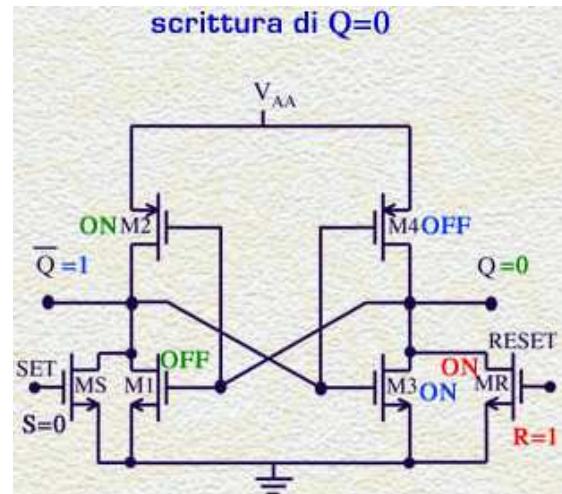
### I bistabili S-R, (SET-RESET)

Iniziamo proponendoci di scrivere  $Q = 1$ . Per portare il circuito in questo stato, qualora già non lo sia, è sufficiente forzare lo spegnimento di M3 e l'accensione di M4. Ciò può essere fatto forzando a massa il gate comune di M3 e M4, ovvero l'elettrodo  $\bar{Q}$ , attraverso l'accensione di un MOSFET MS comandato da un elettrodo d'ingresso chiamato **SET**. Quando SET viene posto alto il MOSFET MS diviene conduttivo portando a potenziale basso il nodo  $\bar{Q}$ , qualora questo si trovi alto. Si noti che, se



il nodo  $\bar{Q}$  si trova alto significa che il MOSFET M2 è conduttivo, quindi, per portare  $\bar{Q}$  basso è necessario che il MOSFET MS sia molto più conduttivo di M2. Vedremo successivamente una soluzione circuitale che non richiede questa condizione su MS. Notiamo ora che portare  $\bar{Q}$  basso attraverso MS provoca lo spegnimento di M3 e l'accensione di M4. M3 spento ed M4 acceso portano Q alto. Abbiamo così mostrato che quando SET viene posto a 1, Q va alto. Inoltre con Q alto M1 si accende e M2 si spegne, condizione coerente con l'uscita  $\bar{Q}$  bassa. In definitiva SET = 1 ha forzato il circuito a stabilirsi nello stato di equilibrio stabile  $Q = 1$ ,  $\bar{Q} = 0$ . A questo punto la variabile SET può essere posta a 0 spegnendo così il transistor MS. Il circuito non si sposterà dal punto di equilibrio raggiunto, infatti pur spegnendo MS il nodo  $\bar{Q}$  è tenuto a massa dal transistor M1 che rimane conduttivo. In conclusione **la scrittura del bit  $Q = 1$  nella cella è avvenuta ponendo  $SET = 1$  per un tempo sufficiente al circuito a raggiungere il corrispondente stato di equilibrio**.

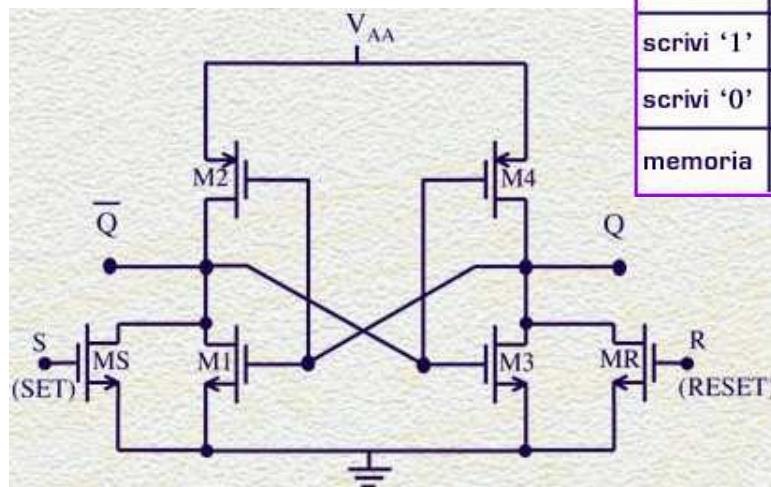
Un analogo discorso può essere fatto per la scrittura del bit  $Q = 0$  nella cella. In tal caso viene forzato a massa il nodo Q attraverso un MOSFET MR comandato da una variabile logica chiamata **RESET** che viene posta alta. Contemporaneamente la variabile SET viene lasciata a livello basso. Con RESET alto Q va basso, M1 si spegne ed M2 si accende portando  $\bar{Q}$  alto.  $\bar{Q}$  alto, a sua volta accende M3 e spegne M4 confermando quindi il valore basso di Q. Raggiunta questa situazione la variabile RESET può essere riportata a livello basso, senza che il circuito si muova dallo stato di equilibrio raggiunto.



Ricapitolando: per poter scrivere il bit nella cella abbiamo aggiunto due MOSFET, MS e MR, al circuito del bistabile. I gate di MS e MR ricevono rispettivamente i segnali di SET e RESET impiegati per svolgere tre funzioni:

- **scrittura del bit ‘1’**; per scrivere il bit ‘1’, si pone SET = 1 e RESET = 0 il risultato è appunto  $Q = 1$ .
- **scrittura del bit ‘0’**; per scrivere il bit ‘0’ si pone SET = 0 e RESET = 1, così che  $Q$  diventi ‘0’.
- **fase di memoria**; la fase di memoria si realizza ponendo SET e RESET entrambi a 0. In tal caso infatti i MOSFET’s MS e MR sono tutte due spenti ed il bistabile rimane indefinitamente nello stato precedentemente impostato.

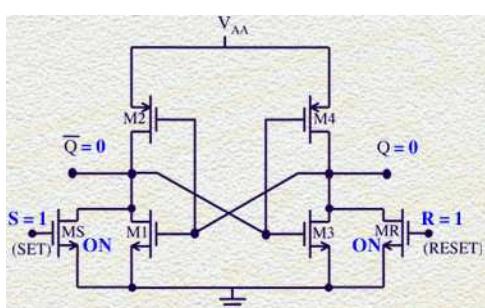
*Il circuito viene chiamato latch o bistabile S-R.*



Funzione	S (Set)	R (Reset)	Q	$\bar{Q}$
scrivi ‘1’	1	0	1	0
scrivi ‘0’	0	1	0	1
memoria	0	0	valori precedentemente scritti	

LO STATO SET=1, RESET=1

Possiamo ora chiederci che cosa accade nel caso in cui sia SET che RESET vengano posti entrambi a valore 1. Possiamo già premettere che un tale comando è di per sé contraddittorio in quanto SET=1 è



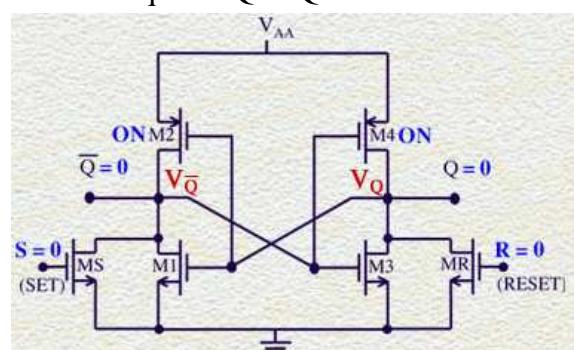
il comando per scrivere nella cella  $Q = 1$  mentre RESET = 1 è il comando per scrivere  $Q = 0$ . Pertanto SET e RESET entrambi 1 che effetto producono? Osservando il circuito possiamo comprenderlo. SET = 1 accende il MOSFET MS portando il nodo  $\bar{Q}$  verso massa, a sua volta RESET = 1 accende il MOSFET MR portando verso massa il nodo Q. Quindi il risultato è  $Q = \bar{Q} = 0$  che evidentemente è in contraddizione con i risultati degli altri casi esaminati, in cui

l’uscita  $\bar{Q}$  del circuito fornisce sempre il complemento logico dell’uscita Q. Concludiamo perciò che la situazione con segnali d’ingresso **SET e RESET alti è da escludere**. Possiamo inoltre evidenziare un altro effetto legato alla situazione SET e RESET entrambi alti e quindi  $Q = \bar{Q} = 0$ . Consideriamo questo come stato iniziale del circuito e vediamo cosa succede una volta che si riportano SET e RESET a 0, ovvero si riporta il circuito in condizione di memoria.

Domandiamoci: Q e  $\bar{Q}$  rimangono a 0 indefinitamente?

La risposta è no! perché lo stato  $Q = \bar{Q} = 0$  non è uno dei due stati di equilibrio stabile del circuito latch. Quindi il circuito inizierà un transitorio che lo porterà in uno dei due stati di equilibrio stabile che sono  $Q = 0$  e  $\bar{Q} = 1$  oppure l’altro stato  $Q = 1$  e  $\bar{Q} = 0$ .

Chiediamoci: in quale dei due stati si porterà il circuito? A tale proposito osserviamo che la situazione di partenza con  $Q$  e  $\bar{Q} = 0$  e  $S = R = 0$  vede tutti i transistor spenti tranne M2 e M4.



Questi inizieranno quindi a caricare le capacità associate ai nodi  $\bar{Q}$  e Q. Le tensioni  $V_{\bar{Q}}$  e  $V_Q$  inizieranno quindi a salire. Nella realtà la velocità con cui  $V_{\bar{Q}}$  e  $V_Q$  cresceranno non è identica perché dei due transistor, M2 e M4, ce ne sarà uno più conduttivo dell'altro, inoltre le capacità dei nodi Q e  $\bar{Q}$  non saranno mai perfettamente identiche. Se, ad esempio, la tensione al nodo Q sale più velocemente, allora raggiunta la tensione di soglia del MOSFET M1, questo si accenderà mentre M3 sarà ancora spento. L'accensione di M1 tende a rallentare ulteriormente la salita del nodo  $\bar{Q}$ . Intanto Q continua a salire, M1 diviene sempre più conduttivo ed M2 sempre meno. In definitiva la maggior velocità di salita della tensione  $V_Q$  porta il circuito verso lo stato  $Q = 1$  e  $\bar{Q} = 0$ .

Se invece salisse più velocemente  $V_{\bar{Q}}$  allora il circuito si porterebbe nello stato complementare. Pertanto si può dire che lo stato che viene raggiunto dipende dalle asimmetrie del circuito che possono essere dovute a

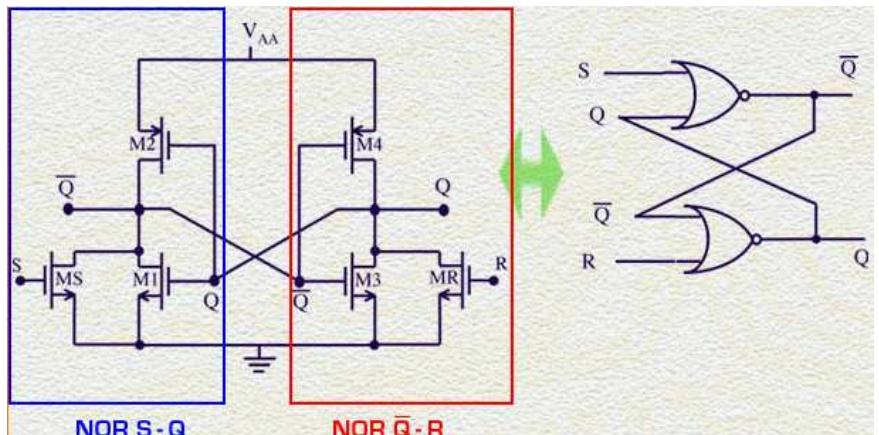
- 1) differenze, anche minime, tra i transistor M2 e M4
- 2) differenze nel carico capacitivo dei nodi Q e  $\bar{Q}$  oppure anche a tempi leggermente diversi in cui SET e RESET vengono riportati a livello basso.

Quindi, in conclusione, *l'ingresso SET = 1 e RESET = 1 produce uno stato del circuito che non è di alcuna utilità, in quanto porta il circuito inizialmente in uno stato contraddittorio e successivamente in uno stato fissato da condizioni anche imponderabili.*

Osserviamo il circuito del latch S-R. L'invertitore M1 – M2 con l'aggiunta di MS svolge una funzione simile ad una porta NOR avente per ingressi S e Q e come uscita  $\bar{Q}$ . Infatti se S oppure Q sono alti l'uscita  $\bar{Q}$  viene portata bassa.

Lo stesso si può dire del 2° invertitore M3 – M4 con l'aggiunta di MR. Il circuito del Latch – SR è quindi costituito da due porte NOR in cui l'uscita della prima è connessa ad uno degli ingressi della seconda, come mostrato in figura.

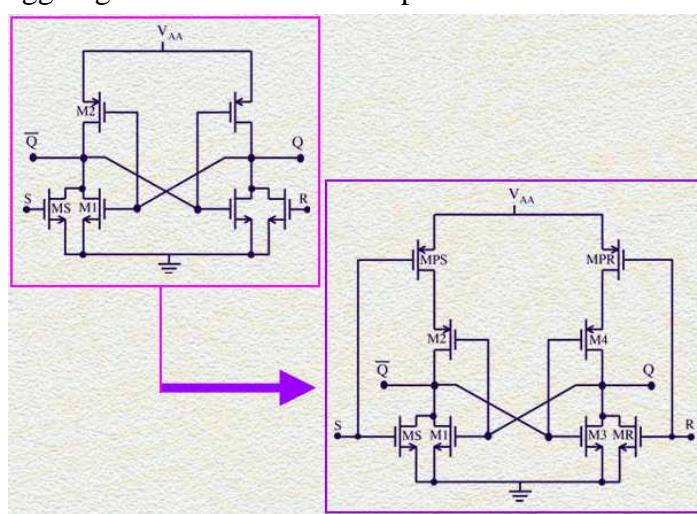
Nel circuito appena discusso le porte NOR non hanno una topologia CMOS in quanto i



transistori MS ed MR non hanno i rispettivi PMOS. Possiamo quindi modificare il circuito aggiungendo i MOSFET complementari di MS e MR come mostrato in figura realizzando così un latch SR con topologia CMOS.

Rispetto al circuito precedente, quello CMOS ha il vantaggio di non presentare mai, almeno in prima approssimazione, un cammino conduttivo dall'alimentazione verso massa, come si poteva verificare nel precedente circuito quando, con  $\bar{Q}$  alto, ovvero M2 acceso, veniva portato il SET alto, accendendo MS e provocando quindi una momentanea conduzione attraverso M2 ed MS.

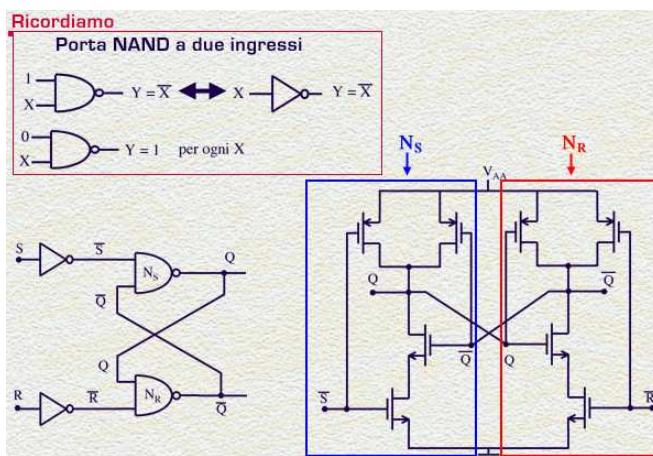
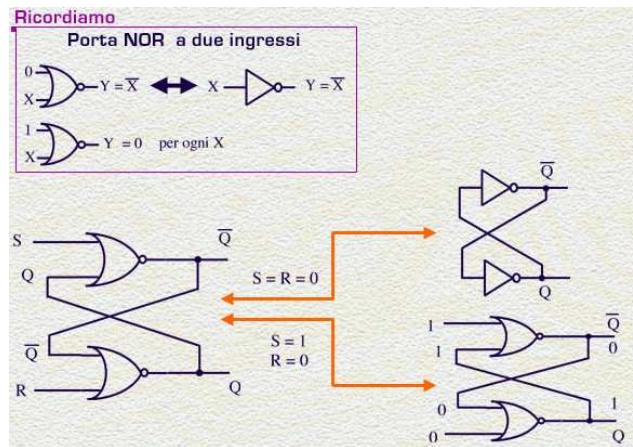
Con porte CMOS ciò non si verifica perché quando si pone SET alto il transistor MPS si spegne non permettendo quindi passaggio di corrente in M2. In figura quindi vediamo il circuito completo del bistabile SR in tecnologia CMOS e realizzato con porte NOR.



Il fatto che due porte NOR realizzino la cella SR di memoria può essere visto direttamente nel seguente modo. Ricordiamo che una porta NOR con uno dei due ingressi posto a 0, si comporta come un invertitore rispetto all'altro ingresso. Quindi il Latch con SET e RESET a 0 è equivalente al circuito di due invertitori, costituente appunto la cella di memoria.

Viceversa, quando una porta NOR ha uno dei due ingressi alto allora la sua uscita è 0, indipendentemente dal valore dell'altro ingresso. Questo ci spiega il caso di SET = 1 e RESET = 0 oppure il caso opposto di SET = 0 e RESET = 1.

È semplice mostrare come si possa realizzare un bistabile SR anche con porte NAND. Basti

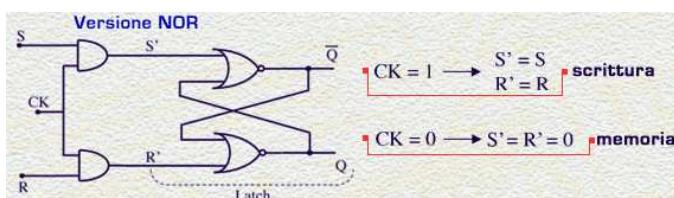


osservare che una NAND con uno dei due ingressi posto a 1 è equivalente ad un invertitore rispetto all'altro ingresso. Quando invece uno dei due ingressi è a 0, l'uscita è pari a 1 indipendentemente dal valore dell'altro ingresso. Quindi un bistabile SR può essere realizzato come mostrato in figura. I due invertitori agli ingressi S e R sono necessari affinché la condizione di memoria si verifichi per S=R=0, come per il bistabile a porte NOR.

In figura possiamo vedere lo schema circuitale costituito dalle due porte NAND, esclusi gli invertitori d'ingresso. La prima NAND,  $N_S$ , ha

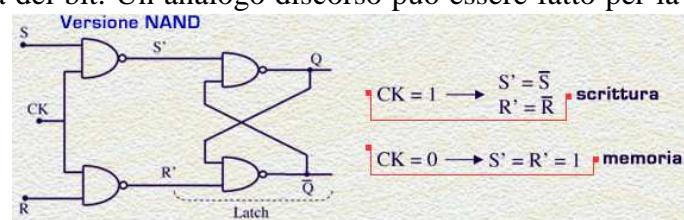
## I circuiti Flip-Flop

I bistabili SR trovano impiego in diversi sistemi digitali svolgendo in generale la funzione di memorizzare un bit per un certo periodo di tempo. In molti sistemi è richiesto che ogni operazione di scrittura o di tenuta di un bit in un bistabile avvenga solo in determinati istanti di tempo stabiliti da un segnale di temporizzazione detto segnale di clock. Per questo motivo il circuito del bistabile SR viene modificato affinché i segnali di SET e RESET giungano al latch solo



in corrispondenza del segnale di clock alto. Per far ciò è sufficiente far precedere il latch da due porte AND con un ingresso collegato al clock. Quando il segnale di clock è alto, le AND si comportano come buffer non invertenti per cui  $S' = S$  e  $R' = R$ . Quando invece il clock è basso  $S'$  e  $R'$  sono a 0 indipendentemente da  $S$  e  $R$ : il latch rimane quindi in fase di memoria. In sintesi il clock alto abilita la scrittura mentre il clock basso comporta la tenuta del bit. Un analogo discorso può essere fatto per la versione del latch con porte NAND. In tal caso il latch viene fatto precedere da porte NAND comandate su un ingresso dal clock.

Quando il clock è alto le NAND si comportano da invertitori rispetto ai bit  $S$  e  $R$  quindi  $S' = \bar{S}$  e  $R' = \bar{R}$  ed il latch viene impostato. Quando invece il clock è basso le NAND d'ingresso forniscono  $S' = 1$  e  $R' = 1$  ed il latch rimane in condizione di memoria. In entrambe le versioni questi bistabili, che operano in sincronismo con il segnale di clock, vengono chiamati FLIP-FLOP S-R.



Avevamo precedentemente visto, per il latch S-R, che l'ingresso  $S = 1$  e  $R = 1$  genera una situazione contraddittoria, ovvero  $Q = \bar{Q} = 0$ . Lo stesso avviene per il flip-flop SR mentre il clock è alto, essendo il circuito equivalente al latch. La situazione  $Q = \bar{Q} = 0$  si mantiene finché il clock è alto, poi, quando il clock va basso, il circuito del latch tende a portarsi in uno dei due suoi possibili stati di equilibrio stabile,  $Q = 1$  oppure  $Q = 0$ , a seconda delle asimmetrie del circuito. Pertanto, **l'ingresso  $S = 1$ ,  $R = 1$ , non producendo alcun risultato utile, viene ritenuto ingresso non valido per il flip-flop S-R.**

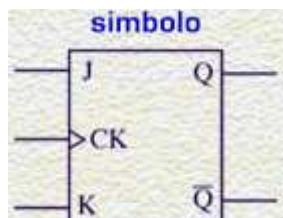
È possibile, con una semplice modifica circuitale, fare in modo che anche l'ingresso  $S = R = 1$  produca un risultato utile. Consideriamo la tavola dei valori del Flip – Flop in cui  $Q_n$  rappresenta il

Funzione	S	R	$Q_n$	$\bar{Q}_n$
scrivi '1'	1	0	1	0
scrivi '0'	0	1	0	1
memoria	0	0	$Q_{n-1}$	$\bar{Q}_{n-1}$
scrivi il complemento	1	1	$\bar{Q}_{n-1}$	$Q_{n-1}$

valore di Q in corrispondenza dell'ennesimo impulso di clock. Faremo in modo che l'ingresso  $S=R=1$  provochi, in corrispondenza del clock, una commutazione dell'uscita Q, ovvero che Q all'istante  $n$  sia uguale al negato di Q all'istante di clock precedente  $n-1$ . Esemplicando se al tempo di clock  $n-1$  il valore  $Q_{n-1}$  era 0, al successivo tempo di clock l'uscita Q commuterà a 1. Lo stesso vale per l'uscita  $\bar{Q}$ . Vediamo come realizzare questa

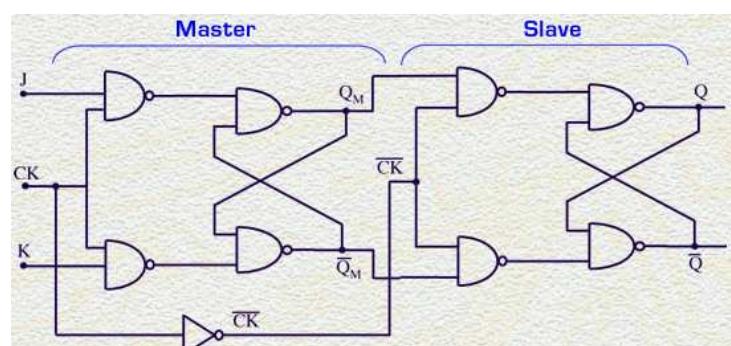
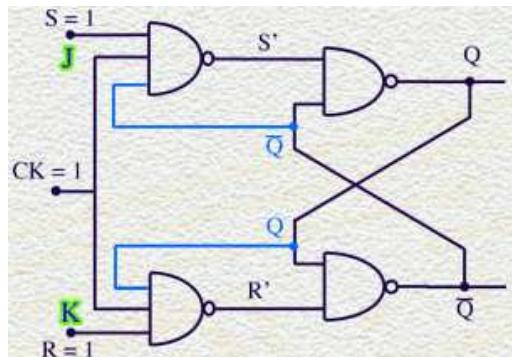
funzione a partire dal Flip – Flop S-R in condizione  $S=R=Clock=1$ . Affinché Q e  $\bar{Q}$  commutino valore e ricordando che Q assume il valore di S' negato, allora è sufficiente rendere S' uguale a Q ed R uguale a  $\bar{Q}$ .

Per far questo è sufficiente prelevare le uscite  $\bar{Q}$  e Q e riportarle ad un ingresso aggiuntivo delle porte NAND. Infatti quando  $S = 1$  e  $CK = 1$  la NAND si comporta da buffer invertente rispetto al 3° ingresso e riporta quindi in  $S'$  il valore Q. Analogamente avviene per  $R'$ , che assume il



valore di  $\bar{Q}$ . Un Flip-Flop di questo tipo viene distinto dal precedente S-R cambiando il nome di S, che diventa J, e di R che diventa K. Il circuito è denominato Flip-Flop J-K. Il circuito visto presenta tuttavia un problema quando si pone  $J=K=1$ . Infatti in corrispondenza del clock alto il circuito commuta in uscita e tale commutazione viene sentita al terzo ingresso delle porte NAND. Pertanto la commutazione delle uscite fa commutare a sua volta  $S'$  e  $R'$  che a loro volta fanno ricommutare  $\bar{Q}$  e Q. Le uscite così continuano a commutare fintanto che il clock è alto, il periodo di commutazione dipende dai tempi di propagazione delle porte. Quando il clock va basso Q e  $\bar{Q}$  non possono più cambiare ma il loro valore dipende da quanto è rimasto alto il clock rispetto ai tempi di propagazione. Non è detto quindi che l'uscita sia il complemento dei valori all'impulso di clock precedente. Questo effetto di commutazioni multiple viene denominato '**Race-around condition**', ovvero 'condizione di inseguimento', in quanto la commutazione dell'uscita influenza gli ingressi che a loro volta fanno ricommutare l'uscita e così via.

Per risolvere questo problema è necessario che la commutazione delle uscite non modifichi gli ingressi del Flip – Flop. Per ottenere questa condizione è stata ideata la configurazione mostrata in figura e costituita da due Flip – Flop S-R in cascata, di cui il primo, detto **MASTER**, è comandato direttamente dal segnale di clock mentre il secondo detto **SLAVE** è pilotato dal segnale di clock complementare. Ne risulta quindi che i due Flip – Flop sono **abilitati in tempi diversi**: quando il clock è alto è abilitato il MASTER mentre lo SLAVE è disabilitato, quando il clock è basso invece è lo SLAVE ad essere abilitato mentre il MASTER è disabilitato.



Vediamo dapprima il funzionamento del circuito nel caso  $J = 1$  e  $K = 0$  corrispondente alla **funzione scrivi Q = 1.**

Quando il clock va alto viene scritto  $Q_M = 1$  nel MASTER mentre lo SLAVE rimane in fase di memoria in quanto il clock negato è basso. Nel momento in cui il clock va basso, il latch del MASTER viene portato in stato di memoria mentre viene abilitato in scrittura lo SLAVE in cui viene quindi scritto il valore di  $Q_M$ , pari a 1.

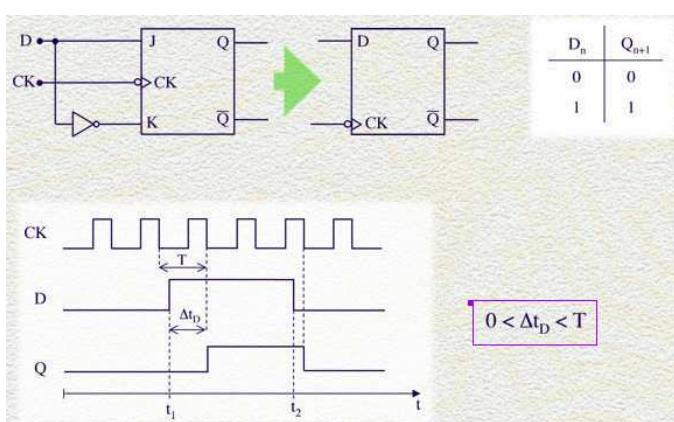
Notiamo che l'uscita  $Q$  assume il valore 1 presente all'ingresso  $J$  solo in corrispondenza del fronte di discesa del clock. Per realizzare un completo Flip – Flop J-K, che operi anche per  $J = K = 1$ , vengono collegate le uscite  $Q$  e  $\bar{Q}$  dello SLAVE alle porte NAND d'ingresso del MASTER.

L'inconveniente delle commutazioni multiple in uscita non si verifica: vediamo perché. Quando il clock va alto e  $J = K = 1$ , il MASTER commuta le sue uscite  $Q_M$  e  $\bar{Q}_M$ . Dato che lo SLAVE è disabilitato i valori  $Q$  e  $\bar{Q}$  presenti all'ingresso del MASTER non cambiano. Quando poi il clock va basso i valori di  $Q_M$  e  $\bar{Q}_M$  vengono memorizzati nello SLAVE comportando la commutazione delle uscite  $Q$  e  $\bar{Q}$ . Queste però non influenzano ulteriormente l'ingresso perché il MASTER è disabilitato avendo il clock basso.

Vediamo quindi che la condizione “race – around” non si verifica e l'uscita commuta una sola volta entro un periodo di clock.

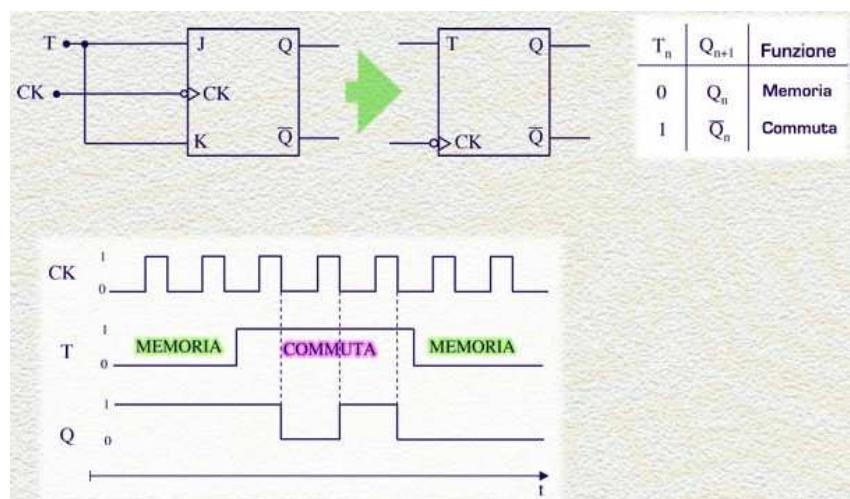
Vediamo ora due Flip – Flop particolari che vengono realizzati a partire dalle configurazioni già viste. Il primo è il **Flip-Flop D** che viene ottenuto a partire dal J-K MASTER-SLAVE in cui un unico segnale D pilota direttamente l'ingresso J e attraverso un invertitore l'ingresso K. In tal modo J e K sono sempre valori complementari e le due condizioni  $J = K = 1$  o  $J = K = 0$  non sono quindi mai verificate. Vediamo che funzione può svolgere questo circuito. Ricordiamo che quando J e K sono complementari, l'uscita Q del latch assume il valore impostato all'ingresso J, nel nostro caso D. Quindi questo flip-flop presenterà all'uscita Q il dato d'ingresso D. Il circuito si comporta quindi da buffer sincronizzato in quanto il dato D appare all'uscita Q solo in sincronismo con il segnale di clock. Ricordiamo infatti che nel flip-flop master-slave l'uscita può modificarsi solo in corrispondenza alla discesa dell'impulso di clock. Facciamo un esempio: in figura è rappresentato il treno di impulsi costituenti il segnale di clock. Il segnale D è inizialmente basso e l'uscita Q del flip-flop uguaglia D. Ad un certo istante  $t_1$  D cambia da 0 a 1. Q non cambia istantaneamente ma solo al completamento

dell'impulso di clock immediatamente seguente  $t_1$ . Infatti è necessario che il clock vada alto affinché il dato D venga memorizzato nel MASTER e che poi vada basso affinché venga trasferito allo SLAVE d'uscita. Il segnale Q passa a 1 esattamente in corrispondenza della discesa dell'impulso di clock successivo a  $t_1$ . All'istante  $t_2$  D torna basso e Q segue D tornando basso in corrispondenza del primo fronte di discesa del clock. Vediamo quindi la funzione svolta da questo circuito. Il circuito ha ricevuto un segnale con fronti di salita e discese



posti in modo arbitrario nel tempo e ha fornito in uscita un segnale identico ma con i fronti sincronizzati con i fronti di discesa del segnale di clock. Il segnale d'uscita risulta quindi sincronizzato con il clock e, nel caso specifico risulta quindi avere una durata pari ad un multiplo esatto di periodi di clock. Notiamo inoltre che il segnale d'uscita risulta ritardato nel tempo rispetto al segnale d'ingresso di un intervallo di tempo  $0 < \Delta t_D < T$ , compreso tra 0 il periodo T di clock. Per questo il Flip – Flop viene denominato **D** che sta per **DELAY** ('ritardo' in inglese). Il Flip – Flop D viene utilizzato come l'elemento di memoria dei registri a scorrimento.

Un altro tipo di Flip – Flop derivato dal J–K MASTER – SLAVE si ottiene connettendo insieme gli ingressi J e K che quindi costituiscono un unico ingresso T (Toggle). L'ingresso  $T = 0$  equivale alla condizione  $J = K = 0$  nella quale il Flip – Flop è in stato di memoria e quindi con le uscite invarianti nel tempo. Se invece  $T = 1$  si verifica la condizione  $J = K = 1$  nella quale il Flip – Flop commuta ad ogni discesa dell'impulso di clock. In figura è rappresentato sul diagramma temporale quanto illustrato. Finché  $T$  è basso, l'uscita Q del flip-flop è fissata. Quando  $T$  viene posto alto, l'uscita Q inizia a commutare ripetutamente in corrispondenza ad ogni fronte di discesa del clock, finché  $T$  non viene posto di nuovo basso portando il flip-flop in posizione di memoria. Questo circuito viene chiamato Flip–Flop T e viene utilizzato come elemento base per la realizzazione dei contatori.



<pagina bianca per data-sheet>

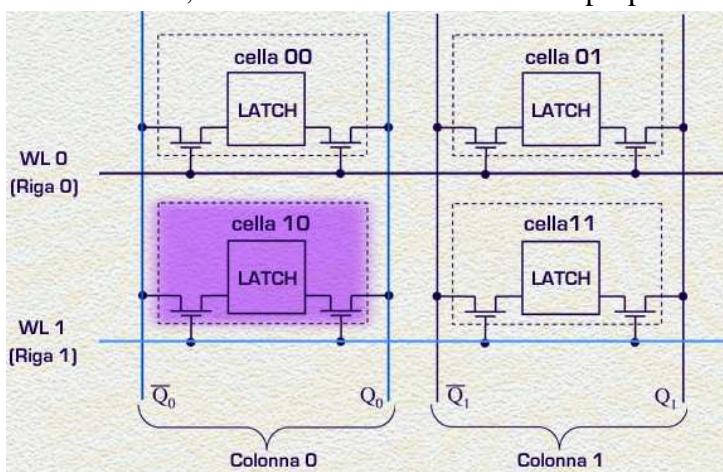
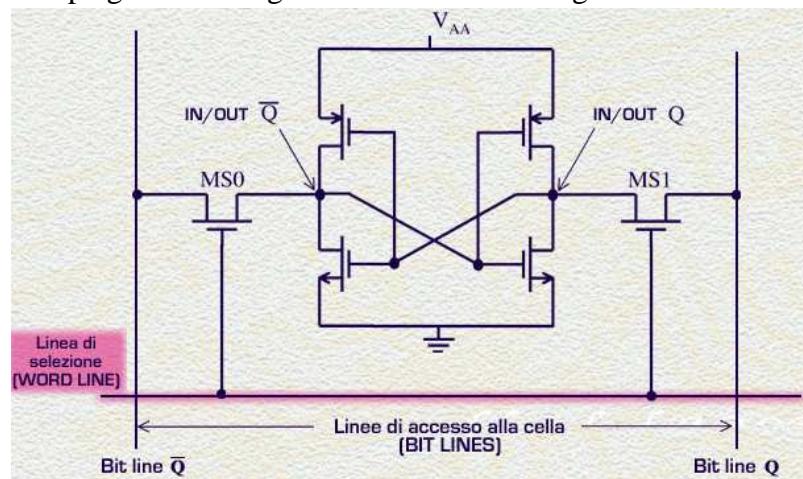
## Lezione n. 23 : Le memorie RAM

### RAM statiche (SRAM)

In una precedente lezione abbiamo studiato il circuito bistabile, o latch, mostrando la sua proprietà di memorizzare un bit. Possiamo quindi pensare di utilizzare il latch come cella elementare per realizzare memorie di molti bit. Per far ciò possiamo impiegare la configurazione mostrata in figura. Osserviamo il latch i cui nodi di ingresso/uscita **IN/OUT  $\bar{Q}$**  e **IN/OUT Q**, nodi a cui occorre accedere per scrivere o leggere il bit memorizzato, vengono collegati a due linee, dette appunto bit-lines, attraverso i due MOSFET MS0 e MS1, che chiameremo **MOSFET di selezione della cella**. Questi due MOSFET sono comandati sul gate da un segnale portato dalla *linea di selezione* o WORD LINE.

Quando il segnale sulla WORD LINE è basso i MOSFET's di selezione sono

spenti e perciò la cella di memoria è isolata e svolge la sua funzione di tenuta del bit. Quando la WORD LINE viene portata alta, allora i MOSFET di selezione divengono conduttori e sulla bit-line Q si può leggere il bit memorizzato nella cella. È ovviamente anche disponibile il suo complementare sulla bit-line  $\bar{Q}$ . La lettura della cella richiede quindi semplicemente di portare alta la WORD LINE e prelevare il dato dalla bit-line. I due MOSFET di selezione agiscono quindi come porte di trasmissione dalla cella alle bit-lines. Se invece si vuole scrivere un bit nella cella, ad esempio  $Q = 1$ , allora è sufficiente forzare a livello alto la bit-line Q e contemporaneamente forzare a livello basso la bit-line  $\bar{Q}$ . Poi si seleziona la cella portando alta la WORD LINE ed accedendo così i MOSFET di selezione. Il latch viene quindi forzato a porsi nello stato di equilibrio  $Q = 1$ ,  $\bar{Q} = 0$ . A questo punto si riporta bassa la WORD LINE e la cella rimane isolata mantenendo il suo stato permanentemente. Si può osservare che tale cella di memoria è volatile perché, se viene a mancare la tensione di alimentazione, lo stato del latch svanisce dopo poco tempo per lo spegnimento di tutti i transistor's.



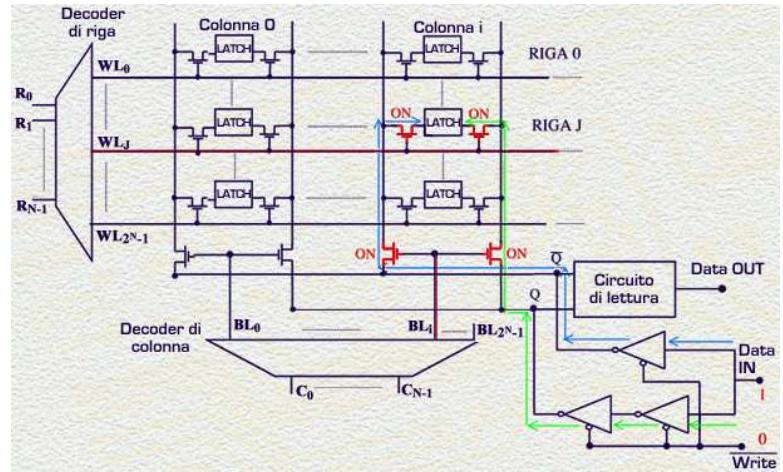
Per realizzare una memoria a molte celle si può ricorrere alla topologia a matrice rappresentata in figura. Ogni cella è singolarmente individuabile da una combinazione di un numero di riga, o WORD LINE, ed un numero di colonna, che identifica una coppia di bit-lines. Ad esempio la cella 1-0, sta sulla riga 1 identificata dalla WORD LINE 1 e sulla colonna 0 identificata dalle bit-lines  $\bar{Q}_0$  e  $Q_0$ . Quindi per leggere o scrivere un bit nella cella di memoria 1-0 è sufficiente portare alta la WORD LINE 1 ed andare a

selezionare le due bit-line della colonna 0. Si comprende perciò come una organizzazione siffatta della memoria permetta in ogni istante un accesso arbitrario a qualsiasi cella della matrice. Per tale motivo questo tipo di memoria viene denominata RAM (RANDOM ACCESS MEMORY) ovvero MEMORIA AD ACCESSO CASUALE.

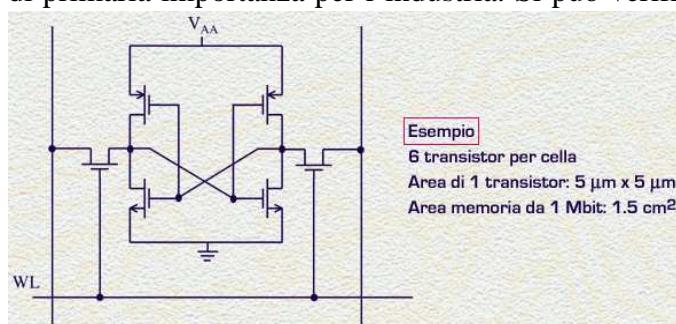
Sono stati sviluppati diversi tipi di memorie RAM che differiscono sostanzialmente tra loro per il tipo di circuito della singola cella, mentre viene mantenuta l'organizzazione a matrice delle celle. La RAM che abbiamo ora introdotto e che impiega come cella un circuito latch è chiamata RAM statica o SRAM. La qualifica STATICÀ data a questa RAM deriva dal fatto che una volta che sono stati scritti

i bit nelle celle, questa memoria mantiene l'informazione staticamente ovvero senza bisogno di altre operazioni. L'attributo STATICHE è necessario per distinguerla da un altro tipo di memoria, detta RAM dinamica, che utilizza una cella elementare diversa e che richiede una periodica riscrittura dei dati nelle celle al fine di poter mantenere l'informazione nel tempo. Studieremo le RAM dinamiche successivamente.

Torniamo ora ad esaminare più in dettaglio la struttura di una memoria RAM statica evidenziando i principali circuiti che la compongono, oltre la matrice di celle. Come abbiamo detto per scrivere o leggere una cella è necessario selezionarla attraverso la combinazione di una RIGA ed una COLONNA. Per selezionare la riga si impiega un decodificatore detto appunto **DECODER DI RIGA** che riceve in ingresso la parola a N bit in grado di identificare una delle  $2^N$  WORD LINES possibili. Per selezionare la colonna si impiega un **DECODIFICATORE DI COLONNA** anch'esso con parola di ingresso ad N bit, supponendo la matrice di memoria quadrata. Il numero totale di celle è quindi  $2^N \times 2^N$ . Il decoder di colonna riceve in ingresso l'indirizzo  $C_0, C_1 \dots C_{N-1}$  e porta alta una delle sue linee d'uscita,  $BL_0 \dots BL_{2^N-1}$  che rende condutti i due MOSFET di selezione di colonna. Le due bit-line della colonna selezionata vengono così collegate al circuito di lettura che rende disponibile in uscita il bit memorizzato. La fase di scrittura viene abilitata attraverso un segnale basso all'ingresso Write che abilita i buffer tri-state permettendo al dato presente su Data IN di raggiungere le linee Q e  $\bar{Q}$  e quindi le bit-lines della colonna selezionata. Il dato in ingresso raggiungerà solo la cella identificata dalla coppia di indirizzi di riga e colonna. Ad esempio, se vogliamo scrivere il bit '1' nella cella **Ji**, ovvero RIGA J – COLONNA i, presentiamo il bit **1** all'ingresso Data IN, portiamo basso Write, quindi selezioniamo la riga J e la colonna i attraverso i corrispondenti indirizzi di riga e colonna. I relativi MOSFET di selezione della cella e di selezione della colonna vengono accesi e il dato verrà quindi diretto verso il latch della cella J-i. Ricordiamo che in una lezione precedente, di introduzione alle memorie a semiconduttore, avevamo evidenziato come uno dei *parametri importanti di una memoria fosse il numero di bit memorizzabili per unità di area di chip*.



In pratica è di grande importanza riuscire a realizzare il maggior numero di celle possibili entro l'area di silicio sulla quale è possibile operare con i processi tecnologici. In generale, a parità di tecnologia, maggiore è la densità di celle nella memoria e minore è il costo unitario di produzione della memoria stessa, fattore questo di primaria importanza per l'industria. Si può verificare, che in una memoria RAM, la maggior parte



della superficie del chip è occupato proprio dalla matrice delle celle. Per aumentare la densità di celle è pertanto necessario ridurre l'area occupata dalla singola cella o il numero e la dimensione delle piste di connessione della matrice. Per quanto riguarda l'area della cella della memoria RAM statica, possiamo osservare che questa è determinata essenzialmente dai sei transistor che la costituiscono: i quattro transistor del latch e i due transistor di selezione. Se ad esempio supponiamo che un transistor occupi una superficie di area pari a  $5 \mu\text{m} \times 5 \mu\text{m}$  allora solo i transistor costituenti la matrice di una memoria da 1 Mbit occuperanno un'area di  $1.5 \text{ cm}^2$ , che è già un valore relativamente alto per un singolo chip. Per poter quindi realizzare memorie con densità di celle, ovvero con numero di celle per unità di superficie del chip, sempre più grande, una via è quella di ridurre le dimensioni dei singoli transistor e delle piste di interconnessione. Ciò è stato reso possibile dai progressi tecnologici che negli anni hanno permesso di realizzare transistor sempre più piccoli. Un'altra via

è quella di ridurre la dimensione delle piste di connessione, aumentando la densità di tracciature sul chip. Questo è stato possibile grazie allo sviluppo delle tecnologie di stampaggio a raggi laser, che permettono di creare piste di dimensioni molto piccole, al punto che oggi siamo in grado di realizzare memorie con densità di celle di circa 100 Gbit/cm<sup>2</sup>.

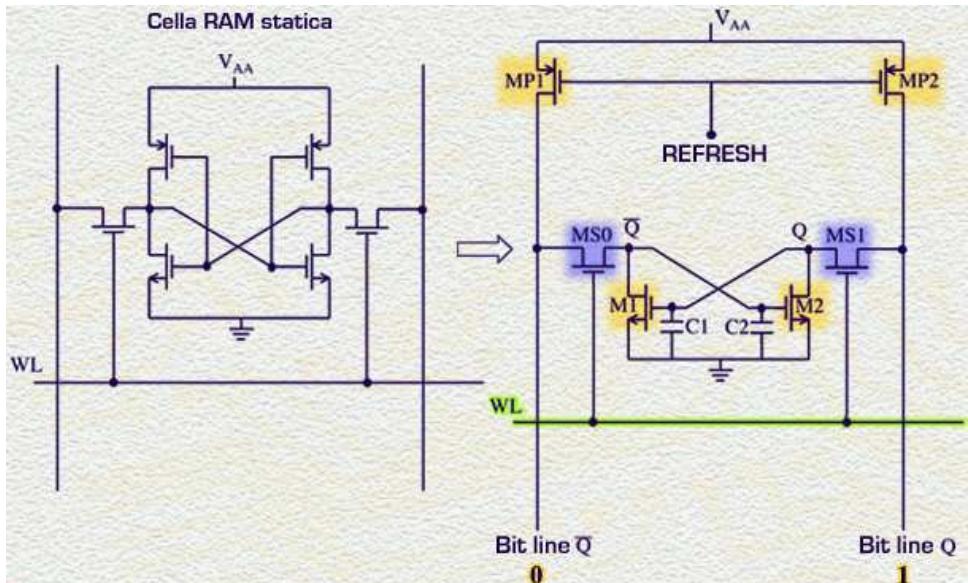
perseguita al fine di aumentare la densità di una memoria è stata quella di cercare soluzioni circuituali che permettessero di ridurre il numero di transistor della singola cella.

### RAM dinamiche (DRAM)

La prima soluzione trovata è stata quella di rimuovere da tutte le celle i due transistor P e di usarne solo due per tutta la colonna, come mostrato in figura. In tal modo il numero di transistor per cella si riduce da 6 a 4 con la conseguente riduzione di area occupata dalla singola cella. Vediamo come funziona una cella di memoria siffatta. Supponiamo di voler scrivere il bit  $Q = 1$  nella cella. Poniamo quindi sulle bit-lines

$Q=1$  e  $\bar{Q}=0$  e selezioniamo la cella portando alta la WORD LINE. Osserviamo il transistor M1 nella cella: il suo GATE risulta quindi collegato alla bit-line  $Q$ , che è allo stato alto, quindi M1 se non è già acceso, si accende.

Il GATE del MOSFET M2 è invece connesso alla bit-line  $\bar{Q}$ , che è bassa, quindi M2 si spegne, se non lo era già



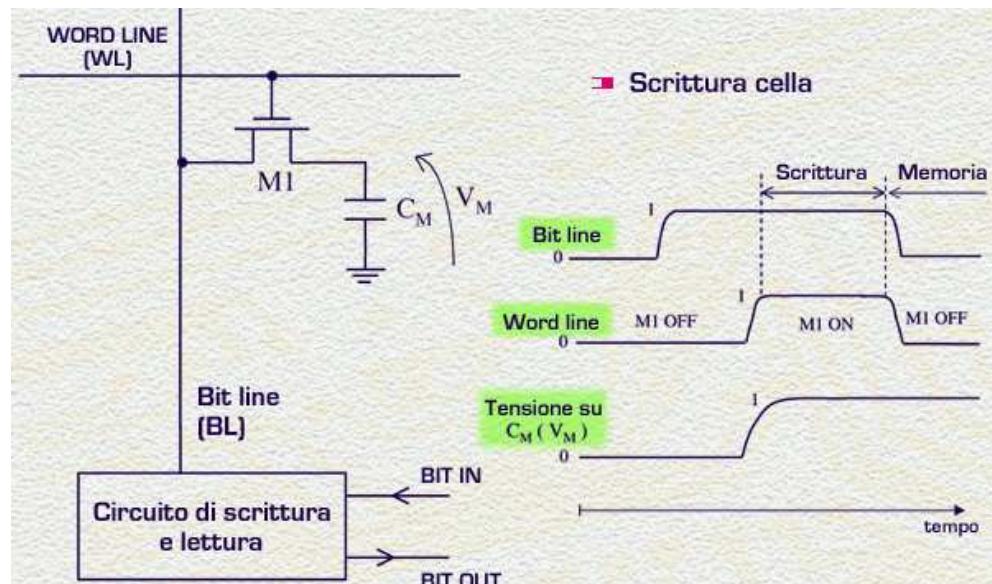
precedentemente. A questo punto il bit è scritto e la WORD LINE può tornare bassa. Il MOSFET M1 rimane nel suo stato di ON fintanto che la capacità C1 al suo gate rimane carica ad una tensione superiore alla tensione di soglia del MOSFET. A sua volta il canale conduttivo di M1 vincola a massa il gate di M2 e quindi tiene scarica la capacità C2 e spento il MOSFET.

La cella mantiene quindi il suo stato di memorizzazione del bit fino a che la capacità C1 non perde la sua carica, provocando lo spegnimento di M1. Lo stato di carica di C1 non può perdurare indefinitamente nel tempo a causa della presenza di una corrente di scarica che scorre attraverso M2 che, sebbene spento, manifesta comunque una residua conduzione di canale detta '*conduzione sotto-soglia*'. In pochi secondi il dato verrebbe quindi perso. Per evitare ciò viene eseguita periodicamente un'operazione detta di REFRESH al fine di ripristinare lo stato di polarizzazione della cella. Per eseguire questa operazione si porta a livello basso il nodo REFRESH della colonna rendendo quindi conduttori i due PMOS MP1 e MP2. A questo punto viene selezionata la cella portando alta la WORD LINE. In questo modo, tramite i transistor di selezione MS0 e MS1 si permette la costituzione di un circuito latch formato da M1 ed MP1, che compongono il primo invertitore, e M2-MP2 che compongono il secondo. I MOSFET di selezione MS0 e MS1 possono essere visti come conduttori interposti tra i transistor M1, M2 e i transistor MP1, MP2 rispettivamente. La loro presenza non compromette la costituzione del circuito latch. Una volta formato il latch questo si porta, dal punto di polarizzazione in cui si trova al momento della connessione, verso il punto di equilibrio stabile più vicino. Dato che la fase di refresh avviene quando ancora M1 è acceso ed M2 è spento, il latch si porterà nel punto in cui  $\bar{Q}$  è basso e  $Q$  è alto, ripristinando quindi la piena carica della capacità C1. A questo punto l'operazione di refresh è completa, la WORD LINE può tornare bassa e si può proseguire con il refresh di altre celle della stessa colonna abilitando ad una ad una tutte le WORD LINE's. L'operazione di refresh per una cella può avvenire con un periodo dell'ordine delle decine di millisecondi, ovvero con frequenze dell'ordine di decine di Hz. La RAM con la struttura vista viene detta **DINAMICA** proprio per il fatto che il suo funzionamento richiede un'operazione, quella di refresh, che si sviluppa periodicamente nel tempo. Un funzionamento statico di questa memoria, ovvero senza refresh, comporterebbe la perdita dei bit. Diversamente una RAM statica tiene memorizzati i dati senza la necessità di operazioni che si sviluppano o ripetono nel tempo.

L'operazione di refresh in una RAM dinamica richiede una logica di controllo aggiuntiva e comporta una maggior complessità di funzionamento, rispetto alla RAM statica. Tuttavia la RAM dinamica presenta il vantaggio di avere una cella più compatta: **4 transistor invece dei 6 della cella statica.** *I due transistor necessari per il refresh sono condivisi da tutte le celle della stessa colonna e quindi sono numericamente influenti, visto che una colonna può contenere centinaia di celle.* Inoltre la cella di RAM dinamica non richiede la pista dell'alimentazione  $V_{AA}$  che invece deve raggiungere ogni cella nella RAM statica, anche questo contribuisce a risparmiare spazio e a raggiungere, con le RAM dinamiche densità di celle superiori rispetto alle RAM statiche.

Un'ulteriore evoluzione nella realizzazione delle memorie RAM ha riguardato la realizzazione di celle costituite da un solo transistor e da un condensatore. Abbiamo già introdotto, in una precedente lezione, il principio di funzionamento di queste celle di memoria. Ora vedremo come vengono organizzate in matrici per formare una RAM. Il condensatore rappresenta l'elemento di memoria: se è carico il bit memorizzato è 1, se scarico il bit è 0. Il MOSFET opera come porta di trasmissione: se è spento, il condensatore di memoria è isolato dal resto del circuito e quindi svolge la sua funzione di tenuta del suo stato di carica, ovvero di memoria del bit; quando invece il MOSFET viene reso conduttivo esso mette in contatto un'armatura del condensatore con la bit-line permettendo quindi la scrittura o la lettura della cella. Facciamo un esempio. Supponiamo di voler scrivere il bit 1 nella cella. Seguiamo il diagramma in figura che mostra gli andamenti delle tensioni nel tempo per la bit-line, la WORD line e della tensione ai capi della capacità di memoria.

Volendo scrivere un '1' nella cella poniamo la bit-line a livello logico alto. Poi portiamo alta la WORD LINE così da rendere conduttivo il transistor M1. Pertanto il condensatore  $C_M$ , trovandosi connesso alla bit-line BL, che è a livello logico alto, si carica a livello alto. A questo punto la WORD LINE viene riportata



bassa, il transistor M1 si spegne ed il condensatore  $C_M$  si trova isolato e carico a livello logico alto memorizzando quindi il bit 1. Anche in tal caso, però, correnti parassite e "correnti di conduzione sotto soglia" del transistor M1 comportano la scarica di  $C_M$  in tempi dell'ordine di pochi secondi. È pertanto necessario periodicamente, in genere ogni qualche decina di ms, effettuare un refresh della cella.

Vediamo ora come si effettua la lettura della stessa cella. Per leggere la cella si porta alta la WORD LINE così da accendere il MOSFET M1. A questo punto lo stato di carica, a livello alto o a livello basso, della capacità  $C_M$  viene rilevato da un circuito di lettura connesso alla bit-line BL.

La lettura dello stato di carica del condensatore di memoria richiede un approfondimento. Infatti, supponiamo che la tensione a cui  $C_M$  è carica, quando è isolata, sia  $V_M$  volt. Chiediamoci: quale valore di tensione viene letto sulla bit-line quando il transistor M1 viene acceso? Per risolvere il problema, dobbiamo premettere che, tra la bit-line e massa, esiste una capacità  $C_L$  somma di tutte le capacità associate a tutti i transistor delle celle connessi alla bit-line stessa e a tutti gli accoppiamenti eletrostatici che la bit-line ha con i conduttori adiacenti. Supponiamo per semplicità che la bit-line si trovi inizialmente a potenziale 0, quindi la carica sulla capacità  $C_L$  è nulla. Quando viene selezionata la cella, accendendo il transistor M1, le due capacità  $C_M$  e  $C_L$  vanno in parallelo e la carica inizialmente presente solo su  $C_M$  si distribuisce tra  $C_M$  e  $C_L$ . Il potenziale a cui si porta la bit-line è facilmente calcolabile dall'equazione che esprime la conservazione della carica totale.

La carica iniziale è pari a quella presente sul condensatore di memoria, quindi  $Q_M = C_M \cdot V_M$ . La carica finale è identica a quella iniziale ma è distribuita sul parallelo di  $C_M$  e  $C_L$  quindi  $Q_F = (C_M + C_L) \cdot V_{BL}$

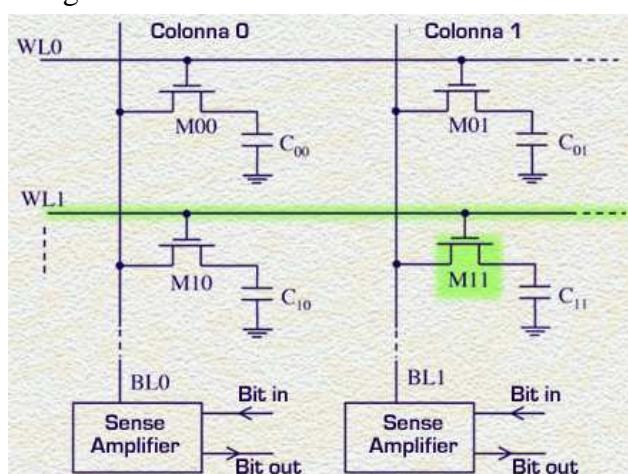
dove  $V_{BL}$  indica il potenziale a cui si porta la bit-line dopo che il MOS è stato acceso.

Uguagliando l'espressione della carica finale con quella della carica iniziale ricaviamo l'espressione della tensione  $V_{BL}$  della bit-line. Discutiamo ora il risultato trovato: occorre osservare che  $C_L$  è molto più grande di  $C_M$  perché  $C_L$  risulta dalla somma di un gran numero di capacità. Quindi il rapporto capacitivo che compare a 2° membro è molto più piccolo di uno e questo ci dice che la tensione  $V_{BL}$  della bit-line sarà molto minore della tensione  $V_M$  a cui si trovava la capacità di memoria quand'era isolata. Facciamo un esempio numerico: supponiamo  $C_M = 20 \text{ fF}$ ,  $C_L = 1 \text{ pF}$ ,  $V_M = 3 \text{ V}$ ; dall'equazione deduciamo che  $V_{BL}$  è circa pari a 60 mV. Questo ci dice che l'operazione di lettura di questa cella comporta un abbassamento notevole del livello logico alto.

Nella cella infatti era stato scritto il bit 1 corrispondente a 3 Volt ed in fase di lettura vengono letti solo 60 mV. Questo supponendo che la capacità  $C_M$  memorizzi il bit 1. Quando invece la cella memorizza il bit 0, la tensione su  $C_M$  è pari a 0 volt. In fase di lettura, sulla bit-line verrà quindi letto ancora 0 volt in quanto sia  $C_M$  che  $C_L$  sono entrambe scaricate.

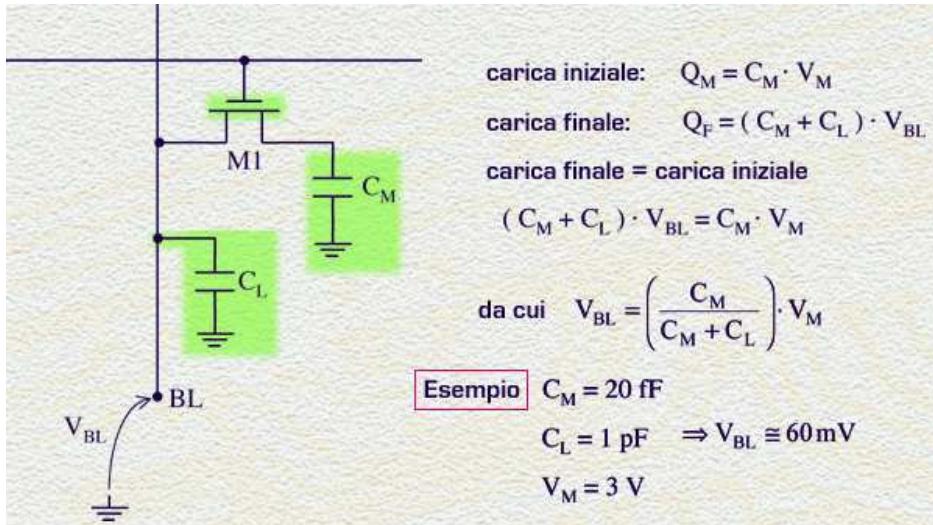
Pertanto, se i livelli logici memorizzati sulla capacità sono 0 e 3 Volt, in fase di lettura i livelli sono 0 e 60 mV. Perciò la lettura di questo tipo di celle richiede un particolare circuito in grado di distinguere due livelli logici che differiscono in tensione di poche decine di millivolt. A tale fine sono stati ideati dei circuiti di lettura chiamati **SENSE AMPLIFIER** in grado di leggere le bit-line e fornire in uscita livelli logici standard. Il circuito di lettura di una RAM dinamica con singolo transistor è pertanto più sofisticato rispetto a quello delle memorie basate su celle latch. Tuttavia dobbiamo notare che **la riduzione dell'area della cella a singolo transistor è notevole rispetto a quella richiesta dalle celle latch**. Infatti, **non solo si è passati da minimo 4 transistors ad un solo transistor più una capacità, ma anche da due bit-linea per colonna necessarie alla cella latch ad una sola bit-line**.

In figura vediamo una sezione della matrice di memoria di una RAM DINAMICA con celle a singolo transistor.



Ogni cella è individuabile univocamente da una coppia di numeri che identificano una riga (WORD LINE) ed una colonna (BIT-LINE). Ad ogni colonna è collegato un SENSE AMPLIFIER. Quando una particolare WORD LINE viene selezionata e portata alta, tutti i transistor della corrispondente riga divengono conduttori e su tutte le BIT-LINES appaiono i bit delle celle della riga selezionata. Ad esempio se viene portata alta la WORD LINE WL1, il contenuto della cella M1-0 sarà presente sulla BIT-LINE BL0 e verrà letto dal SENSE AMPLIFIER corrispondente, contemporaneamente il contenuto della cella M1-1

sarà presente sulla BIT-LINE BL1 e così via, tutte le celle della riga WL1 vengono lette dai rispettivi sense amplifiers. Il decoder di colonna, non mostrato in figura, seleziona una particolare colonna e invia il corrispondente bit verso l'uscita della memoria. Un'analogia operazione avviene per la scrittura: in tal caso attraverso il sense amplifier viene imposto un bit sulla colonna selezionata e viene



$$\text{carica iniziale: } Q_M = C_M \cdot V_M$$

$$\text{carica finale: } Q_F = (C_M + C_L) \cdot V_{BL}$$

**carica finale = carica iniziale**

$$(C_M + C_L) \cdot V_{BL} = C_M \cdot V_M$$

$$\text{da cui } V_{BL} = \left( \frac{C_M}{C_M + C_L} \right) \cdot V_M$$

**Esempio**  $C_M = 20 \text{ fF}$

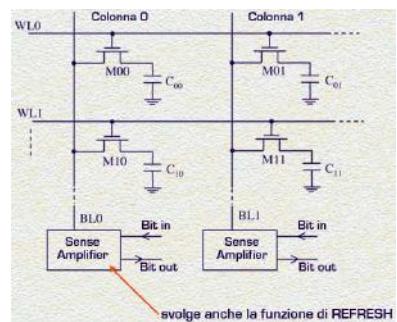
$$C_L = 1 \text{ pF} \Rightarrow V_{BL} \approx 60 \text{ mV}$$

$$V_M = 3 \text{ V}$$

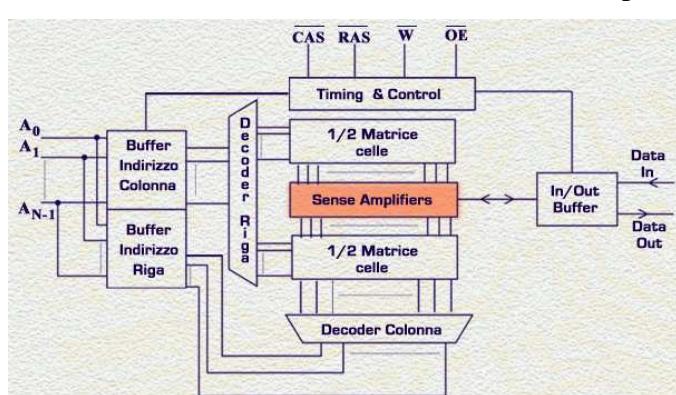
quindi portata alta la word line desiderata. Il SENSE AMPLIFIER svolge anche la funzione di refresh della cella. Infatti il circuito del SENSE AMPLIFIER distingue il livello logico della bit-line, se basso o alto, e ripristina di conseguenza sulla bit-line e quindi anche nella cella selezionata i livelli logici standard. In pratica il refresh di una cella si effettua semplicemente selezionando la cella in lettura.

In figura è mostrato uno schema semplificato della architettura di una RAM DINAMICA. Vediamo che le celle sono suddivise in due matrici che condividono il blocco di sense amplifiers. Nella realtà le

celle sono suddivise in più copie di matrici.



In figura è stata disegnata solo una coppia per semplicità. Vi è un unico insieme di piedini **A<sub>0</sub>, A<sub>1</sub> ... A<sub>N-1</sub>** che ricevono, in tempi diversi, l'indirizzo di riga e quello di colonna, indirizzi che vengono memorizzati in due buffer distinti, che agiscono da registri di memoria. Il sistema è regolato da un circuito logico di temporizzazione e controllo avente 4 segnali d'ingresso. Il **CAS negato**, ovvero attivo quando basso, è il comando COLUMN ADDRESS STROBE e serve per memorizzare l'indirizzo presente agli ingressi **A<sub>0</sub>, A<sub>1</sub> ... A<sub>N-1</sub>** nel BUFFER DI COLONNA. Analogamente il **RAS negato**, ROW ADDRESS STROBE, quando basso memorizza l'indirizzo in ingresso nel buffer di riga. Il bit **W negato**, WRITE ENABLE, quando basso, abilita l'operazione di scrittura mentre, quando è posto alto, abilita l'operazione di lettura. Il bit **OE negato**, OUTPUT ENABLE controlla i buffer d'uscita, ovvero quando è basso rende disponibile il dato letto in uscita.



temporizzazione e controllo avente 4 segnali d'ingresso. Il **CAS negato**, ovvero attivo quando basso, è il comando COLUMN ADDRESS STROBE e serve per memorizzare l'indirizzo presente agli ingressi **A<sub>0</sub>, A<sub>1</sub> ... A<sub>N-1</sub>** nel BUFFER DI COLONNA. Analogamente il **RAS negato**, ROW ADDRESS STROBE, quando basso memorizza l'indirizzo in ingresso nel buffer di riga. Il bit **W negato**, WRITE ENABLE, quando basso, abilita l'operazione di scrittura mentre, quando è posto alto, abilita l'operazione di lettura. Il bit **OE negato**, OUTPUT ENABLE controlla i buffer d'uscita, ovvero quando è basso rende disponibile il dato letto in uscita.

<pagina bianca per esercizio >

<pagina bianca per esercizio >

<pagina bianca per esercizio >

<pagina bianca per data-sheet>

# Lezione n. 24 : Introduzione agli amplificatori. La risposta in frequenza

## L'amplificatore: alcuni concetti introduttivi

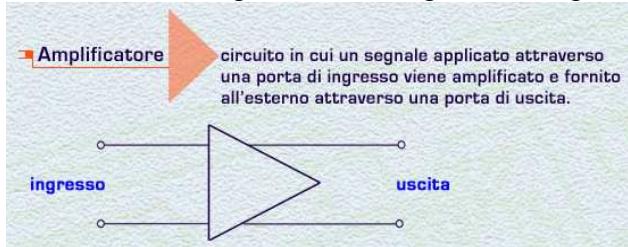
Iniziamo con introdurre alcuni concetti elementari sugli amplificatori, di cui faremo uso in seguito. Un'introduzione generale all'argomento riguardante l'amplificazione dei segnali in elettronica era stata peraltro già presentata nelle prime lezioni del corso.

L'amplificatore è un **circuito in cui un segnale applicato attraverso una porta di ingresso viene amplificato e fornito all'esterno attraverso una porta di uscita**. A volte risulta comodo riferire sia il segnale in ingresso che quello in uscita ad un potenziale comune, ad esempio il potenziale di massa. Questa possibilità viene rappresentata, come indicato nel secondo schema, ponendo uno dei morsetti in ingresso ed uno dei morsetti in uscita allo stesso potenziale di riferimento, nel caso indicato appunto la massa. Si intende in questo caso, che il segnale di ingresso venga applicato al morsetto libero e riferito rispetto a massa, come anche il segnale venga fornito in uscita al morsetto libero e riferito rispetto a massa. Una caratteristica fondamentale dell'amplificatore è rappresentata dal suo **guadagno**, che esprime la misura con cui questo dispositivo effettua l'amplificazione del segnale applicato al suo ingresso. Il guadagno di un amplificatore può essere di definito in diversi modi, essenzialmente a seconda del tipo di grandezza elettrica, tensione o corrente, di cui venga operata l'amplificazione. Se infatti supponiamo di applicare ai morsetti di ingresso dell'amplificatore un segnale di tensione  $V_{in}$  e siamo interessati al corrispondente valore di tensione disponibile ai morsetti di uscita  $V_{out}$ , si definisce guadagno di tensione  $A_V$  il rapporto tra  $V_{out}$  e  $V_{in}$ .

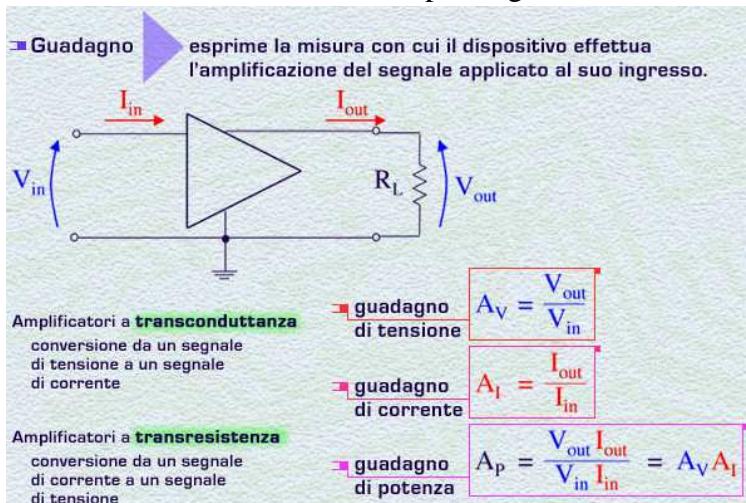
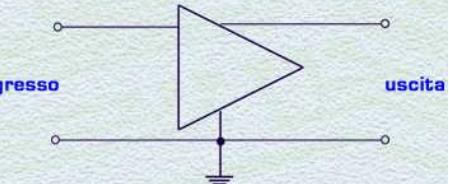
Analogamente, se supponiamo di applicare in ingresso, attraverso un opportuno generatore, un segnale in corrente  $I_{in}$  e siamo interessati al segnale di corrente  $I_{out}$  disponibile in uscita, si definisce guadagno di corrente  $A_I$  dell'amplificatore il rapporto tra  $I_{out}$  e  $I_{in}$ . Ovviamente esistono anche configurazioni di amplificatore che operano una conversione da un segnale di tensione in ingresso a un segnale di corrente in uscita, gli amplificatori a **transconduttanza**, e quelli che operano, al contrario, una conversione da segnale di corrente in ingresso a segnale di tensione in uscita, gli amplificatori a **transresistenza**.

Infine, se la funzione dell'amplificatore non riguarda la sola amplificazione della corrente o della tensione applicate all'ingresso, ma bensì del loro prodotto, cioè della potenza fornita in ingresso, allora è possibile definire anche un guadagno di potenza  $A_P$  che è pari al rapporto tra potenza fornita sul carico di uscita, cioè  $V_{out} \cdot I_{out}$ , diviso per la potenza fornita in ingresso all'amplificatore cioè il prodotto  $V_{in} \cdot I_{in}$ . Il guadagno di potenza risulta quindi pari al prodotto dei guadagni di tensione e di corrente. A determinare il guadagno di un amplificatore concorrono anche le impedenze della sorgente di segnale e del carico connessi al dispositivo, insieme alle impedenze di ingresso e di uscita dell'amplificatore stesso, come è stato già mostrato nelle prime lezioni del corso.

Un'importante classe di amplificatori è quella degli *amplificatori lineari*. Per amplificatore lineare si intende un amplificatore il cui segnale di uscita varia in maniera proporzionale rispetto al segnale di ingresso. Se ad esempio osserviamo la caratteristica di un amplificatore lineare di tensione,

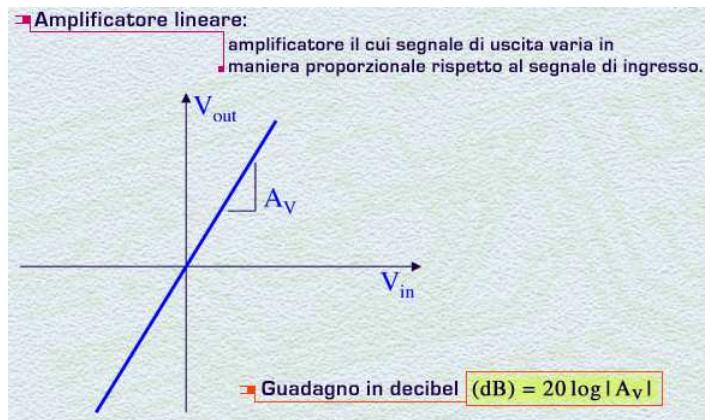


■ Può essere comodo riferire sia il segnale di ingresso che quello in uscita ad un potenziale comune (es. massa).



osserviamo che questa è rappresentata da una retta, il cui coefficiente angolare è dato dal guadagno di tensione  $A_V$  dell'amplificatore.

Occorre per ultimo osservare che il guadagno di un amplificatore viene spesso espresso in decibel. Il guadagno in decibel è definito come 20 volte il logaritmo del modulo del guadagno. Questa notazione è molto comune in elettronica in quanto consente di rappresentare e confrontare con una maggiore facilità valori di guadagno che possono differire anche di diversi ordini di grandezza. Se ad esempio due valori di guadagno differiscono di cinque ordini di grandezza, cioè uno è centomila volte maggiore dell'altro, la loro differenza in decibel è di 100 dB, una differenza che ad esempio può essere facilmente rappresentata in un grafico.



### La risposta in frequenza

Nelle lezioni introduttive sui segnali trattati in elettronica, è stato mostrato come i segnali di tipo analogico possano essere rappresentati come una somma di sinusoidi di diversa frequenza e ampiezza. Un **segnale analogico, cioè, può essere descritto attraverso il suo spettro di frequenza**.

Grazie a questa possibilità di rappresentazione del segnale come somma di sinusoidi di diverse frequenze, ha senso chiedersi, quando questo segnale venga applicato all'ingresso di un amplificatore, quale sia la risposta dell'amplificatore alla somma delle sinusoidi di cui è composto il segnale.

La risposta a questa domanda risiede in una importante caratteristica dell'amplificatore che è la sua **risposta in frequenza**.

Vediamo di affrontare questo argomento, iniziando a valutare innanzitutto quale sia la risposta di un amplificatore ad una singola sinusoide applicata al suo ingresso. Supponiamo di fornire in ingresso ad

un amplificatore un segnale sinusoidale di tensione, di frequenza  $\omega$  e ampiezza  $V_i$ . Se l'amplificatore è lineare, esso possiede un'importante proprietà: il segnale di tensione in uscita sarà ancora una sinusoide alla stessa frequenza  $\omega$  di diversa ampiezza  $V_o$  ed eventualmente sfasata di un termine  $\phi$ . L'amplificatore lineare quindi può cambiare l'ampiezza e la fase di una sinusoide fornita in ingresso ma non ne cambia però la frequenza.

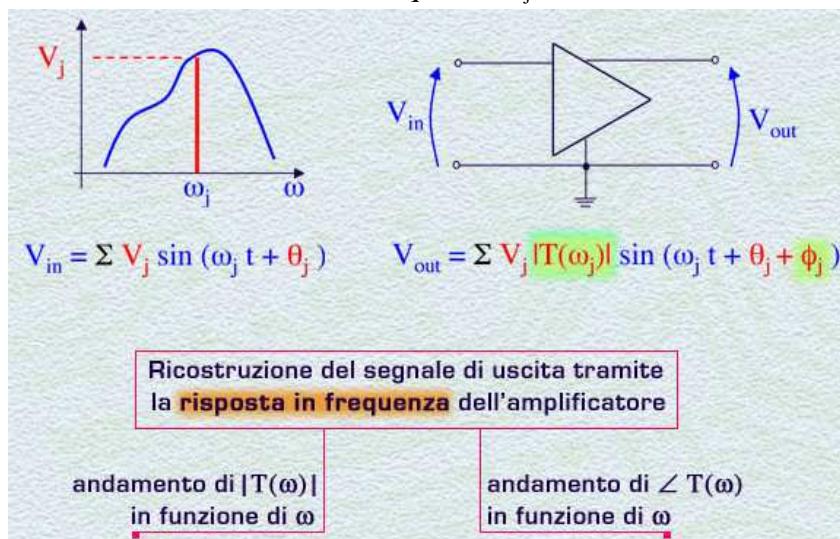
Risulta importante a questo punto conoscere come l'amplificatore,

supposto d'ora in poi sempre lineare, modifichi ampiezza e fase di una qualsiasi sinusoide di frequenza  $\omega$  fornita al suo ingresso. Le proprietà di risposta alle diverse sinusoidi vengono descritte dalla **funzione di trasferimento  $T(\omega)$**  dell'amplificatore.

La funzione di trasferimento  $T(\omega)$  è un numero complesso, funzione di  $\omega$ , il cui modulo è pari al rapporto tra l'ampiezza  $V_o$  della sinusoide in uscita e l'ampiezza  $V_{\text{in}}$  di quella in ingresso mentre l'argomento o fase è pari allo sfasamento  $\phi$  aggiunto alla sinusoide nel suo trasferimento all'uscita dell'amplificatore.

Se ora riconsideriamo, come fatto all'inizio, la rappresentazione di un segnale tramite il suo spettro in frequenza, è possibile comprendere l'importanza di caratterizzare la risposta in frequenza di un amplificatore tramite la determinazione della sua funzione di trasferimento. Consideriamo infatti un generico segnale in ingresso all'amplificatore come rappresentato da una somma di sinusoidi, la cui generica sinusoide abbia frequenza  $\omega_j$ , ampiezza  $V_j$ , e fase  $\theta_j$ . Il segnale in uscita dall'amplificatore

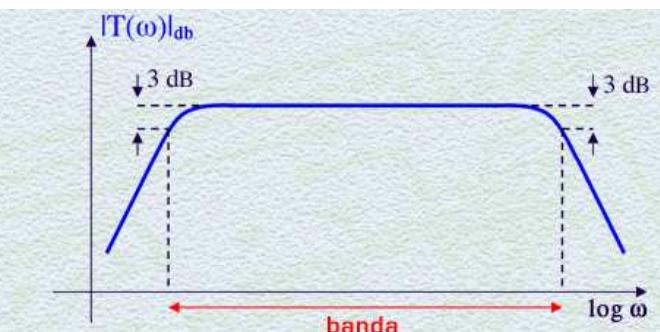
può essere rappresentato ancora come una somma di sinusoidi alle stesse frequenze in cui, l'ampiezza della generica sinusoide di frequenza  $\omega_j$  è ora pari al prodotto di  $V_j$  per il modulo della funzione di trasferimento calcolata alla frequenza  $\omega_j$ , e la cui fase è aumentata del fattore  $\phi_j$ , dato dalla fase della funzione di trasferimento alla frequenza  $\omega_j$ . In sintesi quindi, è possibile ricostruire il segnale di uscita, dato lo spettro in frequenza del segnale in ingresso, noti l'andamento del modulo e della fase della funzione di trasferimento in funzione della frequenza.



amplificato, alterandone lo spettro di frequenza, o, da un altro punto di vista, possa consentire di condizionare in maniera opportuna i segnali. Un esempio di questo ultimo tipo è dato dalla classe dei circuiti di filtraggio, i quali operano una selezione opportuna delle frequenze del segnale da trasmettere rispetto a quelle invece da attenuare.

Molto spesso in elettronica si utilizzano amplificatori che possiedono un guadagno costante in un certo intervallo di frequenze. A tale scopo si definisce in generale la *banda* di un amplificatore, come l'intervallo di frequenze in cui il modulo della funzione di trasferimento dello stesso è costante entro un certo valore, che solitamente è considerato pari a 3 dB.

Introdotta l'importanza del concetto di funzione di trasferimento di un amplificatore, risulta ora necessario determinare nei casi concreti di reti elettriche l'espressione di tale funzione. A tale scopo occorre ricordare come in altri corsi si sia data



**Banda dell'amplificatore**  
intervallo di frequenze in cui il modulo del guadagno dell'amplificatore è all'incirca costante (solitamente entro 3 dB).

Rappresentazione dei componenti mediante la loro impedenza simbolica in funzione della frequenza:

--	--	--

... o in funzione della variabile complessa s (tramite la Trasformata di Laplace):

--	--	--

una rappresentazione in termini di impedenza simbolica ai componenti passivi e reattivi di una generica rete elettrica. Tale rappresentazione descrive il comportamento in frequenza di tali componenti. Secondo questa descrizione, la resistenza risulta avere impedenza pari a  $R$ , la capacità pari  $1/j\omega C$  e l'induttanza pari a  $j\omega L$ .

Analogamente a questa descrizione, è possibile giungere ad una rappresentazione analoga delle impedenze simboliche espresse, questa volta, rispetto alla variabile complessa  $s$ , tramite l'ausilio della trasformata di Laplace. La determinazione di tali impedenze con la Trasformata di

Laplace, consente di associare alla resistenza ancora l'impedenza  $R$ , alla capacità l'impedenza  $1/sC$  e all'induttanza l'impedenza  $sL$ .

Sulla base di tali rappresentazioni, data una generica rete elettrica, è possibile determinare l'espressione della funzione di trasferimento, utilizzando le regole di calcolo tipiche delle reti elettriche. Si perviene così ad una espressione della funzione di trasferimento, se espressa in funzione della variabile complessa  $s$ , pari al rapporto di due polinomi in  $s$ , un polinomio al numeratore di ordine  $m$  ed uno al denominatore di ordine  $n$ , detto anche *ordine della rete*, maggiore o uguale a  $m$ .

Se si determinano le radici di entrambi i polinomi, al numeratore e al denominatore, risulta utile rappresentare la funzione di trasferimento in funzione di tali radici. In particolare, le radici del polinomio al numeratore prendono il nome di *zeri* della funzione di trasferimento, mentre le radici del polinomio al denominatore prendono il nome di *poli*. Se si è determinata l'espressione della funzione di trasferimento in funzione della variabile complessa  $s$ , è possibile passare all'espressione equivalente della funzione di trasferimento in funzione della frequenza, o meglio della pulsazione  $\omega$ , mediante la sostituzione di  $s$  con la variabile  $j\omega$ . In questa maniera si ottiene un'espressione della risposta in frequenza dell'amplificatore in termini di amplificazione  $|T(j\omega)|$  e sfasamento  $\angle T(j\omega)$ .

■ È quindi possibile determinare l'espressione della funzione di trasferimento utilizzando le regole di calcolo delle reti elettriche

$$T(s) = \frac{a_m s^m + a_{m-1} s^{m-1} + \dots + a_0}{b_n s^n + b_{n-1} s^{n-1} + \dots + b_0} \quad m \leq n$$

$n$ : ordine della rete

$$T(s) = a_m \frac{(s - Z_1)(s - Z_2) \dots (s - Z_m)}{(s - P_1)(s - P_2) \dots (s - P_m)}$$

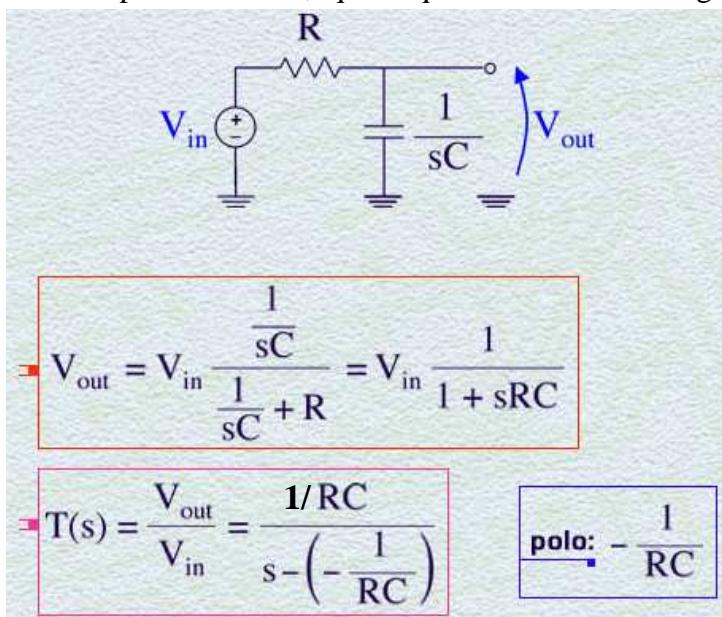
$Z_1, Z_2, \dots, Z_m$ : zeri della funzione di trasferimento

$P_1, P_2, \dots, P_m$ : poli della funzione di trasferimento

■ Una volta determinata  $T(s)$ , la funzione di trasferimento  $T(j\omega)$  può essere valutata nell'ambito delle frequenze fisiche sostituendo  $s=j\omega$ . L'analisi della  $T(j\omega)$  consente di ricavare la risposta in frequenza dell'amplificatore in termini di amplificazione  $|T(j\omega)|$  e sfasamento  $\angle T(j\omega)$ .

### Esempi di funzioni di trasferimento, Diagrammi di Bode.

Analizziamo come esempio di quanto detto nella sezione precedente, la funzione di trasferimento di una semplice rete RC, quale quella illustrata in figura. La funzione di trasferimento  $T(s)$  viene



facilmente determinata come rapporto tra l'uscita  $V_{out}$  e l'ingresso  $V_{in}$ , associando ai componenti della rete l'espressione dell'impedenza in funzione della variabile complessa  $s$ , vale a dire  $R$  per la resistenza e  $1/sC$  per la capacità. In questa maniera l'uscita  $V_{out}$  risulta pari all'ingresso  $V_{in}$  per la partizione tra  $1/sC$  e  $R$  e quindi pari a  $V_{in}$  per  $1/(1+sRC)$ .

La funzione di trasferimento risulta pertanto essere pari a  $1/RC$  diviso per  $[s - (-1/RC)]$ . In questa espressione si è messo in evidenza, così, la presenza di un polo reale di valore pari a  $-1/RC$ .

Al fine di valutare la risposta in frequenza della rete, risulta molto utile una particolare rappresentazione grafica della funzione di

Trasferimento: i Diagrammi di Bode.

I diagrammi di Bode consistono nella rappresentazione grafica sia del modulo che della fase della funzione di trasferimento  $T(j\omega)$ , a cui si perviene come detto sostituendo alla variabile complessa  $s$  la variabile  $j\omega$ . Il modulo di  $T(j\omega)$ , viene espresso in decibel e rappresentato in funzione del logaritmo di  $\omega$ , mentre la fase di  $T(j\omega)$  viene espressa in gradi o radianti e rappresentata anch'essa in funzione del logaritmo di  $\omega$ .

Tornando al caso specifico della nostra rete RC, troviamo quindi che la funzione di trasferimento  $T(j\omega)$  è pari a  $1/(1+j\omega RC)$ . Il modulo di  $T(j\omega)$  risulta pertanto pari a  $1/\sqrt{1+\omega^2(RC)^2}$ . In decibel,  $|T(j\omega)|_{dB}$  è pari a  $-20 \log \sqrt{1+\omega^2(RC)^2}$ .

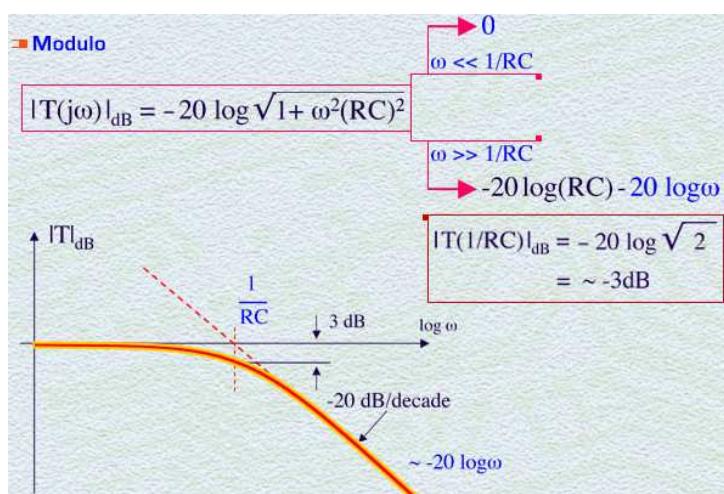
Rappresentiamo ora il diagramma di Bode del modulo della funzione di trasferimento. A tale scopo, più che effettuare una rappresentazione analitica dell'espressione trovata, risulta più significativo ricavare il grafico a partire dall'espressione del modulo della  $T(j\omega)$ , valutato in due zone estreme dell'asse delle

frequenze. Se consideriamo infatti il modulo per pulsazioni molto inferiori al polo  $1/RC$ , nella radice il termine  $\omega^2(RC)^2$  risulta trascurabile rispetto a 1. Pertanto in questo caso il modulo della funzione di

trasferimento, in decibel, tende a essere pari a 0, come mostrato nel grafico. Se consideriamo, all'estremo opposto, la regione di pulsazioni  $\omega$  molto maggiori rispetto al polo  $1/RC$ , allora il termine 1 è trascurabile, sotto radice, rispetto a  $\omega^2(RC)^2$ . In questo caso, l'espressione del modulo della funzione di trasferimento può essere approssimato con  $-20\log(RC) - 20\log(\omega)$ . Notiamo in questa espressione che il modulo in decibel risulta proporzionale al logaritmo di  $\omega$ , che è l'ascissa del nostro grafico. Pertanto l'andamento del modulo della funzione di trasferimento sarà una

retta di pendenza pari a  $-20\text{dB per decade}$ , cioè per ogni decade di aumento della pulsazione, il modulo diminuirà di 20 decibel. Se, infine calcoliamo il valore del modulo della funzione di trasferimento in corrispondenza del polo  $1/RC$ , otteniamo che questo vale  $-20\log(\sqrt{2})$ , cioè  $-3\text{db}$ . Grazie alle osservazioni precedenti è ora possibile tracciare il grafico completo del modulo della funzione di trasferimento della rete (vedi figura sopra). Il grafico così ottenuto ci permette di osservare come questa semplice rete lasci sostanzialmente inalterate le sinusoidi di frequenze inferiori a quella corrispondente al polo  $1/RC$  mentre attenui le sinusoidi di frequenza superiore, con un fattore di attenuazione che cresce con la frequenza. Una rete di questo genere può essere considerata un filtro di tipo passa-basso, in quanto lascia passare inalterate solo le basse frequenze.

Passiamo ora al diagramma di Bode della fase della funzione di trasferimento. Occorre in questo caso determinare l'argomento dell'espressione  $1/(1+j\omega RC)$ . Questo risulta pari a:  $-\arctg(\omega RC)$ . Nel grafico è mostrato l'andamento della fase secondo la funzione arcotangente ricavata. E' da notare che lo sfasamento introdotto dalla rete è pressoché nullo per sinusoidi di frequenza molto inferiore a quella del polo  $1/RC$ . Per una pulsazione dieci volte inferiore al polo  $1/RC$ , lo sfasamento introdotto è di soli  $5.7^\circ$ . Lo sfasamento risulta invece pari a meno  $45^\circ$  in corrispondenza del polo,



I diagrammi di Bode consistono nella rappresentazione grafica del modulo e della fase della funzione di trasferimento.

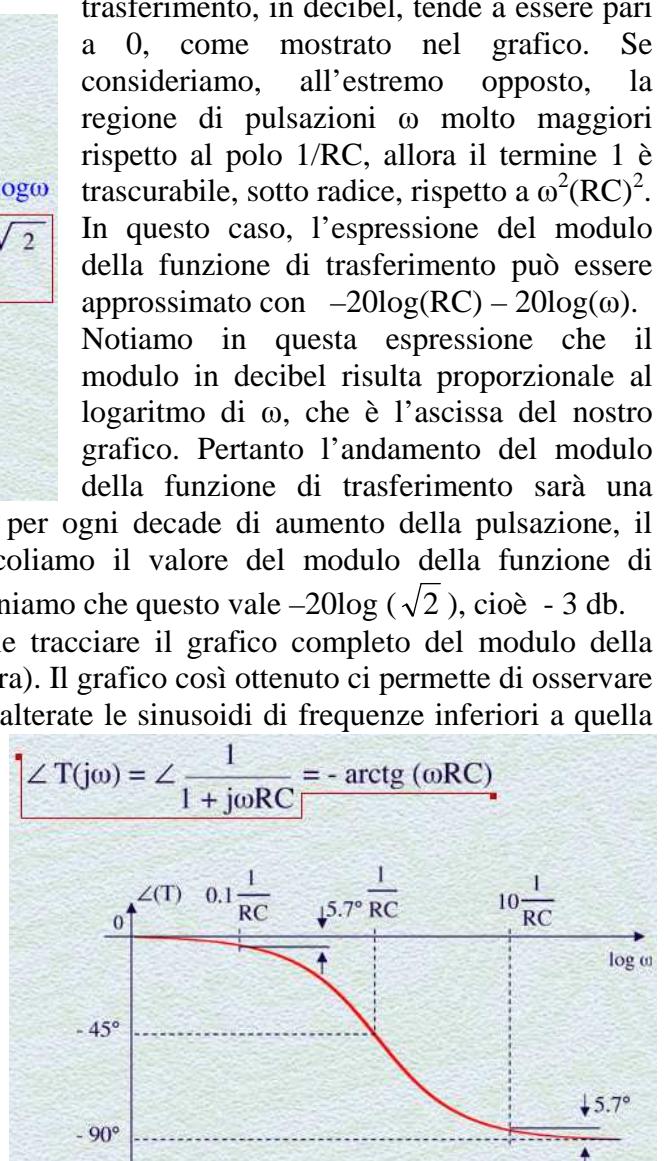
$ T(j\omega)  \text{ in dB, in funzione di } \log \omega$	$\angle T(j\omega) \text{ in funzione di } \log \omega$
---	---

■ Diagrammi di Bode della rete RC

$$T(j\omega) = \frac{1}{1 + j\omega RC}$$

$$|T(j\omega)| = \frac{1}{|1 + j\omega RC|} = \frac{1}{\sqrt{1 + \omega^2(RC)^2}}$$

$$|T(j\omega)|_{dB} = 20 \log |T(j\omega)| = -20 \log \sqrt{1 + \omega^2(RC)^2}$$



mentre diventa sempre più prossimo a  $-90^\circ$  più la pulsazione diventa maggiore di  $1/RC$ . In corrispondenza di una pulsazione pari a dieci volte  $1/RC$ , lo sfasamento raggiunge già  $-84.3^\circ$ .

Vediamo un altro esempio di funzione di trasferimento, questa volta caratterizzata da uno zero ed un polo, entrambi reali. Lo zero è pari a  $-1/\tau_z$  ed il polo è pari a  $-1/\tau_p$ .

Procedendo analogamente a quanto fatto in precedenza, si può determinare il modulo della funzione di trasferimento che risulta pari a  $20\log\sqrt{1+\omega^2\tau_z^2} - 20\log\sqrt{1+\omega^2\tau_p^2}$ .

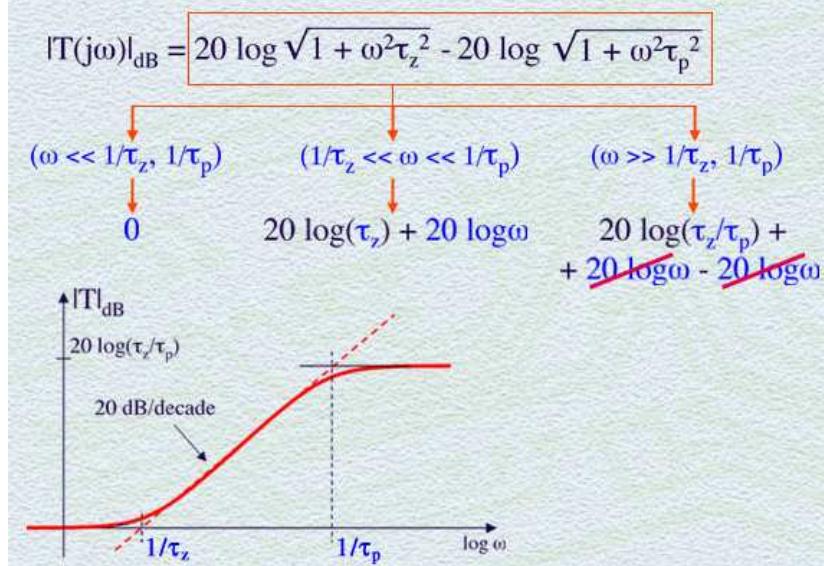
Vediamo ora come rappresentare il diagramma di Bode del modulo, per semplicità nell'ipotesi che le due singolarità siano sufficientemente separate tra loro, ad esempio di almeno una decade.

Supponiamo, per fissare le idee, che lo zero si trovi a ad una frequenza inferiore a quella del polo, cioè  $\tau_z$  sia molto maggiore di  $\tau_p$ .

Anche in questo caso risulta comodo stimare l'andamento del modulo della funzione di trasferimento prendendo in considerazione tre diversi intervalli sull'asse delle frequenze.

Consideriamo innanzitutto l'intervalle in cui  $\omega$  sia molto inferiore sia dello zero che del polo. In questo caso, analogamente all'esempio precedente, il termine sotto entrambe le radici è pressoché pari a 1, e pertanto il modulo della funzione di trasferimento è prossimo a zero. Se ora consideriamo l'intervalle di frequenza compreso tra lo zero ed il polo, è lecito supporre, data la distanza relativa tra zero e polo, che nella prima radice, il termine  $\omega^2\tau_z^2$  domini sull'“1”, mentre nella seconda radice, il termine  $\omega^2\tau_p^2$  sia ancora trascurabile rispetto ad 1. Alla luce di questa osservazione, il modulo di  $T(j\omega)$  si può approssimare con  $20\log(\tau_z) + 20\log(\omega)$ . Osserviamo che per la proporzionalità al logaritmo di  $\omega$ , il modulo della funzione di trasferimento cresce linearmente con pendenza pari a 20 dB/decade.

### Modulo



$$\text{zero: } -\frac{1}{\tau_z}$$

$$\text{polo: } -\frac{1}{\tau_p}$$

$$|T(j\omega)|_{dB} = 20 \log \sqrt{1 + \omega^2\tau_z^2} - 20 \log \sqrt{1 + \omega^2\tau_p^2}$$

Ipotesi  $\tau_z >> \tau_p$  (freq. zero minore della freq. polo)

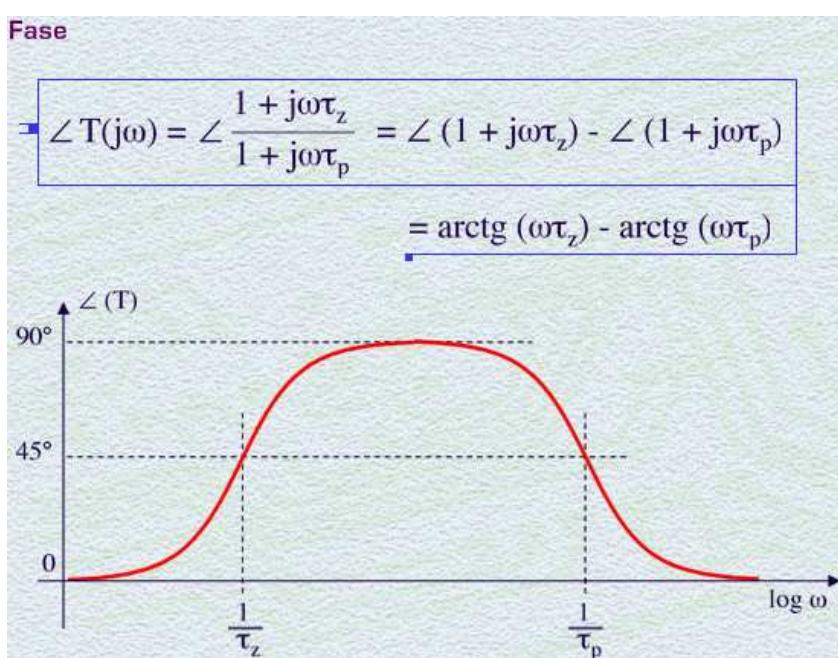
Se infine consideriamo una regione di frequenze molto superiori ad entrambi zero e polo, anche il termine integrando nella seconda radice è praticamente pari al solo  $\omega^2\tau_p^2$ . Pertanto, nell'espressione del modulo di  $T(j\omega)$  che ne consegue, si aggiunge rispetto alla forma precedente, oltre ad un termine costante  $-20\log(\tau_p)$ , anche un termine  $-20\log(\omega)$  che si semplifica con  $20\log(\omega)$ . In questa regione, come mostrato in figura, il modulo diviene pertanto costante, pari a  $20\log(\tau_z/\tau_p)$ . Ciò può essere anche riassunto osservando che la presenza

del polo, come visto nell'esempio precedente, introduce nell'andamento del modulo una pendenza di  $-20\text{dB/decade}$ , che in questo caso va a compensare la pendenza positiva di  $20\text{dB/decade}$  precedentemente introdotta dallo zero. Nel caso di reti più complesse in cui siano presenti più zeri e poli, sufficientemente separati tra loro, il diagramma di Bode del modulo della funzione di trasferimento può essere facilmente tracciato secondo il procedimento visto, inserendo nel diagramma una pendenza di  $20\text{dB/decade}$  in corrispondenza di ogni zero e una pendenza di  $-20\text{dB/decade}$  in corrispondenza di ogni polo.

Vediamo infine lo sfasamento introdotto dalla funzione di trasferimento. L'argomento della  $T(j\omega)$  risulta pari alla differenza tra l'argomento del numeratore e quello del denominatore. Procedendo in maniera analoga a quanto fatto nel primo esempio, l'argomento del termine  $1+j\omega\tau_z$  risulta pari all' $\text{arctg}(\omega\tau_z)$ , mentre l'argomento del termine  $1+j\omega\tau_p$  risulta pari all' $\text{arctg}(\omega\tau_p)$ . Considerando che lo zero è a più bassa frequenza rispetto al polo, la fase di  $T(j\omega)$  crescerà secondo la prima arcotangente raggiungendo un valore pari a  $45^\circ$  in corrispondenza dello zero e un valore prossimo a  $90^\circ$  per frequenze comprese tra lo zero ed il polo. Successivamente interviene la seconda arcotangente che riduce, via via che la frequenza cresce, il valore di fase raggiunto dalla prima arcotangente. Per

frequenze molto maggiori sia dello zero che del polo, la fase ritorna a zero, cioè lo sfasamento di  $90^\circ$  introdotto dallo zero è stato completamente compensato dallo sfasamento di  $-90^\circ$  introdotto dal polo.

Anche per quanto riguarda la fase è opportuno osservare che nel caso di reti con più poli e zeri, è possibile descrivere l'andamento complessivo della fase della funzione di trasferimento considerando che ogni zero negativo introduce uno sfasamento di  $+90^\circ$  e ogni zero positivo introduce uno sfasamento di  $-90^\circ$ .

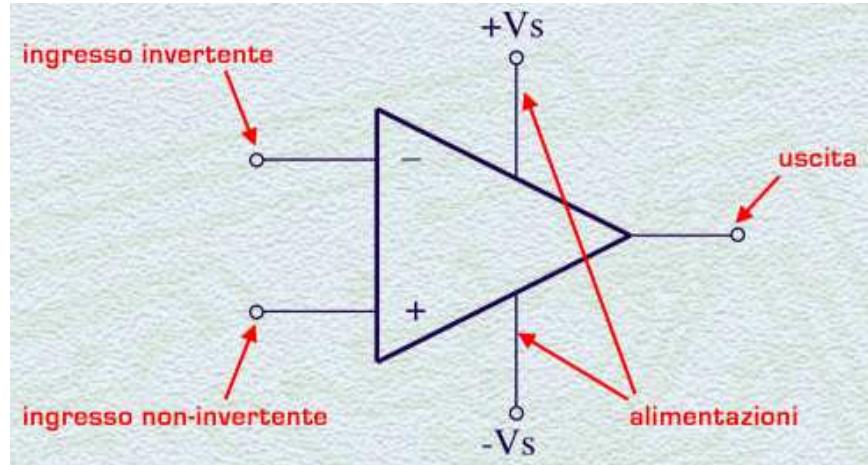


## Lezione n. 25 : L'amplificatore operazionale

### Principio di funzionamento, Modello dell'A.O. ideale

Introduciamo innanzitutto il simbolo circuitale dell'Amplificatore Operazionale. Questo dispositivo è generalmente rappresentato da un triangolo con due morsetti d'ingresso a sinistra ed un morsetto di uscita a destra. Per ragioni che appariranno chiare non appena introdurremo il principio di funzionamento del dispositivo, il morsetto d'ingresso in corrispondenza del segno '-' in figura, prende il nome di **ingresso invertente**, mentre il morsetto in corrispondenza del segno '+', prende il nome di **ingresso non-invertente**. Il morsetto di uscita è unico.

Naturalmente, come ogni circuito elettronico, il dispositivo deve essere opportunamente alimentato.

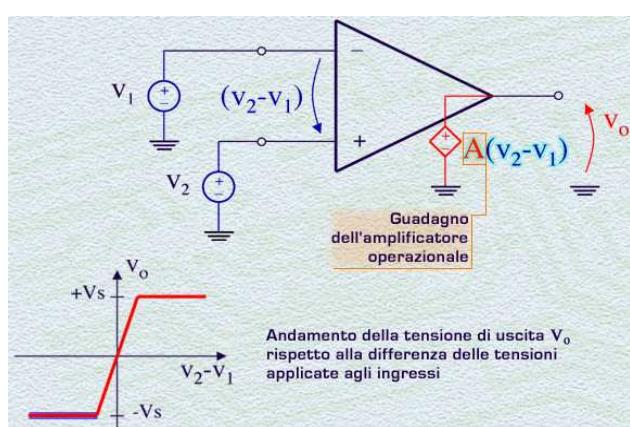


Pertanto, il simbolo completo comprende anche due alimentazioni, una positiva da noi indicata come  $+Vs$  ed una negativa  $-Vs$ . A volte è possibile alimentare l'amplificatore anche nella modalità a **singola polarità** e cioè fissando il morsetto  $-Vs$  a massa ed applicando solo una tensione di alimentazione positiva a  $+Vs$ .

Può capitare, nella rappresentazione schematica di circuiti contenenti uno o più A.O., e in cui l'attenzione sia rivolta solo alla funzionalità del circuito, che i morsetti relativi alle due alimentazioni vengano omessi.

Vediamo ora qual è il *principio di funzionamento* dell'amplificatore operazionale.

Supponiamo di applicare una tensione  $V_1$  tra il morsetto invertente e massa ed una tensione  $V_2$  tra il morsetto non-invertente e massa. Questa operazione può anche essere vista come l'applicazione di una differenza di tensione tra i due ingressi, non invertente e invertente, pari a  $V_2 - V_1$ . Come risposta all'applicazione di queste tensioni, l'A.O. fornisce alla sua uscita un valore di tensione, riferito a massa, pari alla differenza di tensione applicata  $V_2 - V_1$ , moltiplicata per un termine di guadagno  $A$ . In sintesi, quindi, l'A.O. amplifica la differenza di tensione applicata ai suoi ingressi.



Il fattore  $A$  prende il nome di **guadagno dell'amplificatore operazionale**.

Si può osservare che, relativamente alle singole tensioni applicate ai morsetti d'ingresso, l'A.O. ha un funzione invertente per la tensione  $V_1$ , in quanto questa viene moltiplicata per un fattore pari a  $-A$ , mentre ha una funzione non invertente per la tensione  $V_2$ , in quanto questa viene moltiplicata per un fattore pari ad  $A$ . Ciò ci consente di comprendere il significato dei termini invertente e non-invertente che erano stati precedentemente attribuiti ai due morsetti d'ingresso. Sulla base del

principio di funzionamento descritto, tracciamo ora in un grafico l'andamento della tensione di uscita  $V_o$  rispetto alla differenza delle tensioni applicate agli ingressi, cioè  $V_2 - V_1$ . La tensione  $V_o$  cresce all'aumentare della differenza  $V_2 - V_1$ , con una pendenza tanto più ripida quanto maggiore è il valore del guadagno  $A$ . La tensione di uscita non può raggiungere un valore infinito ma si saturerà ad un valore che è prossimo alla tensione di alimentazione positiva del dispositivo  $+Vs$ , che ovviamente non può essere oltrepassata. Analogamente quando la differenza di tensione  $V_2 - V_1$  raggiunge valori sempre più negativi, l'uscita si porterà a tensioni sempre più negative, con una pendenza proporzionale al guadagno  $A$ , fino a saturare ad un valore prossimo alla tensione di alimentazione negativa  $-Vs$ .

Una volta compreso il principio di funzionamento dell'A.O. vediamo di completarne il modello. Spesso, al solo fine di comprendere le funzionalità di un circuito elettronico che faccia uso di un A.O., è sufficiente utilizzare un modello semplificato di questo dispositivo, che chiameremo "**modello ideale**". Questo modello facilita il calcolo delle grandezze elettriche di interesse. Successivamente, al fine di determinare con maggior dettaglio le caratteristiche del circuito in esame, è possibile introdurre un modello più completo dell'A.O., che chiameremo "**modello reale**" e che vedremo più avanti, il quale tiene conto delle non idealità del dispositivo.

La prima idealità dell'A.O. che consideriamo consiste nel supporre che questo dispositivo presenti una **impedenza ai suoi morsetti d'ingresso infinita**.

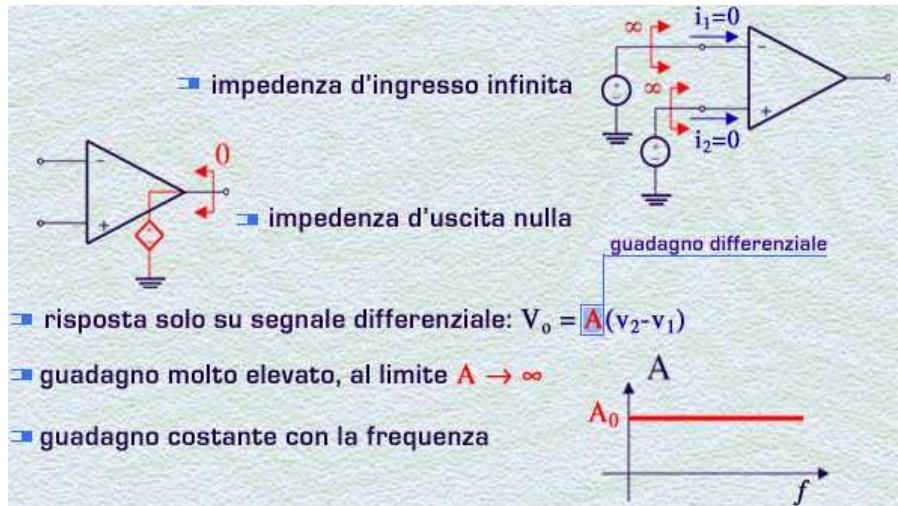
Ciò equivale a dire che quando applichiamo un generatore di tensione su ognuno dei due ingressi, la corrente assorbita dall'amplificatore attraverso i due morsetti risulta sempre nulla.

Al contrario, nel modello ideale dell'A.O. si suppone che **l'impedenza di uscita sia sempre nulla**. Ciò equivale a supporre che la tensione di uscita dell'amplificatore sia fornita da un generatore ideale di tensione, il quale è comandato dalla differenza di tensione all'ingresso.

Un'altra idealità di questo modello consiste nel supporre che l'uscita dell'amplificatore dipenda esclusivamente dalla differenza delle tensioni applicate ai suoi ingressi. Questa proprietà, che di fatto avevamo già assunto implicitamente nel descrivere il principio di funzionamento del dispositivo, esclude la possibilità che la tensione di uscita possa variare qualora venga applicata la stessa tensione su entrambi gli ingressi. Proprio per il fatto che l'uscita varia alla sola applicazione di un segnale differenziale in ingresso, il guadagno **A** dell'A.O. è anche definito **guadagno differenziale**.

Un'altra assunzione che viene fatta sul guadagno differenziale è che esso sia di valore estremamente elevato, al limite infinito.

Infine, si suppone, nel modello ideale, che il guadagno sia costante per ogni frequenza.

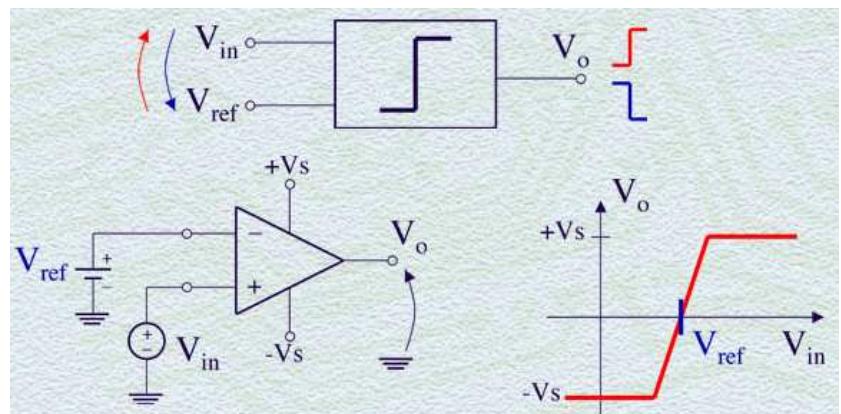


### Esempi di applicazione dell'A.O.: il comparatore, l'amplificatore invertente

Introdotto il principio di funzionamento dell'A.O. e il suo modello ideale, vediamo un primo esempio di applicazione di questo dispositivo; il *circuito comparatore*.

Per circuito comparatore si intende un circuito che opera un confronto tra una tensione in ingresso, che chiamiamo  $V_{in}$ , e una tensione costante di riferimento, che chiamiamo  $V_{ref}$ . Il circuito produce alla sua uscita un valore costante di tensione di un segno, ad esempio positivo, se  $V_{in}$  è maggiore di  $V_{ref}$ , mentre produce un valore di tensione di segno contrario, negativo nel nostro esempio, se  $V_{in}$  è minore di  $V_{ref}$ .

Utilizzando il solo A.O. è possibile realizzare il circuito comparatore appena descritto. Se consideriamo un A.O., in cui applichiamo all'ingresso invertente la tensione di riferimento e applichiamo all'altro ingresso, quello non invertente, la tensione di cui vogliamo fare il confronto, è facile comprendere come questo circuito realizzi di fatto un comparatore. Infatti, se la tensione  $V_{in}$  supera la tensione di riferimento, l'uscita dell'A.O. si porta ad un valore positivo elevato,



in virtù dell'elevato guadagno differenziale, fino a raggiungere un valore prossimo all'alimentazione positiva  $+V_s$ . Quando la tensione  $V_{in}$ , invece, è inferiore alla tensione di riferimento, la tensione di uscita dell'A.O. si porta a valori sempre più negativi, fino a saturare ad un valore prossimo all'alimentazione negativa  $-V_s$ . Nel grafico in figura precedente è mostrato l'andamento della tensione di uscita rispetto alla tensione di ingresso ottenuto nel modo appena descritto. Si può notare come questa curva non sia altro che la curva caratteristica dell'A.O. vista in precedenza, centrata nel punto dell'asse delle ascisse in cui  $V_{in}$  è uguale a  $V_{ref}$ .

Come secondo esempio di applicazione dell'A.O. introduciamo un circuito molto importante in elettronica: *l'amplificatore invertente*.

In questo circuito, l'ingresso non invertente è posto direttamente a massa, mentre una resistenza  $R_1$  collega il generatore di segnale di ingresso  $V_{in}$  con l'ingresso invertente dell'A.O. Infine, una resistenza  $R_2$  collega l'uscita dell'amplificatore con l'ingresso invertente.

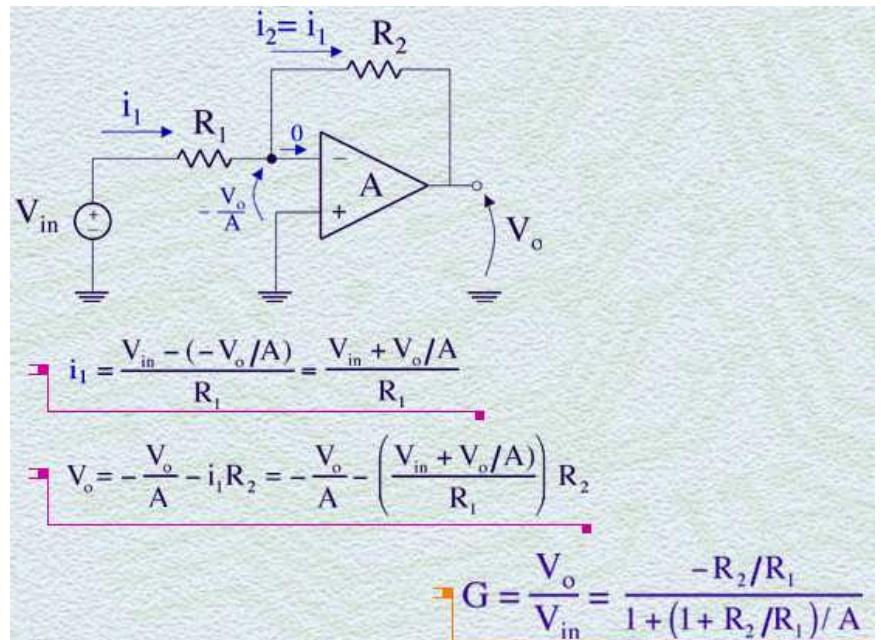
Vediamo ora come calcolare la relazione che lega la tensione di uscita del circuito  $V_o$ , misurata rispetto a massa, con la tensione di ingresso, utilizzando il modello per l'A.O. precedentemente introdotto. Supporremo all'inizio che A.O. sia caratterizzato da un guadagno  $A$  finito e successivamente valuteremo il comportamento del circuito per un valore del guadagno che tenda all'infinito.

Iniziamo con l'osservare che data la relazione costitutiva che lega l'uscita dell'A.O. con la tensione differenziale applicata tra i suoi ingressi, è possibile dire che la differenza di tensione misurata tra ingresso invertente e ingresso non invertente, come indicata in figura, è pari al rapporto, preceduto dal segno meno, tra la tensione di uscita  $V_o$  e il guadagno  $A$ . Dato che l'ingresso non invertente è posto a massa, la tensione al morsetto invertente è pari sempre a  $-V_o/A$ . Di conseguenza, entrambi i valori di tensione ai capi della resistenza  $R_1$  sono noti, ed è quindi possibile determinare la corrente  $i_1$  che scorre nella resistenza  $R_1$ . Questa corrente sarà pari alla differenza tra  $V_{in}$  e  $(-V_o/A)$ , il tutto diviso per  $R_1$  e quindi, semplificando, sarà uguale a  $V_{in} + V_o/A$ , il tutto diviso per  $R_1$ .

Supponendo ora che l'A.O. sia ideale e in particolare che non assorba corrente ai suoi morsetti d'ingresso in quanto ha un'impedenza d'ingresso infinita, risulta evidente che la corrente  $i_1$  non può fluire nel morsetto non invertente e quindi possa fluire interamente nella resistenza  $R_2$ . Nota quindi anche la corrente  $i_2$ , pari a  $i_1$ , è possibile determinare facilmente il valore della tensione di uscita  $V_o$ . Questa è pari alla tensione sul morsetto non invertente,  $-V_o/A$ , più la caduta di tensione sulla resistenza  $R_2$ , che è pari a  $-i_2 R_2$  per  $R_2$ . Introducendo in questa uguaglianza la relazione precedentemente calcolata che lega la corrente  $i_1$  alla tensione di ingresso  $V_{in}$  e alla tensione di uscita  $V_o$  si giunge così ad una uguaglianza in cui compaiono come grandezze elettriche solo  $V_{in}$  e  $V_o$ .

Se ora raccogliamo tutti i termini relativi a  $V_o$  si giunge facilmente a calcolare il rapporto tra la tensione di uscita  $V_o$  e quella di ingresso  $V_{in}$ . Questa relazione, mostrata in figura, rappresenta il guadagno dell'amplificatore invertente che deve questo nome proprio al fatto che tale guadagno, come mostrato nella formula, è negativo, cioè l'uscita è invertita rispetto all'ingresso.

La relazione che abbiamo ricavato vale nel caso generale di un guadagno  $A$  finito. Vediamo ora di valutare le proprietà di questo circuito nel caso ideale in cui il guadagno sia molto elevato, al limite infinito. Risulta evidente dalla relazione appena ricavata, che per  $A$  che tende all'infinito, il



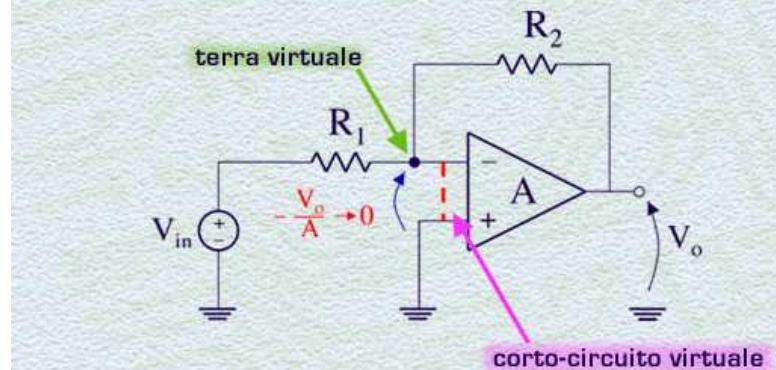
denominatore tende a uno e pertanto il guadagno dell'amplificatore invertente tende a essere pari al rapporto, con il segno meno, tra la resistenza  $R_2$  e la resistenza  $R_1$ .

Questa semplice relazione non dipende più dal valore del guadagno dell'A.O. ma solo dal rapporto delle due resistenze. Se, ad esempio, si vuole progettare un amplificatore invertente che abbia un guadagno pari a  $-100$ , sarà sufficiente scegliere le resistenze  $R_2$  e  $R_1$  in maniera tale che l'una sia rispettivamente cento volte maggiore dell'altra.

Sempre nell'ipotesi di guadagno infinito, ritorniamo al nostro circuito per fare un'importante osservazione.

Consideriamo la differenza di tensione tra gli ingressi invertente e non invertente dell'A.O., precedentemente calcolata come  $-V_o/A$ . Se il guadagno  $A$  tende all'infinito, questa differenza di tensione tende a zero. Questa importante proprietà viene comunemente descritta definendo che tra i due morsetti di ingresso dell'A.O. sia presente un *corto-circuito virtuale*. Questo termine non sta ad indicare che essi siano fisicamente collegati, come nel caso di un corto circuito vero, ma che di fatto, per le proprietà appena viste, essi si trovino ad avere una differenza di tensione nulla tra di loro. Inoltre, se si tiene conto che il morsetto non invertente è posto direttamente a massa, è possibile concludere che il morsetto invertente si trova ad essere posto anch'esso ad una tensione nulla, grazie al corto circuito virtuale. Per questa ragione, questo morsetto, in questa particolare configurazione, viene definito *terra virtuale*.

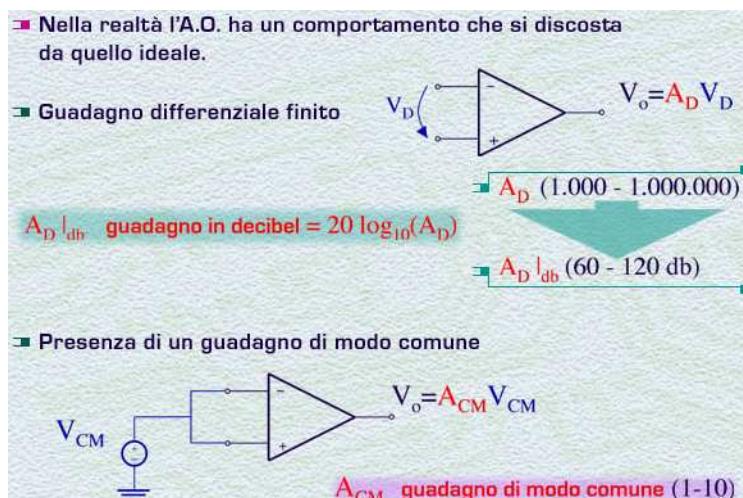
$$G = \frac{V_o}{V_{in}} = \frac{-R_2/R_1}{1 + (1 + R_2/R_1)/A} \xrightarrow{A \rightarrow \infty} -\frac{R_2}{R_1}$$



### Principali non idealità dell'A.O.

Abbiamo visto in precedenza come il modello ideale dell'A.O. sia utile per valutare velocemente le funzioni di un circuito che contenga questo dispositivo. Nella realtà, però, l'A.O. ha un comportamento che si discosta da quello ideale e ciò va tenuto conto in un'analisi più approfondita del circuito.

Noi non ci addentreremo nelle cause delle non-idealità dell'A.O. in quanto non ci interessiamo, in questo ambito, della sua struttura interna.



Introduciamo invece un modello "reale" che consenta di valutare l'effetto di queste non idealità nei circuiti che facciano uso di uno o più A.O.

L'A.O. innanzitutto ha un guadagno differenziale elevato ma non infinito. A seconda del tipo di dispositivo, il guadagno differenziale può avere valori che vanno da qualche migliaio fino anche più di un milione. In elettronica, si usa definire il guadagno differenziale, come molte altre grandezze, in decibel. Il valore in decibel è pari a venti volte il logaritmo

in base dieci del valore del guadagno. Pertanto, valori tipici del guadagno in decibel possono trovarsi compresi tra 60 e 120 dB. Una seconda non idealità è rappresentata dalla presenza anche di un **guadagno di modo comune**. Ciò equivale a dire che quando si applica un segnale di pari valore su entrambi gli ingressi, detto anche **segnale di modo comune**, la tensione di uscita anziché restare invariata, come supposto nel modello ideale, cambia valore. È possibile definire quindi un guadagno di modo comune dell'A.O., che chiameremo  $A_{CM}$ , tale per cui la variazione della tensione di uscita  $V_o$  in seguito all'applicazione di segnale di modo comune in ingresso  $V_{CM}$  vale  $A_{CM}$  per  $V_{CM}$ . Il valore di

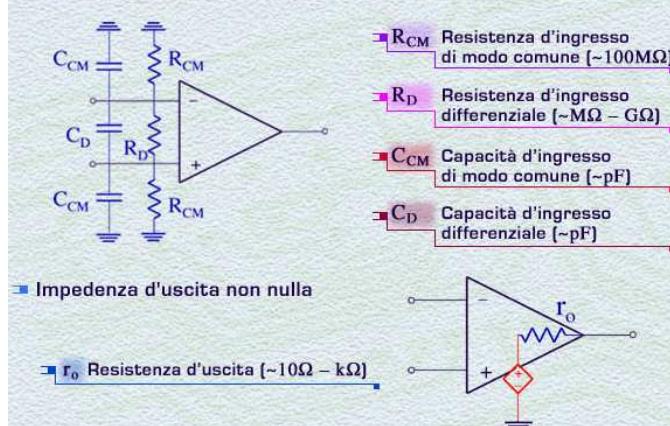
guadagno di modo comune dipende molto anch'esso dal tipo di amplificatore considerato. Tipici valori sono compresi tra uno e dieci.

Se consideriamo il comportamento di un A.O. in presenza di due generatori di tensione  $V_1$  e  $V_2$  applicati rispettivamente al morsetto invertente e a quello non invertente, è possibile pensare equivalentemente all'applicazione di una tensione di modo comune  $V_{CM}$ , pari alla media delle due tensioni applicate, cioè  $(V_1+V_2)/2$ , a cui si sovrappone una tensione differenziale  $V_D$ , pari a  $V_2-V_1$ , applicata tra i due ingressi. La corrispondente tensione di uscita risulta pari a  $V_D$  moltiplicata per  $A_D$ , più  $V_{CM}$  moltiplicata per  $A_{CM}$ . La funzione di interesse dell'A.O. è rappresentata dal suo guadagno differenziale, che quindi si cerca di ottenere di valore più elevato possibile, mentre il guadagno di modo comune è solitamente un parametro indesiderato che pertanto si cerca di tenere ad un valore il più basso possibile. Il parametro di merito dell'A.O. sotto questo aspetto è la **Reiezione di modo comune**, che è definita come il rapporto, calcolato in decibel, tra il guadagno di modo differenziale ed il guadagno di modo comune.

Un'altra non idealità dell'A.O. è rappresentata dall'**impedenza d'ingresso non infinita**.

Il modello reale tiene conto di questa non-idealità sia per il **modo comune** che per il **modo differenziale**. Per l'impedenza di ingresso di modo comune, il modello reale prevede una resistenza d'ingresso  $R_{CM}$  su ognuno degli ingressi. Valori di questa resistenza sono tipicamente dell'ordine dei

#### ■ Impedenza d'ingresso non infinita



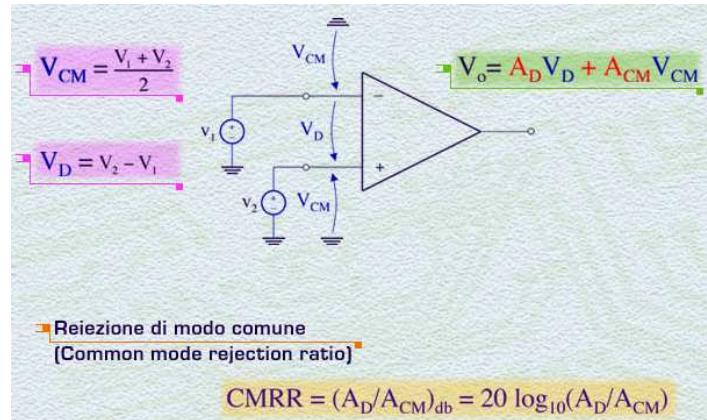
100MΩ. L'impedenza non infinita su modo differenziale è tenuta in conto introducendo nel modello una resistenza tra i due morsetti di ingresso dell'A.O. Essa può assumere valori molto differenti, a seconda della tecnologia di realizzazione dell'A.O., compresi tra qualche MΩ e qualche GΩ. È possibile tenere conto anche di una impedenza di ingresso di tipo capacitivo, introducendo nel modello una capacità di modo comune  $C_{CM}$  su ognuno dei morsetti ed una capacità di modo differenziale  $C_D$  tra i due morsetti. Valori tipici per entrambe sono dell'ordine del pF. L'**impedenza di uscita**

**di un A.O. reale non è nulla**. Per tenere conto di questo, nel modello reale si introduce una resistenza  $r_o$  in serie al generatore di tensione comandato. La resistenza di uscita può assumere valori che vanno da qualche decina di Ω fino a qualche kΩ.

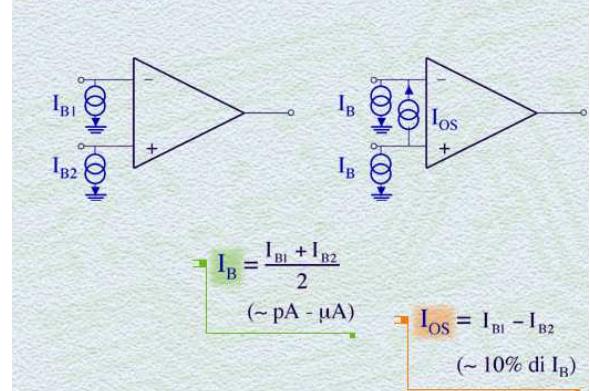
Anche in condizioni statiche, senza l'applicazione di segnali ai suoi ingressi, negli A.O. reali vi sono sempre delle correnti, generalmente molto piccole, che scorrono nei suoi morsetti di ingresso.

Queste correnti vengono definite **correnti di bias** e sono rappresentate nel modello reale da due generatori di corrente  $I_{B1}$  e  $I_{B2}$  applicati ai morsetti di ingresso del modello ideale.

Alternativamente a questa rappresentazione, è possibile anche introdurre due generatori di pari valore  $I_B$ , calcolato come media delle due correnti  $I_{B1}$  e  $I_{B2}$  indicate nella rappresentazione precedente, e aggiungere un generatore tra i due morsetti, di valore  $I_{OS}$  pari alla differenza tra le correnti  $I_{B1}$  e  $I_{B2}$ . Le due rappresentazioni sono equivalenti. Valori di  $I_B$  vanno dal pA al μA, mentre  $I_{OS}$  è generalmente dell'ordine del 10% di  $I_B$ .

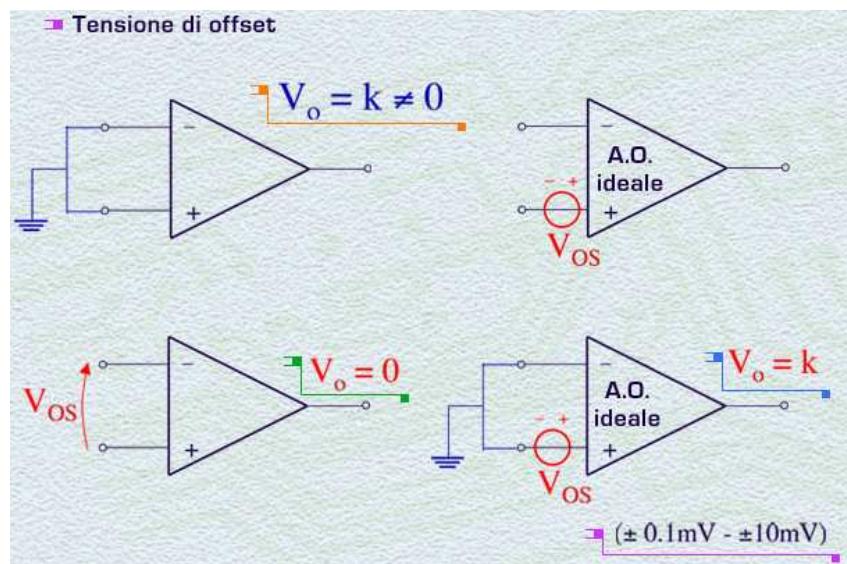


#### ■ Correnti di bias



Un altro parametro del modello reale è la **tensione di offset**. Questo parametro tiene conto del seguente fenomeno che caratterizza un A.O. reale. Quando gli ingressi di un A.O. vengono entrambi messi a massa, idealmente ci si aspetta che la tensione di uscita si trovi anch'essa a zero.

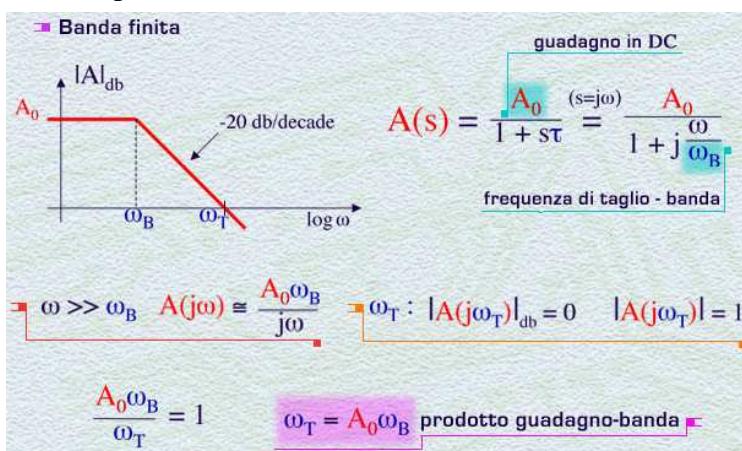
Nella realtà ciò non avviene e generalmente la tensione di uscita si porta ad un valore  $k$  che è, anche se di poco, diverso da zero. Per tenere conto di questa limitazione, si introduce nel modello reale un generatore di tensione, chiamata **tensione di offset**, su uno dei due morsetti del modello ideale dell'A.O. La tensione di offset del generatore introdotto è definita come la tensione differenziale che deve essere applicata in ingresso all'A.O per riportare la tensione di uscita dal valore  $k$  a zero.



Infatti sulla base di questa definizione, ponendo a massa gli ingressi di un A.O. ideale in cui è stato inserito in ingresso il generatore di tensione di offset precedentemente determinato, la tensione di uscita si porterà al valore  $k$  riscontrato nel dispositivo reale. Valori della tensione di offset, di cui generalmente non viene definito il segno dal costruttore, sono compresi tra  $\pm 0.1\text{mV}$  e  $\pm 10\text{mV}$ .

Un'altra limitazione dell'A.O. reale consiste nell'avere una **banda limitata**. Ciò equivale a dire che il guadagno è costante solo fino ad una certa frequenza, che definisce la **banda** del dispositivo. A più alte frequenze, il guadagno inizia a diminuire, con un andamento tipico che dipende da come è costituito internamente l'amplificatore.

Nella figura è mostrato il diagramma di Bode del modulo del guadagno, nel modello più semplice utilizzato per descrivere la risposta in frequenza dell'A.O. In questo modello il guadagno resta costante fino ad una frequenza  $\omega_B$  e per frequenze maggiori decresce con una pendenza costante di  $-20\text{dB/decade}$ . Utilizzando una notazione molto comune per descrivere in generale il comportamento in frequenza dei sistemi elettronici, il modello rappresentato è un tipico sistema ad un **polo**, la cui funzione di trasferimento, espressa rispetto alla variabile complessa  $s$ , risulta pari ad un guadagno costante  $A_0$  diviso per il termine  $(1 + s\tau)$ . Tale espressione è anche esprimibile in funzione delle frequenze fisiche ponendo  $s$  uguale a  $j\omega$  e diventa quindi pari a  $A_0/[1+j(\omega/\omega_B)]$ . La pulsazione  $\omega_B$  definisce la banda o la frequenza di taglio dell'amplificatore.



ci capiterà di chiamare correntemente la variabile  $\omega$  indicata nei diagrammi di Bode come frequenza. E' opportuno ricordare, naturalmente, che la pulsazione  $\omega$  è in verità pari alla frequenza  $f$  moltiplicata per  $2\pi$ . Proseguendo nell'analisi del diagramma di Bode mostrato nella figura, si può osservare che per frequenze maggiori di  $\omega_B$ , l'espressione del guadagno si può approssimare a  $(A_0 * \omega_B)/j\omega$ . Poniamoci ora l'obiettivo di calcolare la frequenza  $\omega_T$  in corrispondenza della quale il guadagno espresso in dB è pari a zero, o equivalentemente il suo valore assoluto è pari a uno. Ciò equivale a determinare la frequenza in cui il diagramma di Bode del guadagno interseca l'asse delle ascisse. Imponendo questa condizione nella relazione appena ricavata, si calcola immediatamente che la frequenza  $\omega_T$  è pari al prodotto tra il guadagno in continua  $A_0$  e la banda  $\omega_B$ . Per questa ragione, la frequenza  $\omega_T$  viene definita come **prodotto guadagno-bandiera**.

Come già accennato in precedenza, la risposta in frequenza dell'A.O. dipende dal tipo di dispositivo.

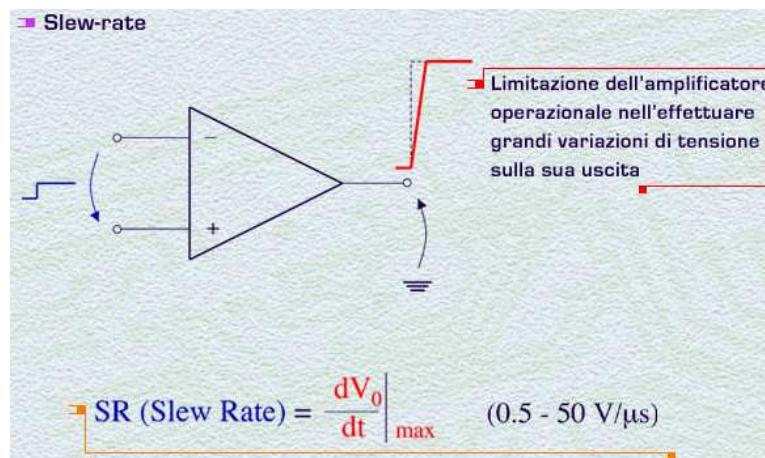
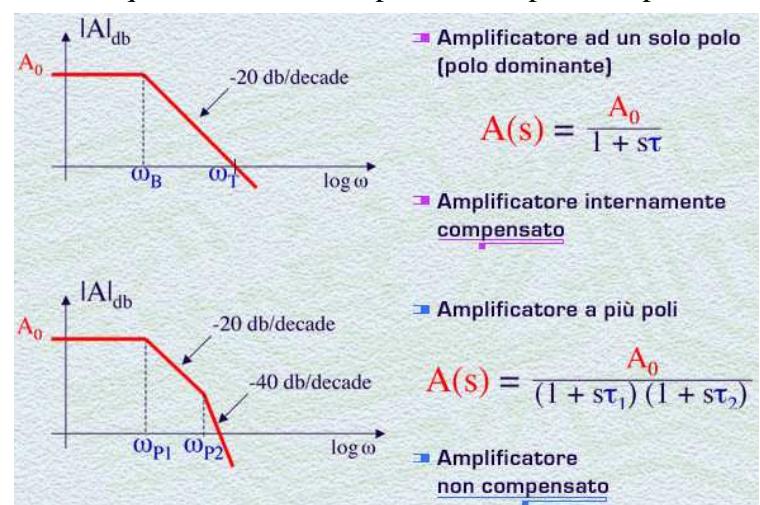
Il modello fin qui descritto, di cui si ripropone il diagramma di Bode, è rappresentativo della classe di amplificatori ad un singolo polo, o a polo dominante. Per ragioni che appariranno più chiare in seguito nel corso l'amplificatore è anche definito **internamente compensato**.

Un altro modello di risposta in frequenza suppone la presenza di due poli, come anche indicato nell'espressione matematica del guadagno in funzione della variabile complessa s. In questo caso il guadagno è costante fino alla frequenza  $\omega_{P1}$ , decresce con una pendenza di -20dB/decade fino in prossimità della frequenza  $\omega_{P2}$  e successivamente decresce con una pendenza pari a -40dB/decade.

Questo tipo di amplificatore si definisce **non compensato**.

Consideriamo infine l'ultima non idealità del nostro modello di A.O. reale: lo *slew-rate*. Questo effetto si osserva negli amplificatori quando si applica un segnale differenziale in ingresso a cui deve corrispondere una grande variazione della tensione di uscita. Nel caso ideale, ci si aspetta che l'uscita sia in grado di variare la sua tensione con un fronte molto rapido. In verità, l'andamento dell'uscita nel tempo è caratterizzato da una pendenza finita, indice di una limitazione del dispositivo ad effettuare grandi variazioni di tensione alla sua uscita.

Lo *slew-rate* è definito come la massima derivata della tensione di uscita rispetto al tempo e può generalmente assumere valori compresi tra 0.5 e 50V/ $\mu$ s.



massima derivata della tensione di uscita rispetto al tempo e può generalmente assumere valori compresi tra 0.5 e 50V/ $\mu$ s.

## Lezione n. 26 : La Controreazione

### Principi base della Controreazione

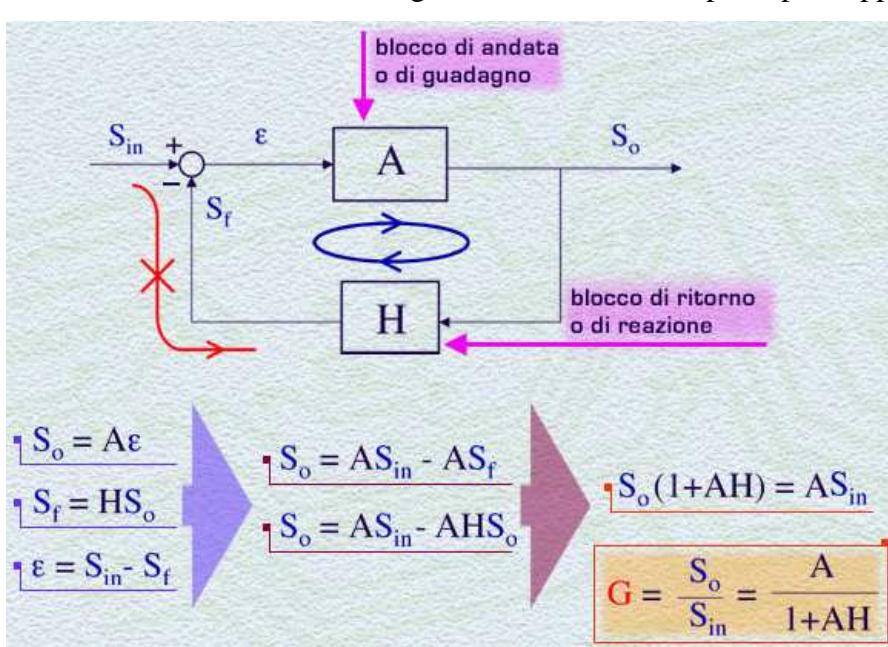
L'idea della Controreazione, o reazione negativa, si deve all'americano Harold Black, che nel 1927, all'età di 29 anni, lavorava come ingegnere presso i laboratori della compagnia telefonica americana (gli attuali Bell Telephone Laboratories) nell'ambito dello sviluppo di sistemi di comunicazione telefonica a grande distanza. Scopo delle ricerche di Black era quello di studiare soluzioni che riducessero gli elevati problemi di **distorsione e instabilità** presenti in sistemi basati su amplificatori caratterizzati da una **non sufficiente linearità**, stabilità e soprattutto da un **comportamento dipendente dalla temperatura e dall'invecchiamento del componente**.

L'introduzione, da parte di Harold Black, del concetto di circuito reazionato negativamente, ha rappresentato non solo una brillante soluzione dei problemi accennati ma più in generale una svolta in diversi settori dell'elettronica moderna. Le proprietà della controreazione hanno consentito di ottenere un elevato controllo dei parametri caratteristici dei circuiti e sistemi elettronici, con la possibilità soprattutto di ridurre gli effetti delle non-linearità introdotte dagli amplificatori utilizzati e di rendere stabili e riproducibili le prestazioni dei circuiti.

Iniziamo con l'introdurre lo schema generale di un circuito reazionato negativamente. Questo è costituito, come indicato in figura, da due blocchi principali opportunamente collegati. Il blocco superiore, indicato con **A**, prende il nome di blocco di andata o di guadagno mentre il blocco inferiore indicato con **H** prende il nome di blocco di ritorno o blocco di reazione. Il guadagno di entrambi i blocchi in questo schema è supposto essere positivo.

Il blocco superiore, indicato con **A**, prende il nome di blocco di andata o di guadagno mentre il blocco inferiore indicato con **H** prende il nome di blocco di ritorno o blocco di reazione. Il guadagno di entrambi i blocchi in questo schema è supposto essere positivo.

Descriviamo ora come avviene la circolazione dei segnali nel sistema indicato. A tale scopo, ci riferiamo genericamente a dei segnali elettrici senza specificare se si tratti di segnali di tensione o di corrente. La trattazione, a questo riguardo, è del tutto equivalente.



Quando un segnale  $S_{in}$  viene applicato all'ingresso del circuito, esso viene all'inizio trasmesso direttamente all'ingresso del blocco di andata. Il segnale di uscita  $S_o$  risulterà pertanto pari al segnale  $S_{in}$  amplificato del fattore **A**. L'uscita  $S_o$  viene prelevata e inviata in ingresso al blocco di reazione, che di conseguenza produrrà alla sua uscita un segnale  $S_f$  che è pari a  $S_o$  per il guadagno **H**. All'ingresso del circuito è presente un nodo sottrattore che opera la differenza tra il segnale  $S_{in}$  ed il segnale  $S_f$ . Questa differenza è riportata in ingresso al blocco di andata **A**. Mentre in corrispondenza della prima applicazione del segnale  $S_{in}$ , il segnale  $S_f$  era nullo e quindi l'intero valore  $S_{in}$  era portato in ingresso al blocco di guadagno, di seguito il blocco di reazione riportare un segnale  $S_f$  non nullo in ingresso, e quindi il segnale differenza all'ingresso del blocco **A**, che chiamiamo  $\epsilon$ , risulterà inferiore al segnale iniziale  $S_{in}$ .

In sintesi quindi, nel sistema appena descritto, l'uscita viene riportata in ingresso tramite il blocco di reazione e sottratta al segnale in ingresso. Per questa ragione il sistema si dice reazionato negativamente. Conseguenza della reazione è quella di **ridurre l'effetto provocato sul sistema dall'applicazione del segnale di ingresso**.

Occorre precisare che il procedimento fin qui descritto si basa sull'ipotesi che il segnale possa percorrere l'anello composto dai blocchi **A**, **H** e dal nodo sottrattore solo secondo il senso orario, indicato in figura. Si esclude invece, che vi possa essere anche un trasferimento diretto del segnale

dall'ingresso all'uscita tramite il blocco **H**. Questo trasferimento è indicato in figura con il percorso tracciato in rosso.

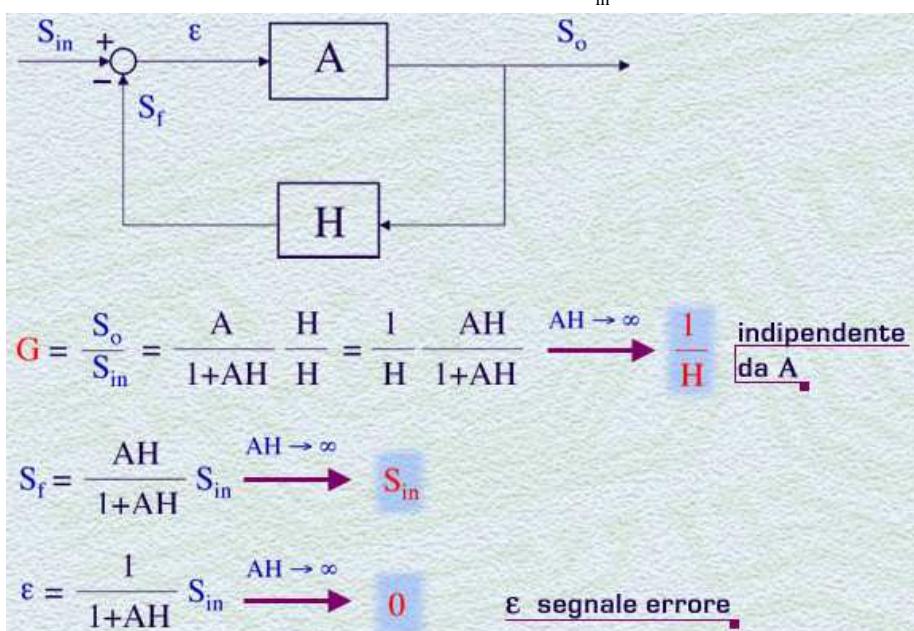
Procediamo ora a calcolare il guadagno del sistema reazionato negativamente e cioè il rapporto tra il segnale di uscita  $S_o$  e quello di ingresso  $S_{in}$ . Ricordiamo innanzitutto che tra i segnali precedentemente indicati sussistono le seguenti relazioni. Il segnale di uscita  $S_o$  è pari al segnale  $\epsilon$  moltiplicato per il guadagno del blocco di andata **A**. Analogamente, il segnale di ritorno  $S_f$  è pari al segnale di uscita  $S_o$  moltiplicato per il guadagno del blocco di ritorno **H**. Infine, per effetto del nodo sottrattore, il segnale  $\epsilon$  è pari alla differenza tra il segnale di ingresso  $S_{in}$  ed il segnale di ritorno  $S_f$ . Utilizzando la terza relazione nella prima, si ottiene che il segnale  $S_o$  è pari ad **A** per  $S_{in}$  meno **A** per  $S_f$ . Inserendo al posto di  $S_f$  la seconda relazione indicata, si ottiene che  $S_o$  è pari ad **A** per  $S_{in}$  meno **H** per  $S_o$ . Questa relazione contiene ora solo termini in  $S_{in}$  e  $S_o$ , che opportunamente raggruppati portano ad una relazione in cui  $S_o$ , che moltiplica  $(1 + A \cdot H)$ , è uguale a  $S_{in}$  per **A**.

Da questa relazione è possibile determinare il guadagno del sistema  $G = \frac{S_o}{S_{in}} = \frac{A}{1 + A \cdot H}$ . Questa relazione esprime in generale il

guadagno di un sistema reazionato negativamente.

Vediamo ora di analizzare ulteriormente la relazione appena ricavata. Se in questa moltiplichiamo sia il numeratore che il denominatore per il guadagno **H**, otteniamo un'espressione del tutto equivalente

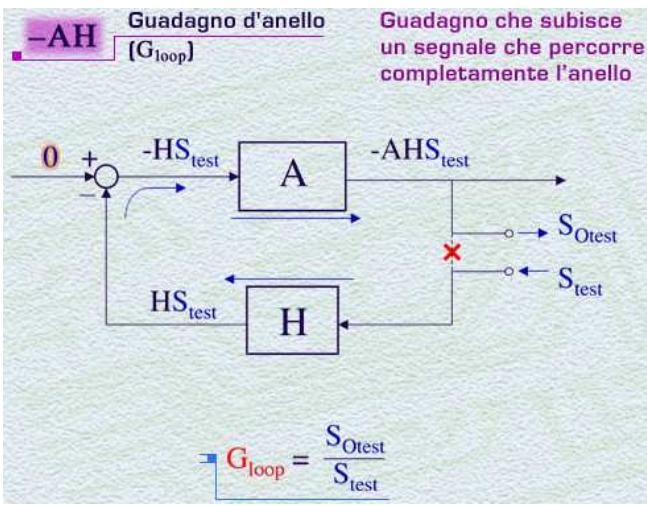
in cui il guadagno è esprimibile come  $G = \frac{S_o}{S_{in}} = \frac{A}{1 + A \cdot H} \frac{H}{H} = \frac{1}{H} \frac{A \cdot H}{1 + A \cdot H}$ . Se ora supponiamo si fare



tendere all'infinito il valore del termine  $A \cdot H$ , otteniamo un'importante proprietà: il guadagno tende ad essere uguale al solo termine  $\frac{1}{H}$ . In questa maniera, quindi, il guadagno tende ad essere indipendente dal valore di **A**, spesso non stabile o differente da esemplare a esemplare di dispositivo utilizzato, e dipende solo dal guadagno del blocco di ritorno **H**, che solitamente viene realizzato con reti, composte da componenti

passivi, maggiormente stabili e riproducibili. Dalle relazioni precedentemente introdotte, è possibile inoltre determinare la relazione che lega il segnale di ritorno  $S_f$  con il segnale di andata  $S_{in}$ .  $S_f$  risulta pari a  $S_{in}$  per il rapporto tra  $A \cdot H$  e  $(1 + A \cdot H)$ . Facendo tendere  $A \cdot H$  all'infinito, il segnale  $S_f$  risulta sempre più uguale a  $S_{in}$ , cioè la reazione tende a sottrarre ad  $S_{in}$  un segnale di valore sempre più simile. Come conseguenza di questo, il segnale  $\epsilon$ , pari alla differenza di questi ultimi due, e di cui si riporta l'espressione calcolata rispetto ad  $S_{in}$ , tenderà a zero all'aumentare di  $A \cdot H$ . Per questa ragione, il segnale  $\epsilon$  prende comunemente il nome di segnale errore.

Abbiamo visto che le conclusioni a cui siamo appena giunti sono legate all'aumento del valore del termine  $A \cdot H$ . Questo termine, o più precisamente il suo opposto -  $A \cdot H$ , possiede anche un preciso significato fisico. Esso infatti rappresenta il **guadagno d'anello** del circuito controreazionato e cioè il guadagno che subisce un segnale applicato in un punto dell'anello, nel percorrere completamente l'anello fino allo stesso punto di partenza. Supponiamo infatti di rompere l'anello in un punto qualsiasi, ad esempio all'ingresso del blocco **H**, e di applicare in quel punto un segnale di prova  $S_{test}$ . Per valutare il guadagno d'anello occorrerà seguire la propagazione di questo segnale lungo tutto l'anello e calcolarne il valore nello stesso punto di partenza, cioè il segnale  $S_{o,test}$ . Il guadagno d'anello risulterà pari ad  $S_{o,test}$  diviso per  $S_{test}$ . Occorre specificare che per effettuare il calcolo del guadagno d'anello occorre seguire la propagazione del segnale di prova nel senso di percorrenza dell'anello,



precedentemente indicato. Inoltre, al fine di valutare la sola propagazione del segnale di prova lungo l'anello, il segnale di ingresso deve essere spento, cioè  $S_{in}$  deve essere posto uguale a zero.

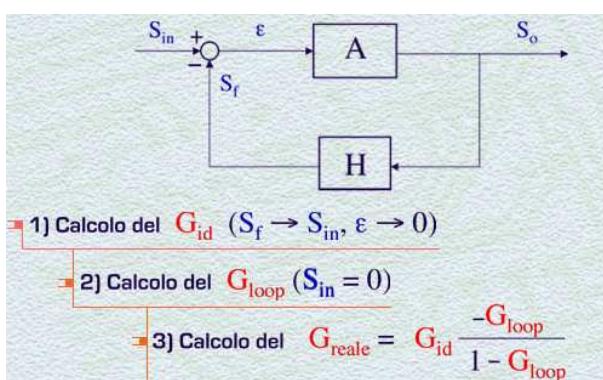
Possiamo ora procedere osservando che il segnale all'uscita del blocco di reazione vale  $\mathbf{H}$  per  $S_{test}$ .

Successivamente questo segnale viene sottratto a 0 nel nodo sottrattore e pertanto il segnale all'ingresso di **A** vale  $-\mathbf{H}$  per  $S_{test}$ . Infine il segnale all'uscita del blocco di guadagno è pari a  $-\mathbf{AH}$  per  $S_{test}$ . Quest'ultimo risulta quindi essere il valore  $S_{Otest}$  del segnale riportato allo stesso punto di partenza del circuito. Di conseguenza il guadagno d'anello, che noi chiameremo di seguito  $G_{loop}$ , risulta pari a  $-\mathbf{AH}$ . Avendo supposto nella nostra trattazione della reazione negativa di avere i blocchi **A** e **H** con guadagno positivo, il guadagno d'anello risulterà pertanto negativo.

La relazione precedentemente indicata per il guadagno **G** del sistema può ora essere espressa in maniera equivalente evidenziando la presenza del guadagno d'anello. Basta sostituire il termine  $\mathbf{AH}$  con  $-G_{loop}$ . Di conseguenza si ottiene che il guadagno **G** è pari a 1 su **H** per il rapporto tra  $-G_{loop}$  ed  $(1 - G_{loop})$ . In maniera analoga a quando detto prima, se facciamo tendere il valore di  $-G_{loop}$  all'infinito otteniamo che il valore del guadagno tenderà ad essere pari a 1 diviso **H**. Come già anticipato in precedenza, questa importante proprietà consente in pratica di progettare sistemi con guadagno stabile, dipendente dal solo dal blocco di reazione, e indipendente dal blocco di guadagno **A**, generalmente non stabile o riproducibile. D'altra parte un valore elevato del blocco di guadagno, anche se non riproducibile o

stabile, è fondamentale per ottenere un guadagno d'anello in modulo elevato, condizione necessaria per ottenere la dipendenza di **G** solo da **H**. A conclusione di quanto detto, indicheremo il termine  $1/\mathbf{H}$  come **guadagno ideale**  $G_{id}$ , del sistema e di conseguenza l'espressione generale del guadagno, che chiameremo per contro guadagno 'reale' sarà esprimibile come  $G_{id}$  per il rapporto tra  $-G_{loop}$  ed  $(1 - G_{loop})$ .

Alla luce di quanto detto, il calcolo del guadagno di un circuito controllato può essere effettuato utilizzando le relazioni tipiche, ricavate per questa



classe di sistemi. Questo calcolo, quando applicato nell'ambito dei circuiti elettronici, può risultare più semplice del calcolo risolutivo della rete elettrica che utilizza le sole leggi di Kirchhoff ed inoltre può aiutare il progettista nella definizione delle caratteristiche fondamentali del circuito. Il calcolo del guadagno di un circuito controllato può quindi essere effettuato nel seguente modo. Si determina innanzitutto il guadagno ideale  $G_{id}$  del circuito. Questo può essere effettuato molto semplicemente per **ispezione del circuito** individuando il percorso dell'anello, ipotizzando che il guadagno d'anello sia infinito, e supponendo quindi che il segnale di ritorno  $S_f$  sia uguale a quello di ingresso. Successivamente, si calcola il guadagno d'anello. A tale scopo il generatore di segnale di ingresso deve essere spento. Infine, si determina il guadagno reale utilizzando la relazione, precedentemente

introdotta, in cui il guadagno reale è uguale a quello ideale moltiplicato per il rapporto tra  $-G_{loop}$  ed  $(1 - G_{loop})$ .

Il procedimento da utilizzarsi per il calcolo del guadagno d'anello segue di fatto la definizione di guadagno d'anello che è stata data in precedenza. Facciamo alcune precisazioni su come procedere nel calcolo. Innanzitutto occorre procedere con lo **spegnimento di tutti i generatori di segnali forzanti**. Ciò equivale a porre in cortocircuito i generatori di tensione e aprire i generatori di corrente. Poi si procede con la scelta del punto dell'anello in cui effettuare il taglio e applicare il generatore di test. È importante a tale riguardo fare una precisazione. **L'operazione di taglio non deve portare ad una modifica dell'impedenza vista dallo stadio che nell'anello si trova immediatamente a monte del taglio.** Infatti, una modifica di tale impedenza potrebbe cambiare la funzione di trasferimento di questo stadio rispetto al suo comportamento con l'anello chiuso.

Pertanto, occorre calcolare il valore dell'impedenza  $Z_0$  vista a valle del taglio e aggiungere un'impedenza di pari valore subito prima del taglio. Infine, si procede all'applicazione del segnale di sonda  $S_{test}$  e come già visto in precedenza si calcola come tale segnale si propaga nell'anello, fino a determinare il segnale  $S_{0test}$  in corrispondenza dello stesso punto in cui si è operato il taglio.

Il guadagno d'anello è pari al rapporto tra  $S_{0test}$  e  $S_{test}$ .

Introdotti i concetti fondamentali che caratterizzano i circuiti Controreazionati, vediamo alcune importanti proprietà di questa classe di circuiti.

Un'importante proprietà consiste nell'**elevata insensibilità del guadagno alle variazioni del guadagno A** del blocco di andata. Data la relazione che esprime il guadagno del circuito reazionato  $G$ , detto anche

**Elevata insensibilità alle variazioni di A.**

$$G = \frac{A}{1+AH} \quad dG = \frac{dA}{(1+AH)^2}$$

$$\frac{\text{variazione relativa del guadagno ad anello chiuso}}{\text{variazione relativa del guadagno ad anello aperto}} = \frac{1}{1+AH} \frac{dA}{A} \quad \text{0 per } AH \rightarrow \infty$$

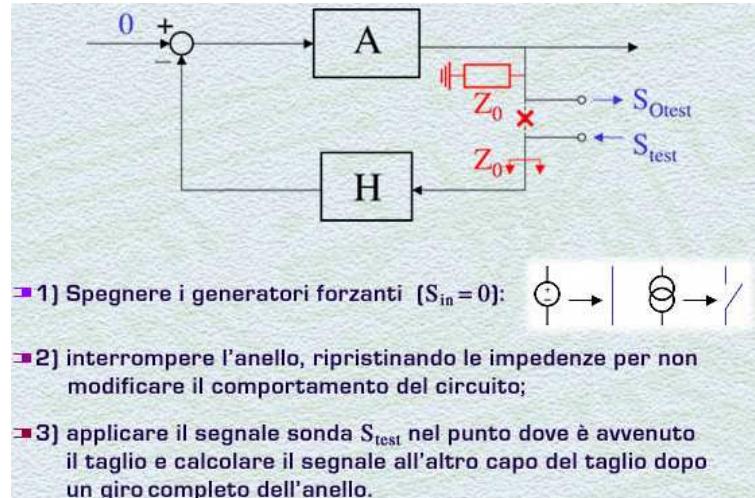
Una variazione del guadagno del blocco di andata, comporta una variazione del guadagno del circuito reazionato percentualmente di minore valore.

Esempio  $G_{loop} = -AH = -100$

$$\frac{dA}{A} = 10\% \quad \frac{dG}{G} = \frac{1}{1+100} \cdot 0.1 \approx 0.001 = 0.1\%$$

a quella del secondo, cioè  $dA$  su  $A$ , divisa per il fattore  $(1 + AH)$ , o se vogliamo,  $(1 - G_{loop})$ . Questa relazione mostra che una variazione del guadagno del blocco di andata, dovuta ad esempio a fattori quali la variazione della temperatura o la sostituzione del blocco di andata con un altro blocco di guadagno diverso, comporta una variazione del guadagno del circuito reazionato percentualmente di minore valore. Se al limite il termine  $AH$  tende all'infinito, una variazione qualsiasi di  $A$  non comporta alcuna variazione del guadagno ad anello chiuso. Questa proprietà fa sì che i circuiti controreazionati abbiano caratteristiche stabili e riproducibili rispetto alle variazioni possibili del blocco di andata.

Vediamo a riguardo un esempio numerico. Supponiamo di avere un sistema Controreazionato in cui il guadagno d'anello  $-AH$  sia pari a  $-100$ . Se il guadagno di andata  $dA$  su  $A$  varia del 10%, corrispondentemente il guadagno ad anello chiuso  $dG$  su  $G$  varierà, per la relazione appena vista, solo di un fattore pari allo 0,1 %.



Una seconda proprietà dei circuiti contoreazionati consiste nella **riduzione della distorsione** introdotta dall'impiego nel blocco di guadagno di amplificatori non lineari. Mostriamo questa proprietà con un esempio. Nella figura è rappresentata la caratteristica ingresso-uscita di un amplificatore utilizzato come blocco di andata di un circuito contoreazionato. Nell'esempio si è schematizzata la non-linearità dell'amplificatore con la presenza di due diverse pendenze della caratteristica, una più ripida corrispondente ad un valore del guadagno  $A_1$  pari a 1000, e una meno ripida corrispondente ad un valore del guadagno  $A_2$  pari a 100. Il rapporto tra i due diversi guadagni è pari a 10. Questa diversità di guadagno è causa di distorsione per segnali applicati all'ingresso all'amplificatore, in quanto a diversi valori di ampiezza del segnale

può corrispondere un diverso valore di guadagno dell'amplificatore. Supponiamo ora di inserire l'amplificatore in un sistema contoreazionato con un blocco di reazione di guadagno  $H$  pari a 0.1. Dalla relazione del guadagno del circuito reazionato si può determinare che in corrispondenza del guadagno ad anello aperto  $A_1$ , il guadagno ad anello chiuso  $G_1$  è pari a 9.9, mentre in corrispondenza del guadagno  $A_2$ , il guadagno ad anello chiuso  $G_2$  risulta pari a 9.1. Come è possibile anche osservare dal grafico, dove sono rappresentati gli andamenti del guadagno del circuito contoreazionato, l'effetto della reazione ha prodotto una maggiore linearizzazione nell'andamento del guadagno rispetto alla condizione ad anello aperto. Il rapporto tra i guadagni  $G_1$  e  $G_2$  risulta di fatti pari a 1.1, quindi con una minore diversità relativa di questi guadagni rispetto ad  $A_1$   $A_2$ , che differivano di un ordine di grandezza. La maggiore linearità del circuito contoreazionato consente di ottenere una minore distorsione del segnale applicato, rispetto alla condizione iniziale. Questa importante proprietà viene pagata naturalmente con una corrispondente riduzione del guadagno totale del sistema.

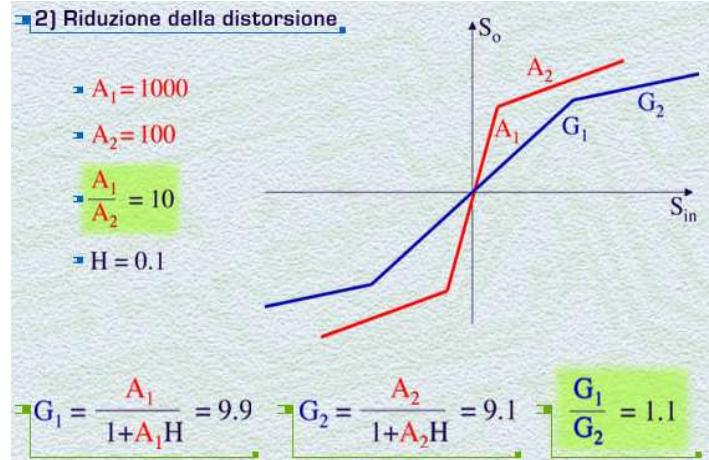
Una terza proprietà dei sistemi retroazionati consiste nella possibilità di poter **estendere notevolmente la banda in frequenza**. Esaminiamo questo aspetto considerando un sistema contoreazionato in cui il blocco di guadagno sia caratterizzato da una funzione di trasferimento a

**3) Estensione della banda**

<b>blocco di guadagno a singolo polo</b>	$A(s) = \frac{A_0}{1+s\tau} = \frac{A_0}{1+\frac{s}{\omega_0}}$	
<b>blocco di reazione indipendente dalla frequenza</b>	$H(s) = H$	
$G(s) = \frac{A(s)}{1+A(s)H} = \frac{A_0}{1+A_0H} \frac{1}{1+\frac{s}{\omega_0}(1+A_0H)}$		
$G(s) = \frac{G_0}{1+\frac{s}{\omega_1}}$	$G_0 = \frac{A_0}{1+A_0H}$	$\omega_1 = \omega_0(1+A_0H)$
<b>Riduzione del guadagno rispetto a <math>A_0</math> di un fattore <math>\sim -G_{loop}(0)</math>.</b>		
<b>Estensione della banda rispetto a <math>\omega_0</math> di un fattore <math>\sim -G_{loop}(0)</math>.</b>		

( $1+A_0 \cdot H$ ), per un termine dipendente dalla frequenza che è pari a 1 diviso  $1 + s/(\omega_0 \cdot (1 + A_0 \cdot H))$ . In conclusione, quindi, anche il guadagno ad anello chiuso risulta avere un polo e si può esprimere come  $G_0$  diviso  $(1 + s/\omega_1)$ , dove  $G_0$  è uguale ad  $A_0$  diviso  $(1+A_0 \cdot H)$  e la frequenza del polo, o meglio la sua pulsazione  $\omega_1$ , risulta pari ad  $\omega_0 \cdot (1+A_0 \cdot H)$ .

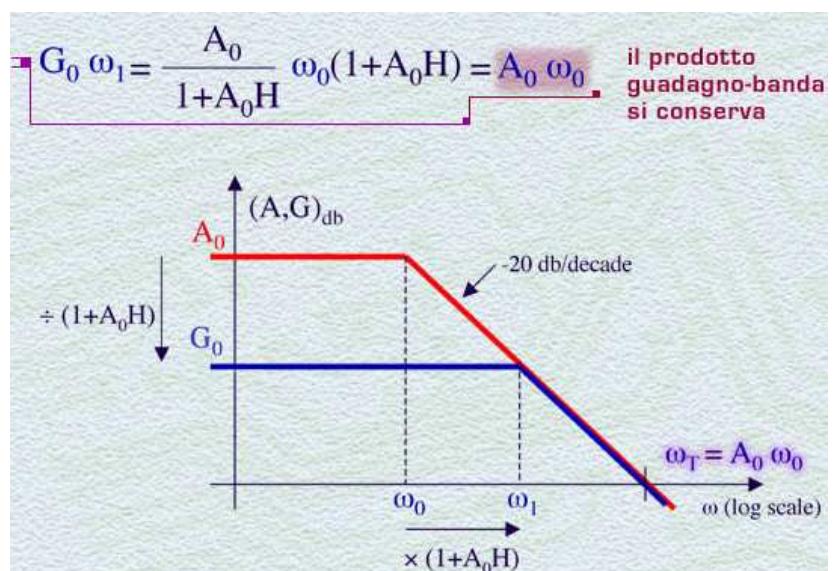
Si può osservare da questa espressione che l'effetto della contoreazione è quello di ridurre il guadagno a bassa frequenza, rispetto a quello del blocco di andata  $A_0$ , di un fattore pari a circa l'opposto del guadagno d'anello calcolato a bassa frequenza, e cioè  $-G_{loop}(0)$ . D'altra parte, la banda del sistema  $\omega_1$  risulta pari alla banda del blocco di andata moltiplicata ancora per un fattore pari a meno il guadagno d'anello a bassa frequenza.



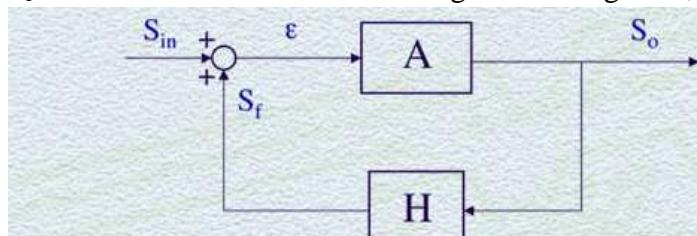
L'effetto benefico della reazione consiste quindi nell'estendere la banda del sistema, rispetto a quella del guadagno ad anello aperto, di un **fattore in modulo pari al guadagno d'anello**. Ciò avviene, naturalmente, a discapito della riduzione del guadagno di uno stesso fattore.

Se calcoliamo il valore del prodotto guadagno-banda del sistema controreazionato, cioè  $G_0 \cdot \omega_1$ , questo risulta pari al termine  $A_0$  diviso  $(1+A_0 \cdot H)$  e moltiplicato per  $\omega_0 \cdot (1+A_0 \cdot H)$ . Semplificando, questo risulta pari a:  $A_0 \cdot \omega_0$ , cioè al prodotto guadagno-banda del blocco di andata. Con ciò si può dire, quindi, che il prodotto guadagno-banda si conserva.

La sintesi di quanto detto finora è illustrata nella figura in cui sono indicati i diagrammi di Bode del guadagno ad anello aperto **A** e di quello ad anello chiuso **G**. La figura mostra la riduzione del guadagno a bassa frequenza  $G_0$  rispetto ad  $A_0$  di un fattore pari a  $(1+A_0 \cdot H)$ , ed il corrispondente aumento della banda ad anello chiuso dello stesso fattore. Il prodotto guadagno-banda resta inalterato.



Concludiamo questa introduzione ai concetti base dei sistemi reazionati, accennando alla reazione positiva. In un sistema reazionato positivamente, come quello indicato in figura, il segnale di ritorno  $S_f$  anziché essere sottratto al segnale di ingresso, come avviene nella reazione negativa, viene



■ La reazione rinforza il segnale all'ingresso del blocco d'andata

■ Inoltre  $G_{loop} = AH > 0$

$$\epsilon = S_{in} + S_f$$

■ Se  $G_{loop} > 1$  il sistema diverge ( $S_o \rightarrow \pm \infty$ )

■ Se  $0 < G_{loop} < 1$  il sistema converge  $G = \frac{S_o}{S_{in}} = \frac{A}{1-AH} > A$

sommato ad esso. Ciò contribuisce a rinforzare, anziché attenuare, il segnale in ingresso al blocco di guadagno, che risulta pari ad  $S_{in} + S_f$ . Inoltre, si può facilmente osservare che il guadagno d'anello è ora positivo ed è uguale ad  $A$  per  $H$ . Da un'analisi di questo tipo di sistema si può ricavare che se il guadagno d'anello è superiore a 1, la continua circolazione di un segnale lungo l'anello porta facilmente il sistema a divergere, cioè l'uscita tende a

portarsi a valori infiniti. Naturalmente, in circuiti reali, ciò corrisponde in verità ad una saturazione dell'uscita a valori prossimi alle tensioni di alimentazione. Nel caso in cui, invece, il guadagno d'anello sia inferiore ad uno, la circolazione del segnale lungo l'anello non porta alla divergenza ed il guadagno **G** ad anello chiuso risulta essere pari ad **A** diviso  $(1 - AH)$ . Considerando che  $AH$  è positivo e minore di uno, ciò significa che il guadagno ad anello chiuso **G** è maggiore del guadagno ad anello aperto **A**.

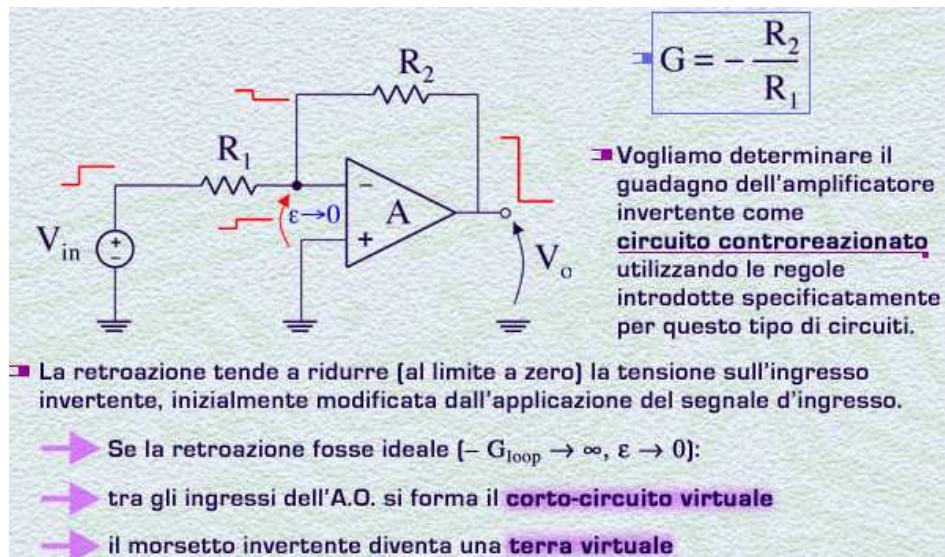
## Lezione n. 27 : L'amplificatore invertente e l'amplificatore non invertente (come circuiti contoreazionati)

### L'amplificatore invertente (analizzato come circuito contoreazionato)

Consideriamo di nuovo un circuito già introdotto in precedenza come esempio di utilizzo dell'amplificatore operazionale. Si tratta dell'*amplificatore invertente*. Questo circuito, di cui viene mostrato nuovamente lo schema, opera un'amplificazione del segnale di ingresso con un guadagno pari a  $-R_2/R_1$ , nell'ipotesi di guadagno infinito dell'amplificatore operazionale. A questa conclusione eravamo giunti dopo aver studiato il circuito utilizzando semplicemente le regole base del calcolo delle reti elettriche.

L'amplificatore invertente appartiene però anche alla classe dei circuiti contoreazionati, della quale abbiamo visto le proprietà generali nella precedente lezione. Consideriamo quindi nuovamente questo circuito sotto questa nuova ottica e vediamo come sia possibile determinare nuovamente il suo guadagno, ma questa volta utilizzando le regole introdotte specificatamente per i circuiti contoreazionati. Ciò potrà sembrare non necessario per questo circuito che risulta estremamente semplice e di cui conosciamo già le caratteristiche. Il procedimento che vedremo in questo esempio potrà però risultare utile qualora vi sia l'interesse a studiare o progettare circuiti contoreazionati più complessi in cui il metodo di calcolo basato sulla teoria delle reti elettriche risulti più difficile da applicare. Inoltre, il procedimento che verrà illustrato consente generalmente al progettista di analizzare il circuito pervenendo alla determinazione delle sue caratteristiche più significative in maniera più intuitiva ed efficace da un punto di vista progettuale. Iniziamo innanzitutto con l'osservare che l'amplificatore invertente risulta essere a tutti gli effetti un circuito contoreazionato.

Supponiamo di applicare un segnale in ingresso, ad esempio un gradino di tensione positivo. Si supponga inizialmente che la tensione di uscita dell'A.O. si trovi a zero. L'effetto del segnale in



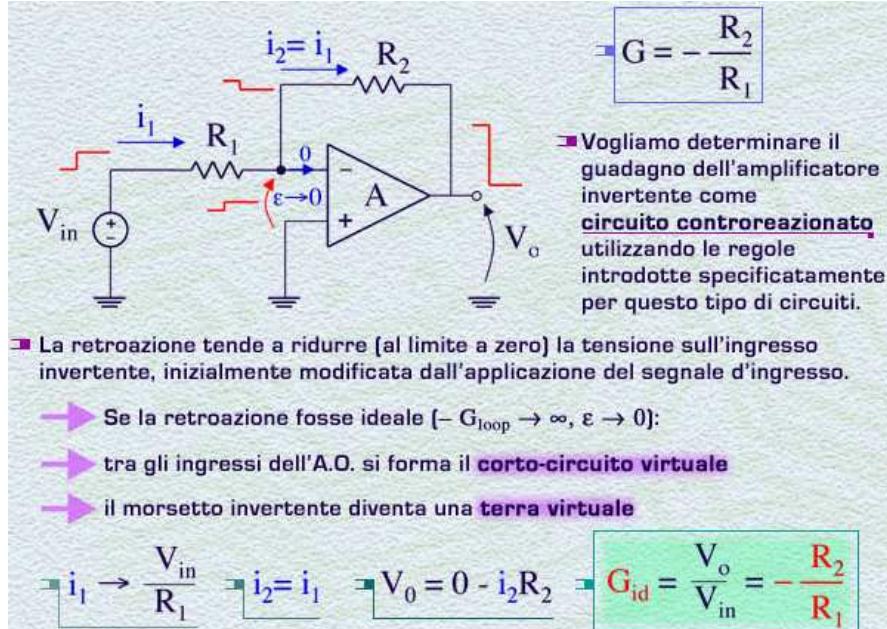
ingresso è quello di produrre, tramite la partizione tra le resistenze  $R_1$  e  $R_2$ , un segnale sempre positivo all'ingresso dell'A.O., cioè tra il morsetto invertente e quello non invertente che è tenuto a massa. A seguito dell'applicazione di questo segnale, sull'uscita dell'A.O. si manifesterà un segnale di tensione negativo rispetto a massa, pari al segnale sul morsetto invertente moltiplicato per il guadagno  $A$ . Osserviamo ora la presenza della resistenza  $R_2$ , il cui ruolo è quello di riportare il segnale dall'uscita dell'amplificatore al suo ingresso. Essa pertanto ci consente di identificare un cammino di ritorno per il segnale, che avevamo visto essere l'elemento caratteristico dei sistemi reazionati negativamente. Il segnale di uscita viene riportato al morsetto invertente, come gradino negativo di tensione, opportunamente ridotto per l'effetto della partizione tra  $R_2$  e  $R_1$ . Di conseguenza, l'effetto prodotto dalla presenza in reazione di  $R_2$  è quello di produrre sul morsetto invertente un gradino di tensione negativo che va quindi a opporsi al segnale positivo indotto sullo stesso morsetto dall'applicazione del segnale di ingresso. In conclusione, il meccanismo appena descritto tende a ridurre l'effetto perturbativo provocato sull'ingresso del blocco di guadagno dal segnale applicato. E proprio sulla base di questa considerazione che possiamo quindi identificare il circuito appena descritto come un circuito **reazionato negativamente**.

Come visto, l'effetto della reazione è quello di ridurre la variazione della tensione in ingresso all'A.O. indotta dall'applicazione del segnale. Ciò ci permette di associare a questa tensione il significato di

segnale errore  $\epsilon$  del sistema reazionato. Migliore è la reazione, cioè maggiore è in modulo il guadagno d'anello del circuito, minore risulterà il valore di questo segnale errore.

Seguendo l'esempio dei segnali indicati in figura, ciò corrisponde al fatto che il segnale negativo riportato tramite  $R_2$  tende a essere uguale a quello positivo indotto, tramite  $R_1$ , dall'applicazione del segnale di ingresso. Il tendere a zero del segnale errore equivale inoltre a dire che tra i due ingressi dell'A.O. tende a formarsi un **corto-circuito virtuale**. Il concetto di corto-circuito virtuale, che avevamo già introdotto quando avevamo trattato per la prima volta l'amplificatore invertente, risulta adesso essere un concetto chiaramente legato alle proprietà di questo circuito come circuito contoreazionato. Analogamente, tenuto conto del fatto che il morsetto non invertente si trova fisso a massa, ciò implica che il morsetto invertente tende a divenire una *terra virtuale*, concetto anch'esso introdotto in precedenza e che assume ora un significato legato alle proprietà della contoreazione.

Dopo questa ampia introduzione volta a riconoscere la presenza della contoreazione nel circuito, risulta ora facile determinare il **guadagno ideale** dello stesso, cioè il guadagno calcolato nell'ipotesi di



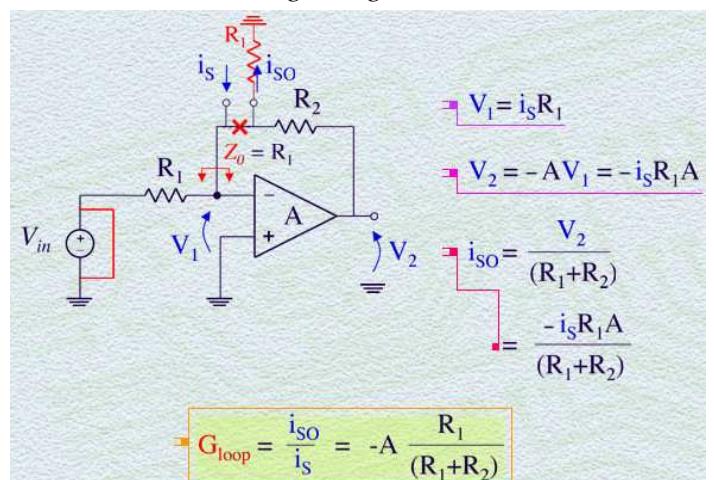
guadagno d'anello infinito. Se il guadagno d'anello è infinito e corrispondentemente il segnale errore tende a zero, la corrente  $i_1$  che scorre nella resistenza  $R_1$  si può determinare come  $V_{in}$  diviso  $R_1$ . Ora, ricordandoci che tale corrente non può fluire nel morsetto invertente dell'A.O., essa risulta quindi scorrere completamente nella resistenza  $R_2$ , cioè la corrente  $i_2$  risulta pari ad  $i_1$ . Infine, tenuto conto che il morsetto invertente è una terra virtuale, è possibile determinare la tensione di uscita grazie alla sola caduta di

tensione su  $R_2$ . Cioè  $V_o$  è pari a  $-i_2 \cdot R_2$ . Il guadagno ideale dell'amplificatore invertente risulta quindi essere pari a  $-R_2 / R_1$ , come già ricavato in precedenza.

Nell'ipotesi ora che il guadagno d'anello sia finito, procediamo al calcolo del **guadagno reale** del circuito reazionato. Per fare questo dobbiamo innanzitutto calcolare il **guadagno d'anello**.

Seguendo il procedimento generale dato nella precedente lezione, spegniamo il generatore di tensione forzante, ponendolo in corto-circuito, e procediamo al taglio dell'anello. Come mostrato in figura, si è scelto di tagliare l'anello appena a valle della resistenza  $R_2$ . Per non modificare la risposta dello stadio a monte di  $R_2$ , occorrerà ripristinare l'impedenza  $Z_0$  vista a valle del taglio.

Questa impedenza è pari a  $R_1$ , in quanto l'impedenza di ingresso dell'A.O. è supposta infinita. Applichiamo ora una corrente sonda  $i_s$  come indicato in figura e determiniamo la corrente  $i_{so}$  uscente da  $R_2$  dopo che il segnale sonda ha percorso un giro completo dell'anello. La tensione  $V_1$  vale  $i_s \cdot R_1$ . La tensione all'uscita dell'A.O.  $V_2$  risulta pari a  $-A \cdot V_1$  e cioè  $-i_s \cdot R_1 \cdot A$ . La corrente  $i_{so}$  può essere calcolata come  $V_2$  diviso  $(R_1 + R_2)$  e risulta quindi pari a  $-i_s \cdot R_1 \cdot A / (R_1 + R_2)$ . Il guadagno d'anello, quindi, pari al rapporto tra  $i_{so}$  e  $i_s$ , risulta uguale a  $-A \cdot R_1 / (R_1 + R_2)$ .



Determinato il guadagno d'anello, possiamo ora calcolare il *guadagno reale* dell'amplificatore invertente. Esso risulta uguale al guadagno ideale moltiplicato per il rapporto  $-G_{\text{loop}} / (1 - G_{\text{loop}})$ .

$$\begin{aligned} G_{\text{reale}} &= G_{\text{id}} \frac{-G_{\text{loop}}}{1 - G_{\text{loop}}} = -\frac{R_2}{R_1} \frac{-A \frac{R_1}{(R_1+R_2)}}{1 + A \frac{R_1}{(R_1+R_2)}} \\ &= -\frac{R_2}{R_1} \frac{AR_1}{R_1+R_2+AR_1} = -\frac{R_2}{R_1} \frac{1}{1 + \frac{R_1+R_2}{R_1} \frac{1}{A}} \end{aligned}$$

**■ La relazione è identica a quella ricavata applicando il solo calcolo delle reti elettriche.**

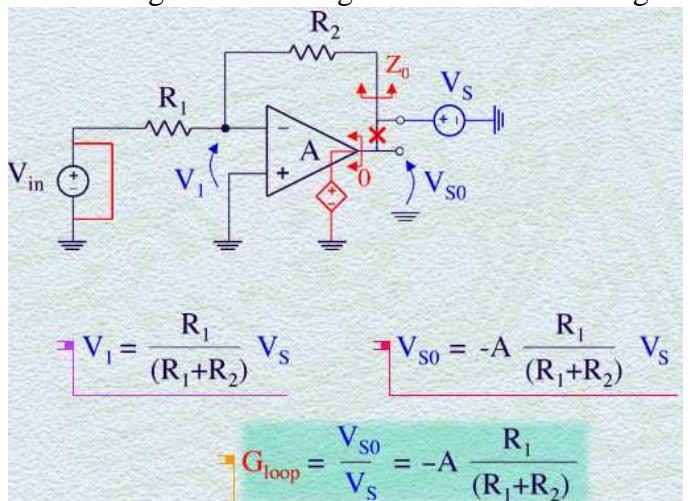
**■ L'aver determinato questa espressione a partire dalle espressioni ricavate per il guadagno ideale e per il guadagno d'anello può portare vantaggi per il progetto del circuito.**

Introducendo le espressioni precedentemente calcolate per il guadagno ideale ed il guadagno d'anello e procedendo alle opportune semplificazioni algebriche, il guadagno reale risulta pari a  $-(R_2/R_1) \cdot (A \cdot R_1) / (R_1 + R_2 + A \cdot R_1)$ . Raccogliendo i termini in maniera differente, questo risulta pari a:  $-(R_2/R_1) \cdot 1 / (1 + [(R_1+R_2)/R_1] \cdot 1/A)$ . Questa relazione è identica a quella ricavata applicando il solo calcolo delle reti elettriche. Vi siamo ora giunti applicando le proprietà dei circuiti controreazionati. In generale,

l'aver determinato questa espressione a partire dalle espressioni ricavate per il guadagno ideale e per il guadagno d'anello può avere anche dei vantaggi dal punto di vista progettuale del circuito. Se ad esempio, il progettista desidera ottenere un guadagno reale prossimo a quello ideale entro un certo margine specificato, può determinare quanto deve essere il valore minimo, in modulo, del guadagno d'anello da ottenere per soddisfare a tale scopo. Per determinare questo valore minimo, può ricorrere all'espressione ricavata per il guadagno d'anello e scegliere opportunamente i componenti da utilizzare.

Prima di abbandonare questo esempio, vediamo come sia possibile effettuare un calcolo alternativo del guadagno d'anello, tagliando l'anello in un punto differente dal caso precedente. Supponiamo ora di tagliare l'anello direttamente all'uscita dell'A.O. Seguendo la regola vista sul reintegro dell'impedenza  $Z_o$  vista a valle del taglio, occorrerebbe ora procedere alla determinazione del valore di tale impedenza. Si osservi, però, che in questo caso la determinazione di  $Z_o$  risulta inutile in quanto qualsiasi sia il suo valore, il guadagno  $A$  dell'A.O., cioè il suo trasferimento dall'ingresso all'uscita, risulta indipendente da  $Z_o$ . Ciò grazie al fatto che l'A.O. si presenta alla sua uscita come un generatore ideale di tensione, e quindi con una impedenza di uscita nulla. Questa proprietà rende questo particolare punto del circuito, come di altri circuiti utilizzanti l'A.O., molto comodo per il calcolo del guadagno d'anello, in quanto non è necessario calcolare l'impedenza vista a valle del taglio. Si è preferito introdurre questa opportunità successivamente all'esempio più generale mostrato in precedenza, in quanto anche se risulta di fatto molto pratico sfruttare il taglio dell'anello all'uscita dell'A.O. ideale senza procedere al ripristino dell'impedenza, non bisogna mai dimenticarsi che più in generale ciò va fatto. Ad esempio, se si considera un A.O. reale dotato di una impedenza di uscita non nulla, in questo caso anche tagliando l'anello all'uscita dell'amplificatore occorre ripristinare l'impedenza vista a valle del taglio per non commettere, a rigore, un errore nella determinazione del guadagno d'anello. Vedremo specificatamente questo esempio in una lezione successiva.

Procedendo nel calcolo, si applichi un generatore di tensione di sonda  $V_s$ . La tensione corrispondente al morsetto invertente  $V_1$  risulta pari a  $V_s$  per il rapporto  $R_1/(R_1+R_2)$ . La tensione di uscita  $V_{s0}$  risulta pertanto uguale a  $V_s$  che moltiplica  $-A \cdot R_1 / (R_1+R_2)$  e quindi il guadagno d'anello risulta uguale a  $-A \cdot R_1 / (R_1+R_2)$ , come calcolato in precedenza.



Al termine dell'analisi dell'amplificatore invertente come circuito contoreazionato, occorre fare un'ultima precisazione di carattere generale. Quando si è introdotta la teoria della contoreazione si è indicato con  $A$  il guadagno del blocco di andata del sistema, occorre però fare attenzione a **NON identificare necessariamente tale guadagno con il solo guadagno dell'A.O.** per il quale si utilizza comunemente lo stesso simbolo  $A$ . Per illustrare meglio ciò si può determinare, ad esempio, il guadagno in andata dell'amplificatore invertente indicato ora, per maggior chiarezza,  $G_A$ . Si è visto che il guadagno d'anello  $G_{loop}$  è pari al prodotto del guadagno d'andata per il guadagno del blocco di reazione cambiato di segno, cioè:  $G_{loop} = -G_A \cdot H$ , ne segue che:  $G_A = -G_{loop}/H$  ossia  $G_A = -G_{loop} \cdot G_{id}$ , con  $G_{id} = 1/H$ , (cfr. lez. precedente).

Sostituendo le espressioni ricavate per il guadagno  $G_{loop}$  e  $G_{id}$ , si ottiene che il guadagno d'andata  $G_A$  risulta pari a:  $G_A = -A \cdot R_2 / (R_1 + R_2)$  in cui  $A$  è *il solo guadagno dell'amplificatore operazionale*. Tale calcolo può essere usato sempre, in generale, per determinare il guadagno di andata di un circuito contoreazionato.

**Osservazione:**  
negli amplificatori contoreazionati basati sull'A.O., il guadagno di andata non coincide necessariamente con il guadagno  $A$  dell'A.O.

■ Ad esempio, nell'amplificatore invertente:

$$G_{loop} = -G_A H$$

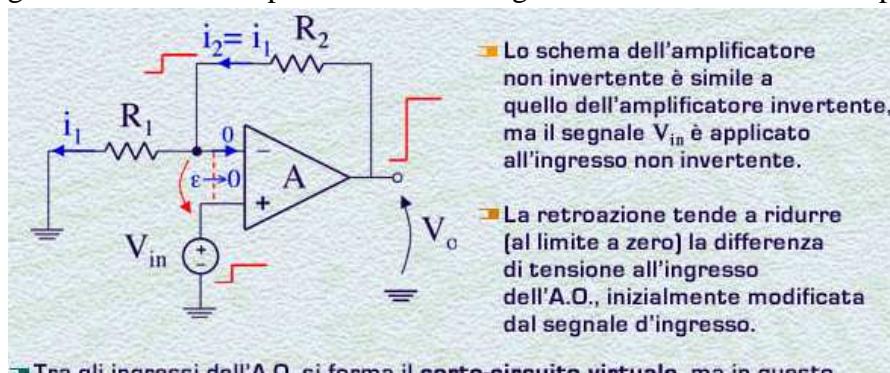
$$G_A = -G_{loop}/H = -G_{loop} G_{id}$$

$$\Rightarrow G_A = -A \frac{R_2}{R_1 + R_2}$$

## L'amplificatore non invertente

Una volta visto il funzionamento dell'amplificatore invertente come circuito contoreazionato, introduciamo ora un altro tipo di amplificatore contoreazionato: *l'amplificatore non invertente*. Lo schema circuitale di questo amplificatore, mostrato in figura, è simile a quello dell'amplificatore invertente. La differenza consiste nel fatto che il segnale  $V_{in}$  è applicato all'ingresso non invertente e non a quello invertente tramite la resistenza  $R_1$ .

Anche in questo caso, procediamo a determinare in maniera intuitiva il funzionamento di questo circuito sulla base della presenza di una reazione negativa. Supponiamo di applicare inizialmente un gradino di tensione positivo  $V_{in}$  sull'ingresso non invertente. Se supponiamo che all'inizio l'ingresso



■ Tra gli ingressi dell'A.O. si forma il **corto-circuito virtuale**, ma in questo caso non esiste una terra virtuale.

■ Se la retroazione fosse ideale ( $-G_{loop} \rightarrow \infty, \varepsilon \rightarrow 0$ )

$$V_- = V_{in} - \varepsilon \rightarrow V_{in}$$

$$i_1 = \frac{V_{in}}{R_1} \quad i_2 = i_1$$

$$V_0 = V_{in} + i_1 R_2$$

$$G_{id} = \frac{V_o}{V_{in}} = 1 + \frac{R_2}{R_1}$$

invertente si trovi a tensione nulla, l'applicazione del segnale sull'ingresso non invertente provocherà un segnale positivo sull'uscita dell'A.O. Tale segnale sarà elevato, in virtù dell'elevato guadagno  $A$  di tensione. Anche in questo caso, come per l'amplificatore invertente, la presenza della resistenza  $R_2$  ha l'effetto di riportare il segnale presente sull'uscita all'ingresso invertente dell'A.O., ridotto per un opportuno fattore di partizione di  $R_2$  con  $R_1$ . Ciò comporta che a seguito

dell'innalzamento della tensione sull'uscita dell'A.O., anche l'ingresso invertente subirà un innalzamento di tensione positivo.

A questo punto possiamo osservare che se l'applicazione iniziale del segnale aveva avuto come effetto quello di aumentare la differenza di tensione tra ingresso non invertente e ingresso invertente, l'azione della contoreazione, esercitata tramite la resistenza  $R_2$ , ha avuto come effetto quello di ridurre tale differenza, innalzando la tensione al morsetto invertente. Ancora una volta, la contoreazione tende a

ridurre l'effetto perturbativo sul circuito provocato dall'applicazione del segnale in ingresso. Risulta quindi intuitivo pensare, se si porta all'estremo questa considerazione, che nel caso di contoreazione ideale, cioè in presenza di un guadagno d'anello in modulo infinito, l'effetto della reazione sia quello di compensare al limite in maniera completa la perturbazione iniziale. Ciò equivale a dire che la reazione tende a mantenere uguale a zero la differenza di tensione tra i due morsetti dell'A.O. Tale differenza risulta quindi rappresentare ancora una volta il segnale errore nel nostro sistema reazionato negativamente.

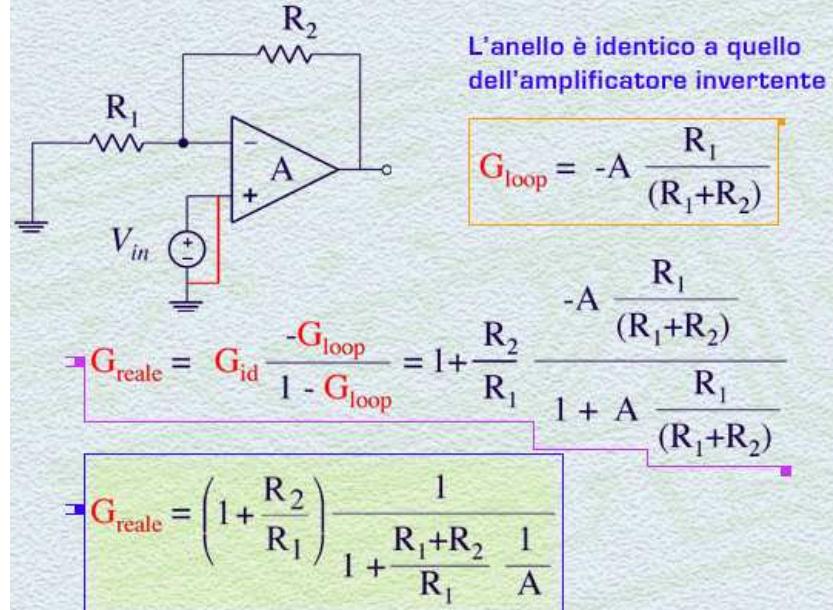
Si può osservare, quindi, come anche in questo circuito sia presente un *corto-circuito virtuale* tra i due ingressi invertente e non-invertente dell'A.O. In questo caso, però, **non esiste una terra virtuale** in quanto il morsetto non invertente è libero di variare la tensione liberamente ed il morsetto invertente lo segue in tensione traslando rigidamente in virtù della presenza del corto-circuito virtuale.

Alla luce dell'ispezione del circuito come circuito contoreazionario, procediamo ora alla determinazione del *guadagno ideale*. Se consideriamo la contoreazione ideale, cioè facciamo tendere il guadagno d'anello all'infinito, il segnale errore, cioè la differenza di tensione tra i due morsetti all'ingresso dell'A.O. tende ad essere pari a zero. Ciò ha come conseguenza che la tensione al morsetto invertente, tende a portarsi al valore della tensione applicata  $V_{in}$ . Di conseguenza, la corrente che scorre in  $R_1$  risulterà pari a  $V_{in}/R_1$ . Come già visto in precedenza, questa corrente scorrerà completamente nella resistenza  $R_2$  e non nel morsetto non invertente. E quindi la tensione all'uscita del circuito  $V_o$  risulterà pari a  $V_{in} + i_1 \cdot R_2$ . Ciò ci consente di determinare il guadagno ideale del circuito ( $V_o/V_{in}$ ) come pari a:  $G_{id} = 1 + (R_2/R_1)$ .

Si può quindi notare che l'amplificatore non invertente debba il suo nome al fatto che l'uscita non cambia segno rispetto al segnale applicato in ingresso. Il guadagno del circuito diviene pressoché identico, in modulo, a quello dell'amplificatore invertente e cioè pari al solo rapporto tra  $R_2$  e  $R_1$ , nel caso in cui tale rapporto risulti molto maggiore di 1.

Procediamo ora alla determinazione del guadagno d'anello e del guadagno reale del circuito. Come già effettuato in precedenza, per calcolare il guadagno d'anello occorre cortocircuitare il generatore di tensione in ingresso  $V_{in}$ . Effettuata questa operazione, osserviamo però che il circuito così ottenuto risulta essere assolutamente identico al circuito con cui avevamo determinato il guadagno d'anello nel caso dell'amplificatore invertente. In virtù di questo, i due amplificatori, invertente e non invertente, risultano quindi avere lo stesso guadagno d'anello che ricordiamo essere pari a  $-A$  per il rapporto tra  $R_1$  e  $R_1+R_2$ .

Possiamo quindi procedere al calcolo del guadagno reale che ricordiamo risulta pari al guadagno ideale per il rapporto tra  $-G_{loop}$  e  $(1 - G_{loop})$ . Inserendo le espressioni corrispondenti e apportando le opportune semplificazioni algebriche, il guadagno reale risulta così pari a  $(1 + R_2 / R_1)$ , il tutto diviso per  $(1 + (R_1+R_2)/(R_1 \cdot A))$ . In questa espressione il guadagno ideale del circuito viene diviso per un fattore che è pari a quello determinato per l'amplificatore invertente, e ciò dovuto al fatto, come già detto, che il guadagno d'anello è lo stesso per entrambi i circuiti.



## Lezione n. 28 : Riproducibilità del guadagno, risposta in frequenza, risposta su segnale

### Riproducibilità del guadagno

Prendiamo nuovamente in considerazione l'amplificatore non invertente, allo scopo di valutare quanto il suo guadagno sia sensibile a variazioni del guadagno A dell'amplificatore operazionale.

Abbiamo già visto nella lezione dedicata ai principi e alle proprietà della controreazione come uno dei benefici della reazione negativa consista nel rendere poco sensibile il guadagno del circuito reazionato a variazioni del guadagno del blocco di andata. Abbiamo visto come questa proprietà sia efficace tanto più il modulo del guadagno d'anello tende ad un valore elevato, al limite all'infinito.

Vediamo di riscontrare quindi questo beneficio nel caso specifico dell'amplificatore invertente, determinando in particolare *quanto risulti riproducibile il guadagno del circuito in presenza di possibili variazioni del guadagno dell'amplificatore operazionale*. La risposta a questa domanda risulta importante in molti casi pratici, in quanto il guadagno dell'amplificatore operazionale può essere soggetto anche a notevoli variazioni, ad esempio per cambi di temperatura di funzionamento del circuito. Non da ultimo, può essere richiesto che un circuito funzioni con le medesime caratteristiche di guadagno, anche in presenza di una possibile sostituzione dell'amplificatore operazionale.

Nella figura viene riproposta la relazione del guadagno reale  $G$  di un amplificatore invertente. Si è, in particolare, evidenziata la dipendenza dal guadagno  $A$  dell'A.O.

Per calcolare la sensibilità del guadagno  $G$  rispetto a variazioni del guadagno dell'amplificatore operazionale, è possibile innanzitutto differenziare rispetto ad  $A$  l'espressione riportata per  $G$ .

Nell'espressione che ne consegue, che viene mostrata in figura, la variazione differenziale  $dG$  viene espressa rispetto ad una variazione differenziale  $dA$ . Successivamente, dal primo termine  $(-R_2/R_1)$  e

dall'espressione che compare dentro la parentesi al denominatore del secondo termine, può essere estratto il guadagno  $G$ , in accordo con l'espressione riportata in precedenza.

A questo punto, nell'espressione che ne consegue, per semplificare i conti, è possibile pensare di trascurare dentro la parentesi il termine  $(R_1+R_2)/R_1 \cdot A$  rispetto ad uno, nell'ipotesi che il guadagno  $A$  sia molto elevato. Questa semplificazione ci permette di giungere facilmente all'espressione conclusiva dove la variazione relativa del guadagno  $G$ , indicata come  $dG/G$ , viene espressa in funzione della variazione relativa del guadagno dell'operazionale  $A$ , indicata come  $dA/A$ . E' da notare, a questo

$$\boxed{dG = -\frac{R_2}{R_1} \frac{1}{\left(1 + \frac{R_1+R_2}{R_1} \frac{1}{A}\right)^2} \frac{R_1+R_2}{R_1} \frac{1}{A^2} dA}$$

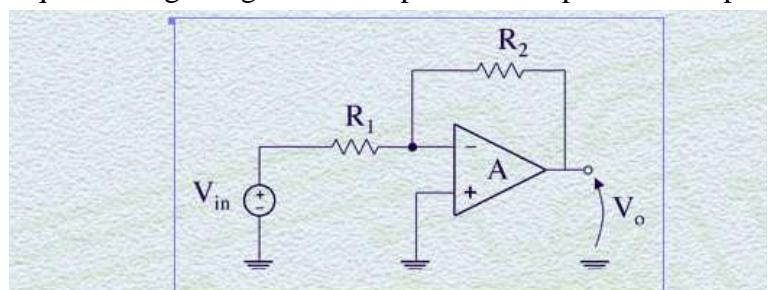
$$\boxed{-\frac{R_2}{R_1} \frac{1}{1 + \frac{R_1+R_2}{R_1} \frac{1}{A}} = G}$$

$$\boxed{dG = G \frac{1}{\left(1 + \frac{R_1+R_2}{R_1} \frac{1}{A}\right)} \frac{R_1+R_2}{R_1} \frac{1}{A^2} dA}$$

$$\frac{dG}{G} = \frac{dA}{A} \frac{R_1+R_2}{R_1} \frac{1}{A} \xrightarrow{A \rightarrow \infty} 0$$

Variazione relativa di  $G$       Variazione relativa di  $A$

punto, che il fattore di proporzionalità tra le variazioni relative dei due guadagni è pari a  $(R_1+R_2)/R_1 \cdot A$ . Questo termine, pari all'inverso del modulo del guadagno d'anello, tende a divenire sempre più piccolo, al limite pari a zero, tanto maggiore è il guadagno dell'amplificatore operazionale  $A$ . Ciò risulta in accordo con quanto visto in

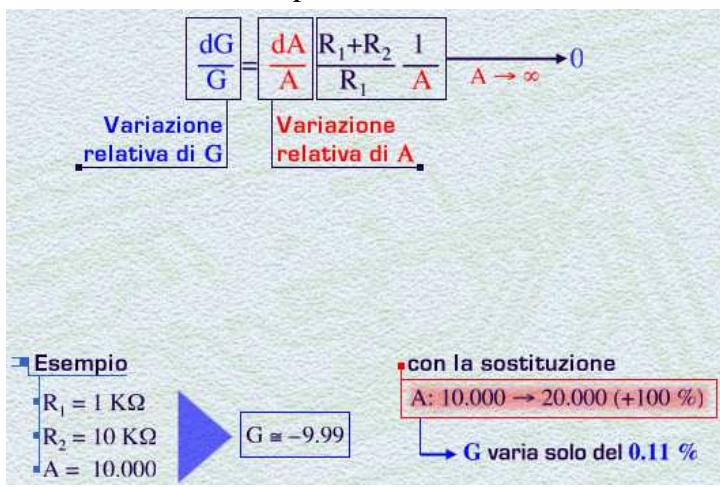


■ Vogliamo determinare quanto è riproducibile il guadagno del circuito in presenza di variazioni del guadagno dell'A.O.

$$G = -\frac{R_2}{R_1} \frac{1}{1 + \frac{R_1+R_2}{R_1} \frac{1}{A}}$$

generale per i sistemi controreazionati. Vediamo un esempio numerico che ci consenta di apprezzare la portata di questo risultato.

Consideriamo un amplificatore invertente in cui  $R_1$  valga  $1\text{ k}\Omega$ ,  $R_2$   $10\text{ k}\Omega$  e il guadagno  $A$  dell'amplificatore operazionale sia pari a 10.000. Il guadagno dell'amplificatore invertente risulta di conseguenza uguale a circa  $-10$ , più precisamente pari a  $-9.99$ . Se ora pensiamo di sostituire nello stesso circuito l'amplificatore operazionale, ad esempio perché non più funzionante, con un altro dispositivo che guadagna il doppio, si può determinare con la relazione ricavata, che a fronte di una variazione così elevata di  $A$ , pari al 100%, il guadagno dell'amplificatore invertente varia solamente dello 0.11% cioè di circa l'uno per mille.

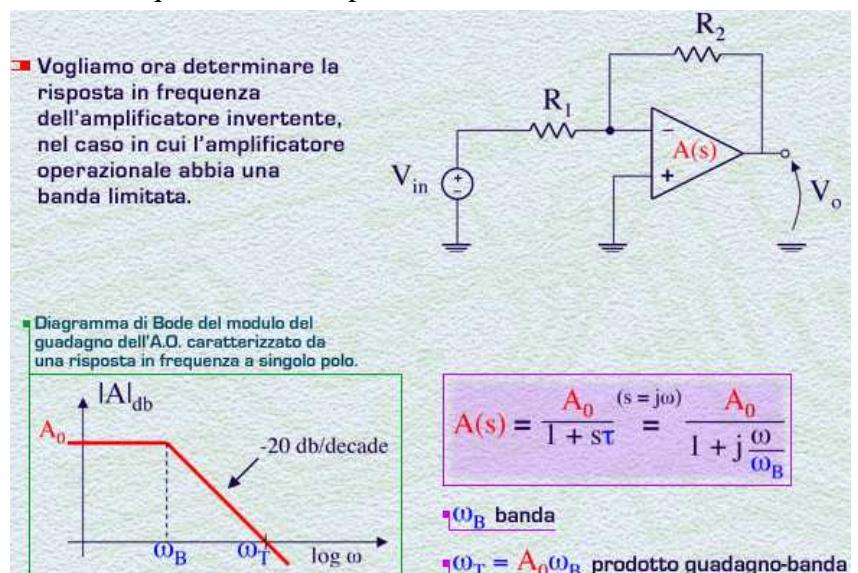


Se si volesse considerare ora *l'amplificatore non invertente* dal punto della riproducibilità del guadagno, si arriverebbe facilmente alla stessa relazione trovata per l'amplificatore invertente. Ciò risulta evidente se si pensa che nell'espressione del guadagno reale dei due circuiti, cambia solo il termine relativo al guadagno ideale, mentre il termine rimanente, in cui è contenuta la dipendenza dal guadagno dell'amplificatore operazionale  $A$ , è il medesimo nei due casi.

### Risposta in frequenza

Vediamo ora di determinare la risposta in frequenza dell'amplificatore invertente, nel caso in cui l'amplificatore operazionale abbia una banda limitata. In figura è mostrato il diagramma di Bode del modulo del guadagno dell'amplificatore operazionale nell'ipotesi che la risposta in frequenza di quest'ultimo sia caratterizzata da un singolo polo. Questo vuole dire che il guadagno  $A(s)$  è esprimibile come  $A_0$ , il guadagno in continua, diviso  $(1+s\tau)$ , ovvero, equivalentemente ponendo  $s=j\omega$ , come  $A_0$  diviso  $(1+j\omega/\omega_B)$ , dove  $\omega_B$  è la banda dell'amplificatore. Come osservato nella lezione introduttiva sull'amplificatore operazionale, la pulsazione  $\omega_T$ , intersezione del diagramma del guadagno con l'asse 0 dB, risulta pari al prodotto guadagno-banda  $A_0 \cdot \omega_B$ .

Consideriamo ora la relazione del guadagno reale dell'amplificatore invertente e introduciamo la dipendenza dalla frequenza per il guadagno  $A$  dell'amplificatore operazionale. Sostituendo ad  $A(\omega)$  l'espressione  $A_0/(1+j\omega/\omega_B)$  e procedendo a sviluppare i conti, si perviene ad un'espressione per  $G$  pari a  $-R_2/R_1$  moltiplicato per una frazione. Il secondo termine al denominatore di questa frazione, cioè  $1/A_0$  per  $(R_1+R_2)/R_1$ , può essere ragionevolmente trascurato, nell'ipotesi che il guadagno  $A_0$  sia sufficientemente elevato e comunque molto maggiore di  $(R_1+R_2)/R_1$ . In questo caso, l'espressione si semplifica notevolmente e diviene pari a  $-R_2/R_1$  diviso  $(1+j\omega/\omega_1)$ . Il guadagno  $G$  è caratterizzato quindi anch'esso da un polo la cui pulsazione è pari a  $\omega_B$ , la banda dell'amplificatore operazionale, che moltiplica il termine  $A_0 \cdot R_1 / (R_1 + R_2)$ . La banda dell'amplificatore invertente  $\omega_1$  è cioè pari alla banda dell'amplificatore operazionale  $\omega_B$  per il modulo del  $G_{loop}$  a frequenza zero. La banda si estende



quindi in maniera proporzionale al modulo del guadagno d'anello, come avevamo peraltro già osservato, in via più generale, nella trattazione dei circuiti controreazionati. Inoltre se supponiamo che

$$G(\omega) = -\frac{R_2}{R_1} \frac{1}{1 + \frac{R_1 + R_2}{R_1}} \frac{1}{A_0} \quad A(\omega) = \frac{A_0}{1 + j \frac{\omega}{\omega_B}}$$

$$G(\omega) = -\frac{R_2}{R_1} \frac{1}{1 + \frac{1}{A_0} \frac{R_1 + R_2}{R_1} + j \frac{\omega}{\omega_1}} \cong -\frac{R_2}{R_1} \frac{1}{1 + j \frac{\omega}{\omega_1}}$$

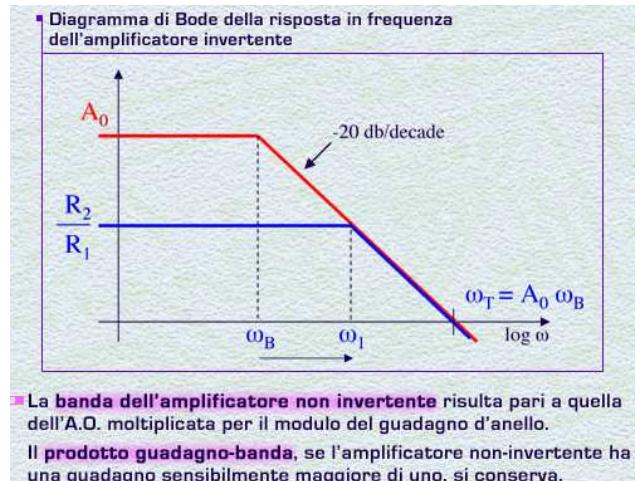
$$\omega_1 = \omega_B \frac{A_0 R_1}{(R_1 + R_2)} = \omega_B (-G_{loop}) \quad \text{La banda si estende in maniera proporzionale al modulo del } G_{loop}$$

se  $\frac{R_2}{R_1} \gg 1$   $\omega_1 = \omega_T \frac{R_1}{R_2}$   $|G| \omega_1 = \omega_T$  Il prodotto guadagno-banda si conserva.

il rapporto tra  $R_2$  e  $R_1$  sia molto maggiore di 1, come è ragionevole che sia nel caso di un amplificatore invertente che possieda un guadagno di una certa entità, allora il polo  $\omega_1$  del circuito diventa pari a  $\omega_T$ , il prodotto guadagno-banda dell'amplificatore operazionale, per  $R_1/R_2$ . Portando, in quest'ultima relazione, il termine  $R_2/R_1$  a sinistra, otteniamo che il prodotto tra il modulo del guadagno  $G$  dell'amplificatore invertente e la pulsazione  $\omega_1$  risulta pari a  $\omega_T$ . Possiamo quindi concludere dicendo che il prodotto guadagno banda dell'amplificatore invertente risulta pari al prodotto guadagno

banda dell'amplificatore operazionale con cui è stato realizzato, cioè che il prodotto guadagno-banda si conserva.

Nel diagramma di Bode della risposta in frequenza dell'amplificatore non invertente viene mostrato l'allargamento della banda rispetto a quella dell'amplificatore operazionale utilizzato, ed il fatto che il prodotto guadagno-banda si conservi. Nel caso di un amplificatore non invertente, si perviene alle stesse conclusioni che abbiamo determinato per l'amplificatore invertente e cioè che la banda dell'amplificatore risulta pari a quella dell'amplificatore operazionale moltiplicata per il modulo del guadagno d'anello e che il prodotto guadagno-banda, supponendo che l'amplificatore non-invertente abbia una guadagno sensibilmente maggiore di uno, si conserva.



## Risposta al segnale

Dopo aver analizzato la risposta in frequenza dell'amplificatore invertente, vediamo di determinare la sua risposta ad un gradino di tensione applicato all'ingresso. Consideriamo innanzitutto un gradino di tensione, ad esempio positivo, applicato in ingresso al circuito e caratterizzato da un'ampiezza  $V_I$  non elevata. Siamo in presenza, cioè di un segnale piccolo. In questo caso la risposta dell'amplificatore al gradino di tensione risulta determinata esclusivamente dalla sua risposta in frequenza. Abbiamo visto che la risposta in frequenza dell'amplificatore è caratterizzata da un polo, analogamente alla risposta in frequenza di una semplice rete RC, o integratore approssimato. E' di un tipo simile di rete è di tipo esponenziale.

noto dall'elettrotecnica che la risposta al gradino di tensione di uscita, che è invertita rispetto all'ingresso, è pari a:

$$V_o(t) = -\frac{R_2}{R_1} \cdot V_I \cdot (1 - e^{-\omega_1 t})$$

$\omega_1$  è la pulsazione del polo della risposta in frequenza del circuito.

L'uscita parte quindi da zero e scende esponenzialmente con costante di tempo pari a  $1/\omega_1$  fino a tendere al valore asintotico pari a  $-R_2/R_1$  per  $V_I$ .

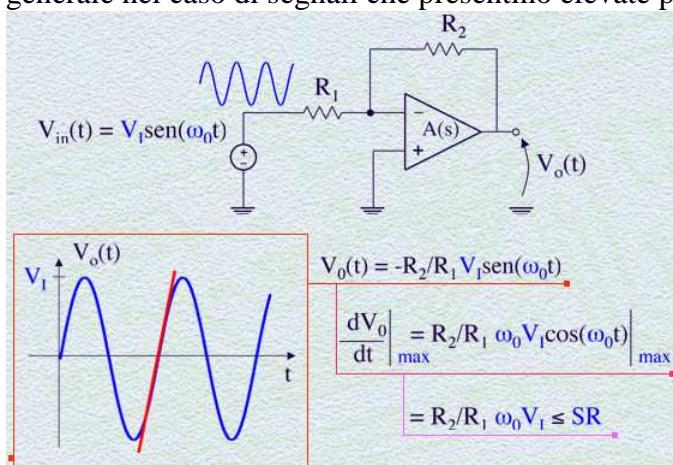
Nel caso in cui il segnale di ingresso abbia un'ampiezza elevata, un ulteriore fenomeno interviene a limitare la velocità di risposta dell'amplificatore al gradino: è lo slew-rate. Abbiamo visto nella lezione sulle caratteristiche dell'amplificatore operazionale, infatti, che questo dispositivo è caratterizzato da una limitata capacità a variare velocemente la sua tensione di uscita su grande segnale. Lo slew rate, il cui valore viene fornito dal costruttore del dispositivo, è definito come la **massima derivata, rispetto al tempo, della tensione di uscita**.

Un gradino di elevata ampiezza applicato in ingresso all'amplificatore implica una grande escursione del segnale all'uscita dello stesso. Nei casi pratici, pertanto, la risposta in uscita risulta al grande segnale risulta limitata più dallo slew-rate che dalla risposta in frequenza del circuito.

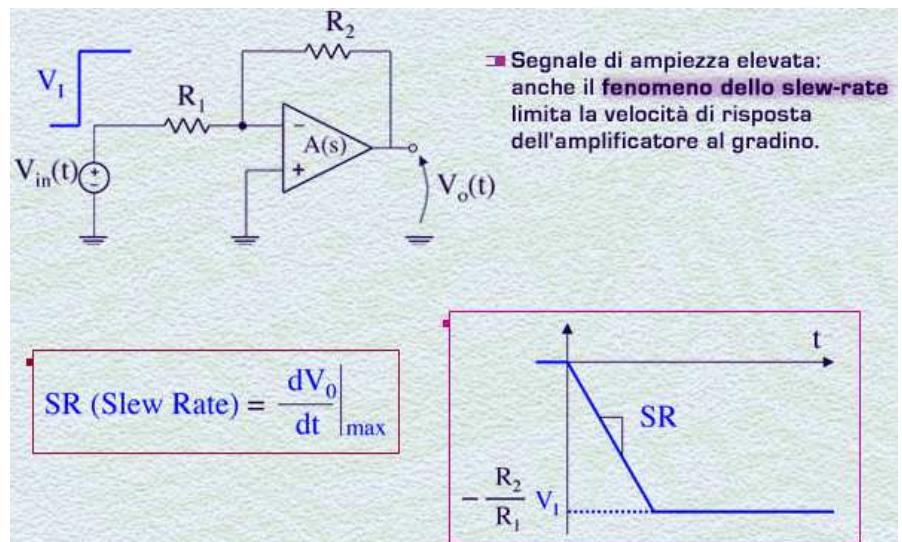
Tipicamente, nel nostro esempio, la tensione in uscita si porta a valori negativi fino a raggiungere il valore finale  $-R_2/R_1$  per  $V_I$ , con un tipico andamento a rampa. La

pendenza di questa rampa è data dal valore dello slew-rate dell'amplificatore operazionale, parametro espresso in volt/ $\mu$ s.

La limitazione nella risposta di un amplificatore legata allo slew-rate deve essere considerata più in generale nel caso di segnali che presentino elevate pendenze. Esaminiamo ad esempio il semplice caso



$R_2/R_1$ , questa limitazione può comportare un massimo valore di ampiezza consentita per una sinusoida di una certa frequenza fissata  $\omega_0$ . Questa ampiezza limite risulta pari a  $V_{I\text{max}} = \text{Slew-Rate}/[(R_2/R_1) \cdot \omega_0]$ . Analogamente è possibile definire la massima frequenza di una sinusoida, questa volta di ampiezza fissata  $V_I$ . Tale frequenza limite risulta pari a  $\omega_{0\text{max}} = \text{Slew-Rate}/[(R_2/R_1) \cdot V_I]$ . Naturalmente risulta evidente da queste relazioni che maggiore è l'ampiezza  $V_I$  di una sinusoida che si vuole amplificare senza distorsione, minore risulterà la frequenza  $\omega_0$  massima concessa dallo Slew-Rate, e analogamente maggiore è la frequenza  $\omega_0$  della sinusoida che si vuole amplificare senza distorsione, minore risulterà l'ampiezza  $V_I$  massima concessa dallo Slew-Rate.



di una sinusoida di ampiezza elevata applicata all'ingresso dell'amplificatore. La sinusoida può essere espressa come un segnale  $V_{in}(t) = V_I \sin(\omega_0 t)$ . La sinusoida in uscita dell'amplificatore invertente sarà pari a  $-R_2/R_1$  per  $V_I \sin(\omega_0 t)$ . La massima pendenza di questa sinusoida è pari al massimo valore della sua derivata, in modulo  $(R_2/R_1) \cdot \omega_0 \cdot V_I \cdot \cos(\omega_0 t)$ . Il massimo valore del modulo della derivata è uguale a  $(R_2/R_1) \cdot \omega_0 \cdot V_I$ . Il massimo valore di questo prodotto è limitato dal massimo valore dello slew-rate del circuito.

Fissato un certo guadagno dell'amplificatore

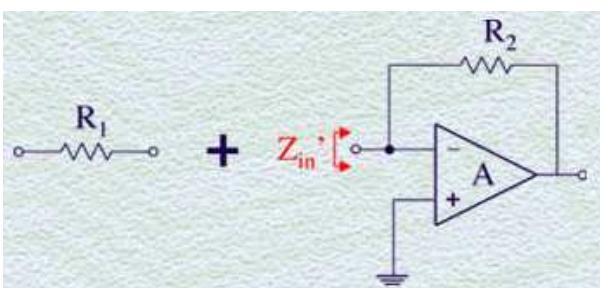
$$\begin{aligned} & R_2/R_1 \omega_0 V_I \leq SR \\ \downarrow & \quad \omega_0 \text{ fissato} \quad \quad \quad V_I \text{ fissato} \\ & V_{I\text{max}} = \frac{SR}{R_2/R_1 \omega_0} \\ & \omega_{0\text{max}} = \frac{SR}{R_2/R_1 V_I} \end{aligned}$$

## Lezione n. 29 : Impedenze di ingresso e di uscita

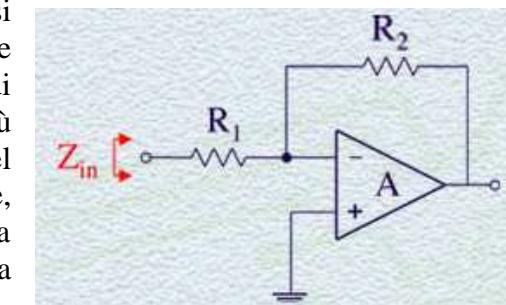
### L'impedenza di ingresso

Iniziamo con il determinare l'impedenza di ingresso di un amplificatore invertente.

Possiamo innanzitutto osservare che il problema che ci apprestiamo ad esaminare può di fatto essere semplificato. Infatti dato che la prima impedenza che si incontra all'ingresso del circuito risulta essere  $R_1$ , è possibile pensare di valutare l'impedenza di ingresso come la somma di due impedenze disposte in serie: la resistenza  $R_1$  più l'impedenza  $Z'_\text{in}$  vista a valle di  $R_1$  cioè all'ingresso del circuito composto ora dal solo amplificatore operazionale,



controreaziona to con la resistenza  $R_2$ .

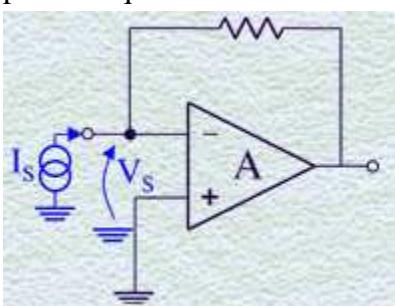


Nota  $R_1$ , il problema si sposta ora alla determinazione dell'impedenza  $Z'_\text{in}$ .

Per determinare l'impedenza  $Z'_\text{in}$  occorre utilizzare il metodo noto dall'elettrotecnica che consiste nell'imporre in ingresso al circuito un generatore di

sonda, di tensione o di corrente, e determinare alla stessa morsetto l'altra grandezza elettrica, che sarà una corrente se si è imposta una tensione o il viceversa.

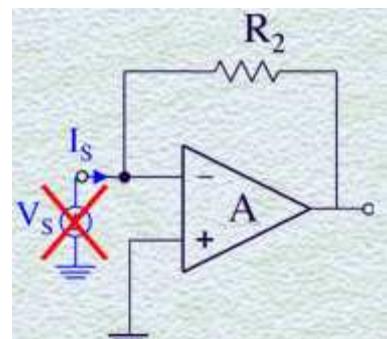
Occorre però osservare, nel caso di questo circuito come più in generale nel caso di circuiti controreazionati, che *un'opportuna scelta del tipo di generatore da utilizzarsi come sonda, generatore di corrente o di tensione, risulti fondamentale ai fini di una corretta determinazione dell'impedenza*. Supponiamo infatti di utilizzare come generatore sonda un generatore di tensione. Questa scelta non risulta opportuna, in quanto in questa maniera viene forzato direttamente in tensione il nodo della terra virtuale. Se viene imposta una tensione al nodo della terra virtuale, quest'ultimo non si può più considerare come tale, in quanto non è più la reazione che ne determina la tensione, prossima allo zero. L'impedenza che determineremmo in questo modo non è pertanto quella del circuito in cui la reazione è libera di agire.



La scelta corretta del generatore di sonda ricade quindi su un generatore di corrente.

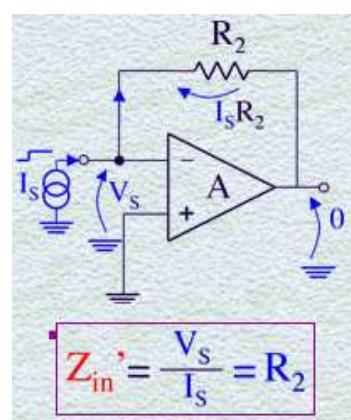
Questo infatti può introdurre nel circuito la perturbazione necessaria al fine di determinare l'impedenza di ingresso, ma senza impedire alla reazione di esercitare il suo effetto, in quanto non viene imposta una tensione sul nodo di terra virtuale.

Per determinare l'impedenza  $Z'_\text{in}$ , consideriamo due situazioni estreme: una situazione in cui la reazione sia spenta, cioè il guadagno d'anello sia pari a 0, e una seconda situazione in cui la reazione sia



ideale, cioè il guadagno d'anello tenda la limite all'infinito.

Consideriamo che nel primo caso la reazione venga spenta ad esempio supponendo che il guadagno dell'amplificatore operazionale sia nullo. In questo caso, qualunque variazione di tensione si sviluppi sul morsetto invertente in seguito all'applicazione del segnale di corrente sonda  $I_s$ , la tensione di uscita dell'amplificatore resterà sempre immutata e pari a zero. In questa situazione, l'estremo di  $R_2$  connesso all'uscita dell'amplificatore resterà sempre fisso e sulla resistenza si svilupperà una caduta di tensione pari a  $I_s$  per  $R_2$ . Tale caduta è pari anche alla tensione  $V_s$  sul morsetto invertente dell'amplificatore operazionale. L'impedenza di ingresso calcolata come rapporto tra  $V_s$  e  $I_s$  risulta pertanto pari a  $R_2$ .



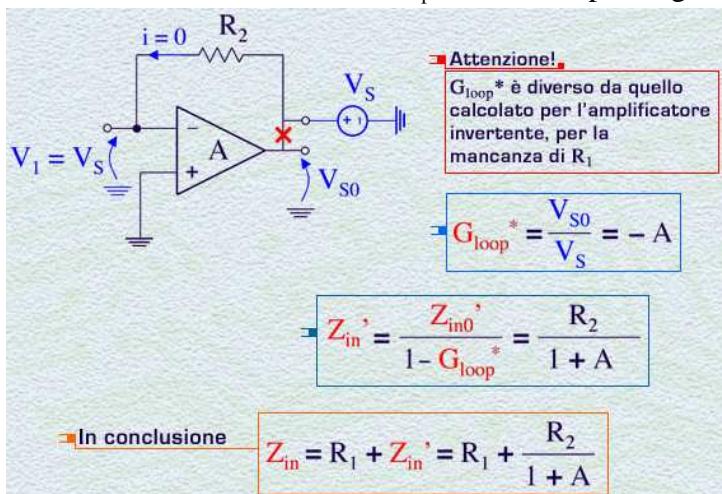
Se ora consideriamo il caso opposto a quello appena trattato, in cui si supponga una **reazione ideale**, ad esempio considerando un guadagno A dell'amplificatore operazionale tendente all'infinito, vediamo quali sono gli effetti sul circuito introdotti dall'applicazione del segnale sonda  $I_s$ . In seguito all'applicazione di  $I_s$ , il morsetto non invertente tenderà a crescere in tensione. Di conseguenza l'uscita dell'amplificatore tenderà a scendere in tensione e ciò provocherà, grazie alla presenza di  $R_2$  in reazione, una discesa anche del morsetto invertente, salito inizialmente per l'effetto dell'applicazione del segnale. Risulta intuitivo pensare che nel caso di reazione ideale, l'effetto della reazione tenderà a compensare completamente ogni variazione di tensione sul morsetto invertente, che pertanto tenderà a restare a tensione nulla. Questo fenomeno, già visto nella trattazione dell'amplificatore invertente come circuito controreazionato, consente di identificare questo morsetto come una terra virtuale e pertanto l'impedenza  $Z'_\text{in}$  risulterà prossima a zero.

L'aver valutato l'impedenza di ingresso in questi due casi estremi ci consente di estrarre, anche se in maniera puramente intuitiva, una relazione generale che esprime l'impedenza di ingresso per questo tipo di circuito. Infatti, visto che nel caso di reazione 'spenta' l'impedenza risulta pari all'impedenza fisica, nel nostro caso  $R_2$ , connessa al morsetto in esame, e visto che l'impedenza tende a zero più ideale diventa la reazione, cioè maggiore è il modulo del guadagno d'anello, risulta intuitivo pensare che nel caso generale l'impedenza risulti pari all'impedenza  $Z'_{\text{in}0}$  che si misura a reazione spenta, divisa per il termine oramai familiare che identifica la bontà della reazione, pari a  $1 - G_{\text{loop}}$ . Questa espressione, che può essere dimostrata rigorosamente, giustifica le impedenze calcolate nei due casi estremi illustrati in precedenza.

Per il calcolo dell'impedenza a reazione spenta  $Z'_{\text{in}0}$ , è sufficiente determinare la risultante delle impedenze fisiche connesse al morsetto di ingresso, nel nostro caso la sola  $R_2$ . Per quanto riguarda il guadagno d'anello occorre determinare quello del circuito specifico che stiamo considerando.

A tale proposito, occorre osservare che il circuito che stiamo analizzando non è più l'amplificatore invertente completo, in quanto in questo circuito non compare più la resistenza  $R_1$ . Per questa ragione, nell'espressione dell'impedenza di ingresso abbiamo indicato il guadagno d'anello con  $G_{\text{loop}}^*$ , che è diverso dal  $G_{\text{loop}}$  senza asterisco calcolato per l'amplificatore invertente, con la presenza di  $R_1$ .

Procedendo al calcolo di  $G_{\text{loop}}^*$ , ad esempio tagliando l'anello sull'uscita dell'amplificatore, e



procedendo nel modo già visto nelle lezioni precedenti, si osserva che nella resistenza  $R_2$  non può scorrere corrente perché corrente non può scorrere nel morsetto invertente. Pertanto la tensione sul morsetto invertente vale  $V_s$  ed il guadagno d'anello risulta semplicemente pari a  $-A$ .

Di conseguenza, l'impedenza  $Z'_{\text{in}}$  risulta pari  $R_2$  diviso  $(1+A)$ .

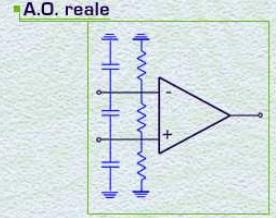
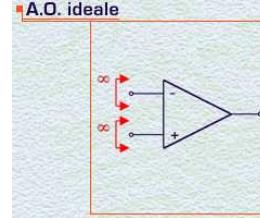
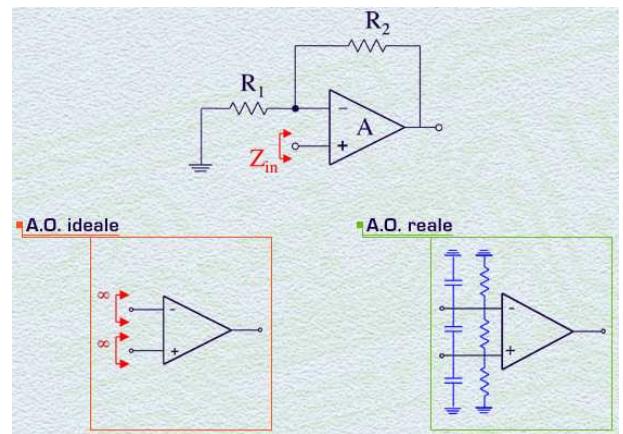
Concluso l'esame del sottocircuito ottenuto dall'amplificatore invertente togliendo la resistenza  $R_1$ , come visto all'inizio, risulta ora semplice mettere insieme i risultati conclusivi dell'analisi effettuata. Come

visto all'inizio, infatti, l'impedenza di ingresso dell'amplificatore invertente risulta pari a  $R_1$  più l'impedenza  $Z'_{\text{in}}$  appena calcolata. L'impedenza  $Z_{\text{in}}$  risulta pertanto pari a  $R_1 + R_2/(1+A)$ .

Occorre osservare, in conclusione, che maggiore è il guadagno  $A$  dell'amplificatore operazionale, più il morsetto invertente diviene terra virtuale e quindi l'impedenza di ingresso del circuito risulta sempre più uguale alla sola resistenza  $R_1$ .

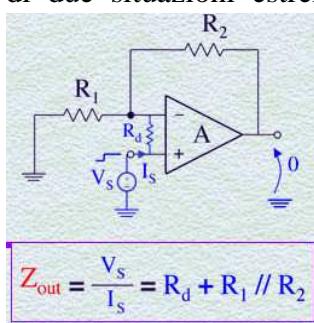
Consideriamo ora il calcolo dell'impedenza di ingresso dell'amplificatore non invertente. Occorre innanzitutto osservare che se si considera un amplificatore operazionale ideale, dotato di impedenza di ingresso infinita, risulta inutile effettuare questo calcolo in quanto l'impedenza di ingresso dell'amplificatore non invertente risulterà anch'essa infinita.

Nel caso, invece, di un amplificatore operazionale reale, l'impedenza di ingresso non è infinita ma mostra diverse componenti resistive e capacitiva, sia di modo comune che di modo differenziale, come visto nella lezione relativa alle caratteristiche dell'amplificatore operazionale. Per semplicità, procediamo al calcolo dell'impedenza di ingresso considerando la sola resistenza differenziale di ingresso  $R_d$ .



Occorre osservare che al fine di determinare l'impedenza di ingresso, occorre anche in questo caso utilizzare un generatore di sonda che non comprometta l'effetto dovuto alla contoreazione. In questo caso la scelta obbligata è di un **generatore di tensione** in quanto un generatore di corrente imporrebbe una tensione tra i morsetti invertente e non invertente, tramite la caduta di tensione provocata dalla corrente sonda sulla resistenza  $R_d$ . Questa tensione imposta tra i morsetti di ingresso sarebbe incompatibile con la creazione di un corto-circuito virtuale tra gli stessi per effetto della contoreazione. Occorre quindi usare un generatore sonda di tensione.

Procedendo, conviene anche in questo caso determinare l'impedenza di ingresso a partire dall'esame di due situazioni estreme, come fatto nel caso della determinazione dell'impedenza di ingresso dell'amplificatore invertente.



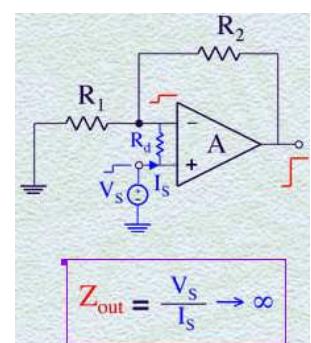
$$Z_{out} = \frac{V_s}{I_s} = R_d + R_1 // R_2$$

Consideriamo innanzitutto la situazione in cui la **reazione** sia **spenta**, ad esempio supponendo nullo il guadagno  $A$  dell'amplificatore operazionale. In questo caso l'applicazione di un segnale sonda di corrente  $V_s$  determina una corrente  $I_s$  che è legata solamente all'impedenza fisica totale connessa al morsetto non invertente, cioè la somma di  $R_d$  con  $R_1//R_2$ . Nessun ulteriore effetto è introdotto dalla reazione in quanto l'amplificatore operazionale è spento, cioè guadagna zero e quindi la sua uscita resta a tensione nulla.

Nella situazione opposta, cioè considerando ad esempio un **guadagno A infinito**, il segnale sonda  $V_s$  provoca un segnale di tensione positivo sull'uscita dell'amplificatore operazionale, che viene riportato anche sul morsetto invertente, tramite la partizione tra  $R_1$  e  $R_2$ . Come conseguenza di questo, la tensione sviluppata sulla resistenza  $R_d$ , inizialmente alzata per effetto dell'innalzamento del morsetto non-invertente, tende a ridursi per l'innalzamento anche del morsetto invertente. In caso di reazione ideale, è possibile pensare che la tensione sviluppata su  $R_d$ , e di conseguenza la corrente circolante nella stessa, tenda ad essere nulla. L'impedenza di uscita in questo caso risulta quindi pari a  $V_s$  diviso per una corrente  $I_s$  che tende a zero. Di conseguenza l'impedenza tende all'infinito.

Al contrario di quanto dedotto per l'impedenza di ingresso dell'amplificatore invertente, è possibile pensare in questo caso, sulla base di quanto visto nelle due situazioni estreme descritte, che in generale l'impedenza di uscita risulti pari all'impedenza  $Z_{in0}$  calcolata a reazione spenta, moltiplicata per il termine  $(1-G_{loop})$ , dove il  $G_{loop}$  è il guadagno d'anello calcolato per l'amplificatore non-invertente (uguale a quello dell'amplificatore invertente).

Inserendo le espressioni corrispondenti per  $Z_{in0}$  e per il guadagno d'anello, l'impedenza di ingresso  $Z_{in}$  risulta quindi pari a  $(R_d + R_1//R_2)$ , moltiplicato per  $[1+A*R_1/(R_1+R_2)]$ .



$$Z_{out} = \frac{V_s}{I_s} \rightarrow \infty$$

**Impedenza calcolata a reazione spenta**

$$Z_{in} = (R_d + R_1 // R_2) \left( 1 + A \frac{R_1}{(R_1+R_2)} \right) \xrightarrow{A \rightarrow \infty} \infty$$

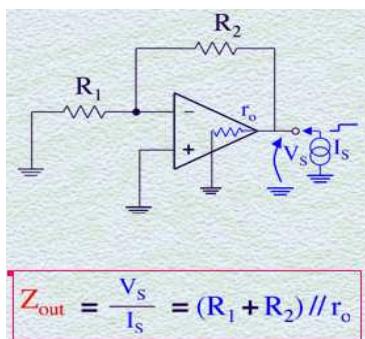
Per cui in corrispondenza di un guadagno **A** che tende all'infinito, l'impedenza di uscita tende anch'essa all'infinito, come visto nella seconda situazione estrema analizzata in precedenza.

Nell'amplificatore non invertente, in conclusione, **la controreazione tende ad alzare l'impedenza di ingresso**.

### L'impedenza di uscita

Consideriamo ora il calcolo dell'impedenza di uscita dell'amplificatore invertente. Occorre innanzitutto osservare che se si considera un amplificatore operazionale ideale, dotato di impedenza di uscita nulla, risulta inutile effettuare questo calcolo in quanto l'impedenza di uscita dell'amplificatore invertente risulta ovviamente anch'essa nulla.

Il calcolo in questione pertanto assume significato nel caso in cui si consideri un **amplificatore operazionale reale**, dotato di una impedenza di uscita non nulla pari a  $r_o$ . Anche in questo caso, conviene determinare l'impedenza di uscita a partire dall'esame di due situazioni estreme, come fatto in precedenza.



Consideriamo innanzitutto la situazione in cui la **reazione sia spenta**, ad esempio supponendo nullo il guadagno **A** dell'amplificatore operazionale. In questo caso l'applicazione di un segnale sonda di corrente  $I_s$  sviluppa una tensione  $V_s$  sul morsetto di uscita che è determinata solamente dall'impedenza risultante dalle resistenze connesse a tale morsetto, cioè  $(R_1 + R_2) // r_o$ . Nessun effetto su  $V_o$  è provocato dalla reazione in quanto anche se un segnale viene riportato sul morsetto invertente tramite la partizione tra  $R_1$  e  $R_2$ , ciò non comporta nessun effetto sull'uscita in quanto l'amplificatore

operazionale è spento, cioè guadagna zero.

Nella situazione opposta, cioè considerando ad esempio un **guadagno A infinito**, il segnale sonda  $I_s$  provoca un segnale di tensione positivo sull'uscita, che viene riportato anche sul morsetto invertente, tramite la partizione tra  $R_1$  e  $R_2$ . Come conseguenza di questo, l'amplificatore operazionale, che adesso è acceso, determinerà un segnale negativo sulla sua uscita che tende a compensare la perturbazione positiva iniziale. In caso di reazione ideale, è possibile pensare che tale compensazione sia perfetta e cioè che il morsetto di uscita tenda a non spostarsi in tensione.

L'impedenza di uscita in questo caso risulta quindi tendere a zero. Analogamente a quanto fatto nel caso dell'impedenza di ingresso dell'amplificatore invertente, è quindi possibile pensare, sulla base di quanto visto nelle due situazioni, che nel **caso generale** l'impedenza di uscita risulti pari all'impedenza  $Z_{out0}$  calcolata a reazione spenta, divisa per il termine  $(1 - G_{loop})$ , dove il guadagno

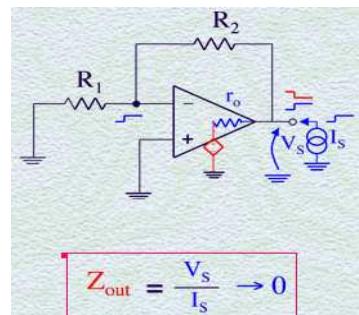
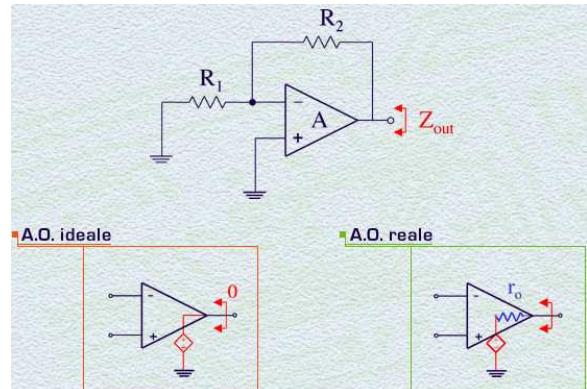
$$Z_{out} = \frac{Z_{out0}}{1 - G_{loop}}$$

Impedenza calcolata  
a reazione spenta

d'anello è quello già calcolato per l'amplificatore invertente nell'ipotesi che la presenza di  $r_o$  non lo modifichi in maniera sostanziale. Inserendo le espressioni corrispondenti per  $Z_{out0}$  e per il guadagno

d'anello, l'impedenza di uscita  $Z_{out}$  risulta quindi pari a  $(R_1 + R_2) // r_o$ , diviso  $[1 + A * R_1 / (R_1 + R_2)]$ . In corrispondenza di un guadagno **A** che tende all'infinito, l'impedenza di uscita tende a zero, come visto nel secondo caso analizzato in precedenza.

La **controreazione** tende pertanto a **ridurre l'impedenza di uscita**. Questa conclusione risulta valida anche per l'amplificatore non invertente, il cui circuito per il calcolo dell'impedenza di uscita è identico a quello dell'amplificatore invertente.



$$Z_{out} = \frac{(R_1 + R_2) // r_o}{1 + A \frac{R_1}{(R_1 + R_2)}} \xrightarrow{A \rightarrow \infty} 0$$

## Lezione n. 30 : Esempi ed esercizi A

### Riproducibilità del guadagno

In questo esercizio, verrà considerato un amplificatore invertente allo scopo di determinare il legame che intercorre tra il massimo guadagno ottenibile per l'amplificatore ed un richiesto grado di riproducibilità dello guadagno stesso. Questi due parametri, infatti, risultano legati fra di loro.

Nel caso particolare mostrato in figura si consideri un amplificatore invertente realizzato utilizzando un amplificatore operazionale di guadagno pari a 100 dB. Si supponga che questo guadagno possa subire una variazione, in termini relativi,  $dA/A$  pari al 10%. Si supponga, invece, che una volta fissati i valori delle resistenze  $R_1$  e  $R_2$  per ottenere un certo valore di guadagno, queste resistenze non cambino il loro valore. Cioè  $dR/R$  sia pari a 0.

Ci poniamo, in queste condizioni, l'obiettivo di determinare quanto possa valere il massimo valore di guadagno dell'amplificatore invertente, compatibilmente con la richiesta di avere una stabilità di questo guadagno, a seguito della variazione di  $A$ , migliore dello 0.01 %.

Abbiamo visto in una delle lezioni scorse che un effetto benefico della contoreazione sull'amplificatore invertente, consiste nel ridurre sensibilmente le possibili variazioni del guadagno del circuito provocate da possibili variazioni del guadagno  $A$  dell'amplificatore operazionale.

Avevamo, in particolare, determinato che la variazione relativa del guadagno ad anello chiuso  $dG/G$  è

$$\frac{dG}{G} = \frac{dA}{A} - \frac{1}{G_{\text{loop}}} = \frac{dA}{A} \frac{R_1 + R_2}{R_1} \frac{1}{A}$$

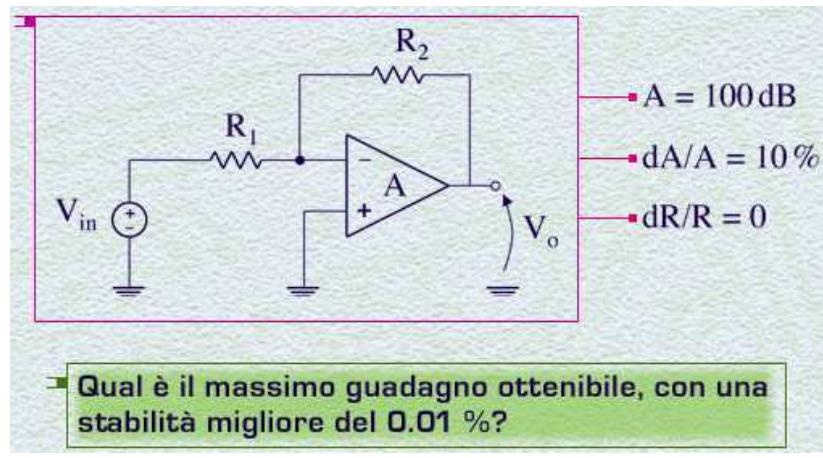
$$= \frac{dA}{A} \left( 1 + \frac{R_2}{R_1} \right) \frac{1}{A} = \frac{dA}{A} \left( 1 - G \right) \frac{1}{A}$$

$$\frac{dG}{G} = \frac{dA}{A} \left( 1 - G \right) \frac{1}{A} < 0.01 \%$$

$$G < 0.01 / 10 \times 100.000 = 100$$

**Osservazione:** maggiore è la stabilità richiesta, minore è il guadagno

$$\frac{dG}{G} < 0.01 \% \rightarrow G < 10$$



pari alla variazione relativa del guadagno dell'amplificatore operazionale  $dA/A$  ridotta di una quantità pari al modulo del guadagno d'anello,  $-G_{\text{loop}}$ . Inserendo in questa relazione l'espressione del guadagno d'anello, si ottiene che  $dG/G$  è uguale a  $dA/A$  per  $(R_1 + R_2)/(R_1 * A)$ , e quindi uguale a  $dA/A$  per  $(1 + R_2/R_1)/A$ .

Ricordando che il rapporto  $-R_2/R_1$  corrisponde al guadagno  $G$  del circuito, si ottiene che  $dG/G$  è uguale a  $dA/A$  per  $(1 - G)/A$ .

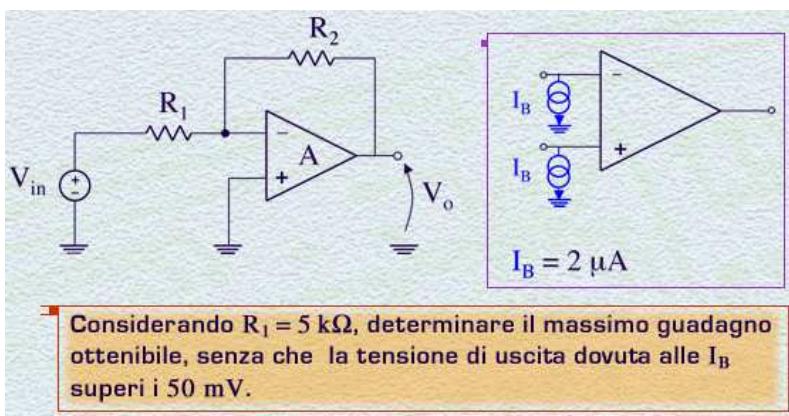
Imponendo che la variazione del guadagno sia inferiore al valore

richiesto, lo 0.01%, e inserendo i valori di  $dA/A$ , pari al 10%, e di  $A$ , pari a 100 dB, cioè a 100.000, è possibile determinare il massimo guadagno  $G$  che soddisfi alla condizione imposta. Da questo calcolo risulta che  $G$  deve essere minore di  $0.01/10$  per 100.000, cioè  $G$  deve essere minore di 100.

A conclusione di questo esercizio osserviamo che maggiore è la stabilità richiesta per il guadagno del circuito, minore sarà il massimo valore consentito per tale guadagno. Se ad esempio si fosse richiesta una riproducibilità del guadagno 10 volte migliore rispetto a quella imposta nell'esercizio, cioè  $dG/G$  minore dello 0.01 per mille anziché percento, il guadagno massimo sarebbe stato inferiore a 10 anziché a 100.

## Correnti di bias

Consideriamo in questo esercizio un amplificatore invertente con l'obiettivo di valutare l'effetto provocato sulla tensione di uscita del circuito dalla presenza delle correnti di bias all'ingresso dell'amplificatore operazionale.



La presenza di queste correnti era stata introdotta nella lezione relativa all'amplificatore operazionale. Supponiamo di considerare correnti di bias entranti dai morsetti di ingresso dell'amplificatore operazionale e, per il momento, di uguale valore  $I_B$  pari a  $2\mu\text{A}$ .

Consideriamo il valore della resistenza  $R_1$  fisso a  $5 \text{ k}\Omega$  e determiniamo quale sia il massimo guadagno ottenibile per il circuito, imponendo che la tensione di uscita dovuta alla presenza delle correnti di bias non superi i 50 mV.

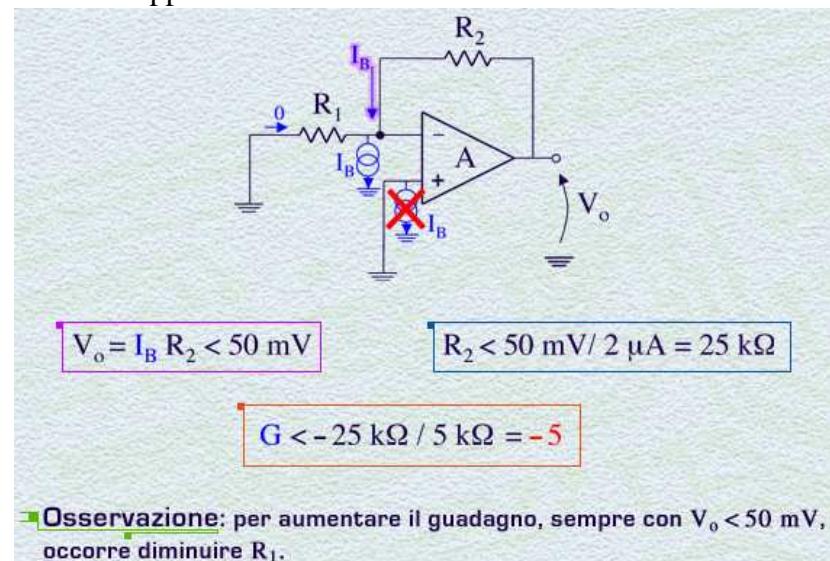
Per valutare il solo effetto delle correnti di bias sull'uscita, spegniamo il generatore di segnale in ingresso. Osserviamo innanzitutto che il generatore di bias applicato al morsetto non invertente non ha alcun effetto sulla tensione di uscita. Infatti tutta la sua corrente  $I_B$ , fluisce verso massa e non provoca alcuna modifica alla tensione del morsetto non invertente che è fissato alla massa stessa.

Analizziamo invece l'effetto del generatore applicato al morsetto non invertente. Osserviamo innanzitutto che tra ai capi della resistenza  $R_1$  si trova una differenza di tensione nulla. Infatti un estremo di  $R_1$  si trova a massa mentre l'altro è fissato anch'esso a tensione nulla dalla massa virtuale del circuito. Pertanto, nella resistenza  $R_1$  non scorre corrente. L'intera corrente  $I_B$  del generatore di bias scorre quindi nella resistenza  $R_2$ , sviluppando sull'uscita una tensione pari a  $I_B$  per  $R_2$ . Tale tensione deve essere inferiore a 50mV. Questo implica un massimo valore della resistenza  $R_2$  utilizzabile pari a  $50\text{mV}/2\mu\text{A}$  e cioè pari a  $25 \text{ k}\Omega$ .

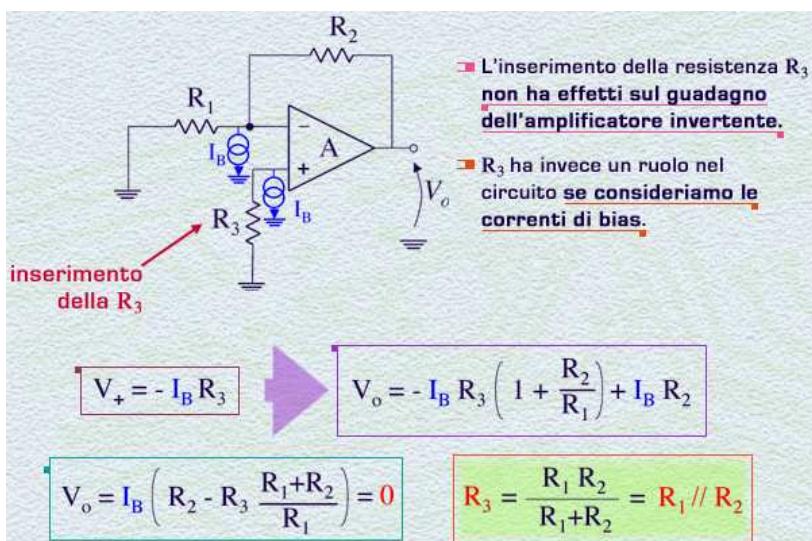
Il massimo guadagno del circuito, fissata  $R_1$  a  $5 \text{ k}\Omega$ , risulta pertanto pari a  $-25 \text{ k}\Omega/5 \text{ k}\Omega$  cioè uguale a -5.

Si può osservare, in conclusione, che la massima tensione in uscita dovuta alla corrente di bias limita il massimo valore di  $R_2$  e quindi il massimo guadagno del circuito, per una fissata  $R_1$ . Per aumentare il guadagno una volta fissata  $R_2$  per soddisfare alla massima tensione di uscita tollerata per la presenza delle correnti di bias, è opportuno invece cercare di ridurre il valore di  $R_1$ , che nell'esempio era invece fissato. La resistenza  $R_1$  infatti non ha alcun effetto sul trasferimento in uscita delle correnti di bias, mentre una sua riduzione aumenta il valore del guadagno.

Consideriamo sempre l'amplificatore invertente, ma introduciamo ora un accorgimento per annullare l'effetto delle correnti di bias. Abbiamo appena visto che dei due generatori di corrente di bias, uno solo di questi provoca un effetto sull'uscita dell'amplificatore. Riuscendo a far entrare in gioco opportunamente anche l'altro generatore, quello connesso al morsetto non invertente, è possibile creare una compensazione degli effetti di entrambi i generatori, rendendo nullo l'effetto globale sull'uscita.



Per fare questo viene inserita una resistenza  $R_3$  sul morsetto non invertente. L'aver introdotto questa resistenza non comporta alcune effetto sul guadagno dell'amplificatore invertente, in quanto in presenza di un segnale in ingresso, nessuna corrente può scorrere in  $R_3$ , connessa unicamente al morsetto non invertente dell'amplificatore, e quindi tale morsetto si troverà a tensione nulla come nella configurazione senza  $R_3$ . Se invece consideriamo la presenza delle correnti di bias, allora  $R_3$  gioca un ruolo nel circuito. La corrente  $I_B$  del generatore di bias connesso a tale morsetto fluirà interamente in

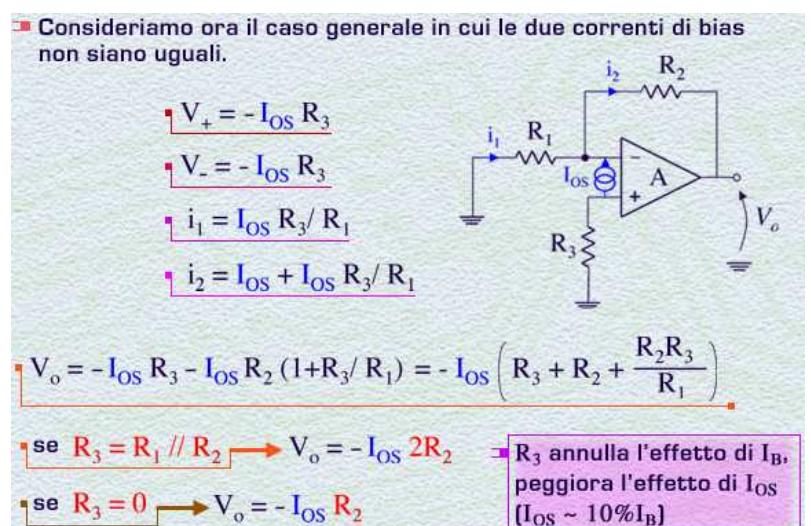


$R_3$  sviluppando sul morsetto non invertente una tensione pari a possibile determinare la corrispondente tensione all'uscita dell'amplificatore utilizzando il trasferimento da questo punto all'uscita che, si noti, non è altro che il guadagno della configurazione non invertente. La tensione di uscita vale pertanto  $-I_B * R_3$  per  $(1+R_2/R_1)$ . A questa tensione va sommata, per sovrapposizione degli effetti, quella provocata dall'altro generatore di bias  $I_B$  connesso al morsetto invertente. Come già visto in precedenza, il contributo di questo generatore alla tensione di uscita vale  $I_B * R_2$ . Se si sommano i due contributi si può notare che gli effetti provocati sulla tensione di uscita dai due generatori sono di segno opposto. Questi effetti possono quindi essere compensati tra loro tramite un'opportuna scelta del valore di  $R_3$ . La tensione di uscita risulta infatti pari a  $I_B$  che moltiplica  $(R_2 - (R_3 * (R_1+R_2)/R_1))$  e affinché questa tensione sia nulla, occorre che sia nullo il termine entro parentesi. Questo si ottiene ponendo  $R_3 = R_1 * R_2 / (R_1 + R_2)$ , cioè  $R_3$  pari al parallelo di  $R_1$  con  $R_2$ . In questa maniera gli effetti dei due generatori di bias vengono completamente compensati.

Abbiamo finora considerato la presenza di due correnti di bias di pari valore  $I_B$ . Consideriamo ora il caso più generale in cui le due correnti non siano uguali. Abbiamo visto nella lezione sull'amplificatore operazionale come questa situazione possa essere trattata supponendo la presenza di due generatori di uguale corrente  $I_B$  pari al valor medio delle due correnti di bias, e la presenza di un ulteriore generatore di corrente di offset  $I_{OS}$ , piazzato tra morsetto invertente e non invertente, pari alla differenza delle correnti di bias.

Abbiamo appena visto gli effetti provocati dai due generatori di uguale corrente  $I_B$ , analizziamo ora quale sia l'effetto provocato dal generatore di offset  $I_{OS}$ . Nel fare questa valutazione, consideriamo ancora il caso in cui sia presente la resistenza  $R_3$  ad annullare gli effetti dovuti ai generatori di corrente  $I_B$ , come visto in precedenza.

Possiamo innanzitutto osservare che la corrente  $I_{OS}$  scorre nella resistenza  $R_3$ , sviluppando su di essa una tensione pari a  $-I_{OS} * R_3$ . In virtù del corto circuito virtuale tra i due morsetti di ingresso dell'amplificatore operazionale, anche la tensione sul morsetto invertente vale  $-I_{OS} * R_3$ . Questa tensione cadrà anche tra i capi della resistenza  $R_1$  e pertanto la corrente in quest'ultima sarà pari a  $I_{OS} * R_3 / R_1$ . La corrente  $i_2$  che fluisce nella resistenza  $R_2$  risulta data dalla somma tra la corrente del generatore di offset  $I_{OS}$  e la corrente  $i_1$ , cioè è pari a  $I_{OS} + I_{OS} * R_3 / R_1$ . La tensione  $V_o$  all'uscita



dell'amplificatore risulta pari alla tensione sul morsetto invertente  $-I_{OS} \cdot R_3$ , più la caduta di tensione sulla resistenza  $R_2$ , pari a  $-I_{OS}$  per  $R_2 \cdot (1 + R_3/R_1)$ .

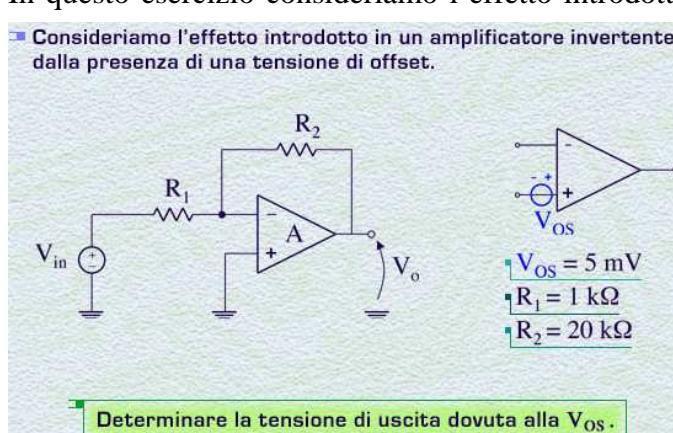
Raccogliendo i termini che moltiplicano  $I_{OS}$ , si ha che  $V_o$  risulta pari a  $-I_{OS}$  per  $(R_3 + R_2 + (R_2 \cdot R_3)/R_1)$ . Se consideriamo  $R_3$  uguale al parallelo tra  $R_1$  e  $R_2$ , condizione trovata in precedenza per annullare l'effetto dei generatori di corrente di bias  $I_B$ , la tensione di uscita risulta pari a  $-I_{OS} \cdot 2R_2$ .

E' da notare che se non si fosse introdotta la resistenza  $R_3$ , cioè ponendo  $R_3=0$ , la tensione di uscita dovuta alla presenza della corrente di offset sarebbe stata pari a  $-I_{OS} \cdot R_2$ , cioè alla metà del valore trovato con la presenza di  $R_3$ .

In conclusione, la presenza di  $R_3$ , pari al parallelo tra  $R_1$  e  $R_2$ , annulla l'effetto dovuto ai generatori di corrente  $I_B$ , ma aumenta l'effetto dovuto alla corrente di offset  $I_{OS}$ . La presenza di  $R_3$  è complessivamente da valutarsi comunque positiva, se si considera che tipicamente la corrente di offset è dell'ordine del 10% della corrente media di bias  $I_B$ , è pertanto un annullamento dell'effetto provocato da quest'ultima è da preferirsi anche se comporta una aumento dell'effetto provocato dalla corrente di offset  $I_{OS}$ .

## Tensione di offset

In questo esercizio consideriamo l'effetto introdotto in un amplificatore invertente dalla presenza di una tensione di offset.



Consideriamo l'effetto introdotto in un amplificatore invertente dalla presenza di una tensione di offset. A questo proposito, è opportuno ricordare che nella lezione sull'amplificatore operazionale era stato introdotto nel modello dell'amplificatore reale un generatore di tensione di offset che dà ragione del fatto che ponendo alla stessa tensione nulla i morsetti di ingresso di un dispositivo reale, la tensione di uscita si porta ad un valore diverso da zero.

Consideriamo quindi un amplificatore invertente in cui sia presente, ad esempio sul morsetto non invertente dell'amplificatore

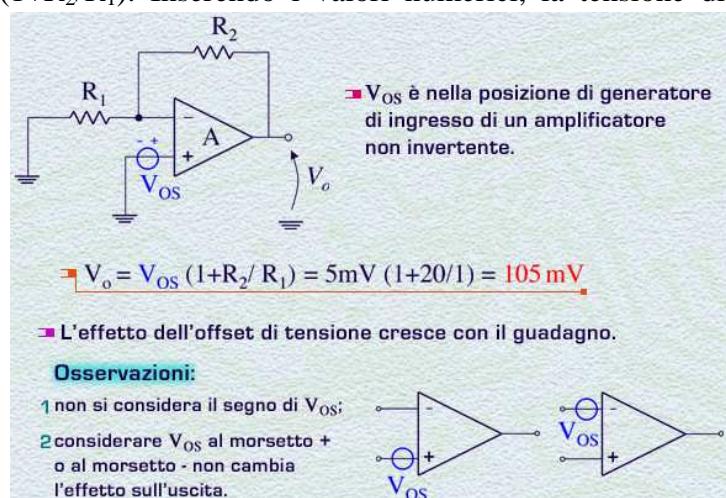
operazionale, un generatore di tensione di offset di valore pari a 5 mV.

Supponendo come valori per le resistenze 1 kΩ per  $R_1$  e 20 kΩ per  $R_2$ , determiniamo quale sia la tensione di uscita dovuta alla presenza del generatore di offset.

Anche in questo caso, al fine di valutare l'effetto del generatore di offset, spegniamo il generatore di tensione di ingresso  $V_{in}$ . Possiamo subito notare che la posizione del generatore di offset  $V_{OS}$  si trova ad essere già quella di un generatore di segnale in ingresso ad un amplificatore non-invertente.

E' quindi immediato determinare la tensione di uscita come pari alla tensione di offset  $V_{OS}$  per il guadagno dell'amplificatore non invertente  $(1+R_2/R_1)$ . Inserendo i valori numerici, la tensione di uscita risulta essere quindi essere pari a 5mV per  $(1+20)$  e cioè pari a 105mV. Si può quindi osservare che l'effetto della tensione di offset cresce proporzionalmente con il guadagno dell'amplificatore invertente  $-R_2/R_1$ , che quindi non può essere aumentato eccessivamente se la tensione in uscita dovuta all'offset di tensione deve essere mantenuta sotto un certo valore prefissato.

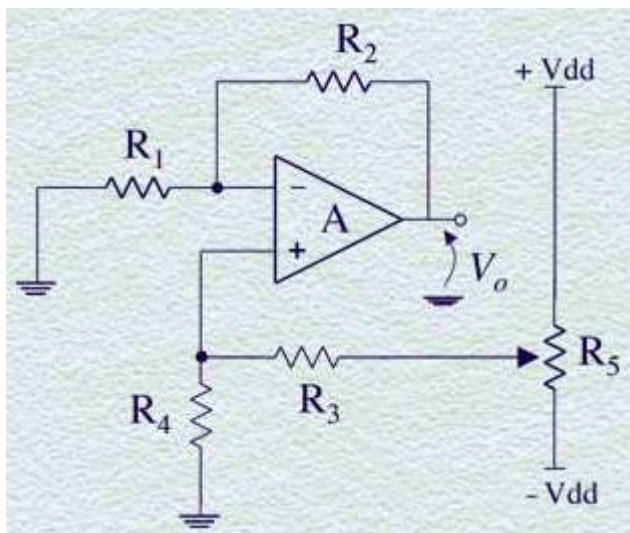
E' da notare, infine, che generalmente il valore della tensione di offset viene fornito dal costruttore in valore assoluto, senza cioè specificarne il segno. Il dare quindi un verso al generatore di offset può risultare utile nel calcolo del suo effetto in un circuito, ma di fatto il risultato conclusivo del calcolo deve essere considerato in valore assoluto. Inoltre si può facilmente verificare che l'applicare il generatore di



offset su un morsetto dell'amplificatore operazionale piuttosto che sull'altro non comporta nessuna variazione ai fini dell'effetto di questo generatore sul circuito.

Diversi amplificatori operazionali commerciali sono forniti di alcuni morsetti di servizio che consentono all'utilizzatore di compensare l'effetto della tensione di offset sull'uscita. Mediante l'applicazione di un potenziometro tra questi morsetti, risulta così possibile effettuare una regolazione attraverso i circuiti interni del dispositivo, al fine di regolare la tensione di uscita in DC dell'amplificatore fino a portarla a valori sempre più prossimi a zero.

Anche se un amplificatore operazionale non possiede questi morsetti per una compensazione interna dell'offset, è sempre possibile creare una rete esterna resistiva connessa all'amplificatore che consenta all'utilizzatore del dispositivo di compensare l'effetto della tensione di offset.



Nella figura è mostrato un esempio di rete resistiva da utilizzarsi nel caso di un amplificatore invertente. Grazie alla partizione che viene realizzata con le resistenze  $R_3$  e  $R_4$  ed il potenziometro  $R_5$ , è possibile regolare la tensione del morsetto non invertente ad una tensione DC di poco superiore o inferiore allo zero. Quando il potenziometro  $R_5$  è regolato nel suo punto medio, cioè con una resistenza pari a  $R_5/2$  verso entrambe le alimentazioni  $+V_{dd}$  e  $-V_{dd}$ , il terzo terminale, quello indicato con la freccia e connesso a  $R_3$ , si trova a zero e quindi anche la tensione al morsetto invertente, risulta di conseguenza nulla. Regolando il potenziometro  $R_5$  al di sopra e al di sotto di questo punto, è possibile portare la tensione del

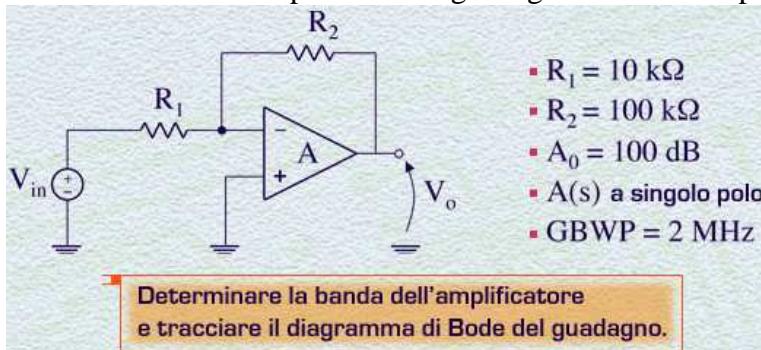
morsetto non invertente, tramite la partizione con  $R_3$  e  $R_4$  a valori in DC positivi e negativi. In questa maniera, l'eventuale offset presente in uscita può essere compensato mediante una opportuna regolazione della tensione del morsetto non invertente che comporta una modifica della tensione in continua in uscita del circuito, per mezzo del guadagno non invertente dello stesso.

La presenza della rete introdotta per la compensazione non ha invece alcun effetto sul guadagno dell'amplificatore che, dal punto di vista del segnale, è utilizzato nella configurazione non invertente.

## Lezione n. 31 : Esempi ed esercizi B

### Risposta in frequenza di un amplificatore invertente

Consideriamo in questo esercizio un amplificatore invertente con le seguenti caratteristiche. Le resistenze  $R_1$  e  $R_2$  hanno valore rispettivamente di  $10\text{ k}\Omega$  e  $100\text{ k}\Omega$ . L'amplificatore operazionale utilizzato nel circuito possiede un guadagno in continua pari a  $100\text{ dB}$ . La sua risposta in frequenza è



inoltre caratterizzata dalla presenza di un singolo polo. Dal costruttore viene fornito il prodotto guadagno-banda dell'amplificatore, definito come GBWP dall'inglese Gain-BandWidth-Product, che nel caso specifico vale  $2\text{ MHz}$ .

Noti questi dati sul circuito, ci si pone l'obiettivo di determinarne la banda, nonché di tracciare il Diagramma di Bode del suo guadagno.

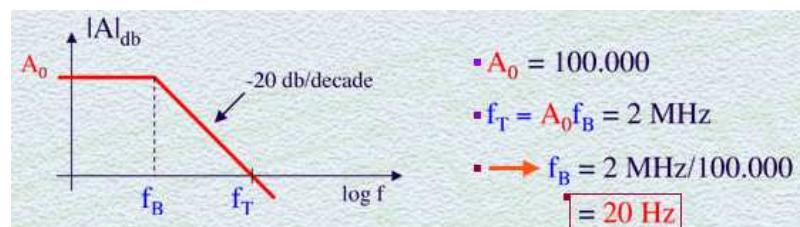
Nella figura viene riportato il diagramma di Bode della operazionale, in cui viene mostrata la presenza del singolo polo, la banda del circuito ad anello chiuso, si estende rispetto a quella dell'amplificatore operazionale di un fattore pari al modulo del guadagno d'anello, cioè  $-G_{loop}$ . L'espressione del modulo del guadagno d'anello è già nota dalle lezioni precedenti è vale  $A_0$  per  $R_1$ , diviso  $(R_1+R_2)$ . Inserendo i valori del guadagno  $A_0$  e delle resistenze, il modulo del guadagno d'anello  $-G_{loop}$  vale 9090.

Abbiamo visto in una lezione precedente, che per l'amplificatore invertente come per quello non invertente, se l'amplificatore operazionale possiede un singolo polo, la banda del circuito ad anello chiuso, si estende rispetto a quella dell'amplificatore operazionale di un fattore pari al modulo del guadagno d'anello, cioè  $-G_{loop}$ . L'espressione del modulo del guadagno d'anello è già nota dalle lezioni precedenti è vale  $A_0$  per  $R_1$ , diviso  $(R_1+R_2)$ . Inserendo i valori del guadagno  $A_0$  e delle resistenze, il modulo del guadagno d'anello  $-G_{loop}$  vale 9090.

Possiamo così determinare il valore del polo ad anello chiuso, che risulta uguale alla banda dell'amplificatore operazionale, pari a  $20\text{ Hz}$ , moltiplicata per 9090. La banda ad anello chiuso risulta così pari a circa  $180\text{ kHz}$ .

Determiniamo ora il prodotto guadagno-banda dell'amplificatore invertente. Il modulo del guadagno è pari al rapporto tra  $R_2$  e  $R_1$ , che vale 10 o, in decibel,  $20\text{ dB}$ . Il prodotto guadagno banda ad anello chiuso è pari pertanto a 10 per  $180\text{ kHz}$ , in cui  $180\text{ kHz}$  è la banda precedentemente determinata, e risulta quindi uguale a  $1.8\text{ MHz}$ .

Possiamo notare come questo valore sia molto prossimo a  $2\text{ MHz}$ , che era il valore del prodotto guadagno-banda dell'amplificatore operazionale utilizzato. Si dimostra così sostanzialmente verificata la proprietà che avevamo già presentato per questo tipo di amplificatori, nel caso in cui il guadagno sia sensibilmente superiore ad uno, e cioè che il prodotto guadagno banda si conserva.



**f<sub>1</sub> = f<sub>B</sub> (-G<sub>loop</sub>)** La banda ad anello chiuso è estesa di un fattore pari al modulo del G<sub>loop</sub>

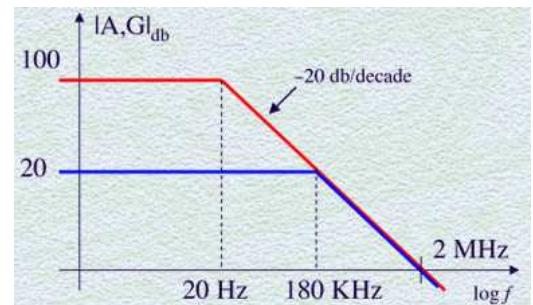
$$-G_{loop} = \frac{A_0 R_1}{R_1 + R_2} = 9090$$

$$f_1 = 20\text{ Hz} \times 9090 \approx 180\text{ kHz}$$

$ G  = \frac{R_2}{R_1} = 10$	$GBWP_{anello\ chiuso} = 10 \times 180\text{ kHz} = 1.8\text{ MHz}$
$\text{anello chiuso}$	$\approx 2\text{ MHz}$ $GBWP_{anello\ aperto}$
	Il prodotto guadagno-banda si conserva

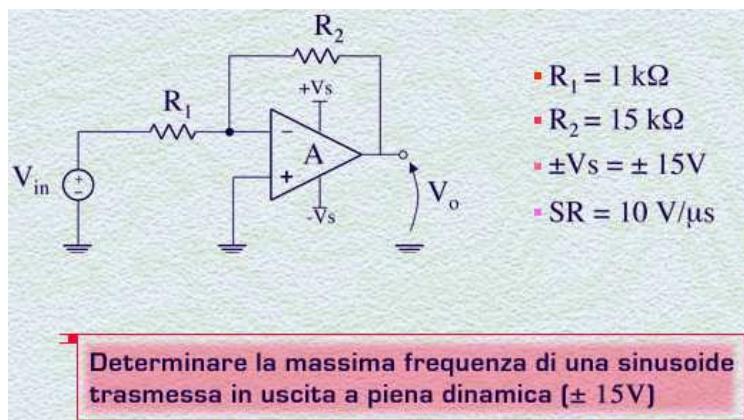
valore del prodotto guadagno-banda dell'amplificatore operazionale utilizzato. Si dimostra così sostanzialmente verificata la proprietà che avevamo già presentato per questo tipo di amplificatori, nel caso in cui il guadagno sia sensibilmente superiore ad uno, e cioè che il prodotto guadagno banda si conserva.

Nel grafico è riportato il diagramma di Bode del modulo del guadagno dell'amplificatore, in cui sono indicati i valori caratteristici del guadagno e della banda precedentemente determinati. Per confronto, anche il diagramma di Bode dell'amplificatore operazionale, è riportato, non in scala, sullo stesso grafico.



### parte A: Slew rate.

In questo esercizio valuteremo con alcuni esempi, la capacità di un amplificatore invertente di amplificare e rendere disponibili in uscita con la massima dinamica disponibile segnali applicati al suo ingresso.



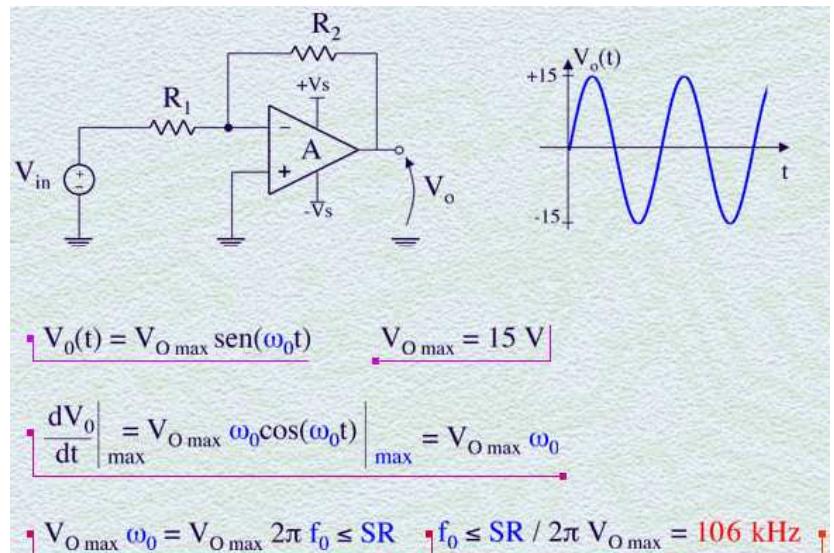
dinamica di uscita del circuito, cioè il massimo e minimo valore che può avere un segnale all'uscita del circuito. Infine consideriamo che la risposta al grande segnale dell'amplificatore sia limitata in uscita da uno Slew-Rate dell'amplificatore operazionale di valore pari a  $10 \text{ V}/\mu\text{s}$ .

In queste condizioni ci poniamo l'obiettivo di determinare la massima frequenza di una sinusoide applicata in ingresso che possa essere trasmessa nella piena dinamica di uscita, cioè  $\pm 15\text{V}$ , senza distorsioni a causa dello slew-rate.

Nella figura viene mostrata la sinusoide che vogliamo ottenere correttamente all'uscita del circuito nella piena dinamica. Questa sinusoide può essere descritta come  $V_o(t) = V_{0\max} \sin(\omega_0 t)$ . In cui  $V_{0\max}$  rappresenta la massima ampiezza della sinusoide ed è posta pari a  $15\text{V}$ . Per valutare la massima pendenza assunta dalla sinusoide, determiniamo il massimo della sua derivata nel tempo, che vuol dire il massimo valore di  $V_{0\max} * \omega_0 * \cos(\omega_0 t)$ . Questo risulta pari a  $V_{0\max} * \omega_0$ .

Per determinare la massima frequenza della sinusoide trasmessa senza distorsione, imponiamo che il massimo valore della pendenza appena determinato sia inferiore al valore dello *slew-rate* dell'amplificatore operazionale, cioè che  $V_{0\max} * \omega_0$ , che è pari a  $V_{0\max} * 2\pi f_0$  sia inferiore al valore dello slew-rate, diviso  $2\pi * V_{0\max}$ . Inserendo i valori numerici,  $10\text{V}/\mu\text{s}$  per lo slew-rate e  $15\text{V}$  per  $V_{0\max}$ , otteniamo che la massima frequenza della sinusoide che può essere trasmessa compatibilmente con lo slew-rate è di  $106 \text{ kHz}$ .

Per ottenere la corrispondente forma d'onda della sinusoide in ingresso, è sufficiente dividere l'espressione di  $V_o(t)$  così ricavata per il guadagno dell'amplificatore invertente che è pari a  $-R_2/R_1$  uguale a  $-15$ .



## parte B: Minimo carico resistivo in uscita.

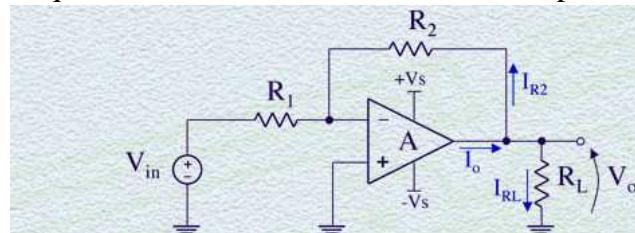
Consideriamo ora lo stesso circuito visto in precedenza ma per esaminare un diverso problema. Supponiamo che all'uscita dell'amplificatore invertente sia collegato un carico resistivo  $R_L$  e supponiamo di voler avere su questo carico, a seguito dell'applicazione di un segnale in ingresso, il massimo e minimo segnale di tensione disponibile nella dinamica in uscita, cioè  $\pm 15V$ .

A limitare questa possibilità nei casi pratici può intervenire un'ulteriore parametro dell'amplificatore operazionale reale: la sua limitata capacità a fornire un'elevata corrente in uscita  $I_o$ . Consideriamo, nel nostro esempio, un amplificatore operazionale che possa fornire in uscita, sia in un verso che nell'altro, una corrente massima  $I_{o\max}$  pari a 20 mA.

Ora, se consideriamo di applicare una medesima tensione sul carico  $R_L$ , nel nostro caso i valori estremi della dinamica in uscita  $\pm 15V$ , minore sarà il valore di  $R_L$ , maggiore sarà il valore della corrente che dovrà fluire in essa a pari tensione. Pertanto, considerando la limitazione di corrente in uscita dell'amplificatore operazionale, ci poniamo l'obiettivo di determinare il minimo valore del carico  $R_L$  che può essere utilizzato con questo circuito.

Valutiamo questo limite nel caso di una tensione positiva  $V_o$  sul carico  $R_L$ , e quindi, corrispondentemente, per una corrente  $I_o$  uscente dall'amplificatore. Nel caso di tensione negativa su  $R_L$ , la trattazione è completamente analoga.

In questo caso, la corrente uscente dell'amplificatore operazionale  $I_o$  è pari alla corrente che scorre nel carico  $I_{RL}$  più la corrente che scorre nella resistenza  $R_2$ . La corrente  $I_o$  risulta quindi pari a  $V_o/R_L + V_o/R_2$  e quindi pari a  $V_o$  diviso per il parallelo tra  $R_L$  e  $R_2$ . Questa corrente deve essere minore della massima corrente disponibile all'uscita all'amplificatore operazionale,  $I_{o\max}$ .



$$I_o = I_{RL} + I_{R2} = V_o/R_L + V_o/R_2 = V_o/(R_L//R_2) < I_{o\max}$$

$$\frac{V_o}{R_L + R_2} < I_{o\max}$$

$$R_L(I_{o\max} R_2 - V_o) > V_o R_2$$

$$R_L > \frac{V_o R_2}{I_{o\max} R_2 - V_o} = 789 \Omega$$

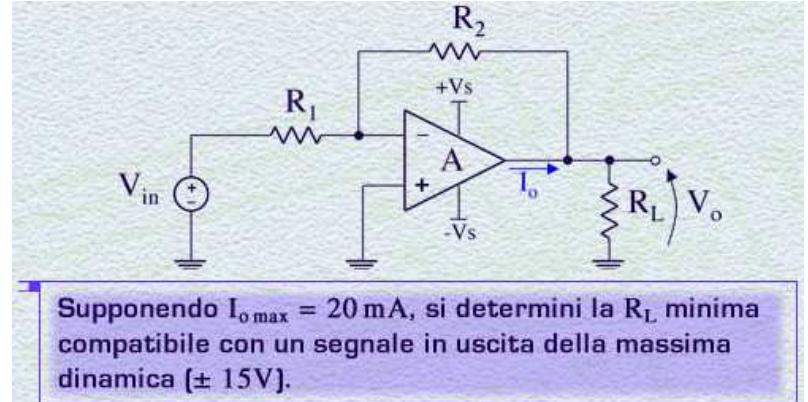
**Osservazione:**  $I_{o\max} R_2 - V_o > 0$     $R_2 > V_o/I_{o\max}$     $(I_{o\max} > I_{R2})$

deve essere maggiore di  $V_o * R_2 / (I_{o\max} * R_2 - V_o)$ .

Inserendo i valori numerici, 15V per  $V_o$ , la massima tensione che vogliamo applicare sul carico, 15 kΩ per  $R_2$  e 20 mA per  $I_{o\max}$ , si ottiene un valore minimo per  $R_L$  pari a 789 Ω.

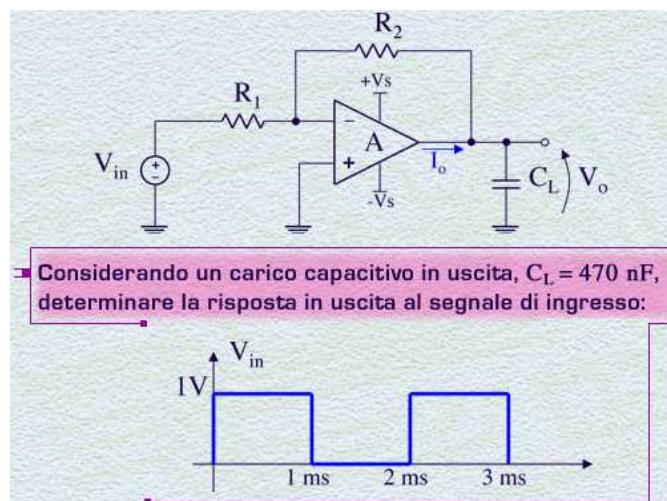
Prima di concludere questa parte dell'esercizio, è opportuno osservare che affinché la relazione matematica appena determinata per  $R_L$  abbia senso fisico, e cioè che  $R_L$  sia positiva, occorre che il denominatore dell'espressione,  $(I_{o\max} * R_2 - V_o)$ , sia positivo. Ciò impone quindi che  $R_2$  debba essere maggiore di  $V_o/I_{o\max}$ . Questa condizione impone quindi un limite inferiore al valore da scegliere per la resistenza  $R_2$ . Il significato fisico di questa condizione è intuitivo.

Occorre infatti che l'amplificatore operazionale, che è limitato nella sua corrente di uscita, riesca almeno a fornire la corrente che deve scorrere nella resistenza  $R_2$ , cioè la corrente  $I_{o\max}$  deve essere superiore a  $I_{R2}$ . Dato che la tensione sull'uscita vale  $V_o$ , la corrente  $I_{R2}$  risulta pari a  $V_o/R_2$  e quindi  $I_{o\max}$  deve essere maggiore di  $V_o/R_2$  o  $R_2$  deve essere maggiore di  $V_o/I_{o\max}$ , che è la condizione matematica prima indicata.



### parte C: Carico capacitivo in uscita.

Nell'ultima parte di questo esercizio, consideriamo invece la situazione in cui il nostro amplificatore invertente debba applicare il suo segnale di uscita su un carico capacitivo.



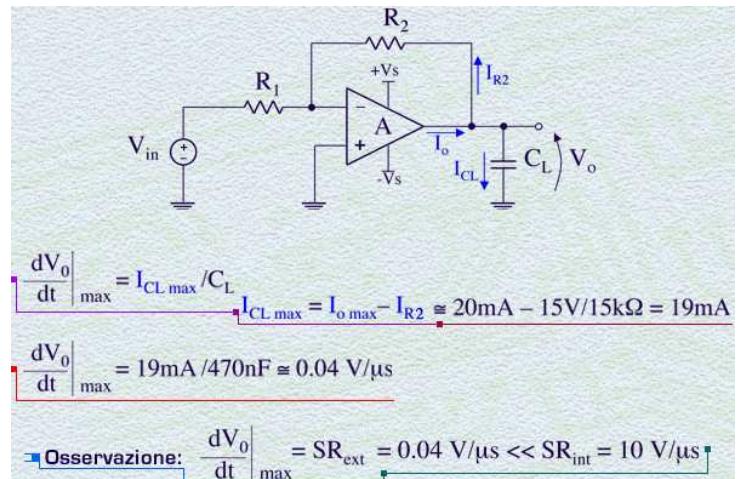
di ingresso, consideriamo innanzitutto la massima pendenza con cui la tensione può variare in uscita all'amplificatore. Essendoci sull'uscita un carico capacitivo, la massima pendenza della tensione in uscita risulta limitata dal massimo valore di corrente erogabile dall'amplificatore. Supponendo che la massima corrente disponibile per la carica della capacità  $C_L$  sia pari a  $I_{CLmax}$ , risulta evidente, come in un tipico caso di corrente costante integrata su di una capacità, che la tensione sulla capacità varierà con un andamento a rampa la cui pendenza, ch'è la derivata di  $V_o$  rispetto al tempo, è pari a  $I_{CLmax}$  diviso il valore della capacità  $C_L$ . La massima corrente  $I_{CLmax}$  che può scorrere nella capacità  $C_L$  risulta pari alla differenza tra la massima corrente erogata all'uscita dell'amplificatore operazionale,  $I_{o max}$ , e la corrente  $I_{R2}$  che scorre nella resistenza  $R_2$ . Consideriamo, per semplicità, una corrente  $I_{R2}$  costante e pari al suo valore massimo, che è uguale alla massima tensione in uscita  $V_{o max}$  divisa  $R_2$ . In verità questa corrente varia con il variare della tensione in uscita, ma l'approssimazione fatta consente di giungere velocemente ad una stima del valore della corrente  $I_{CLmax}$ , comunque per difetto.

Inserendo i valori numerici, la corrente  $I_{CLmax}$  risulta quindi pari a 20 mA meno  $15V/15k\Omega$  e quindi pari a 19 mA. Da questo risultato si può concludere, peraltro, che il contributo della corrente in  $R_2$  è comunque praticamente trascurabile rispetto alla massima corrente erogata  $I_{o max}$ . Nota la corrente  $I_{CLmax}$ , è possibile ora determinare la massima pendenza del segnale in uscita che vale  $19 \text{ mA}/470 \text{ nF}$  pari a circa  $0.04 \text{ V}/\mu\text{s}$ .

E' opportuno notare che il fenomeno appena descritto, cioè la carica della capacità  $C_L$  con andamento della tensione a rampa, risulta avere una certa rassomiglianza con il fenomeno dello slew-rate che caratterizza l'andamento dell'uscita di un amplificatore operazionale per elevati segnali. Anche per lo slew-rate, infatti, si ipotizza una variazione a rampa della tensione di uscita, con una pendenza costante espressa anch'essa in  $\text{V}/\mu\text{s}$ . Data questa analogia, ha senso quindi definire la limitazione nella velocità di carica della capacità  $C_L$  come uno Slew-rate esterno, che chiamiamo  $S_{Rext}$ , da confrontare con lo slew-rate proprio dell'amplificatore operazionale, che chiamiamo invece slew-rate interno o  $S_{Rint}$ . Se confrontiamo numericamente gli effetti dei due fenomeni, possiamo notare che il valore dello slew-rate esterno, che è pari a  $0.04 \text{ V}/\mu\text{s}$ , è notevolmente inferiore a quello dello slew-rate interno, che ricordiamo vale  $10 \text{ V}/\mu\text{s}$ . La velocità di risposta dell'uscita su grande segnale è pertanto limitata dallo slew-rate esterno, cioè quello da carico capacitivo sull'uscita.

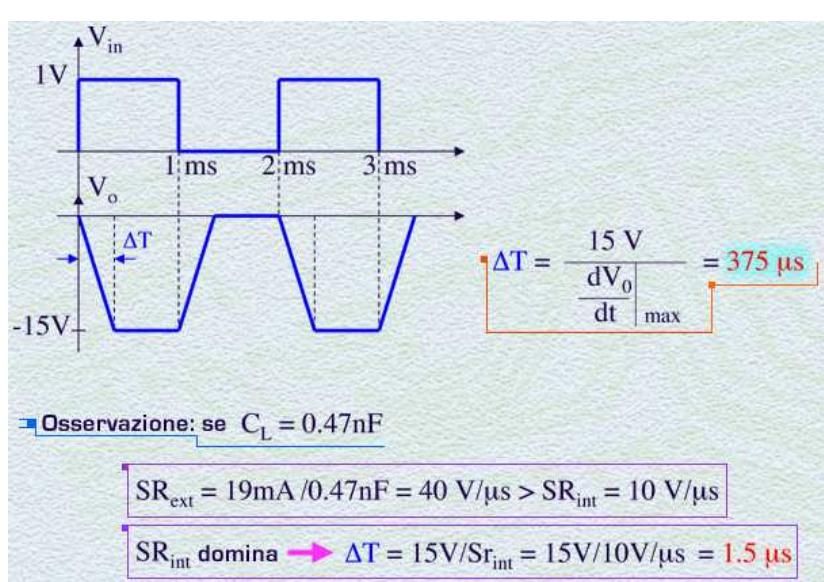
Consideriamo un caso specifico in cui all'uscita dell'amplificatore si collega una capacità di valore pari a 470 nF. Consideriamo l'amplificatore operazionale sempre limitato nella sua capacità di erogare corrente all'uscita ad un valore massimo di 20 mA. Ci poniamo l'obiettivo di determinare la risposta in uscita dell'amplificatore nel caso in cui all'ingresso sia applicato un segnale come quello mostrato in figura, consistente in un'onda quadra di ampiezza 1 V e periodo pari a 2 ms.

Per determinare la risposta in uscita al segnale



Concludiamo l'esercizio tracciando l'andamento della tensione in uscita in funzione dell'andamento dell'onda quadra applicata in ingresso. Osserviamo che in corrispondenza del fronte positivo del

segnale in ingresso al tempo  $t=0$ , l'uscita tenderà a portarsi a valori negativi fino a  $-15V$  con un andamento a rampa la cui pendenza è data dallo slew-rate esterno. L'intervallo di tempo  $\Delta T$  con cui l'uscita si porta a  $-15V$  può essere calcolato come rapporto tra  $15 V$  e la pendenza  $dv/dt$  pari al valore dello slew-rate esterno. Tale intervallo risulta pari a  $375\mu s$ . Analogamente a quanto succede al tempo  $t=0$ , al tempo  $t=1ms$  il segnale in ingresso si riporta a 0 e quindi l'uscita inizierà a salire verso il valore zero che raggiungerà dopo un tempo pari



ancora a  $375\mu s$ . L'andamento visto si ripeterà sul successivo fronte di salita al tempo  $t=2ms$  e così via per gli altri fronti dell'onda quadra.

Si può notare come il tempo  $\Delta T$  determinato, rappresenti un ritardo con cui l'uscita raggiunge il valore nominale, rispetto al tempo in cui è applicato il fronte dell'onda quadra in ingresso.

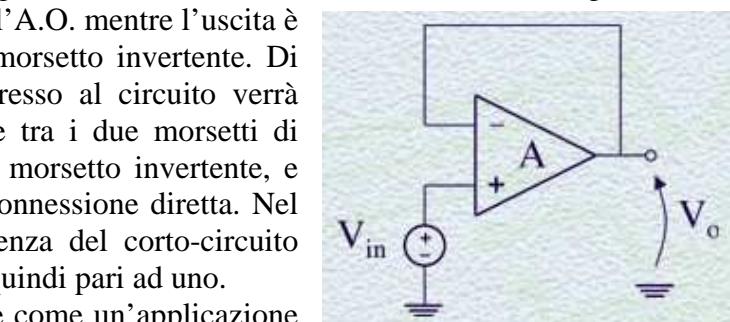
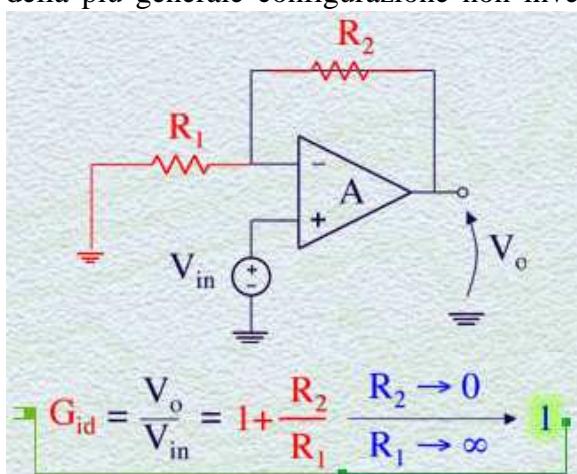
Nel caso in cui il carico capacitivo abbia un valore 1.000 volte inferiore rispetto al valore considerato, cioè di  $0.47 nF$ , si può notare che il valore dello slew-rate esterno risulta pari a  $19mA/0.47nF$  e cioè a  $40V/\mu s$ . Questo valore risulta questa volta superiore a quello dello slew-rate interno, pari a  $10V/\mu s$ . Lo slew-rate interno sarà quindi il fenomeno limitante la velocità di risposta dell'uscita, la quale varierà con una pendenza pari  $10V/\mu s$ . Analogamente a quanto fatto in precedenza, è possibile calcolare il tempo con cui la tensione in uscita raggiunge l'ampiezza nominale, che è pari a  $15V/(10V/\mu s)$ . Questo intervallo di tempo, molto più corto rispetto al caso precedente, risulta pari a  $1.5 \mu s$ .

## Lezione n. 32 : Altre configurazioni con l'Amplificatore Operazionale

### Il buffer

Esaminiamo un circuito molto utilizzato in elettronica: il **buffer non invertente di tensione**. Lo schema di questo semplice circuito è mostrato nella figura. Possiamo osservare che la tensione di ingresso  $V_{in}$  viene applicata al morsetto non invertente dell'A.O. mentre l'uscita è collegata in contoreazione direttamente al morsetto invertente. Di conseguenza, una tensione applicata in ingresso al circuito verrà trasferita, in virtù del corto circuito virtuale tra i due morsetti di ingresso dell'amplificatore operazionale, sul morsetto invertente, e da questo all'uscita del circuito, tramite la connessione diretta. Nel caso di contoreazione ideale, cioè in presenza del corto-circuito virtuale, il guadagno di tensione del buffer è quindi pari ad uno.

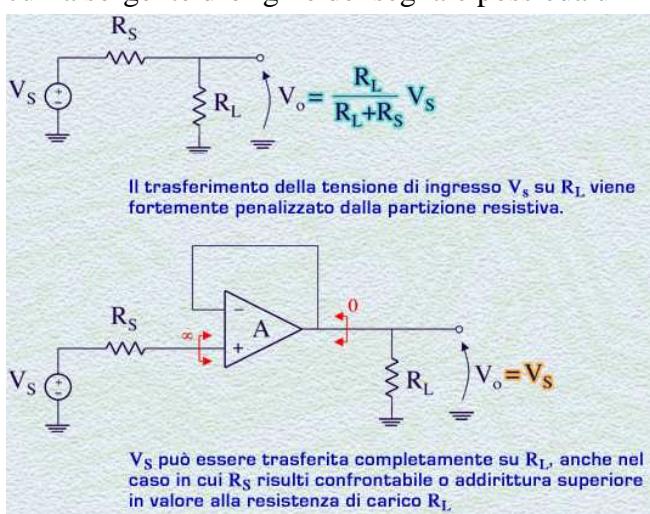
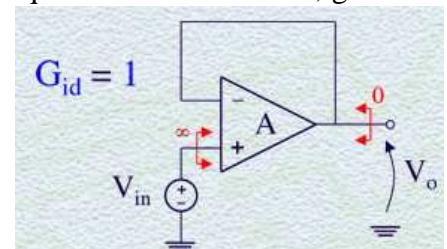
Questa configurazione può essere vista anche come un'applicazione della più generale configurazione non invertente, che viene mostrata nella figura inserendo le due resistenze  $R_1$  e  $R_2$ .



$$R_1 \rightarrow \infty \quad R_2 \rightarrow 0$$

Possiamo infatti osservare che graficamente il circuito a Buffer può essere ottenuto dalla configurazione invertente facendo tendere all'infinito il valore di  $R_1$ , cioè sostituendo a questa resistenza un circuito aperto, e facendo tendere a zero il valore di  $R_2$ , cioè sostituendo a questa resistenza un corto-circuito. Se ora consideriamo l'espressione del guadagno dell'amplificatore non invertente, pari a  $(1 + R_2/R_1)$ , si può notare che facendo tendere rispettivamente  $R_1$  all'infinito e  $R_2$  a zero si ottiene effettivamente che il guadagno del circuito tende ad essere pari ad uno, come avevamo dedotto in precedenza.

Il Buffer è quindi un circuito che possiede un guadagno di tensione pari a uno e che viene utilizzato per le sue importanti proprietà riguardanti l'impedenza di ingresso e quella di uscita. Infatti, grazie al fatto che l'ingresso viene applicato direttamente al morsetto non invertente dell'A.O., l'impedenza di ingresso di questo circuito risulta essere elevatissima, al limite infinita nel caso di un A.O. ideale. Se invece consideriamo l'impedenza di uscita del circuito, questa risulta essere bassissima, al limite nulla nel caso di un A.O. ideale. Per queste caratteristiche, il Buffer viene utilizzato come "**trasformatore di impedenza**", nel senso può essere introdotto in un circuito per applicare su di un carico un segnale con bassa impedenza di sorgente, anche nel caso in cui la sorgente d'origine del segnale possieda un'impedenza di uscita elevata.



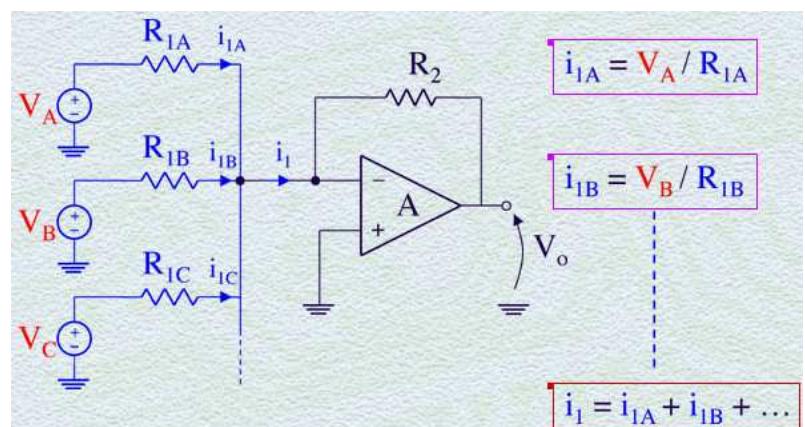
Vediamo di illustrare meglio questo impiego caratteristico del Buffer attraverso un esempio. Si supponga di voler applicare su un carico  $R_L$  la tensione fornita da una sorgente  $V_s$  dotata di una resistenza di sorgente  $R_s$ . In virtù della partizione resistiva che viene operata tramite  $R_s$  e  $R_L$ , il valore di tensione trasmesso sul carico  $R_L$  risulta pari a  $V_s$  ridotto per il fattore  $R_L/(R_L + R_s)$ . Nel caso in cui la resistenza di sorgente abbia valore prossimo a  $R_L$ , il trasferimento della tensione di ingresso  $V_s$  su  $R_L$  viene fortemente penalizzato dalla partizione resistiva. Si consideri ora lo stesso problema ma nel caso in cui tra  $R_s$  e  $R_L$  venga interposto un Buffer di tensione ideale. La

presenza del Buffer ha come effetto quello di disaccoppiare le due resistenze  $R_S$  ed  $R_L$ . Per l'effetto dell'impedenza d'ingresso idealmente infinita del Buffer, il segnale d'ingresso  $V_S$  viene trasmesso interamente all'ingresso non invertente dell'A.O., in quanto non scorre corrente su  $R_S$ . Tale segnale viene fornito inalterato all'uscita del Buffer in virtù del guadagno unitario di quest'ultimo. Infine, dato che il Buffer, in virtù della sua impedenza nulla di uscita, si comporta come un generatore ideale di tensione, il valore della tensione  $V_S$  viene applicato direttamente sulla resistenza  $R_L$ . In questa maniera la tensione di ingresso  $V_S$  può essere trasferita completamente su  $R_L$ , anche nel caso in cui  $R_S$  risulti confrontabile o addirittura superiore in valore alla resistenza di carico  $R_L$ .

## L'amplificatore Sommatore

Consideriamo come ulteriore applicazione della configurazione invertente un circuito in cui siano presenti più segnali di ingresso di cui sia necessario effettuare la somma. Per tale ragione, il circuito che ci apprestiamo a studiare prende il nome di amplificatore sommatore.

Lo schema di questo circuito è mostrato in figura. Come si può notare, esso è molto simile all'amplificatore invertente, con la differenza che in ingresso sono presenti non uno ma diversi generatori di tensione  $V_A$ ,  $V_B$ ,  $V_C$  e così via. I generatori sono tutti connessi ad un unico nodo, il morsetto invertente dell'amplificatore operazionale, ciascuno tramite una propria resistenza,  $R_{1A}$  per  $V_A$ ,  $R_{1B}$  per  $V_B$ ,  $R_{1C}$  per  $V_C$ , e così via. Per quanto riguarda la parte rimanente dell'amplificatore, questa consiste nella solita configurazione con la resistenza  $R_2$  in contoreazione.



Analizziamo ora il funzionamento del circuito. Per quanto già visto per l'amplificatore invertente, per l'effetto della contoreazione il morsetto invertente si trova ad essere una terra virtuale. Di conseguenza la tensione ai capi di ogni singola resistenza, ad esempio  $R_{1A}$ , sarà pari alla corrispondente tensione di ingresso, in questo caso  $V_A$ . Ciò consente di determinare il valore delle singole correnti,  $i_{1A}$  come  $V_A/R_{1A}$ ,  $i_{1B}$  come  $V_B/R_{1B}$  e così via.

$$V_o = -R_2 (1/R_{1A} V_A + 1/R_{1B} V_B + 1/R_{1C} V_C \dots)$$

↓

$$V_o = \sum_j [a_j V_j]$$

effettua la somma pesata (e invertente) degli ingressi

aj forniscono i pesi della somma

Se

$$R_{1A} = R_{1B} = R_{1C} = \dots$$

↓

$$V_o = a \sum_j V_j$$

somma (invertente) degli ingressi

La corrente totale  $i_1$  che fluisce nella terra virtuale è pari alla somma di queste correnti. La terra virtuale, quindi, si comporta come un pozzo di corrente, cioè assorbe e somma le correnti provenienti dai vari rami di ingresso al circuito.

Questa corrente, come già noto per la configurazione invertente, fluisce interamente nella resistenza  $R_2$ , e ci consente così di determinare la tensione di uscita  $V_o$  dell'amplificatore tramite la caduta di tensione su  $R_2$ . Giungiamo così all'espressione della tensione  $V_o$  che risulta pari a:  $-R_2 (1/R_{1A} \cdot V_A + 1/R_{1B} \cdot V_B + 1/R_{1C} \cdot V_C \dots)$ , sommando cioè i contributi di tutti i generatori di tensione applicati agli ingressi del circuito.

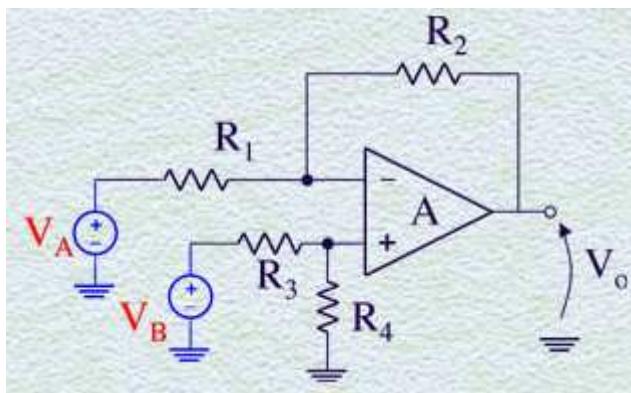
E' semplice a questo punto osservare come la funzione di questo circuito sia quella di effettuare una somma delle tensioni  $V_A$ ,  $V_B$ ,  $V_C$ , eccetera applicate agli ingressi del circuito. Questa operazione risulta invertente in quanto i contributi vengono sommati con lo stesso segno, ma il risultato della somma viene invertito di segno. Inoltre in questa operazione di somma è possibile dare un diverso peso alle varie tensioni di ingresso da sommare, utilizzando valori opportuni per le resistenze  $R_{1A}$ ,  $R_{1B}$ , eccetera. La somma risulta essere quindi una somma pesata.

Se le resistenze vengono scelte tutte dello stesso valore, cioè  $R_{1A} = R_{1B} = R_{1C}$  e così via, allora la tensione in uscita del circuito risulta come una vera somma (sempre invertente) delle tensioni in ingresso, cioè tutti gli ingressi vengono sommati con lo stesso peso.

### L'amplificatore Differenziale

Abbiamo visto che le configurazioni invertente e non invertente amplificano e riportano in uscita la tensione applicata in ingresso rispettivamente invertita e non invertita di segno. Vediamo ora un circuito basato sulla **presenza contemporanea delle configurazioni invertente e non invertente allo scopo di effettuare l'operazione di differenza tra due tensioni applicate in ingresso**. Si tratta per l'appunto dell'**amplificatore differenziale**.

Nella figura è mostrato lo schema del circuito. Si può notare che il generatore di tensione  $V_A$  è applicato all'ingresso del circuito nella configurazione invertente, mentre il generatore  $V_B$  è applicato al morsetto non-invertente, tramite la partizione tra le resistenze  $R_3$  e  $R_4$ .

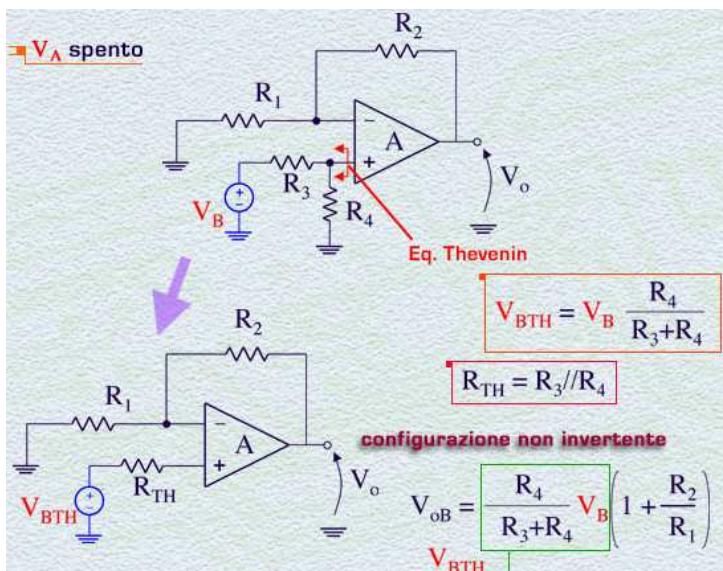


Determiniamo la tensione in uscita al circuito in seguito all'applicazione dei due generatori in ingresso, applicando il principio di sovrapposizione degli effetti.

Iniziamo considerando il solo effetto del generatore  $V_A$ , spegnendo il generatore  $V_B$ . Se il generatore  $V_B$  è spento, non vi è alcuna ragione perché della corrente circoli nella maglia formata dalle resistenze  $R_3$  e  $R_4$ . Pertanto, il morsetto non

invertente dell'amplificatore operazionale si troverà a tensione nulla. Osserviamo a questo punto che il circuito così ottenuto è di fatto un amplificatore invertente per quanto riguarda il generatore di ingresso  $V_A$ . Di conseguenza risulta immediato determinare la tensione di uscita relativa a questo generatore che vale:  $V_{oA} = -V_A \cdot R_2 / R_1$ , dove si è utilizzato il guadagno della configurazione invertente pari a  $-R_2/R_1$ .

Determiniamo ora il solo effetto del generatore  $V_B$ , spegnendo il generatore  $V_A$ . Risulta pratico a questo scopo determinare il circuito equivalente Thevenin al morsetto non invertente del circuito,



$$\boxed{V_B \text{ spento} \rightarrow V_+ = 0 \text{ configurazione invertente}}$$

$$\boxed{V_{oA} = -R_2 / R_1 V_A}$$

come mostrato nella figura. Il generatore di tensione Thevenin si calcola come la tensione riportata sul morsetto non invertente per effetto della partizione della tensione  $V_B$  effettuata da  $R_3$  e  $R_4$ .

Questa tensione  $V_{BTH}$  vale  $V_B \cdot R_4 / (R_3 + R_4)$ . La resistenza equivalente Thevenin  $R_{TH}$  risulta pari al parallelo tra le resistenze  $R_3$  e  $R_4$ . Nella figura è mostrato il circuito ottenuto sostituendo alla maglia formata da  $V_B$ ,  $R_3$  e  $R_4$  il circuito equivalente Thevenin appena determinato. La resistenza  $R_{TH}$  può essere trascurata in quanto in essa non può scorrere corrente, poiché connessa al morsetto non invertente che non assorbe corrente. Il circuito così ottenuto risulta

quindi un amplificatore non invertente rispetto al generatore di ingresso  $V_{BTH}$ . Risulta semplice così determinare la tensione di uscita  $V_{oB}$  dovuta a questo generatore che risulta uguale alla tensione  $V_{BTH}$ , cioè a  $V_B \cdot R_4 / (R_3 + R_4)$ , moltiplicata per il guadagno della configurazione non invertente  $(1 + R_2 / R_1)$ .

Sommando gli effetti dei due generatori in ingresso, la tensione in uscita  $V_o$  risulta pari a  $-(R_2 / R_1) \cdot V_A + R_4 / (R_3 + R_4) \cdot (1 + R_2 / R_1) \cdot V_B$ . Raccogliendo la frazione  $R_4 / R_3$  nel secondo termine, la tensione  $V_o$  risulta uguale a  $-(R_2 / R_1) \cdot V_A + (R_4 / R_3) \cdot [(1 + R_2 / R_1) / (1 + R_4 / R_3)] \cdot V_B$ . Se ora scegliamo i

valori delle resistenze in maniera tale che il rapporto tra  $R_4$  e  $R_3$  sia uguale al rapporto tra  $R_2$  e  $R_1$ , allora la tensione di uscita  $V_o$  risulta pari a  $-(R_2/R_1) \cdot V_A + (R_2/R_1) \cdot V_B$ , in quanto numeratore e denominatore della frazione al secondo termine si semplificano.

Giungiamo così al risultato conclusivo, in cui la tensione di uscita  $V_o$  risulta pari alla differenza delle tensioni applicate in ingresso ( $V_B - V_A$ ) moltiplicata per un fattore di amplificazione  $R_2/R_1$ .

Abbiamo così ottenuto un circuito che amplifica solo la differenza dei segnali applicati al suo ingresso, da cui il nome di amplificatore differenziale. Naturalmente se si vuole ottenere in uscita la differenza delle tensioni applicate in ingresso senza il fattore di amplificazione, è sufficiente scegliere le resistenze  $R_2$  e  $R_1$  dello stesso valore.

**■ comando gli effetti:**

$$V_o = -R_2/R_1 V_A + \frac{R_4}{R_3+R_4} \left( 1 + \frac{R_2}{R_1} \right) V_B$$

$$V_o = -R_2/R_1 V_A + R_4/R_3 \frac{1+R_2/R_1}{1+R_4/R_3} V_B$$

**■ ponendo**  $R_4/R_3 = R_2/R_1$  **■**  $V_o = -R_2/R_1 V_A + R_4/R_3 \frac{1+R_2/R_1}{1+R_4/R_3} V_B$

$\downarrow R_2/R_1$

**■**  $V_o = R_2/R_1 (V_B - V_A)$  **effettua la differenza dei segnali in ingresso**

**■ se**  $R_2 = R_1$  **■**  $V_o = V_B - V_A$

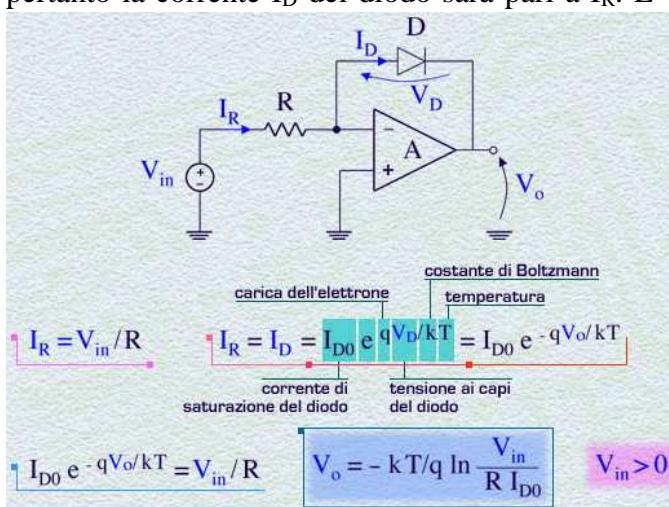
### L'amplificatore Logaritmico e L'amplificatore Esponenziale

Consideriamo un altro tipo di amplificatore, chiamato comunemente **amplificatore logaritmico**, ottenuto dall'amplificatore invertente sostituendo alla resistenza in reazione un diodo, orientato come mostrato in figura.

Per comprendere il funzionamento del circuito supponiamo di essere nelle condizioni ideali in cui l'amplificatore operazionale abbia guadagno infinito e supponiamo che il diodo sia in conduzione, ipotesi che vedremo in seguito essere verificata. Il diodo in conduzione, grazie alla sua bassa resistenza dinamica, agisce effettivamente come un efficace elemento in reazione, riportando la tensione dall'uscita dell'amplificatore al morsetto invertente. Ha senso quindi pensare che, analogamente a quanto avviene in un amplificatore invertente, il morsetto invertente si trovi ad essere una terra virtuale. Supponiamo ora di avere in ingresso una tensione  $V_{in} > 0$ . La conseguente corrente  $I_R$  che scorre nella resistenza  $R$  vale  $V_{in}/R$ . Questa corrente scorrerà interamente nel ramo in reazione, pertanto la corrente  $I_D$  del diodo sarà pari a  $I_R$ . E' opportuno a questo punto notare che l'ipotesi del

diodo in conduzione risulta verificata, in quanto il segno di  $I_D$ , per  $V_{in}$  positivo e di conseguenza  $I_R$  positivo, risulta consistente con la conduzione del diodo. Questa ipotesi, e quindi il funzionamento del circuito, non vale più per valori di  $V_{in}$  negativi.

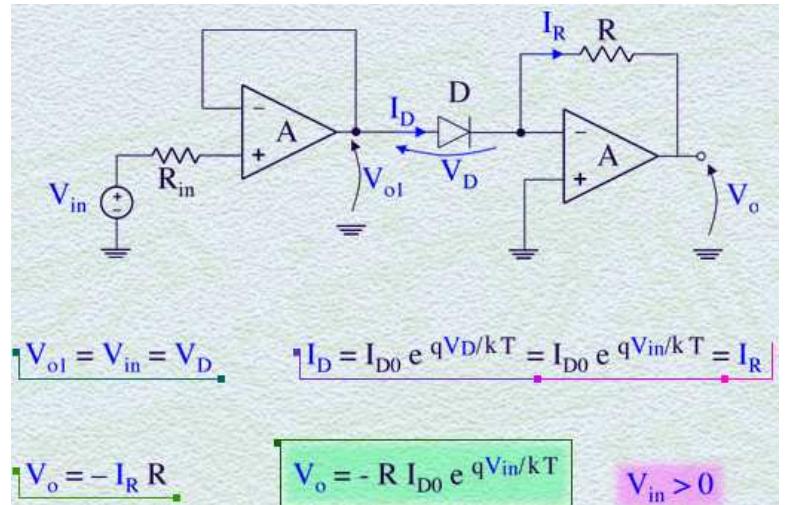
Considerando ora la caratteristica del diodo, supposto in conduzione, possiamo dire che la corrente  $I_D$  è pari a  $I_{D0}$  per l'esponenziale di  $q$  per  $V_D/kT$ , dove  $I_{D0}$  è la corrente di saturazione del diodo,  $q$  è la carica dell'elettrone,  $V_D$  è la tensione ai capi del diodo con il verso mostrato in figura,  $k$  è la costante di Boltzmann e  $T$  è la temperatura. Se si osserva che la tensione



all'uscita del circuito  $V_o$  è l'opposto della tensione  $V_D$  ai capi del diodo, la corrente  $I_D$  risulta pari a  $I_{D0}$  per l'esponenziale di  $-q$  per  $V_o/kT$ . Ponendo questa corrente pari alla corrente  $I_R$ , cioè a  $V_{in}/R$ , ed esplicitando la dipendenza della tensione di uscita  $V_o$  rispetto a quella di ingresso  $V_{in}$ , si giunge alla relazione caratteristica del circuito, in cui la tensione  $V_o$  risulta pari a  $-kT/q$  per il logaritmo naturale della tensione  $V_{in}$  divisa  $R \cdot I_{D0}$ . Risulta evidente da questa relazione che il circuito opera quindi un'amplificazione logaritmica della tensione applicata al suo ingresso. E' opportuno ricordare che **questa proprietà del circuito sussiste solo per valori positivi della tensione  $V_{in}$** .

Vediamo ora un circuito che effettua l'operazione contraria a quella dell'amplificatore logaritmico. Si tratta dell'**amplificatore antilogaritmico o esponenziale**. Nella figura è mostrato lo schema del circuito. Si può osservare che all'ingresso del circuito è stato posto un buffer di tensione. La seconda parte del circuito consiste in una configurazione invertente in cui si è sostituita la resistenza in ingresso con un diodo orientato come mostrato in figura. Il ruolo del buffer di tensione è quello di riportare interamente la tensione in ingresso  $V_{in}$  all'uscita  $V_{o1}$  e di applicarla interamente all'ingresso del diodo, a bassa impedenza. Questo avviene anche in presenza di una resistenza di sorgente  $R_{in}$ . Grazie alla bassa impedenza di uscita del buffer e grazie al fatto che il morsetto invertente dell'amplificatore operazionale è una terra virtuale, la tensione  $V_{o1}$  risulta interamente applicata ai capi del diodo, cioè  $V_D$  è uguale a  $V_{o1}$  e quindi a  $V_{in}$ . Se introduciamo ora la relazione caratteristica del diodo, la corrente  $I_D$  risulta pari a  $I_{D0}$  per l'esponenziale di  $q$  per  $V_D$  diviso per  $kT$  e quindi uguale a  $I_{D0}$  per l'esponenziale di  $q$  per  $V_{in}$  diviso per  $kT$ . Questa corrente scorre interamente nella resistenza  $R$ . Possiamo quindi determinare, grazie alla caduta di tensione sulla resistenza  $R$  provocata dalla corrente  $I_R = I_D$ , la tensione all'uscita del circuito che risulta pari a  $-R \cdot I_{D0}$  per l'esponenziale di  $q$  per  $V_{in}$  diviso per  $kT$ . La dipendenza della tensione di uscita dalla tensione di ingresso è quindi di tipo esponenziale.

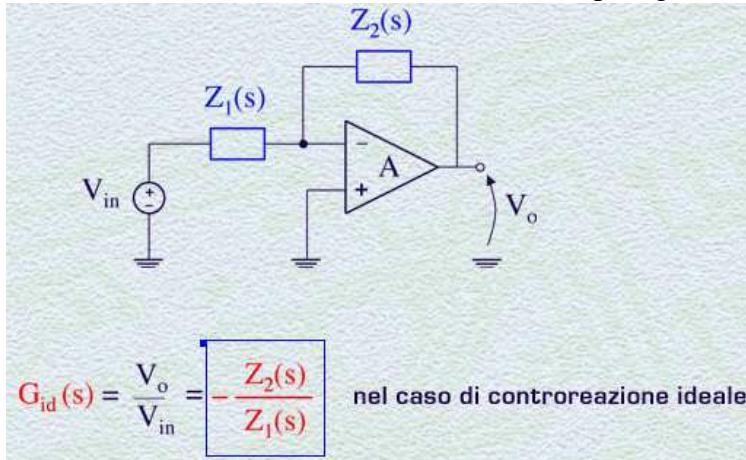
E' opportuno notare che ***anche in questo caso, il diodo conduce e quindi il circuito esercita la sua funzione correttamente solo se la tensione di ingresso  $V_{in}$  è positiva.***



## Lezione n. 33 : L'integratore

### Configurazione invertente con impedenze simboliche

Iniziamo con il considerare uno schema più generale



della configurazione invertente, trattata ampiamente nelle precedenti lezioni. Consideriamo un amplificatore invertente, come quello mostrato in figura, in cui però al posto delle resistenze  $R_1$  ed  $R_2$  siano introdotte due impedenze simboliche  $Z_1(s)$  e  $Z_2(s)$ , funzioni della variabile complessa  $s$ . I bipoli introdotti nel circuito, possono avere quindi oltre alla componente resistiva della loro impedenza, anche una componente reattiva, dipendente dalla frequenza. Questa rappresentazione più generale della configurazione invertente consente in cui componenti reattivi quali capacità o induttanze vengano utilizzati oltre alle resistenze.

Se venisse ripetuta tutta la trattazione che abbiamo visto in occasione dello studio dell'amplificatore invertente, semplicemente considerando la presenza delle impedenze simboliche  $Z_1(s)$  e  $Z_2(s)$ , al posto delle resistenze  $R_1$  e  $R_2$ , sarebbe possibile giungere alle stesse conclusioni a cui eravamo giunti in quell'occasione. In particolare, nella stessa maniera con cui avevamo determinato un guadagno ideale dell'amplificatore invertente pari al rapporto, col segno meno, tra le resistenze  $R_2$  e  $R_1$ , è possibile ora determinare anche nel caso del circuito con impedenze simboliche un guadagno ideale che risulta pari al rapporto, con il segno meno, tra l'impedenza  $Z_2(s)$  e l'impedenza  $Z_1(s)$ .

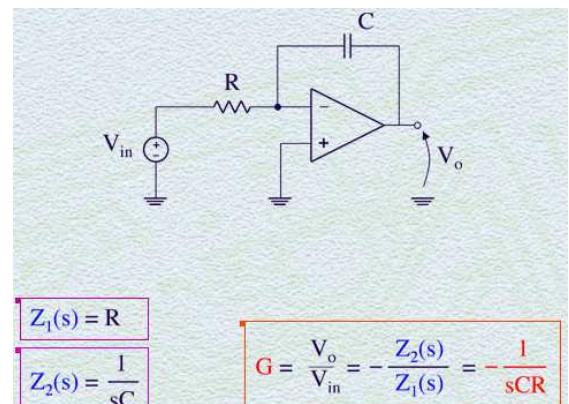
In questa maniera, dato un generico circuito in cui sia riconoscibile la presenza di una configurazione invertente, per il calcolo della sua funzione di trasferimento sarà sufficiente determinare le impedenze simboliche  $Z_1(s)$  e  $Z_2(s)$  e utilizzare semplicemente la relazione che esprime il guadagno del circuito in funzione di queste due impedenze.

### L'integratore ideale

Nella figura è mostrato lo schema del circuito integratore ideale. Lo schema è simile a quello dell'amplificatore invertente, con la differenza che una capacità  $C$  è stata inserita al posto della resistenza in reazione. La funzione di trasferimento del circuito può essere facilmente determinata utilizzando le impedenze simboliche  $Z_1(s)$  e  $Z_2(s)$ , rispettivamente per la resistenza  $R$  e la capacità  $C$ .  $Z_1(s)$  è pari alla resistenza  $R$ , mentre  $Z_2(s)$  è pari all'impedenza complessa della capacità  $C$  espressa in funzione di  $s$ , cioè  $1/sC$ . Se ora utilizziamo la relazione che esprime il guadagno ideale per una configurazione invertente con impedenze simboliche come  $-Z_2(s)/Z_1(s)$ , sostituendo nella relazione le impedenze prima indicate si ottiene che la funzione di trasferimento dell'integratore ideale è pari a  $-1/sCR$ . Valutiamo ora questa funzione di trasferimento nel

$$s = j\omega \rightarrow G(j\omega) = \frac{V_o(j\omega)}{V_{in}(j\omega)} = -\frac{1}{j\omega CR}$$

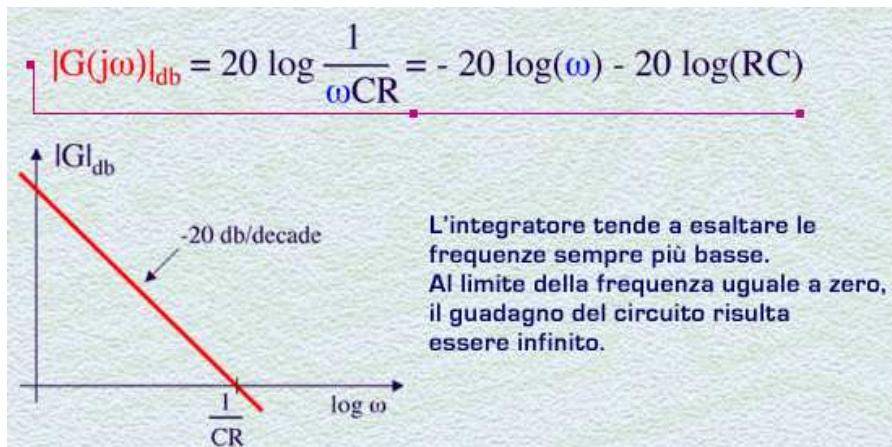
$$|G(j\omega)| = \frac{1}{\omega CR} \quad \angle G(j\omega) = \angle \frac{j}{\omega CR} = +90^\circ$$



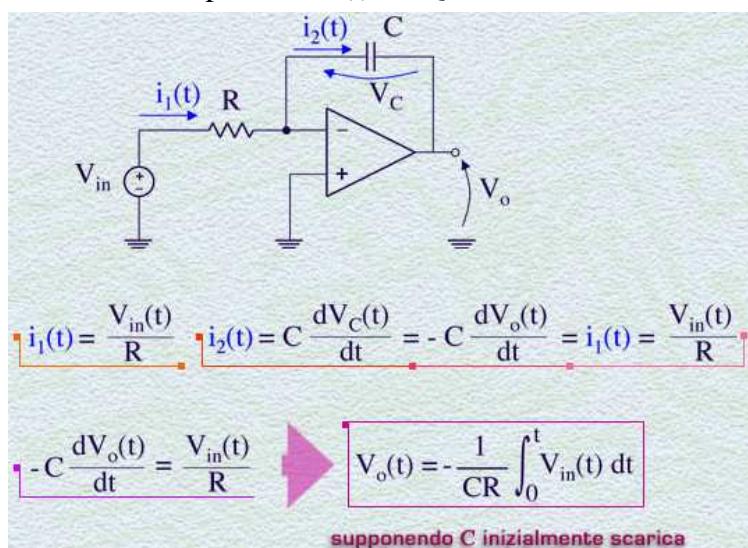
campo delle frequenze fisiche sostituendo ad  $s$  la variabile  $j\omega$ . L'espressione del guadagno diventa, con questa sostituzione, pari a  $-1/j\omega CR$ .

Il modulo della funzione di trasferimento  $G(j\omega)$  risulta pari a  $1/\omega CR$ , mentre per quanto riguarda la fase, il guadagno risulta essere un numero immaginario puro,  $j$  diviso per  $\omega CR$ , e pertanto la fase è costante e pari a  $90^\circ$ .

L'espressione della funzione di trasferimento in dB risulta pari a  $20\log \frac{1}{\omega CR}$ . Questa si può esprimere anche come  $-20\log\omega - 20\log RC$ . Il diagramma di Bode del modulo del guadagno risulta pertanto una retta a pendenza costante pari a -20 dB per decade. L'intersezione della retta con l'asse a 0 dB avviene in corrispondenza di una pulsazione  $\omega$  pari a  $1/CR$ . Dal diagramma di Bode notiamo quindi che l'integratore tende ad esaltare le frequenze sempre più basse. Al limite della frequenza uguale a zero, il guadagno del circuito risulta essere infinito. Si può pervenire a questa conclusione anche considerando che a frequenza nulla la capacità in reazione risulta essere di fatto un circuito aperto e pertanto il circuito funziona ad anello aperto con un guadagno di tensione idealmente infinito dato dall'amplificatore operazionale.



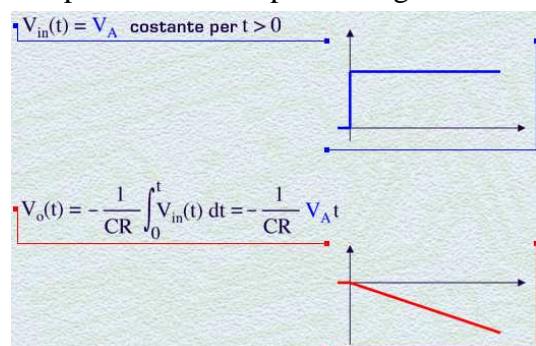
Determinata la funzione di trasferimento dell'integratore ideale nel dominio della frequenza, valutiamo ora il comportamento del circuito nel dominio del tempo. Supponiamo di applicare al circuito un segnale di ingresso  $V_{\text{in}}$  variabile nel tempo e determiniamo l'andamento del corrispondente segnale in uscita  $V_o(t)$ . Consideriamo innanzitutto che la corrente  $i_1(t)$  che scorre nella resistenza  $R$  è pari a  $V_{\text{in}}(t)/R$ . Questa corrente risulta uguale alla corrente  $i_2(t)$  che fluisce verso la capacità  $C$ .



Considerando la relazione caratteristica che lega corrente e tensione in una capacità, la corrente  $i_2(t)$  risulta pari alla capacità  $C$  per la derivata della tensione  $V_C$  rispetto al tempo. Dato che la tensione  $V_C$  è pari all'opposto della tensione di uscita, dato che il morsetto invertente è una terra virtuale, la corrente  $i_2(t)$  risulta uguale a  $-C$  per la derivata di  $V_o$  nel tempo. Questa corrente è pari a  $i_1(t)$ , uguale a  $V_{\text{in}}/R$ . Possiamo pertanto determinare l'equazione differenziale contenente le sole variabili  $V_{\text{in}}(t)$  e  $V_o(t)$ . Il termine  $-C$  per la derivata di  $V_o$  rispetto al tempo è uguale a  $V_{\text{in}}/R$ . Portando il termine  $dt$  a destra nell'equazione e integrando entrambi i termini della stessa si perviene alla relazione conclusiva in cui la tensione di uscita  $V_o(t)$  risulta pari a  $-1/CR$  per l'integrale della tensione in ingresso  $V_{\text{in}}$  da 0 all'istante  $t$ .

Il circuito opera quindi un'integrazione della tensione applicata in ingresso e deve proprio a questa proprietà il nome di integratore ideale. E' da osservare per ultimo che nell'integrazione dell'equazione differenziale **è stata considerata inizialmente scarica la capacità C**.

Consideriamo un esempio. Supponiamo di applicare all'ingresso di un integratore ideale un gradino di tensione di ampiezza  $V_A$ . La tensione di uscita, come abbiamo visto, può essere ricavata come  $-1/CR$  per l'integrale del segnale in ingresso. Integrando la tensione costante di ingresso  $V_A$ , si ottiene che l'uscita è funzione lineare del tempo e cioè è pari a  $-1/CR$  per  $V_A$  per  $t$ . L'uscita pertanto scende a tensione sempre più negativa con una pendenza costante pari a  $-V_A/CR$ .



Consideriamo ora l'effetto sul circuito integratore provocato dalla presenza delle correnti di bias all'ingresso dell'amplificatore operazionale. Consideriamo quindi un circuito integratore ideale in cui

introduciamo due generatori di corrente di bias  $I_{B1}$  e  $I_{B2}$  applicati agli ingressi dell'amplificatore operazionale. Il generatore  $I_{B2}$  non provoca alcun effetto in quanto la sua corrente scorre completamente verso la massa non modificando la tensione del morsetto non invertente che è fissato alla massa stessa. Vediamo invece l'effetto del generatore  $I_{B1}$ . Visto che il morsetto invertente si trova ad

essere una terra virtuale, la tensione ai capi di  $R$  risulta essere nulla. Pertanto, nessuna corrente scorre nella resistenza e quindi la corrente  $I_{B1}$  proviene interamente dalla capacità  $C$ . La corrente costante  $I_{B1}$  viene quindi integrata sulla capacità  $C$  producendo in uscita una salita a rampa della tensione verso valori sempre più positivi. La pendenza di questa rampa vale  $I_{B1}$  diviso per  $C$ . In conclusione, quindi, la presenza di una corrente di bias al morsetto invertente **porta inesorabilmente la tensione di uscita del circuito a salire a valori sempre più positivi fino a saturare ad un valore prossimo all'alimentazione positiva del circuito**. In queste condizioni, **il circuito risulta pertanto inutilizzabile**.

Consideriamo ora l'effetto provocato sul circuito dalla presenza di una tensione di offset dell'amplificatore operazionale. Se consideriamo un generatore di tensione di offset  $V_{OS}$  applicato al morsetto non-invertente, anche la tensione al morsetto invertente si porterà al valore  $V_{OS}$  in virtù della presenza del corto circuito virtuale. La corrente  $i_1$  che scorre nella resistenza  $R$  risulterà pertanto pari a  $V_{OS}/R$ . Questa corrente è pari alla corrente  $i_2$  che proviene dalla capacità  $C$ . La tensione di uscita del circuito  $V_o$  risulta pertanto pari alla tensione  $V_{OS}$  del morsetto invertente più la caduta di tensione sulla capacità  $C$ , determinata dall'integrazione della corrente  $i_2$ , che vale  $i_2/C$  per  $t$ .

Sostituendo a  $i_2$  il valore di  $i_1$ , la tensione  $V_o$  risulta pari a  $V_{OS} + V_{OS}/RC$  per  $t$ . La tensione in uscita risulta quindi pari ad un termine costante a cui si somma un termine che cresce nel tempo con andamento a rampa. Anche per la presenza della tensione di offset, quindi, **l'uscita del circuito tenderà a saturare ad un valore prossimo all'alimentazione positiva, rendendo inutilizzabile il circuito**.

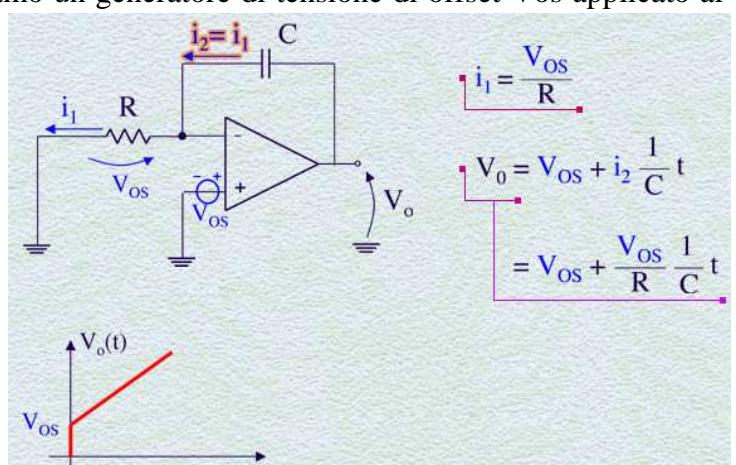
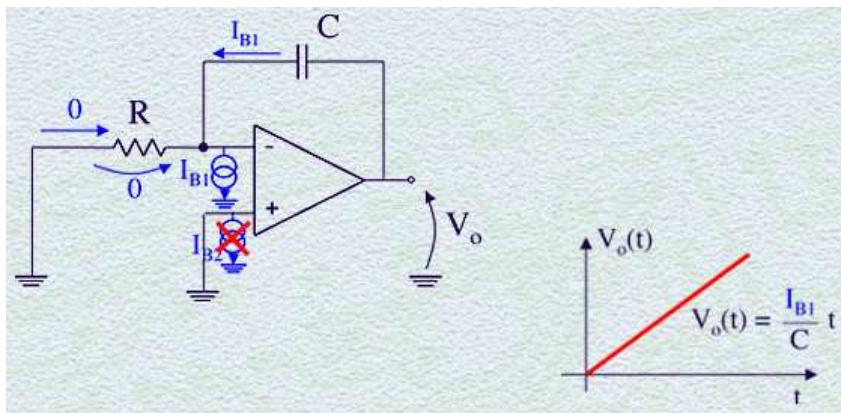
Per questo effetto, come per il precedente, risulta quindi evidente che il circuito integratore, così per come lo abbiamo considerato finora, non può essere utilizzato in un impiego reale.

## L'integratore approssimato

Un circuito che può risolvere i problemi evidenziati per il circuito ideale è l'integratore approssimato. Questo circuito consiste in un integratore ideale in cui è stata aggiunta una resistenza in parallelo alla capacità in reazione. La presenza della resistenza fornisce un cammino parallelo a quello della capacità per le correnti generate dalle imperfezioni in DC dell'amplificatore operazionale, vale a dire correnti di bias e tensione di offset.

Prima di valutare questi effetti sul circuito, determiniamo la funzione di trasferimento.

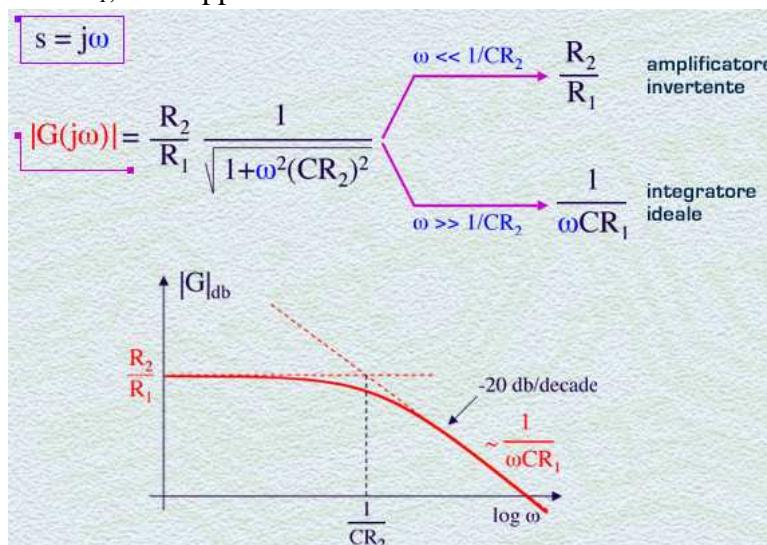
A tale scopo consideriamo ancora la relazione del guadagno ideale di una configurazione invertente con impedanze simboliche. L'impedenza  $Z_1(s)$  risulta pari a  $R_1$ , mentre l'impedenza  $Z_2(s)$  risulta pari al parallelo tra l'impedenza  $1/sC$  con l'impedenza  $R_2$ . L'impedenza simbolica risultante è pari al



rapporto tra  $R_2$  e  $1+sCR_2$ . Di conseguenza, l'espressione del guadagno  $G$ , pari a  $-Z_2/Z_1$ , risulta uguale a  $-R_2/1+sCR_2$ , diviso per  $R_1$ . L'espressione di  $G$  è quindi uguale a  $-R_2/R_1$  diviso per  $1+sCR_2$ . La funzione di trasferimento presenta pertanto un polo a pulsazione  $1/CR_2$ .

Esaminiamo l'andamento in frequenza di questa funzione di trasferimento operando la sostituzione di  $s$  con  $j\omega$ . Il modulo della funzione di trasferimento risulta pari a  $R_2/R_1$ , diviso per la radice di  $1+\omega^2 \cdot (CR_2)^2$ . Consideriamo questa funzione di  $\omega$  nei due casi estremi a bassa e ad alta frequenza. Per pulsazioni molto inferiori a  $1/CR_2$ , il temine  $\omega^2(CR_2)^2$  risulta trascurabile rispetto ad uno e pertanto il modulo della funzione di trasferimento risulta pari a  $R_2/R_1$  che rappresenta il guadagno di un amplificatore invertente. Questo è abbastanza intuitivo se si considera che a bassa frequenza l'impedenza della capacità  $C$  diventa elevata rendendo la presenza della capacità trascurabile rispetto alla resistenza  $R_2$ .

Per pulsazioni superiori a  $1/CR_2$  il temine  $\omega^2(CR_2)^2$  risulta dominante rispetto al termine 1 e pertanto, semplificando  $R_2$  nella relazione così ottenuta, il modulo della funzione di trasferimento risulta pari a  $1/\omega CR_1$ , che rappresenta la funzione di trasferimento dell'integratore ideale. Ad alta frequenza, infatti,

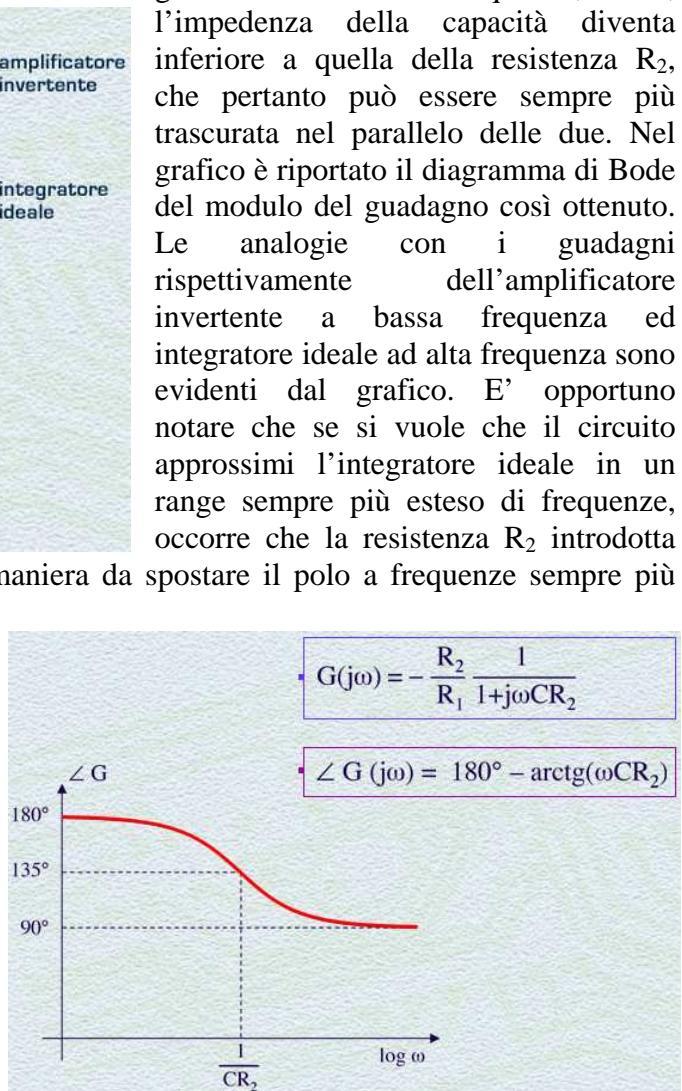
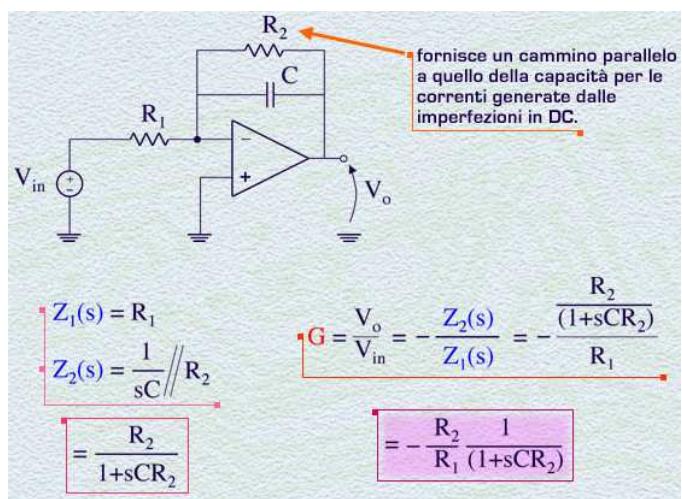


l'impedenza della capacità diventa inferiore a quella della resistenza  $R_2$ , che pertanto può essere sempre più trascurata nel parallelo delle due. Nel grafico è riportato il diagramma di Bode del modulo del guadagno così ottenuto. Le analogie con i guadagni rispettivamente dell'amplificatore invertente a bassa frequenza ed integratore ideale ad alta frequenza sono evidenti dal grafico. E' opportuno notare che se si vuole che il circuito approssimi l'integratore ideale in un range sempre più esteso di frequenze, occorre che la resistenza  $R_2$  introdotta

nel circuito sia di valore sempre maggiore, in maniera da spostare il polo a frequenze sempre più basse.

Consideriamo ora la fase della funzione di trasferimento. Questa risulta pari a un termine costante pari a  $180^\circ$  corrispondente al segno meno nella relazione, a cui si somma lo sfasamento variabile con la frequenza dato dalla presenza del polo. Come già visto in altri esempi, lo sfasamento del polo è pari a  $-\arctg(\omega CR_2)$ . Nel grafico relativo all'andamento con la frequenza della fase del guadagno si può notare quindi che la fase parte dai  $180^\circ$  a bassa frequenza per ridursi a  $135^\circ$  alla pulsazione  $1/CR_2$  del polo e raggiungere il valore di  $90^\circ$  ad alta frequenza.

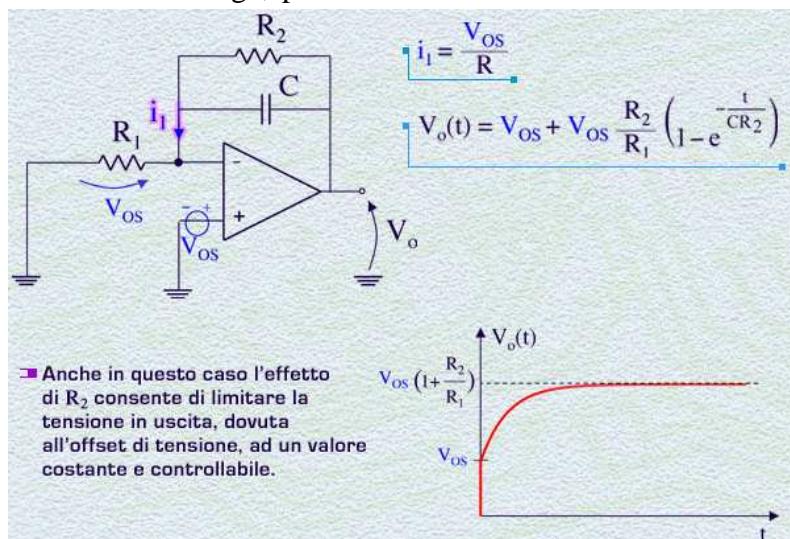
Consideriamo ora l'effetto sul circuito integratore approssimato introdotto dalla presenza delle correnti di bias. Analogamente a quanto visto per l'integratore ideale, la sola  $I_{B1}$  gioca un ruolo e in



questo caso fluisce nel parallelo della capacità C con la resistenza R<sub>2</sub>. Il calcolo della tensione sviluppata su questo parallelo da una corrente costante in ingresso può essere determinato con le regole viste in elettrotecnica, pertanto ci limitiamo a riportarne il risultato conclusivo. La tensione in uscita del circuito V<sub>o</sub> risulta pari a I<sub>B1</sub>R<sub>2</sub> per  $(1-e^{-t/CR_2})$ . La tensione di uscita sale quindi con andamento esponenziale e tende a saturare al valore I<sub>B1</sub>R<sub>2</sub>. La presenza di R<sub>2</sub> limita quindi la salita della tensione di uscita ad un valore costante, che può essere sempre più ridotto dall'utilizzatore del circuito, riducendo il valore di R<sub>2</sub>.

Vale la pena ricordare, però, che come visto in precedenza, questo circuito approssima sempre meglio l'integratore ideale, **maggior è il valore di R<sub>2</sub>**. Pertanto un compromesso nella scelta del valore di R<sub>2</sub> va ricercato per mantenere il comportamento come integratore del circuito in un range di frequenze sufficientemente elevato e nel contempo consentire una riduzione soddisfacente dell'effetto della corrente di bias.

In maniera analoga, può essere valutato l'effetto sul circuito dovuto alla presenza della tensione di offset.



Analogamente al caso dell'integratore ideale, la corrente i<sub>1</sub> è pari a V<sub>OS</sub>/R<sub>1</sub>. Questa corrente scorre nel parallelo tra la capacità C e la resistenza R<sub>2</sub>. La tensione di uscita risulta pari a V<sub>OS</sub> per la caduta di tensione sul parallelo di C ed R<sub>2</sub>. Utilizzando il conto appena visto nel caso della corrente di bias, la tensione in uscita risulta pari a V<sub>OS</sub> + V<sub>OS</sub> per R<sub>2</sub>/R<sub>1</sub> per  $(1 - e^{-t/CR_2})$ .

La tensione di uscita risulta quindi ad un termine costante più un termine che cresce esponenzialmente fino a portare la tensione a saturare al valore

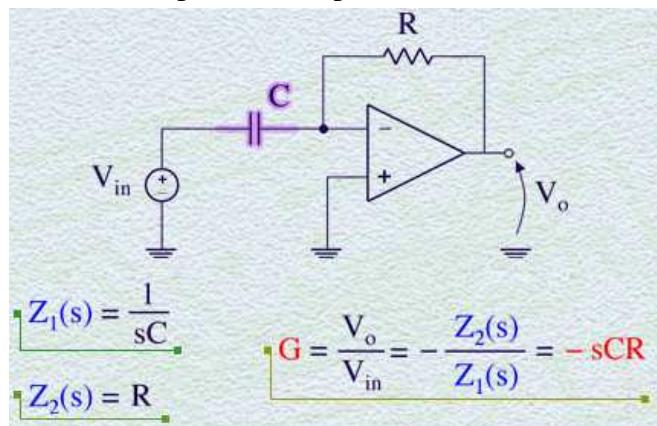
V<sub>OS</sub>(1+R<sub>2</sub>/R<sub>1</sub>). Anche in questo caso, quindi, l'effetto benefico di R<sub>2</sub> consente di limitare la tensione di uscita, dovuta all'offset di tensione, ad un valore costante e controllabile dall'utilizzatore.

## Lezione n. 34 : Il derivatore

### Il derivatore ideale

Il derivatore ideale è un circuito basato sulla configurazione invertente, con una capacità  $C$  disposta tra il generatore di ingresso  $V_{in}$  ed il morsetto invertente dell'amplificatore operazionale. La funzione di trasferimento del circuito può essere facilmente determinata utilizzando le impedenze simboliche  $Z_1(s)$  e  $Z_2(s)$ , rispettivamente per la capacità  $C$  e per la resistenza  $R$ . L'impedenza  $Z_1(s)$  è pari all'impedenza complessa della capacità  $C$ , cioè  $1/sC$ , mentre  $Z_2(s)$  è pari alla resistenza  $R$ . Se ora utilizziamo la relazione che esprime il guadagno ideale per una configurazione invertente con impedenze simboliche come  $-Z_2(s)/Z_1(s)$ , sostituendo nella relazione le impedenze prima indicate si ottiene che la funzione di trasferimento del derivatore ideale è pari a  $-sCR$ .

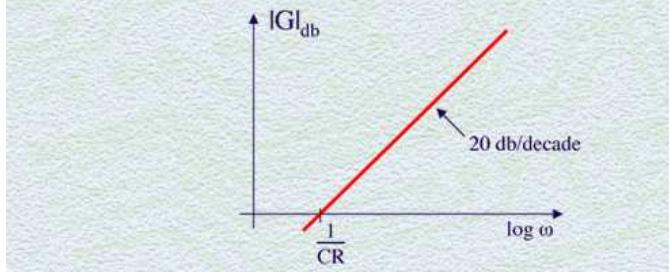
Determiniamo la risposta in frequenza del circuito sostituendo a  $s$  la variabile  $j\omega$ . L'espressione del guadagno diventa, con questa sostituzione, pari a  $-j\omega CR$ .



$$s = j\omega \rightarrow G(j\omega) = \frac{V_o(j\omega)}{V_{in}(j\omega)} = -j\omega CR$$

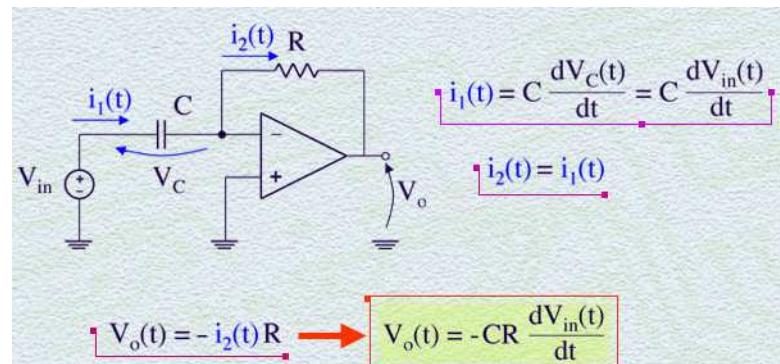
$$|G(j\omega)| = \omega CR \quad \angle G(j\omega) = -90^\circ$$

$$|G(j\omega)|_{db} = 20 \log(\omega CR) = 20 \log(\omega) + 20 \log(RC)$$



che **il derivatore tende ad esaltare le frequenze sempre più alte**. Al limite per una frequenza che tende all'infinito, il guadagno del circuito risulta essere infinito. Intuitivamente, al tendere all'infinito della frequenza, la capacità in ingresso al circuito tende a diventare un corto circuito e pertanto il segnale di tensione viene applicato direttamente in ingresso all'amplificatore operazionale e viene così amplificato con un guadagno di tensione idealmente infinito. Determinata la funzione di trasferimento del derivatore ideale, valutiamo ora il comportamento del circuito nel dominio del tempo. Supponiamo di applicare al circuito un segnale di ingresso  $V_{in}$  variabile nel tempo e determiniamo l'andamento del corrispondente segnale in uscita  $V_o(t)$ .

Considerando la relazione caratteristica che lega corrente e tensione in una capacità, la corrente  $i_1(t)$  risulta pari alla capacità  $C$  per la derivata della tensione  $V_C$  rispetto al tempo. Dato che la tensione  $V_C$  è pari alla tensione in ingresso, visto che il morsetto invertente è una terra virtuale, la corrente  $i_1(t)$  risulta uguale a  $C$  per la derivata di  $V_{in}$  nel tempo. La corrente  $i_2(t)$  che fluisce nella resistenza  $R$  in reazione risulta uguale alla corrente  $i_1(t)$ . La tensione in uscita è uguale a  $-i_2(t)R$  e, sostituendo a  $i_2$  la corrente  $i_1$  espressa in funzione di  $V_{in}$ , possiamo quindi pervenire ad una

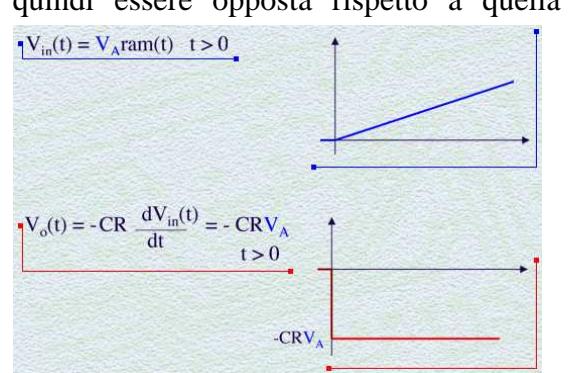


relazione differenziale contenente le sole variabili  $V_{in}(t)$  e  $V_o(t)$ . La tensione di uscita  $V_o$  è pari a  $-CR$  per la derivata di  $V_{in}$  rispetto al tempo. L'**uscita del circuito risulta quindi essere proporzionale alla derivata della tensione applicata in ingresso** e proprio per questo il circuito prende il nome di derivatore ideale.

La funzione implementata con questo circuito risulta quindi essere opposta rispetto a quella implementata nell'integratore ideale, in cui le posizioni della capacità e della resistenza sono invertite rispetto al circuito derivatore.

Consideriamo un esempio. Supponiamo di applicare all'ingresso di un derivatore ideale una tensione con un andamento a rampa nel tempo, come illustrato nella figura.

La tensione di uscita, come abbiamo visto, risulta uguale a  $-CR$  per la derivata del segnale in ingresso. Dato che la rampa in ingresso ha una pendenza pari a  $V_A$ , la tensione di uscita risulta quindi essere un gradino di tensione negativa di ampiezza pari a  $-CR \cdot V_A$ .



## Il derivatore approssimato

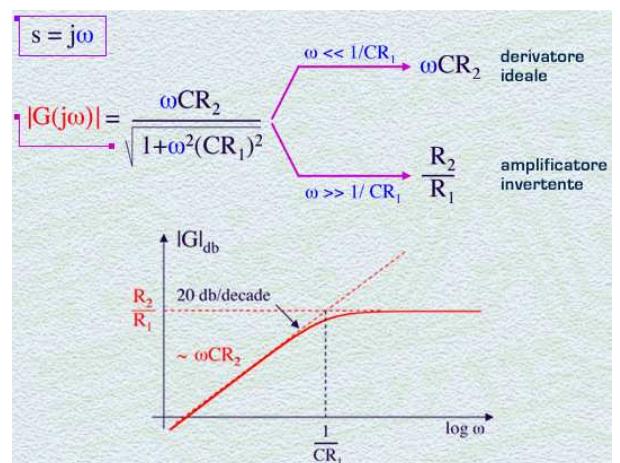
Analogamente a quanto visto nel caso del circuito integratore, anche per quanto riguarda il circuito derivatore è possibile introdurne una versione approssimata, che chiameremo derivatore approssimato. Questo circuito può avere diverse applicazioni, ad esempio come *filtro in frequenza di tipo passa-alto*, come vedremo meglio in seguito.

Il derivatore approssimato consiste in un circuito derivatore ideale in cui è stata aggiunta una resistenza in serie alla capacità collegata al generatore in ingresso.

Al fine di determinare la funzione di trasferimento del circuito, consideriamo ancora la relazione del guadagno ideale di una configurazione invertente con impedanze simboliche. L'impedenza  $Z_1(s)$  risulta pari a  $R_1 + \frac{1}{sC}$ . L'impedenza simbolica risultante è pari a  $\frac{1+sCR_1}{sC}$ . L'impedenza  $Z_2(s)$  è pari alla resistenza  $R_2$ . L'espressione del guadagno  $G$ , pari a  $\frac{V_o}{V_{in}}$ , è uguale a  $-\frac{R_2}{(1+sCR_1)}$ , e quindi risulta uguale a  $-\frac{sCR_2}{(1+sCR_1)}$ .

La funzione di trasferimento presenta pertanto uno zero nell'origine ed un polo a pulsazione  $1/CR_1$ .

Esaminiamo l'andamento in frequenza di questa funzione di trasferimento operando la sostituzione di  $s$  con  $j\omega$ . Il modulo della funzione di trasferimento risulta pari a  $\omega CR_2$ , diviso per la radice di  $1+\omega^2(CR_1)^2$ . Consideriamo questa funzione di  $\omega$  nei due casi estremi a bassa e ad alta frequenza. Per pulsazioni molto inferiori a  $1/CR_1$ , il termine  $\omega^2(CR_1)^2$  risulta trascurabile rispetto ad uno, sotto radice, e pertanto il modulo della funzione di trasferimento risulta pari a  $\omega CR_2$  che rappresenta il guadagno di un derivatore ideale. Questo è abbastanza intuitivo se si considera che a bassa frequenza l'impedenza della capacità  $C$  diventa elevata e quindi dominante nella serie con la resistenza  $R_1$ , introdotta nel derivatore approssimato.



Per pulsazioni superiori a  $1/CR_1$  il termine  $\omega^2(CR_1)^2$  risulta dominante rispetto ad uno e pertanto, semplificando  $\omega C$  nella relazione così ottenuta, il modulo della funzione di trasferimento risulta pari a  $R_2/R_1$ , che rappresenta la funzione di trasferimento dell'amplificatore invertente. Ad alta frequenza, infatti, la capacità tende a comportarsi come un corto circuito e quindi nella serie con la resistenza  $R_1$  tende a restare solo quest'ultima.

Nel grafico è riportato il diagramma di Bode del modulo del guadagno così ottenuto. Si può notare l'approssimazione del circuito con un derivatore ideale a bassa frequenza e con un amplificatore invertente ad alta frequenza. E' opportuno notare che se si vuole che il circuito approssimi il derivatore ideale in un range sempre più esteso di frequenze, occorre che la resistenza  $R_1$  introdotta nel circuito sia di valore sempre minore, in maniera da spostare il polo  $1/CR_1$  a frequenze sempre più elevate.

Consideriamo ora la fase della funzione di trasferimento. Questa risulta pari a un termine costante  $-90^\circ$ , a cui si somma lo sfasamento variabile con la frequenza dato dalla presenza del polo. Lo sfasamento introdotto dal polo è pari a meno l'arcotangente di  $\omega CR_1$ . Nel grafico relativo alla fase del guadagno si può notare quindi che la fase parte da  $-90^\circ$  a bassa frequenza per portarsi a  $-135^\circ$  in corrispondenza della pulsazione  $1/CR_1$  del polo e raggiungere il valore di  $-180^\circ$  ad alta frequenza.

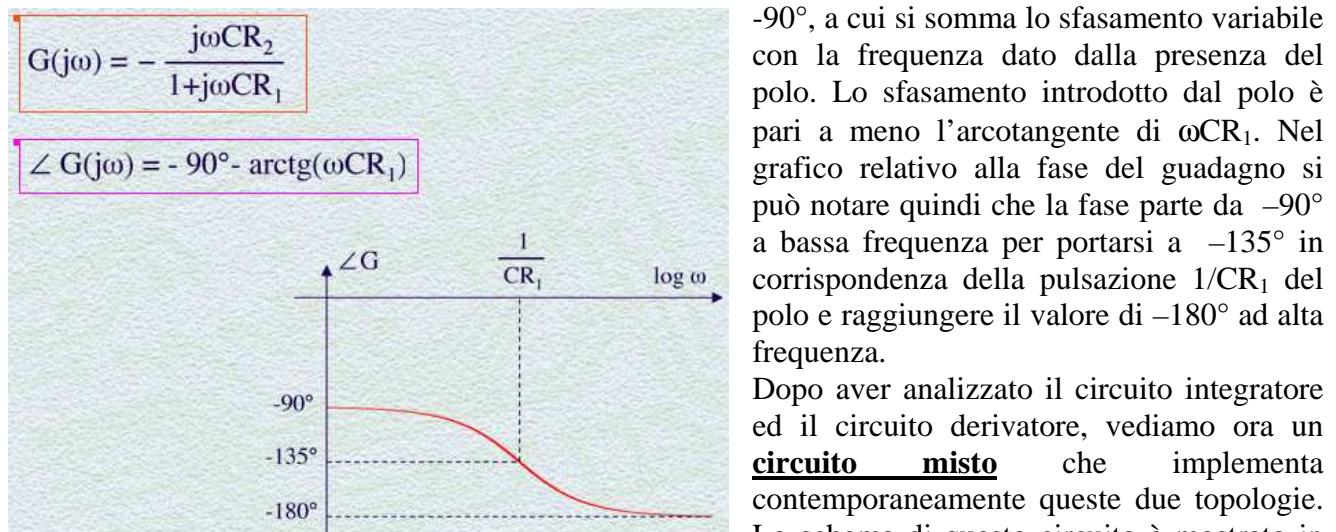
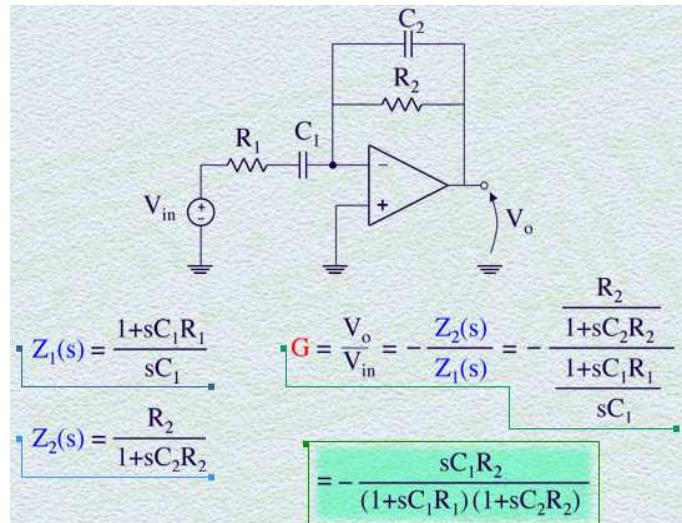


figura. Si può notare che per la presenza di una resistenza  $R_1$  in serie alla capacità  $C_1$  in ingresso e per la presenza di una resistenza  $R_2$  in parallelo alla capacità  $C_2$  sul ramo di reazione, riconosciamo in questo circuito la presenza degli elementi caratteristici sia del derivatore approssimato che dell'integratore approssimato.

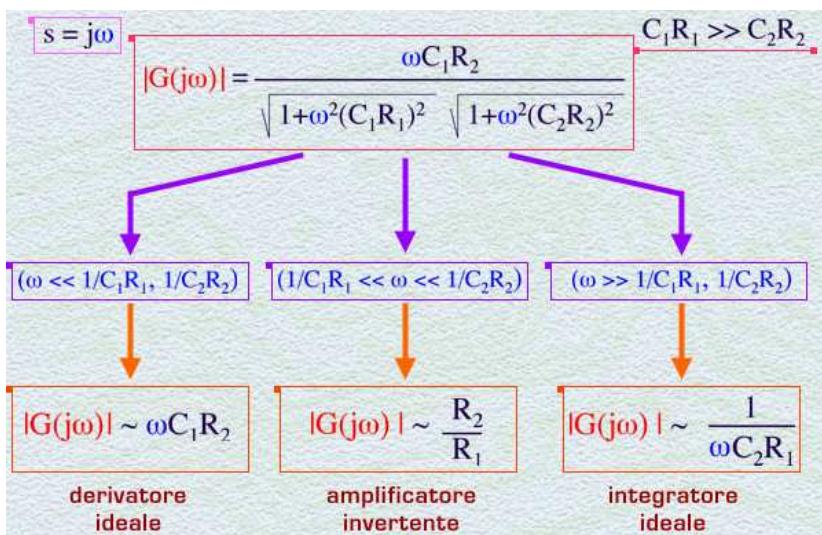
Determiniamo la funzione di trasferimento del circuito mediante l'utilizzo delle impedanze simboliche. L'impedenza  $Z_1(s)$ , come già visto per il derivatore approssimato, risulta pari a  $1+sC_1R_1$  diviso  $sC_1$ . L'impedenza  $Z_2(s)$ , come già visto invece per l'integratore approssimato, risulta pari a  $R_2$  diviso  $1+sC_2R_2$ . L'espressione del guadagno  $G$ , pari a  $-Z_2/Z_1$ , risulta uguale a  $-[R_2/(1+sC_2R_2)]/[(1+sC_1R_1)/sC_1]$ . Il guadagno  $G$  risulta quindi uguale a  $-sC_1R_2/[(1+sC_1R_1)\cdot(1+sC_2R_2)]$ . La funzione di trasferimento presenta pertanto uno zero nell'origine e due poli alla pulsazione rispettivamente di  $1/C_1R_1$  e  $1/C_2R_2$ .

Esaminiamo l'andamento in frequenza di questa funzione di trasferimento operando la sostituzione di  $s$  con  $j\omega$ . Il modulo della funzione di trasferimento risulta pari a  $\omega C_1 R_2$ , diviso per il prodotto della radice di  $1+\omega^2(C_1 R_1)^2$  per la radice di  $1+\omega^2(C_2 R_2)^2$ . Nelle tipiche condizioni di utilizzo di questo circuito, si fa sì che il polo del derivatore, cioè  $1/C_1 R_1$  sia a frequenza inferiore rispetto al polo dell'integratore  $1/C_2 R_2$ . Questo impone la condizione che  $C_1 R_1$  sia molto maggiore di  $C_2 R_2$ .

Risulta comodo stimare l'andamento del modulo della funzione di trasferimento prendendo in considerazione tre diversi intervalli sull'asse delle frequenze. Consideriamo innanzitutto l'intervallo in cui  $\omega$  sia molto inferiore ad entrambi i poli. In questo caso, entrambi i termini  $\omega^2(C_1 R_1)^2$  e  $\omega^2(C_2 R_2)^2$



sono trascurabili rispetto ad 1 sotto le radici, e pertanto il modulo della funzione di trasferimento risulta pari a  $\omega C_1 R_2$  che rappresenta il guadagno del derivatore ideale. Se ora consideriamo

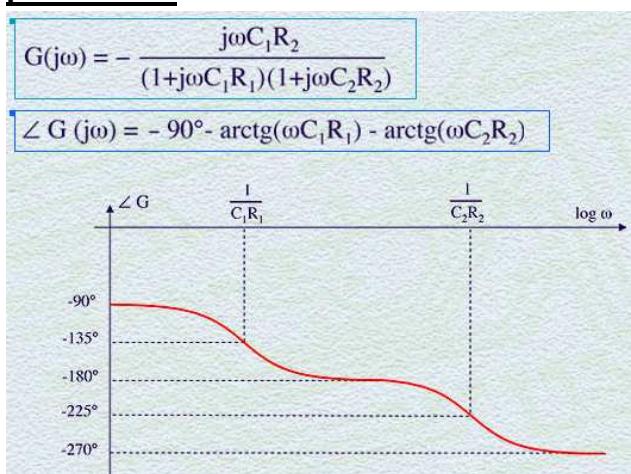
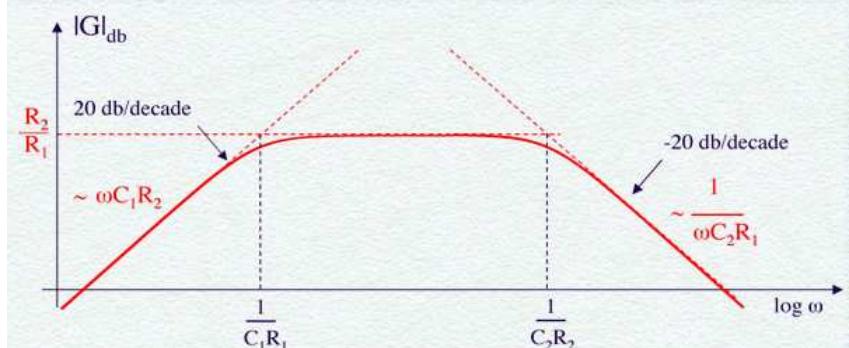


di frequenze molto superiori ad entrambi i poli, anche il termine integrando nella seconda radice è praticamente pari al solo  $\omega^2(C_2 R_2)^2$ . Pertanto, l'espressione del modulo del guadagno che ne consegue, con le opportune semplificazioni, risulta pari a  $1/\omega C_2 R_1$ , che è il guadagno dell'integratore ideale.

In conclusione, questo circuito misto si comporta come un derivatore ideale a bassa frequenza, come un amplificatore invertente a frequenze comprese tra quelle dei poli e infine si comporta come un integratore ideale alle alte frequenze.

Questa peculiarità del circuito è riassunta nel diagramma di Bode del guadagno, mostrato nella figura. Questo diagramma può essere visto come la **sovraposizione dei diagrammi di Bode rispettivamente del derivatore approssimato e dell'integratore approssimato**, considerando che il polo del primo,  $1/C_1 R_1$  è inferiore al polo del secondo,  $1/C_2 R_2$ .

Si può notare che questo circuito amplifica con un guadagno costante le frequenze comprese tra i due poli mentre attenua le frequenze al di fuori di questo intervallo. Una funzione del circuito è quella quindi di effettuare un **filtraggio di tipo passa-banda**.



Consideriamo infine la fase della funzione di trasferimento. Con ragionamenti analoghi a quelli fatti per il derivatore approssimato e per l'integratore approssimato, la fase risulta essere pari a  $-90^\circ$  meno l'arcotangente di  $\omega C_1 R_1$  meno l'arcotangente di  $\omega C_2 R_2$ . Considerando la posizione relativa dei poli  $1/C_1 R_1$  e  $1/C_2 R_2$ , la fase parte quindi a  $-90^\circ$  a bassa frequenza, per raggiungere i  $-135^\circ$  in corrispondenza della pulsazione  $1/C_1 R_1$ . Poi scende verso i  $-180^\circ$  a frequenze comprese tra i poli, raggiunge i  $-225^\circ$  alla pulsazione del secondo polo e infine tende a portarsi verso i  $-270^\circ$  a frequenze più elevate di entrambi i poli.

## Lezione n. 35 : Il filtraggio

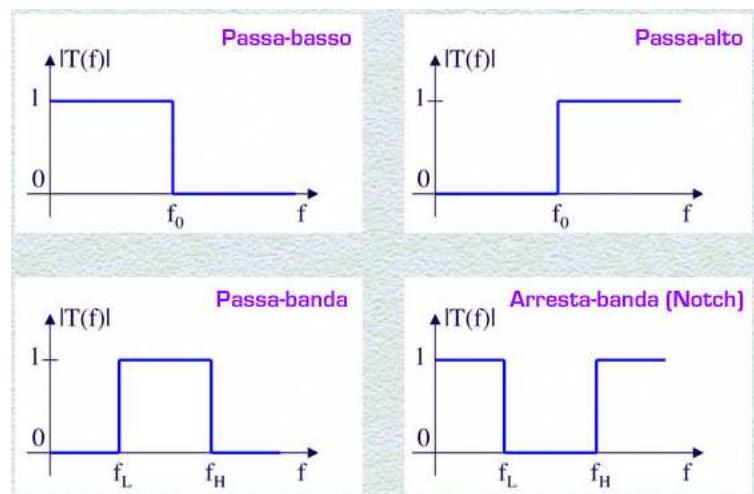
Il filtro è un circuito che opera su un segnale una selezione tra le frequenze da trasmettere inalterate e quelle invece su cui operare un'attenuazione più o meno elevata. I filtri sono molto utilizzati nel settore delle comunicazioni e della strumentazione elettronica. Una delle loro principali applicazioni in un sistema di trasmissione consiste, ad esempio, nel trasmettere solo le frequenze che fanno parte della banda del segnale e nel contempo rigettare o attenuare il più possibile le frequenze nello spettro del rumore o i disturbi fuori dalla banda del segnale.

Naturalmente, una trattazione completa del filtraggio e dell'implementazione di filtri va oltre gli scopi di questo corso. Noi ci limiteremo ad introdurre alcuni concetti introduttivi sul filtraggio e su alcuni esempi di filtri attivi che si possono realizzare attraverso l'uso di circuiti basati su amplificatori operazionali.

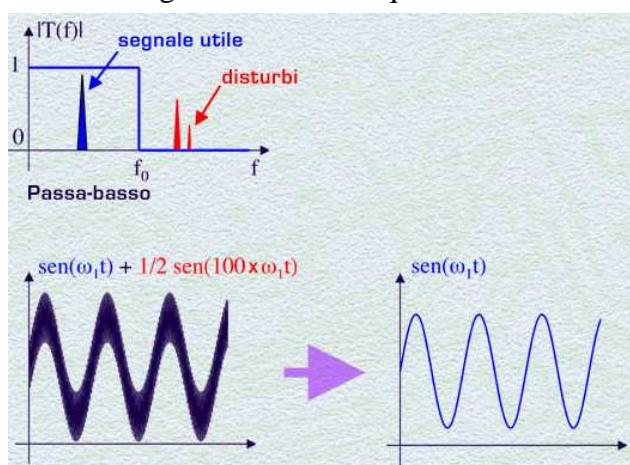
### Classificazione dei filtri

I filtri si possono classificare, in base alla loro risposta in frequenza, in quattro categorie principali: filtri passa-basso, filtri passa-alto, filtri passa-banda, filtri arresta-banda.

Consideriamo il modulo della risposta in frequenza  $|T(f)|$  di questi filtri, normalizzato rispetto al suo valore massimo. Possiamo osservare che un filtro passa-basso ideale trasmette inalterate le sinusoidi di frequenza inferiore a  $f_0$ , detta frequenza di taglio del filtro. Le sinusoidi a frequenza maggiore di  $f_0$  subiscono invece una completa attenuazione, cioè vengono completamente eliminate dallo spettro in frequenza del segnale. Nel filtro passa-alto ideale, avviene invece il contrario. Le sinusoidi a frequenza inferiore di  $f_0$  vengono completamente eliminate, mentre quelle di frequenza superiore vengono trasmesse inalterate. Il filtro passa-banda ideale trasmette invece solo le sinusoidi la cui frequenza è compresa in una banda con estremo inferiore a frequenza  $f_L$  ed estremo superiore a frequenza  $f_H$ . Le sinusoidi con frequenza esterna a questa banda vengono completamente attenuate. Al contrario, nel filtro arresta banda, vengono completamente attenuate le sinusoidi con la frequenza contenuta nella banda tra  $f_L$  e  $f_H$ , mentre vengono trasmesse quelle esterne a tale banda. Nel caso il filtro arresta-banda operi la selezione in una banda  $f_L-f_H$  molto stretta, prende anche il nome di filtro Notch.

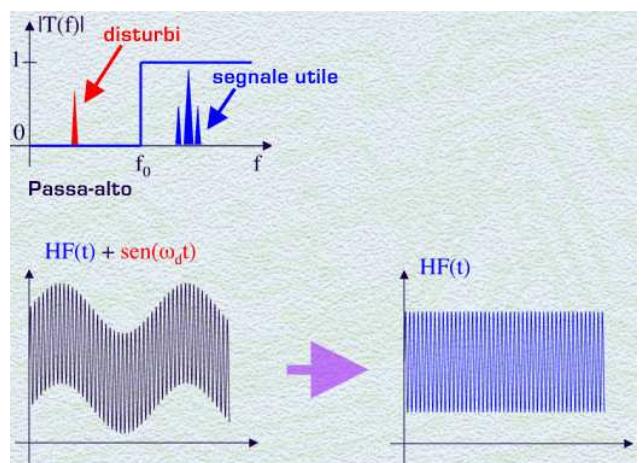


Le possibilità di impiego dei filtri in elettronica sono molteplici. Ci limitiamo a vedere due semplici esempi di applicazione che ce ne illustrano la loro importanza. Supponiamo, in un primo caso, di avere un segnale utile di tipo sinusoidale in un sistema in cui siano presenti anche dei disturbi a più alta frequenza. Nell'esempio specifico, ad una sinusoide a frequenza  $\omega_1$  si aggiunge un disturbo a frequenza 100 volte maggiore di quella della sinusoide utile e di ampiezza pari alla metà. Come si vede dalla figura, il segnale risultante, in cui si intravede la forma d'onda della sinusoide utile a bassa frequenza, risulta decisamente influenzato dal disturbo aggiunto. Operando un filtraggio con un filtro passa-basso ad una frequenza  $f_0$  maggiore di quella del segnale utile ma minore della frequenza del disturbo, è possibile estrarre la sola sinusoide a frequenza  $\omega_1$ , avendo rigettato il disturbo a frequenza maggiore.



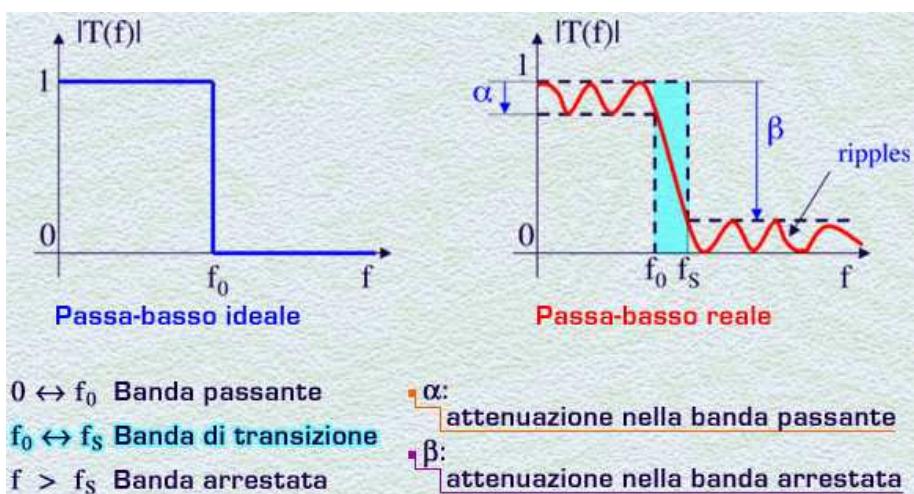
intravede la forma d'onda della sinusoide utile a bassa frequenza, risulta decisamente influenzato dal disturbo aggiunto. Operando un filtraggio con un filtro passa-basso ad una frequenza  $f_0$  maggiore di quella del segnale utile ma minore della frequenza del disturbo, è possibile estrarre la sola sinusoide a frequenza  $\omega_1$ , avendo rigettato il disturbo a frequenza maggiore.

Nel secondo esempio consideriamo una situazione dalle caratteristiche opposte rispetto alla precedente. Supponiamo di avere un segnale utile  $HF(t)$  composto da uno spettro ad alta frequenza a cui si sovrappone un disturbo sinusoidale a bassa frequenza, ad esempio dovuto alla tensione di rete a 50 Hz. Dal grafico del segnale complessivo si nota come il disturbo a bassa frequenza provochi un'evidente ondeggiamiento del segnale utile. In questo caso, l'uso di un filtro di tipo passa-alto, consente di trasmettere inalterate le alte frequenze di cui è composto il segnale utile, eliminando la fastidiosa oscillazione a bassa frequenza.



### Caratteristiche dei filtri reali

I filtri che abbiamo introdotto finora presentano una risposta in frequenza di tipo ideale, e cioè caratterizzata da un **guadagno costante nella banda da trasmettere e da un guadagno rigorosamente nullo nella banda da attenuare**. Filtri come questi rappresentano solo un'idealizzazione teorica, in quanto questa elevata selettività nella risposta in frequenza non è realizzabile con i circuiti reali. Analizziamo quindi quali siano le caratteristiche tipiche della risposta in frequenza di filtri reali. Nella figura è mostrato un confronto tra la risposta di un filtro passa-basso ideale



l'attenuazione massima del guadagno nell'intervallo di frequenza, tra 0 e  $f_0$ . Questo intervallo di frequenza definisce la banda passante del filtro.

Per frequenze maggiori di  $f_0$ , il guadagno del filtro ideale è rigorosamente uguale a zero. Nel caso del filtro reale, invece, per frequenze maggiori di  $f_0$ , il guadagno inizia a scendere e diventa inferiore di un certo valore, ritenuto accettabile a seconda dello scopo di utilizzo del filtro, solo ad una certa frequenza  $f_s$  maggiore di  $f_0$ . L'intervallo delle frequenze superiori a  $f_s$  definisce la banda arrestata del filtro. Come già detto, nella banda arrestata il guadagno non è nullo, se non per alcune frequenze specifiche, ma può essere mantenuto solo al di sotto di un certo valore. Il livello di attenuazione richiesto per il guadagno nella banda arrestata rispetto al suo valore massimo nella banda passante viene identificato con un parametro  $\beta$ .

I parametri  $\alpha$  e  $\beta$ , quindi, rappresentano dei **parametri di merito** del filtro e permettono di identificare quanto meglio un filtro reale approssimi la risposta di un filtro ideale. Volendo approssimare sempre meglio il filtro ideale, l'attenuazione  $\alpha$  nella banda passante deve essere sempre più prossima allo zero. Valori tipici di  $\alpha$ , che viene comunemente indicato in decibel, vanno da 0.05 a 3 dB. Per approssimare meglio la risposta del filtro ideale nella banda arrestata, invece, il fattore di attenuazione  $\beta$  deve essere invece il più grande possibile. Valori tipici di  $\beta$  in decibel sono compresi tra i 20 dB ed i 100 dB.

E' da osservare che sia nella banda passante che nella banda arrestata, il guadagno può mostrare un andamento non monotono, ma al contrario alcune in alcune classi di filtri l'andamento può presentare delle oscillazioni che vengono comunemente indicate con il termine inglese 'ripples'.

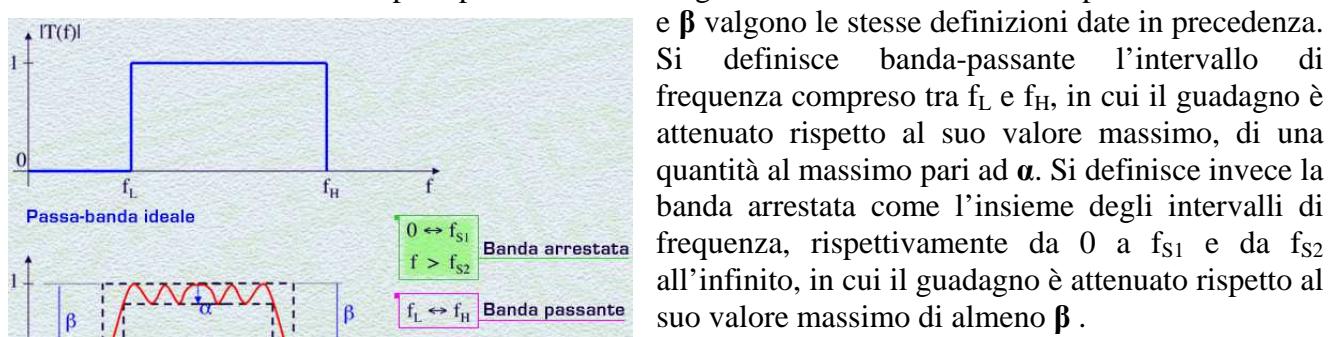
ideale e quella di un filtro passa-basso reale. Si può notare innanzitutto che mentre nel filtro ideale la risposta è costante nella banda da trasmettere, nel filtro reale il guadagno non è costante ma varia all'interno di un certo intervallo di ampiezze. La massima deviazione del guadagno del filtro dal suo valore massimo viene identificata con un parametro  $\alpha$ , definito come

Da ultimo, notiamo che nella banda di frequenze comprese tra  $f_0$  e  $f_s$ , il guadagno assume valori che non sono ben lontani sia dal guadagno massimo che dallo zero. Questo intervallo di frequenze viene definito come **banda di transizione**.

Se consideriamo ora un filtro passa-alto reale, la sua risposta in frequenza mostra le stesse differenze rispetto al corrispondente filtro ideale, appena riscontrate per il filtro passa-basso. In particolare, si definisce banda passante del filtro l'insieme delle frequenze superiori alla frequenza di taglio  $f_0$ . In questo intervallo, il guadagno può risultare attenuato rispetto al suo valore massimo di una quantità pari ad  $\alpha$ , che deve assumere il valore più piccolo possibile per approssimare meglio il filtro ideale. La banda arrestata, invece, corrisponde all'intervallo di frequenze da 0 alla frequenza  $f_s$ , inferiore a  $f_0$ . Il guadagno in corrispondenza di  $f_s$  è attenuato rispetto al valore massimo di una quantità  $\beta$ , il più grande possibile per meglio approssimare il filtro ideale.

Anche per il filtro passa-alto si definisce banda di transizione l'intervallo di frequenza tra  $f_s$  e  $f_0$ , in cui il guadagno non è né più attenuato del fattore  $\beta$  (come nella banda arrestata) né meno attenuato del fattore  $\alpha$  (come nella banda passante).

Le stesse considerazioni fatte per i primi due filtri valgono anche nel caso del filtro passa-banda. Per  $\alpha$  e  $\beta$  valgono le stesse definizioni date in precedenza.

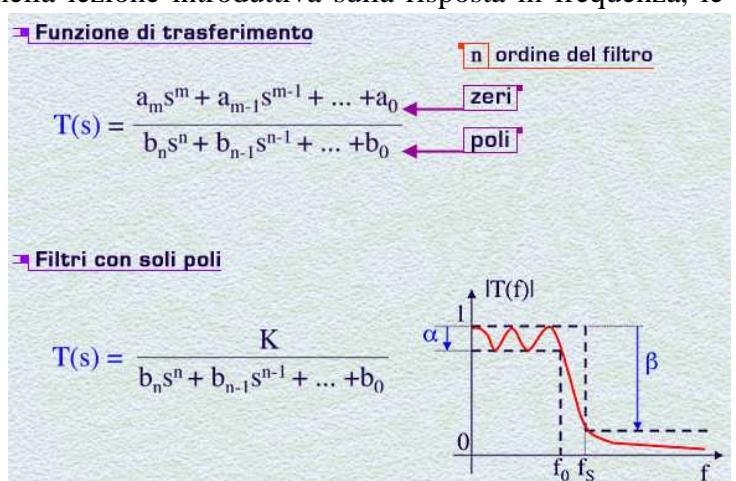


Si definisce banda-passante l'intervallo di frequenza compreso tra  $f_L$  e  $f_H$ , in cui il guadagno è attenuato rispetto al suo valore massimo, di una quantità al massimo pari ad  $\alpha$ . Si definisce invece la banda arrestata come l'insieme degli intervalli di frequenza, rispettivamente da 0 a  $f_{S1}$  e da  $f_{S2}$  all'infinito, in cui il guadagno è attenuato rispetto al suo valore massimo di almeno  $\beta$ .

La definizione dei parametri caratteristici per il filtro arresta-banda, segue lo stesso schema visto per gli altri tre tipi di filtri.

La funzione di trasferimento di un filtro reale è rappresentabile, come già visto per una generica rete elettrica, come un rapporto tra due polinomi. Il grado  $n$  del polinomio al denominatore definisce l'ordine del filtro. Come abbiamo già visto nella lezione introduttiva sulla risposta in frequenza, le radici del numeratore definiscono gli zeri della rete, mentre le radici del denominatore ne definiscono i poli. Esistono alcune classi di filtri in cui la risposta in frequenza non presenta zeri ma solo poli. In questo caso, non esistono delle frequenze in cui la funzione di trasferimento ha guadagno nullo e pertanto l'andamento del guadagno nella banda arrestata è solitamente monotono decrescente senza che esso raggiunga mai un valore nullo.

Sebbene esistano molte classi di filtri con caratteristiche differenti, nell'ambito di questo corso ci limiteremo a considerare semplici filtri attivi del I ordine e, come esempio di una classe di filtri di ordine superiore, i filtri di Butterworth.

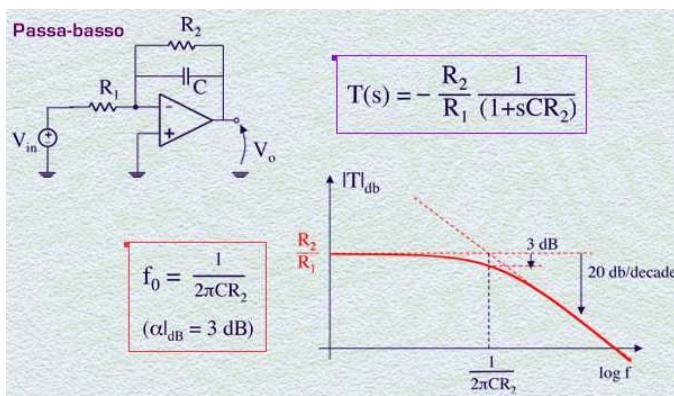


## Filtri attivi del primo ordine

Si definiscono filtri attivi, **circuiti che utilizzano stadi amplificanti opportunamente controllati** in maniera tale da implementare funzioni di trasferimento corrispondenti ad una certa classe di filtri. Questi filtri si differenziano dai cosiddetti filtri passivi che sono invece basati su reti passive utilizzanti induttanze e capacità.

Una prima semplice classe di filtri attivi sono quelli del 1° ordine basati sulla configurazione invertente con impedenze simboliche che abbiamo già analizzato nelle scorse lezioni. Rivediamo quindi brevemente le caratteristiche di questi circuiti, ponendo particolare attenzione questa volta alla loro funzione di filtraggio.

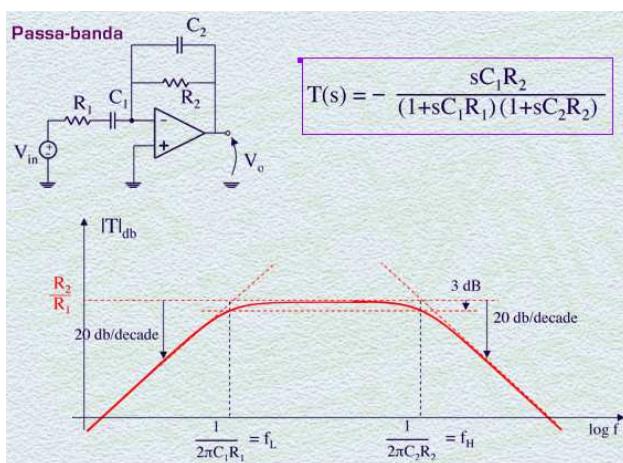
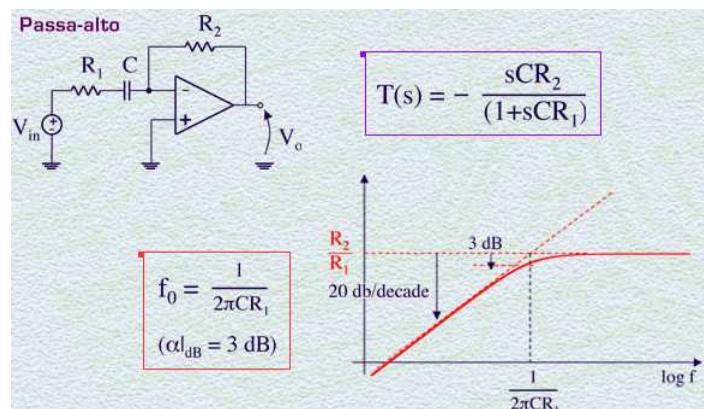
Un filtro **passa-basso** del I ordine può essere realizzato con il circuito **integratore approssimato**, di cui si ripropone lo schema nella figura. La funzione di trasferimento presenta un polo determinato dai



dai due componenti in reazione, la resistenza  $R_2$  e la capacità  $C$ . Come si può osservare dal diagramma di Bode della risposta in frequenza, il circuito amplifica con uguale guadagno  $-R_2/R_1$  fino alla frequenza  $1/2\pi CR_2$ , mentre attenua le frequenze superiori con un fattore di attenuazione che cresce di 20 dB per ogni decade. La banda passante del filtro, definita entro un'attenuazione  $\alpha$  di 3 dB, è pari a  $1/2\pi CR_2$ . Questo filtro è molto semplice da realizzare. Il principale svantaggio, essendo un

filtro solo del I ordine, è che l'attenuazione delle frequenze indesiderate avviene in maniera poco selettiva. Infatti, l'attenuazione è di soli 20 dB ad una decade oltre la frequenza di taglio e raggiunge i 100 dB solo cinque decadi oltre la frequenza di taglio.

Un filtro **passa-alto** del I ordine può essere invece realizzato con il circuito **derivatore approssimato**, di cui si ripropone lo schema nella figura. La funzione di trasferimento presenta uno zero nell'origine ed un polo determinato dai due componenti sul ramo in ingresso, la resistenza  $R_1$  e la capacità  $C$ . Dal diagramma di Bode si osserva che il circuito amplifica con uguale guadagno  $-R_2/R_1$  oltre alla frequenza  $1/2\pi CR_1$ , mentre attenua le frequenze inferiori, con un fattore di attenuazione di 20 dB per ogni decade. La banda passante del filtro, definita entro un'attenuazione  $\alpha$  di 3 dB, si estende dalla frequenza  $1/2\pi CR_1$  all'infinito. La limitata capacità di attenuazione di questo filtro nella banda da arrestare è identica a quella del filtro passa-basso precedente.



Concludiamo riportando lo schema ed il diagramma di Bode del circuito passa-banda ottenuto con la configurazione mista derivatore-integratore. La banda passante del circuito, entro un'attenuazione di 3 dB, è definita tra le frequenze  $f_L = 1/2\pi C_1 R_1$  e la frequenza  $f_H$  pari a  $1/2\pi C_2 R_2$ .

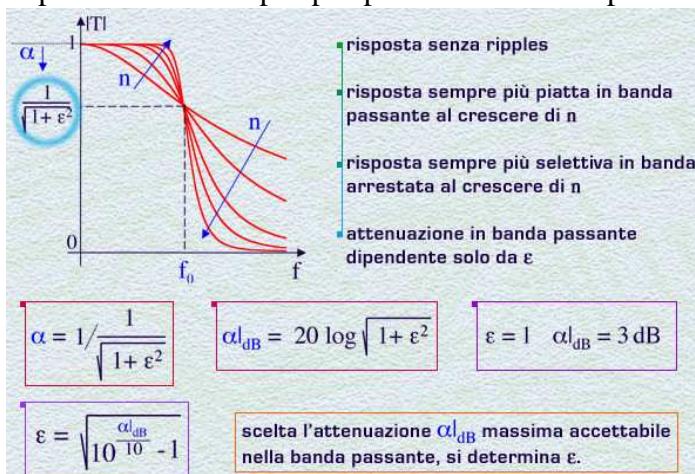
Le attenuazioni fuori banda passante sono le stesse che abbiamo visto per i circuiti passa-basso e passa-alto.

## Filtri Butterworth

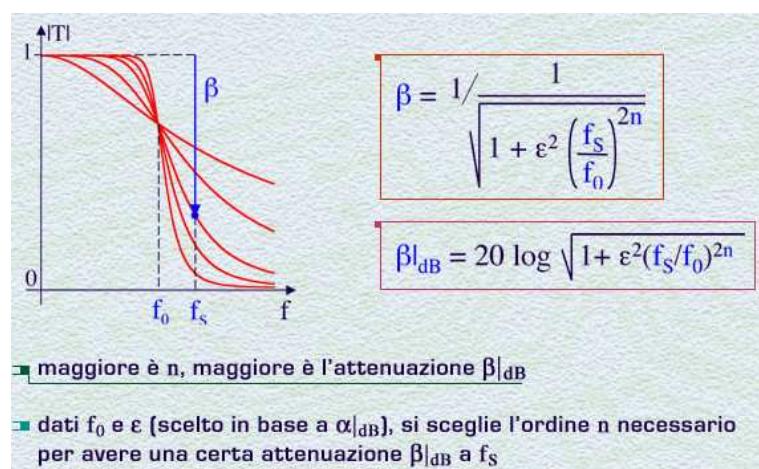
I Filtri di Butterworth, che devono il proprio nome all'ingegnere britannico che li introdusse nel 1930, sono caratterizzati da una risposta in frequenza, mostrata in figura, monotona decrescente e quindi priva di oscillazioni o *ripples*. La funzione di trasferimento  $T(f)$  di un filtro passa-basso di Butterworth con pulsazione di taglio  $f_0$  è esprimibile matematicamente come 1 sulla radice di  $[1+\epsilon^2 \cdot (f/f_0)^{2n}]$ , dove  $n$  rappresenta l'ordine del filtro. Un'altra caratteristica di questo filtro è che la risposta alla frequenza di taglio  $f_0$ ,  $T(f_0)$ , è pari a 1 sulla radice di  $[1+\epsilon^2]$ , per qualsiasi valore di  $n$ . Questa peculiarità risulta evidente dal grafico mostrato in figura, in cui la risposta del filtro è tracciata per diversi valori di  $n$ .

Oltre alla già citata assenza di *ripples*, occorre notare che, all'aumentare dell'ordine del filtro, la risposta risulta sempre più piatta nella banda passante, cioè fino alla pulsazione di taglio  $f_0$ . Inoltre, per frequenze superiori a  $f_0$  la risposta diventa sempre più ripida, e quindi selettiva nella banda da arrestare, sempre all'aumentare dell'ordine del filtro. La risposta in frequenza approssima quindi sempre meglio la risposta di un filtro passa-basso ideale, maggiore è l'ordine  $n$  del filtro.

Il termine 1 sulla radice di  $[1+\epsilon^2]$ , corrisponde al minimo valore del guadagno del filtro nella banda passante e, come detto dipende dal solo parametro  $\epsilon$ , e non dall'ordine del filtro. Minore è il parametro  $\epsilon$ , minore risulta l'attenuazione del guadagno nella banda passante rispetto al suo valore massimo. Il termine di attenuazione del guadagno '1' ed il guadagno in corrispondenza della frequenza di taglio, pari a 1 sulla radice di  $[1+\epsilon^2]$ . L'attenuazione  $\alpha$ , espressa in decibel, risulta quindi pari a  $20 \log \sqrt{1+\epsilon^2}$ . Se, ad esempio,  $\epsilon$  è uguale ad uno, il guadagno risulta attenuato rispetto al suo valore massimo di 3 dB. Esplicitando la relazione trovata rispetto ad  $\epsilon$ , si trova che  $\epsilon$  è uguale alla radice di 10 elevato al valore di  $\alpha$  in dB, diviso per 10, meno 1. In questa maniera, scelto il valore di attenuazione massima che si ritenga accettabile nella banda passante, si determina il corrispondente valore di  $\epsilon$ . Analogamente a quanto fatto per l'attenuazione nella banda passante, è possibile determinare l'attenuazione  $\beta$  nella banda arrestata a partire da una frequenza  $f_S$ . Il fattore di attenuazione  $\beta$  è esprimibile come rapporto tra il massimo guadagno e il guadagno calcolato alla frequenza  $f_S$ , che vale radice di  $[1+\epsilon^2 \cdot (f_S/f_0)^{2n}]$ . Esprimendo questa attenuazione in decibel otteniamo che  $\beta_{dB}$  è uguale a 20 volte il logaritmo della radice di  $[1+\epsilon^2 \cdot (f_S/f_0)^{2n}]$ . Si può osservare dalla relazione, come è già evidente dal grafico mostrato sopra, che **al crescere dell'ordine del filtro, aumenta l'attenuazione  $\beta$**  alla frequenza  $f_S$  e quindi corrispondentemente l'attenuazione complessiva nella banda arrestata.



valore massimo. Il termine di attenuazione del guadagno '1' ed il guadagno in corrispondenza della frequenza di taglio, pari a 1 sulla radice di  $[1+\epsilon^2]$ . L'attenuazione  $\alpha$ , espressa in decibel, risulta quindi pari a  $20 \log \sqrt{1+\epsilon^2}$ . Se, ad esempio,  $\epsilon$  è uguale ad uno, il guadagno risulta attenuato rispetto al suo valore massimo di 3 dB. Esplicitando la relazione trovata rispetto ad  $\epsilon$ , si trova che  $\epsilon$  è uguale alla radice di 10 elevato al valore di  $\alpha$  in dB, diviso per 10, meno 1. In questa maniera, scelto il valore di attenuazione massima che si ritenga accettabile nella banda passante, si determina il corrispondente valore di  $\epsilon$ . Analogamente a quanto fatto per l'attenuazione nella banda passante, è possibile determinare l'attenuazione  $\beta$  nella banda arrestata a partire da una frequenza  $f_S$ . Il fattore di attenuazione  $\beta$  è esprimibile come rapporto tra il massimo guadagno e il guadagno calcolato alla frequenza  $f_S$ , che vale radice di  $[1+\epsilon^2 \cdot (f_S/f_0)^{2n}]$ . Esprimendo questa attenuazione in decibel otteniamo che  $\beta_{dB}$  è uguale a 20 volte il logaritmo della radice di  $[1+\epsilon^2 \cdot (f_S/f_0)^{2n}]$ . Si può osservare dalla relazione, come è già evidente dal grafico mostrato sopra, che **al crescere dell'ordine del filtro, aumenta l'attenuazione  $\beta$**  alla frequenza  $f_S$  e quindi corrispondentemente l'attenuazione complessiva nella banda arrestata.



In conclusione, scelti la frequenza di taglio del filtro  $f_0$  ed il parametro  $\epsilon$ , determinato in base all'attenuazione massima tollerata nella banda passante, è possibile determinare l'ordine  $n$  del filtro necessario per avere una soddisfacente attenuazione nella banda arrestata, corrispondente ad un certo valore di  $\beta$  alla frequenza di interesse  $f_s$ .

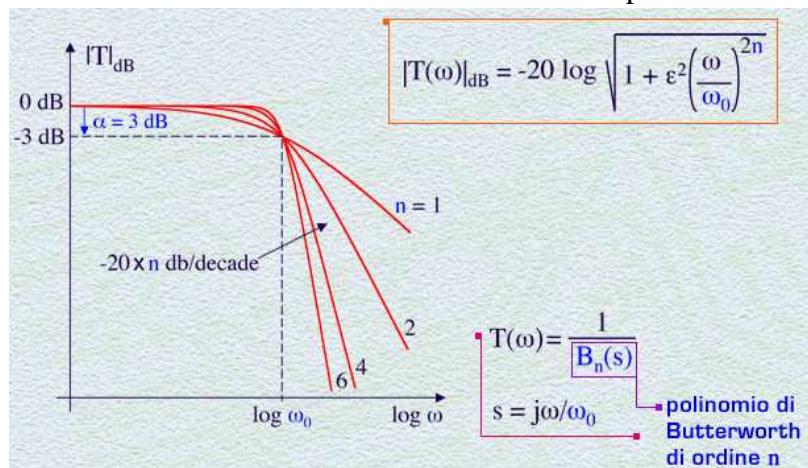
Nella parte conclusiva di questa lezione, abbiamo quindi visto, le caratteristiche principali dei filtri Butterworth.

La prossima lezione sarà dedicata a mostrare delle configurazioni circuitali che consentono di implementare questa classe di filtri.

## Lezione n. 36 : Filtri Butterworth

### Descrizione dei filtri tramite i Polinomi di Butterworth

Dopo aver introdotto nella scorsa lezione la caratteristica risposta in frequenza di un filtro di tipo Butterworth, vediamo ora come procedere alla sua realizzazione pratica. Nella figura viene riproposto il diagramma di Bode del modulo della funzione di trasferimento di un filtro passa-basso di Butterworth. Rispetto alla scorsa lezione, questa volta il grafico è mostrato per comodità in scala bilogaritmica con il modulo di  $T$  in decibel tracciato rispetto al logaritmo della pulsazione. Le curve sono tracciate per diversi valori dell'ordine del filtro, per uno stesso fattore di attenuazione  $\alpha$  nella banda passante pari a 3 dB. Dal grafico possiamo osservare che per frequenze superiori alla frequenza di taglio  $f_0$ , il modulo decresce all'aumentare della frequenza con una pendenza costante che è pari a  $-20$  per  $n$ , l'ordine del filtro, dB per decade. Questo andamento è anche deducibile dall'espressione del modulo di  $T$ , riportata in altro a destra nella figura ed espressa in funzione di  $\omega$ , se la consideriamo per frequenze superiori alla frequenza di taglio, in cui il termine  $\epsilon^2(\omega/\omega_0)^{2n}$  è dominante rispetto ad uno.



Al fine di utilizzare configurazioni circuituali che consentano di implementare un filtro Butterworth descritto dall'espressione della funzione di trasferimento che abbiamo visto finora, è necessario determinare un'espressione equivalente per  $T(\omega)$  che possa essere confrontata direttamente con le espressioni delle funzioni di trasferimento deducibili dai circuiti con il calcolo delle reti elettriche. Senza inoltrarci nel modo con cui può essere calcolata questa espressione equivalente, ci limitiamo a riportarne il risultato per poi invece concentrare la nostra attenzione su come utilizzare alcune configurazioni circuituali per implementare questa funzione di trasferimento.

Procedendo con questo approccio quindi, esprimiamo  $T(\omega)$ , in maniera equivalente, come uno diviso per un polinomio di ordine  $n$ ,  $B_n(s)$ , espresso in funzione della variabile  $s = j\omega/\omega_0$ . Il polinomio  $B_n(s)$  prende il nome di polinomio di Butterworth di ordine  $n$ . La variabile complessa  $s$  è analoga alla variabile complessa  $s$  finora utilizzata per la descrizione della risposta in frequenza dei circuiti, con la differenza che in questo caso è stata direttamente normalizzata rispetto alla pulsazione di taglio del filtro  $\omega_0$ .

I polinomi di Butterworth sono stati calcolati per i diversi ordini del filtro.

Riportiamo qui la tabella con le espressioni dei polinomi di Butterworth per filtri passa-basso di ordine fino a 5, e tutti caratterizzati da un'attenuazione della funzione di trasferimento alla frequenza di taglio  $f_0$  pari a 3 dB. Possiamo notare che i polinomi di ordine superiore al secondo sono espressi come prodotti di polinomi del I e del II ordine. Ad esempio, il polinomio del IV ordine è ottenuto come prodotto di due polinomi del II ordine, mentre quello del V ordine come prodotto di un polinomio del I ordine e di due del II ordine. Solo i coefficienti numerici dei polinomi del II ordine utilizzati nelle espressioni di polinomi di ordine superiore sono differenti da caso a caso. Da quanto visto possiamo quindi intuire che una volta che saremo in grado di utilizzare due configurazioni circuituali di base, una per implementare un polinomio del I ordine ed una per un polinomio del II ordine, sarà sufficiente utilizzare in cascata una sequenza opportuna di questi circuiti per ottenere filtri di Butterworth di ordine superiore al secondo.

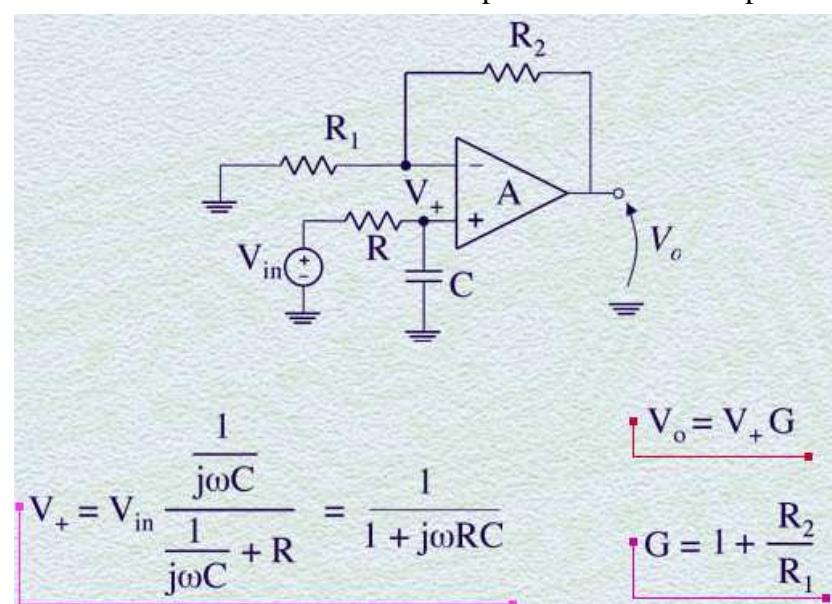
Riprenderemo meglio questo punto più avanti.

## Realizzazione dei Filtri passa-basso del I e II ordine

Consideriamo innanzitutto una configurazione circuitale che consente di implementare un filtro passa-basso di Butterworth del I ordine. Lo schema del circuito è mostrato nella figura. Esso consiste in una configurazione non invertente in cui il generatore in ingresso è applicato tramite una partizione con una resistenza  $R$  e una capacità  $C$ . La tensione  $V_+$  al morsetto non invertente dell'amplificatore operazionale è uguale alla tensione di ingresso  $V_{in}$  per  $1/j\omega C$ , diviso per  $(1/j\omega C + R)$  e quindi uguale a  $1/(1+j\omega RC)$ . La tensione di uscita  $V_o$  è uguale alla tensione  $V_+$  moltiplicata per il guadagno  $G$  della configurazione non invertente. Ricordiamo che  $G$  è uguale ad  $1 + R_2/R_1$ . La funzione di trasferimento del circuito è quindi uguale a  $(1 + R_2/R_1)$ , diviso  $(1+j\omega RC)$ .

$$\begin{aligned} T(j\omega) &= \left(1 + \frac{R_2}{R_1}\right) \frac{1}{1 + j\omega RC} \\ &= \left(1 + \frac{R_2}{R_1}\right) \frac{1}{1 + j\frac{\omega}{\omega_0}} \\ &= \left(1 + \frac{R_2}{R_1}\right) \frac{1}{1 + s} \end{aligned}$$

definendo  
 $\omega_0 = 1/RC$   
e introducendo  
 $s = j\frac{\omega}{\omega_0}$



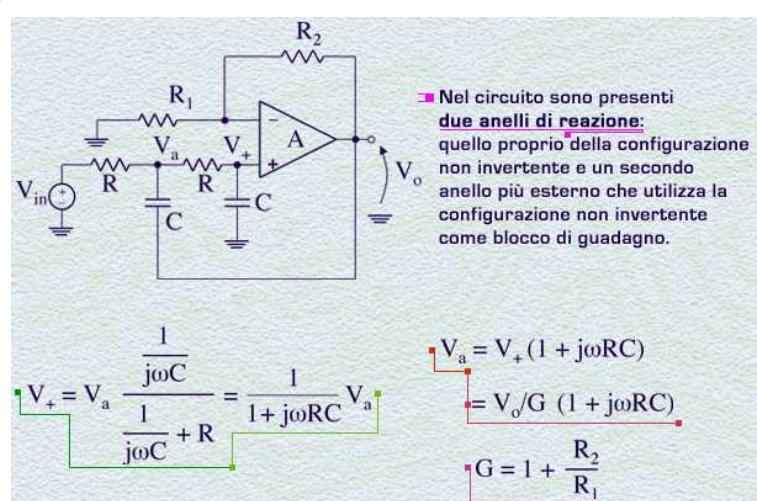
Definendo la pulsazione di taglio  $\omega_0$  come  $1/RC$ , l'espressione diventa  $(1+R_2/R_1)$  diviso  $(1+j\omega/\omega_0)$  e introducendo in questa la variabile  $s = j\omega/\omega_0$ , la funzione di trasferimento  $T$  diventa pari a  $(1 + R_2/R_1)$  diviso  $(1+s)$ . In conclusione, riconosciamo che il circuito considerato implementa un filtro di Butterworth del I ordine, descritto dal polinomio di I grado  $(1+s)$ , in accordo con la tabella vista in precedenza. Occorre precisare che, a differenza di quanto visto finora, in cui i filtri sono stati considerati con guadagno a bassa frequenza pari a 1,

che il filtro ottenuto possiede un guadagno in continua pari a  $(1 + R_2/R_1)$ .

Consideriamo ora un circuito che consente di implementare un filtro di Butterworth del II ordine.

Lo schema del circuito è mostrato nella figura. Possiamo individuare nel circuito un nucleo centrale che consiste nel filtro del I ordine appena visto al cui ingresso  $V_a$  viene applicato il segnale di ingresso  $V_{in}$  tramite un'ulteriore rete di partizione  $R-C$ . Il secondo estremo della capacità  $C$  è collegato direttamente all'uscita dell'amplificatore. Nel circuito sono quindi presenti due anelli di reazione, uno è quello proprio della configurazione non invertente, un secondo anello più esterno utilizza la configurazione non invertente come blocco di guadagno.

Procediamo ora al calcolo della funzione di trasferimento. Analogamente a quanto visto in precedenza, la tensione  $V_+$  al morsetto non invertente dell'amplificatore operazionale è uguale alla tensione di ingresso  $V_a$  per  $1/j\omega C$ , diviso per  $(1/j\omega C + R)$  e quindi uguale a  $V_a$  per  $1/(1+j\omega RC)$ . La tensione  $V_a$  è quindi esprimibile come  $V_+$  per  $(1+j\omega RC)$ , ed è uguale alla tensione di uscita  $V_o$ , divisa per il guadagno  $G$ .



della configurazione non invertente e moltiplicata per  $(1+j\omega RC)$ . Il guadagno  $G$  è sempre uguale a  $1+R_2/R_1$ .

La corrente  $i_a$ , indicata nello schema del circuito, è pari alla differenza di tensione  $V_a - V_o$  diviso per l'impedenza della capacità  $1/j\omega C$ , e quindi uguale a  $j\omega C$  per  $(V_a - V_o)$ .

Sostituendo l'espressione precedentemente trovata per  $V_a$  si ottiene che  $i_a$  è uguale a  $j\omega C$  che moltiplica  $(V_o/G + j\omega RC \cdot V_o/G - V_o)$ . La corrente  $i_b$ , indicata in figura, è uguale alla tensione  $V_+$  diviso l'impedenza della capacità  $1/j\omega C$  ed è quindi uguale a  $j\omega CV_+$  ovvero a  $j\omega C \cdot V_o/G$ .

Possiamo, a questo punto, calcolare la corrente  $i_c$  come somma di  $i_a + i_b$ . Utilizzando le espressioni ricavate per

queste ultime correnti, otteniamo che  $i_c$  è uguale a  $-\omega^2 C^2 R \cdot V_o/G + j\omega C \cdot (2V_o/G - V_o)$ . Guardando il ramo di ingresso, la corrente  $i_c$  è anche uguale alla differenza tra  $V_{in}$  e  $V_a$ , diviso  $R$ . Inserendo in quest'ultima relazione l'espressione di  $V_a$  rispetto a  $V_o$ , otteniamo che  $V_{in}/R - (V_o/GR) \cdot (1+j\omega RC)$  è uguale, per l'equivalenza appena vista, a  $-\omega^2 C^2 R \cdot V_o/G + j\omega C \cdot (2V_o/G - V_o)$ . Abbiamo così ottenuto un'equivalenza in cui compaiono ora solo le variabili  $V_{in}$  e  $V_o$ . Possiamo così determinare la funzione di trasferimento del circuito. Dopo alcune semplificazioni algebriche, si ottiene che la funzione di trasferimento  $T(j\omega)$ , uguale a  $V_o/V_{in}$  è pari a  $G$  diviso  $[-\omega^2 C^2 R^2 + j\omega CR \cdot (3-G) + 1]$ .

Analogamente a quanto fatto prima, definendo la pulsazione di taglio  $\omega_0$  come  $1/RC$  e sostituendo la variabile complessa  $s = j\omega/\omega_0$ , si ottiene che  $T$  è uguale a  $G/(s^2 + s(3-G) + 1)$ .

Possiamo notare, in conclusione, che il circuito considerato implementa un filtro di Butterworth del II ordine, descritto da un polinomio di II grado in  $s$ . Affinché vi sia una corrispondenza esatta con il polinomio di Butterworth che risulta pari, secondo la tabella mostrata in precedenza, a  $s^2 + 1,414s + 1$ , occorre che il coefficiente del termine in  $s$ ,  $3-G$ , sia posto uguale a 1,414. Questo requisito pone quindi una condizione sul guadagno non invertente  $G$  del circuito che deve essere posto uguale a 1,586, mediante un'opportuna scelta dei valori delle resistenze  $R_2$  e  $R_1$ , il cui rapporto deve essere uguale a 0,586.

### Filtri Passa-basso di ordine qualsiasi.

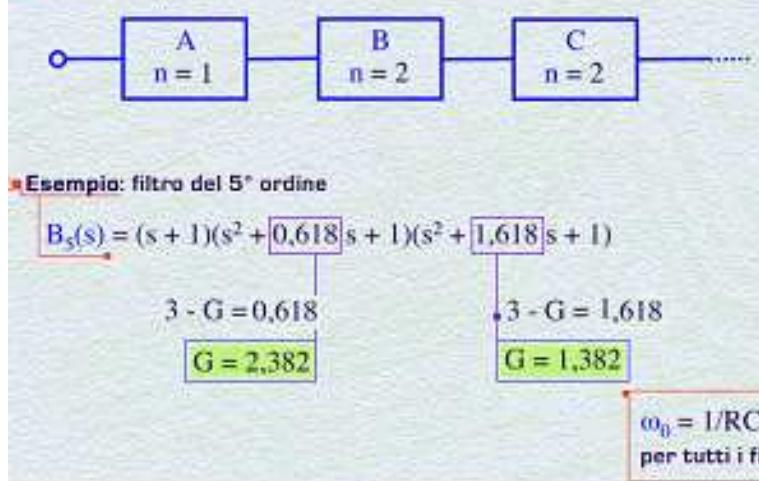
Una volta che abbiamo individuato due configurazioni circuitali con le quali è possibile implementare filtri di Butterworth del I e del II ordine, è facile realizzare anche filtri di Butterworth di ordine superiore. E' sufficiente infatti mettere in cascata un numero necessario di filtri del I o del II ordine fino a raggiungere l'ordine richiesto del filtro complessivo.

Nel caso si consideri un filtro di ordine pari, occorre utilizzare in cascata più filtri del II ordine. Per ognuno di questi blocchi il guadagno  $G$  deve essere progettato in maniera tale da ottenere il polinomio del II ordine corrispondente a quello richiesto nell'espressione del Polinomio di Butterworth del filtro complessivo.

Vediamo un esempio. Si supponga di voler implementare un filtro di Butterworth del IV ordine. In questo caso, occorre mettere in cascata due filtri del II ordine ognuno dei quali implementi uno dei polinomi di II grado di cui è composta la risposta complessiva  $B_4(s)$ . Il primo polinomio è pari a  $s^2 + 0,77s + 1$ . Utilizzando un circuito del II ordine visto in precedenza, occorre che il guadagno  $G$  soddisfi la condizione  $3 - G = 0,77$  e quindi che  $G$  sia pari a 2,23. Per quanto riguarda il secondo polinomio, uguale a  $s^2 + 1,85s + 1$ , occorre che il guadagno  $G$  soddisfi la condizione  $3 - G = 1,85$  e quindi che  $G$  sia uguale a 1,15. Per entrambi i circuiti, la pulsazione di taglio  $\omega_0$  deve essere uguale a  $1/RC$ . In conclusione, per realizzare il filtro richiesto, possono essere utilizzati due circuiti con stessa frequenza di taglio e con un diverso guadagno non invertente  $G$ , che deve soddisfare le condizioni indicate.

In maniera analoga, è possibile anche realizzare filtri di ordine dispari. In questo caso occorrerà utilizzare un filtro del I ordine ed un numero adeguato di filtri del II ordine. Consideriamo ad esempio

#### Ordine n dispari

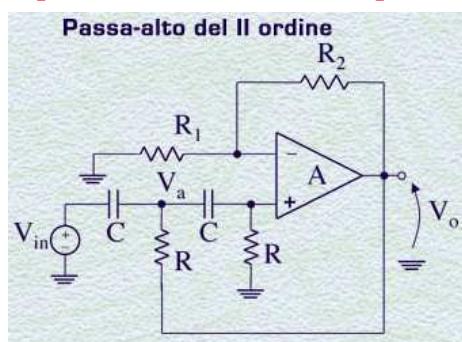


$G$  del secondo circuito del II ordine sia posto uguale a 1,382. Anche in questo caso, la frequenza di taglio di tutti e tre i filtri elementari dovrà essere la stessa.

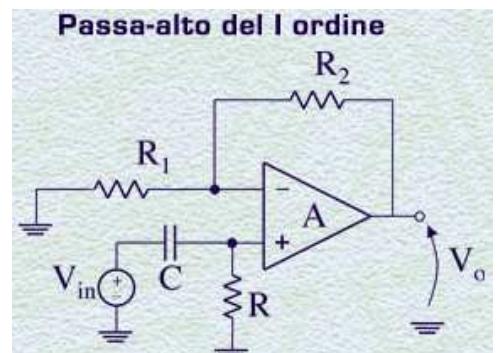
### Filtri passa-alto, passa-banda e arresta-banda

Una volta compreso come realizzare filtri Butterworth di tipo passa-basso, è possibile realizzare anche filtri di tipo passa-alto. Si può facilmente verificare, infatti che

*funzioni di trasferimento di tipo passa-alto dotate di caratteristiche di attenuazione nella banda arrestata analoghe a quelle viste per filtri passa-basso, possono essere ottenute dalle configurazioni viste, semplicemente invertendo la posizione relativa delle capacità C con le resistenze R.*

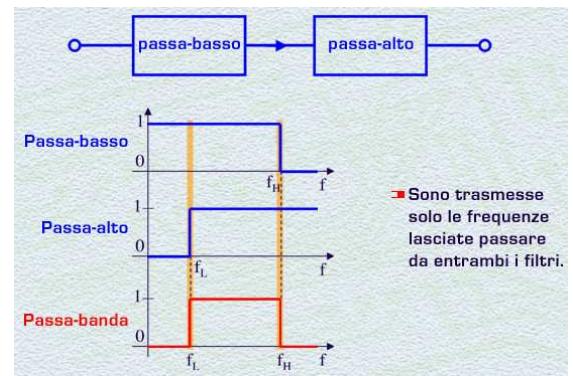
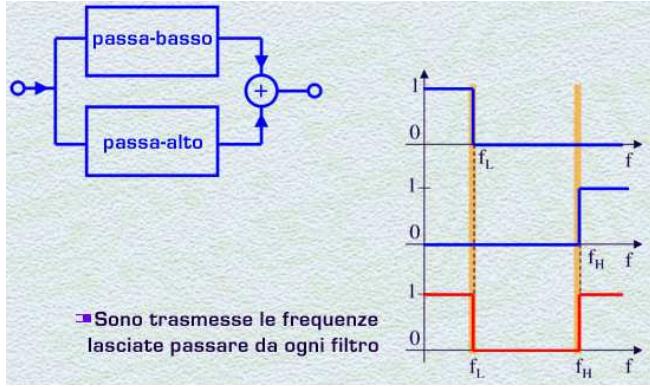


Nella figura è mostrato un filtro di Butterworth del I ordine ottenuto dalla configurazione vista in precedenza semplicemente mettendo la capacità  $C$  al posto della resistenza  $R$  e viceversa. Analogamente, anche un filtro passa-alto del II ordine può essere facilmente ottenuto dalla configurazione passa-basso del II ordine vista in precedenza, invertendo le posizioni delle capacità  $C$  con le resistenze  $R$ .



Un filtro di tipo passa-banda può essere ottenuto ponendo in cascata due filtri, uno di tipo passa-basso e uno di tipo passa-alto. Infatti, ponendo in sequenza i due filtri, solo le frequenze lasciate passare da entrambi verranno trasmesse dal filtro complessivo. Perchè ciò avvenga, occorre, come mostrato in figura, che la frequenza di taglio  $f_H$  del filtro passa-basso, sia maggiore della frequenza di taglio  $f_L$  del filtro passa-alto. In questa maniera solo le frequenze intermedie, comprese tra  $f_L$  e  $f_H$  verranno lasciate passare inalterate da entrambi i filtri.

Per realizzare un filtro arresta-banda, occorre invece utilizzare due filtri passa-basso e passa-alto a cui venga applicato in parallelo il segnale in ingresso e le cui uscite vengano sommate in un'unica uscita, mediante un circuito sommatore. L'uscita rappresenta così la somma delle sinusoidi lasciate passare da ogni singolo filtro. Per realizzare un filtro arresta banda, occorrerà in questo caso che la frequenza di taglio  $f_L$  del passa-basso sia inferiore della frequenza di taglio  $f_H$  del passa-alto. In questa maniera, esiste un intervallo di frequenza, tra  $f_L$  e  $f_H$ , in cui entrambi i filtri non lasciano passare nulla e quindi nessuna sinusode verrà trasmessa all'uscita del circuito. Negli altri intervalli, ogni filtro trasmetterà all'uscita le sinusoidi che appartengono alla propria banda passante,



andando così a comporre la risposta in frequenza del filtro arresta-banda mostrata in figura.

## Lezione n. 37 : La stabilità

### La stabilità nei circuiti controreazionati

Consideriamo lo schema generale di un circuito reazionato negativamente che avevamo introdotto nella lezione specifica sulla controreazione. Consideriamo ora la dipendenza dalla frequenza dei blocchi componenti il circuito indicando con  $A(s)$  la funzione di trasferimento del blocco di andata e con  $H(s)$  la funzione di trasferimento del blocco di reazione.

Introducendo la dipendenza dalla frequenza di  $A$  e di  $H$  nella relazione generale del guadagno del circuito controreazionato, si ottiene che  $G(s)$  è uguale a  $A(s)/[1+A(s)H(s)]$ . **La dipendenza dalla frequenza del guadagno  $G(s)$  può portare il circuito in un regime di funzionamento instabile, e cioè con un andamento del segnale di uscita divergente, indipendente dal segnale applicato in ingresso.**

Il problema della stabilità in un circuito controreazionato è un argomento piuttosto esteso che non può essere esaurito in maniera esaustiva nell'ambito di una sola lezione. Ci limitiamo in questa lezione a determinare in maniera per lo più intuitiva il **criterio di stabilità per un circuito controreazionato**, che è basato sull'analisi in frequenza del guadagno d'anello.

Consideriamo il guadagno d'anello del circuito Gloop(s) che è uguale a  $-A(s) \cdot H(s)$ . Valutando la

$$G_{\text{loop}}(s) = -A(s)H(s)$$

$$s = j\omega$$

$$G_{\text{loop}}(j\omega) = -A(j\omega)H(j\omega) = -|A(j\omega)H(j\omega)| e^{j\angle A(j\omega)H(j\omega)}$$

Supponendo, per semplicità, che  $A(j\omega)$  e  $H(j\omega)$  siano costanti a bassa frequenza (poli e zeri intervengono solo all'aumentare della frequenza)

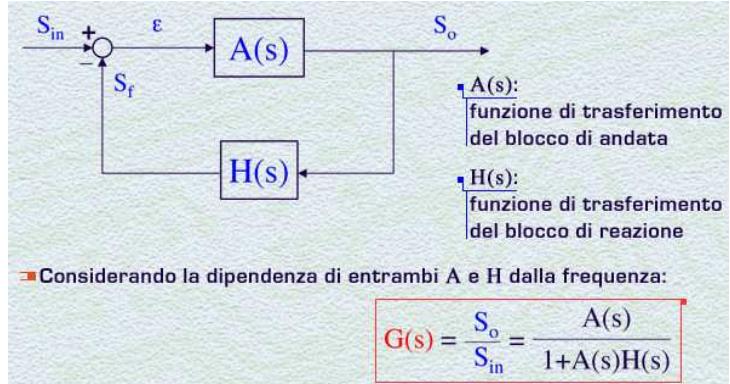
e valutiamo come la **stabilità o l'instabilità** del circuito sia determinata dall'**andamento di  $A(j\omega)H(j\omega)$  all'aumentare della frequenza**.

a bassa frequenza e che invece possiedano poli e zeri che esercitino il loro effetto di sfasamento all'aumentare della frequenza.

Esaminando ora l'andamento con la frequenza del modulo e della fase di  $A(j\omega) \cdot H(j\omega)$ , è possibile comprendere se il circuito ha un comportamento stabile o instabile.

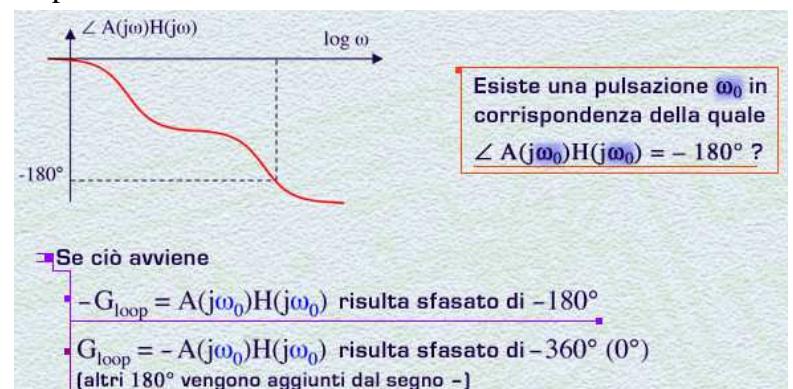
Consideriamo un diagramma della fase del prodotto  $A(j\omega) \cdot H(j\omega)$ , come quello mostrato in figura. Tipicamente, la fase parte da zero per portarsi verso valori sempre più negativi, a causa della

presenza dei poli nella rete, in numero superiore ad eventuali zeri. Dobbiamo osservare in questo diagramma se esiste una pulsazione  $\omega_0$  in corrispondenza della quale si ha che la  $\angle A(j\omega_0)H(j\omega_0) = -180^\circ$ . Se ciò avviene allora possiamo dire che in corrispondenza di questa pulsazione, il valore negato del guadagno d'anello cioè  $A(j\omega_0)H(j\omega_0)$  risulta sfasato di  $-180^\circ$ , e cioè è un numero reale negativo. In maniera del tutto equivalente possiamo anche dire che il Gloop in corrispondenza della pulsazione  $\omega_0$ , pari a  $-A(j\omega_0)H(j\omega_0)$  risulta sfasato di  $-360^\circ$  o equivalentemente di  $0^\circ$ , e cioè risulta un numero



dipendenza dalla frequenza del Gloop con la solita sostituzione  $s=j\omega$ , questo risulta uguale a  $-A(j\omega) \cdot H(j\omega)$ . Dato che il Gloop è un numero complesso, può essere anche espresso con la notazione che ne evidenzia modulo e fase. Il Gloop risulta così uguale a  $-|A(j\omega) \cdot H(j\omega)| \cdot e^{j\angle A(j\omega)H(j\omega)}$ .

Per comprendere meglio il ruolo del modulo e della fase del guadagno d'anello sulla stabilità del circuito, supponiamo per semplicità che  $A(j\omega)$  e  $H(j\omega)$  siano pari ad un valore costante



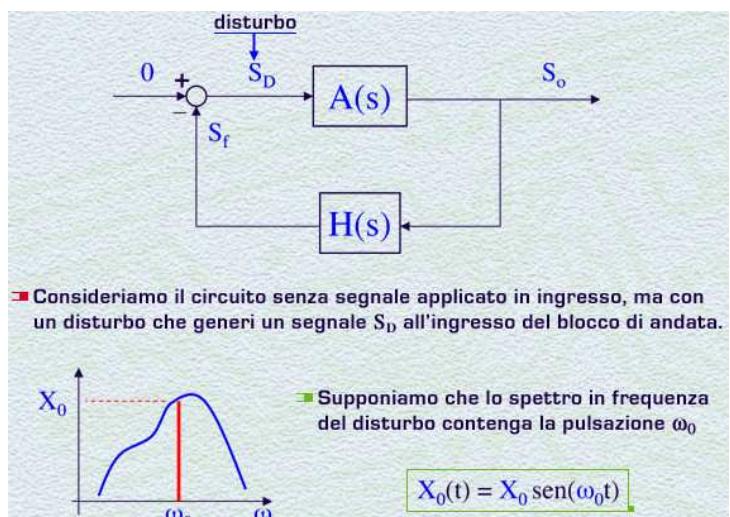
reale positivo. Questo poiché la presenza del segno meno aggiunge ai  $-180^\circ$  di  $A \cdot H$  un ulteriore sfasamento di  $-180^\circ$ .

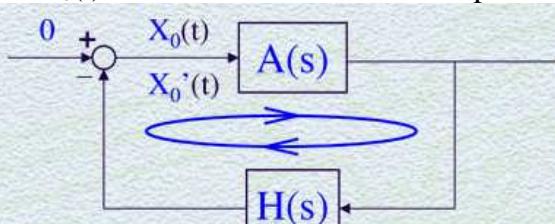
Questa condizione trovata in corrispondenza della pulsazione  $\omega_0$  assume un importante significato fisico e cioè che la reazione che era nata, a bassa frequenza, come reazione negativa, cioè caratterizzata da una sfasamento del segnale nel percorrere l'anello di  $-180^\circ$ , diventa

- La reazione da negativa diventa positiva.
- La sinusoide a  $\omega_0$  torna in fase con se stessa dopo aver percorso un giro dell'anello.

positiva a più alta frequenza, cioè la sinusoide di pulsazione  $\omega_0$  nel percorrere completamente l'anello ritorna in fase con se stessa.

Vediamo ora di approfondire il comportamento del circuito in presenza di questa sinusoide di pulsazione  $\omega_0$  in corrispondenza della quale la reazione è diventata positiva. Valutiamo questo problema in un caso pratico e cioè considerando il circuito senza segnale applicato in ingresso ma con invece la presenza di un disturbo introdotto dall'esterno, ad esempio in seguito all'accensione delle alimentazioni del circuito stesso o per effetto di un accoppiamento di un'interferenza elettromagnetica. L'introduzione del disturbo in un qualche punto dell'anello, genererà presto o tardi un segnale  $S_D$  all'ingresso del blocco di andata. Supponiamo che lo spettro in frequenza del disturbo abbia una banda sufficientemente ampia da contenere anche la pulsazione  $\omega_0$ , a cui ci siamo riferiti in precedenza, in corrispondenza della quale il termine  $A \cdot H$  risulta sfasato di  $-180^\circ$ . Esisterà quindi una sinusoide  $X_0(t)$  che circola nell'anello esprimibile come  $X_0 \cdot \sin(\omega_0 t)$ . Consideriamo ora questa sinusoide dopo un giro completo dell'anello. Valutiamo quindi la sinusoide  $X_0'(t)$  all'ingresso del blocco di andata, ottenuta dalla  $X_0(t)$  applicata allo stesso punto dopo aver percorso completamente un giro dell'anello. La  $X_0'(t)$  risulterà pari a  $X_0(t)$ , moltiplicata per il guadagno d'anello calcolato a quella pulsazione, pari a  $-A(j\omega_0) \cdot H(j\omega_0)$ . Come visto in precedenza, il termine  $-A \cdot H$  a questa frequenza risulta avere fase di  $-360^\circ$  cioè risulta un numero reale positivo e quindi è pari al modulo di  $A(j\omega_0) \cdot H(j\omega_0)$ . In conclusione, la sinusoide a pulsazione  $\omega_0$ ,





$$X_0'(t) = -A(j\omega_0)H(j\omega_0)X_0(t) = |A(j\omega_0)H(j\omega_0)| X_0(t)$$

sfasato di  $-360^\circ$

Dopo un giro dell'anello, la sinusoide a  $\omega_0$  ritorna in fase con ampiezza moltiplicata per  $|A(j\omega_0)H(j\omega_0)| = |G_{loop}(j\omega_0)|$

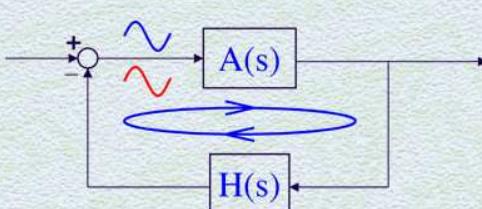
dopo aver percorso un giro dell'anello, ritorna in fase con l'ampiezza moltiplicata per il modulo di  $A(j\omega_0) \cdot H(j\omega_0)$  che è pari al modulo del guadagno d'anello a  $\omega_0$ .

Esaminiamo ora in tre diversi casi il propagarsi di questa sinusoide nel percorrere più volte l'anello. Si consideri, nel primo caso, che il modulo di  $A(j\omega_0) \cdot H(j\omega_0)$  sia inferiore ad uno. In questo caso l'ampiezza  $X_0'$  della sinusoide dopo aver percorso un giro dell'anello risulterà inferiore all'ampiezza  $X_0$  della sinusoide di partenza. La sinusoide ritorna quindi allo stesso punto con ampiezza ridotta. Risulta intuitivo pensare in questo caso che nel percorrere continuamente l'anello l'ampiezza della sinusoide si riduca sempre più e che quindi dopo diversi giri la sinusoide si smorzi completamente. In questo caso possiamo quindi concludere che nonostante il circuito presenti una condizione in cui la

reazione è positiva, questo non è sufficiente a far divergere l'uscita e quindi il circuito si può dire stabile.

Consideriamo ora un secondo caso in cui il modulo di  $A(j\omega_0) \cdot H(j\omega_0)$  sia pari a uno. In questo caso l'ampiezza  $X'_0$  della sinusoide dopo aver percorso un giro dell'anello risulterà uguale all'ampiezza  $X_0$  della sinusoide iniziale. La sinusoide ritorna quindi allo caso possiamo facilmente intuire che la condizione esaminata rappresenta una condizione di oscillazione per il circuito. Esiste, cioè, una

**2° caso**  $|A(j\omega_0)H(j\omega_0)| = 1 \rightarrow X'_0 = X_0$



■ La sinusoide torna allo stesso punto con uguale ampiezza e quindi si mantiene nel percorrere l'anello anche infinite volte.

#### Il circuito oscilla [con pulsazione $\omega_0$ ]

questo caso, che la sinusoide si rinforzerà man mano che percorrerà diversi giri dell'anello portando velocemente l'uscita del circuito a divergere verso valori prossimi alle tensioni di alimentazione. Il circuito si dice in questo caso instabile.

Alla luce di quanto visto finora, possiamo trarre una conclusione riguardo al criterio di stabilità di un circuito controllato. Anche se siamo giunti a questo criterio senza una trattazione teorica rigorosa, esso pur tuttavia risulta valido in maniera del tutto generale.

Possiamo riassumere questo criterio come segue.

Se dall'analisi della risposta in frequenza del guadagno d'anello

- **Il criterio di stabilità a cui siamo giunti ci fornisce una condizione sufficiente per la stabilità.**

■ Se esiste una pulsazione  $\omega_0$  in corrispondenza della quale

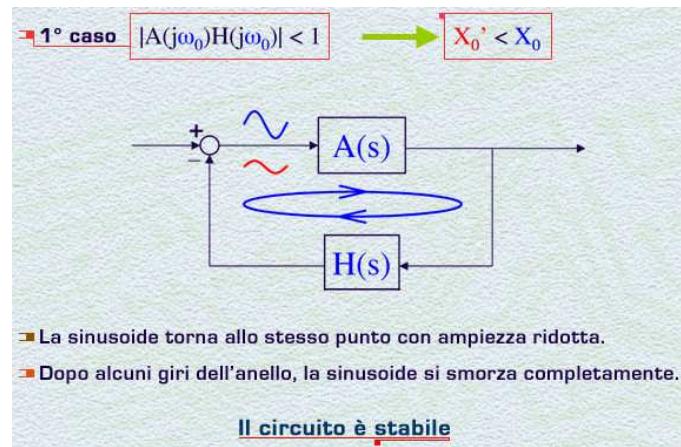
$$A(j\omega_0)H(j\omega_0) = -G_{loop} \text{ risulta sfasato di } -180^\circ$$

■ allora affinché il circuito sia **stabile** occorre che

$$|A(j\omega_0)H(j\omega_0)| < 1$$

■ Questo criterio può essere enunciato equivalentemente indicando la pulsazione  $\omega_0$  in corrispondenza della quale il  $G_{loop}$  risulti sfasato di  $-360^\circ$ .

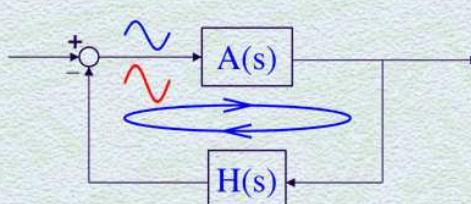
E' da osservare, a puro titolo formale, che questo criterio può essere enunciato equivalentemente indicando la pulsazione  $\omega_0$  in corrispondenza della quale il  $G_{loop}$  risulti sfasato di  $-360^\circ$  piuttosto



sinusoide a pulsazione  $\omega_0$  che una volta introdotta nel circuito, ad esempio in seguito all'accoppiamento di un disturbo, si mantiene anche nel percorrere l'anello un numero infinito di volte.

Consideriamo ora il terzo ed ultimo caso, in cui il modulo di  $A(j\omega_0) \cdot H(j\omega_0)$  sia maggiore di uno. In questo caso l'ampiezza  $X'_0$  della sinusoide dopo aver percorso un giro dell'anello risulterà maggiore dell'ampiezza  $X_0$  della sinusoide iniziale. La sinusoide ritorna quindi allo stesso punto con un'ampiezza maggiore. E' chiaro, in

**3° caso**  $|A(j\omega_0)H(j\omega_0)| > 1 \rightarrow X'_0 > X_0$



■ La sinusoide torna allo stesso punto con una più elevata ampiezza. Dopo alcuni giri dell'anello, l'oscillazione diverge portando il circuito alla saturazione.

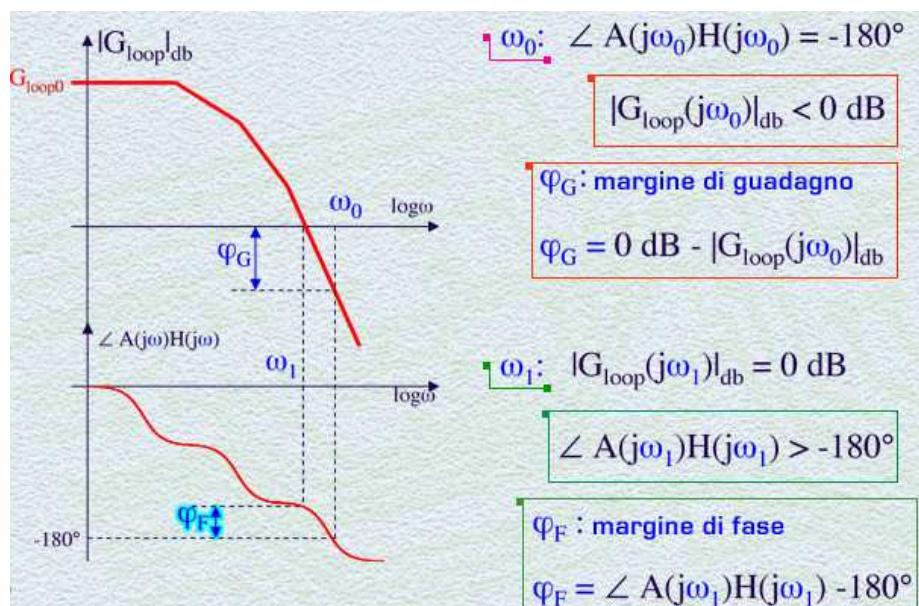
#### Il circuito è instabile

del circuito si osserva che esiste una pulsazione  $\omega_0$  in corrispondenza della quale il termine  $A(j\omega_0) \cdot H(j\omega_0)$  pari a  $-G_{loop}$  risulti sfasato di  $-180^\circ$ , allora affinché il circuito sia stabile occorre che il modulo di  $A(j\omega_0) \cdot H(j\omega_0)$ , pari al modulo del  $G_{loop}$ , sia inferiore a 1. Questa condizione corrisponde alla situazione esaminata precedentemente nel 1° caso, in cui qualsiasi oscillazione a pulsazione  $\omega_0$  instaurata nel circuito tende naturalmente a smorzarsi nel percorrere l'anello.

che, come indicato, il termine A·H, pari a meno Gloop, sia sfasato di  $-180^\circ$ . Le due notazioni sono del tutto equivalenti nel significato. In seguito verrà utilizzata, per comodità, la notazione con lo sfasamento del termine A·H cioè del  $-Gloop$ , in quanto questa notazione consente di porre l'attenzione sul solo sfasamento introdotto dalle singolarità del circuito (poli e zeri), non considerando a priori lo sfasamento di  $-180^\circ$  che si ha sempre in quanto il Gloop, come abbiamo visto tante volte in circuiti contoreazionati, parte a bassa frequenza come numero reale negativo.

Vediamo ora come verificare il criterio di stabilità dato un generico circuito contoreazionato. Occorre innanzitutto tracciare il diagramma di Bode del modulo del Gloop e della fase di  $-Gloop$  cioè del termine A·H. Osservando il diagramma della fase, occorre valutare se esiste una pulsazione  $\omega_0$  in corrispondenza della quale la fase è pari a  $-180^\circ$ . Se ciò avviene, come nel caso mostrato in figura, occorre allora verificare se in corrispondenza di tale pulsazione, il Gloop è inferiore ai zero dB, cioè graficamente se il diagramma del Gloop in corrispondenza di  $\omega_0$  si trova sotto l'asse delle ascisse. Se così è, come nel caso mostrato in figura, allora possiamo dire che il circuito è stabile, altrimenti il circuito è instabile. **Nel caso il circuito sia stabile**, è possibile anche definire un parametro di merito  $\varphi_G$ , o margine di guadagno che è definito come la differenza tra il modulo del Gloop al limite della stabilità, e cioè 0 dB, ed il modulo del Gloop nella situazione specifica, cioè modulo di  $Gloop(j\omega_0)$ . Il margine di guadagno indica, in sostanza, di quanto potrebbe aumentare il modulo del guadagno d'anello, mantenendo comunque il circuito in condizione di stabilità.

Il criterio di stabilità può essere verificato anche in un'altra maniera equivalente. Osservando il diagramma del modulo del guadagno d'anello, si può determinare la pulsazione  $\omega_1$  in corrispondenza della quale il modulo risulta pari a 0 dB, cioè il diagramma interseca l'asse delle ascisse. Per verificare la stabilità del circuito, occorre ora valutare la fase in corrispondenza



della pulsazione  $\omega_1$ . Se la fase per  $\omega_1$  è meno negativa di  $-180^\circ$ , come nel caso mostrato in figura, allora il circuito è stabile. Possiamo infatti osservare che se la fase alla pulsazione  $\omega_1$  non ha ancora raggiunto i  $-180^\circ$ , allora questo implica indirettamente che tale sfasamento limite verrà raggiunto ad una pulsazione  $\omega_0$  maggiore di  $\omega_1$ , in corrispondenza della quale il modulo del Gloop, che era pari a 0 dB a  $\omega_1$ , sarà necessariamente inferiore ai 0 dB. Anche in questa maniera, se vogliamo indiretta rispetto al criterio enunciato, è quindi possibile verificare la stabilità del circuito. La differenza tra la fase determinata alla frequenza  $\omega_1$  e il valore limite per la stabilità,  $-180^\circ$ , viene definita come **margine di fase**. Il margine di fase indica, pertanto, di quanto potrebbe diventare più negativa la fase alla pulsazione  $\omega_1$ , in cui il modulo del Gloop è uguale a 0 dB, senza che il circuito raggiunga il limite dell'instabilità.

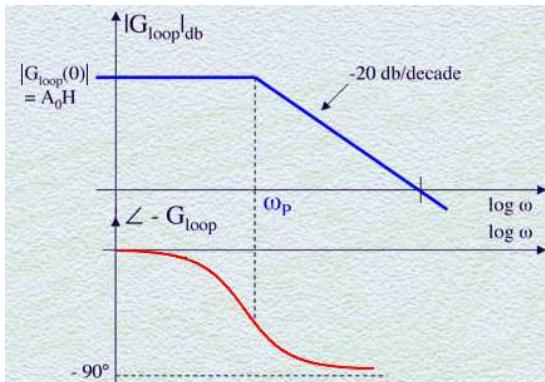
### Stabilità in amplificatori con uno o più poli

In questa parte della lezione, analizziamo la stabilità di amplificatori caratterizzati da un diverso numero di poli. Iniziamo con il considerare gli amplificatori ad un polo. Si tratta, ad esempio, degli amplificatori invertente e non invertente in cui abbiamo considerato, nelle lezioni passate, la presenza di un polo nella risposta in frequenza dell'amplificatore operazionale.

Ricordiamo che se consideriamo il blocco di andata  $A(s)$  è caratterizzato da un singolo polo di pulsazione  $\omega_p$  ed il blocco di reazione invece di guadagno costante  $H$ , il guadagno si può esprimere come  $G(s) = A_0/(1+A_0H)$  per  $1/\{1 + s/[\omega_p(1+A_0H)]\}$ .

Se determiniamo ora il guadagno d'anello, questo risulta pari a  $-A_0H/(1+s/\omega_p)$ , e cioè il Gloop è caratterizzato dallo stesso polo del guadagno d'andata.

Tracciando il diagramma di Bode del modulo del Gloop e della fase di -Gloop, osserviamo che lo sfasamento di quest'ultimo risulta sempre meno negativo di  $-90^\circ$  e pertanto, qualsiasi sia il valore del modulo del guadagno d'anello, il circuito risulta sempre incondizionatamente stabile.



- Lo sfasamento massimo di  $-G_{\text{loop}}$  è meno negativo di  $-90^\circ$ .
- Il sistema è incondizionatamente stabile.

guadagno di andata, come indicato nella relazione mostrata in figura.

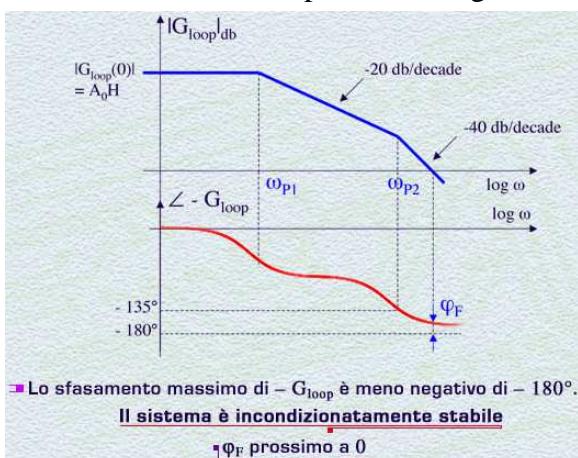
Se tracciamo il diagramma di Bode del modulo del Gloop e della fase di -Gloop, osserviamo che lo sfasamento risulta sempre meno negativo di  $-180^\circ$ . Anche in questo caso, quindi, il sistema risulta incondizionatamente stabile. Osserviamo, però, a differenza dell'amplificatore ad un polo, che il margine di fase è decisamente ridotto in questo caso rispetto al precedente. All'aumentare della frequenza, il margine di fase, pari a  $-45^\circ$  in corrispondenza della pulsazione  $\omega_{p2}$ , tende a diventare nullo. E' da aspettarsi, quindi, che l'aggiunta di un ulteriore polo nella risposta in frequenza del guadagno d'anello possa portare l'amplificatore ad operare non più in condizioni di stabilità.

Analizziamo ora il caso di un amplificatore in cui il blocco di andata sia caratterizzato dalla presenza di 2 poli, mentre il blocco di reazione abbia guadagno costante. Anche in questo caso, il guadagno d'anello acquisirà gli stessi due poli del

$$A(s) = \frac{A_0}{\left(1 + \frac{s}{\omega_{p1}}\right)\left(1 + \frac{s}{\omega_{p2}}\right)}$$

$$H(s) = H$$

$$G_{\text{loop}}(s) = -\frac{A_0 H}{\left(1 + \frac{s}{\omega_{p1}}\right)\left(1 + \frac{s}{\omega_{p2}}\right)}$$



- Lo sfasamento massimo di  $-G_{\text{loop}}$  è meno negativo di  $-180^\circ$ .
- Il sistema è incondizionatamente stabile
- $\varphi_F$  prossimo a 0

moltiplicato per il numero dei poli e quindi potenzialmente può essere più negativo del limite per la stabilità di  $-180^\circ$ .

Consideriamo, come esempio, un amplificatore il cui blocco di guadagno sia caratterizzato da tre poli, come indicato nella figura. Il guadagno in continua  $A_0$  è pari a 100 dB, mentre le frequenze dei tre poli sono rispettivamente a 100 kHz, 1 MHz e 10 MHz. Avendo la libertà di poter scegliere il guadagno  $H$  del blocco di reazione, studiamo la possibilità di ottenere un circuito contoreazionato stabile che utilizzi questo blocco di guadagno.

Lo sfasamento introdotto dai poli può rendere instabile il circuito.

Esempio: circuito con blocco di guadagno a tre poli

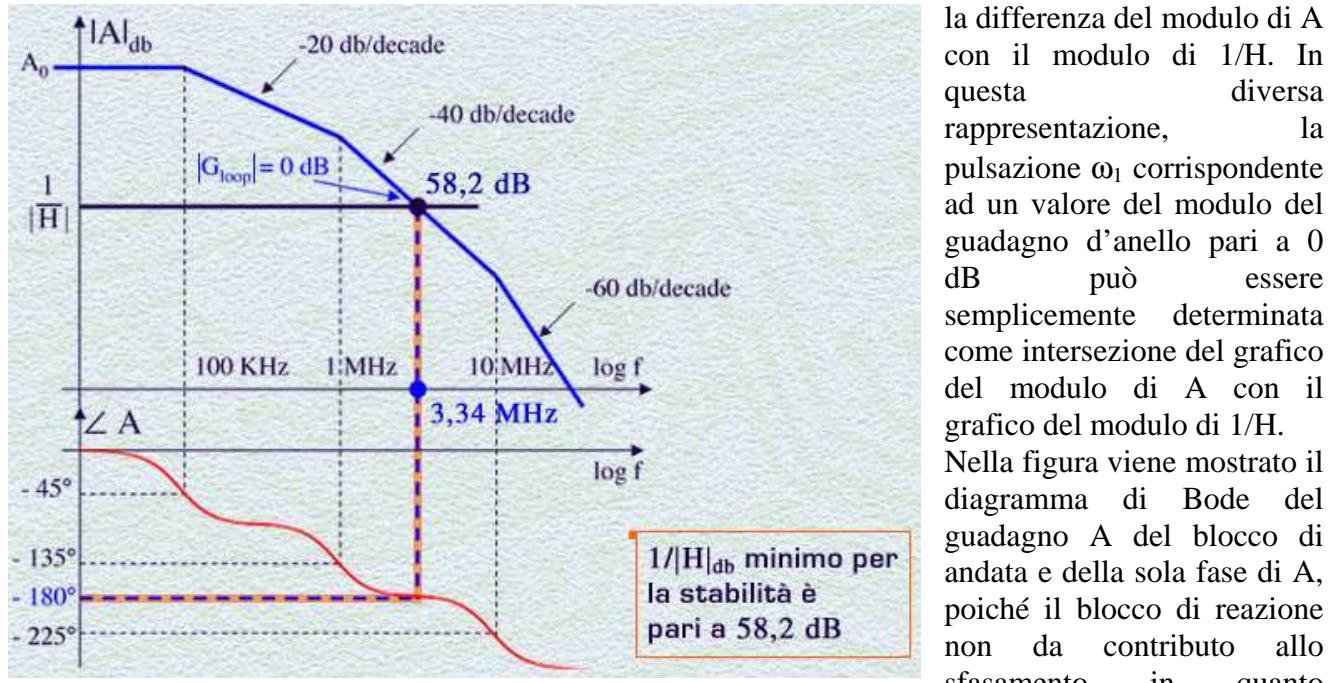
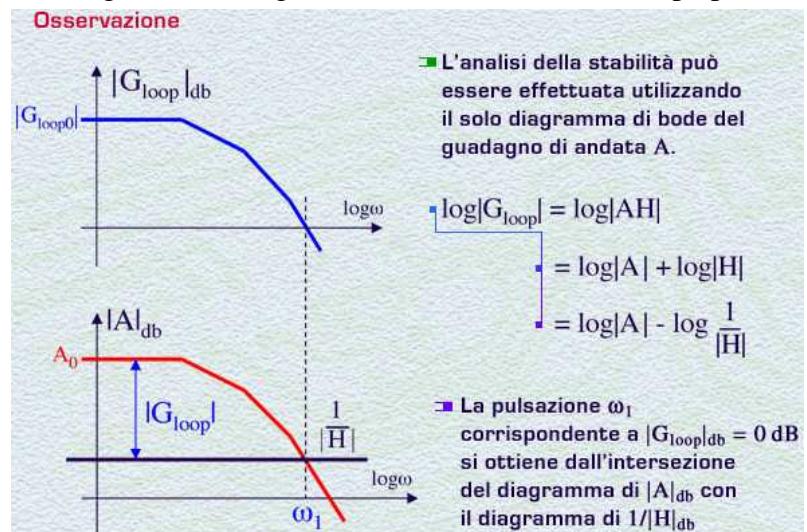
$$A(s) = \frac{A_0}{\left(1 + \frac{s}{\omega_{p1}}\right)\left(1 + \frac{s}{\omega_{p2}}\right)\left(1 + \frac{s}{\omega_{p3}}\right)}$$

- $A_0 = 100 \text{ dB}$
- $f_{p1} = 100 \text{ KHz}$
- $f_{p2} = 1 \text{ MHz}$
- $f_{p3} = 10 \text{ MHz}$

Potendo scegliere  $H$ , studiamo la possibilità di ottenere un circuito contoreazionato stabile con questo blocco di guadagno.

Prima di procedere, vediamo una metodo alternativo di analisi dei diagrammi di Bode, che ci consenta di trovare facilmente la soluzione al nostro problema. Finora il problema della stabilità è stato analizzato facendo uso del diagramma di Bode del guadagno d'anello. Osserviamo però, che la stessa analisi può essere effettuata utilizzando anche il solo diagramma di Bode del guadagno di andata A. Osserviamo infatti che per le proprietà dei logaritmi, il logaritmo del modulo del Gloop, pari al logaritmo del modulo di  $A \cdot H$ , risulta uguale alla somma del logaritmo del modulo di A + il logaritmo del modulo di H, e quindi può essere anche espresso come il logaritmo del modulo di A meno il logaritmo di 1 diviso il modulo di H. Quest'ultimo termine rappresenta il logaritmo del modulo del guadagno ideale  $1/H$  del circuito contoreazionato che può essere determinato con un'opportuna scelta del guadagno H.

Se ora consideriamo il diagramma del modulo del guadagno di andata A e sullo stesso grafico tracciamo la retta del guadagno ideale  $1/H$ , supposto costante nel nostro esempio, possiamo osservare che per la proprietà dei logaritmo appena vista, il modulo del Gloop può essere semplicemente individuato come

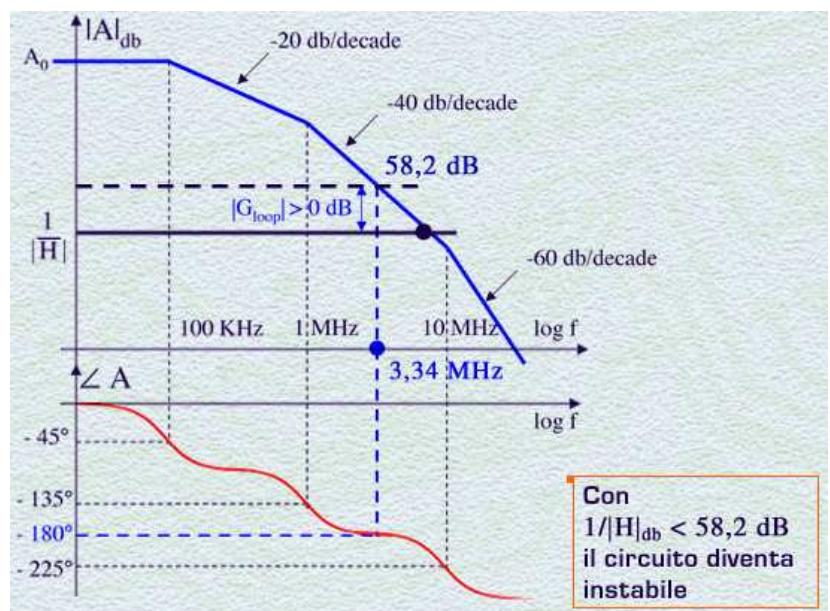


la differenza del modulo di A con il modulo di  $1/H$ . In questa diversa rappresentazione, la pulsazione  $\omega_1$  corrispondente ad un valore del modulo del guadagno d'anello pari a 0 dB può essere semplicemente determinata come intersezione del grafico del modulo di A con il grafico del modulo di  $1/H$ . Nella figura viene mostrato il diagramma di Bode del guadagno A del blocco di andata e della sola fase di A, poiché il blocco di reazione non da contributo allo sfasamento in quanto

supposto di guadagno costante positivo H. Nello stesso grafico del modulo di A è indicato il modulo del guadagno ideale  $1/H$ . Per quanto appena detto, in corrispondenza della frequenza di intersezione tra A e  $1/H$ , il modulo del Gloop è pari ad uno. Il diagramma della fase contiene alcune frequenze di riferimento calcolate secondo la posizione dei diversi poli. Ad esempio, alla frequenza di 1 MHz lo sfasamento introdotto è all'incirca pari a  $-135^\circ$ , somma dei  $-90^\circ$  introdotti dal polo a 100 kHz e dei  $-45^\circ$  introdotti dal polo a 1 MHz. Analogamente, alla frequenza di 10 MHz lo sfasamento introdotto dai tre poli è di  $-225^\circ$ , somma di  $-90^\circ$  del I polo,  $-90^\circ$  del II polo e  $-45^\circ$  del III polo. Esiste quindi una frequenza in corrispondenza della quale lo sfasamento è di  $-180^\circ$ . Facendo i conti utilizzando le funzioni arcotangente che forniscono lo sfasamento dei vari poli, questa frequenza risulta pari a 3,34 MHz. Se andiamo a determinare il valore del guadagno di andata A in corrispondenza di questa frequenza, questo risulta pari a 58,2 dB. E' evidente che affinché il circuito sia stabile occorre che l'intersezione della retta  $1/|H|$  con il modulo di A non avvenga ad una frequenza maggiore di quella determinata, o in altre parole, il modulo del guadagno ideale  $1/H$  non può essere inferiore a 58,2 dB.

Se la retta  $1/|H|$  fosse invece inferiore a 58,2 dB, alla frequenza di 3,34 MHz il guadagno d'anello risulterebbe superiore ad uno ed il circuito diventerebbe instabile.

In conclusione a questo esempio, osserviamo che dato un certo blocco di andata, la scelta di un opportuno valore del guadagno ideale deve tenere conto della stabilità del circuito. In particolare la stabilità tende ad essere garantita se risulta maggiore il guadagno ideale del circuito mentre tende ad essere compromessa se è minore il valore richiesto per il guadagno ideale.



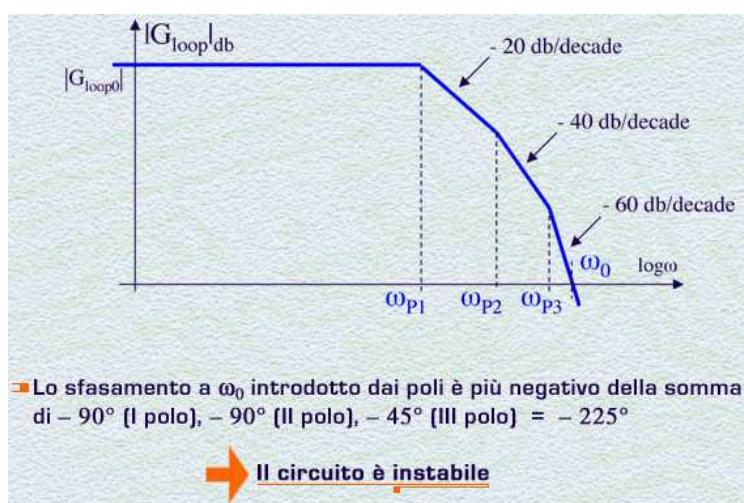
## Lezione n. 38 : Tecniche di compensazione

### Compensazione a polo dominante

Nella scorsa lezione abbiamo introdotto il problema della stabilità nei circuiti contoreazionati. Abbiamo in particolare visto come **amplificatori in cui il guadagno d'anello sia caratterizzato dalla presenza di più poli possano risultare facilmente instabili**. Un circuito instabile può essere stabilizzato utilizzando opportune tecniche che prendono il nome di tecniche di compensazione. Queste tecniche consentono, mediante la modifica di componenti stessi del circuito o tramite l'aggiunta di componenti o reti esterne, di riportare il circuito entro i parametri stabiliti dal criterio di stabilità. Inoltre, la compensazione viene effettuata in maniera tale da garantire un soddisfacente **margine di fase** (tipicamente superiore ai 45°), in maniera da mantenere soddisfatta la condizione di stabilità anche in presenza di possibili variazioni delle caratteristiche del circuito, ad esempio per una variazione della temperatura o il cambio di componente.

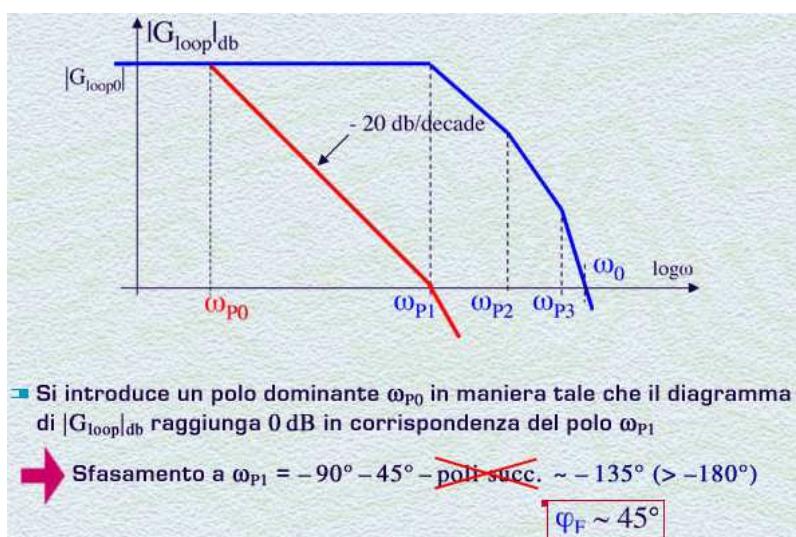
In questa lezione, introdurremo due delle tecniche di compensazione più utilizzate, la compensazione a polo dominante e la compensazione polo-zero.

Iniziamo con la compensazione a polo dominante. Consideriamo il diagramma di Bode del guadagno d'anello di un circuito, caratterizzato dalla presenza di tre poli. Supponiamo, come mostrato nella figura, che la pulsazione  $\omega_0$ , corrispondente all'intersezione del modulo del Gloop con l'asse delle ascisse a 0 dB sia maggiore della pulsazione di tutti e tre i poli. Anche senza effettuare conti dettagliati, è quindi possibile stimare che lo sfasamento introdotto dai tre poli sia più negativo della somma di  $-90^\circ$  (I polo),  $-90^\circ$  (II polo),  $-45^\circ$  (III polo) =  $-225^\circ$ .



Il terzo polo  $\omega_{p3}$  introduce uno sfasamento superiore a  $-45^\circ$  ad una pulsazione  $\omega_0$  alla sua destra nel diagramma di Bode. Lo sfasamento totale di  $-225^\circ$  è comunque più negativo di  $-180^\circ$  e pertanto, secondo il criterio di stabilità, il circuito è instabile.

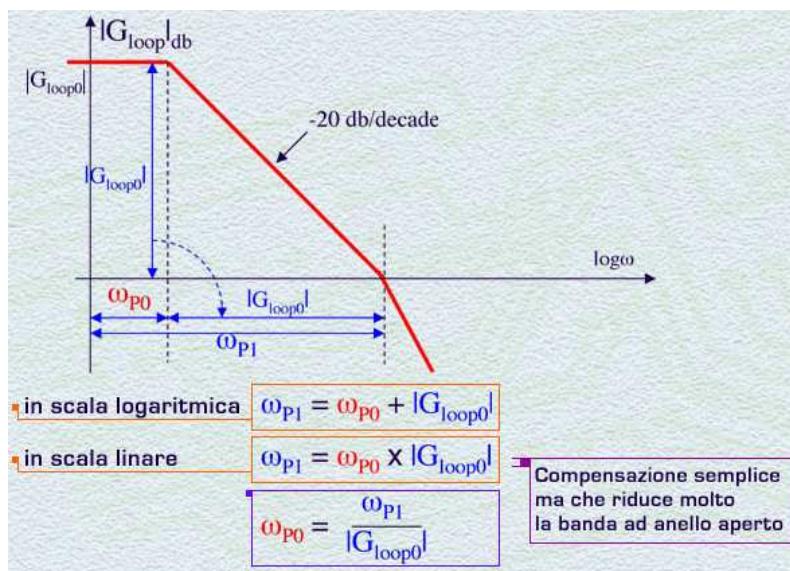
La compensazione a polo dominante consiste nell'introdurre nel cammino del guadagno d'anello un polo a pulsazione  $\omega_{p0}$ , inferiore alla pulsazione di tutti e tre i poli. In questa maniera, il diagramma del modulo del Gloop inizia a scendere con una pendenza di  $-20$  dB/decade a partire da questo polo a bassa frequenza, e raggiunge con questa pendenza l'asse a 0 dB in corrispondenza del polo  $\omega_{p1}$ . La frequenza  $\omega_{p1}$  diventa quindi la nuova pulsazione in corrispondenza della quale il modulo del Gloop risulta pari a 0 dB. Se ora stimiamo lo sfasamento in corrispondenza di questo nuovo punto di intersezione, esso risulta pari alla somma di  $-90^\circ$  dovuti al polo  $\omega_{p0}$ ,  $-45^\circ$  dovuti al polo  $\omega_{p1}$ , meno un contributo dovuto ai poli a frequenza superiore. E' da notare però che poiché tali poli si trovano a frequenze superiori rispetto a  $\omega_{p1}$ , il loro contributo allo sfasamento sarà modesto, al limite trascurabile se distanti almeno una decade rispetto a  $\omega_{p1}$ . Trascurando quindi questo contributo, lo sfasamento totale può essere stimato pari a  $-135^\circ$ , valore



dovuti al polo  $\omega_{p0}$ ,  $-45^\circ$  dovuti al polo  $\omega_{p1}$ , meno un contributo dovuto ai poli a frequenza superiore. E' da notare però che poiché tali poli si trovano a frequenze superiori rispetto a  $\omega_{p1}$ , il loro contributo allo sfasamento sarà modesto, al limite trascurabile se distanti almeno una decade rispetto a  $\omega_{p1}$ . Trascurando quindi questo contributo, lo sfasamento totale può essere stimato pari a  $-135^\circ$ , valore

meno negativo di  $-180^\circ$ . Il circuito così compensato risulta essere adesso stabile. Il margine di fase, pari alla differenza tra la fase ottenuta in corrispondenza dell'intersezione a 0 dB, cioè  $-135^\circ$ , e la fase nella condizione limite, pari a  $-180^\circ$  risulta quindi di  $45^\circ$ .

Procediamo ora al calcolo della pulsazione  $\omega_{P0}$  del polo dominante introdotto nel guadagno d'anello.



logaritmi, questa somma in scala logaritmica risulta pari ad un prodotto in scala lineare, e cioè la pulsazione  $\omega_{P1}$  è uguale alla pulsazione  $\omega_{P0}$  per il modulo del  $G_{loop0}$ . Data questa relazione possiamo così concludere che la pulsazione del polo introdotto  $\omega_{P0}$  è uguale alla pulsazione  $\omega_{P1}$  divisa per il modulo del guadagno d'anello a bassa frequenza.

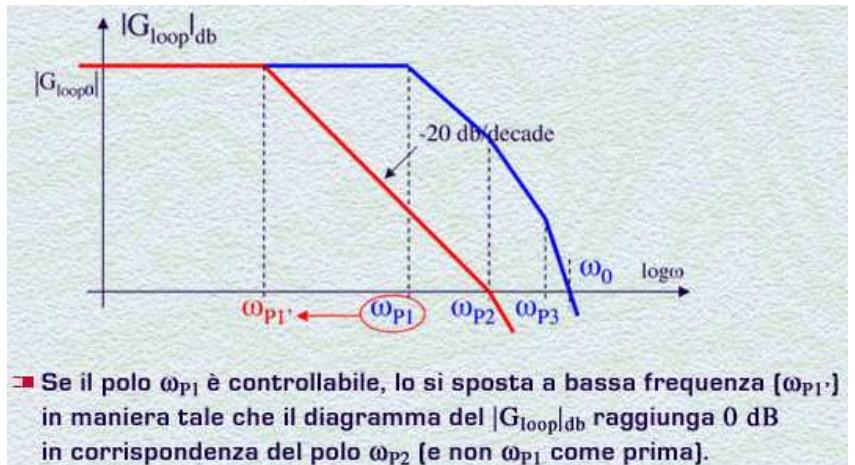
Da questa relazione possiamo notare che la nuova banda ad anello aperto del circuito,  $\omega_{P0}$  risulta notevolmente ridotta rispetto alla banda iniziale  $\omega_{P1}$ . La tecnica di compensazione a polo dominante, caratterizzata da una notevole semplicità, presenta quindi lo svantaggio di ***ridurre pesantemente la banda ad anello aperto del circuito***.

Vediamo ora una variante della tecnica appena vista che consiste nell'utilizzare come polo dominante il 1° polo del blocco di andata, spostandolo in maniera adeguata a più bassa frequenza. Questo metodo si basa sulla possibilità di controllare la costante di tempo associata a questo polo. Questa opportunità non è sempre disponibile nei circuiti reali. Si pensi ad esempio ad un amplificatore che utilizzi come blocco di guadagno un circuito integrato le cui caratteristiche, tra cui la risposta in frequenza, siano fissate a priori dal costruttore.

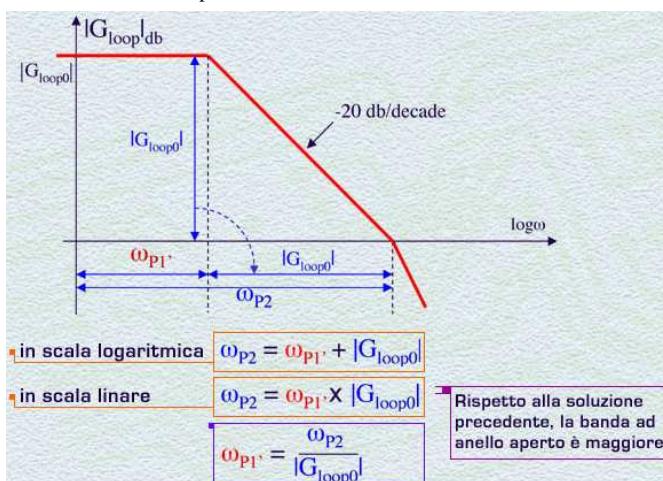
Se invece il polo  $\omega_{P1}$  è controllabile, è possibile spostarlo a più bassa frequenza alla pulsazione  $\omega_{P1'}$ , come mostrato in figura, in maniera tale che il diagramma del modulo del  $G_{loop}$  raggiunga i 0 dB in corrispondenza del polo  $\omega_{P2}$ . Come nella compensazione vista precedentemente, la nuova configurazione dei poli fa sì che la fase in corrispondenza dell'intersezione del modulo del  $G_{loop}$  con 0 dB sia circa pari a  $-135^\circ$ , in quanto il contributo del polo  $\omega_{P3}$  si può considerare in prima approssimazione trascurabile. Il margine di fase risulta anche in questo caso pari quindi a circa  $45^\circ$ .

Il calcolo della pulsazione  $\omega_{P1'}$  del polo spostato ricalca il procedimento già visto in precedenza. Considerando le quantità di interesse in scala logaritmica, la pulsazione  $\omega_{P2}$  risulta uguale a  $\omega_{P1'}$  più il modulo del guadagno d'anello a bassa frequenza  $G_{loop0}$ . In scala lineare, la pulsazione  $\omega_{P2}$  risulta uguale a  $\omega_{P1'}$  per il modulo di  $G_{loop0}$ . Dall'ultima relazione otteniamo che  $\omega_{P1'}$  è uguale a  $\omega_{P2}$  diviso il

Se osserviamo il diagramma di Bode ottenuto, possiamo determinare la relazione che lega  $\omega_{P0}$  a  $\omega_{P1}$  e al guadagno d'anello a bassa frequenza  $G_{loop0}$ . In maniera del tutto analoga a quanto avevamo fatto nel caso del prodotto guadagno-banda nel diagramma di Bode di un amplificatore a singolo polo, anche in questo caso sfruttando il grafico bilogaritmico, possiamo facilmente dedurre che in scala logaritmica la pulsazione  $\omega_{P1}$  risulta pari alla somma della pulsazione  $\omega_{P0}$  più un ulteriore segmento che è uguale al modulo di  $G_{loop0}$ . Per le note proprietà dei



modulo di  $G_{loop0}$ . Possiamo osservare da quest'ultima relazione che in questa variante della tecnica a polo dominante, la banda risulta maggiore che nel caso mostrato in precedenza, in cui la banda ad anello aperto era pari a  $\omega_{P1}/|G_{loop0}|$ .

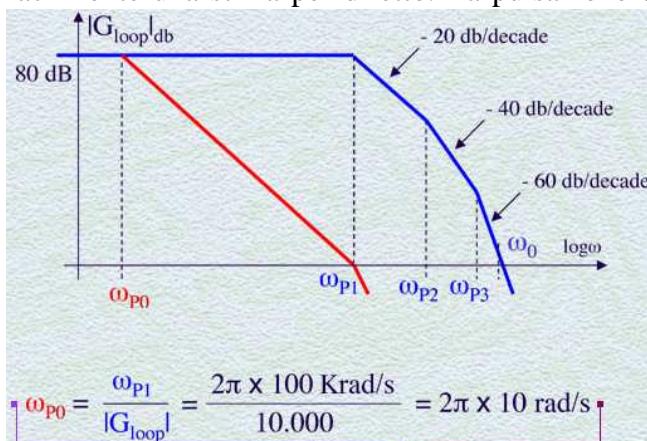


100 KHz, 1 MHz e 10 MHz.

Nella figura è mostrato il diagramma di Bode del modulo del guadagno d'anello del circuito. Si può facilmente verificare che il guadagno d'anello del circuito è uguale a meno il guadagno  $A(s)$  dell'amplificatore operazionale.

Nel grafico sono mostrati i valori del

Gloop in corrispondenza della pulsazione dei tre poli. Anche senza procedere al calcolo preciso della fase di -Gloop in corrispondenza dell'intersezione  $\omega_0$  del Gloop con l'asse a 0 dB, è possibile darne facilmente una stima per difetto. La pulsazione  $\omega_0$  risulta infatti molto superiore a entrambi i primi



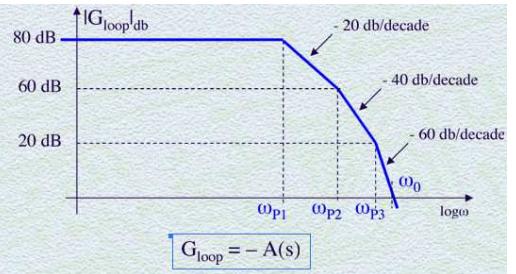
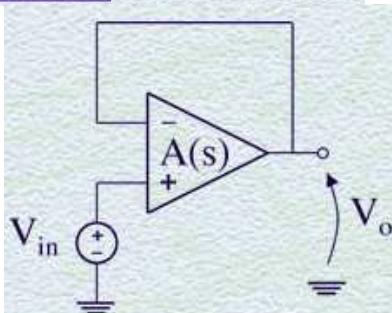
polo a bassa frequenza,  $\omega_0$ , in maniera tale che il nuovo diagramma di Bode del Gloop intersechi l'asse a 0 dB in corrispondenza della pulsazione  $\omega_{P1}$ . La pulsazione del polo dominante  $\omega_0$  è pari, in virtù della formula vista in precedenza, a  $2\pi \cdot 100$  krad/s diviso per il  $|G_{loop0}|$  che è pari a 10.000.

Il polo dominante risulta avere così pulsazione pari a  $2\pi \cdot 10$  rad/s, cioè frequenza di 10 Hz.

Si noti il notevole abbattimento della banda ad anello aperto. Questo rappresenta il prezzo da pagare per aver raggiunto la stabilità del circuito.

polo dominante, la banda risulta maggiore che nel caso mostrato in precedenza, in cui la banda ad anello aperto era pari a  $\omega_{P1}/|G_{loop0}|$ .

Vediamo un esempio di applicazione della compensazione a polo dominante. Consideriamo un buffer di tensione come quello mostrato in figura, realizzato utilizzando un amplificatore operazionale la cui risposta in frequenza sia caratterizzata dalla presenza di tre poli. Il guadagno in continua  $A_0$  dell'amplificatore operazionale è pari a 80 dB. I poli della funzione di trasferimento  $A(s)$  si trovano rispettivamente alle frequenze



Il circuito è instabile in quanto lo sfasamento a  $\omega_0$  introdotto dai tre poli è più negativo di  $(-90^\circ \times 2) - 45^\circ = -225^\circ (< -180^\circ)$

$$A(s) = \frac{A_0}{\left(1 + \frac{s}{\omega_{P1}}\right)\left(1 + \frac{s}{\omega_{P2}}\right)\left(1 + \frac{s}{\omega_{P3}}\right)}$$

- $A_0 = 80 \text{ dB}$
- $\omega_{P1} = 2\pi \times 100 \text{ KHz}$
- $\omega_{P2} = 2\pi \times 1 \text{ MHz}$
- $\omega_{P3} = 2\pi \times 10 \text{ MHz}$

poli  $\omega_{P1}$  e  $\omega_{P2}$ , ed è appena superiore al polo  $\omega_{P3}$ , data l'elevata pendenza di  $-60 \text{ dB/decade}$  con cui il diagramma del Gloop passa dal valore corrispondente a  $\omega_{P3}$  a quello corrispondente a  $\omega_0$ . Si può stimare così che lo sfasamento in corrispondenza di  $\omega_0$  risulta circa pari  $-90^\circ$  moltiplicato per 2, per i primi due poli,  $-45^\circ$  per il terzo polo. Il valore così stimato di  $-225^\circ$  è più negativo di  $-180^\circ$  e pertanto il circuito è instabile.

Procediamo quindi alla compensazione a polo dominante. Introduciamo nel circuito un nuovo

Vediamo ora come implementare in pratica questa compensazione nel nostro buffer. Il polo dominante viene introdotto nel circuito tramite una rete RC. Questa rete, posizionata all'uscita dell'amplificatore operazionale, introduce un polo nel guadagno d'anello. Si noti tuttavia, che il circuito conserva la sua funzione di buffer di tensione, in quanto il segnale in ingresso viene riportato comunque direttamente all'uscita. L'introduzione del polo aggiuntivo modificherà tuttavia la risposta in frequenza del buffer.

Calcoliamo il nuovo guadagno d'anello in seguito all'introduzione della rete RC. Tale guadagno risulta pari a quanto riportato in figura. Come vediamo, la rete RC ha introdotto nel  $G_{loop}$  un  $4^{\circ}$  polo di pulsazione  $\omega_{P0}$  pari a  $1/RC$ .

Dimensioniamo i componenti introdotti nel circuito al fine di soddisfare la condizione per la compensazione trovata in precedenza. Ricordiamo infatti che la pulsazione  $\omega_{P0}$  deve essere uguale a  $2\pi \cdot 10$  rad/s. L'inverso di questa pulsazione è pari alla costante di tempo  $RC$  della rete introdotta. Abbiamo libertà di scelta su entrambi i componenti  $R$  e  $C$ . Supponendo di scegliere una resistenza  $R$  pari a  $1 \text{ M}\Omega$ , ne consegue che la capacità deve essere pari a  $1/(\omega_{P0} \cdot R)$  e cioè deve essere uguale a  $16 \text{ nF}$ .

### Compensazione polo-zero

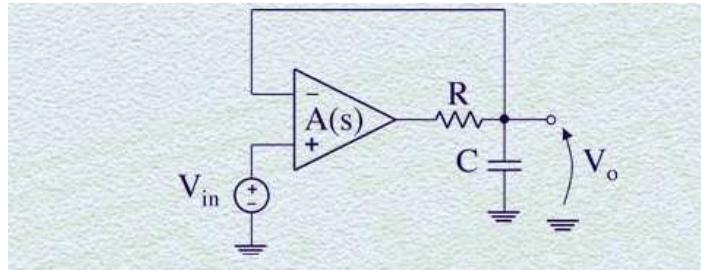
La tecnica di compensazione polo-zero consiste nell'introdurre in cascata al blocco di guadagno  $A(s)$  una rete la cui risposta in frequenza  $B(s)$  sia caratterizzata dalla presenza di un polo e di uno zero.

Un tipico andamento della funzione di trasferimento  $B(s)$  introdotta è mostrato in figura. Il guadagno del blocco aggiunto è costante a bassa ed ad alta frequenza ed è caratterizzato dalla presenza di un polo  $\omega_{P0}$  e da uno zero  $\omega_{Z0}$ , con la frequenza del polo inferiore a quella dello zero.

In questa tecnica, lo zero introdotto viene utilizzato per compensare il polo  $\omega_{P1}$  a più bassa frequenza nel blocco di guadagno. Il polo introdotto viene invece utilizzato per ottenere la stabilità del circuito in maniera simile a quella vista per la compensazione a polo dominante.

L'effetto della cancellazione del polo a pulsazione  $\omega_{P1}$  con lo zero  $\omega_{Z0}$  viene mostrato nella relazione analitica ottenuta moltiplicando la funzione di trasferimento  $A(s)$  per quella della rete polo-zero aggiunta  $B(s)$ . Affinché tale cancellazione possa avere luogo, occorre che la pulsazione  $\omega_{Z0}$  dello zero venga posta pari alla pulsazione  $\omega_{P1}$ . La funzione di trasferimento risultante diventa quindi caratterizzata dalla presenza di tre poli, come la funzione di trasferimento di partenza  $A(s)$ .

Ora però abbiamo acquisito la possibilità di compensare il circuito scegliendo opportunamente il valore di uno di questi poli, e cioè quello di pulsazione  $\omega_{P0}$  introdotto dalla rete polo-zero.



■ La rete RC introduce un polo nel guadagno d'anello.

$$G_{loop} = -\frac{\frac{1}{sC}}{R + \frac{1}{sC}} A(s) = -\frac{1}{1+sCR} A(s)$$

IV polo

$\omega_{P0} = 1/RC$

### Compensazione polo-zero

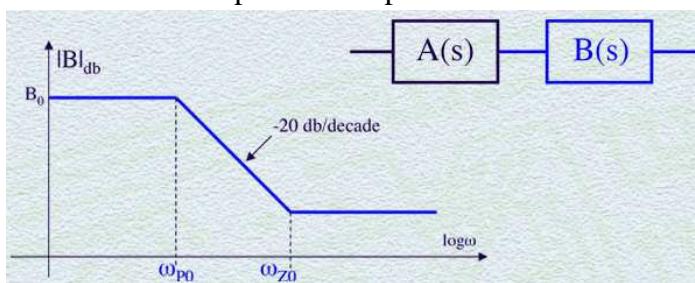
La tecnica di compensazione polo-zero consiste nell'introdurre in cascata al blocco di guadagno  $A(s)$  una rete la cui risposta in frequenza  $B(s)$  sia caratterizzata dalla presenza di un polo e di uno zero.

Un tipico andamento della funzione di trasferimento  $B(s)$  introdotta è mostrato in figura. Il guadagno del blocco aggiunto è costante a bassa ed ad alta frequenza ed è caratterizzato dalla presenza di un polo  $\omega_{P0}$  e da uno zero  $\omega_{Z0}$ , con la frequenza del polo inferiore a quella dello zero.

In questa tecnica, lo zero introdotto viene utilizzato per compensare il polo  $\omega_{P1}$  a più bassa frequenza nel blocco di guadagno. Il polo introdotto viene invece utilizzato per ottenere la stabilità del circuito in maniera simile a quella vista per la compensazione a polo dominante.

L'effetto della cancellazione del polo a pulsazione  $\omega_{P1}$  con lo zero  $\omega_{Z0}$  viene mostrato nella relazione analitica ottenuta moltiplicando la funzione di trasferimento  $A(s)$  per quella della rete polo-zero aggiunta  $B(s)$ . Affinché tale cancellazione possa avere luogo, occorre che la pulsazione  $\omega_{Z0}$  dello zero venga posta pari alla pulsazione  $\omega_{P1}$ . La funzione di trasferimento risultante diventa quindi caratterizzata dalla presenza di tre poli, come la funzione di trasferimento di partenza  $A(s)$ .

Ora però abbiamo acquisito la possibilità di compensare il circuito scegliendo opportunamente il valore di uno di questi poli, e cioè quello di pulsazione  $\omega_{P0}$  introdotto dalla rete polo-zero.



$$A(s)B(s) = \frac{A_0}{(1 + \frac{s}{\omega_{P1}})(1 + \frac{s}{\omega_{P2}})(1 + \frac{s}{\omega_{P3}})} B_0 \frac{(1 + \frac{s}{\omega_{Z0}})}{(1 + \frac{s}{\omega_{P0}})}$$

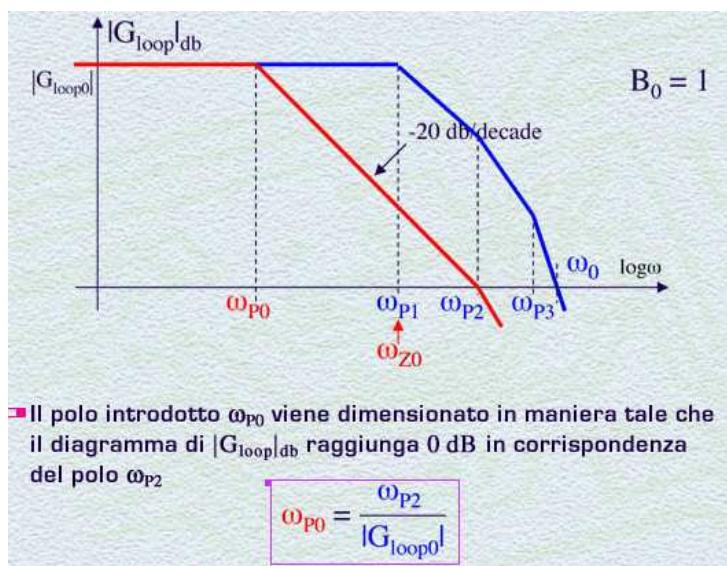
rete polo-zero

■ Lo zero  $\omega_{Z0}$  viene utilizzato per compensare il polo  $\omega_{P1}$

→  $\omega_{Z0} = \omega_{P1}$

■ Il polo aggiunto  $\omega_{P0}$  viene dimensionato in maniera tale che il circuito risultante sia stabile.

Nella figura è mostrato il diagramma di Bode del modulo del guadagno d'anello prima e dopo la compensazione con la rete polo-zero. Per semplicità di rappresentazione, abbiamo supposto che il guadagno in continua della rete  $B_0$  sia pari ad uno. In questa maniera, il modulo del  $G_{loop}$  a bassa frequenza non è cambiato per l'introduzione della rete di compensazione. Dal diagramma di Bode possiamo osservare la compensazione del polo  $\omega_{P1}$  tramite lo zero  $\omega_{Z0}$ . Il polo  $\omega_{P0}$  introdotto dalla rete aggiuntiva si trova invece a più bassa frequenza. Il procedimento per il dimensionamento di questo polo è del tutto analogo alla tecnica di compensazione vista in precedenza. Cioè  $\omega_{P0}$  viene dimensionato opportunamente in maniera tale che il modulo del  $G_{loop}$



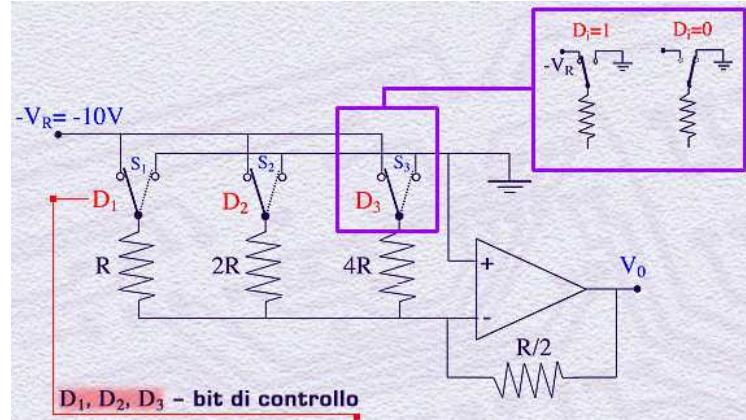
raggiunga l'asse a 0 dB in corrispondenza del polo  $\omega_{P2}$ . La pulsazione  $\omega_{P0}$  risulta quindi uguale a  $\omega_{P2}$  diviso per il modulo del  $G_{loop0}$ .

## Lezione n. 39 : Introduzione ai convertitori digitali-analogici (DAC)

### Circuito sommatore usato come DAC

Consideriamo il circuito sommatore rappresentato nella figura, in cui tutti i componenti, cioè l'amplificatore operazionale, i deviatori  $S_i$  e le resistenze  $R$ , siano, per ora, ideali.

Chiamiamo  $D_1, D_2, D_3$  i bit di controllo, rispettivamente, dei deviatori  $S_1, S_2, S_3$ . La posizione di ciascun deviatore è controllata dal valore 0 oppure 1 del bit  $D_i$ : al valore 0 corrisponde la condizione di deviatore a massa, al valore 1 la condizione di deviatore collegato alla tensione fissa di  $-V_R$ . Si noti che i valori delle resistenze sono diversi a seconda del deviatore a cui sono collegate.



Parola digitale di 3 bit:  $D_1 D_2 D_3$

$2^3 = 8$  configurazioni distinte  $N_k$  ( $K=0, \dots, 7$ )

Config.	Numero N	D <sub>1</sub>	D <sub>2</sub>	D <sub>3</sub>
$N_0$	0	0	0	0
$N_1$	1	0	0	1
$N_2$	2	0	1	0
$N_3$	3	0	1	1
$N_4$	4	1	0	0
$N_5$	5	1	0	1
$N_6$	6	1	1	0
$N_7$	7	1	1	1

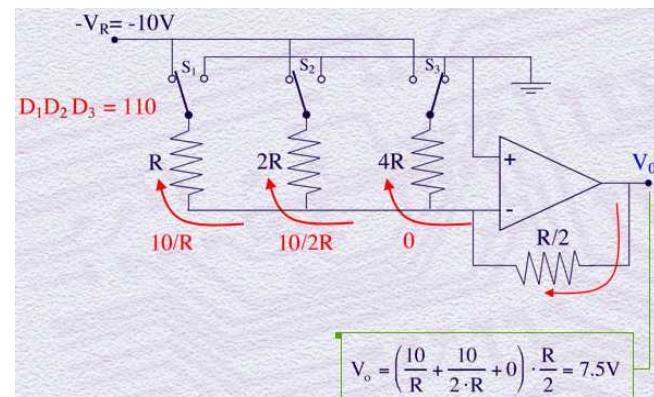
corrisponde la configurazione dei deviatori riportata nella figura. Si trova che le correnti afferenti alla terra virtuale sono diverse nei vari rami per via dei diversi valori delle resistenze e che l'uscita si porta a 7.5V.

E' buon esercizio, ragionando in maniera analoga, calcolare tutti i valori di tensione dell'uscita corrispondenti a tutte le possibili parole di ingresso.

Possiamo così aggiungere alla tabella una ulteriore colonna  $V_o$ , dove riportiamo i valori dell'uscita in corrispondenza a ciascuna delle

Ci chiediamo quale valore assume la tensione di uscita  $V_o$  dell'operazionale a seconda della posizione dei deviatori del circuito. La parola digitale costituita dall'insieme dei tre bit  $D_i$  definisce l'informazione data all'ingresso del circuito. Essa può assumere  $2^3 = 8$  configurazioni distinte  $N_K$  ( $K=0, \dots, 7$ ) individuabili con le cifre decimali da 0 a 7 codificate in binario, come si vede dalla tabella.

Per abituarci al calcolo del valore di tensione  $V_o$  a cui si porta l'uscita del circuito, applichiamo ad esempio la parola 110, corrispondente al numero 6, all'ingresso del sommatore. Ad esso



possibili configurazioni dei deviatori. Si noti che, facendo corrispondere a ciascuna parola  $D_1 D_2 D_3$  il valore decimale  $N_K$  equivalente, possiamo anche

$$\text{scrivere che: } V_o = V_R \frac{N_k}{2^n} = 10 \cdot \frac{N_k}{8}$$

Nel caso dell'esempio di prima, la parola 110 corrispondente al numero 6, determina una tensione di uscita pari a  $10 \cdot 6 / 8$ , cioè appunto 7.5V. Riassumendo, il

circuito sommatore in esame converte una parola digitale di 3 bit (il nostro codice binario in ingresso) in una tensione proporzionale al valore decimale  $N_K$  corrispondente.

Il circuito realizza quindi una conversione digitale-analogica e per questo viene chiamato DAC (Digital-to-Analog Converter).

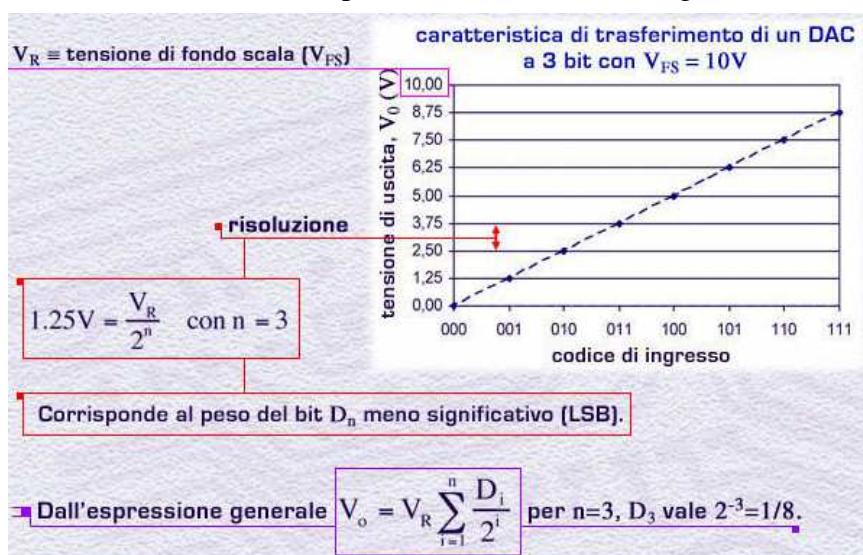
codice binario in ingresso	tensione in uscita				
Config.	N	D <sub>1</sub>	D <sub>2</sub>	D <sub>3</sub>	$V_o(V)$
$N_0$	0	0	0	0	0
$N_1$	1	0	0	1	1.25
$N_2$	2	0	1	0	2.5
$N_3$	3	0	1	1	3.75
$N_4$	4	1	0	0	5
$N_5$	5	1	0	1	6.25
$N_6$	6	1	1	0	7.5
$N_7$	7	1	1	1	8.75

Se si volesse ricavare l'espressione letterale generale della tensione  $V_o$  di uscita in funzione del valore Di dei singoli bit della parola di ingresso, riprendendo l'esempio del caso particolare della parola 110 fatta prima, si ha che:  $V_o = V_R \left( \frac{D_1}{2} + \frac{D_2}{2^2} + \frac{D_3}{2^3} \right) = 7.5V$ . In generale quindi, con n bit si ha:  $V_o = V_R \sum_{i=1}^n \frac{D_i}{2^i}$ .

La rappresentazione in un piano cartesiano della tensione di uscita  $V_o$  in funzione del codice di ingresso è detta "caratteristica di trasferimento del DAC".

In sostanza il sommatore presenta in uscita uno tra gli 8 valori che si ottengono dalla suddivisione in  $2^n$  parti uguali della tensione  $V_R=10V$ . La tensione  $V_R$  è chiamata normalmente tensione di fondo scala  $V_{FS}$  del DAC.

La più piccola variazione della tensione di uscita, pari all'incremento dell'uscita da un codice al successivo, prende il nome di "risoluzione" del DAC e vale, nel nostro caso, 1.25 V, cioè  $V_R/2^3$ . La risoluzione corrisponde quindi al peso del bit meno significativo (LSB) nell'espressione generale prima trovata. Per  $n=3$  il peso di  $D_n$  vale infatti proprio  $2^{-3}=1/8$ .



## Effetti delle non idealità del sommatore

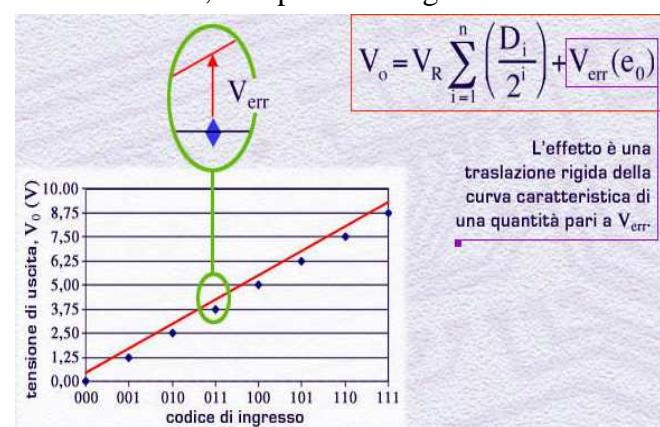
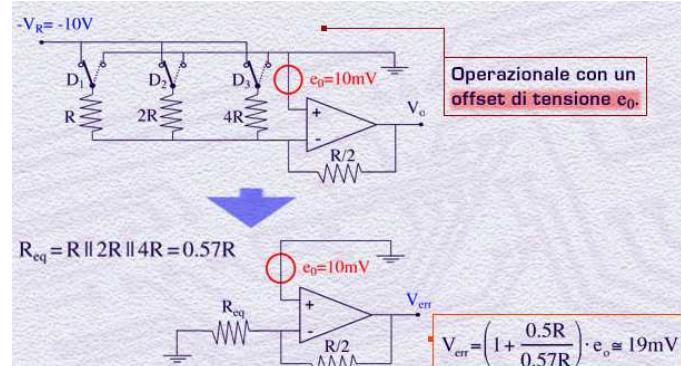
Si sono sin qui considerati ideali tanto l'amplificatore operazionale quanto i deviatori. Esaminiamo ora gli effetti delle non idealità partendo da quelle dell'operazionale, soffermandoci in particolare sulla presenza di un offset di tensione  $e_0$ . Questo offset di tensione si riflette su  $V_o$  in modo indipendente dal codice di ingresso. Infatti, utilizzando la sovrapposizione degli effetti tra i generatori  $e_0$  e  $V_R$ , ci si rende immediatamente conto che, indipendentemente dal valore del bit di controllo, i resistori  $R$ ,  $2R$  e  $4R$  sono tutti a massa quando è presente solo  $e_0$ .

Pertanto, per trovare la tensione di uscita della configurazione non invertente equivalente, la resistenza  $R_{eq}$  è il parallelo tra  $R$ ,  $2R$  e  $4R$ . Nel caso ad esempio che la tensione di offset  $e_0$  sia di 10mV, la tensione  $V_{err}$  in uscita dal DAC è di 19mV. In questo caso l'errore è piccolo rispetto all'LSB, ma il suo peso cresce al crescere del numero n di bit. Infatti, aumentando il numero di bit e quindi di rami all'ingresso, il valore di  $R_{eq}$  diminuisce e conseguentemente aumenta  $V_{err}$ . Tenendo conto dell'offset, l'espressione generale diventa:

$$V_o = V_R \sum_{i=1}^n \left( \frac{D_i}{2^i} \right) + V_{err}(e_0)$$

L'effetto che questo errore ha sulla caratteristica di trasferimento di un DAC è di trasstrarla rigidamente di una quantità proprio pari a  $V_{err}$ .

La regolazione dell'offset con gli appositi trimmer per riportare a zero l'uscita non risolve completamente il problema. Infatti la regolazione dipende dalle condizioni operative (temperatura, alimentazioni, età del componente) e quindi deve essere periodicamente rifatta.



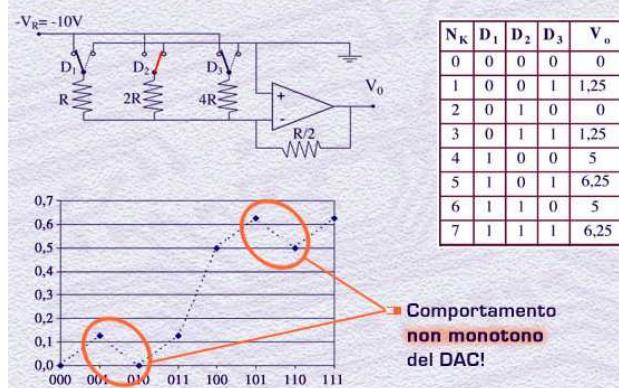
Immaginiamo ora che siano i deviatori a non comportarsi idealmente. I deviatori reali si comportano come resistenze  $r_i$  in serie al ramo in cui sono inseriti, cambiando così il peso del relativo bit. Ricalcolando la tensione di uscita  $V_o$  nella nuova situazione si ottiene la nuova espressione della funzione di trasferimento:

$$V_R \sum_{i=1}^n \frac{D_i}{2^i} \left( 1 - \frac{r_i}{R} \cdot \frac{1}{2^{i-1}} \right).$$

Le  $r_i$  determinano una non linearità nella caratteristica di trasferimento, variando l'incremento dell'uscita da un codice al successivo, incremento che differisce così dal valore ideale di un LSB.

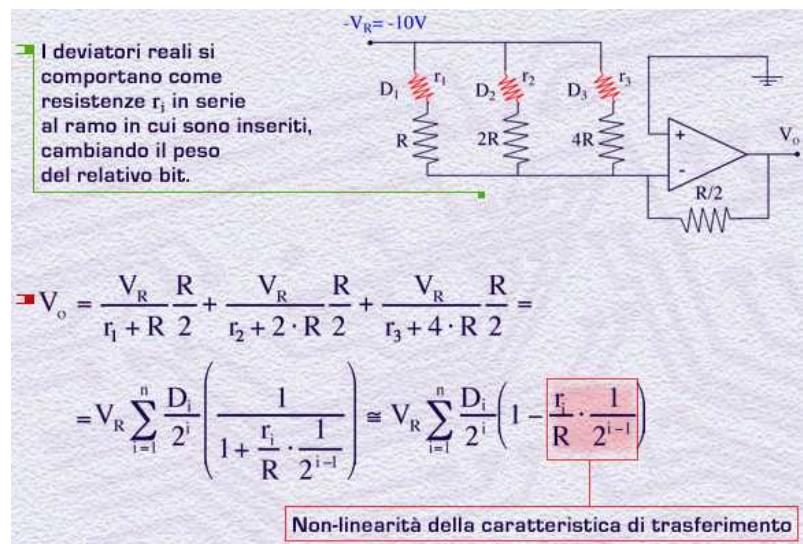
Nel nostro sommatore reale potrebbe capitare che un deviatore si guasti, rimanendo sempre collegato a  $V_R$  o a massa, indipendentemente dal valore che assume il suo bit di controllo.

■ Si supponga che il deviatore  $S_2$  sia bloccato a massa.



del codice. Questo comportamento non monotono dell'uscita in alcuni casi è molto pericoloso: si pensi ad esempio ai sistemi di controllo dove un comportamento non monotono può portare ad instabilità.

Un caso analogo al precedente è quello in cui un deviatore sia bloccato nella posizione 1, cioè con il collegamento alla tensione  $V_R$  sempre attivo. L'effetto è ovviamente analogo ed è lasciato al lettore verificare la tabella dei valori dell'uscita in funzione del codice di ingresso.



## Esercizio 1

- Avrete notato che la tensione di uscita del DAC non dipende dal valore di  $R$ , come pure non vi dipende la sua risoluzione. Ma allora con quale criterio scegliereste il valore della resistenza  $R$ ?
- Quanti interruttori dovrebbe avere il convertitore appena studiato per avere una risoluzione di almeno 10mV?

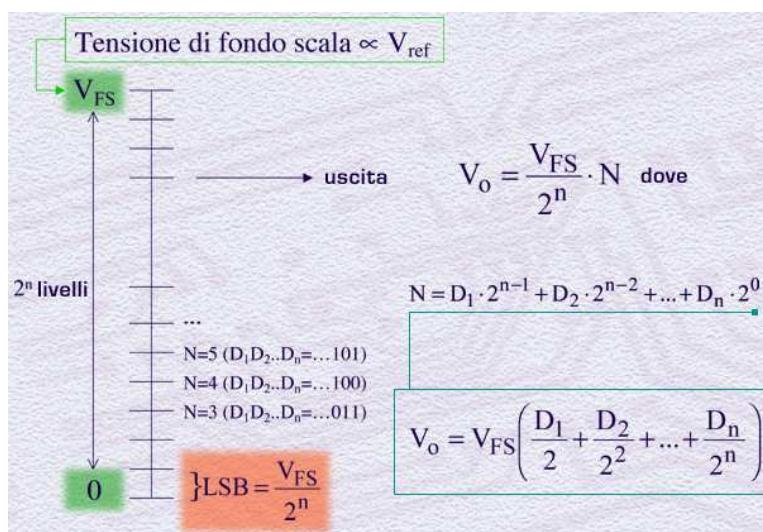
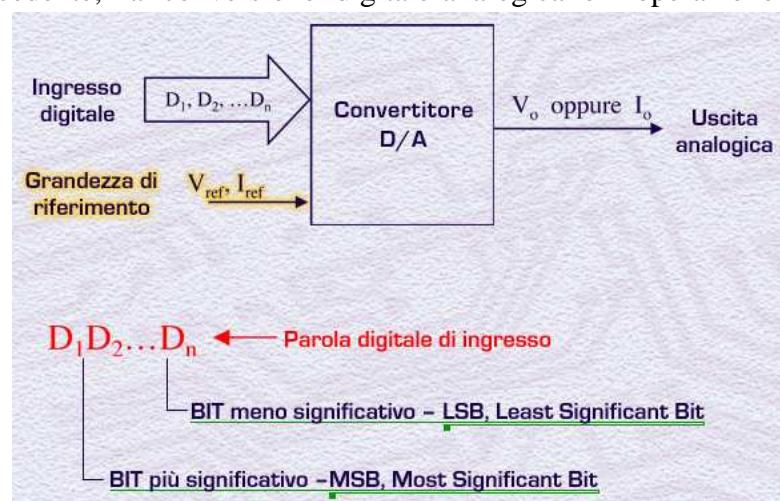
## Lezione n. 40 : Proprietà dei convertitori digitali-analogici (DAC)

### Generalità sui DAC

Come abbiamo visto nella lezione precedente, la conversione digitale-analogica è l'operazione mediante la quale un apposito circuito genera in uscita una grandezza elettrica (una tensione o una corrente) di valore proporzionale a un numero fornito in ingresso, rappresentato con una parola digitale, normalmente codificata in binario, di  $n$  bit indicati con  $D_1, D_2, \dots, D_n$ .

Assumeremo che  $D_1$  rappresenti il bit più significativo della parola digitale e lo indichiamo con il termine MSB, acronimo di Most Significant Bit, mentre  $D_n$  rappresenti il bit meno significativo, indicato con LSB, Least Significant Bit.

Il circuito viene chiamato convertitore digitale-analogico o DAC dal termine inglese Digital to Analog Converter.

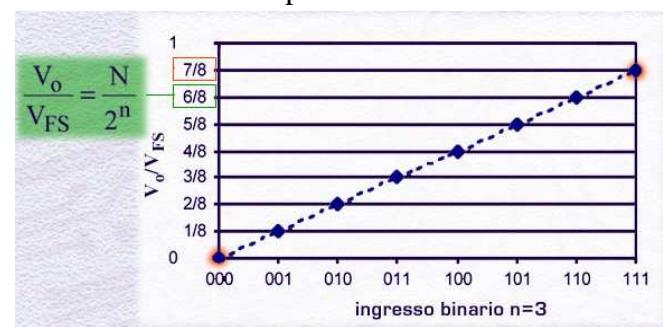


$V_{FS}/2^n$  e coincidente con il LSB, per il numero decimale intero  $N$  rappresentato dalla parola digitale. Poiché, in una codifica binaria,  $N=D_1 \cdot 2^{n-1} + D_2 \cdot 2^{n-2} + D_n \cdot 2^0$  dal prodotto si ottiene la relazione

generale tra ingresso ed uscita:  $V_o = V_{FS} \left( \frac{D_1}{2} + \frac{D_2}{2^2} + \dots + \frac{D_n}{2^n} \right)$ . Nel grafico è rappresentata la

caratteristica di trasferimento di un DAC ideale con codice di ingresso a tre bit. L'asse delle ascisse riporta le combinazioni dell'ingresso mentre l'asse delle ordinate riporta la tensione di uscita normalizzata al valore di fondo scala. Ad ogni codice di ingresso corrisponde un livello normalizzato di uscita che è una frazione di valore compreso tra 0 e 7/8. La frazione si ottiene dividendo il numero decimale intero  $N$  corrispondente alla parola digitale per  $2^n$ .

Nel caso ideale tutti i valori stanno su una retta indice questo di un comportamento lineare del DAC e sono separati l'uno dall'altro di 1/8 del fondo scala, cioè di un LSB. Al codice binario minimo, 000, corrisponde un'uscita nulla; a quello massimo, cioè 111 (7 in decimale), corrisponde una tensione pari a 7/8 del fondo scala: la tensione di uscita non raggiunge mai il valore di fondo scala, ma ne rimane distanziata di un LSB.



I parametri principali che caratterizzano un DAC sono:

- Il numero **n** dei **bit nella parola di ingresso**. Esso è un indice del numero di livelli discreti in uscita, pari a  $2^n$ , o, equivalentemente, del numero di intervalli in cui viene suddivisa la tensione di fondo scala;
- è quindi sinonimo di **risoluzione**, concetto che quantifica la più piccola variazione che può subire, nel caso ideale, la grandezza in uscita. La risoluzione, oltre che in bit, si trova espressa in molti modi: in parti su  $2^n$ ; in % (pari a  $2^{-n} \times 100$ ); in volt o milliampère (pari al valore di fondo scala diviso per  $2^n$ , corrispondente ad un LSB).
- Inoltre un DAC è caratterizzato da un definito valore di **fondo scala**. Valori tipici di  $V_{FS}$  nei DAC di tensione sono: 2.5V, 5V, o 10V mentre per i DAC in corrente  $I_{FS}$  può valere qualche mA.

### Errori e non linearità nei DAC reali

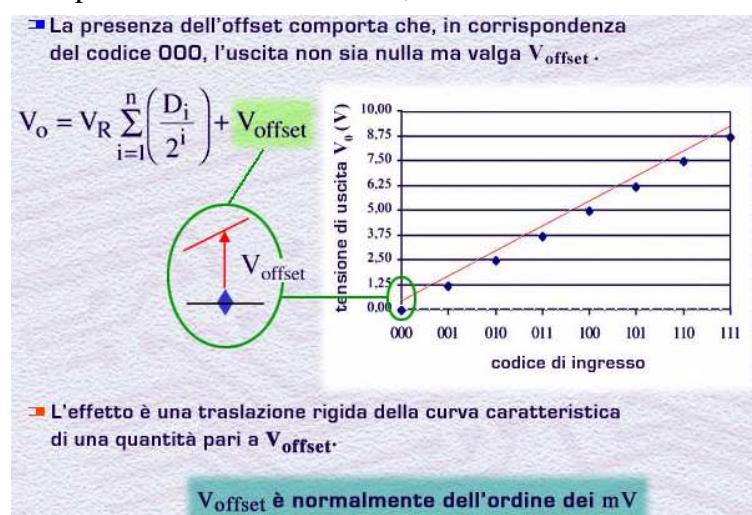
Il comportamento dei DAC reali si discosta da quello ideale per la presenza di svariate cause che si riflettono in scostamenti della caratteristica reale rispetto a quella ideale. Gli errori principali sono quelli di:

- offset,
- guadagno
- non linearità.

La presenza di un offset comporta che, in corrispondenza del codice 000, l'uscita non sia nulla ma valga  $V_{offset}$ . Come già visto nella lezione precedente, l'effetto dell'offset è una traslazione rigida della curva caratteristica di una quantità pari a  $V_{offset}$ . L'utilizzatore può normalmente regolare a zero mediante dei potenziometri questo valore, ma la sua dipendenza dalla temperatura, dall'età, e, più in generale, dalle condizioni operative del DAC, fa sì che la correzione sia efficace solo per tempi limitati e vada rieseguita periodicamente.

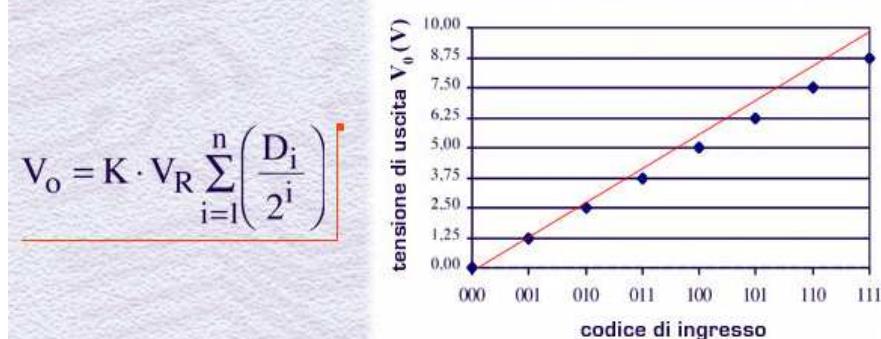
Un buon DAC ha un offset dell'ordine di pochi millivolt.

Un'altra causa di scostamento dell'uscita analogica dalla situazione ideale è la



■ La presenza di un **guadagno non costante** al variare del codice di ingresso comporta che il valore dell'uscita analogica differisca dal valore ideale.

■ Normalmente lo scostamento aumenta all'aumentare del codice di ingresso e può essere rappresentato con un fattore moltiplicativo K nella relazione ingresso/uscita.



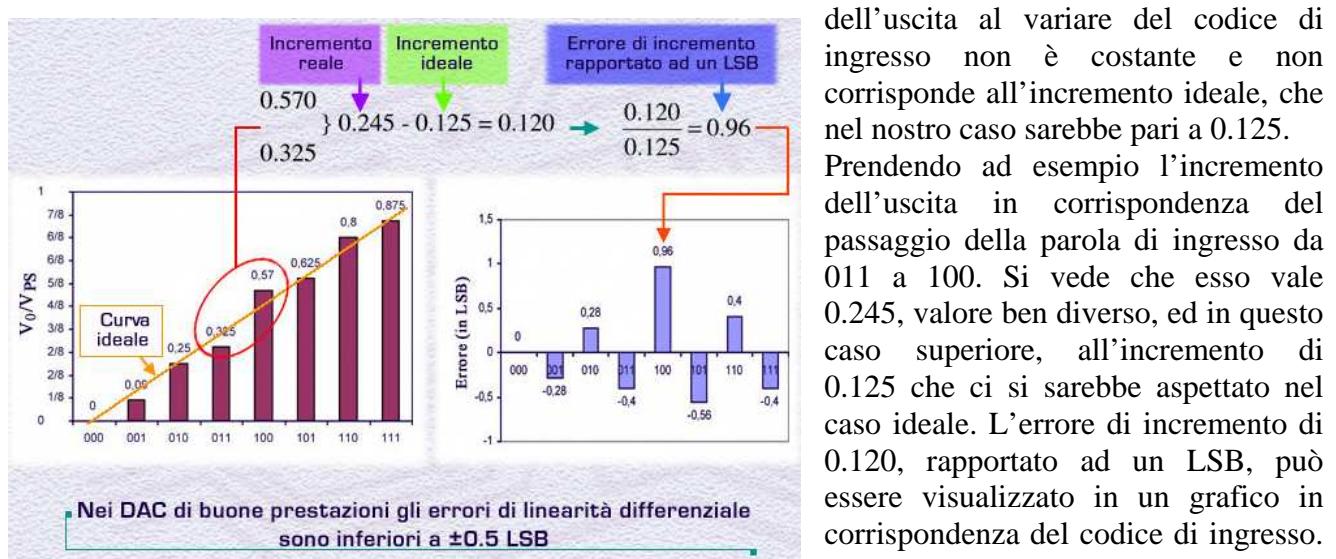
presenza di un **guadagno non costante al variare del codice di ingresso**. Normalmente lo scostamento aumenta all'aumentare del codice di ingresso e può essere rappresentato con un fattore moltiplicativo K nella relazione ingresso/uscita. L'errore di guadagno ha l'effetto di modificare l'inclinazione della caratteristica lineare reale rispetto a quella ideale. Anche per il fattore K valgono le stesse considerazioni sulla sua controllabilità fatte prima per l'offset.

Prendiamo ora in considerazione la caratteristica di trasferimento di un DAC reale a 3 bit come quella nella figura, il cui asse verticale è stato normalizzato al valore di fondo scala. In cima ad ogni colonna dell'istogramma è riportato il valore effettivo dell'uscita normalizzata. Il grafico riporta anche, per confronto, la retta che visualizza la risposta che avrebbe avuto il DAC se fosse stato ideale.

Si nota innanzitutto che i valori reali della tensione di uscita differiscono da quelli ideali. La semplice differenza tra il valore reale e il valore ideale dell'uscita dà luogo all'**errore di linearità integrale** di un DAC. Essa

può venire riportata in un grafico per ogni codice di ingresso ed è normalmente espressa in termini di LSB, cioè rapportata al peso di un LSB (0,125 nel nostro caso). I DAC di buone prestazioni hanno errori di linearità integrale inferiori a  $\pm 1$  LSB, cioè mai lo scostamento del valore reale dell'uscita rispetto al valore ideale supera 1 LSB.

Un altro aspetto che si nota nella caratteristica di funzionamento di un DAC reale è che l'incremento dell'uscita al variare del codice di ingresso non è costante e non corrisponde all'incremento ideale, che nel nostro caso sarebbe pari a 0,125.



codici visualizza quello che viene chiamato, per come è stato costruito, l'**errore di linearità differenziale** del DAC.

In un DAC ideale tutti gli incrementi tra un codice e il successivo sarebbero uguali tra loro e pari a 1 LSB e quindi il grafico sarebbe costantemente a zero. Nel DAC reale invece, in corrispondenza ad esempio del codice 110 trovo un errore di linearità differenziale di +0,96 LSB e così via per gli altri codici.

I DAC di buone prestazioni hanno errori di linearità differenziale contenuti entro  $\pm 0,5$  LSB. Si noti che *l'errore di linearità integrale discusso prima è, per ogni codice di ingresso, pari alla somma degli errori di linearità differenziale del codice considerato e di quelli che lo precedono, da cui il suo nome.*

Un DAC si dice monotono quando la sua uscita cresce al crescere del numero rappresentato dalla parola digitale in ingresso. Quando ciò non accade, il DAC presenta un errore di linearità differenziale superiore a un LSB. L'esempio nella figura a pagina seguente riporta la caratteristica di trasferimento di un DAC non monotono in cui l'uscita diminuisce da 0,27 a 0,25 al crescere del codice da 001 a 010. Il corrispondente grafico degli errori di linearità differenziale presenta effettivamente valori superiori a 1 LSB in corrispondenza di quei codici.

Naturalmente, non è necessariamente vero che un DAC con un errore di linearità differenziale superiore a 1 LSB sia non monotono.

La monotonicità è caratteristica essenziale per i DAC usati nei sistemi di controllo, dove un comportamento non monotono può portare ad una instabilità del sistema.

Tante possono essere le cause che producono non monotonicità nei DAC reali. Una di queste era stata vista nella lezione precedente e dovuta al blocco di un deviatore nel circuito di ingresso del sommatore.

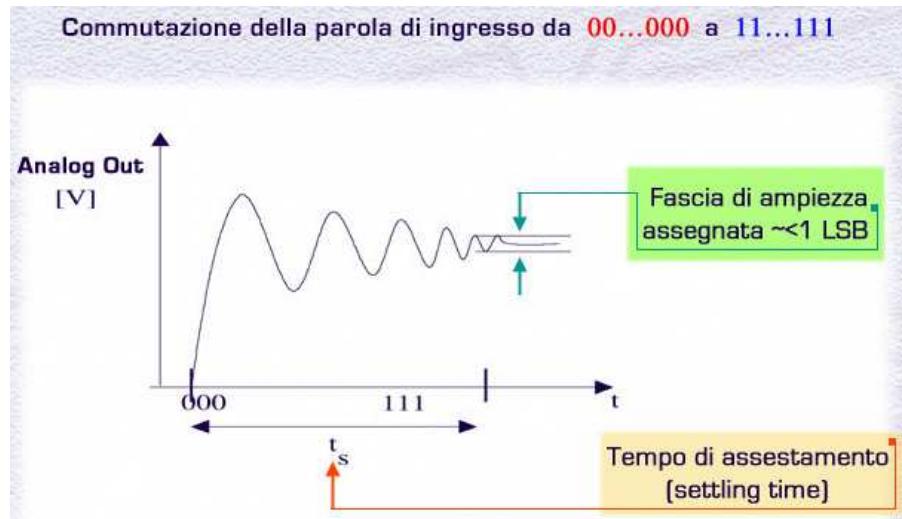
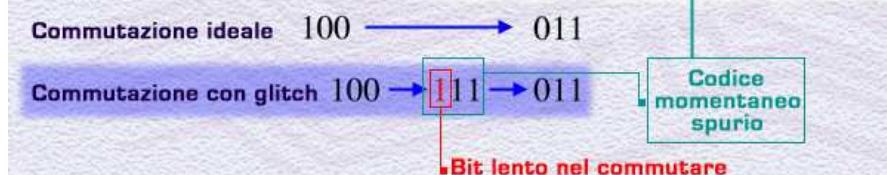
Un altro aspetto importante di un DAC è il tempo che impiega l'uscita analogica a raggiungere il nuovo

valore imposto da una commutazione all'ingresso. Questa informazione sulla velocità di risposta del DAC è normalmente indicata come **"tempo di assestamento"** del DAC o, con terminologia inglese, **"settling time"** del DAC.

Operativamente, esso indica l'intervallo di tempo  $t_s$  impiegato dall'uscita del DAC per raggiungere il valore finale quando tutti i bit del codice di ingresso commutano istantaneamente da 0 a 1. Il tempo è preso quando il valore analogico dell'uscita rientra in una fascia di ampiezza assegnata a cavallo del valore asintotico finale. L'ampiezza della fascia viene normalmente espressa in LSB o percentuale di  $V_{FS}$ . In uscita ad un DAC, nella transizione della risposta da un codice di ingresso al seguente, può succedere che compaia un impulso di breve durata. Esso è dovuto alla commutazione non istantanea dei vari bit della parola, che produce codici di ingresso transitori e quindi uscite momentaneamente differenti da quella finale. L'impulso

**Il glitch** è un impulso di breve durata che può comparire all'uscita del DAC nella transizione della risposta da un codice di ingresso al seguente.

E' dovuto alla **commutazione non istantanea** dei bit della parola, che produce codici di ingresso transitori e quindi uscite momentaneamente differenti da quella finale.



presente in uscita è spesso indicato con il termine inglese di **glitch**.

Nel caso ad esempio in cui il codice di ingresso passi dal valore 100 al valore 011, il **glitch** potrebbe manifestarsi se il primo bit commutes con ritardo rispetto agli altri due. Questo fatto genererebbe infatti un codice momentaneo spurio 111 che tenderebbe per un brevissimo tempo a far crescere l'uscita ben oltre il valore corretto a cui si assesterebbe solo in un secondo tempo.

Ulteriori parametri caratterizzanti le prestazioni di un DAC reale sono:

- la stabilità
- l'accuratezza
- la precisione.

Vediamone il significato.

- La **stabilità** è un indice di quanto il comportamento di un DAC sia sensibile all'età, alla temperatura, alle tensioni di alimentazione. Si quantifica tramite coefficienti di temperatura per guadagno, offset, errori di linearità differenziale e integrale e con la reiezione alle alimentazioni.
- L'**accuratezza** (assoluta o inaccuratezza) è la peggior differenza che si può riscontrare tra l'uscita del convertitore reale e la corrispondente uscita del convertitore ideale. Viene normalmente espressa in frazioni di LSB.
- La **precisione** è un indice della capacità del DAC di fornire sempre lo stesso valore analogico di uscita a parità di codice di ingresso.

## Esercizio 1

- Quale è la risoluzione di un DAC a 12 bit alimentato a 3.3V ?
- Se lo stesso DAC fosse alimentato a 10V, quale risoluzione avrebbe ?
- C'è motivo di scegliere l'uno piuttosto che l'altro ?

## Lezione n. 41 : Esempi di architetture di DAC

### Convertitori D/A ad impedenze pesate

Le tipologie circuituali impiegate nella realizzazione dei convertitori digitali-analogici commerciali sono numerose e differiscono tra loro per:

- la tecnologia di fabbricazione
- la velocità di conversione
- l'area occupata sul chip
- il numero massimo di bit convertibili.

In questa lezione vedremo alcune di queste topologie circuituali, scegliendo tra quelle più utilizzate nei moderni DAC integrati.

Per iniziare soffermiamoci di nuovo ancora un momento sulla struttura del DAC a resistenze pesate che abbiamo già introdotto nelle lezioni precedenti, e riportata nella figura per comodità.

Il suo principio di funzionamento, ricordate, si basa sulla possibilità di inviare correnti sempre più elevate, secondo una relazione binaria, alla terra virtuale di un operazionale in modo che vengano sommate nel ramo di reazione a dare la tensione analogica di uscita.

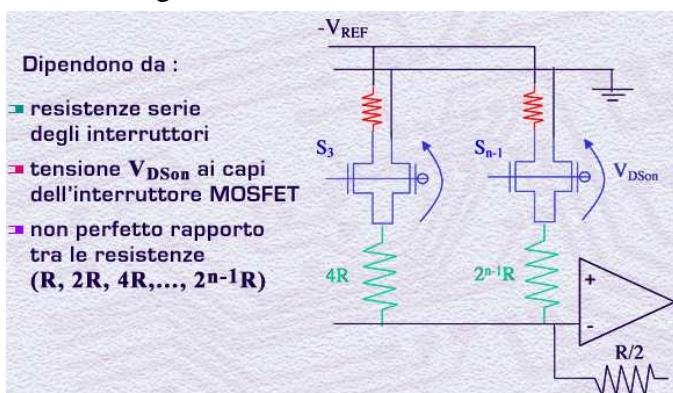
Il problema principale di questa struttura deriva proprio dalla difficoltà di realizzare resistori di valore crescente secondo la sequenza  $R, 2R, 4R, \dots, 2^{n-1}R$ , dove  $n$  è il numero dei bit nella parola di ingresso del DAC. Per

esempio, con  $n=10$  e  $R=5\text{k}\Omega$ , la tecnologia dovrebbe consentire la realizzazione di resistenze crescenti fino al megaohm, con valori successivi in rigoroso rapporto di 2. Questo requisito è molto difficile da garantire. Inoltre, resistenze di valori così elevati occupano sul chip molto spazio e quindi

portano a dispositivi con costi troppo elevati in caso di  $n$  superiore a 6 o 8.

Gli errori di linearità integrale e differenziale in questo DAC derivano dalle resistenze serie degli interruptori, che pesano diversamente a seconda del bit cui sono associati, dalla tensione  $V_{DS}$  ai capi degli interruptori MOS chiusi e proprio dai non perfetti rapporti tra le resistenze.

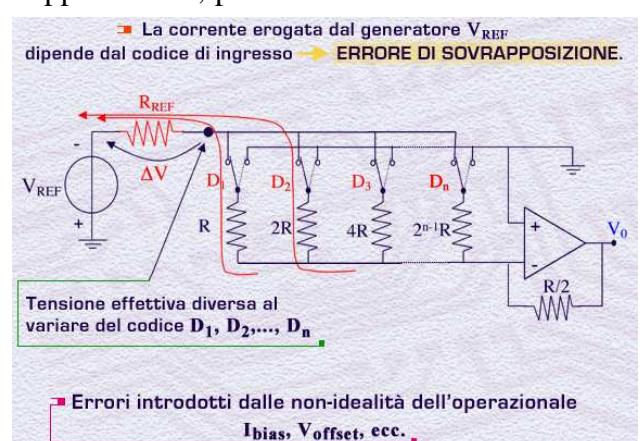
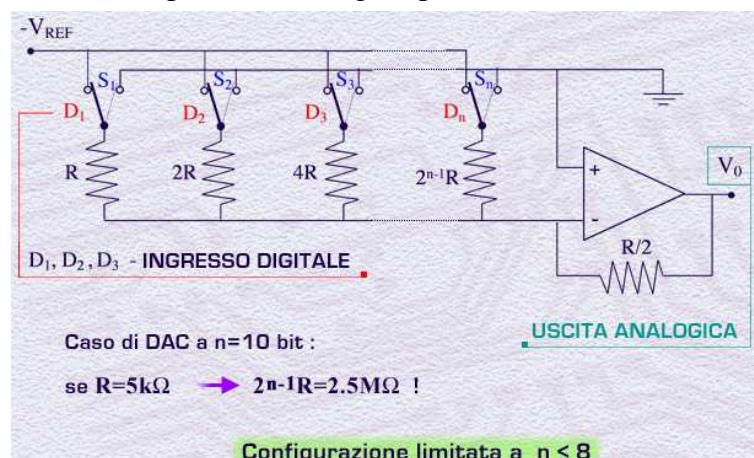
Si noti infine che la corrente erogata dal generatore di riferimento  $V_{ref}$  dipende dal codice di ingresso, il che comporta errori detti di "sovraposizione", prodotti dalla caduta di tensione



codice di ingresso, il che comporta errori detti di "sovraposizione", prodotti dalla caduta di tensione sulla resistenza di uscita di  $V_{ref}$ , caduta che varia da codice a codice. Altri errori vengono introdotti dalle non idealità dell'operazionale, come la corrente di polarizzazione di ingresso e la tensione di offset, i cui effetti sull'uscita variano anch'essi con la parola digitale di ingresso.

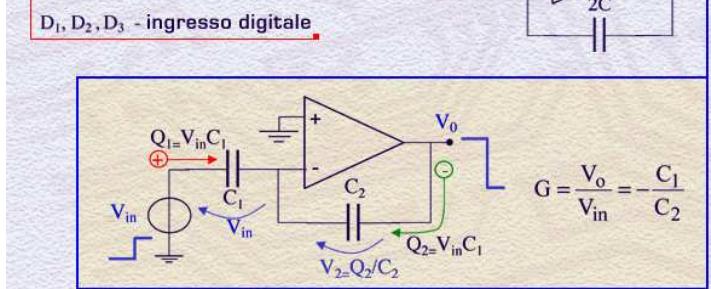
L'architettura del DAC a resistenze pesate può essere anche proposto con delle capacità al posto delle resistenze.

Per comprenderne il funzionamento basta ricordare che in una configurazione invertente realizzata con un amplificatore operazionale in cui sia presente



una capacità  $C_1$  serie verso il nodo di terra virtuale ed una capacità  $C_2$  di retroazione, il guadagno di tensione vale:

Grazie alla terra virtuale, infatti, quando in ingresso viene applicato un segnale  $V_{in}$ , la carica  $Q_1$  immagazzinata sulla capacità  $C_1$ , pari a  $C_1 V_{in}$ , sarà identica a quella  $Q_2$  immagazzinata sul condensatore di retroazione  $C_2$  ai cui capi si stabilirà quindi una tensione  $V_2$  pari a  $Q_2/C_2$ , da cui il guadagno.

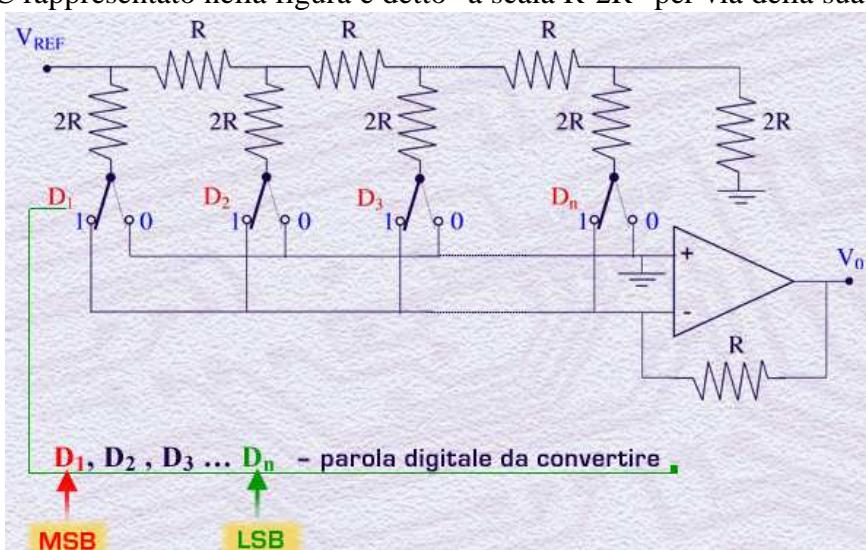


contributi, presenti o meno a seconda dello stato (chiuso o aperto) del relativo interruttore, vanno a sommarsi o meno sulla capacità di retroazione dell'operazionale e determinano un'uscita analogica proporzionale al codice di attivazione dei deviatori.

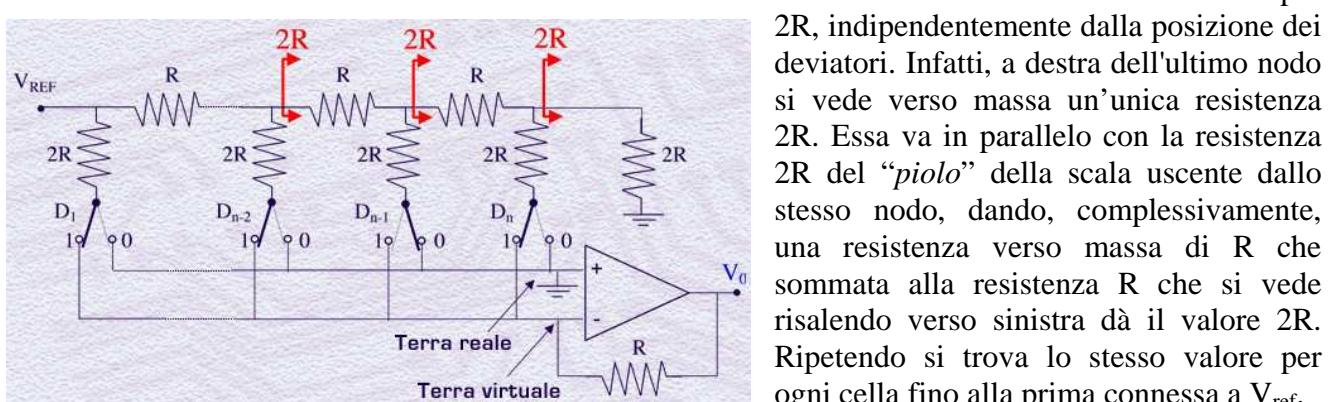
Poiché utilizza solo condensatori e interruttori, questa topologia si adatta bene ad essere impiegata nei sistemi MOS VLSI anche se, come nel caso resistivo, l'esigenza di costruire capacità sempre più grandi ne limita la realizzabilità a convertitori fino a 8 o 10 bit.

### Convertitori D/A a scala

Il problema delle resistenze di valore crescente presente nel precedente convertitore viene eliminato con la struttura di convertitore DAC rappresentato nella figura e detto "a scala R-2R" per via della sua configurazione. Esso è costituito infatti da una cascata di resistenze  $R$  che partono da una tensione di riferimento e arrivano fino a massa, a cui sono connessi come "pioli" dei resistori di valore  $2R$ . Su ogni piolo è sistemato un deviatore che dirige la corrente o verso la terra virtuale di un amplificatore a transresistenza o direttamente verso massa. I deviatori sono comandati dal valore 0 o 1 dei bit  $D_1 D_2 \dots D_n$  della parola digitale da convertire.



Si osservi innanzitutto che la resistenza a destra di ciascun nodo della scala verso massa vale sempre  $2R$ , indipendentemente dalla posizione dei deviatori. Infatti, a destra dell'ultimo nodo si vede verso massa un'unica resistenza  $2R$ . Essa va in parallelo con la resistenza  $2R$  del "piolo" della scala uscente dallo stesso nodo, dando, complessivamente, una resistenza verso massa di  $R$  che sommata alla resistenza  $R$  che si vede risalendo verso sinistra dà il valore  $2R$ .



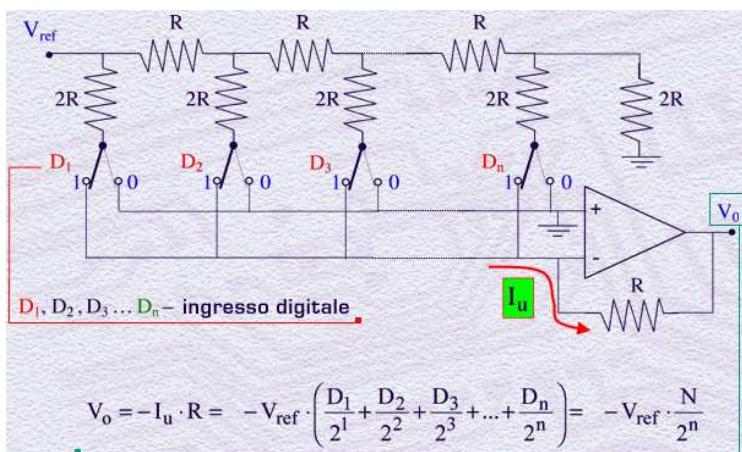
Ripetendo si trova lo stesso valore per ogni cella fino alla prima connessa a  $V_{ref}$ .

Se ora pensiamo alla corrente che circola nei rami, vediamo che quella che entra in un generico nodo si ripartisce egualmente nei due rami, quello con  $R$  verso il nodo a valle e quello con  $2R$  verso il deviatore. La corrente entrante nel nodo è quindi doppia di quella che scorre nella resistenza  $2R$  verso il deviatore. Ne consegue che le correnti nei vari rami dei deviatori sono, a partire da sinistra, nei seguenti rapporti:  $I_1, I_1/2, I_1/4, \dots, I_1/2^{n-1}$ .

Pertanto, come nei DAC a resistenze pesate, anche ora le correnti controllate dagli interruttori hanno valori che variano secondo pesi binari. Notando che  $I_1$  è data da  $V_{ref}/2R$ , la somma  $I_u$  delle correnti trasmesse dai deviatori ed afferenti alla terra virtuale dell'operazionale a seconda che il deviatore sia a 1 od a 0, è data da:  $\frac{V_{ref}}{2R} \left( \frac{D_1}{2^0} + \frac{D_2}{2^1} + \frac{D_3}{2^2} + \dots + \frac{D_n}{2^{n-1}} \right)$ ,

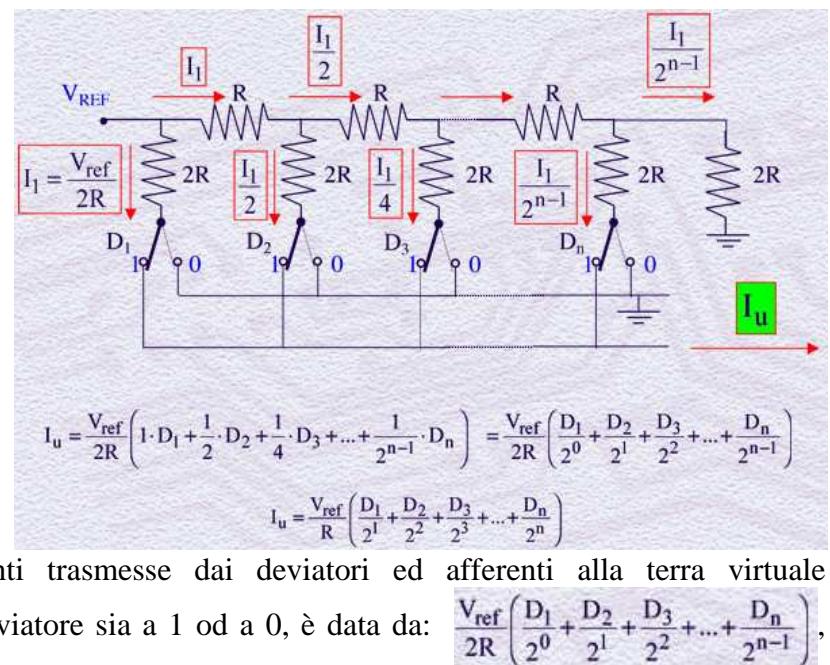
dove  $D_i$  è il generico bit nella parola digitale di ingresso con  $D_1$  il bit più significativo e  $D_n$  quello meno significativo, come al solito. Essa può essere ripresa nella forma più elegante:  $I_u = \frac{V_{ref}}{R} \left( \frac{D_1}{2^1} + \frac{D_2}{2^2} + \frac{D_3}{2^3} + \dots + \frac{D_n}{2^n} \right)$ .

Quando la corrente  $I_u$  viene inviata nella terra virtuale del convertitore corrente/tensione realizzato con un amplificatore operazionale reazionato con una resistenza  $R$  uguale a quelle impiegate nella rete a scala, la tensione di uscita  $V_o$  del convertitore assume il valore riportato in figura.



In questo tipo di DAC **non si hanno errori di sovrapposizione** poiché la corrente fornita dal generatore  $V_{ref}$  è sempre pari a  $V_{ref}/R$  indipendentemente dalla parola digitale di ingresso. Infatti, grazie al fatto che negli interruttori scorre sempre una corrente che, o va alla terra virtuale dell'operazionale o va direttamente a massa, la loro somma totale non cambia al cambiare della parola digitale.

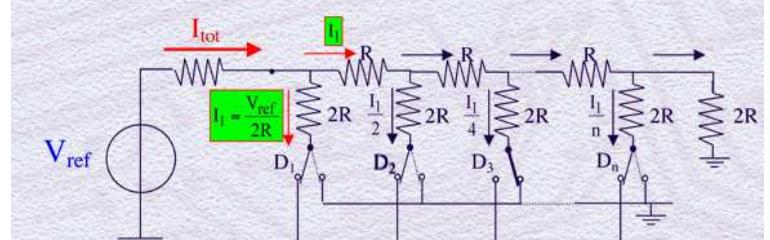
Anche in questo caso, come per i DAC a resistenze pesate, giocano un ruolo fondamentale nel definire le prestazioni del dispositivo l'accuratezza di  $V_{ref}$  e le caratteristiche dei deviatori (bassa resistenza di "on" e offset dell'operazionale).



Abbiamo così realizzato la conversione di una parola digitale ( $D_1D_2..D_n$ ) fornita in ingresso in una tensione analogica in uscita, frazione della tensione di fondo scala fissata con  $V_{ref}$  secondo il valore del numero intero corrispondente alla parola digitale. Per avere una tensione di uscita positiva, basta utilizzare una  $V_{ref}$  negativa.

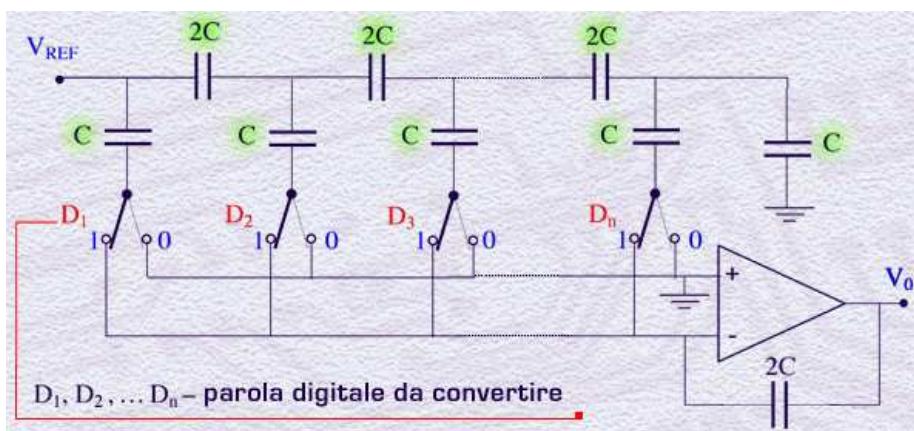
#### Assenza di errori di sovrapposizione

$$I_{tot} = \frac{V_{ref}}{R} \quad I_{tot} = I_1 + I_2 \quad \text{sempre uguale qualunque sia il codice}$$



Giocano un ruolo le resistenze di **on** dei deviatori, l'accuratezza di  $V_{ref}$  e l'offset dell'operazionale.

Anche l'architettura del DAC a scala R-2R può essere proposta con delle capacità al posto delle resistenze. Le capacità che costituiscono la scala hanno valore  $2C$  mentre i pioli hanno valore  $C$ . Lascerei a voi come esercizio studiarne il funzionamento e verificare che il livello di tensione dell'uscita è effettivamente proporzionale al codice di ingresso.



Anche in questo caso, poiché si utilizzano solo condensatori ed interruttori, la topologia si adatta benissimo ad essere impiegata nei sistemi MOS VLSI e non pone limitazioni al numero di bit utilizzabili perché tutte le capacità sono di dimensioni contenute.

### Convertitori D/A a partitore bipolare

Un altro dispositivo interessante è il DAC con struttura a partitore di tensione. Esso è molto semplice dal punto di vista costruttivo ed ha un principio di funzionamento facile da capire.

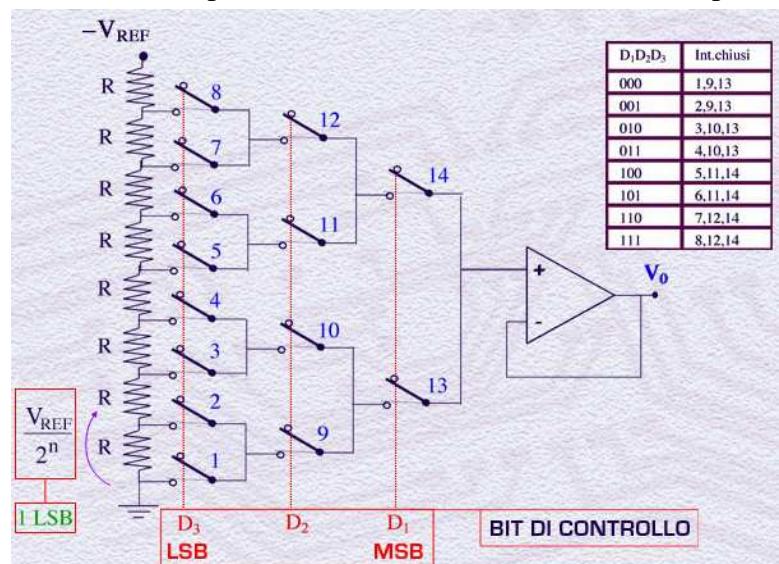
Una fila di  $2^n$  resistenze uguali in serie, collegate tra la tensione di riferimento e massa, suddivide la tensione Vref in  $2^n$  valori. Spostandosi da massa a Vref, si vede che ogni terminale della rete di resistenze presenta una tensione crescente con incrementi pari a  $V_{ref}/2^n$ , corrispondenti a 1 LSB. Il valore corrispondente alla codifica numerica in ingresso al DAC viene selezionato attraverso la chiusura opportunamente combinata di interruttori MOS comandati dagli n bit della parola, come sintetizzato nella tabella della figura. Il buffer di tensione in uscita garantisce la possibilità di erogare una sufficiente corrente al carico senza che ciò penalizzi la precisione della partizione resistiva.

Questo circuito offre il vantaggio di garantire una **intrinseca monotonicità** della conversione perché, almeno fintanto che gli interruttori funzionano correttamente, la tensione di uscita non può che crescere al crescere del codice di ingresso.

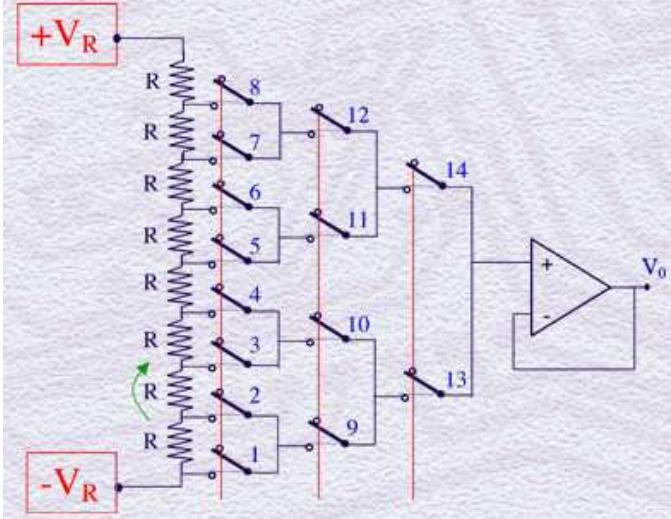
Anche in questo caso, l'elevato numero di resistori e di deviatori limita l'impiego di questa architettura a **non più di 12 bit**. Si consideri che già con 10 bit, c'è bisogno di 1024 resistori uguali che, per dissipare poca potenza, devono anche essere di valore medio-alto e quindi occupare uno spazio sul chip considerevole.

Importante è anche valutare l'effetto prodotto sull'uscita dalla corrente di polarizzazione ( $I_{bias}$ ) dell'operazionale e dalla corrente di perdita degli interruttori MOSFET che a seconda che siano chiusi o aperti perturbano la caratteristica in maniera non uniforme introducendo non linearità. Da ultimo, non si pensi che i DAC siano necessariamente unipolari, nel senso che la grandezza di uscita mantenga sempre lo stesso segno al variare del codice, come fino ad ora supposto. I DAC possono benissimo essere resi **bipolari**, con l'uscita analogica che assume sia valori positivi che negativi, mediante semplici accorgimenti circuituali, così da adattarsi alle più svariate esigenze d'impiego.

L'esempio più semplice è ovviamente quello del convertitore a partitore resistivo. Basterà che un estremo del partitore sia collegato ad una alimentazione positiva e l'altro ad una alimentazione

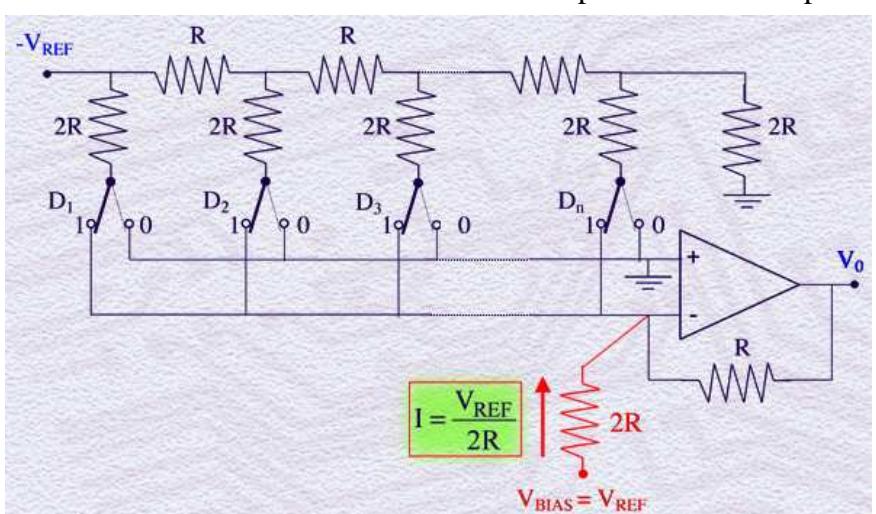


negativa  $-V_{ref}$ , per attivare in uscita, in base al codice di ingresso, valori di tensione compresi tra  $-V_{ref}$  e quasi  $+V_{ref}$ .



l'intervallo di uscita da  $-V_{ref}/2$  a  $+V_{ref}/2$  invece dell'intervallo originario  $0 \div V_{ref}$ , basterà fare in modo che la corrente di offset produca una tensione di uscita  $-V_{ref}/2$  quando il codice di ingresso contiene tutti zeri. Questo è ottenuto iniettando una corrente esattamente pari a  $V_{ref}/2R$  nella terra virtuale scegliendo  $V_{bias} = -V_{ref}$ .

Per meglio adattarsi all'uso con convertitori digitali-analogici bipolari, sono stati sviluppati opportuni codici della parola d'ingresso diversi dal semplice codice binario considerato finora.



Uscita	
Codice complementare	0111 0110 0101 0100 0011 0010 0001 0000 1111 1110 1101 1100 1011 1010 1001 1000
Codice complementare a 1	1111 1110 1101 1100 1011 1010 1001 1000 0111 0110 0101 0100 0011 0010 0001 0000 1111 1110 1101 1100 1011 1010 1001 1000
7/8 $V_{FS}$	3/4 5/8 1/2 3/8 1/4 1/8 +0 -0 -1/8 -1/4 -3/8 -1/2 -5/8 -3/4 -7/8
assente	1111 1110 1101 1100 1011 1010 1001 1000 0111 0110 0101 0100 0011 0010 0001 0000 1111 1110 1101 1100 1011 1010 1001 1000
Stesso codice	
Codice binario traslato	
Codice segno-valore	

Un altro codice è quello a **complemento a uno** che usa il solito codice binario per i valori positivi, mentre i negativi sono ottenuti facendone il complementare a 1 bit per bit.

### Esercizio 1

Disegnare un DAC a capacità pesate a 3 bit e calcolare i valori ideali della tensione di uscita in corrispondenza di tutte le combinazioni di ingresso possibili, quando  $V_{ref}=5V$  e  $C=1pF$ .

### Esercizio 2

Progettare un DAC a scala R-2R con uscita bipolare che sia compresa tra  $-5V$  e  $+5V$ , utilizzando un codice segno-valore.

Si noti a questo punto che il peso del bit meno significativo è raddoppiato rispetto al caso unipolare per riuscire, con lo stesso numero di bit, a coprire una dinamica del DAC doppia. Laddove l'uscita è fornita da una struttura operazionale a convertitore corrente-tensione, come ad esempio nei convertitori a scala o a resistenze pesate, si può facilmente rendere bipolare la risposta del sistema iniettando in terra virtuale una opportuna corrente di offset. La resistenza aggiunta al circuito a scala R-2R della figura sotto, svolge appunto questa funzione: se si vuole raddoppiare la dinamica di uscita del convertitore portandolo a coprire

Questo tuttavia può ancora fornire una uscita bipolare con una semplice traslazione pari al fondo scala.  
A titolo di esempio, citiamo la **codifica segno-valore** in cui il MSB rappresenta il segno del numero e tutti gli altri bit rappresentano la grandezza del numero, uguale sia che sia negativo o positivo. Nel caso del codice riportato nella tabella il primo bit della parola a 0 prelude ad un numero positivo, mentre un 1 prelude ad una parola negativa. Ne risultano due codici per l'uscita analogica a 0.

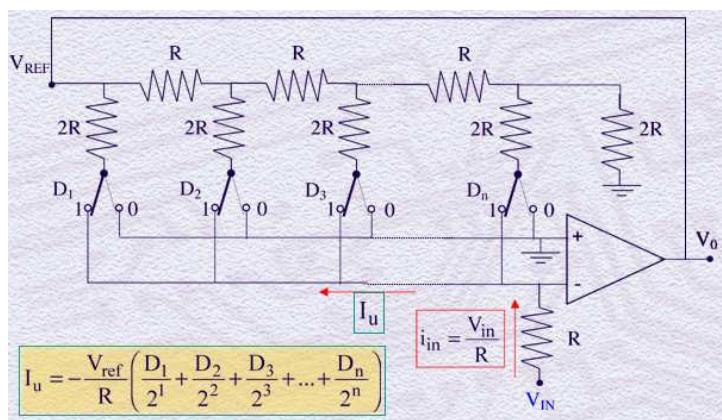
## Lezione n. 42 : Applicazioni dei DAC

### DAC usato come amplificatore a guadagno variabile

I DAC trovano applicazione non solo “direttamente” nella generazione di un segnale analogico funzione di un segnale di ingresso digitale, ma anche come parti di circuiti elettronici più complessi dove se ne sfrutta la topologia interna senza direttamente sfruttarne la funzione. Un primo esempio tra queste possibili applicazioni non convenzionali di un DAC è mostrato nella figura dove si utilizza un DAC a scala R-2R per realizzare un amplificatore a guadagno variabile digitalmente.

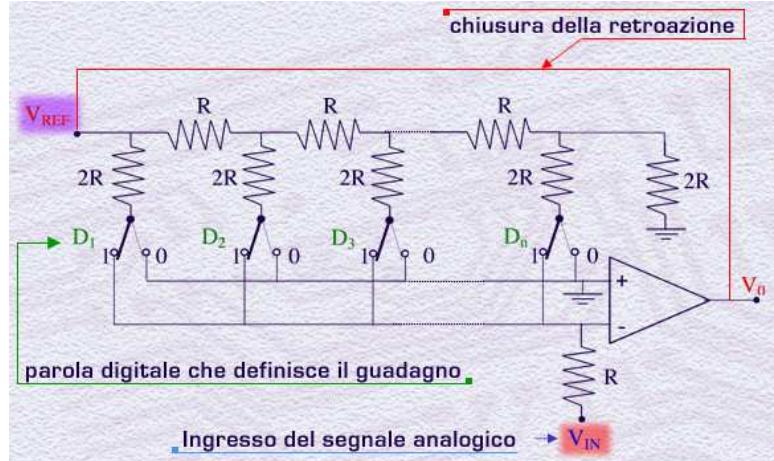
Il circuito del DAC è modificato, rispetto a quanto visto nella lezione scorsa, solo per quanto riguarda due connessioni esterne. Innanzitutto la resistenza  $R$ , che prima era collegata all'uscita dell'operazionale e formava la resistenza di retroazione dell'amplificatore a transresistenza, è ora direttamente accessibile dall'esterno e ad essa applichiamo il segnale di tensione analogico  $V_{IN}$  che vorremo amplificare.

La seconda modifica stà nell'uscita dell'operazionale che ora è collegata direttamente al morsetto  $V_{REF}$ , in modo da realizzare una retroazione negativa in cui tutta la rete di resistenze a scala costituisce l'elemento di retroazione. Tutto è diventato quindi equivalente ad un amplificatore reazionato invertente come quelli già visti molte volte in questo corso, solo un po' più complicati circuitalmente. Vediamone il funzionamento.



applicata al circuito. Ricordando che  $V_{REF}$  corrisponde all'uscita  $V_o$  del circuito, si ottiene la relazione finale tra ingresso ed uscita pari al valore di  $V_o$  espresso in figura. Il circuito quindi amplifica il segnale  $V_{IN}$  di ingresso da un minimo di 1 ad un massimo di  $2^n$  volte a seconda che la parola digitale corrisponda a  $2^n$  o ad 1. Ad esempio, utilizzando un DAC a 8 bit, si può realizzare un amplificatore invertente il cui guadagno può assumere 255 valori finiti diversi, da 256 a circa 1, variabili con passo unitario. Si noti che la condizione con tutti i bit a 0 non dovrebbe essere ammessa, in quanto produrrebbe una uscita saturata a una delle due alimentazioni dell'operazionale, a seconda della polarità dell'ingresso, perché corrisponde all'apertura dell'anello di retroazione.

Come esercizio, provate ad immaginare possibili utilizzazioni di questi amplificatori variabili digitalmente ed a riconoscerli negli apparati elettronici che vengono usati quotidianamente.



Il segnale di tensione  $V_{IN}$  attiva nella resistenza  $R$ , che è collegata alla terra virtuale dell'amplificatore, una corrente  $I_{in} = V_{IN}/R$ . Questa dovrà essere identica alla corrente che è prelevata alla terra virtuale da  $V_{REF}$  attraverso la rete a scala R-2R, e che già avevamo calcolato essere pari al valore in figura.

Eguagliando le due correnti ed accorpando gli addendi, si ottiene per  $V_{IN}$  il valore in figura dove  $N$  è il numero intero corrispondente alla parola digitale di  $n$  bit

Uguagliando le due correnti ... :

$$V_{in} = -\frac{V_{ref}}{R} \left( \underbrace{\frac{D_1}{2^1} + \frac{D_2}{2^2} + \frac{D_3}{2^3} + \dots + \frac{D_n}{2^n}}_{\frac{N}{2^n}} \right) \rightarrow V_{in} = -V_{ref} \frac{N}{2^n}$$

... si ottiene il guadagno dell'amplificatore :

$$V_o = -V_{in} \frac{2^n}{N}$$

con DAC a 8 bit, si ottiene un amplificatore con  $-1 < G < 256$  a passi unitari.

## Potenziometri digitali

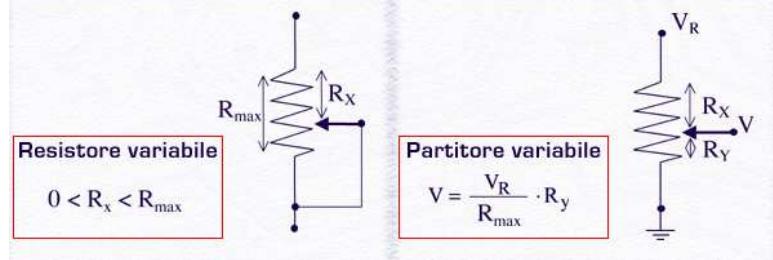
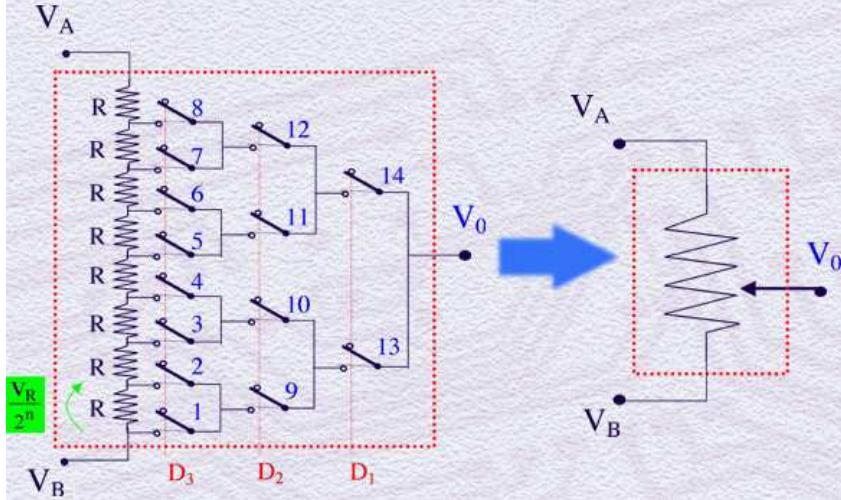
Nei circuiti elettronici sono a volte presenti delle resistenze variabili, la cui regolazione può essere fatta dall'esterno dopo che il circuito è stato costruito. Esse sono indicate negli schemi circuitali con il simbolo di una resistenza e di un terzo morsetto che rappresenta il pattino scorrevole che ne definisce la reale partizione. In effetti il simbolo riprende il modo in cui questi dispositivi sono costruiti: lungo il corpo del resistore (di solito di polvere di carbone o di film metallico) scorre un pattino metallico azionato da una piccola vite senza fine. Il pattino contatta quindi un punto intermedio del resistore e consente l'estrazione del segnale secondo una frazione variabile.

Se il pattino viene collegato ad uno dei due terminali del resistore, si realizza un bipolo resistore di valore variabile da zero (quando il pattino è contro al terminale a cui non è collegato elettricamente) ad un valore massimo fissato dal costruttore quando è a fine corsa. Se invece si usano tutti e tre i morsetti indipendentemente, si realizza un partitore di tensione variabile.

L'utilità di questi dispositivi nei circuiti spazia dalle regolazioni fini dei valori di tensione di polarizzazione di un punto del circuito alla modifica dei rapporti di resistenze in un amplificatore per variarne il guadagno.

Utilizzando, potremmo dire "impropriamente", un DAC a partitore di tensione si può realizzare un potenziometro digitale. Vediamo come.

Riportiamo innanzitutto la rete resistiva e l'albero degli interruttori di un DAC a partitore di tensione.



Si vede facilmente che il terminale  $V_o$  si porta alla tensione del nodo selezionato dalla posizione degli interruttori. Poiché la differenza di potenziale  $V_A - V_B$  applicata agli estremi del DAC viene divisa in 8 parti dal partitore resistivo, avremo a disposizione sugli 8 nodi del partitore la tensione di partizione accessibile alla rete di interruttori che, in base al codice di ingresso, collega un solo nodo per volta all'uscita.

Il terminale  $V_o$  si comporta quindi

in maniera simile al cursore centrale di un potenziometro i cui estremi siano collegati alle tensioni  $V_A$  e  $V_B$ , con la differenza che in questo caso la variazione di tensione del cursore è discreta tra i due estremi  $V_A$  e  $V_B$  invece che continua come in un potenziometro meccanico. In verità aumentando il numero di bit  $n$ , il numero di passi in cui viene suddivisa la resistenza complessiva ai capi del partitore aumenta e la differenza con i potenziometri meccanici sfuma fino, in pratica, a scomparire.

Il sistema mostrato nella figura prende effettivamente il nome **potenziometro digitale** perché il valore della resistenza di partizione viene modificata tramite un parola digitale fornita gli in ingresso.

I vantaggi principali di questi dispositivi sono:

- Assenza di usura meccanica e insensibilità alle vibrazioni, grazie alla regolazione tutta elettronica
- Tempi di regolazione più rapidi
- Facilità e certezza di riproduzione della posizione del "cursore", che non si ha mai quando si usa un potenziometro meccanico
- Minor ingombro rispetto ai trimmer meccanici.

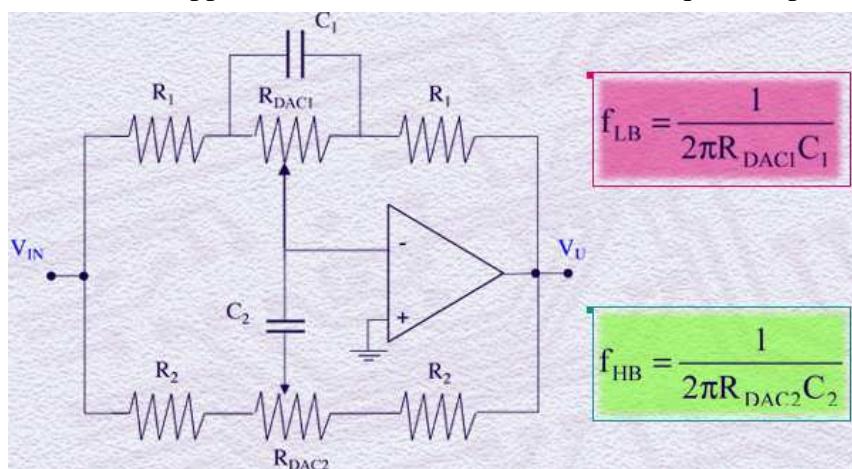
- Trovano largo impiego soprattutto nei sistemi elettronici di più recente progettazione che lavorano a bassa tensione.
- Traggono beneficio dalla presenza nel sistema di microcontrollori, che possono facilmente gestire via bus la parola di selezione della posizione del “cursoro”.
- Sono spesso dotati di “servizi” aggiuntivi, come ad esempio un reset interno che forza la posizione del cursore al centro della dinamica quando si accende il sistema.
- Lo **svantaggio** maggiore di un loro uso in circuiti analogici discreti è nella limitata dinamica della differenza di potenziale  $V_A - V_B$  applicabile, imposta dai vincoli della tecnologia digitale adottata per la costruzione del chip e quindi normalmente limitata a circa 5V, anche se si trovano componenti che possono lavorare a tensioni maggiori.

Il potenziometro digitale trova impiego in numerosissime applicazioni. A parte quelle “ovvie” di partitore di tensione o di resistenza a due terminali, già accennate, una applicazione interessante è quella indicata nella figura.

Si tratta di un amplificatore operazionale invertente con la rete resistiva interna al DAC nel ramo di retroazione. Il guadagno di tensione del circuito è, come noto e nel caso ideale, pari a  $-R_B/R_A$  dove  $R_A$  ed  $R_B$  sono le frazioni del partitore del DAC tra il morsetto A ed il cursore e tra il morsetto B ed il cursore, rispettivamente. Al variare della parola digitale di comando del DAC, varia il rapporto  $R_B/R_A$  e quindi varia il guadagno dell'amplificatore. Il circuito svolge quindi la funzione di controllo digitale del guadagno ed è molto utilizzato nelle applicazioni audio.

Poiché i valori di  $R_A$  ed  $R_B$  sono vincolati tra loro dall'essere  $R_A + R_B = R_{max}$ , calcolare la relazione tra il guadagno dell'amplificatore ed il codice digitale di ingresso, verificando che essa è quasi logaritmica.

Un'ulteriore applicazione dei resistori variabili è quella riportata in questo circuito che svolge la funzione di filtro variabile per le frequenze audio, ed è chiamato “**controllo programmabile dei toni**”.

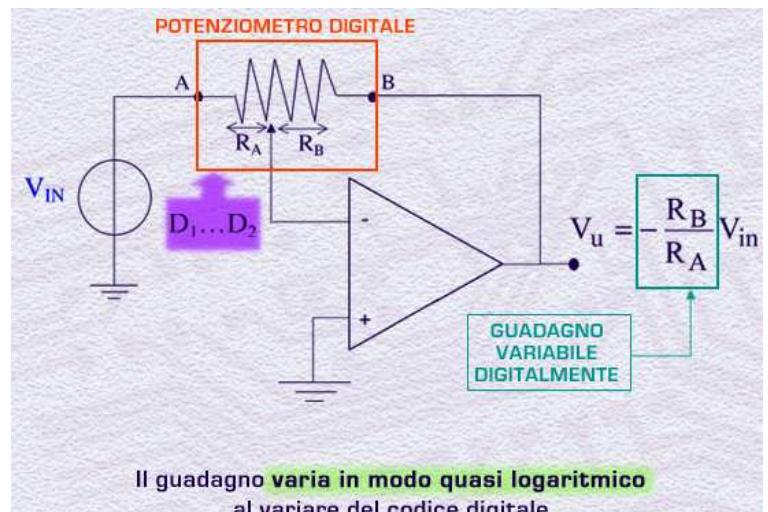


Complessivamente il circuito definisce un filtro di banda variabile, che esalta o attenua alcune frequenze dello spettro audio invece di altre.

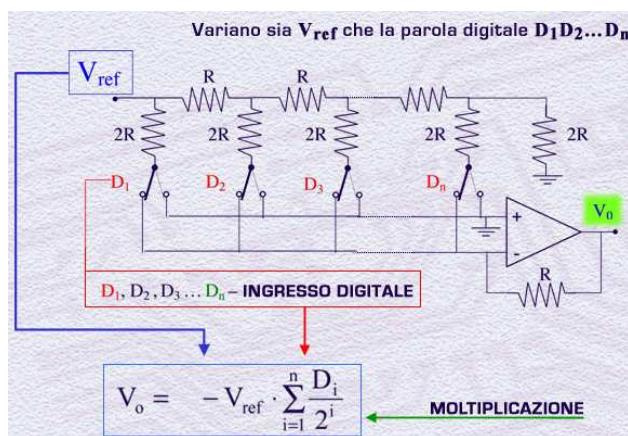
### Moltiplicatori con DAC

Abbiamo spesso fatto notare come sia fondamentale, per le prestazioni dei DAC, la stabilità e la accuratezza della tensione di riferimento utilizzata per “costruire” la risposta del convertitore che è proporzionale al prodotto della tensione di riferimento per la parola digitale di ingresso.

Tuttavia anche su questo punto si può assumere un punto di vista completamente opposto e pensare addirittura di lasciare variare nel tempo la tensione di riferimento durante il funzionamento del convertitore. Così facendo otteniamo che il segnale di riferimento variabile viene moltiplicato per la



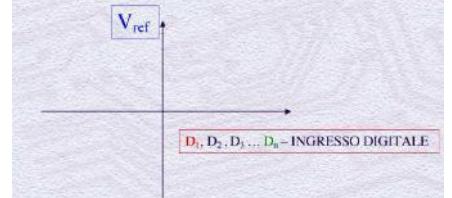
frazione rappresentata dal codice digitale di ingresso, cioè realizziamo un prodotto tra un segnale analogico ed un segnale digitale. Un tale uso del DAC gli fa prendere il nome di “**DAC moltiplicatore**”.



Si possono avere DAC moltiplicatori a uno, due, quattro quadranti, a seconda che riferimento e codice siano entrambi unipolari, almeno uno dei due sia bipolare, o lo siano entrambi.

Nei DAC che utilizzano l’architettura R-2R, ad esempio, non ci sono restrizioni sulla polarità di  $V_{ref}$ , che può

- Si possono avere moltiplicatori
  - a 1 quadrante ( $V_{ref}$  e codice unipolari).
  - a 2 quadranti ( $V_{ref}$  o codice bipolarì).
  - a 4 quadranti ( $V_{ref}$  e codice bipolarì).



essere sia positiva sia negativa, come pure sui codici, per cui questi DAC sono potenzialmente in grado di lavorare come moltiplicatori a due o quattro quadranti.

L’amplificatore a guadagno variabile digitalmente prima visto costituisce effettivamente una interessante applicazione del moltiplicatore a due quadranti.

## Esercizio 1

Utilizzando un DAC a 3 bit collegato in modo da formare un amplificatore a guadagno variabile come quello appena visto, calcolare il guadagno quando la parola digitale è "010" e calcolare le correnti in tutti i nodi del circuito quando  $V_{in}=1V$  e  $R=1k\Omega$ .

## Esercizio 2

- Come suggerito nel testo, provate a verificare che il guadagno del circuito amplificatore invertente che utilizza un potenziometro digitale varia in modo quasi logaritmico al variare del codice digitale.
- Come si sarebbe potuto realizzare un amplificatore a guadagno variabile linearmente, sempre utilizzando un potenziometro digitale.

## Esercizio 3

Provate a pensare a qualche applicazione in cui può risultare utile il moltiplicatore di un segnale analogico con un valore digitale come visto nel testo.

## Lezione n. 43 : Introduzione ai convertitori analogico-digitale (ADC).

L'operazione inversa alla conversione digitale/analogica vista nelle precedenti lezioni è la conversione di un segnale analogico in uno digitale. Un convertitore analogico-digitale (detto anche ADC dall'inglese Analog to Digital Converter) è un sistema che riceve in ingresso un segnale analogico continuo, normalmente una tensione oppure qualche volta anche una corrente, e lo converte in una parola digitale a  $n$  bit codificata in binario  $D_1D_2\dots D_n$ . Manteniamo sempre la convenzione che  $D_1$  sia il bit più significativo della parola e  $D_n$  quello meno significativo, e che  $N$  sia il numero decimale intero corrispondente alla parola digitale.

L'idea su cui si basa la conversione di un valore analogico in una parola digitale è molto semplice: dividere la tensione di fondo scala  $V_{FS}$  in  $2^n$  intervalli, o canali, ciascuno individuato da un codice univoco, ed associare il livello analogico  $V_{in}$  al corrispondente canale. Il sistema presenterà in uscita il codice del canale in cui cade  $V_{in}$ .

Lo schema già evidenzia la principale particolarità degli ADC, quella di attribuire lo stesso codice ad

un "intervallo di valori"  $V_{in}$ , esattamente tutti quelli che cadono in un singolo canale.

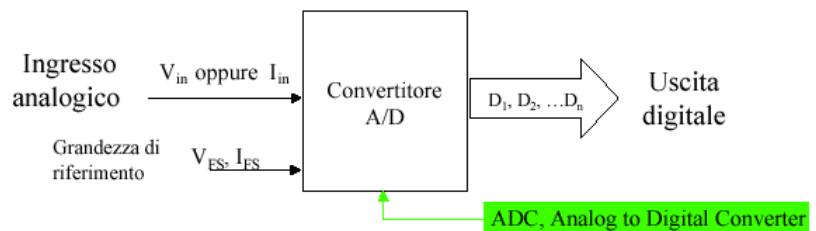
All'ampiezza di ogni canale, pari a  $V_{FS}/2^n$  si dà perciò il nome di RISOLUZIONE del convertitore: maggiore è il numero di canali, cioè maggiore è il numero  $n$  di bit del convertitore, migliore sarà la sua risoluzione perché si riuscirà ad attribuire uno stesso codice ad un insieme più ristretto di valori di tensione analogica di ingresso.

Le tecniche che si possono utilizzare per realizzare questa operazione di "associazione"

ampiezza-codice sono numerose, ciascuna con proprie caratteristiche di velocità e accuratezza di conversione.

Prima di esaminare e confrontare le più comuni architetture di convertitori A/D, cosa che faremo nelle prossime lezioni, vediamo ora le caratteristiche più generali degli ADC e come si definiscono i vari tipi di errore che ne caratterizzano le prestazioni.

Vediamo come si presenta la caratteristica ingresso-uscita di un ADC ideale. Sull'asse orizzontale vengono rappresentati i valori della tensione di ingresso  $V_{in}$  spesso, come nel nostro caso, riferiti alla tensione di fondo scala  $V_{FS}$ ; l'asse orizzontale ha quindi valori che variano con continuità tra 0 e 1. Sull'asse verticale vengono rappresentate le diverse parole del codice di uscita, otto nel nostro esempio di tre bit.



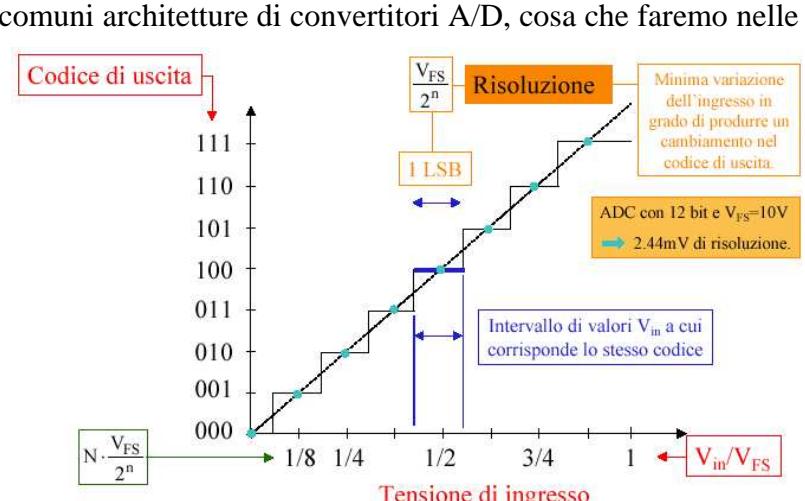
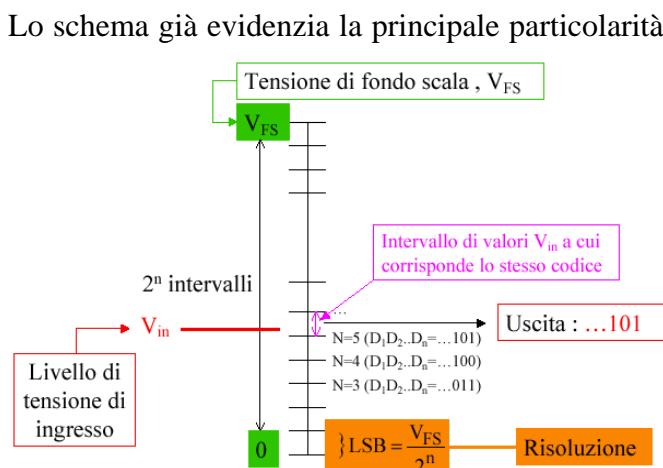
$D_1D_2\dots D_n$  ← Parola digitale di uscita

BIT meno significativo – LSB, Least Significant Bit

BIT più significativo – MSB, Most Significant Bit

$N$  è il numero decimale intero corrispondente alla parola digitale  $D_1D_2\dots D_n$ :

$$N = D_1 \cdot 2^{n-1} + D_2 \cdot 2^{n-2} + \dots + D_n \cdot 2^0$$



La curva che mette in relazione la tensione di ingresso con i codici di uscita è una scalinata: infatti come abbiamo già visto, lo stesso codice viene attribuito a un intervallo di valori di  $V_{IN}$ , valori a 2 cavallo delle tensioni multiple di  $1/8$  del fondo scala, intervallo che corrisponde al peso di un LSB del convertitore.

I punti di transizione da un codice al successivo sono in questo esempio posti al centro di ciascuna delle otto parti in cui è diviso il fondo scala così da centrare l'intervallo attorno al valore di riferimento. Questa scelta è intuitiva e comoda ma certamente arbitraria e si sarebbe potuto decidere di fare cadere i punti di transizione in una posizione qualsiasi dell'intervallo, inclusi gli estremi.

Con la scelta qui assunta i codici di uscita rappresentano fette di  $V_{IN}$  pari a  $1/8$  di  $V_{FS}$ , tranne i codici 000 (che copre solo il primo sedicesimo di  $V_{FS}$ ) e 111 (che copre i  $3/16$  finali di  $V_{FS}$ ). Per avere codici con pesi tutti uguali avremmo dovuto spostare il primo punto di transizione nella posizione  $V_{IN}/V_{FS}=1/8$ . Naturalmente, quando i bit del codice diventano 10, 12 o più, e quindi i gradini alcune migliaia, l'avere il primo e l'ultimo gradino di ampiezze diverse diventa assolutamente inessenziale e d'ora in avanti ci disinteresseremo di essi.

Nel convertitore ideale come da noi rappresentato i punti medi dei gradini stanno quindi tutti sulla stessa retta. La larghezza dei gradini è uguale a un LSB. In generale, con una parola di  $n$  bit, avremo che 1 LSB vale  $V_{FS}/2^n$ .

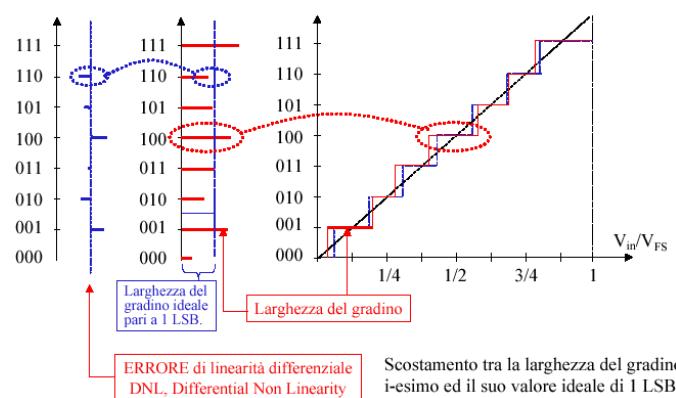
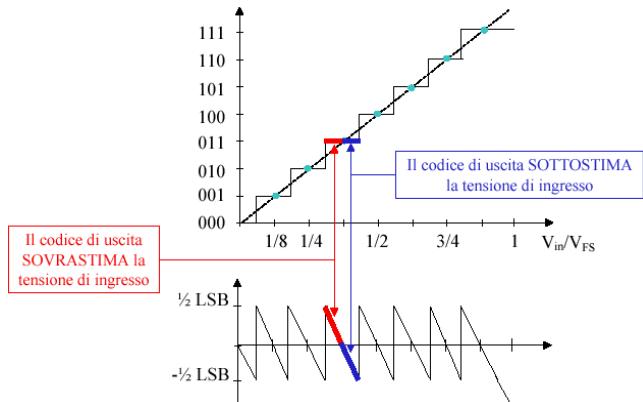
Come già accennato, l'LSB rappresenta la risoluzione del convertitore, ovvero la minima variazione dell'ingresso in grado di produrre un cambiamento nel codice di uscita.

Ad esempio, con  $V_{FS}=10V$  e  $n=12$ , tenendo conto che  $2^{12}=4096$ , avremo una risoluzione di (10 volt diviso 4096), pari a 2.44 mV: ogni volta che il segnale di ingresso varia di 2.44mV, il codice di uscita varia di una unità.

Nell'ADC ideale, i codici di uscita restano dunque costanti per un intervallo di tensioni di ingresso pari a un LSB. Perciò, al crescere con continuità di  $V_{IN}$ , ciascun codice di uscita dapprima sovrastima e poi sottostima la tensione di ingresso. Questo errore imputabile alla discretizzazione del segnale, e perciò detto errore di quantizzazione, è ineliminabile ed è rappresentato da un andamento a dente di sega come nella figura. L'escursione del dente di sega è di  $\pm 1/2$  LSB per tutti gli intervalli eccetto che per i due alle estremità.

Riassumendo, possiamo dire che, per un dato codice di uscita, la tensione di ingresso può trovarsi ovunque all'interno di un intervallo di quantizzazione largo un LSB.

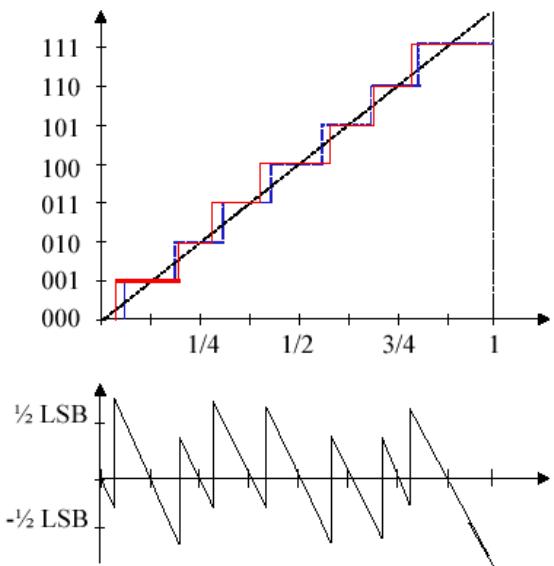
Contrariamente a quanto avviene nel caso ideale, per un convertitore reale i punti di transizione tra un codice e il successivo non cadono tutti sulla stessa retta perché i gradini non sono tutti larghi uguali tra loro e pari alla larghezza ideale. Per quantificare questa differenza tra il caso ideale ed il caso reale si introduce l'errore di linearità differenziale definito come la differenza tra la larghezza del gradino reale di un dato codice e la larghezza ideale del gradino, cioè un LSB.



giusto segno) tra i due gradini. Il grafico mostra proprio l'errore di linearità differenziale associato a ciascun codice. Esso è il dato fornito dai costruttori ed è indice della regolarità della scalinata.

Accidentalmente l'errore di linearità differenziale dei due codici estremi, 000 e 111 nel nostro caso, viene calcolato sottraendo alla larghezza del gradino reale rispettivamente 0.5 LSB e 1.5 LSB, coerentemente con la posizione scelta per i punti di transizione del primo e dell'ultimo codice.

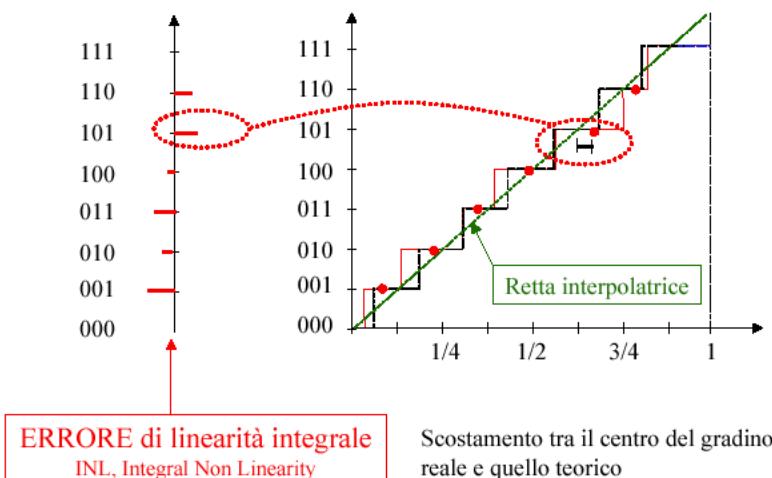
Il conseguente grafico dell'errore di quantizzazione della situazione reale appena descritta risulterà molto più irregolare di quello dell'ADC ideale. Infatti ad ogni gradino più largo di quello ideale corrisponde un errore di quantizzazione maggiore di  $\pm\frac{1}{2}$ LSB perché maggiore è l'intervallo dei valori di  $V_{in}$  a cui corrisponde quello stesso codice. Viceversa, per gradini più stretti, corrisponderà un errore minore di  $\pm\frac{1}{2}$ LSB. L'entità dell'errore di quantizzazione in eccesso a  $\pm\frac{1}{2}$ LSB rende l'idea degli scostamenti dei gradini dalla situazione ideale ed è anch'esso fornito dal costruttore. Un altro parametro che quantifica lo scostamento della gradinata di un ADC reale da quella ideale è l'errore di linearità integrale definito come la distanza tra il centro del gradino reale e la retta interpolatrice. Quest'ultima può essere o la bisettrice ideale del quadrante, o



la retta interpolatrice ai minimi quadrati dei centri dei gradini (detta anche di best fitting) o anche la retta interpolatrice che congiunge i punti corrispondenti alla prima e all'ultima transizione.

Come nei DAC, l'errore di linearità integrale è uguale alla somma dei suoi errori di linearità differenziale, fatto che ne giustifica il nome.

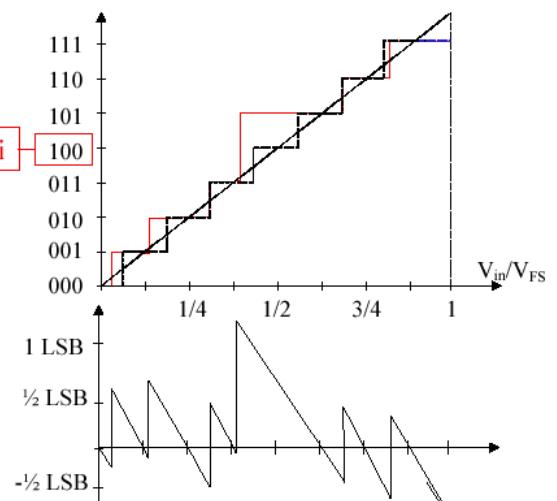
Un errore peculiare dei convertitori ADC è quello per un codice mancante. Nell'esempio della figura il codice di uscita passa da 011 a 101 quando l'ingresso supera di poco la tensione



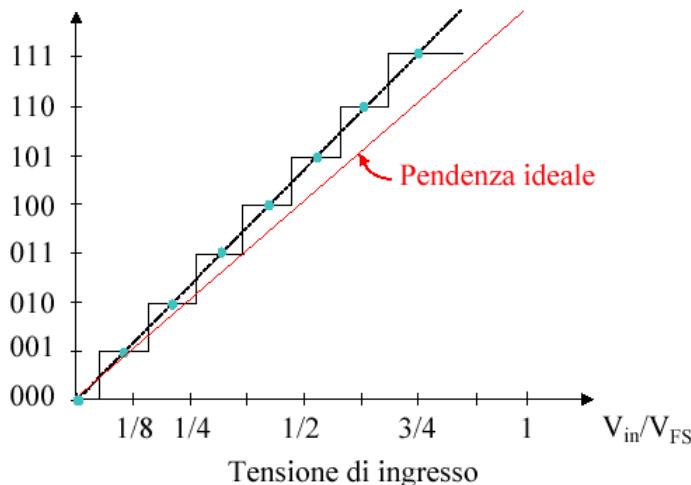
di  $3/8$  di  $V_{FS}$ . Il codice 100 non esce mai, per cui si suole dire che questo convertitore presenta un codice mancante. Un convertitore con un codice mancante ha necessariamente un errore di quantizzazione superiore ad 1 LSB, ed è proprio tramite l'analisi degli errori che si scoprono gli eventuali bit mancanti. I costruttori dichiarano se il loro prodotto soffre del problema di avere alcuni codici mancanti, anche se spesso tale specifica è riferita ad un numero di bit inferiore a quelli effettivamente disponibili nell'ADC. Ad esempio un ADC a 10 bit può essere garantito non avere codici mancanti su 9 bit, intendendo che ci potranno essere codici mancanti ma non adiacenti tra di loro in modo che si passerà da un codice ad un altro distante 2 LSB ma non di più.

**Codice che non uscirà mai**

Se manca un codice, l'errore di quantizzazione è necessariamente maggiore di 1 LSB



Un altro errore tipico degli ADC è "l'errore di guadagno", che si riscontra quando la pendenza della retta congiungente i punti della prima e dell'ultima transizione differisce da quella ideale: il convertitore presenta pertanto un errore di guadagno, dato dalla differenza tra la pendenza reale e quella ideale unitaria. Nella figura ad esempio è riportato il caso in cui il guadagno dell'ADC è maggiore dell'ideale perché i gradini della scalinata sono più piccoli di 1 LSB.

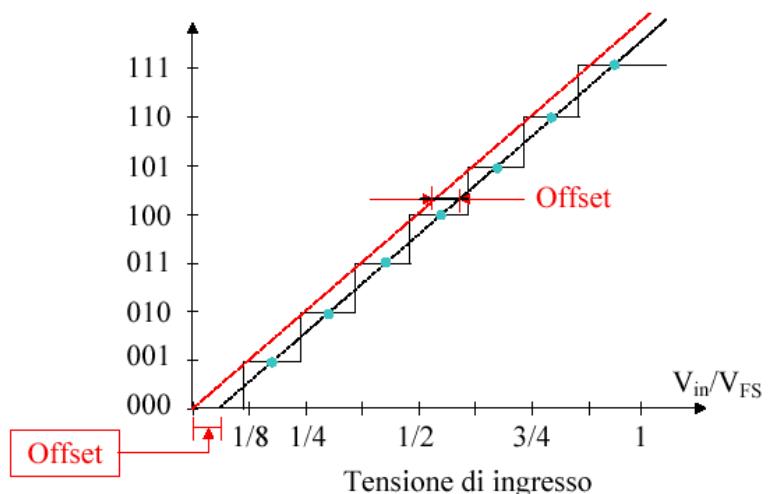


ingresso più alta del dovuto di quasi 0.5 LSB: questo determina un offset per il convertitore di quasi 0.5 LSB.

Tutti gli scostamenti visti rispetto al comportamento ideale dipendono poi dalla **temperatura** di funzionamento dell'ADC: le specifiche di un convertitore elencano pertanto anche i coefficienti di temperatura per gli errori di guadagno, offset e linearità.

Un buon convertitore avrà errori di linearità inferiori a 0.5 LSB, non presenterà codici mancanti e risulterà **monotono** su tutto l'intervallo di temperatura di funzionamento, cioè il codice di uscita cresce sempre al crescere della tensione di ingresso.

Anche un offset nella conversione produce una scalinata che non si appoggia su quella ideale. Lo scostamento in questo caso è lo stesso per ogni ampiezza del segnale di ingresso. Nell'esempio della figura la prima transizione di codice si ha per una tensione di



## Lezione n. 44 : ARCHITETTURE BASE dei CONVERTITORI ANALOGICO-DIGITALI (ADC).

La struttura della maggior parte dei convertitori analogico – digitale si basa sul confronto, attraverso un comparatore analogico, del segnale incognito  $V_x$ , supposto costante per tutto il tempo della conversione, con una tensione di riferimento variabile  $V_R(t)$ .

Durante la conversione,  $V_R$  viene fatto variare secondo criteri che differiscono da convertitore a convertitore, ma sempre con l'obiettivo di uguagliare, entro l'errore di quantizzazione del convertitore, la tensione incognita di ingresso.  $V_R(t)$  si stabilizzerà nell'istante  $t'$  in cui:

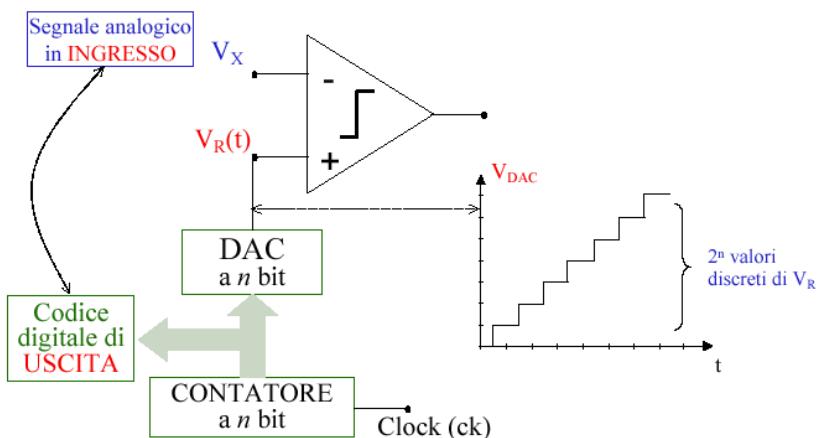
$$|V_x - V_R(t')| \leq 0.5 \text{ LSB}$$

A questo punto la logica del convertitore A/D

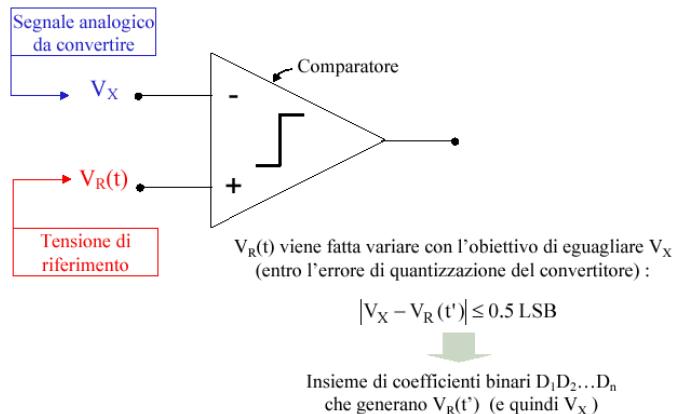
avrà individuato un insieme di coefficienti binari  $D_1D_2...D_n$  in grado di generare il valore  $V_R(t')$  e quindi in definitiva di rappresentare il valore analogico incognito  $V_x$ .

I convertitori differiscono tra loro proprio per la metodologia impiegata per generare l'insieme dei coefficienti  $D_i$ .

Il modo più ovvio e naturale di generare una tensione di riferimento variabile è di usare un DAC:



del convertitore. Quando l'uscita del DAC supera  $V_x$ , la logica del convertitore blocca il contatore, nel cui registro di uscita troviamo il codice cercato. Per come funziona, un convertitore di questo tipo è indicato come "*convertitore a contatore-rampa*". Uno schema a blocchi più completo del convertitore è mostrato nella figura: la conversione ha inizio quando il segnale di reset azzera il contatore e contemporaneamente alza l'uscita negata di un flip-flop set-reset che apre la porta al clock verso il contatore. L'uscita del contatore conta gli impulsi di clock: a ogni impulso l'uscita del DAC si incrementa di un LSB, assumendo l'aspetto della scalinata vista prima. Quando la tensione in uscita dal DAC supera  $V_x$ , il comparatore commuta, attiva il comando di Set del FF che abbassa l'uscita negata chiudendo la porta al clock e bloccando così il conteggio del contatore. Contemporaneamente il FF attiva, tramite l'uscita Q, un segnale di fine conversione (E.O.C. - End of Conversion) che può essere utilizzato per leggere il codice di uscita del contatore, che rappresenta il valore numerico del segnale analogico di ingresso. Data la metodologia di conversione impiegata, il convertitore ha la caratteristica

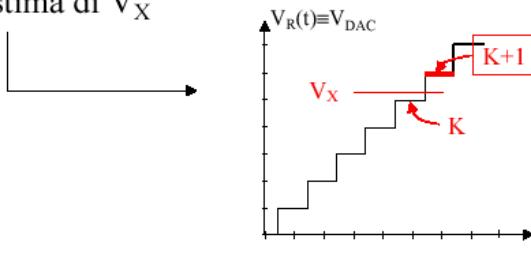


che il tempo di conversione varia proporzionalmente all'ampiezza di  $V_x$ ; si va da un tempo minimo di praticamente 0 secondi ad un tempo di conversione massimo  $T_{MAX}$  quando  $V_x$  è uguale alla tensione di fondo scala  $V_{FS}$  del convertitore, corrispondente a  $2^n$  periodi di clock. Detta  $f_{ck}$  la frequenza di clock, questo tempo vale:

$$T_{MAX} = 2^n / f_{ck}$$

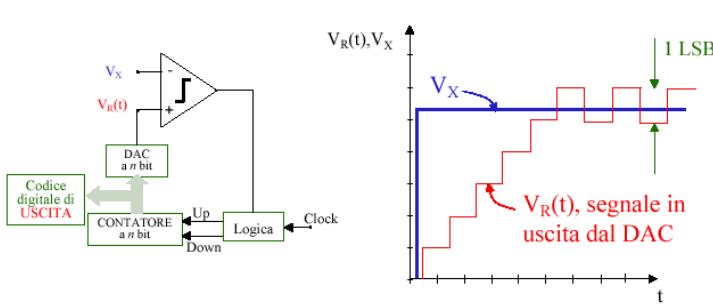
Dovendo aspettare tanti tempi di clock, la velocità di conversione è relativamente bassa e questo costituisce il principale svantaggio di questo schema di convertitore. Per esempio, con un DAC da 10 bit e un clock da 1MHz occorrono 1.024 ms per completare la conversione di un segnale  $V_x$  uguale al fondo scala. Con un simile ADC non possono quindi essere previste più di 1000 conversioni al secondo.

- Semplicità circuitale
- Poco costoso
- Sovrastima di  $V_x$



contatore resterà immagazzinato il codice corrispondente al gradino  $k+1$ .

L'architettura del convertitore a contatore-rampa può facilmente essere modificata in modo tale che la parola di uscita segua le variazioni del segnale di ingresso, restando agganciata ad esso entro  $\pm 1$  LSB invece che ripartire da 0 ogni volta che si vuole campionare il nuovo valore  $V_x$ . Lo schema a blocchi di un convertitore di questo tipo, chiamato "convertitore a inseguimento" è mostrato nella figura. Un contatore Up-Down comanda il DAC, la cui uscita viene confrontata con il segnale di ingresso incognito  $V_x$  da un comparatore. L'uscita del comparatore, tramite un blocco logico, stabilisce se il segnale di clock debba arrivare al morsetto Up o al morsetto Down del contatore: se l'uscita del DAC risulta minore di  $V_x$  il contatore dovrà incrementare il suo conteggio, in caso contrario dovrà decrementarlo.



Se  $V_x$  resta costante dopo l'aggancio,  $V_{DAC}$  continua a "cavalcarlo" sovrastimandolo o sottostimandolo per non più di 1 LSB.

Il tempo di conversione,  $T_{conv}$ , varia proporzionalmente a  $V_x$ :

$$T_{MIN} \text{ per } V_x = 0V \rightarrow T_{MIN} = 0 \text{ s}$$

$$T_{MAX} \text{ per } V_x \approx V_{FS} \rightarrow T_{MAX} = \frac{2^n}{f_{ck}}$$

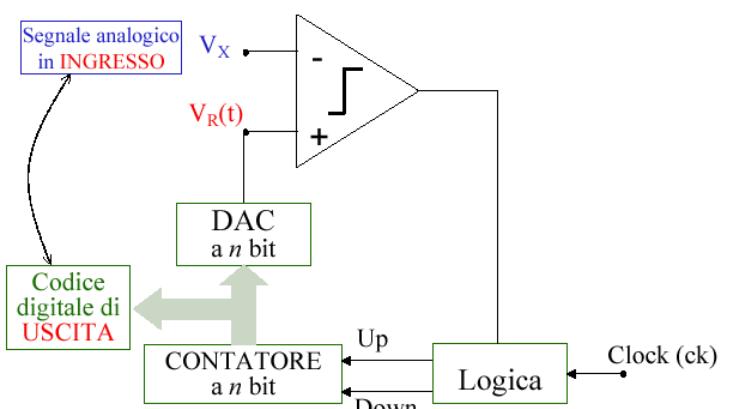
Velocità di conversione relativamente bassa

$$\text{Esempio : ADC a 10 bit e } f_{ck}=1 \text{ MHz} \rightarrow T_{MAX}=1.024 \text{ ms}$$

Al massimo possono essere previste ~1000 conversioni al secondo

Il principale vantaggio di questo convertitore risiede nella sua semplicità architettonale che richiede, per la realizzazione circuitale, una minima quantità di blocchi funzionali, risultando quindi poco costoso.

Si noti inoltre che il codice binario nel contatore rappresenta il più piccolo valore della tensione  $V_{DAC}$  che risulta maggiore di  $V_x$ , ma non è detto che questo sia anche il valore di  $V_{DAC}$  più vicino a  $V_x$ , entro 0.5 LSB. Se guardiamo la figura, infatti, vediamo che  $V_x$  è più vicino al gradino  $k$  che non al gradino  $k+1$ , ma nel



All'inizio di una sessione di conversione, come per esempio all'accensione del convertitore, l'uscita del DAC appare come una scalinata che sale o scende fino a che non "aggancia" il segnale di ingresso  $V_x$ .

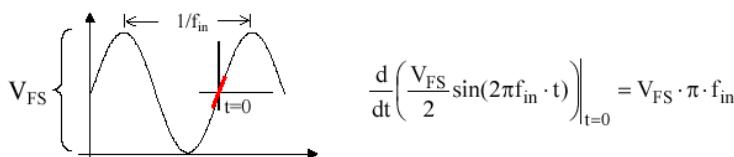
Se  $V_x$  resta costante dopo l'aggancio,  $V_{DAC}$  continua a "cavalcarlo" sovrastimandolo o sottostimandolo per non più di 1 LSB.

Se  $V_x$  varia lentamente,  $V_{DAC}$  ne segue le variazioni ad ogni colpo di clock. Fintanto che l'uscita del DAC resta agganciata all'ingresso, la parola digitale immagazzinata dal contatore ne rappresenta la conversione numerica, che può essere letta in uscita.

Se  $V_x$  cambia troppo rapidamente, il DAC perde l'aggancio perché la sua uscita può variare solo di 1 LSB ogni colpo di clock. Durante questo periodo di tempo la parola immagazzinata dal contatore non è più rappresentativa del valore  $V_x$  dell'ingresso.

La situazione tipica in cui si sia in presenza di brusche variazioni dell'ingresso è quella di segnali provenienti da più sensori, canalizzati da un multiplatore verso il convertitore. Questo convertitore non è quindi adatto all'impiego in questo tipo di contesti.

Massima velocità di variazione dell'ingresso sinusoidale  $V_x$ :



Massima velocità di variazione dell'uscita del DAC:

$$\frac{1 \text{ LSB}}{\text{Periodo di clock}} = 1 \text{ LSB} \cdot f_{ck} = \frac{V_{FS}}{2^n} \cdot f_{ck}$$

$$V_{FS} \cdot \pi \cdot f_{in} < \frac{V_{FS}}{2^n} \cdot f_{ck} \quad \text{da cui} \quad f_{in} < \frac{f_{ck}}{\pi \cdot 2^n}$$

la derivata rispetto al tempo della sinusoide e calcolandola per  $t=0$ , si trova facilmente che la pendenza massima vale

$$V_{FS} \cdot \pi \cdot f_{in}$$

La massima velocità di variazione dell'uscita del DAC è, come già detto, pari a 1LSB in un periodo di clock, cioè  $1 \text{ LSB} \cdot f_{ck}$ , ovvero  $(V_{FS}/2^n) \cdot f_{ck}$ , ricordando che 1 LSB è per l'appunto uguale a  $V_{FS}/2^n$ . Deve quindi valere la condizione:

$$V_{FS} \cdot \pi \cdot f_{in} < (V_{FS}/2^n) f_{ck}$$

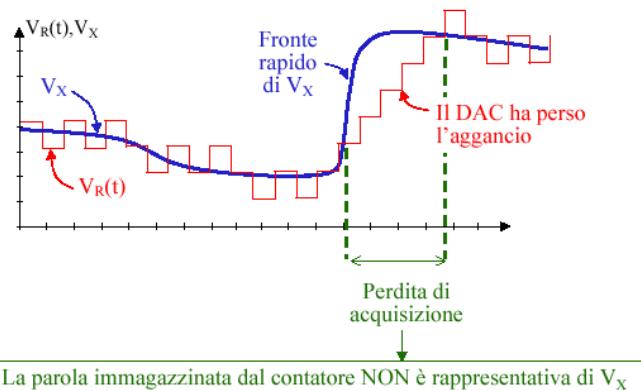
da cui si ricava la relazione:

$$f_{in} < f_{ck} / \pi \cdot 2^n$$

Per un convertitore da 10 bit con un clock da 1 MHz, la massima frequenza inseguibile di una sinusoide che copre tutta la dinamica di ingresso è di appena 310 Hz.

La lentezza degli ADC esaminati fino ad ora deriva sostanzialmente dalla strategia di impostazione del segnale di riferimento disponibile all'uscita del DAC, che varia per salti di un solo LSB. Ma questa non è la sola strategia possibile.

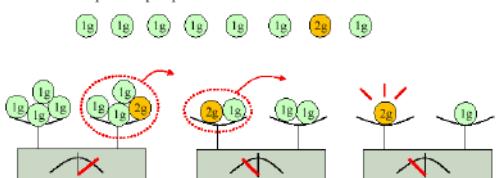
Si pensi al ben noto e semplice rompicapo in cui si chiede quale sia il numero minimo di pesate con cui si può estrarre, da un insieme di 8 palline identiche per forma, di cui sette hanno lo stesso peso e una pesa maggiore delle altre, la pallina a peso maggiore. Il procedimento da seguire è il seguente: si suddividono le palline in due gruppi di 4, si mettono sui piatti di una bilancia e si trova il gruppo che pesa di più. Lo si suddivide a sua volta in due insiemi di due palline, si mettono sui piatti della bilancia e si trova di nuovo l'insieme più pesante.



Per evitare la perdita di acquisizione è importante assicurarsi che la massima velocità di variazione prevista per  $V_x$  non superi quella che il DAC è in grado di seguire.

Per calcolare questo limite, confrontiamo la frequenza sinusoidale massima del segnale all'ingresso con la frequenza di clock del convertitore. Una sinusoide di frequenza  $f_{in}$  e ampiezza  $V_{FS}/2$ , in grado quindi di coprire tutta la dinamica di ingresso del convertitore, presenta la sua pendenza massima nell'origine. Facendo

Trovare la pallina più pesante ?

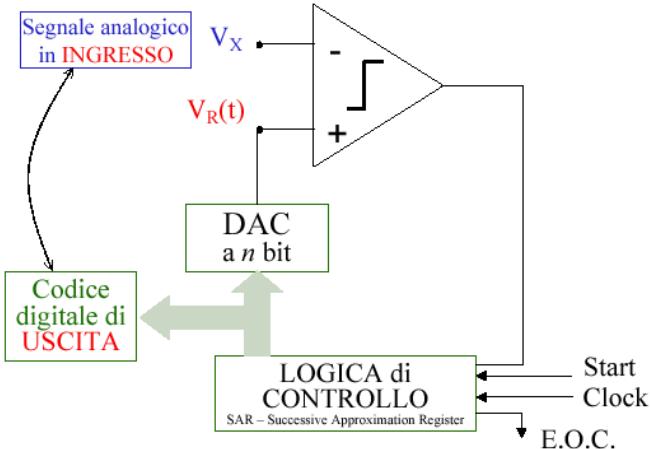


Partendo da 8 palline, sono bastate 3 pesate !

In generale : partendo da  $2^n$  elementi, bastano  $n$  passaggi.

Si dividono le sue due palline sui due piatti e si trova la più pesante. Sono bastate solo 3 pesate per giungere alla fine ! Questo tipo di strategia è detta di "ricerca binaria". Data la sua efficienza, la

ricerca binaria è effettivamente applicata in una architettura di convertitore A/D chiamato convertitore binario o più frequentemente convertitore ad approssimazioni successive. La struttura di un ADC ad approssimazioni successive è mostrata schematicamente nella figura. Il convertitore è sempre costituito da un comparatore che confronta il segnale analogico da convertire  $V_x$  con il segnale di riferimento prodotto da un DAC, il quale però ora è gestito da una logica di controllo che, per trovare la miglior approssimazione a  $V_x$ , utilizza appunto una metodologia a ricerca binaria. Proprio per via del suo modo di operare, la logica di



controllo viene normalmente chiamata SAR, dall'inglese Successive Approximation Register.

Per capirne il funzionamento, mettiamoci nella condizione semplice di avere un convertitore a 3 bit e supponiamo di avere un segnale analogico  $V_x$  da convertire che vale poco di più della metà del fondo scala del convertitore.

Come si vede dalla figura, dopo l'azzeramento, al primo colpo di clock la logica di controllo imposta la parola corrispondente a metà del fondo scala. Nel nostro esempio di ADC a tre bit questa parola è 100. Si noti che in generale, prendere metà del fondo scala corrisponde ad avere il primo bit di una parola a 1 e tutti gli altri a 0.

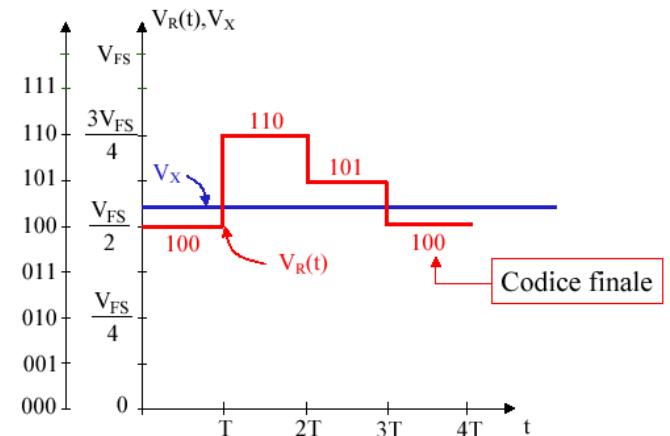
Il DAC si incarica di generare il corrispondente livello di tensione, che viene inviato al comparatore per il confronto con il segnale di ingresso. In base all'esito del confronto la logica stabilisce se tenere o togliere il primo bit dalla parola: lo tiene se il segnale di ingresso supera quello generato dal DAC, lo toglie in caso contrario. Nel nostro esempio, data la posizione di  $V_x$ , il MSB viene conservato.

Al secondo colpo di clock, la logica va a guardare in quale metà della metà sopra sta il segnale. Per posizionarsi in questa nuova metà, bisogna semplicemente mettere a 1 il successivo bit della parola digitale, cioè generare nel nostro caso la parola 110, incrementando così di  $V_{FS}/4$  la  $V_{DAC}$  precedente. Essa ora risulta maggiore di  $V_x$  e quindi il comparatore scatta ed impone alla logica di togliere questo 1 perché  $V_x$  sta nella nuova metà sotto. Al terzo colpo di clock la logica prova a vedere in quale delle due nuove metà sta il segnale  $V_x$  andando ad accendere l'ultimo bit, generando cioè la parola 101. Anche in questo caso poiché l'uscita del DAC risulta ancora superiore a  $V_x$ , la logica esclude questo ultimo bit e presenta il codice finale 100, che rappresenta la conversione numerica dell'ingresso. Accidentalmente in questo esempio 100 corrisponde al primo codice provato, ma solo dopo avere provato tutti gli uni nella parola, abbiamo potuto sapere che è quello giusto !

La conversione si conclude quindi dopo tre periodi di clock per il nostro convertitore a tre bit. In generale si conclude dopo  $n$  periodi per un convertitore a  $n$  bit. Il tempo di conversione è quindi molto piccolo e pari a:

$$T_{conv} = n/f_{ck}$$

Ad esempio, con un convertitore da 10 bit e un clock da 1 MHz, per concludere una conversione ci vogliono solo 10 microsecondi. Di conseguenza con questo convertitore si potrebbe raggiungere un tasso di 100.000 conversioni al secondo.



I fattori che limitano la velocità di questi ADC sono il tempo di assestamento del DAC al livello di uscita nominale del codice in ingresso e il tempo di risposta del comparatore per eseguire il confronto tra i segnali di ingresso, in particolare quando differiscono di poco. Un ulteriore ritardo viene introdotto dalla SAR, soprattutto al crescere di n, ma è normalmente trascurabile rispetto ai primi due.

Per convertitori ad n bit  $\rightarrow T_{Conv} = \frac{n}{f_{ck}}$

Velocità di conversione elevata

Esempio : ADC a 10 bit e  $f_{ck}=1$  MHz  $\rightarrow T_{Conv} = 10 \mu s$

Si potrebbe raggiungere un tasso di 100.000 conversioni/s

Fattori limitanti la frequenza di clock,  $f_{ck}$  :

- tempo di assestamento del DAC
- tempo di risposta del comparatore  
(in particolare quando  $V_x$  e  $V_R$  differiscono di poco)
- tempo di risposta della SAR

questo resti costante entro  $\pm\frac{1}{2}$ LSB durante il tempo di conversione. Si tratta di un requisito molto stringente, che si traduce in una frequenza massima di una eventuale sinusoide in ingresso che si volesse convertire.

Considerando la massima pendenza di una sinusoide che occupi tutta la dinamica del convertitore pari, come abbiamo già visto, a:

$$V_{FS} \cdot \pi \cdot f_{in}$$

E moltiplicandola per la durata della conversione pari a:

$$n/f_{ck}$$

si ottiene la massima variazione ammissibile del segnale durante la conversione, che deve risultare inferiore a  $\pm\frac{1}{2}$ LSB. Impostando la diseguaglianza si ottiene:

$$f_{in} \leq f_{ck} \cdot n \cdot (2\pi)^2$$

Questa limitazione è molto stringente e se il segnale di ingresso non rispetta questa condizione durante la conversione, l'uscita digitale non ha alcuna relazione con il segnale di ingresso.

Ad esempio, per un ADC da 10 bit e un clock da 1 MHZ, viene accettata una frequenza di ingresso  $f_{in}$  massima di 16 Hz per un segnale sinusoidale con ampiezza picco-picco pari a  $V_{FS}$  (corrispondente, per una  $V_{FS}$  di 5 volt, a un tasso massimo di variazione del segnale di ingresso di 0.25 volt al millisecondo).

Per evitare questa limitazione, a

E' fondamentale che il segnale di ingresso resti costante entro  $\pm\frac{1}{2}$ LSB durante il tempo di conversione

Massima velocità di variazione di un ingresso sinusoidale :

$$\left. \frac{d}{dt} \left( \frac{V_{FS}}{2} \sin(2\pi f_{in} \cdot t) \right) \right|_{t=0} = V_{FS} \cdot \pi \cdot f_{in}$$

Durata della conversione :

$$T_{Conv} = \frac{n}{f_{ck}}$$

$\rightarrow [V_{FS} \cdot \pi \cdot f_{in}] \cdot \frac{n}{f_{ck}} < .5 \cdot \frac{V_{FS}}{2^n}$  da cui

$$f_{in} \leq .5 \cdot \frac{f_{ck}}{\pi \cdot n \cdot 2^n}$$

Esempio : ADC a 10 bit e  $f_{ck}=1$  MHz  $\rightarrow f_{in} < 16$  Hz

per un segnale sinusoidale con ampiezza picco-picco pari a  $V_{FS}$

monte del convertitore si inserisce un blocco analogico di campionamento e memorizzazione (Sample and Hold) che ha appunto il compito di mantenere  $V_x$  costante entro  $\pm\frac{1}{2}$  LSB durante tutta la fase di conversione. Questi circuiti, di utilità generale, verranno presentati in una prossima lezione.

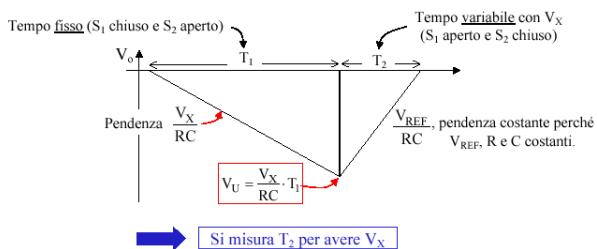
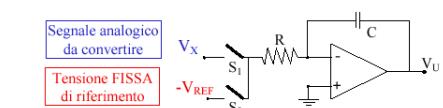
## Lezione n. 45 : Convertitori A/D ad elevate prestazioni

In questa lezione vedremo alcuni convertitori che, per le loro caratteristiche, riescono a soddisfare esigenze particolari. Queste sono per esempio quella di fornire il dato convertito con una elevatissima risoluzione (di 20 bit o anche di più), oppure di fornirlo in un tempo brevissimo in modo da raggiungere frequenze di campionamento di parecchie decine di Msample/s, oppure di adattarsi a segnali che variano da pochi mV a parecchi Volt, cioè tali da avere una elevata dinamica di ingresso. Diciamo subito che è difficilissimo, se non impossibile, avere tutte queste esigenze soddisfatte contemporaneamente da un solo convertitore. L'esigenza di prestazioni particolari infatti fa sì che questi abbiano architetture interne specializzate ed ottimizzate nel raggiungimento di un solo obiettivo a scapito degli altri. Da ultimo vedremo brevemente come adattare le configurazioni viste al caso di segnali di ingresso sia positivi che negativi.

Le misure che richiedono una elevata risoluzione (una parte su un milione o più) impiegano un

- Elevatissima risoluzione (20 bit e oltre)
- Elevatissima velocità (10 Msample/s e oltre)
- Elevata dinamica del segnale di ingresso (3 decadi o più)
- Segnali bipolar (positivi e negativi)

Difficilissimo, se non impossibile, avere tutte queste esigenze soddisfatte contemporaneamente !



particolare tipo di convertitore, chiamato "convertitore a doppia rampa" che può arrivare a generare parole di 20 bit e oltre, anche se paga queste elevate prestazioni in risoluzione con tempi di conversione molto lunghi, anche di decine di millisecondi.

Il suo principio di funzionamento può essere capito con riferimento al circuito integratore ad operazionale mostrato nella figura, al cui ingresso viene applicato, in base alla posizione dei due interruttori S<sub>1</sub> ed S<sub>2</sub>, o il segnale analogico V<sub>x</sub> o una

tensione di riferimento fissa V<sub>ref</sub>. Il ciclo di conversione è costituito da due integrazioni successive: dapprima viene integrato il segnale incognito V<sub>x</sub> per un intervallo di **tempo fisso T<sub>1</sub>**. La figura mostra

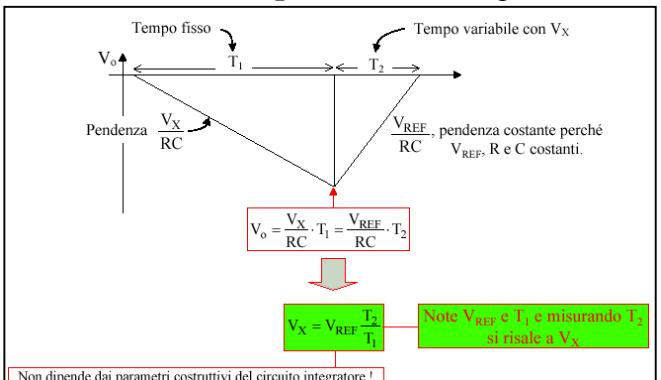
l'uscita dell'integratore nel caso in cui il segnale incognito V<sub>x</sub> sia una tensione costante positiva: durante T<sub>1</sub> la sua integrazione produce una rampa con pendenza V<sub>x</sub>/RC. Alla fine della prima fase di integrazione l'uscita dell'integratore raggiunge il livello V<sub>U</sub>=(V<sub>x</sub>/RC)T<sub>1</sub>. Questo livello è quindi funzione del valore V<sub>x</sub> dell'ingresso da convertire. Poi l'integratore commuta il suo ingresso sul segnale di riferimento fisso V<sub>REF</sub> di polarità opposta a V<sub>x</sub>. Da questo momento in poi viene integrata la corrente prodotta dalla tensione costante V<sub>REF</sub> che produce una rampa con pendenza V<sub>REF</sub>/RC che fa tornare a zero l'uscita dell'integratore. Questa seconda fase di integrazione dura un **tempo T<sub>2</sub> variabile** perché dipende dal valore di partenza V<sub>U</sub> dell'uscita dell'integratore. Misurando questo tempo T<sub>2</sub>, si risale quindi al valore di V<sub>x</sub>. Se infatti V<sub>x</sub> fosse stato diverso, ad esempio una tensione minore, proporzionalmente T<sub>2</sub> sarebbe stato minore. Dalla semplice proporzione tra i triangoli si ricava che

$$(V_x/RC)T_1 = (V_{REF}/RC)T_2$$

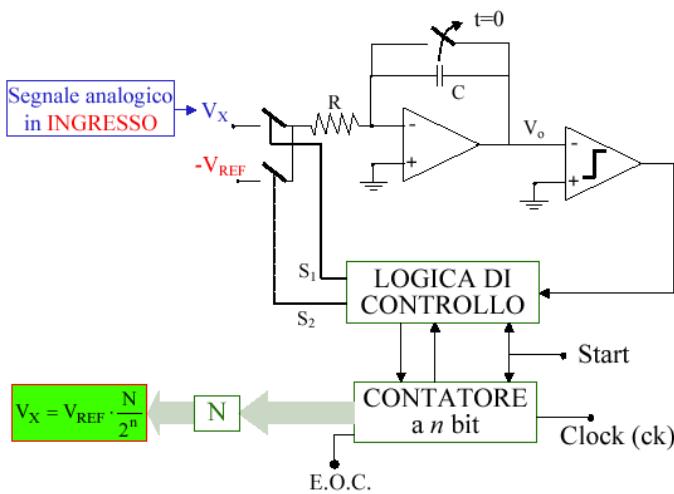
Da questa relazione si vede che si può eliminare la costante RC dell'integratore: succedendosi infatti le due integrazioni l'una immediatamente dopo l'altra, è ragionevole pensare che i valori R e C non cambino durante il ciclo completo di due integrazioni successive. Questo fatto assicura una grande stabilità al convertitore perché lo rende indipendente dai suoi componenti costruttivi. La relazione diventa allora

$$V_x = V_{REF} \cdot \frac{T_2}{T_1}$$

Misurando i tempi T<sub>2</sub> e T<sub>1</sub>, e supponendo nota V<sub>REF</sub>, si risale direttamente a V<sub>x</sub>.



Rivediamo il funzionamento di questo ADC analizzando più in dettaglio il suo schema a blocchi.  
All'inizio della conversione il contatore viene azzerato e l'uscita dell'integratore viene portata a un livello leggermente positivo con una connessione non mostrata nella figura. Si chiude l'interruttore  $S_1$  e  $V_x$  viene applicato all'ingresso dell'integratore. L'integrazione di  $V_x$  dura per un tempo  $T_1$  che inizia dal momento in cui l'uscita dell'integratore, passando per lo zero, fa scattare il comparatore.



L'uscita del comparatore attiva la logica di controllo che fa partire il conteggio del contatore. La fine di  $T_1$  coincide con l'accumulo nel contatore del conteggio massimo  $2^n$ , condizione che viene rilevata dalla logica di controllo che provvede ad aprire  $S_1$  e a chiudere  $S_2$ , connettendo così  $V_{REF}$  all'ingresso dell'integratore. Ha inizio la fase  $T_2$ , durante la quale, mentre il contatore

ricomincia a contare, l'uscita dell'integratore inverte la pendenza e diminuisce fino a che non attraversa di nuovo lo zero. Il comparatore si accorge di ciò cambiando nuovamente stato e segnalando la fine dell'intervallo  $T_2$ . La logica di controllo sente la commutazione del comparatore e provvede a bloccare il contatore, a rilevarne il conteggio  $N$  accumulato e a segnalare la fine del ciclo di conversione. Il conteggio  $N$  accumulato dal contatore costituisce una misura di  $T_2$  e viene usata per la conversione numerica di  $V_x$  secondo la relazione  $V_x = V_{REF}(N/2^n)$ .

L'elevata risoluzione raggiungibile con questi convertitori deriva dalla elevata risoluzione con cui si può contare il tempo  $T_2$  con  $N$  impulsi di clock .

Il tempo di conversione complessivo  $T_c$ , somma di  $T_1$  e di  $T_2$ , varia in funzione del segnale di ingresso; il tempo di conversione massimo  $T_{max}$  si ha quando  $N$  raggiunge il valore massimo  $N_{max}=2^n$ ; dato che per accumulare il conteggio massimo con un clock di frequenza  $f_{ck}$  occorre un tempo  $=2^n/f_{ck}$ , avremo che  $T_{max}=2 \cdot 2^n/f_{ck}$ . Un convertitore a doppia rampa a 10 bit che utilizzi un clock di 1 MHz (periodo 1 microsecondo), avrà un  $T_{max} = 2048$  microsecondi, cioè di circa 2 millisecondi. La sua velocità di conversione massima sarà poco meno di 500 conversioni al secondo, quindi molto minore di quella di un ADC ad approssimazioni successive con ugual numero di bit.

Il principio di conversione si basa in definitiva sull'uguaglianza delle integrazioni nei due intervalli di tempo  $T_1$  e  $T_2$ . Il segnale costante  $V_x$  produce una corrente che viene integrata per un tempo fisso  $T_1$ , cosicché l'integrale è proporzionale a  $V_x \cdot T_1$ , come già visto;  $V_{REF}$  viene integrato per un tempo  $T_2$  tale da far sì che il valore dell'integrale, pari a  $V_{REF} \cdot T_2$  uguagli  $V_x \cdot T_1$ .

In base a queste considerazioni non è necessario che  $V_x$  resti costante durante  $T_1$  perché la conversione abbia valore.

Se  $V_x$  varia, come mostrato nelle curve in rosso, la parola digitale che si ottiene alla fine del ciclo di conversione ha sempre senso e rappresenta il **valor medio** di  $V_x$  nell'intervallo  $T_1$ .

Infatti, per definizione di valor medio, vale la relazione

$$\frac{1}{RC} \int_0^{T_1} V_x(t) dt = \frac{1}{RC} \bar{V}_x \cdot T_1$$

Avremo quindi

$$V_{x\text{medio}} = V_{REF}(N/2^n)$$

Il tempo di conversione,  $T_c=T_1+T_2$ , varia proporzionalmente a  $V_x$ :

$$T_{c\min} \text{ quando } V_x = 0V \rightarrow T_{c\min} = 0 \text{ s}$$

$$T_{c\max} \text{ quando } N=2^n, \text{ cioè } T_1=T_2 \rightarrow T_{c\max} = 2 \cdot \frac{2^n}{f_{ck}}$$

Velocità di conversione relativamente bassa

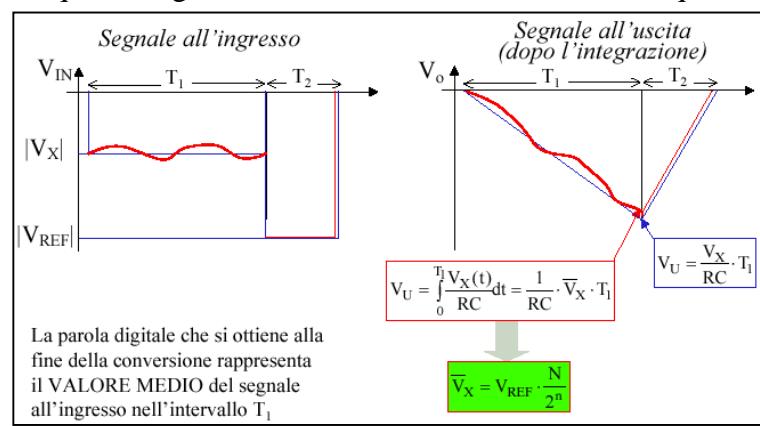
Esempio: DAC a 10 bit e  $f_{ck}=1 \text{ MHz}$  (periodo  $1 \mu\text{s}$ )

$$\rightarrow T_{c\max} \approx 2 \text{ ms} \rightarrow \approx 500 \text{ conversioni/s}$$

Il fatto che la parola digitale N non perde quindi significato anche se  $V_x$  varia, rende questo convertitore molto interessante ed usatissimo e gli fa assumere eccellenti caratteristiche di linearità differenziale e integrale.

L'avere l'uscita digitale che rappresenta il valore medio del segnale di ingresso durante l'intervallo di tempo  $T_1$  può essere utilmente sfruttato per molti scopi. Uno è quello di eliminare dalla misura l'effetto di disturbi sinusoidali sovrapposti al segnale incognito.

Se facciamo in modo che  $T_1$  sia uguale al periodo del disturbo sinusoidale, la prima fase di integrazione eliminerà dalla misura l'effetto di questo disturbo, dato che il suo integrale in un periodo è nullo. Verranno analogamente eliminati gli effetti di disturbi con frequenza multipla intera di quello mostrato. L'integratore si comporta quindi



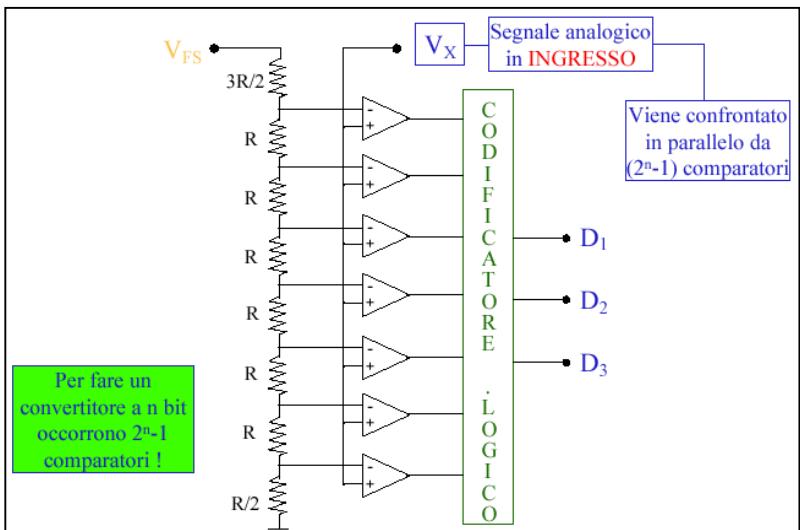
come un filtro passa basso che elimina dalla sua uscita segnali sinusoidali con frequenza  $=1/T_1$  e suoi multipli. Questa proprietà viene utilmente impiegata in molti voltmetri digitali di qualità, che montano un convertitore a doppia rampa in cui  $T_1$  viene posto uguale a 20 millisecondi, cioè il periodo della frequenza di rete (50 Hz): i disturbi alla frequenza di rete (ed i suoi multipli), assai frequenti come sorgenti di rumore sovrapposto al segnale utile, vengono così eliminate dalla misura, rendendo quest'ultima estremamente accurata.

Esempio : nei voltmetri digitali di precisione  $T_1=20\text{ms}$ , così da togliere i disturbi di rete (50Hz e multipli).

Quando si ha la necessità di campionare e convertire un segnale con elevata frequenza bisogna usare convertitori speciali che impiegano una metodologia di conversione di tipo parallelo invece che seriale come visto fino ad ora. I più veloci che si possano realizzare sono i cosiddetti "Convertitori flash", il cui nome rende bene l'idea della loro intrinseca velocità di conversione.

La struttura di un convertitore di questo tipo è mostrata nella figura, limitata per semplicità a tre bit. Il segnale incognito  $V_x$  viene contemporaneamente confrontato con sette diverse tensioni di riferimento, ottenute ripartendo la tensione di fondo scala con un partitore resistivo di precisione. I valori delle resistenze sono scelti in modo da riprodurre la caratteristica statica ingresso–uscita di un ADC ideale.

A seconda del valore di  $V_x$  tutti i comparatori il cui riferimento dato dal partitore è più basso di  $V_x$ , hanno l'uscita alta, mentre tutti gli altri sopra rimangono con l'uscita bassa. Le uscite dei comparatori, che rappresentano  $V_x$  in una scala detta termometrica, vengono interpretate dal "codificatore logico" per fornire la parola digitale di tre bit codificata in binario.



Si noti che per un convertitore a tre bit servono sette comparatori; in generale, per un convertitore a  $n$  bit occorrono  $2^n-1$  comparatori. Il costo di queste strutture cresce quindi esponenzialmente al crescere del numero dei bit e il loro impiego viene normalmente limitato a convertitori a 8-10 bit.

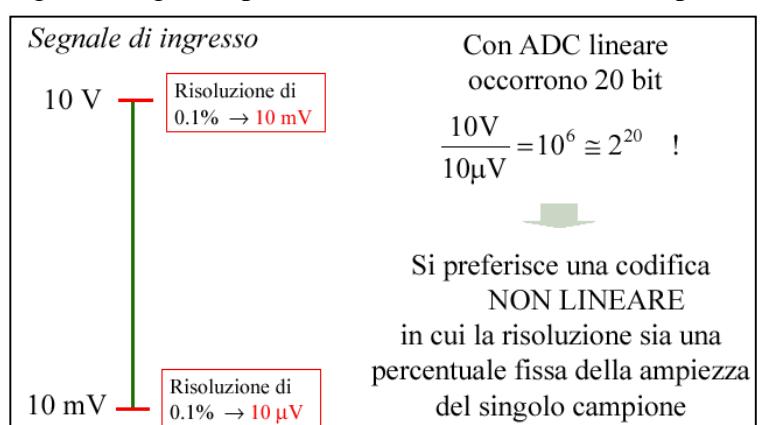
Nei convertitori flash, quindi, la velocità di conversione è molto alta; essa viene limitata infatti solo dai **tempi di ritardo dei comparatori e della rete logica**. Con questi convertitori si possono raggiungere tassi di conversione da 10 a 100 milioni di conversioni al secondo.

A parte i problemi di **costo**, il numero dei bit viene anche limitato dalla difficoltà a realizzare un partitore preciso con  $2^n$  resistenze identiche.

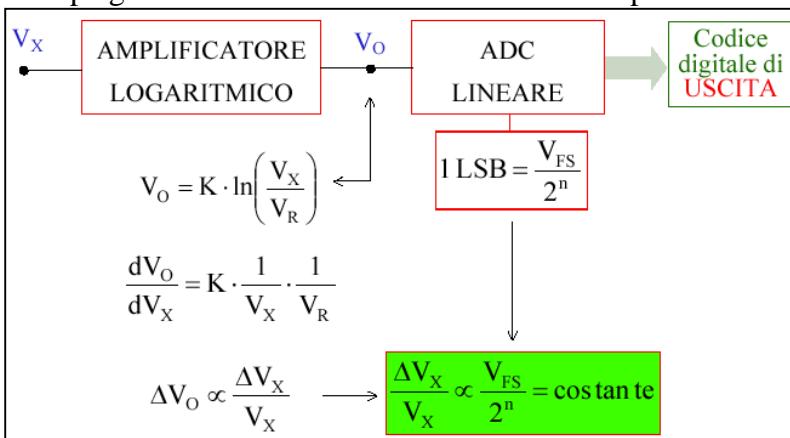
Inoltre occorre tener presente anche che i comparatori non sono dispositivi ideali, ma hanno un loro offset, dell'ordine tipicamente dei millivolt, per cui la soglia impostata dal partitore viene spostata dall'offset del comparatore e, se la risoluzione è troppo spinta, le soglie possono accavallarsi e degradare la linearità differenziale del convertitore.

Numerosi sono i casi in cui la grandezza analogica di ingresso può variare entro intervalli che coprono

alcune decadi, ad esempio dai millivolt ai volt e si vogliono risoluzioni elevate (ad esempio di una parte su mille) anche per la misura dei bassi livelli. Poniamo per l'appunto di dover misurare un segnale  $V_i$  compreso tra 10 mV e 10 V, volendo che la risoluzione sia almeno dell'un per mille su tutta la gamma. Rispetto alla tensione 10 mV dobbiamo essere in grado di leggere i 10  $\mu$ V. Se tenessimo questa risoluzione su tutta la gamma occorrerebbe un ADC lineare da 20 bit



perché rispetto alla tensione di fondo scala di 10 volt questa risoluzione rappresenta una parte su un milione. Questa risoluzione risulterebbe ampiamente sovrabbondante per i 10 V: infatti, se volessimo mantenere la stessa risoluzione dell'un per mille anche per i 10 V, basterebbe poter leggere i 10 mV. L'impiego di un convertitore lineare da 20 bit per risolvere questo problema di misura risulterebbe



quindi inutilmente dispendioso. Si preferisce optare per una codifica **NON LINEARE**, in cui la risoluzione non sia una percentuale fissa del fondo scala (10V nel nostro esempio) ma una percentuale fissa dell'ampiezza del singolo campione. Per fare questo, un modo semplice è quello di comprimere logaritmicamente il segnale e poi convertirlo linearmente.

Il segnale  $V_x$  quindi passa attraverso

un amplificatore logaritmico la cui uscita soddisfa la relazione  $V_o = K \cdot \ln(V_x/V_r)$  dove  $K$  e  $V_r$  sono parametri regolabili dell'amplificatore. Nota la dinamica utile di  $V_x$ , si scelgono  $K$  e  $V_r$  in modo da adattare  $V_o$  alla dinamica da 0 al fondo scala  $V_{FS}$  dell'ADC, tenendo presente che  $V_o=0$  quando  $V_x=V_r$ .

Differenziando rispetto a  $V_x$  l'espressione di  $V_o$  ed esprimendo il risultato in termini di incremento finiti possiamo scrivere

$$\Delta V_o \approx K \cdot \Delta V_x / V_x$$

Poiché 1 LSB di un ADC lineare a  $n$  bit corrisponde a una variazione di tensione di  $V_{FS}/2^n$  al suo ingresso, otteniamo che

$$\Delta V_x / V_x \propto V_{FS} / 2^n$$

ed è costante. Come avevamo desiderato, l'incremento  $\Delta V_x$  della tensione di ingresso che corrisponde a 1 LSB dell'ADC lineare cresce al crescere di  $V_x$ . O detto in modo diverso, 1 LSB corrisponde a una percentuale fissa di  $V_x$ .

Per tornare al nostro problema di misura, ci basta un ADC a **10 bit**, preceduto da un amplificatore logaritmico, per avere misure di  $V_X$  con una risoluzione di una parte su mille su tutto l'intervallo di variazione utile di  $V_X$ .

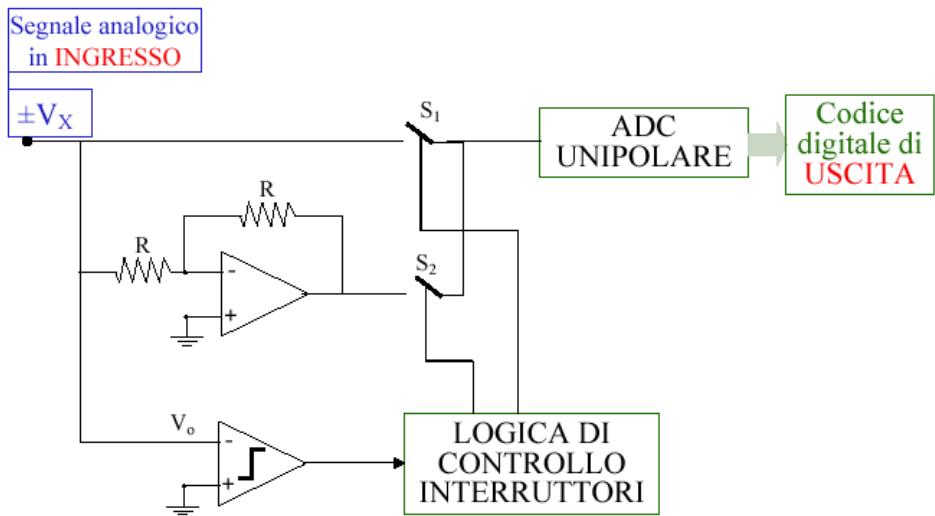
I convertitori discussi precedentemente ricevevano tutti in ingresso segnali unipolari che potevano variare tra 0 e  $+V_{FS}$  oppure 0 e  $-V_{FS}$ . Naturalmente si può presentare il caso che il segnale sia bipolare, cioè sia positivo e negativo.

Diversi sono i modi in cui si può ottenere un comportamento bipolare dal convertitore. Il più semplice è quello di sdoppiare il segnale di ingresso, inserendo su un ramo un amplificatore invertente a guadagno unitario. Un

comparatore rileva la polarità del segnale incognito e attiva un blocco logico che chiude l'interruttore del ramo su cui è presente la polarità giusta per entrare in un ADC unipolare a valle.

Gli ADC a rampa o ad approssimazioni successive possono operare in modalità bipolare pur di impiegare DAC bipolarì nel circuito. Da ultimo, un

convertitore Flash funziona in modo bipolare quando si collega l'estremo inferiore del partitore di resistenze a  $-V_{FS}$  invece che a massa.

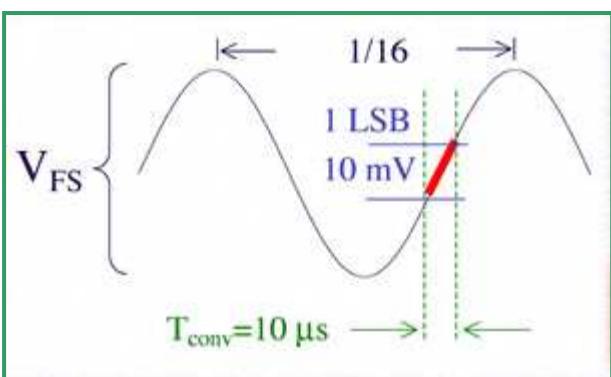
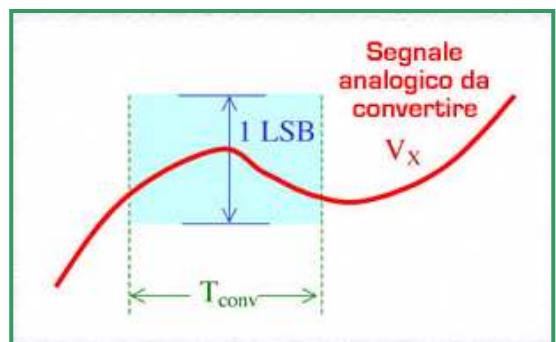


## Lezione n. 46 : Circuiti di Sample & Hold

### Struttura circuitale di un Sample & Hold

Più volte parlando degli ADC nelle lezioni precedenti abbiamo sottolineato la necessità che il segnale analogico da convertire non modifichi il suo valore di più di 1 LSB durante il tempo di conversione, in modo che il codice prodotto in uscita rappresenti al meglio il valore di ingresso.

Questa condizione, se applicata direttamente al segnale di ingresso, ne limiterebbe la frequenza ad un valore bassissimo. Riguardando il caso del convertitore ad approssimazioni successive a 10 bit e frequenza di clock di 1 MHz, si era visto che per poter convertire una sinusoide ampia come tutta la dinamica dell'ADC (10V),



poteva essere convertita correttamente solo se la sua frequenza era al massimo di 16 Hz. Oltre questa frequenza, la massima variazione del segnale di ingresso durante l'intervallo di conversione di  $10\mu\text{s}$  risulterebbe essere maggiore di 10 mV, cioè maggiore di 1 LSB del campionatore che è pari appunto a  $V_{FS}/2^n = 10\text{mV}$ .

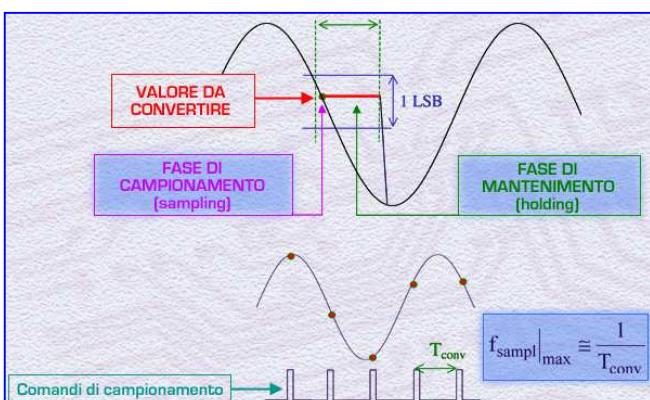
Questo fatto di poter trattare solo frequenze di segnale molto basse disturba, tanto più che il breve tempo di conversione di soli  $10\mu\text{s}$  dà all'ADC la potenzialità di convertire un valore di tensione con

una frequenza di campionamento di 100 kHz. Il base al teorema del campionamento, questo corrisponderebbe a poter trattare segnali con frequenze fino a 50 kHz, ben maggiori dei 16 Hz ! Per poter effettivamente convertire anche segnali a frequenza elevata bisogna che pur variando il segnale di ingresso, il valore da convertire sia trattenuto costante durante il tempo di conversione.

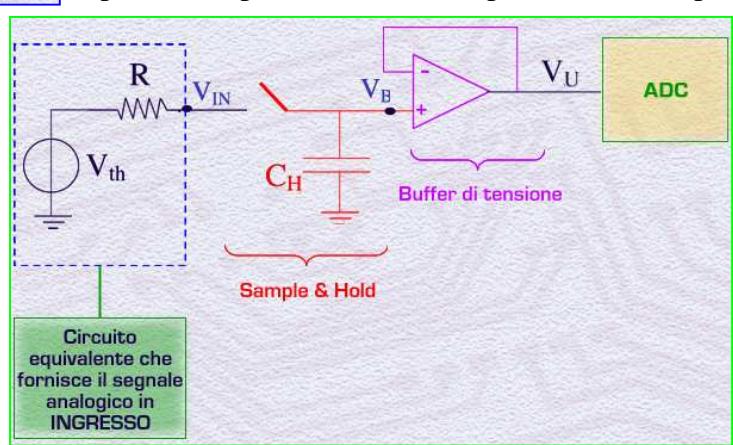
$$\text{Teorema di campionamento } f_{in} = \frac{f_{sampler}}{2}$$

di conversione. Il circuito elettronico che svolge tale compito è chiamato "*Circuito di campionamento e mantenimento*" o, ormai comunemente, "*Circuito di Sample&Hold*".

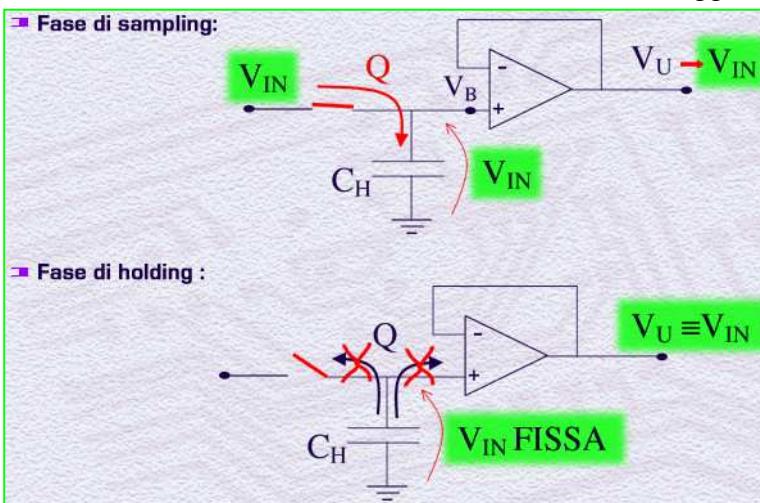
In sostanza esso è un circuito che, in corrispondenza di un segnale di comando, cattura il valore della tensione di ingresso (fase di campionamento o di sampling) e lo tiene memorizzato per tutto il tempo di conversione (fase di mantenimento o di holding). Queste due operazioni possono venire ripetute nel tempo



anche in continua successione e permettere così di campionare il segnale con una frequenza sostanzialmente pari al reciproco del tempo di conversione e di convertirlo con grande precisione. Il più semplice circuito che svolga queste funzioni è costituito da un interruttore ed un condensatore. A sinistra dell'interruttore è connesso il circuito elettronico che fornisce il segnale analogico che si vuole campionare ed a destra del condensatore è posto un buffer di tensione per pilotare l'ADC.



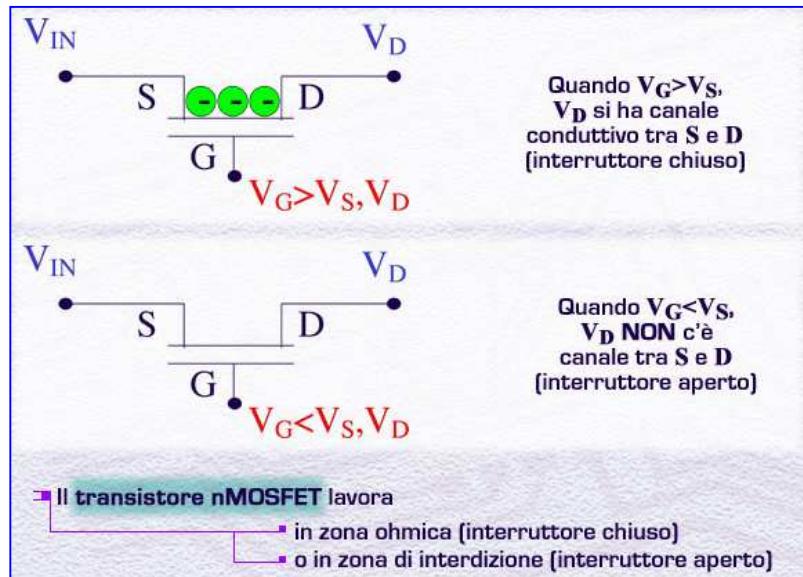
Nella fase di *sampling* l'interruttore viene chiuso, cosicché il condensatore  $C_H$  viene caricato al valore della tensione  $V_{IN}$ . Tramite il buffer anche  $V_U$  si aggiorna al valore  $V_{IN}$ . Quando ciò è avvenuto, l'interruttore si apre iniziando la fase di "holding" durante la quale il condensatore  $C_H$ , rimasto isolato, trattiene la carica prima immagazzinata e mantiene costante il valore  $V_{IN}$  ai suoi capi e lo rende disponibile tramite il buffer al circuito dell'ADC per la conversione. Il buffer, solitamente realizzato ad operazionali, ha proprio la funzione di fornire in uscita la tensione con tutta la corrente necessaria senza per questo assorbire carica all'ingresso da  $C_H$ . Si noti che durante questa fase il segnale analogico può variare senza che



la tensione ai capi del condensatore ne risenta. Nei circuiti di S&H reali l'interruttore viene realizzato usualmente con un MOSFET opportunamente pilotato con una tensione al Gate. Quando al Gate viene applicata una tensione positiva maggiore della tensione presente al Source ed al Drain, nel canale sono accumulati elettroni così da formare un percorso conduttivo a bassa resistenza tra Source e Drain. Questo rende il MOSFET equivalente ad un interruttore chiuso.

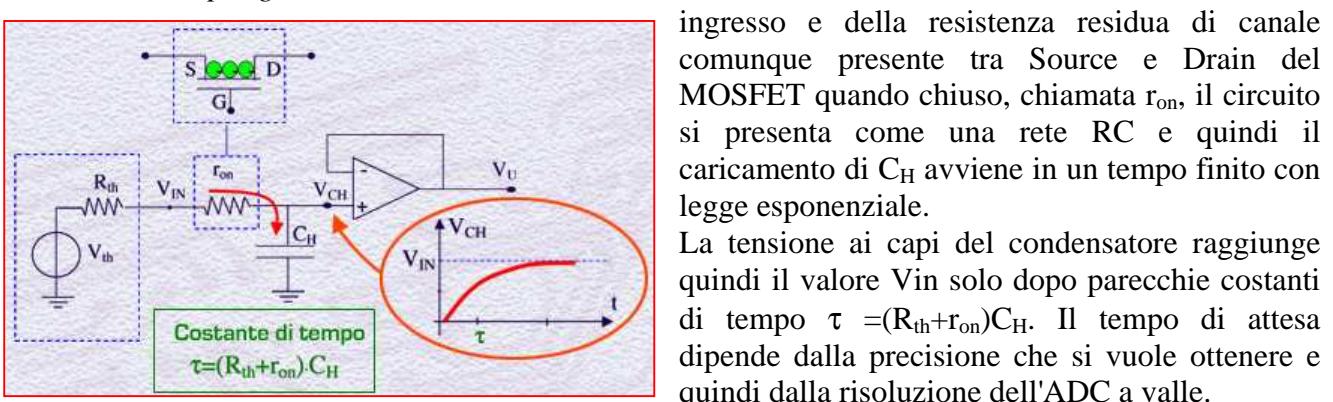
Quando invece al Gate viene applicata una tensione minore della tensione presente al Source ed al Drain, il canale viene completamente svuotato di portatori con la conseguenza di avere un isolamento elettrico tra Source e Drain. Questo rende il MOSFET equivalente ad un interruttore aperto.

Poiché la tensione di Source ha la stessa ampia escursione del segnale di ingresso, sostanzialmente pari alla dinamica dell'ADC, anche il segnale di comando del Gate deve avere una ampia escursione, spesso maggiore di 10V, per essergli sempre sopra quando si vuole chiudere l'interruttore (facendo lavorare il MOSFET in zona ohmica) o sempre sotto quando si vuole aprirlo (facendo lavorare il MOSFET in interdizione).



## Prestazioni di un Sample & Hold

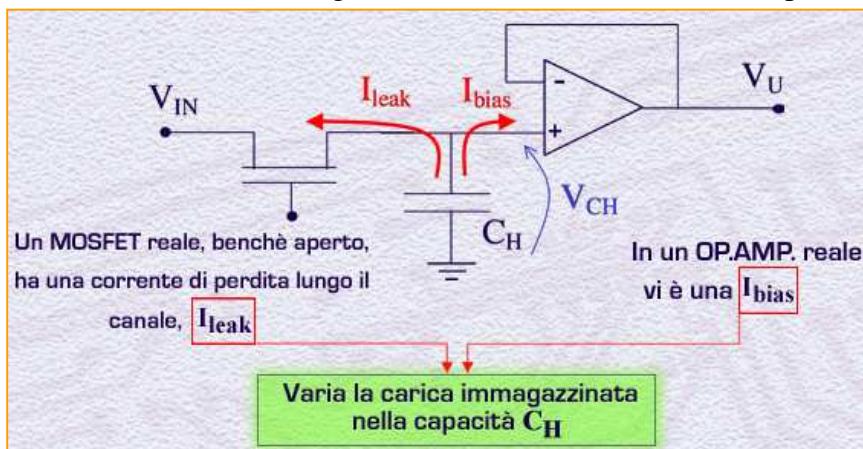
La fase di *sampling* non è in realtà istantanea. A causa della resistenza Thevenin del circuito di ingresso e della resistenza residua di canale comunque presente tra Source e Drain del MOSFET quando chiuso, chiamata  $r_{on}$ , il circuito si presenta come una rete RC e quindi il caricamento di  $C_H$  avviene in un tempo finito con legge esponenziale.



La tensione ai capi del condensatore raggiunge quindi il valore  $V_{IN}$  solo dopo parecchie costanti di tempo  $\tau = (R_{th} + r_{on})C_H$ . Il tempo di attesa dipende dalla precisione che si vuole ottenere e quindi dalla risoluzione dell'ADC a valle.

Ad esempio, pur nell'ipotesi che il circuito che fornisce il segnale sia ideale ( $R_{th}=0 \Omega$ ), un MOSFET con  $r_{on}=100 \Omega$  ed una  $C_H=100 \text{ pF}$  comportano una  $\tau = 10 \text{ ns}$ . Se l'ADC ha 10 bit ed un fondo scala di 10V, ed il segnale  $V_{in}$  deve passare da 0V a 10V (è certamente la situazione più penalizzante), la tensione  $V_{CH}$  si assesta entro 1/1000 di 10V dopo circa 7 costanti di tempo e quindi la fase di campionamento dura circa 100 ns. Per ridurre questo tempo, che comunque normalmente è trascurabile rispetto al tempo di conversione dell'ADC a meno di utilizzare convertitori flash, bisognerebbe avere  $r_{on}$  più piccola (ma è difficile andare sotto 10  $\Omega$ ) oppure  **$C_H$  più piccola, ma questo è in contrasto con le esigenze in fase di holding**, o meglio ancora bisognerebbe usare un circuito S&H retroazionato, come vedremo tra poco.

Durante la fase di holding, si vorrebbe che la tensione ai capi del condensatore rimanesse costante al

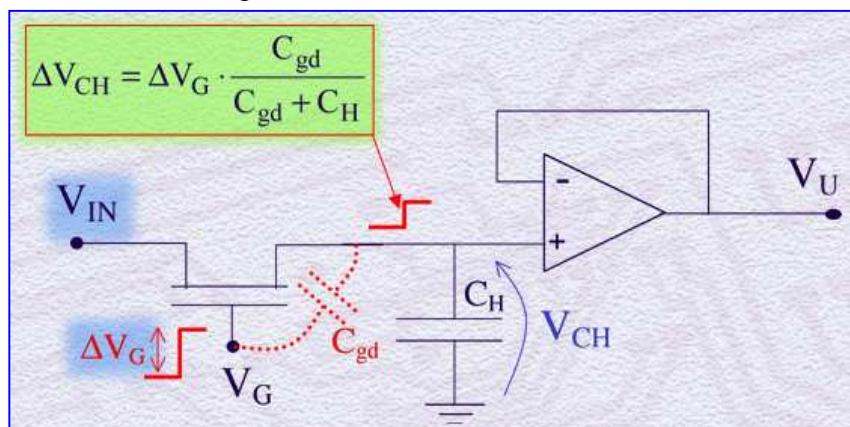


valore campionario, cioè che la carica immagazzinata in  $C_H$  non cambi nel tempo. In un S&H reale la capacità seppur lentamente si scarica. Infatti, l'interruttore MOSFET reale aperto, presenta un residuo di canale attraverso cui scorre una ancorché piccola corrente, indicata con  $I_{leak}$ . Inoltre, l'amplificatore operazionale del buffer ha una corrente non nulla di polarizzazione al suo morsetto

di ingresso indicata con  $I_{bias}$ . La presenza di queste correnti fa variare nel tempo la carica accumulata in  $C_H$ , tendendo normalmente a scaricarla.

Ad esempio, con  $C_H=100\text{pF}$  e  $I_{leak}=100\text{pA}$  e  $I_{bias}=1\text{nA}$ , dopo un tempo di  $10\mu\text{s}$  (che è il tempo di conversione di un ADC ad approssimazioni successive) la carica su  $C_H$  è variata di  $2 \cdot 10^{-14} \text{ C}$  e conseguentemente la tensione  $V_{CH}$  è scesa di  $200\mu\text{V}$ . Questa variazione è trascurabile fintantoché rimane inferiore a 1 LSB dell'ADC. E' questo il caso se si utilizzano ADC con meno di 16 bit. Degli ADC a maggior risoluzione invece richiederebbero o di scegliere  $C_H$  più grandi o di utilizzare un operazionale con una  $I_{bias}$  più piccola.

Purtroppo la struttura reale del MOSFET introduce degli errori nell'accuratezza del trasferimento del segnale  $V_{in}$  all'ADC. Il più significativo è originato dalla presenza della capacità  $C_{gd}$  del MOSFET. Quando si applica il segnale al Gate, ad esempio positivo, per chiudere l'interruttore e campionare il segnale, la capacità  $C_{gd}$  e la capacità  $C_H$  si trovano su una maglia con la tensione di Gate e quindi tramite  $C_{gd}$  la tensione ai capi di  $C_H$  si sposta della quantità indicata in figura.



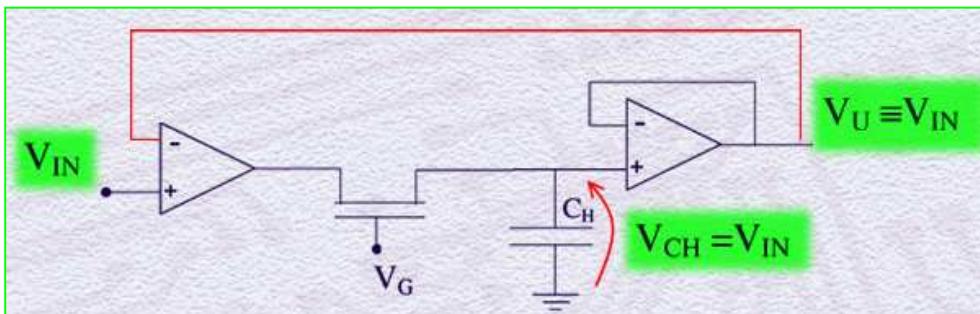
Questo spostamento di tensione può causare un errore nella lettura del valore da parte dell'ADC se l'effetto non viene limitato sufficientemente. A questo scostamento spurio corrisponde infatti una iniezione di carica nella capacità  $C_H$  che va ad aggiungersi o sottrarsi a quella che lì dovrà immagazzinarsi in base al valore  $V_{in}$  da campionare.

Per limitare l'entità dell'errore introdotto bisogna scegliere MOSFET con  $C_{gd}$  piccolissima facendo piccoli gli interruttori MOSFET e  $C_H$  grandi, quest'ultima scelta penalizzante per quanto riguarda il tempo di assestamento del segnale di ingresso durante la fase di sampling, come visto prima.

Ad esempio, se  $C_{gd}=100\text{fF}$  ed il segnale al Gate è di 10V e si vuole mantenere l'errore entro 1/10.000 del fondo scala dell'ADC che supponiamo essere 10V, ne risulterà che  $C_H$  deve essere di 1nF.

Se lo scostamento finale  $\Delta V_{CH}$  prodotto dopo che si è esaurito il ciclo di apertura e chiusura del MOSFET fosse costante al variare di  $V_{IN}$  si potrebbe pensare di eliminarlo via software dopo l'acquisizione. In realtà, la variazione  $\Delta V_{CH}$  introdotta dal Gate si accoppia a quella imposta da  $V_{IN}$  in modo complesso per cui l'errore non sarà costante ma dipenderà dalla tensione di ingresso e quindi determina una non linearità nella funzione di trasferimento del S&H, e per questo va assolutamente minimizzato.

Per migliorare sia l'accuratezza che la velocità della fase di sampling, si usano circuiti di S&H retroazionati. Un esempio tra i più semplici ma molto usato, è quello rappresentato schematicamente nella figura. Esso è costituito dalla struttura di andata che ben conosciamo a cui è stato aggiunto un operazionale all'ingresso su cui si chiude la retroazione proveniente dall'uscita. Nell'ipotesi che i due Op.Amp. siano ideali, il circuito assicura **durante la fase di campionamento**, cioè di interruttore chiuso, che la tensione ai capi di  $C_H$  (e quindi all'uscita) sia uguale a  $V_{IN}$  con **grandissima precisione**.



Se infatti così non fosse, ad esempio perché  $V_{CH}$  ancora non ha raggiunto il valore di  $V_{IN}$ , anche l'uscita sarebbe inferiore a  $V_{IN}$  come pure il valore al morsetto invertente dell'operazionale di

ingresso. L'uscita dell'operazionale tenderebbe quindi ad essere grande e positiva, tendendo così a caricare  $C_H$  e ad alzarne la sua tensione. Solo quando  $V_{CH}=V_{IN}$ , il circuito raggiunge una situazione di equilibrio. Nel caso di guadagno dell'anello infinito, questa è proprio la situazione finale raggiunta, che assicurerrebbe una accuratezza elevatissima.

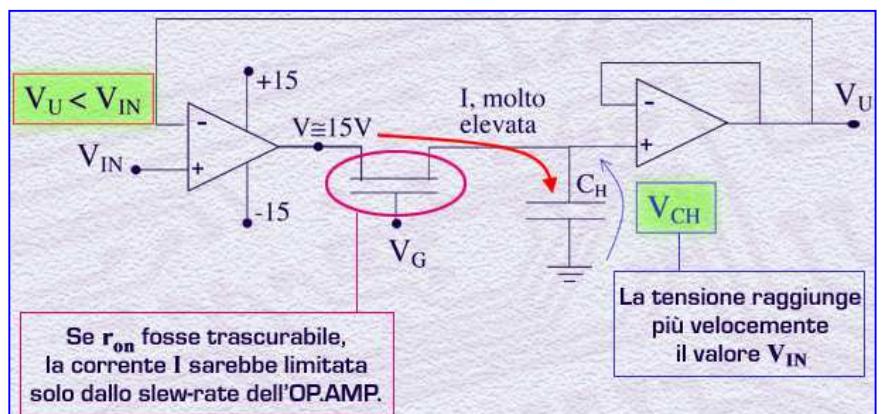
Nel caso di strutture reali, l'uguaglianza tra le due tensioni,  $V_{CH}$  e  $V_{IN}$ , non è perfetta ma tuttavia elevata e direttamente proporzionale al guadagno di anello del circuito, che quindi bisogna assicurarsi che sia sufficientemente elevato.

Si noti che la **retroazione non apporta invece nessuna miglioria durante la fase di holding**, perché l'anello è aperto e quindi rimane inattiva.

Un altro vantaggio della struttura retroazionata è la sua **velocità nella fase di sampling**. Infatti, come appena descritto, fintanto che  $V_{CH} < V_{IN}$  l'uscita dell'operazionale di ingresso è alta quasi quanto l'alimentazione positiva dell'operazionale così da caricare  $C_H$  con una corrente ben più elevata che senza retroazione perché ai capi di  $r_{on}$  cade una tensione elevata. Addirittura, se

$r_{on}$  fosse trascurabile, il tempo di acquisizione sarebbe brevissimo, limitato solo dallo slew-rate dell'Op.Amp. cioè dalla massima corrente da esso fornibile.

Grazie a queste proprietà di precisione e velocità, i circuiti S&H realmente utilizzati sono del tipo retroazionato.



## Esercizio 1

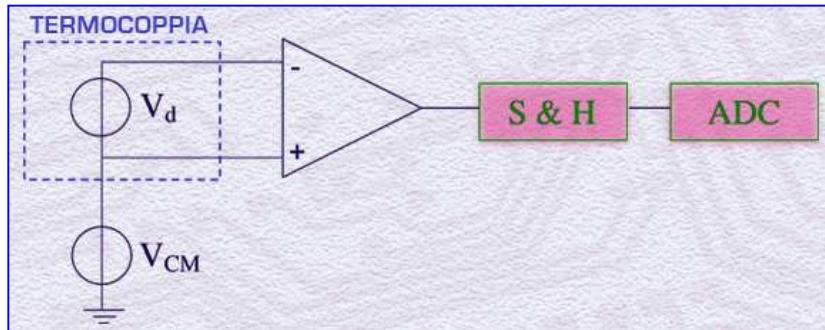
Supponendo che il segnale  $V_{IN}$  da convertire abbia una dinamica compresa tra 0V e +10V, e che l'interruttore MOSFET abbia una soglia  $V_T=0.8V$ , quale valore bisogna dare a  $V_G$  affinché il MOSFET operi da interruttore aperto ? E quale affinché operi da interruttore chiuso ?

## Lezione n. 47 : Progetto di un sistema di acquisizione dati

### Un sistema di misura della temperatura

A scopo anche di ripasso e di esercizio, proviamo in questa lezione a legare insieme alcune delle conoscenze acquisite durante il corso, scorrendo il progetto di un sistema completo di acquisizione e conversione digitale di un segnale elettrico prodotto da un sensore. Per semplicità, supponiamo di esserci dati l'obiettivo di voler realizzare un sistema per la misura della temperatura tramite il segnale fornito da una termocoppia, nell'intervallo tra 0 °C e 450 °C. Vogliamo condurre uno studio di fattibilità allo scopo di stabilire la possibilità di misurare la temperatura con una risoluzione di 0.5 °C. Una possibile struttura del sistema elettronico di misura è schematicamente indicata nella figura.

Essa comprende il generatore di segnale  $V_d$ , che rappresenta l'equivalente Thevenin ideale della termocoppia, il generatore  $V_{CM}$ , che rappresenta la tensione di modo comune sovrapposta al segnale, il blocco di amplificazione analogica del segnale, il blocco sample and hold (S&H) e il convertitore analogico digitale (ADC) per generare il valore numerico all'uscita.



Per coprire l'intervallo di temperatura tra 0°C e 450°C si può utilizzare una termocoppia ferro-costantana, che copre l'intervallo fornendo una tensione a vuoto  $V_d$  compresa tra 0 e 25 mV.

La sua sensibilità è quindi pari a  $55\mu\text{V}/^\circ\text{C}$ . Ne conseguono che 0.5 °C corrispondono a  $26.5 \mu\text{V}$ .

Questi sono i segnali più piccoli che dobbiamo elaborare e riconoscere per cui **gli errori di misura dovranno essere tenuti al di sotto di questo valore**.

Le condizioni di misura danno anche un valore massimo di 10 V per  $V_{CM}$ .

Analizziamo i requisiti che devono avere i blocchi analogici e il convertitore ADC per consentire la misura della temperatura

Benchè sia l'ultimo stadio del sistema, può essere utile fin d'ora individuare il numero dei bit del convertitore A/D in grado di soddisfare la risoluzione ipotizzata per la misura. Osserviamo che la risoluzione di 0.5 °C desiderata corrisponde alla lettura di 5 parti su 4500, cioè di una parte su 900.

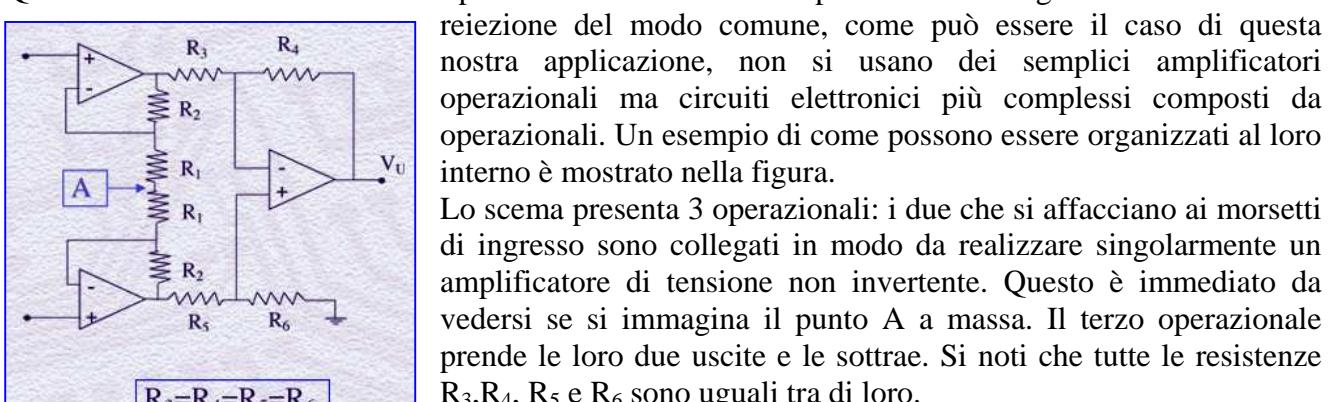
Il bit meno significativo del convertitore deve soddisfare questi requisiti, per cui dovremo avere che:  $1/2^n < 1/900$ . Ricaviamo così per  $n$  il valore di 10 ( $2^{10} = 1024$ ).

Questo valore medio di numero di bit necessari e l'assenza di una reale esigenza di velocità suggeriscono la scelta tra ADC ad approssimazioni successive oppure a doppia rampa. Terremo a mente, nel seguito della discussione, questa doppia alternativa, esaminando, all'occorrenza, vantaggi e svantaggi delle due possibili scelte.

Per fissare le idee assumiamo che il convertitore lavori con una **dinamica di ingresso di 5 V**.

### Amplificatore per strumentazione

Quando si richiedono elevatissime prestazioni in termini di amplificazione di segnali differenziali e di reiezione del modo comune, come può essere il caso di questa nostra applicazione, non si usano dei semplici amplificatori operazionali ma circuiti elettronici più complessi composti da operazionali. Un esempio di come possono essere organizzati al loro interno è mostrato nella figura.



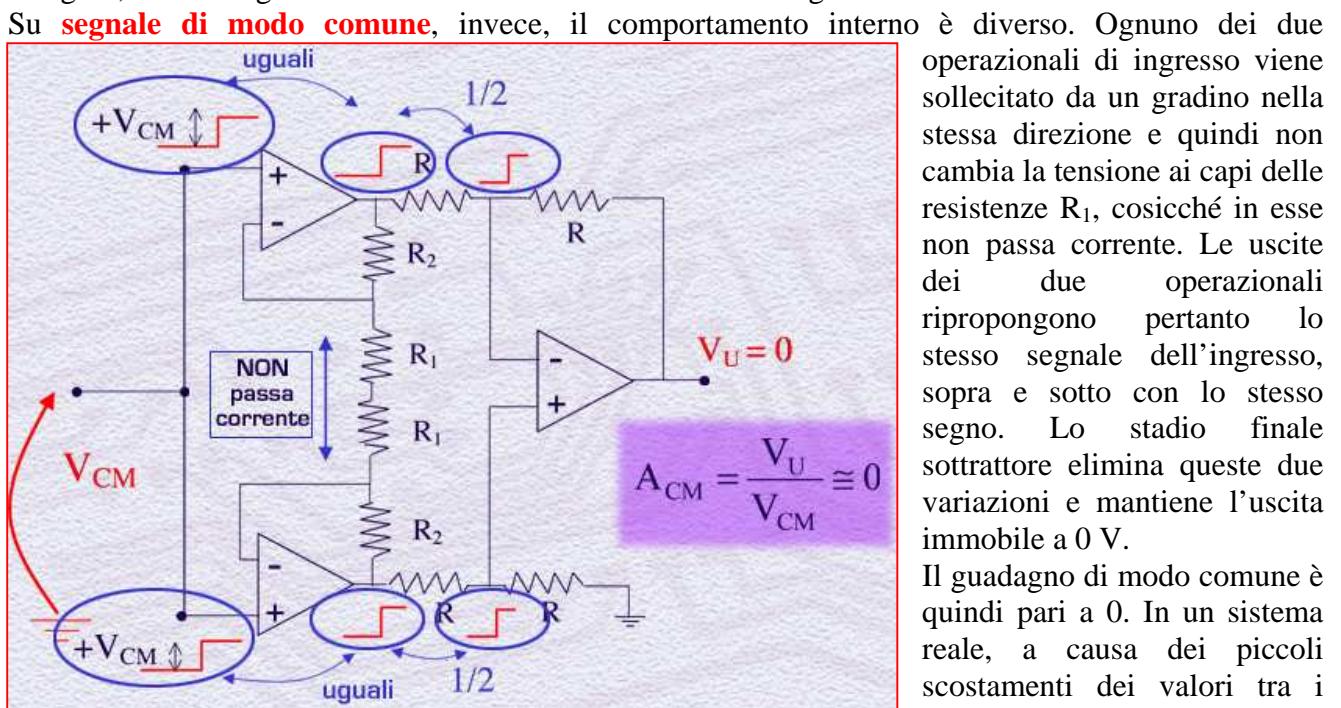
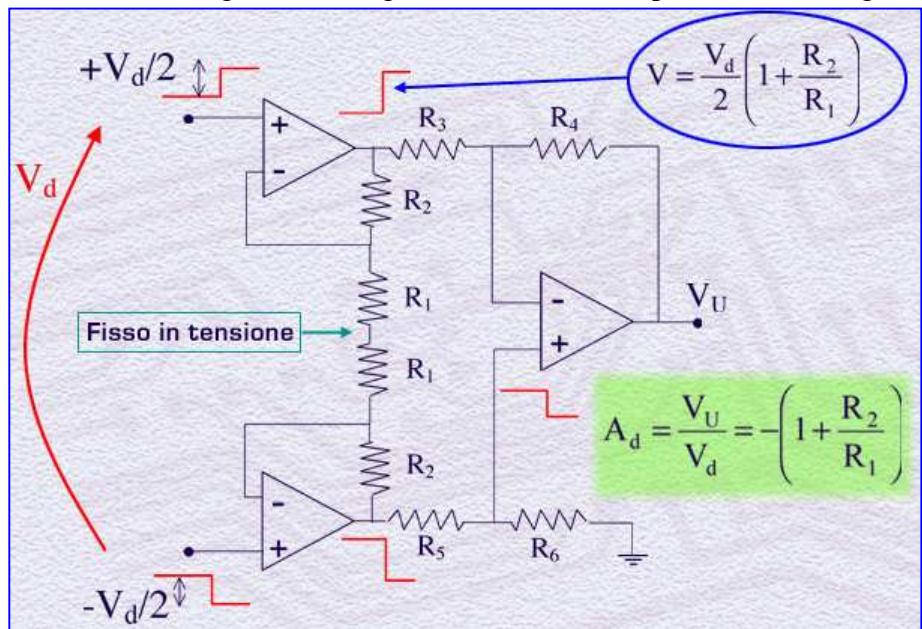
Per differenziarsi, questi amplificatori composti di elevate caratteristiche ed adatti a segnali differenziali sono spesso chiamati **"Amplificatori da strumentazione"** o, in inglese, **"Instrumentation Amplifiers"**.

Vediamone innanzitutto il comportamento su **segnale differenziale**.

Poiché l'amplificatore è simmetrico ed ha in ingresso un segnale simmetrico, il punto A delle figure precedente rimane effettivamente fisso in tensione comportandosi da terra virtuale. Pertanto le due parti sopra e sotto possono essere analizzate separatamente. E' immediato trovare che la tensione V, in uscita all'operazionale sopra, vale quanto indicato in figura. Lo stesso avviene sotto ma con segno diverso. Questo comporta che l'ultimo operazionale avrà il morsetto "+" che si dovrà spostare in giù di metà di questo valore come imposto dalla partizione sotto tra R<sub>5</sub> e R<sub>6</sub>. Questo spostamento si ripresenta necessariamente al morsetto "-" determinando una tensione ai capi della resistenza R<sub>3</sub> doppia rispetto al caso che il morsetto "-" stesse fisso, cosicché la corrente prodotta, che scorre in R<sub>4</sub>, fornisce una tensione di uscita doppia rispetto ad ogni singolo canale. Il guadagno corrispondente per segnali differenziali è quindi pari ad A<sub>d</sub> il cui valore è sempre indicato in figura, dove il segno meno rende conto dell'inversione di segnale dello stadio finale.

Su **segnale di modo comune**, invece, il comportamento interno è diverso. Ognuno dei due operazionali di ingresso viene sollecitato da un gradino nella stessa direzione e quindi non cambia la tensione ai capi delle resistenze R<sub>1</sub>, cosicché in esse non passa corrente. Le uscite dei due operazionali ripropongono pertanto lo stesso segnale dell'ingresso, sopra e sotto con lo stesso segno. Lo stadio finale sottrattore elimina queste due variazioni e mantiene l'uscita immobile a 0 V.

Il guadagno di modo comune è quindi pari a 0. In un sistema reale, a causa dei piccoli scostamenti dei valori tra i resistori della parte sopra



rispetto a quelli sotto ed a causa delle non idealità dei 3 operazionali usati, il guadagno di modo comune non è nullo ma ha un valore finito. Esso comunque risulterà molto migliore del singolo valore di ogni operazionale usato, per cui gli **"Instrumentation Amplifiers"** forniscono il più elevato valore di CMRR (Common Mode Rejection Ratio) che si può pensare di ottenere in un circuito elettronico. Vediamo nella nostra applicazione quali requisiti di reiezione al modo comune l'amplificatore deve assicurare nel suo complesso. L'uscita V<sub>U</sub> dell'amplificatore può essere rappresentata come una

combinazione lineare del segnale differenziale  $V_d$  e del segnale di modo comune  $V_{CM}$ , moltiplicati per i rispettivi guadagni  $A_d$  e  $A_{CM}$ . Possiamo riscrivere questa relazione raccogliendo  $A_d$ , come indicato in figura, così da mettere in evidenza il termine di  $CMRR = A_d / A_{CM}$ , ottenendo l'espressione di  $V_U$  come indicata in figura.

La tensione  $V_d$  rappresenta il segnale utile della termocoppia, mentre il termine  $V_{CM}/CMRR$  può essere interpretato come il segnale errore prodotto dalla reiezione finita al modo

comune dell'amplificatore che si somma, in ingresso, al segnale utile.

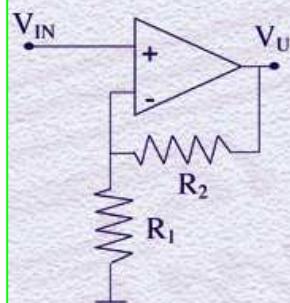
Nel nostro caso, ricordiamo,  $V_{CM} = 10 \text{ V}$  ed il segnale utile più piccolo con cui dobbiamo confrontare l'errore è di  $26.5 \mu\text{V}$ .

Pertanto il minimo valore di CMRR che si rende necessario, cioè quello per cui  $V_d = V_{CM}/CMRR$ , è di  $4 \times 10^5$ , vale a dire  $20 \log_{10} 4 \times 10^5 = 112 \text{ dB}$ .

Valori di 115-120dB sono molto elevati: li si può riscontrare in amplificatori realizzati in tecnologia ibrida, particolarmente curati e costosi, o solo nei migliori componenti integrati.

Dato che abbiamo ipotizzato di disporre di ADC con 5 V di fondo scala ed il massimo segnale del

$$\left. \begin{array}{l} \text{Massimo segnale dal sensore: } 25 \text{ mV} \\ \text{Fondo scala ADC: } 5 \text{ V} \end{array} \right\} G = 200$$



$$G_{id} = 1 + \frac{R_2}{R_1} = 200 \quad R_1 = 1 \text{ k}\Omega, R_2 = 199 \text{ k}\Omega$$

$$G_{reale} = \frac{A_0}{1 - G_{loop}} = \frac{A_0}{1 + A_0 \frac{R_1}{R_1 + R_2}} = 200 \quad \frac{R_2}{R_1} = 199.25$$

se  $A_0 = 100.000$  (100dB)

sensore è di 25 mV, il guadagno  $G$  del nostro amplificatore dovrà essere di  $G=200$ .

Dobbiamo quindi scegliere  $R_2$  e  $R_1$  in modo che il loro rapporto sia pari a 199, per esempio  $R_2=199\text{k}$  e  $R_1=1\text{k}$ .

A causa del valore finito del guadagno di anello Gloop (che in questo caso vale  $A_0 R_1 / (R_1 + R_2)$ ), il guadagno reale del nostro stadio non invertente così progettato sarebbe minore di 200. Per non avere alcun errore di guadagno basta semplicemente modificare opportunamente il rapporto  $R_2/R_1$ .

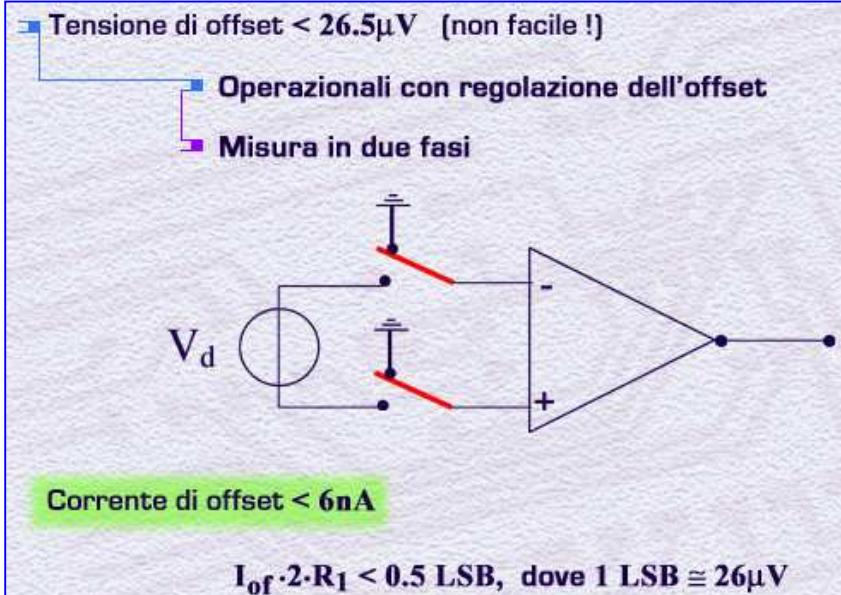
Supponendo che il guadagno  $A_0$  di ogni singolo operazionale sia di 100dB, cioè  $A_0 = 100.000$ , troveremmo  $R_2/R_1 = 199.25$ . Basta quindi regolare opportunamente il rapporto delle resistenze con un trimmer per avere il guadagno desiderato.

### Errori e tempi di misura

La tensione di offset degli operazionali impiegati deve essere minore di 1 LSB, cioè di  $26.5 \mu\text{V}$ .

Si tratta di un valore piuttosto basso, non sempre facilmente garantibile, se non a costi elevati, dai costruttori. Occorrerà quindi scegliere un AO che disponga della possibilità di regolazione dell'offset, oppure prevedere che la misura completa della temperatura avvenga in due fasi: una prima fase per la

misura dell'offset, in cui gli interruttori della figura sono chiusi verso terra, e una fase per la misura di



segna più offset, in cui gli interruttori collegano gli amplificatori alla termocoppia.

Per quanto riguarda la corrente di offset, essa sviluppa un errore in tensione pari a  $I_{of} \cdot 2 \cdot R_1$  che deve essere minore di 0.5 LSB (cioè di circa 12 $\mu$ V), da cui risulta che  $I_{of} < 6 \text{ nA}$ . Questo valore, pur piccolo, è disponibile tra gli AO integrati di buona qualità. Qualora si optasse per una misura a due fasi come indicato per la tensione di offset, si può osservare che nella fase in cui gli interruttori sono a terra si misura l'effetto di tutti gli offset, non solo di quello

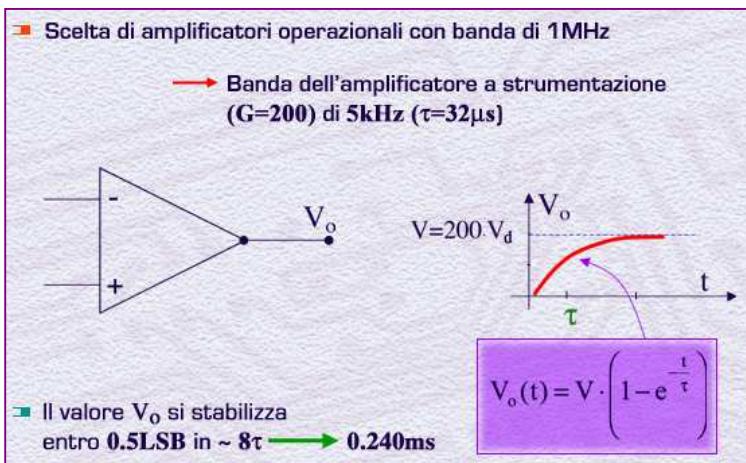
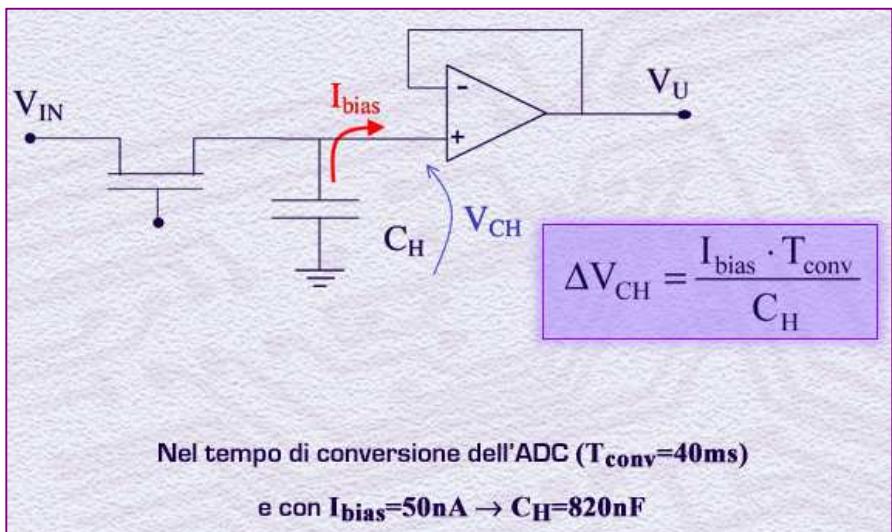
di tensione.

Per quanto riguarda il S&H ci limitiamo a considerare, come abbiamo visto nella relativa lezione, la variazione della tensione memorizzata sul condensatore di memoria  $C_H$  durante il tempo di conversione, prodotta dalla corrente di bias dell'operazionale utilizzato come buffer di tensione.

Tale variazione di tensione sappiamo che vale quanto indicato in figura.

Ricordando che 1 LSB vale, per il nostro ADC a 10 bit con 5 V di dinamica, 4.88 mV, ne risulta che  $\Delta V_{CH} < 2.44 \text{ mV}$ .

Assumendo di scegliere l'ADC con  $T_{conv}$  più lungo, cioè un ADC a doppia rampa con  $T_{conv}$  pari a 40 ms e di utilizzare un buffer con  $I_{bias} = 50 \text{ nA}$ , avremo che la capacità di holding dovrà valere  $C_H > (50 \text{ nA} * 40 \text{ ms}) / 2.44 \text{ mV} = 819.6 \text{ nF}$ . Potremmo quindi scegliere per  $C_H$  il valore conservativo di 1 microfarad.



Anche la banda passante degli amplificatori operazionali può giocare un ruolo nel definire i tempi legati alla misura. Assumiamo che gli AO utilizzati nella catena di misura abbiano un prodotto GBW = 1MHz (AO con buone prestazioni statiche di CMRR, guadagno e offset non hanno normalmente bande particolarmente elevate). In una configurazione non invertente che guadagna 200 la banda sarà allora di 5 kHz, a cui corrisponde una costante di tempo  $\tau$  di 32  $\mu$ s.

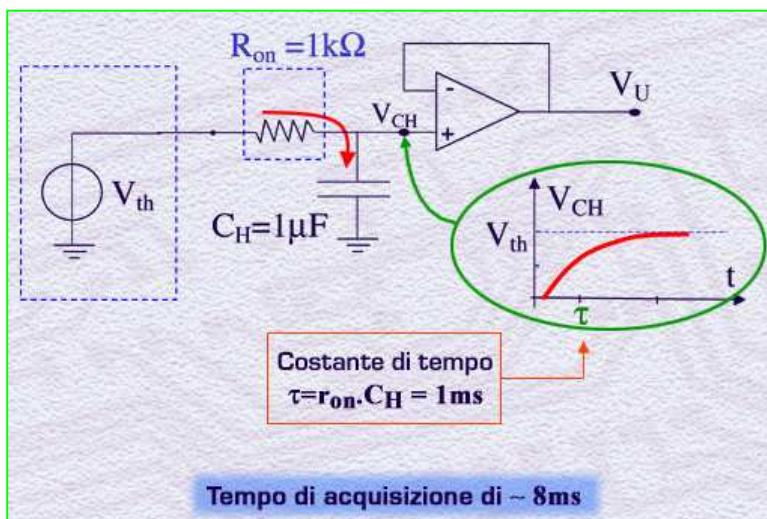
Se consideriamo la variazione di temperatura come un gradino di tensione applicato all'ingresso dell'amplificatore a strumentazione, e quindi alle due strutture amplificatrici non invertenti, ci aspettiamo una risposta, dell'amplificatore per strumentazione,  $V_o(t)$  di tipo esponenziale a singola

costante di tempo  $\tau$ , del tipo:  $V_o(t) = V(1 - e^{-t/\tau})$ , dove  $V$ = tensione della termocoppia\*200 (in cui 200 è il guadagno).

A causa di ciò, il comando di acquisizione al S&H dovrà essere dato solo dopo che la risposta dell'amplificatore ha raggiunto un valore che si discosta da quello asintotico finale per meno di 0.5 LSB. Poichè 0.5 LSB corrispondono allo 0.05% del fondo scala, dovrà essere soddisfatta la relazione  $[V - V(1 - e^{-t/\tau})] * 100/V < 0.049\%$ .

Possiamo facilmente trovare l'istante  $t$ , dopo circa 8 costanti di tempo, oltre il quale la disegualanza è verificata:  $t = -\tau \cdot \ln(0.049/100) = -\tau \cdot (-7.6) = 0.243 \text{ ms}$ .

Anche il blocco S&H, sappiamo, risponde in modo esponenziale nella fase di acquisizione.



Assumendo che l'interruttore del S&H abbia una resistenza  $r_{on}=1\text{k}\Omega$ , con la  $C_H=1\mu\text{F}$  scelta avremo una costante di tempo della rete RC del S&H di 1 ms. Il tempo di acquisizione del segnale con una precisione dello 0.05% del fondo scala è quindi di 8 ms circa, predominante rispetto al tempo di risposta dell'amplificatore. Come ci aspettavamo, su  $C_H$  pesano due esigenze contrapposte: deve essere piccolo per ridurre il tempo di acquisizione, ma deve essere grande per ridurre l'errore dovuto alla corrente di bias del buffer. L'abilità del progettista sta proprio

nell'individuare con lucidità le esigenze contrapposte e nel mediane al meglio in base alla specifica applicazione.

Con le ipotesi fatte, dopo circa 8 ms dalla chiusura degli interruttori, possiamo iniziare la conversione vera e propria. Un ADC a 10 bit ad approssimazioni successive può agevolmente compiere la conversione in 10  $\mu\text{s}$ . Considerata anche la necessità di trasferire poi la parola a un microcontrollore, possiamo assumere che tutto il ciclo di misura sia contenuto entro 10 ms. Questo sistema potrebbe quindi lavorare a un tasso massimo di 100 misure al secondo o, alternativamente, potrebbe leggere il segnale di 100 termocoppe diverse, ciascuna delle quali verrebbe letta una volta al secondo. Un ADC a doppia rampa impiegherebbe invece circa 40 ms per la conversione, a cui si devono aggiungere gli 8 ms della catena a monte. Il ciclo completo di misura durerebbe quindi circa 50 ms, a cui corrisponderebbe un numero massimo di 20 conversioni al secondo.

■ **Usando un ADC ad approssimazioni successive ( $T_{conv}=10\mu\text{s}$ )**

**ciclo di misura esaurito in 10ms**  
**(100 misure/s -inutile!**  
**o 100 termocoppe diverse lette ogni secondo)**

■ **Usando un ADC a doppia rampa ( $T_{conv}=40\text{ms}$ )**

**ciclo di misura esaurito in ~ 50ms**  
**(20 misure/s)**

## Esercizio 1

Riuscite ad immaginare quanti circuiti di acquisizione dati da sensori ci sono in un moderno Airbus passeggeri? Sapreste dire con quali sensori si misurano le seguenti grandezze:

- la velocità del velivolo rispetto a terra;
- la pressione dell'ambiente interno;
- l'umidità della stanza passeggeri;
- la quantità di carburante nel serbatoio.

Di ognuno di essi, provate a cercarne le caratteristiche tecniche ed immaginatevi il circuito equivalente elettrico migliore ai fini del progetto di un sistema di acquisizione dati.

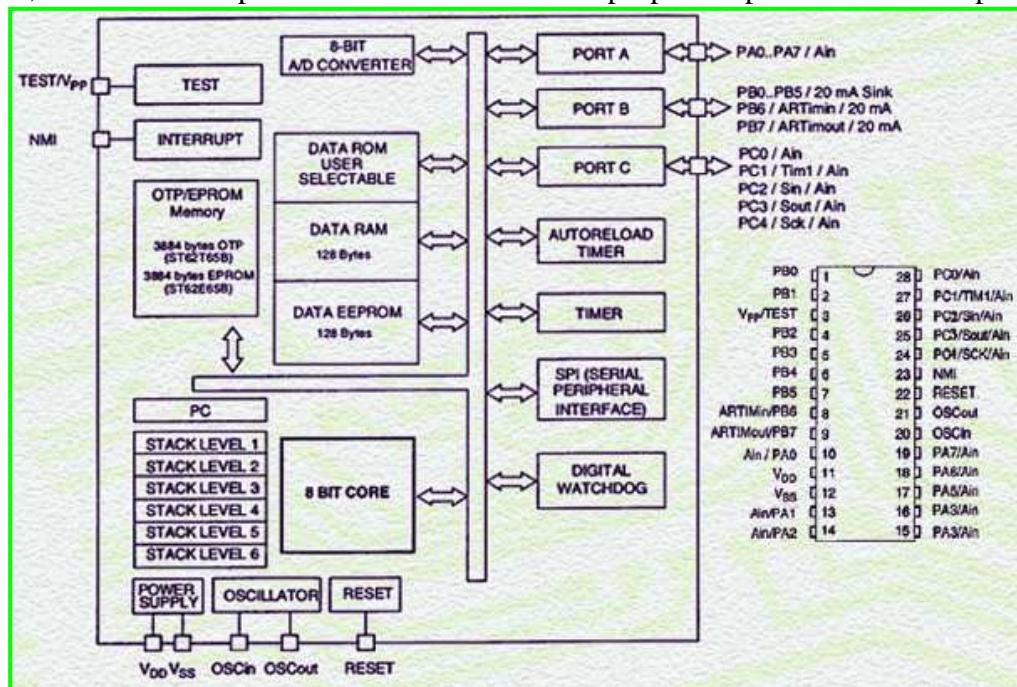
# Lezioni n. 48, 49 : Microcontrollori

In questa lezione si introdurranno le principali architetture interne dei microcontrollori. Come esempi, si tratteranno prodotti commerciali di fascia medio-bassa. Ciò permetterà di familiarizzare con i vari blocchi interni e di iniziare ad apprezzarne le potenzialità. In particolare si descriveranno le due tipologie di memoria presenti nei microcontrollori: quella di programma, in cui il progettista dovrà inserire il firmware, e quella per i dati temporanei, a disposizione dell'utente.

## Architettura Interna

I microcontrollori ( $\mu$ C) sono dei piccoli microprocessori ( $\mu$ P), che hanno molte funzioni accessorie on-chip a scapito, a volte, di una minore potenza di calcolo. Saranno proprio le periferiche on-chip a fornire loro elevata flessibilità di impiego e vastità di applicazioni.

Allo scopo di semplificarne l'analisi, tra i vari prodotti commerciali si prenderanno in considerazione soprattutto quelli di fascia medio-bassa, cioè con architetture ad 8 o 16 bit e con piedinature da 8 a 100 pin. Il tipo di prestazioni e il genere di periferiche disponibili sono diverse, a seconda della complessità del prodotto, ma i concetti e le strutture fondamentali sono comuni.

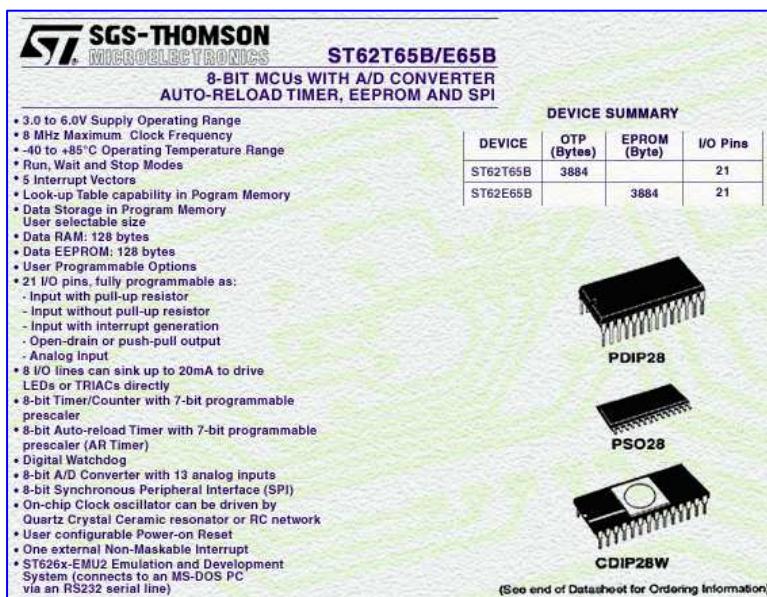


Non volendo illustrare tutti i possibili  $\mu$ C in commercio, ma neanche parlare di blocchi ideali, verranno trattati due  $\mu$ C di fascia bassa e prezzo molto contenuto (inferiore a 10 US\$). Essi sono i prodotti della famiglia PIC16C7X della Microchip e quelli della famiglia ST6xx della STMicroelectronics. La figura mostra il diagramma a blocchi e la piedinatura del microcontrollore ST62T65 di STMicroelectronics; si può notare come il cuore vero e proprio sia circondato dalle periferiche on-chip. Dalla piedinatura di questo microcontrollore si nota come alcuni pin abbiano due o addirittura tre funzioni, impostabili dall'utente via firmware, per rendere più compatto, e quindi più economico, il dispositivo.

Le caratteristiche di questo microcontrollore sono:

- PORTE (PA, PB, PC): porte di Input/Output (I/O) anche di tipo analogico (Ain), grazie alla presenza di un convertitore analogico-digitale (ADC) interno.
- TIMER (TIMin, TIMout): utilizzato per effettuare conteggi.
- PORTE di comunicazione (Sin, Sout, Sck): di tipo seriale per ridurre il numero di pin.
- RAM: di piccole dimensioni, può servire per memorizzare pochi dati (insufficiente per memorizzare i dati di un campionamento completo).
- OTP/EPROM: memoria programma. Può essere di tipo OTP (one-time programmable) oppure programmabile e cancellabile (otticamente) dal progettista. Si possono per esempio realizzare prototipi su  $\mu$ C riprogrammabili (e più costosi) per poi usare quelli non riprogrammabili per la produzione. Gli OTP possono essere programmati dal progettista (spendendo tempo e soldi per l'apparecchiatura di programmazione) oppure chiedendo alla stessa Casa madre di effettuare un "mask programming" in fase di design della maschera di metal (cosiddetta versione ROM). Quest'ultima scelta risulta conveniente quando si deve produrre un volume elevato di esemplari. Si possono altresì usare le E<sup>2</sup>PROM: uno dei vantaggi è la facile riprogrammabilità on-board, senza togliere il chip dalla scheda, poiché la cancellazione avviene elettricamente e non attraverso esposizione agli UV.

La figura riporta il frontespizio del data-sheet di questo componente, nel quale si possono notare i diversi package in cui può essere fornito. In particolare la versione con memoria programma interna di tipo EPROM, necessita del package con finestra trasparente per permettere la cancellazione, mediante esposizione a raggi UV.



(al più 4k) e come i package siano di dimensioni contenute (al massimo 40 pin). Per limitare l'ingombro del package ed il costo del  $\mu$ C, molte linee di I/O sono condivise sullo stesso piedino. Per questo motivo non è possibile usare

contemporaneamente tutte le porte (che nel caso di ST6 si chiamano PA, PB, PC), gli ingressi analogici (Ain) e quelli ausiliari di controllo delle periferiche (ad esempio per la comunicazione seriale: Sin, Sout e Sck). Il progettista dovrà scegliere cosa usare, impostando via firmware

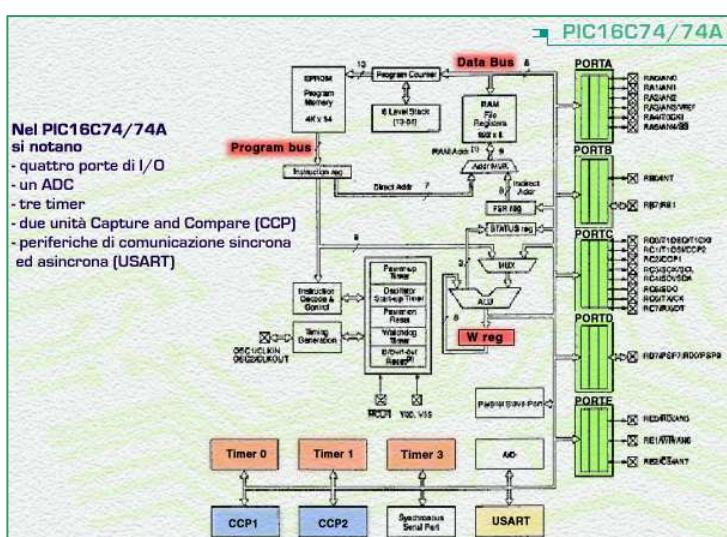
il  $\mu$ C in modo tale che il pin preso in

considerazione sia configurato come Input o Output, digitale o analogico.

Ad esempio il  $\mu$ C in figura è quello della famiglia dotato di un maggior numero di periferiche interne. Il registro all'uscita dell'ALU (che qui prende il nome di Working Register, W) di solito è noto con il nome di accumulatore. Si può notare la presenza di due bus: uno comandato dal Program Counter, che serve per accedere alle istruzioni (che verranno poi opportunamente 'decodificate' ed eseguite da una logica opportuna); l'altro, il bus dati, che permette la comunicazione tra la RAM e le periferiche (I/O, ADC, timer, ecc.).

PIC16C7x		Clock Frequency of Operation (MHz)	EPROM/Program Memory (x-tokens)	Data Memory (bytes)	Timer Module(s)	Peripherals			Features			Voltage Range (Volts)	In-Circuit Serial Programming	Brown-out Reset	Packages
Device	Pin Count					Serial Parallel 2-Wire Interface	Serial Port (SPI/C, USART)	Parallel Slave Port	ADC Converter (8-bit) Channels	Internal Sources	I/O Pins				
PIC16C710	20	512	36	TMRO	—	—	—	4	4	13	3.0-6.0	Yes	Yes	18-pin DIP SOIC 20-pin SSOP	
PIC16C71	20	1K	36	TMRO	—	—	—	4	4	13	3.0-6.0	Yes	—	18-pin DIP, SOIC	
PIC16C711	20	1K	68	TMRO	—	—	—	4	4	13	3.0-6.0	Yes	Yes	18-pin DIP, SOIC 20-pin SSOP	
PIC16C72	20	2K	128	TMRO, TMRI, TMR2	1	SPI/I <sup>2</sup> C	—	5	8	22	2.5-6.0	Yes	Yes	28-pin SDIP, SOIC, SSOP	
PIC16C73	20	4K	192	TMRO, TMRI, TMR2	2	SPI/I <sup>2</sup> C, USART	—	5	11	22	3.0-6.0	Yes	—	28-pin SDIP, SOIC	
PIC16C73A <sup>(1)</sup>	20	4K	192	TMRO, TMRI, TMR2	2	SPI/I <sup>2</sup> C, USART	—	5	11	22	2.5-6.0	Yes	Yes	28-pin SDIP, SOIC	
PIC16C74	20	4K	192	TMRO, TMRI, TMR2	2	SPI/I <sup>2</sup> C, USART	Yes	8	12	33	3.0-6.0	Yes	—	40-pin DIP 44-pin PLCC, MQFP	
PIC16C74A <sup>(1)</sup>	20	4K	192	TMRO, TMRI, TMR2	2	SPI/I <sup>2</sup> C, USART	Yes	8	12	33	2.5-6.0	Yes	Yes	40-pin DIP: 44-pin PLCC, MQFP, TQFP	

All PIC16/17 Family devices have Power-on Reset, selectable Watchdog Timer, selectable code protect and High I/O current capability.  
All PIC16C7X Family devices use serial programming with clock pin RB6 and data pin RB7.  
Note 1: Please contact your local sales office for availability of these devices.

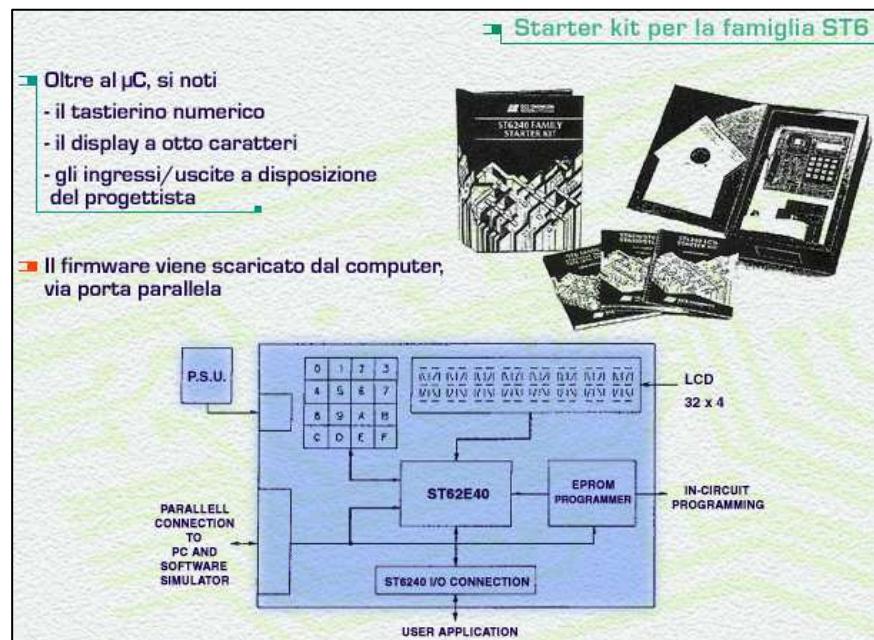


## Sistemi di Sviluppo per µC

Per poter acquisire confidenza con un particolare µC, con la gestione delle sue periferiche interne ed il suo insieme di istruzioni, molti Costruttori offrono al progettista un vero e proprio sistema di sviluppo, chiamato anche Demo Board, Starter Kit, Evaluation Board, Development System. Questi kit consistono solitamente di una scheda general purpose comprendente il µC e completa di tastierino numerico, display, pulsanti, alcuni ingressi ed uscite digitali, eventualmente qualche ingresso analogico, e del software che permette di sviluppare il firmware e trasferirlo nel µC della scheda. In questo modo, il progettista può prendere confidenza con il µC senza dover ideare una scheda prototipale, bensì può utilizzare quella già predisposta e verificata dal Costruttore. Il progettista dovrà dedicarsi esclusivamente alla stesura di firmware più o meno complessi, in grado di gestire tutte le periferiche interne e quelle addizionali esterne previste sulla evaluation board.

La figura sopra, mostra lo starter kit della STMicroelectronics per i µC della famiglia ST6xx, completo di manuali, scheda, cavo parallelo di connessione al computer e software di gestione e programmazione. Il costo di questi sistemi di sviluppo può variare da uno fino a qualche milione di lire, a seconda della fascia del µC e della componentistica accessoria presente sulla scheda (display, OpAmp per il condizionamento dei segnali provenienti da sensori, ADC o DAC esterni, buffer, ...).

Il progettista ha a disposizione diversi pacchetti software che facilitano la programmazione del µC e la stesura del codice. E' proprio questo software che



genera il vero e proprio file finale di programmazione (caratterizzato dai veri codici operativi in formato digitale) traducendolo dai codici mnemonici del listato, dalle macro e dalle label utilizzate via via nel programma. L'assemblatore è il software che permette di non scrivere direttamente in ASCII. Solitamente il programma viene scritto con i codici mnemonici (add, goto, ...), come mostrato nella videata in figura. Spesso è addirittura possibile scrivere routine direttamente in linguaggio ad alto livello (solitamente C). In quest'ultimo caso, tuttavia, si deve notare che la traduzione da linguaggio C ad assembler non è sempre ottimizzata e, quindi, questa possibilità risulta praticabile solo per µC di fascia alta. Per quelli di fascia medio bassa è preferibile prendere dimestichezza con l'assembler ed implementare il programma direttamente con i codici mnemonici.

Infine, lo stesso software permette di effettuare una simulazione del firmware appena sviluppato (l'hardware non è ancora stato realizzato). Con questa simulazione sarà possibile seguire passo passo l'esecuzione delle singole istruzioni, controllare il flusso di dati tra i vari registri, monitorare i bit di interesse nei registri "special function" e nelle porte di ingresso/uscita, misurare i tempi di esecuzione



e le durate temporali dei cicli. In questo modo sarà possibile effettuare un debug del firmware, prima di andarlo ad implementare nel vero e proprio µC e nell'hardware in cui esso sarà inserito.

In rete si trovano le versioni freeware o shareware di alcuni assemblatori, traduttori e compilatori per i µC di fascia bassa. Mentre per i pacchetti più complessi e per i µC di fascia medio alta, il costo può spaziare da pochi milioni a decine di milioni di lire. Il costo dipende della potenza del simulatore logico, alla possibilità di effettuare simulazioni in real-time ed, eventualmente, alla possibilità di inserire vettori di stimoli (sotto forma di file) agli ingressi digitali ed analogici, per simulare l'applicazione reale.

Per potere inserire il firmware all'interno del microcontrollore o, in gergo comune, per “bruciare il micro”, si usa un programmatore. E' lo stesso apparato che, solitamente, è in grado di programmare anche altri dispositivi digitali, quali memorie EPROM, E<sup>2</sup>PROM, FLASH o logiche programmabili PLD ed FPGA. La comunicazione tra computer, su cui si è progettato il firmware, e programmatore avviene solitamente attraverso seriale **RS232**.

Nel caso di produzione industriale di grandi volumi, è conveniente passare ad un gang-programmer, in grado di programmare in parallelo, contemporaneamente, più chip, riducendo sensibilmente il tempo

totale. Uno di questi prodotti è mostrato in figura.

Una volta programmato, il µC può essere inserito nel circuito prototipo e fatto funzionare. Solitamente compariranno problemi e malfunzionamenti dovuti sia al firmware del µC che all'hardware della scheda. Il debug di tutto il sistema può diventare molto difficile e dispendioso (in termini di tempo e costo), perché spesso è estremamente difficile identificare la causa del malfunzionamento. In tutti questi casi, sarebbe comodo potere “vedere” cosa sta accadendo all'interno del µC, monitorando i suoi

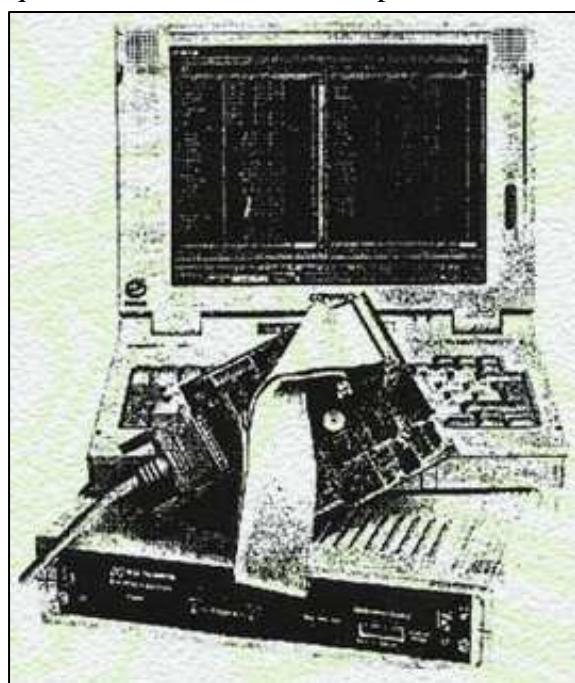
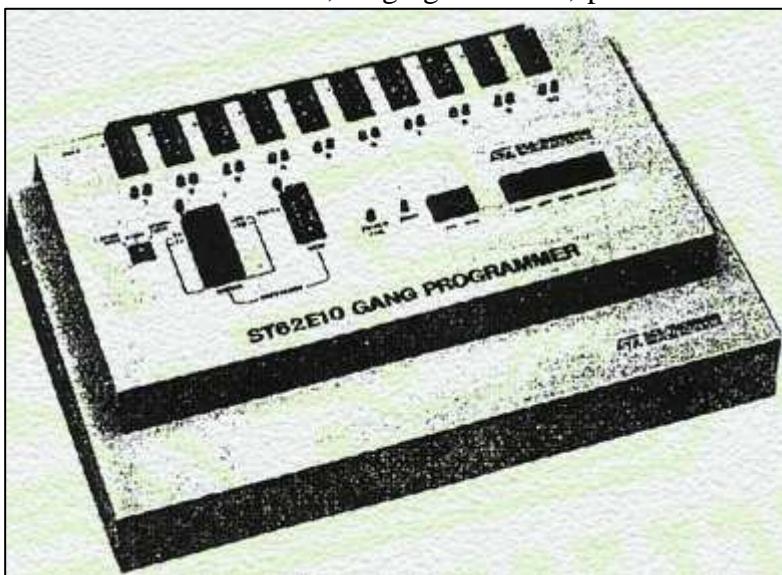
registri, le sue porte di ingresso/uscita, le sue periferiche interne, per scoprire il problema e risolverlo. In questo caso viene in aiuto l'emulatore, che è una scheda hardware anche molto complessa che fa le veci del microcontrollore vero e proprio e può essere inserita (tramite opportuno zoccolo adattatore) al posto del µC durante la fase di prototipazione della scheda e di verifica delle funzionalità. Un simile prodotto è mostrato in figura. Il firmware sarà inserito nella memoria dell'emulatore, il quale avrà ai suoi piedini lo stesso comportamento del µC che esso emula.

Così facendo si può evitare di programmare un µC vero, ma soprattutto è possibile monitorare il contenuto di tutti i registri e di tutte le periferiche interne al µC emulato.

Alcuni emulatori sono anche dotati di sonde analogiche e digitali per acquisire e controllare in tempo reale tensioni o livelli logici nella scheda, in modo da confrontarle con quanto atteso, durante lo svolgimento

delle operazioni. In questo modo è ancora più facile capire se il problema risiede nel firmware o nell'hardware della scheda.

Il costo di questi prodotti può variare da alcuni milioni di lire, per gli emulatori più semplici, a decine di milioni per quelli più accessoriati e per quelli in grado di emulare µC veloci e di fascia elevata.



## Memoria Programma

La memoria programma contiene il firmware che il µC dovrà eseguire. Il requisito principale che deve avere tale memoria è quello di essere non volatile. Ne esistono diversi tipi, brevemente descritti qui di seguito.

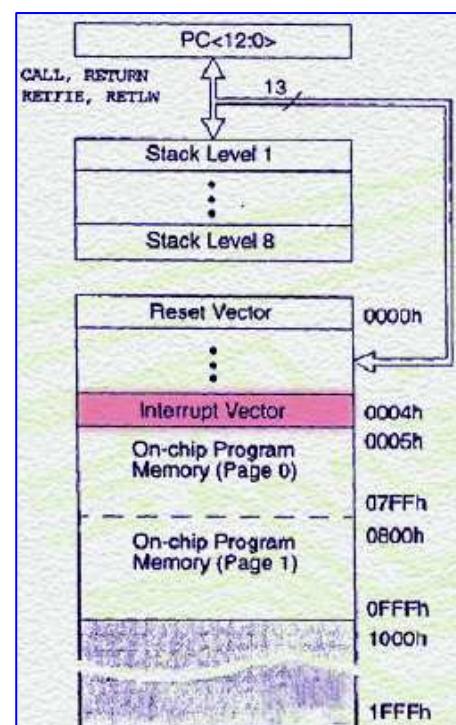
- **ROM:** si sottintende che sia “mask-programmed”, ovvero direttamente programmata sul chip dalla Casa costruttrice del µC, in accordo al listato fornito dal progettista. Questo tipo di soluzione è conveniente solo nel caso di grosse produzioni (dell'ordine di migliaia di pezzi).
- **OTP:** One-Time Programmable, è una memoria programmabile dal progettista stesso mediante un semplice programmatore di PROM/µC. Questo tipo di soluzione è conveniente nel caso di piccole produzioni, cioè fino a poche centinaia di pezzi; per maggiori quantitativi risulterebbe troppo dispendiosa in termini di tempo richiesto per la programmazione e di costo dell'operatore. Queste memorie vengono usate quando si è completata la messa a punto del programma e quindi si è certi del suo corretto funzionamento.
- **EPROM:** cancellabile mediante UV, richiede un package comprensivo di finestra trasparente agli ultravioletti per la cancellazione, pertanto risultano più costose delle OTP ma hanno l'ovvio vantaggio di poter essere totalmente cancellate e riprogrammate. Il loro impiego è fondamentale nella fase di sviluppo del firmware e di prototipazione, ovvero di messa a punto del software.
- **E<sup>2</sup>PROM:** sono programmabili e cancellabili elettricamente in modo veloce (non è necessario utilizzare la luce UV come per le EPROM). Sono implementate nei µC più costosi, ma per certe applicazioni il loro impiego risulta vantaggioso, in quanto permette un facile aggiornamento del firmware anche quando il µC è montato sulla scheda (mediante il cosiddetto In-System Programming).

Per programmare µC ci si avvale del linguaggio assembler costituito da istruzioni che gestiscono il funzionamento dell'integrato. Dato che l'assembler può risultare alquanto macchinoso e di difficile sviluppo, a volte si preferisce usare linguaggi ad alto livello (come il C) che verranno poi tradotti in assembler da un apposito software da PC (il cosiddetto compilatore). Nel seguito, indicheremo con ‘software’ il programma ad alto livello mentre con ‘firmware’ quello che gestisce più a basso livello l’hardware interno ed esterno al µC.

La memoria programma generalmente non è molto estesa, da 1k a 4k circa, con parole di 8 bit (per ST) o di 14bit (per PIC); solitamente è solo interna al µC e non offre la possibilità di espansione all'esterno. Vedremo che l'approccio a 14 bit permette di poter scrivere operazioni complesse in un'unica locazione di memoria, anche se ciò non si traduce direttamente in una maggiore velocità di esecuzione delle istruzioni rispetto alla soluzione ad 8 bit (che avrà bisogno spesso di due parole di memoria).

Come mostra la figura non tutta la memoria programma è utilizzabile dal progettista. Spesso le Case costruttrici mappano la memoria in diverse aree; talune riservate (perché destinate alla gestione delle periferiche o perché non implementate nel chip), altre contenenti espressamente i vettori di interrupt ed infine le rimanenti disponibili all'utente/progettista. Le prime righe sono utilizzate per la partenza del µC: nell'indirizzo 0000H (Reset vector) vi è la prima istruzione che il µC esegue quando viene resettato, che sarà un **jump** alla riga della prima istruzione. Vi sono poi altre righe privilegiate dove vengono memorizzati i vettori di interrupt: ogni volta che avviene una richiesta di interrupt viene interrotta l'istruzione in corso, il program counter viene salvato nello stack e viene effettuato un salto alla cella di memoria relativa all'interrupt intervenuto, per avviare l'esecuzione della relativa routine.

Nel caso del PIC16C71 (ovvero µC semplici), è disponibile 1k di memoria programma. Per indirizzare tale memoria sarebbero necessari 10 bit di address che salirebbero a 13 nel caso di 8k. Per risparmiare su di un bus di address così esteso, viene utilizzata una memoria mappata in pagine, come nell'esempio della figura, ad esempio, per il

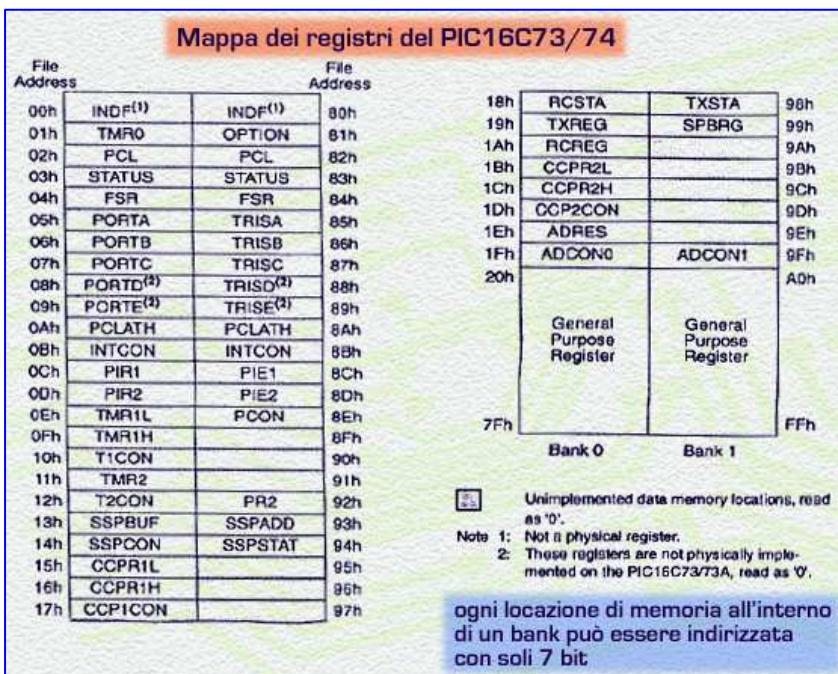


PIC16C74A. Lo scopo è quello di utilizzare un numero inferiore di bit per accedere all'interno di una pagina, (bit di page-select); cosicché il program-counter possa essere più corto e le istruzioni di **jump** vadano ad occupare un solo op-code.

Come si vede, nel caso del PIC la memoria è completamente indirizzabile con 13 bit (l'ultima cella è appunto 1FFFH pari a  $1+4+4+4=13$ bit). Nonostante questo, per accedere a ciascuna pagina bastano 11 bit (infatti 07FFH è pari a 00'0111'1111'1111). Il passaggio da una pagina all'altra avviene asserendo il bit di pagina (Page Select). Se in futuro il Costruttore potrà (e vorrà) integrare più memoria on-chip, non servirà modificare il program counter, bensì basterà aggiungere un bit al registro di pagina, che diverrà a due bit.

# Memoria Dati

Questa memoria è solitamente utilizzata per contenere i risultati temporanei delle elaborazioni effettuate dal µC, o i valori numerici acquisiti dalle porte di ingresso, dal convertitore ADC o dai



PIC16C7X questo bit è il quinto contenuto dal registro STATUS ed è chiamato **RP0**: impostandolo a 0 oppure ad 1, il µC accederà al bank 0 o 1.

Gli indirizzi della memoria dati vanno da  $00_H$  ad  $FF_H$  ma a disposizione dell'utente vi sono solo 96 bytes di memoria per ogni banco. Nell'eventualità in cui si cerchi di leggere una cella che non c'è (ad esempio indirizzando il Bank 3), il programma potrebbe causare degli errori, perché si leggerebbero cose insensate (poichè la cella in questione viene vista in alta impedenza) oppure si scriverebbe a vuoto un dato importante, perdendolo.

Procedendo nel modo appena visto, ogni locazione di memoria all'interno di un bank può essere indirizzata con soli 7 bit. Questi bit possono far parte del medesimo op-code ed essere trattati come veri e propri registri privilegiati.

Questo tipo di architettura è denominata register file e permette di indirizzare le celle in RAM mediante il loro “nome” (codice a 7 bit), come fossero registri aggiuntivi all’accumulatore.

Solo il working register **W** (nel PIC) è fisicamente implementato all'interno dell'ALU, esso contiene i risultati delle operazioni aritmetiche (SUM, SUB), logiche (AND, OR, NOT) e di shift. L'altro operando invece, può essere un registro qualsiasi della RAM (f sta per file register) oppure una costante passata immediatamente nell'istruzione. Ad esempio la somma di W ed un generico registro f occupa un solo op-code: ADDWF f ,d op-code = 00 0111 dfff ffff, dove l'indirizzo a 7 bit di f specifica la locazione nella RAM dati, mentre il bit d indica se salvare il risultato nel registro W (d=0) oppure in f (d=1).

Con l'istruzione: ADDWF 45h, 0, si somma a W il contenuto del registro 45H ed il risultato viene memorizzato in W.

timer interni. Nei  $\mu$ C di fascia bassa, solitamente la memoria dati è di piccole dimensioni, ossia di poche decine o centinaia di bytes. Solo nei  $\mu$ C di fascia elevata si può arrivare a qualche kbyte di RAM.

Sempre per risparmiare spazio, l'architettura interna della memoria dati è organizzata a banchi o register file, in modo da realizzare l'indirizzamento con un numero inferiore di bit ed avere op-codes più compatti. La figura mostra la tipica memoria dati di un µC della famiglia PIC. Per passare da un bank ad un altro basta impostare a 0 o ad 1 un bit in un registro speciale. Ad esempio nel

Mediante: ADDLW K op-code = 11 111x kkkk kkkk, la costante K ad otto bit viene sommata al working register W. Il bit x significa “don’t care”, cioè può assumere indifferentemente il valore 1 o 0. Risulta così evidente come con una sola istruzione (a 14 bit) si possano rappresentare dei codici operativi anche complessi. Inoltre, anche le operazioni tra più celle di memoria risultano semplificate utilizzando l’architettura register-file. Per esempio se si desidera sommare all’operando A il valore dell’operando B, non è necessario passare attraverso l’accumulatore, con le istruzioni:

LDA operando A ADD operando B STA operando B

bensì è possibile usare solo due istruzioni:

MOVFW operando A ADDWF operando B,1.

Nella memoria RAM sono implementati anche i registri speciali, la cui presenza è necessaria per il controllo delle periferiche interne al μC, degli interrupt, dei dati letti e scritti sulle porte di I/O, dei contenuti dei timer, ecc., così come mostrato in figura. Impostando il valore 1 oppure 0 nel rispettivo bit, si modificano i percorsi interni dei segnali nel μC o si modifica la struttura interna di ciascuna periferica on chip. Per questo motivo, il significato di ogni bit deve essere ben chiaro al progettista che dovrà necessariamente saperli usare, leggendoli o

PIC16C7X										Registri speciali	
PIC16C73/73A/74/74A SPECIAL FUNCTION REGISTER SUMMARY											
Address	Name	Bit 7	Bit 6	Bit 5	Bit 4	Bit 3	Bit 2	Bit 1	Bit 0	Value on POR, BOR	Value on all other power reset (V)
<b>Bank0</b>											
00H	INDF	Addressing this location uses contents of PSW to address data memory bank or physical register									
01H	TMR0	Timer Module 0 Register									
02H	PCL	Program Counter (PC) Least Significant Byte									
03H	STATUS	WIF	RIF	PIR0	TO	PE	Z	DC	C	0001 3FFF	0000 1000
04H	PER	Indirect data memory address pointer									
05H	PORTA	PORTA Data Latch when written: PORTA pins when read									
06H	PORTB	PORTB Data Latch when written: PORTB pins when read									
07H	PORTC	PORTC Data Latch when written: PORTC pins when read									
08H	PORTD	PORTD Data Latch when written: PORTD pins when read									
09H	PORTF	PORTF Data Latch when written: PORTF pins when read									
0AH	POLATH	Writing this location uses contents of PSW to address data memory bank or physical register									
0BH	INTCON	OE	PEIE	TOIE	INTE	PIE1	TOP	INTF	PIEF	0000 9000	0000 0000
0CH	PIR1	ADIF	RCIF	TXIF	SPIIF	CPIF	TMR1IF	TMR1OF	—	0000 0000	0000 0000
0DH	PIR2	—	—	—	—	—	—	—	DCPINF	0000 0000	0000 0000
0EH	TMR1L	Holding register for the Least Significant Byte of the 16-bit TMR1 register									
0FH	TMR1H	Holding register for the Most Significant Byte of the 16-bit TMR1 register									
10H	T1CON	SEL1	SEL0	T1OSCEN	T1TEN	T1M1CS	T1M1ON	—	—	0000 0000	1000 0000
11H	TMR2	Timer modulus register									
12H	T2CON	SEL1	SEL0	T2OSC1	T2OSC2	T2OUTPS1	T2OUTPS2	T2MRON	T2CKPS1	0000 0000	0000 0000
13H	SIPBUF	Synchronous Serial Port Receive Buffer/Transmit Register									
14H	SIPCON	WCOL	SIPOV	SIPEN	CAP	SIPM2	SIPM3	SIPM4	SIPM5	0000 0000	0000 0000
15H	COPR1H	Capture/Compare/PWM Register (LSD)									
16H	COPR1L	Capture/Compare/PWM Register (MSB)									
17H	COPCON	CP0IF	CP0EN	CP0PS	CP1IF	CP1EN	CP1PS	CP2IF	CP2EN	0000 0000	0000 0000
18H	PCON	SPEN	PO0	SREN	CPEN	—	PEAR	OCER	PCIO	0000 0000	0000 0000
19H	TXREG	USART Transmit Data Register									
1AH	RCREG	USART Receive Data Register									
1BH	COPR2H	Capture/Compare/PWM Register (LSD)									
1CH	COPR2L	Capture/Compare/PWM Register (MSB)									
1DH	ADRES	A/D Result Register									
1EH	ADCON	ADP1	ADC06	CM1S2	CM1S1	CM0S2	CM0S1	GOV1	GOV0	0000 0000	0000 0000
1FH									ADON	0000 0000	0000 0000

impostandoli al momento opportuno all’interno del firmware. Le altre locazioni sono invece destinate ai registri general-purpose a disposizione del progettista. Ad esempio, il registro di Stato, riportato in figura seguente, contiene i bit di flag, quali il carry, l’overflow, lo zero, tutti asseriti dopo ogni operazione che coinvolge la ALU.

Registri speciali											
RW0	RW1	RW2	RW3	R1	R2	R3	RW4	RW5	RW6	RW7	RW8
RP0	RP1	RP0	TO	PD	Z	DC	RP0	RP1	RP0	RP1	RP0
<b>bit 7:</b> RP0: Register Bank Select bit (used for indirect addressing)											
1 = Bank 2, 3 (10H - FFH)	0 = Bank 0, 1 (00H - FFH)	0 = Bank 2, 3 (10H - FFH)	0 = Bank 0, 1 (00H - FFH)	0 = Bank 1 (00H - FFH)	0 = Bank 0 (00H - FFH)						
<b>bit 6-5:</b> RP1:RP0: Register Bank Select bits (used for direct addressing)											
11 = Bank 3 (10H - FFH)	10 = Bank 2 (00H - FFH)	10 = Bank 2 (00H - FFH)	10 = Bank 1 (00H - FFH)	10 = Bank 1 (00H - FFH)	10 = Bank 0 (00H - FFH)						
<b>bit 4:</b> TO: Time-out bit											
1 = After power-up, CLRTO instruction, or SLEEP instruction	0 = A WDT time-out occurred										
<b>bit 3:</b> PD: Power-down bit											
1 = After power-up, or by the CLRPD instruction	0 = By execution of the SLEEP instruction	0 = By execution of the SLEEP instruction	0 = By execution of the SLEEP instruction	0 = By execution of the SLEEP instruction	0 = By execution of the SLEEP instruction	0 = By execution of the SLEEP instruction	0 = By execution of the SLEEP instruction	0 = By execution of the SLEEP instruction	0 = By execution of the SLEEP instruction	0 = By execution of the SLEEP instruction	0 = By execution of the SLEEP instruction
<b>bit 2:</b> Z: Zero bit											
1 = The result of an arithmetic or logic operation is zero	0 = The result of an arithmetic or logic operation is not zero	0 = The result of an arithmetic or logic operation is not zero	0 = The result of an arithmetic or logic operation is not zero	0 = The result of an arithmetic or logic operation is not zero	0 = The result of an arithmetic or logic operation is not zero	0 = The result of an arithmetic or logic operation is not zero	0 = The result of an arithmetic or logic operation is not zero	0 = The result of an arithmetic or logic operation is not zero	0 = The result of an arithmetic or logic operation is not zero	0 = The result of an arithmetic or logic operation is not zero	0 = The result of an arithmetic or logic operation is not zero
<b>bit 1:</b> DC: Digit carry/borrow bit (ADDW, ADDWF, SUBWF, SUBWF instructions)(or borrow the polarity is reversed)											
1 = A carry-out from the 4th low order bit of the result	0 = No carry-out from the 4th low order bit of the result	0 = No carry-out from the 4th low order bit of the result	0 = No carry-out from the 4th low order bit of the result	0 = No carry-out from the 4th low order bit of the result	0 = No carry-out from the 4th low order bit of the result	0 = No carry-out from the 4th low order bit of the result	0 = No carry-out from the 4th low order bit of the result	0 = No carry-out from the 4th low order bit of the result	0 = No carry-out from the 4th low order bit of the result	0 = No carry-out from the 4th low order bit of the result	0 = No carry-out from the 4th low order bit of the result
<b>bit 0:</b> C: Carry/borrow bit (ADDW, ADDWF, SUBWF, SUBWF instructions)											
1 = A carry-out from the most significant bit of the result occurred	0 = No carry-out from the most significant bit of the result occurred	0 = No carry-out from the most significant bit of the result occurred	0 = No carry-out from the most significant bit of the result occurred	0 = No carry-out from the most significant bit of the result occurred	0 = No carry-out from the most significant bit of the result occurred	0 = No carry-out from the most significant bit of the result occurred	0 = No carry-out from the most significant bit of the result occurred	0 = No carry-out from the most significant bit of the result occurred	0 = No carry-out from the most significant bit of the result occurred	0 = No carry-out from the most significant bit of the result occurred	0 = No carry-out from the most significant bit of the result occurred
<b>Contenuto del registro Status (indirizzi 03H e 03H) del PIC16C73/74</b>											

INSTRUCTION SET											
Mnemonic, Operande	Description	Cycles	14-Bit Opcode	status Affected							
		M68	LSD								
<b>BYTE-ORIENTED FILE REGISTER OPERATIONS</b>											
ADDWF	I, d	Add W and f	1	00 0111 dfff ffff	0, DC, Z	1, 2					
ANDWF	I, d	AND W with f	1	01 0001 dfff ffff	Z	2					
CLRF	I	Clear f	1	00 0001 0000 0000	Z	2					
CLRF	-	Clear W	1	00 0001 0000 0000	Z	2					
COMP	I, d	Complement I	1	00 0001 dfff ffff	Z	2					
DISCF	I, d	Decrement f	1	00 0011 dfff ffff	Z	2					
DISCFZ	I, d	Decrement, Skip if C	1(2)	00 1011 dfff ffff	Z	2, 2.5					
INCFF	I, d	Increment f	1(2)	00 1111 dfff ffff	Z	2, 2.5					
INCF	I, d	Increment W	1(2)	00 0001 dfff ffff	Z	2, 2.5					
MOVWF	I, f	Move I	1	00 1000 0fff 0fff	Z	2					
MOVWF	I, f	Move W to I	1	00 0000 0fff 0fff	Z	2					
NOP	-	No Operation	1	00 0000 0000 0000	Z	2					
RLF	I, f	Rotate Left through Carry	1	00 1101 dfff ffff	C	1, 2					
RRF	I, f	Rotate Right through Carry	1	00 1100 dfff ffff	C	1, 2					
SUBWF	I, f	Subtract W from f	1	00 1000 dfff ffff	C, DC, Z	1, 2					
SUBWF	I, f	Subtract W from f	1	00 1100 dfff ffff	Z	1, 2					
XORWF	I, f	Exclusive OR W with f	1	00 1100 dfff ffff	Z	1, 2					
XORWF	I, f	Exclusive OR W with f	1	00 1100 dfff ffff	Z	1, 2					
<b>BIT-ORIENTED FILE REGISTER OPERATIONS</b>											
ADDLW	k	Add literal and W	1	11 111x kkkk kkkk	0, DC, Z						
ANDLW	k	AND literal with W	1	11 1001 kkkk kkkk	Z						
CALL	k	Call subroutine	2	10 000x kkkk kkkk							
CLRWF	k	Clear W	1	00 0000 0010 0000	TO, PD						
GOTO	k	Go to address	2	10 100x kkkk kkkk							
IORLW	k	Inclusive OR literal with W	1	11 1000 kkkk kkkk							

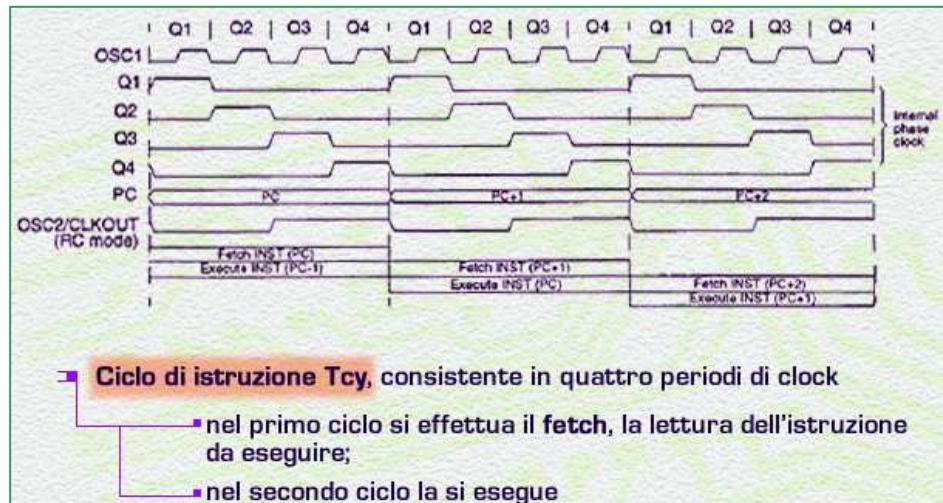
processori, ad esempio l'ST6, utilizzano un op-code di due parole da 8 bit ciascuna: il primo byte specificherà il tipo di operazione, mentre il secondo specificherà il dato.

Per indirizzare una certa locazione di memoria RAM è possibile specificarne l'indirizzo all'interno dell'op-code (ad esempio con i 7 bit ffff visti negli esempi precedenti): questo è il cosiddetto indirizzamento diretto. Tuttavia, se si dovesse scandire una zona di RAM senza conoscerne a priori l'estensione oppure il punto di partenza o di arrivo, sarebbe preferibile avere a disposizione una modalità diversa di indirizzamento, detto indiretto.

### Ciclo Istruzione

Le operazioni del  $\mu$ C sono cadenzate da un clock, essendo una macchina microprogrammata e, in definitiva, sequenziale. Non è comunque vero che in ogni periodo di clock venga eseguita un'intera operazione. Infatti, il ciclo istruzione viene eseguito in più periodi di clock. Ad esempio nel PIC, in ciascun periodo del clock viene svolta solo una fase di un'operazione.

Come mostrato nello schema temporale, nei primi quattro periodi di clock, il  $\mu$ C va a prendere (fetch) la prossima istruzione indirizzata dal program-counter; nei

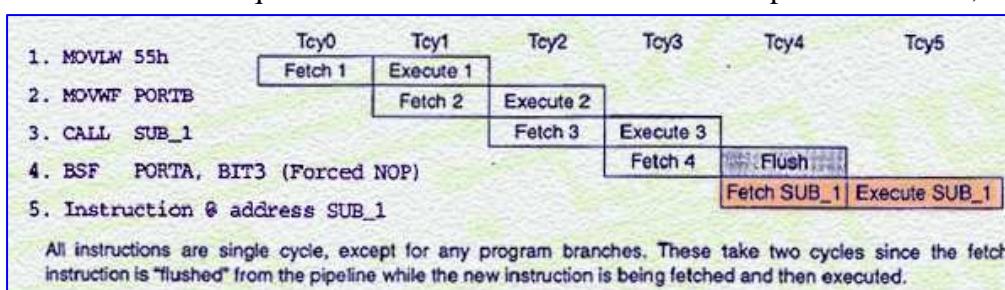


successivi quattro, decodifica l'istruzione, ovvero comprende di che tipo di comando si tratta e di quali operandi eventualmente necessita, andandoli ad indirizzare ed inviandoli all'ALU o alle periferiche desiderate. Quindi un ciclo istruzione (indicato con Tcy, time for cycle) consiste di quattro periodi di clock ed un'istruzione completa necessita di due cicli istruzione, come illustrato in figura precedente.

Proprio per sfruttare il parallelismo della fase di fetch dell'istruzione attuale, con quella di execute dell'istruzione precedente, è conveniente impiegare l'architettura di tipo pipeline. Così in Tcy1 viene eseguita l'istruzione caricata nel ciclo precedente, ma si procede anche alla lettura della nuova istruzione (Fetch 2). In questo modo il throughput del  $\mu$ C risulta raddoppiato: nel caso del PIC, utilizzando la frequenza massima consentita di clock pari a 20MHz, si ha un throughput di

un'istruzione ogni 200ns, nonostante il tempo di esecuzione di un'istruzione sia di 400ns.

Nel flusso mostrato nello schema si nota la presenza di una



particularità: la chiamata a subroutine `CALL SUB_1` viene letta al secondo ciclo istruzione (Tcy2). Pertanto l'acquisizione Fetch 4 risulta inutile: l'istruzione 4 non verrà eseguita (si ha un Flush invece dell'Execute 4) ed in Tcy4 si effettua il salto andando a leggere l'istruzione all'indirizzo `SUB_1`. In questo caso le istruzioni effettive sono quattro, ma i cicli trascorsi sono stati cinque (tralasciando il Tcy0 di fetch), perché uno è andato a vuoto.

**Pag.1**

*In questa lezione si introdurranno le porte di Ingresso/Uscita dei microcontrollori ed il convertitore ADC interno al µC. Infine si mostreranno alcuni esempi di impiego delle porte di ingresso/uscita del µC.*

## Porte di I/O

**Pag.2**

Le porte di I/O sono l'esempio più semplice di interfaccia interna al µC, che gli permette di venire a contatto con il mondo esterno. La **Pag.2** schematizza alcune possibili impostazioni dei pin di I/O del microcontrollore ST62xx. Il singolo pin di I/O può essere un ingresso (I) o un'uscita (O) del µC in base a come viene impostata la porta stessa; di conseguenza può essere configurato ed utilizzato in diversi modaliità, così riassumibili:

- Input digitale: può essere un ingresso CMOS normale o a trigger di Schmidt (che consente di eliminare l'incertezza nella commutazione ed aumenta i Noise Margin). È possibile disporre di pull-up interno (si risparmia la resistenza esterna), in modo da poter lasciare flottante il piedino.
- Input con Interrupt: oltre che come semplice ingresso dati, può anche essere abilitato per la generazione di un interrupt all'interno del µC.
- Input analogico: permette di acquisire grandezze analogiche ed inviarle al convertitore Analogico/Digitale (ADC, *veda la Lezione ad esso dedicata*) interno al µC, per successive elaborazioni.
- Output digitale: normalmente a livelli CMOS; a volte può essere anche open drain (con il solo n-MOS di pull-down in uscita, per permettere la connessione cablata, wired-and); può o meno avere resistenza interna di pull-up; può essere three-state (in grado cioè di andare in alta impedenza e rimanere flottante).

**Pag.3**

La porta di I/O digitale è semplicemente schematizzabile come mostrato a **Pag.3**. Se l'ingresso di selezione è a 1, i MOS sono entrambi spenti, indipendentemente dal valore di output; l'uscita è in alta impedenza, quindi il pin è configurato per essere un ingresso. Viceversa se l'ingresso selezione è mantenuto basso, le porte diventano trasparenti al valore di output (negato) applicato alla porta ed il pin agisce come uscita. Si può notare la presenza dei due diodi di protezione.

**Pag.4**

### Registri di Configurazione

Per poter configurare ciascun I/O, vi è un struttura abbastanza complessa per ciascun pin che dovrà essere impostata via firmware, come mostrato a **Pag.4** per l'ST62xx. Ciascuna porta di I/O (A, B, C) possiede tre registri di 8 bit, un bit per ogni linea della porta. Questi registri sono il **Data Register** (contiene il valore letto o da scrivere sulla porta corrispondente), il **Direction Register** (specifica se la linea è di input o di output) e l'**Option Register** (se la linea è di input, specifica se l'interrupt è abilitato o meno; se la linea è di output, specifica se essa è open-drain oppure normale).

Nei µC della Microchip la scelta di come usare il pin di I/O avviene mediante lo Special Function Register chiamato **TRIS**: se è a 1a forza il pin in three-state (quindi può agire da input), mentre se è a 0 riversa il contenuto del latch sull'uscita (quindi il pin agisce da output). Ogni registro **TRIS** da 8 bit avrà ciascun bit collegato al relativo pin della porta corrispondente (**TRISA**, **TRISB** e così via). Il bit letto dalla porta sarà scritto nel relativo bit del registro corrispondente **PORT**. Si noti che all'accensione del µC, i registri **TRIS** sono precaricati ad 1, configurando tutti i pin come ingressi, in modo da evitare eventuali cortocircuiti all'esterno del µC.

### Ingressi/Uscite di Periferiche Interne

Taluni pin di I/O possono essere impiegati dalle periferiche on-chip. Per esempio, un ingresso può essere utilizzato come segnale di trigger per il Timer interno: è proprio questo il ruolo alternativo del bit4 della porta A nel PIC 16C7x, che si chiama appunto RA4/TOCK1. In riferimento alle operazioni di configurazione, la presenza del buffer RD TRIS rende possibile conoscere l'attuale impostazione dei bit della porta.

Nei µC più complessi, esistono intere porte compatibili come bus da 8/16 bit per interfacciarsi a microprocessori o memorie esterne al µC. In tal caso, oltre al bus dati di I/O, saranno disponibili anche piedini predisposti al controllo di scrittura (WR) e lettura (RD) delle periferiche esterne.

### Ingressi di Interrupt

Infine vi sono ingressi che possono essere utilizzati per generare interrupt all'interno del µC. In questo caso vi sarà un registro per definire il tipo di evento che può scatenare l'interrupt, ed in particolare:

- interrupt-on-change (avviene indifferentemente sul fronte di salita o di discesa);
- interrupt-on-rising edge (avviene sul fronte di salita);
- interrupt-on-falling edge (avviene sul fronte di discesa).

## Gestione degli Interrupt

### Pag.5

Quando si verifica un evento asincrono di interrupt (Interrupt Request, IRQ) il programma principale viene interrotto e viene eseguita una routine dedicata. Le richieste di interrupt possono essere generate da diverse sorgenti:

- dai pin di una qualunque porta di I/O opportunamente abilitata;
- dal pin privilegiato di Non-Maskable Interrupt (NMI), proveniente dall'esterno;
- dalle periferiche interne al µC (per esempio dall'ADC, per comunicare l'avvenuta conversione e la disponibilità del dato, o dal timer, per segnalare la fine del conteggio).

### Pag.6

Per gestire un interrupt è necessario che il µC sappia l'indirizzo in memoria programma dove si trova la routine di gestione firmware vera e propria, la quale può essere diversa a seconda della sorgente che ha scatenato l'interrupt. Questa selezione è implementabile in due modalità diverse: a vettore di interrupt oppure a polling.

#### Vettore di interrupt

Si tratta di una zona di memoria composta da alcune celle attigue, solitamente 7, ciascuna delle quali contiene l'indirizzo di inizio della routine di interrupt relativa. A seconda della sorgente che scatena la chiamata, il core del µC andrà a prelevare nella relativa cella del vettore l'indirizzo a cui saltare. Nei µC di fascia elevata vi è una cella per ogni possibile causa di interrupt, ovvero per ogni periferica interna e pin dedicato.

Per evitare che il µC venga interrotto troppo spesso dall'occorrenza di diversi interrupt, spesso l'architettura interna permette di "mascherare" quelli meno significativi, ad esempio originati dalle periferiche più lente o di minore importanza. In questo modo, a seconda di ciò che si vuole fare elaborare al µC, il progettista deciderà di mascherare o meno alcuni interrupt di priorità inferiore ad una certa soglia impostabile via firmware.

Magari per un errore nel firmware o per un vero e proprio bug, potrebbe capitare che venga impostata una priorità troppo elevata, per cui tutti gli interrupt vengono mascherati: in questo modo non sarebbe più possibile interrompere la normale esecuzione del µC. Per evitare ciò, i Costruttori assicurano comunque la presenza di un piedino NMI, non mascherabile appunto, sempre in grado di scatenare la chiamata alla routine relativa, facendo così uscire il µC da una condizione di stallo o malfunzionamento in cui era finito.

#### Polling

I µC di fascia bassa hanno un'unica cella di interrupt, come mostrato in figura a **Pag.6**, da cui leggere l'indirizzo di inizio della sola routine di interrupt implementata. Tutte le periferiche interne ed i pin dedicati vengono gestiti da questa unica routine, qualunque sia la sorgente che ha scatenato l'interrupt. Per identificare la periferica chiamante, dovrà essere cura del progettista sviluppare una routine che vada a testare una per una le periferiche in modo da localizzare l'origine dell'interrupt. Ad esempio dovrà controllare in successione se e quale piedino dedicato sia stato asserito, oppure se si è trattato della fine conversione dell'ADC, o dell'overflow del timer, o dell'arrivo di un dato dalla porta seriale, e così di seguito.

Si nota che questa modalità è sicuramente più dispendiosa in termini di tempo di esecuzione della precedente, la quale instradava direttamente il core all'esecuzione della routine specifica per la sorgente che aveva determinato l'interrupt.

### Pag.7

Oltre agli interrupt provenienti dalle periferiche on-chip, i µC permettono di configurare alcuni piedini per interrupt esterni. Per abilitare tutti oppure soltanto alcuni interrupt, è possibile agire su alcuni registri speciali contenuti nel µC. Impostando i singoli bit di questi registri è possibile escludere o abilitare gli eventi di interrupt provenienti dalle periferiche (come per esempio TIMER, porte RB, convertitore A/D, periferiche seriali sincrone o asincrone, USART, ecc.). Inoltre, interrogando tali registri è possibile fare il polling della situazione attuale di una periferica. Nel PIC è presente il registro INTCON (**Pag.7**).

### Pag.8

Esiste anche un registro di abilitazione degli interrupt provenienti da altre periferiche on-chip (Peripheral Interrupt Enable register, PIE). Infine, vi è il Peripheral Interrupt Register PIR (**Pag.8**), che contiene i "flag bit" degli stati delle singole periferiche on-chip che possono provocare un interrupt se abilitate dal precedente registro.

La possibilità di gestire interrupt nidificati richiede degli stack sufficientemente profondi per salvare i contenuti del program counter e degli altri registri importanti ed un core sufficientemente complesso da permettere questi "context switching", tutte le volte che venga invocato un nuovo interrupt. E' per questi motivi che solo i µC di fascia alta permettono la nidificazione di due (o poco più) livelli di interrupt, mentre in quelli di fascia bassa è compito del progettista scrivere un firmware che eviti di rispondere ad una chiamata di interrupt quando si sta già eseguendo la routine di gestione dell'interrupt precedente.

## Convertitore ADC

### Pag.9

Molti microcontrollori di fascia media hanno al loro interno un convertitore analogico-digitale, solitamente ad 8 bit, con multiplexer analogico e Sample&Hold, come mostrato a [Pag.9](#). I µC più costosi possono avere ADC interni fino a 10 bit e mux fino a 12 ingressi. ([I concetti alla base del campionamento, multiplexing analogico e conversione analogico-digitale e viceversa, verranno trattati estesamente nelle Lezioni ad essi dedicati](#)).

Mediante le linee interne `CHS0:CHS2` viene selezionato uno solo degli interruttori (pass-transistor a MOSFET), in modo che l'ingresso analogico corrispondente sia inviato all'unico ADC interno al µC per essere convertito. In taluni µC, come nel caso del PIC, la tensione di riferimento dell'ADC può essere impostata anche dall'esterno, mediante uno degli otto ingressi analogici a disposizione (il quarto dal basso, a [Pag.9](#)).

Nell'istante in cui viene specificata la porta d'ingresso (mediante `CHS0`, `CHS1` e `CHS2`), il µC provvede a chiudere il Sample&Hold. Successivamente, dopo aver atteso un tempo sufficientemente lungo affinché si sia assestata la tensione ai capi del condensatore sarà possibile iniziare la conversione (Start of Conversion, SoC) impostando ad 1 il bit `GO/DONE` del registro di configurazione dell'ADC, l'`ADCON0`). A questo punto il µC necessiterà di un certo numero di periodi di clock per terminare la conversione (uno per ogni bit del convertitore). Solo alla fine della conversione è possibile leggere il risultato dell'operazione nel registro `ADRES`. Si può chiedere, via firmware, al µC di interrogare l'ADC mediante polling del bit `GO/DONE` del registro di configurazione, in modo da verificare se il µC ha completato la conversione (`GO/DONE=0`). Ovvero può essere lo stesso ADC a scatenare un interrupt nel momento in cui finisce la conversione del dato.

Se durante la conversione non si devono eseguire altre istruzioni, si può mandare il µC in "Sleep Mode", mediante l'asserzione di un bit nel registro specifico, ossia disattivare alcune se non quasi tutte le periferiche interne. In questo modo è possibile ridurre sensibilmente il consumo di corrente e, anche, ridurre l'azione "disturbante" delle periferiche digitali sulla conversione vera e propria. In questo modo, se viene abilitata tale opzione, è l'ADC stesso al termine della conversione a forzare il "Wake-up" dell'intero µC.

### Registri di Configurazione

### Pag.10

La figura a blocchi del convertitore vista in precedenza evidenzia come i piedini del multiplexer debbano essere configurati come ingressi analogici per poter essere utilizzati a tale scopo e non come I/O digitali. Per fare ciò si deve configurare via firmware il registro `ADCON1`. L'altro registro importante è `ADCON0` (mostrato a [Pag.10](#)), che permette di impostare la frequenza di conversione (scegliendo  $T_{AD}$  come multiplo del periodo di clock del µC, oppure ricorrendo all'oscillatore RC interno), di scegliere il canale che si vuole leggere, di iniziare la conversione stessa o di spegnere l'intero convertitore ADC.

### Pag.11

#### Prestazioni

Nella descrizione delle prestazioni dell'ADC e del relativo multiplexer analogico, vengono specificati sia i parametri dinamici (timing e ritardi minimi) che quelli statici (errori differenziali, errori integrali, offset, valori di full scale range, ecc.). La figura a [Pag.11](#) mostra la tempistica seguita dal core Microchip durante la conversione ed i valori suggeriti dal Costruttore.

### Pag.12

Le prestazioni statiche sono elencate a [Pag.12](#). L'accuratezza di  $\pm 1\text{LSB}$  è garantita solo con un'alimentazione nominale di 5V, anzi addirittura di 5.12V. Questo evidenzia ancora di più in fatto che, usando una  $V_{REF}$  inferiore a  $V_{DD}$ , la precisione della conversione è destinata a degradare sensibilmente.

Inoltre, si noti che un convertitore che si possa definire di 8bit, dovrebbe avere una accuratezza migliore di  $\frac{1}{2}\text{LSB}$ , e non solo di 1LSB. Tutti questi indizi fanno ritenere che l'accuratezza del convertitore in esame è circa quella di un ADC a 7 bit, nonostante la risoluzione offerta sia quella di un 8 bit; ovvero sono disponibili 256 codifiche, ma l'ultimo bit può essere errato (se una lettura fosse 123, il valore corretto potrebbe essere indifferentemente 122, 123 o 124).

## Esempi di impiego degli I/O

### Pag.13

Si voglia gestire una tastiera di 16 tasti (ad esempio 10 numerici e 6 ausiliari) mediante il µC. Si descriveranno di seguito alcune soluzioni adottabili.

### Pag.14

### Esempio 1: Multiplexing digitale

Ovviamente, è inefficiente utilizzare 16 ingressi digitali per monitorare individualmente ciascun tasto; molto più economico risulta operare un multiplexing hardware, scandendo la tastiera per righe e colonne, come mostrato a **Pag.14**, mediante un opportuno firmware. Si collegano quindi i pulsanti in una matrice quadrata di 4 righe per 4 colonne. Le righe vengono pilotate in successione da 4 pin della PORT A, che saranno configurati come uscite digitali. Le colonne, adeguatamente messe in pull-down da quattro resistenze, verranno invece collegate a 4 bit della PORT B, impostati come ingressi digitali. In totale vengono così impiegati otto piedini anziché sedici.

Il firmware dovrà inizialmente impostare tutti gli 8 pin come ingressi e poi, ordinatamente, dovrà configurare un pin della PORT A alla volta come uscita, portandolo ad 1. In questo modo, leggendo i quattro bit della PORT B sarà possibile riconoscere al tasto premuto. Se dopo una scansione delle quattro righe si fosse rivelato più di un tasto premuto, si dovrà scartare l'acquisizione per evitare la possibile equivocazione della lettura causata dalla pressione simultanea di più tasti.

Per evitare che il  $\mu$ C debba continuamente scandire la tastiera, anche quando nessun tasto è premuto, è possibile fare in modo che appena si prema un tasto si scateni un interrupt. In questo modo il  $\mu$ C saprà di dovere effettuare la scansione per rilevare quale tasto è stato premuto; invece, in mancanza di pressioni, il  $\mu$ C può eseguire altre operazioni del programma principale. Infatti, se il tempo richiesto da ogni istruzione fosse ad esempio di 200ns, sarebbe molto dispendioso monitorare la tastiera più spesso di una volta ogni 10'000 istruzioni (ossia un polling ogni 2ms) se questa venisse per esempio utilizzata una volta al secondo, o peggio all'ora. Anche perché la pressione di un tasto ad opera di un operatore umano dura diverse decine (se non addirittura centinaia) di millisecondi.

Quindi nel programma principale (main), tra le altre istruzioni si dovranno inserire le seguenti comandi, per permettere la corretta gestione dell'interrupt, come mostrato a **Pag.14**. Nella subroutine di gestione dell'interrupt, non essendovi vettori di interrupt, le istruzioni dovranno verificare che la richiesta provenga proprio dalla tastiera.

L'intervallo di tempo compreso fra l'istante in cui viene premuto un tasto e l'istante in cui si scatena l'interrupt, può per esempio essere pari a 2 $\mu$ s. A questo si deve sommare il tempo che intercorre fra la richiesta di interrupt e l'inizio della scansione: supponendo che questo sia quantificabile in 3 $\mu$ s, il ritardo fra l'istante in cui viene premuto un tasto e l'effettiva rilevazione può valere circa 5 $\mu$ s. Quindi è da escludere che il  $\mu$ C non rilevi la pressione di un tasto, a causa del rilascio troppo veloce del medesimo, poiché l'operazione non può essere più veloce di qualche decina di millisecondi.

Vi sono poi altri problemi da analizzare. Uno di questi è legato al *bouncing* (rimbalzo) dei tasti che può avvenire a distanza di decine di millisecondi: il rischio è lo scatenamento di interrupt in successione, la conseguente esecuzione della routine e, in definitiva, la rilevazione di più pressioni dello stesso tasto. Una soluzione al problema è quella di filtrare passa-basso il segnale d'ingresso alla porta B mediante l'inserimento di quattro piccoli condensatori, con costante di tempo di qualche decina di millisecondi. A ciò si può preferire una soluzione firmware che scarti la seconda segnalazione se essa il medesimo elemento della tastiera. Solamente una scansione completa senza la rivelazione di alcun tasto premuto permette di ritenere valide pressioni successive. Infine, la sensibilità al solo "rising-edge" risolve il problema derivante dall'eventuale pressione continua del tasto.

## **Pag.15**

### Esempio 2: Acquisizione analogica

Utilizzando l'ADC è possibile ancora ricorrere al multiplexing delle righe, ma con un'unica linea di ritorno; è quindi possibile risparmiare pin di I/O. Lo schema è mostrato a **Pag.15**. Con questo tipo di implementazione viene abilitato un solo piedino all'“interrupt on rising edge”; esso viene tramutato in ingresso analogico, per misurare la tensione corrispondente al nodo. Una volta ricevuto l'interrupt si devono scandire le colonne con sufficiente lentezza, in modo da garantire la necessaria stabilizzazione dell'ADC, in alcune decine di microsecondi. Ovviamente questo sistema risulta più lento del precedente, perché viene richiesta una conversione (di durata pari a qualche decina di microsecondi).

Viene predisposta un'opportuna configurazione di resistenze: il loro valore, intorno al k $\Omega$ , è stato scelto per evitare che la corrente di leakage dell'ingresso dell'ADC determini un errore eccessivo in termini di LSB. La tensione sul pin AN2/RB0 sarà legata alla colonna alla quale appartiene il tasto premuto, come mostrato nella tabella.

Con il tasto della prima colonna premuto (si ricorda che le altre righe devono essere impostate in three-state, ovvero come ingressi) la corrente in uscita è la massima e vale  $5V/10k\Omega=0.5mA$ , valore sufficientemente piccolo da non determinare un significativo abbassamento della V<sub>OH</sub> della porta; non è quindi necessario utilizzare un buffer.

Da notare comunque che un buffer esterno sarebbe difficilmente utilizzabile volendo garantire l'abilitazione di una sola riga alla volta. Infatti, con un quad-buffer le uscite delle quattro porte assumerebbero valore 0 oppure 1 e non sarebbe previsto il three-state; la pressione contemporanea di più tasti su righe diverse provocherebbe cortocircuiti e danni.

## **Pag.16**

I problemi di questa configurazione derivano dalla tolleranza delle resistenze che impediscono di avere dei valori precisi di tensione. Nel caso di resistenze al 10%, il caso peggiore di fluttuazione si ha premendo la quarta colonna; in questo caso la tensione potrebbe assumere i valori estremi:

$$V_{\max,\min} = 5 \cdot \frac{(10K \pm 10\%)}{(10K \pm 10\%) + (3.3K + 6.7K + 20K \mp 10\%)}$$

Gli estremi di variabilità della tensione sono pertanto:  $V_{\max}=1.45V$  e  $V_{\min}=1.07V$ . Premendo la terza colonna si avrebbe:

$$V_{\max,\min} = 5 \cdot \frac{(10K \pm 10\%)}{(10K \pm 10\%) + (3.3K + 6.7K \mp 10\%)}$$

da cui si ottiene:  $V_{\max}=2.75V$  e  $V_{\min}=2.25V$ . Infine, per la seconda colonna si avrebbe:

$$V_{\max,\min} = 5 \cdot \frac{(10K \pm 10\%)}{(10K \pm 10\%) + (3.3K \mp 10\%)}$$

da cui si ottiene  $V_{\max}=3.94V$  e  $V_{\min}=3.56V$ . E' necessario fare in modo che tutti i valori compresi in questi intervalli siano fatti corrispondere al relativo tasto premuto. Ad esempio, riferendosi sempre all'esempio in esame basta scegliere le soglie dell'ADC nel seguente modo: tra  $0 \div 1V$  ossia  $00_H \div 33_H$  (nessun tasto),  $1 \div 2V$   $34_H \div 66_H$  (quarta colonna),  $2 \div 3V$   $64_H \div 99_H$  (terza colonna),  $3 \div 4V$   $9A_H \div CC_H$  (seconda colonna) ed infine  $4 \div 5V$   $CD_H \div FF_H$  (prima colonna). Non serve quindi una risoluzione ad 8 bit dell'ADC, ma è sufficiente monitorare i primi  $\log_2 5 \approx 3$  bit della codifica.

In base a quanto detto, si comprende che solo in via teorica si potrebbe aumentare il numero di colonne, al limite proprio a 256 colonne; purtroppo, l'intervallo dei valori diverebbe troppo piccolo e sarebbe soggetto a numerosi errori: tasti di colonne vicine potrebbero essere interpretati come lo stesso tasto o, addirittura, essere fraintesi. Bisogna anche tener presente che la tensione da convertire dipende direttamente dalla tensione di riga: solo teoricamente essa è di 5V mentre più realisticamente scende all'aumentare della corrente prelevata. Per esempio, la corrente prelevata quando si preme un tasto della prima colonna è 4 volte più grande di quella che si ha premendo un tasto dell'ultima.

Altro problema dell'implementazione in esame è la gestione del caso in cui vengano premuti più tasti contemporaneamente. Nel caso di tasti su colonne diverse (ma non righe), prevale il tasto della colonna più "bassa". Invece, nel caso di tasti su righe diverse può manifestarsi il problema del "tasto fantasma" (premendo tre tasti, sembra che sia premuto anche i "quarto" posto a quadrato con gli altri tre) nel circuito a multiplexing di righe e colonne degli [Esempi 1](#) e [2](#) ma non nel caso di circuito in esame, con singolo ingresso analogico.

## Pag.17

### [Esempio 3: Unica linea analogica](#)

Questo metodo è il più economico in termini di utilizzo di pin del µC. Esso utilizza ancora un pin abilitato all'"interrupt on falling-edge", che poi viene reimpostato come ingresso analogico, in modo da leggerne la tensione. Lo schema è riportato a [Pag.17](#).

I problemi di questo metodo sono la scarsa immunità a disturbi e spikes (a meno di mettere il condensatore C di filtraggio) ed il problema della pressione simultanea dei tasti.

Le resistenze devono essere scelte in modo da realizzare intervalli di codifica uniformi, ad esempio proprio come mostrato in figura. Anche in questo caso non si possono riconoscere più tasti premuti contemporaneamente, in quanto prevale quello più a sinistra. La tolleranza delle resistenze deve essere del 2%, ottenendo una distanza fra le codifiche di circa 20 codici. La corrente massima assorbita è  $5V/10k\Omega=0.5mA$ ; tale valore può essere diminuito aumentando il valore di tutte le resistenze, ma così facendo diverrebbe importante la corrente  $I_{leakage}$  dell'ADC.

L'inserimento di una capacità di filtraggio rallenta notevolmente l'aquisizione: per esempio, utilizzando  $C=47nF$ , nel caso peggiore il dato si stabilizzerebbe con la costante di tempo  $10k\Omega \cdot 47nF \approx 0.5ms$ . Quindi, per raggiungere il codice corretto entro 3LSB sarebbe necessario attendere un tempo pari a  $T=0.47ms \cdot \ln(256/3)=2ms$ . L'attesa sarebbe estremamente dispendiosa, visto che in tale tempo è possibile effettuare ben 10'000 istruzioni (supponendo  $f_{ck}=20MHz$  ed un ciclo istruzione di 200ns). E' quindi preferibile rinunciare al condensatore di filtraggio, rischiando chiamate a vuoto dovute a spikes.

- 9.1 Un grave problema di questo circuito è l'impossibilità di implementare un "interrupt on falling edge" dell'ingresso, perché la pressione dei tasti lontani non è sufficiente a scatenarlo. Si è quindi costretti ad utilizzare un polling continuo da parte del µC.

## Pag.1

*In questa lezione si introdurranno le periferiche Timer e Counter, nonchè le periferiche di comunicazione seriale presenti in un µC. Infine si presenteranno le periferiche interne per la generazione del clock e dei reset del µC.*

# Timer e Counter

## Pag.2

Sia il timer che il counter fanno capo alla stessa periferica e compiono la stessa funzione di conteggio di eventi. Tuttavia mentre il timer conta il passare del tempo, scandito dal clock del µC (o da un suo sottomultiplo), il counter conta degli impulsi che giungono ad uno dei pin d'ingresso, opportunamente abilitato. Quest'ultimo è utilizzato per contare eventi esterni al µC, provenienti da sensori di presenza, fotocellule, microswitch, prossimetri, ecc. Solitamente nei µC vi sono da uno a tre periferiche timer/counter che possono essere impiegate nei modi più svariati. Ciascuna di esse può generare un proprio interrupt al termine del conteggio (overflow). Un timer normale può essere di 8 bit con un bit aggiuntivo di overflow, eventualmente con prescaler programmabile o disattivabile durante la modalità SLEEP (nel qual caso non può agire da wake-up), ed unità di sincronizzazione. Un tipico schema a blocchi del [TIMERO](#) dei µC Microchip è mostrato a [Pag.2](#). Il registro che controlla il [TIMERO](#) nel PIC16Cxx è l'Option register. Via firmware, tramite il bit [TOCS](#) ed il mux da esso pilotato, è possibile selezionare fra modalità timer o counter. In modalità timer viene usato un sottomultiplo della frequenza di clock del µC ( $f_{osc}/4$ ), mentre in modalità counter si utilizza il piedino [TOCKI](#) per contare gli eventi esterni, nella commutazione rising-edge (se [TOSE=0](#)) oppure falling-edge (se [TOSE=1](#)), attraverso la porta EXOR.

## Pag.3

Mediante prescaler programmabili (tramite i bit [PS2:PS0](#)) è possibile dividere ulteriormente la frequenza degli impulsi di conteggio; in questo modo, sarà possibile avere un timer che conti i secondi, nonostante il clock  $f_{osc}$  lavori a megahertz, o un counter che conti eventi in ingresso a multipli di potenze di due. Il prescaler opera da divisore di frequenza ed è ottenuto mediante un contatore seguito da un multiplexer digitale in uscita, come mostrato a [Pag.3](#): con 3 bit di selezione, contenuti nel registro Option, si potranno ottenere 8 divisioni differenti, da un rapporto minimo di 1:1 ad uno massimo di 1:128.

## Pag.4

### Watchdog

Il µC, come ogni altra macchina a stati, corre il rischio di uscire dal flusso corretto di elaborazione e, al limite, di entrare in stallo a causa di un errore firmware, hardware, oppure per disturbi presenti sull'alimentazione (spike), o sui pin di I/O. Per uscire da questo stato di stallo viene fornito un circuito interno al µC chiamato Watchdog che, accorgendosi del funzionamento anomalo del µC, lo resetta evitando di lasciarlo procedere in modo incontrollato. La scelta se abilitarlo o meno viene fatta asserendo un bit in fase di programmazione del µC.

Questo dispositivo è costituito essenzialmente da un timer autonomo, che effettua continuamente un conteggio alla rovescia e, nel caso in cui raggiunga lo zero (cioè nel caso in cui si verifichi l'underflow) resetta il µC. Durante il normale funzionamento, cioè quando tutto sta procedendo correttamente, il firmware deve impedire che ciò avvenga, semplicemente reimpostando ad [FFH](#) questo timer, mediante un'istruzione apposita. Se ad un certo punto il µC andasse in stallo o iniziasse ad elaborare delle istruzioni indefinite, la mancata reimpostazione del watchdog lo porterebbe a raggiungere il fondo scala, scatenando il reset di tutta la macchina. Tipicamente il periodo del watchdog è abbastanza grande; per esempio nei PIC16Cxx è di 18ms e può raggiungere i 2 secondi nel caso in cui venga utilizzato un postscaler 1:128 (simile al prescaler visto a [Pag.3](#)).

Il watchdog può essere anche usato per risvegliare un µC dallo stato di SLEEP (wake-up programmato) in alternativa al metodo dell'interrupt. Si ricordi che lo 'SLEEP' è uno stato a basso consumo del µC, in cui vengono disabilitate e disconnesse dall'alimentazione intere periferiche interne (quali l'ADC, i vari timer, tutte quelle non utilizzate). Si supponga di voler utilizzare l'ADC interno al µC, allo scopo di acquisire segnali lentamente variabili: si può decidere di abbassare il clock a 100kHz (ogni istruzione dura 40µs) per eseguire conversioni lente e precise. In questo modo una conversione richiederà tipicamente diverse centinaia di millisecondi; volendo risparmiare la potenza (e quindi la durata della batteria) è inaccettabile implementare via firmware un loop d'attesa così lungo. Allora, dopo aver impostato la conversione è possibile mandare il µC in SLEEP e fare in modo che non sia l'ADC bensì il watchdog a scatenare l'interrupt per il risveglio ed il proseguimento delle operazioni del µC stesso.

## Periferiche di Comunicazione Seriale

### Pag.5

Le porte di comunicazione seriali sono estremamente utili in quanto permettono al µC di colloquiare e scambiare dati con altre periferiche esterne, sia presenti sulla stessa scheda (collegamenti intrascheda con EEPROM seriali, shift register, display drivers, convertitori A/D, ecc.) che su schede diverse (collegamenti interscheda). Utilizzando porte seriali invece che parallele è possibile risparmiare sul numero di pin impiegati e sui collegamenti.

Esistono sostanzialmente tre protocolli che differiscono fra loro per la modalità di trasferimento dei dati, per il numero di interconnessioni utilizzate e per la possibilità di possedere più master e di effettuare l'arbitraggio o meno del bus. Essi sono: Serial Peripheral Interface (SPI) di Motorola simile al Microwire di National Semiconductor; l'Inter-Integrated Circuit (I<sup>2</sup>C) di Philips; ed il Serial Communication Interface (SCI) più noto come USART (Universal Synchronous/Asynchronous Receiver/Transmitter).

### Pag.6

#### Porta Seriale Sincrona SPI

Si inizino a considerare le prime e più semplici: la SPI e la gemella Microwire. Entrambe hanno il vantaggio di essere molto semplici, poiché utilizzano solo tre fili per trasmettere 8 bit sincroni con un clock, ma hanno lo svantaggio di prevedere un solo master ed una linea di clock unidirezionale (dal master verso lo slave). I pin d'interesse sono: Serial Data Output (SDO); Serial Data Input (SDI); Serial Clock (SCK). Inoltre vi possono essere ulteriori collegamenti di Chip Enable (CE) nel caso siano presenti più Slave.

Il master è il µC che gestisce il clock e, in definitiva, il trasferimento dati; lo slave è la periferica esterna (al limite un altro µC, che rimarrà sempre slave). Come mostrato a **Pag.6**, entrambi sono dotati di shift-register collegati in modo da formare una catena chiusa, come se si trattasse di un unico shift-register a 16 bit condiviso tra i due integrati. E' proprio lo shift-register che permette la conversione del dato da seriale a parallelo e viceversa: da un caricamento parallelo di 8 bit, i dati escono serialmente (con in testa il MSB) e vengono riconvertiti in parallelo nel ricevitore. In pratica funziona come uno ski-lift con i comandi posti in mano all'operatore a monte, ma non a quello a valle.

Per inizializzare la porta seriale sincrona (SSP) in modalità SPI è necessario specificare se il µC svolge la funzione di master (in questo caso **SCK** è una uscita) oppure quella di slave (**SCK** diventa input), la polarità del clock (rising o falling edge), la frequenza del clock (solo per il master) ed infine la modalità di intervento dello slave (solo per lo slave). Il registro con il dato è doppio: c'è quello da trasmettere nello shift register **SSPSR** e quello in cui viene trasferito il contenuto ricevuto **SSPBUF**.

Una volta ricevuti tutti gli 8 bit di interesse (ovvero quando lo slave avrà contato gli 8 impulsi di clock), viene impostato il flag di buffer full ed il flag di interrupt (nel registro **PIR1<3>**) il dato potrà, quindi, essere prelevato dal core del µC. Ovviamente è preferibile utilizzare la tecnica degli interrupt per evitare di continuare ad effettuare il polling.

### Pag.7

Nel caso di più slave, il master ne dovrà abilitare uno alla volta, mediante linee aggiuntive di chip enable (CE), come mostrato a **Pag.7**.

Vi sono tre diverse modalità di trasmissione dei dati, nonostante il master sia unico, a seconda di chi deve leggere e/o scrivere:

- Write – il master trasmette dati utili, mentre lo slave rimanda dati inutili (dummy);
- Write and Read – sia il master che lo slave trasmettono dati utili;
- Read – il master invia dati dummy, mentre lo slave rimanda dati utili.

Risulta quindi evidente come questo sia un protocollo full-duplex, poiché i dati possono essere scritti e letti contemporaneamente.

Il master detiene costantemente il comando delle operazioni: decide l'istante in cui iniziare la trasmissione (abilitando il clock), anche quando è lo slave a dovere (non potendo "volere") trasmettere. E' dunque necessario creare un particolare protocollo firmware di interrogazione dello slave, che permetta di sapere quando i dati provenienti da quest'ultimo siano 'buoni' e quando invece 'dummy'. Inoltre, nel momento in cui si verifichi un malfunzionamento, lo slave non ha la possibilità di informare il master tramite la stessa linea seriale, ma deve attendere che quest'ultimo lo interroghi, per poi passargli un opportuno codice di emergenza, concordato a priori. Nel caso in cui lo slave si trovi in modalità SLEEP, sarà l'inizio del trasferimento dati a causare il suo 'wake-up'.

### Pag.8

#### Porta Seriale Sincrona I<sup>2</sup>C

Questo protocollo è più complicato dei precedenti, perché permette di gestire anche l'arbitraggio e la sincronizzazione del bus seriale, utilizzando soltanto due fili: il Serial Clock (SCL) ed il Serial Data (SDA). Entrambe sono linee bidirezionali poiché ciascuna periferica può svolgere la funzione di master, impadronendosi del bus ed inviando l'indirizzo della periferica con cui vuole comunicare; a questo punto, altre richieste volte ad ottenere il

controllo del bus verranno disattese. Si tratta quindi di un sistema di comunicazione seriale adatto alla connessione di periferiche “intelligenti”, in grado di gestire il traffico sul bus e riconoscere la chiamata al proprio indirizzo. Non saranno più necessarie le linee di chip-enable come per il protocollo SPI, le connessioni sono soltanto due, qualunque sia il numero delle periferiche, al contrario di quanto accade con la SPI.

Le due linee devono necessariamente essere di tipo open-collector, con resistenza di pull-up esterna, in modo che il suo dimensionamento venga effettuato in funzione del numero di periferiche (sia master che slave) collegate. In questo modo, ogni piedino afferente a ciascuna linea potrà essere o un’uscita (open drain) o un ingresso, a seconda di quale periferica ne abbia il controllo. La **Pag.8** mostra la struttura del bus seriale I<sup>2</sup>C.

I problemi risiedono nella gestione dei protocolli, infatti è necessario un hardware complesso. Un µC di fascia bassa che non abbia l’interfaccia dedicata può comunicare con questo protocollo solo attraverso un firmware creato ad hoc. Anche le periferiche che rimarranno sempre slave dovranno essere di tipo intelligente, per garantire la corretta gestione dell’indirizzamento da parte del master e la risposta al chiamante.

Lo standard I<sup>2</sup>C prevede velocità massime fino a 100kbps (400kbps per la nuova modalità fast-mode). Il massimo carico capacitivo delle linee è di 400pF e questo limita il numero massimo di periferiche collegabili. Quando non vi è trasferimento dei dati, sia la linea di clock (SCL) che quella di dato (SDA) sono portate a livello alto dalle resistenze esterne di pull-up. La richiesta d’utilizzo del bus, cioè la condizione di START si verifica quando una periferica abbassa SDA mantenendo il clock SCL alto. Successivamente, nel momento in cui il clock ‘sale e poi scende’ il dato viene memorizzato (attenzione: in questo lasso di tempo il dato non deve cambiare). Il rilascio del bus, cioè la condizione di STOP, avviene quando si alza SDA con clock alto.

Dopo aver preso il controllo del bus, il master trasmette nei successivi 10 bit l’address della periferica con cui desidera comunicare; poi rilascia il bus ed attende l’acknowledge da quest’ultima. Per inviare l’indirizzo della periferica che il master vuole contattare, esistono due formati. Il più semplice è quello a 7 bit seguiti dal bit R/W, utilizzato per stabilire se il master vuole leggere o scrivere nella periferica. Il secondo formato invece utilizza un indirizzo a 10 bit, inizialmente vengono trasmessi 5 bit (11110), poi 2 bit di indirizzo, il bit R/W ed infine, una volta ricevuto l’acknowledge, i restanti bit di indirizzo.

Tutti i dati devono essere trasmessi per byte (8 bit); non vi sono limitazioni sul numero di bytes che possono essere trasmessi per ogni trasferimento. Dopo che il master ha spedito ciascun byte, lo slave ricevente deve sempre generare un bit di acknowledge /ACK. Se ciò non avviene, il master abortisce il trasferimento e genera la condizione di STOP.

Le due modalità di trasferimento dati sono:

- 1) il master trasmette e lo slave riceve;
- 2) il master riceve e lo slave trasmette.

Se il master sta ricevendo i dati, esso stesso provvede a generare l’acknowledge per ogni byte ricevuto, tranne che per l’ultimo, per il quale non invia la conferma, in modo da segnalare allo slave l’interruzione della trasmissione. Di conseguenza, lo slave rilascia la linea SDA, permettendo al master di generare la condizione di STOP.

Se lo slave ha la necessità di ritardare la trasmissione del prossimo byte, è sufficiente che mantenga la linea SCL bassa, forzando così il master nello stato di wait. Questa tecnica può anche essere utilizzata per ogni singolo bit.

Quando più master vogliono trasferire dati contemporaneamente è necessario utilizzare un arbitraggio sulla linea SDA, durante il periodo in cui la linea SCL si mantiene alta. Il master che trasmette un high quando l’altro trasmette un low perde l’arbitraggio e, subito dopo aver terminato la lettura dei bit componenti il byte, deve commutare in slave (perché il nuovo master lo potrebbe indirizzare come slave).

## Pag.9

### Porta USART

Il protocollo SCI (Serial Communication Interface) è conosciuto anche come USART (Universal Synchronous/Asynchronous Receiver/Transmitter). Mentre gli standard SPI ed I<sup>2</sup>C sono usati soprattutto per connessioni intrascheda, lo SCI è impiegato per collegamenti interscheda. Il protocollo SCI/USART può essere configurato in full-duplex asincrono, in grado di comunicare con periferiche quali monitor e PC, oppure in half-duplex sincrono, per comunicare con ADC, DAC, EEPROM seriali, ecc. (sia come master che come slave).

Tipici esempi sono i protocolli seriali RS232, oppure lo standard industriale 485 per trasmissioni sia asincrone che sincrone. Dai livelli CMOS di 0V e +5V, lo standard RS232 impone l’impiego di +12V e -12V, mediante opportuni integrati traslatori di livello.

La frequenza del clock di trasmissione (espresso in baud, ovvero bit/sec) viene imposta dal firmware, assieme alle altre impostazioni (polarità, presenza del bit di parità, trasmissione su 7 o 8 bit, ecc.). In particolare, per stabilire il baud-rate del generatore di sincronismo, è necessario impostare un registro con un opportuno valore, in modo da dividere adeguatamente la f<sub>osc</sub> del clock del µC. A tal proposito, esistono delle tabelle riportate nei data-book dei vari µC. In modalità USART asincrono, il protocollo di trasmissione è lo standard NRZ (no return-to-zero) caratterizzato dalla struttura mostrata a **Pag.9**: 1 bit di start; 8 o 9 bit di dati; infine 1 bit di stop. In certi hardware dove la parità (vale 1 se il numero di 1 del numero da inviare è dispari) non è implementata la si può creare via firmware e memorizzare come nono bit.

## Pag.10

Il trasmettitore ed il ricevitore dell'USART implementati nei µC sono indipendenti, anche se utilizzano lo stesso formato dei dati e lo stesso baud-rate. La **Pag.10** riporta il diagramma a blocchi dell'unità di trasmissione USART della Microchip. L'elemento base è lo shift-register [TSR](#) (Transmit Shift Register) che preleva oppure inserisce il dato nel registro [TXREG](#).

## Pag.11

### Oscillatore

L'oscillatore è il circuito che genera il clock necessario al funzionamento del µC. Alcuni µC possono lavorare con frequenze molto basse, fino alla continua (in funzionamento cosiddetto one-step o DC); tuttavia di solito esiste una  $f_{osc}$  minima, al di sotto della quale non si può scendere, ed una massima, che non può essere superata pena la mancata garanzia di corretto funzionamento del µC. La frequenza minima è vincolata dalla presenza o meno di memoria dinamica all'interno del µC e, quindi, alla necessità di un refresh periodico con cadenza sufficientemente elevata. La frequenza massima di clock è invece legata alle limitazioni nella velocità di commutazione della logica interna.

All'interno del µC è già inclusa la circuiteria necessaria per realizzare la retroazione positiva necessaria a garantire l'oscillazione di un semplice circuito esterno. Tipicamente si ha la presenza di un piedino [OscIn](#) ed un piedino [OscOut](#), come mostrato a **Pag.11**. A questi si può connettere un quarzo esterno per ottenere la frequenza di oscillazione desiderata. E' sempre raccomandata la presenza di un parallel-cut, cioè di due condensatori per fissare l'esatta frequenza di oscillazione. Se si desidera un'oscillazione stabile, precisa e di elevato valore (dell'ordine dei MHz) si utilizza un quarzo; altrimenti se si vuole lavorare a frequenze più basse è sufficiente impiegare una rete RC, perdendo ovviamente in precisione e stabilità.

## Pag.12

Il vantaggio di una frequenza di clock più bassa è naturalmente un minor consumo (che sale al crescere della frequenza, in tutte le reti CMOS) ed una inferiore generazione di disturbi sia condotti che irradiati. Microchip divide l'intervallo delle frequenze di clock in tre fasce: LP (low power application, fino a 200kHz), XT (intermedio) e HS (high speed, fino a 20MHz). L'andamento del consumo al variare della frequenza e della tensione di alimentazione (quando non funzionano le periferiche interne) è mostrato a **Pag.12**.

## Pag.13

Per applicazioni in cui le temporizzazioni non siano critiche ed in cui non sia necessario raggiungere elevate velocità, è possibile utilizzare al posto del quarzo una semplice ed economica rete RC esterna. A tale scopo è necessario impostare opportunamente il µC, in modo che esso modifichi la configurazione hardware interna del pin [OscIn](#), come mostrato a **Pag.13**. Quando la tensione sale, il buffer interno commuta e cortocircuita il condensatore. Al riguardo, il Costruttore fornisce delle tabelle con i valori  $R_{ext}$  e  $C_{ext}$  ottimali in funzione della frequenza di clock desiderata.

## Pag.14

### Reset

Il reset serve per l'inizializzazione corretta della macchina sequenziale rappresentata dal core del µC. Innanzitutto, quando lo si alimenta, il µC deve iniziare dalla prima riga di codice ed eseguire in modo ordinato tutte le inizializzazioni. Le sorgenti di reset sono diverse, poiché l'impulso può essere fornito dall'esterno oppure essere generato internamente al µC stesso. Le origini principali di reset sono:

- Power-on Reset;
- Master-Clear Reset, durante la normale esecuzione, mediante il piedino esterno [/MCLR](#);
- Master-Clear Reset, durante la modalità SLEEP, sempre mediante il piedino [/MCLR](#);
- Watchdog-timer reset, al raggiungimento dell'overflow del timer di watchdog ([WDT](#));
- Brown-Out Reset, se si verificano fluttuazioni dell'alimentazione.

Per analizzare tutte le modalità di reset è utile analizzare lo schema a blocchi semplificato riportato a **Pag.14**, dedicato appunto a tutte queste funzioni. Il flip/flop finale evita che l'impulso sia troppo breve e che quindi non venga adeguatamente interpretato dal core. Saranno poi le singole unità che hanno asserito il Set del flip/flop a doverlo disattivare, agendo sul Reset. Si considerano ora in maggior dettaglio le varie tipologie di reset.

## Pag.15

### [Master Clear](#)

Questo reset può avvenire durante la normale esecuzione del firmware, oppure durante la fase di SLEEP. Al piedino dedicato a tale scopo il progettista può collegare una semplice rete RC che garantisca il reset automatico all'accensione, come indicato a **Pag.15**.

Il diodo esterno serve ad aiutare il diodo di protezione interno al pin **MCLR**, quando avviene lo spegnimento dell'alimentazione: infatti il condensatore carico si troverebbe ad una tensione più elevata dell'alimentazione del  $\mu$ C, finendo col danneggiare l'integrato; in questo modo, il condensatore è rapidamente cortocircuitato a massa mediante questo diodo aggiuntivo. La resistenza R1 ha lo scopo di limitare la corrente all'ingresso della porta del  $\mu$ C, in presenza di scariche elettrostatiche.

#### **Power-on Reset**

In aggiunta alla rete esterna appena vista, spesso i  $\mu$ C possiedono normalmente un circuito interno che genera un reset tutte le volte che viene applicata la tensione di alimentazione (power-on). Vi è un particolare blocco interno che registra il fronte di salita della  $V_{DD}$  e provvede a resettare il chip: in pratica si ha un reset tutte le volte che l'alimentazione attraversa una soglia posta a circa  $\frac{1}{2}V_{DD}$ .

#### **Timer di Power-on e di timer di Oscillator Start-up**

Una volta stabilizzata l'alimentazione, è opportuno attendere che anche il clock inizi ad oscillare e si stabilizzi alla frequenza desiderata. Per questo motivo, alcuni  $\mu$ C includono al loro interno il Power-on Reset Timer (PWRT), avente un oscillatore RC interno che, trascorso un tempo di 72ms, permette al  $\mu$ C di uscire dallo stato di reset in cui era forzato. Oppure, l'uscita dal reset può essere ulteriormente ritardata attraverso il blocco Oscillator Start-Up Timer (OST) che, solo dopo aver contato 1024 impulsi di clock provenienti dall'oscillatore esterno, da il consenso al core e rilascia il reset.

## **Pag.16**

#### **Brown-Out**

Se l'alimentazione subisse delle fluttuazioni significative (anche di circa 1V) il power-on reset non interverrebbe, nonostante il  $\mu$ C possa avere subito errori firmware (dovuti alla mancata lettura o scrittura di alcuni bit), hardware (pilotaggi o letture errate dalle periferiche esterne) o software (alcuni bit dei registri o della memoria programmi potrebbero aver subito modifiche). Si rende pertanto necessaria la presenza di un circuito di brown-out detect che interviene quando la tensione di alimentazione fluttua, di una certa quantità rispetto al valore nominale. Nei PIC, per esempio, il reset avviene quando la  $V_{DD}$  scende sotto i 3.8V per un tempo maggiore o uguale a 100 $\mu$ s; il reset rimane attivo per altri 72ms dall'ultima volta in cui l'alimentazione è risalita sopra i 4.2V, come mostra la **Pag.16**.