Cenni sul transistor bipolare a giunzione

fuso@df.unipi.it; http://www.df.unipi.it/∼fuso/dida

(Dated: version 4 - FF, 21 marzo 2015)

Quella del transistor è probabilmente una delle scoperte più importanti che la fisica ha conseguito nel secolo scorso, almeno in termini di conseguenze pratiche. Vista l'importanza dell'argomento, esistono voluminosi trattati che lo riguardano. Qui ci limitiamo a puntualizzare alcuni aspetti costruttivi e funzionali del transistor bipolare a giunzione (*BJT*), che avrete modo di approfondire e ampliare nel prosieguo della vostra carriera. In ogni caso, dato il carattere necessariamente stringato e semplificato di questa nota, siete invitati a riferirvi a uno dei tantissimi testi di elettronica che si trovano in biblioteca, o in rete, per chiarire eventuali punti oscuri.

I. INTRODUZIONE

Il transistor *BJT* rappresenta il paradigma dei dispositivi *attivi* usati in elettronica, dove l'aggettivo si riferisce alla possibilità di *controllare* qualcosa (potenziali, correnti) con qualcos'altro (altri potenziali, altre correnti) nell'ambito dello stesso componente. Allo stato attuale, a causa di limitazioni nel funzionamento e nelle tecnologie, il transistor *BJT* in quanto tale ha una diffusione piuttosto limitata. Di certo esso si incontra meno frequentemente dei transistor di tipo *MOS-FET*, che sono presenti in milioni o miliardi di unità all'interno di qualsiasi dispositivo elettronico di tipo digitale (computer, telefonini, etc.).

Tuttavia, a parte il fatto che questa tipologia di transistor è impiegata nell'esperienza pratica di laboratorio, ci sono vari ottimi motivi che spingono verso l'analisi e l'interpretazione, anche semplificata, del funzionamento del transistor *BJT*. In quanto paradigma, esso si presta infatti piuttosto bene a mostrare cosa si intende con dispositivo attivo e come un dispositivo attivo possa essere impiegato praticamente.

II. DOPPIA GIUNZIONE

Il concetto di base di un transistor *BJT* (d'ora in avanti solo transistor) è la presenza di una "doppia giunzione". Le due giunzioni, ad esempio dello stesso tipo di quelle di un diodo bipolare al silicio, sono messe in serie una dopo l'altra. Dal punto di vista schematico, possiamo immaginare di avere una successione di tre semiconduttori drogati in modo diverso: si può avere una successione di tipo *npn* oppure di tipo *pnp*, come rappresentato in Fig. 1, che mostra anche il simbolo circuitale per le due varianti. In questa nota faremo riferimento alla variante *npn*, sia perché di questo tipo è il transistor usato nell'esperienza pratica (modello 2N1711, o equivalente), sia perché di fatto la tipologia *npn* è quella di impiego più frequente e dotata in genere di caratteristiche migliori.

La presenza delle due giunzioni potrebbe far pensare a due diodi collegati tra di loro in un montaggio "back-to-back", cioè anodo-anodo o catodo-catodo, ma questa descrizione non è sufficiente per spiegare l'effetto di "trans-resistenza" (trans-resistor, da cui il nome di transistor) specifico del dispositivo.

Prima di esaminare queste caratteristiche specifiche, soffermiamoci su alcune considerazioni molto generali:

- se abbiamo due giunzioni è evidente che possiamo ipotizzare di polarizzarle direttamente o inversamente l'una indipendentemente dall'altra. Nel seguito vedremo un po' meglio cosa si verifica nelle varie situazioni. Per il momento, possiamo dare un nome alle tre possibilità di interesse pratico: si dice che il transistor è *in interdizione* quando le due giunzioni sono entrambe polarizzate inversamente, *in saturazione* quando tutte e due sono polarizzate direttamente, *in regime attivo* quando una è polarizzata inversamente e l'altra direttamente (vedremo in seguito quale giunzione si trova generalmente in una condizione e quale nell'altra). Quest'ultima configurazione è quella di maggiore interesse per l'esperienza pratica.

- Il transistor è un componente, o dispositivo, con tre elettrodi (tre filini che portano o prendono corrente) che hanno dei nomi caratteristici di origine "storica": *emettitore (E), base (B), collettore (C)*, indicati anche in Fig. 1.

- I tre elettrodi fanno subito pensare a un dispositivo "attivo": in sostanza, uno dei fili serve a controllare quello che succede tra gli altri due fili. In realtà, poiché i segnali devono essere riferiti a un valore comune (la linea di massa, in genere), il dispositivo può essere classificato come un "quadripolo", seguendo una definizione in uso in elettronica per indicare, ancora una volta, che ci sono due "segnali", di cui uno controlla l'altro.

- La considerazione appena fatta sulla necessità di riferire i potenziali a un valore comune fa anche capire che, nell'uso pratico, uno dei tre elettrodi deve essere messo in comune fra "ingresso" e "uscita" (ovvero fra "controllante" e "controllato"). Infatti è possibile montare un transistor nelle configurazioni cosiddette a *emettitore comune, base comune, collettore comune*. Qui non esamineremo l'intera casistica, argomento di per sé interessante, dato che le

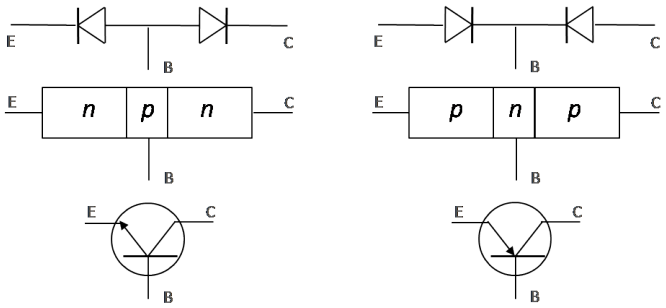

Figura 1. Rappresentazione schematica della doppia giunzione che si trova in un transistor *BJT* di tipo *npn* o *pnp* e simboli circuitali (le lettere E, B, C si riferiscono ai tre elettrodi del dispositivo).

diverse configurazioni comportano diverse caratteristiche di impiego, ma troppo avanzato per i nostri scopi. Infatti ci limiteremo a esaminare la configurazione a emettitore comune, anche se alcune descrizioni iniziali faranno riferimento a quella a base comune.

- Infine c'è un'ulteriore conseguenza del fatto che il transistor è un componente con tre elettrodi, dunque diverso dagli altri incontrati in precedenza. In quei componenti (ad esempio resistori, ma anche diodi) era possibile ipotizzare la presenza di semplici relazioni in grado di legare le due grandezze di interesse, per esempio intensità di corrente attraverso il dispositivo e differenza di potenziale applicata. Qui, invece, sarà necessario trovare delle forme più complicate per rappresentare o predire il comportamento, come ad esempio i grafici delle curve (notate il plurale) di risposta.

III. EFFETTO TRANSISTOR

Il semplice disegno rappresentato in Fig. 1 non tiene conto dell'effettiva realtà costruttiva dei transistor. Infatti:

- le tre regioni di materiale con diverso drogaggio non hanno le stesse dimensioni: la base (la regione drogata p in un transistor *npn*) può essere considerata molto sottile, sicuramente molto più della regione corrispondente al collettore (spesso lo spessore in questione è sub-micrometrico);

- il drogaggio delle tre regioni non è lo stesso (in termini di densità di drogante: è ovvio che nella regione n il drogaggio è con donori e nella p è con accettori): in particolare l'emettitore è molto più drogato della base;

- la geometria rappresentata in Fig. 1, che è sicuramente molto poco probabile dal punto di vista tecnologico (immaginate di inserire un numero gigantesco di bacchette di quel tipo all'interno di un chip...), non rappresenta affatto la realtà costruttiva. Un esempio più ragionevole è mostrato in Fig. 2(a), dove si vede uno schema compatibile con la tecnologia *planare* in uso nella microelettronica e da cui si può capire che, di fatto, l'emettitore è circondato dal materiale della base, a sua volta circondata dal materiale del collettore.

Queste caratteristiche costruttive determinano una forte interazione tra la corrente che interessa le due giunzioni. Per capire quali siano le conseguenze di questa interazione facciamo riferimento allo schema di Fig. 2(b), dove si vedono due distinti generatori di differenza di potenziale continua collegati in modo da polarizzare *direttamente la giunzione EB* attraverso la d.d.p. V_{BE} e *inversamente la giunzione CB* attraverso la d.d.p. V_{CB}. Questo si capisce facilmente ricordando che polarizzare direttamente una giunzione *np* richiede di collegare il polo negativo del generatore alla regione n e quello positivo alla p (e, come sapete, facendo in modo che tale differenza di potenziale sia superiore alla tensione di soglia della giunzione, V_{thr}, cosa che diamo qui per scontata), e viceversa. Poiché supponiamo di avere a che fare con un transistor *npn*, la disposizione delle polarità per ottenere lo scopo desiderato è quella che si vede in figura.

In queste condizioni c'è un flusso di elettroni che il generatore V_{BE} invia alla giunzione emettitore-base. Se si trattasse di una giunzione "isolata", cioè di un semplice diodo, questo flusso di elettroni che entra (viene "emesso") dall'emettitore si ritroverebbe a uscire dalla base. Invece l'effetto transistor fa sì che una gran parte di questo flusso scavalchi la giunzione tra base e collettore, e quindi fuoriesca dall'elettrodo collegato al collettore (sia "raccolto" dal collettore).

Questo si verifica fondamentalmente per tre motivi principali:

1. il drogaggio p della base ha una densità molto minore del drogaggio n dell'emettitore: il grande (denso) flusso di elettroni che si muove nell'emettitore e supera la giunzione EB non può ricombinarsi in modo "completo" nella regione di base, dove non trova una sufficiente densità di lacune.

2. A causa della polarizzazione inversa, nella regione del collettore esiste un campo elettrico che accelera gli elettroni che superano la barriera BC in modo che essi possano percorrere la regione del collettore e di qui fuoriuscire.

3. Forma e spessore del dispositivo aiutano a rendere particolarmente efficace l'effetto: infatti lo spessore ridotto della regione di base permette agli elettroni di essere facilmente catturati dal campo presente nel collettore e la geometria del dispositivo [vedi Fig. 2(a)] rende il processo di superamento della barriera BC e la raccolta degli elettroni nel collet-

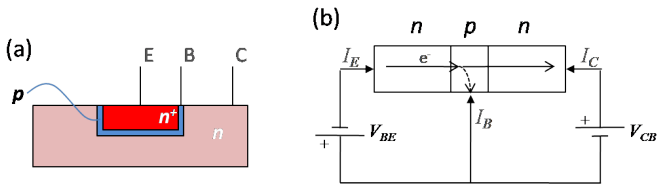

Figura 2. Geometria costruttiva tipica per un transistor *BJT npn* (a); rappresentazione schematica del passaggio di elettroni dall'emettitore al collettore responsabile per l'effetto transistor. Notate che le frecce riportate nella figura per le correnti rappresentano il segno convenzionalmente positivo (il verso effettivo delle correnti potrebbe essere diverso, secondo quanto discusso nel testo).

tore quasi isotropo (non esistono direzioni in cui il processo sia meno efficace).

A. Correnti nel transistor

Cerchiamo ora di quantificare l'effetto transistor. Allo scopo definiamo I_E, I_B, I_C le intensità di corrente che scorrono rispettivamente verso l'emettitore, la base e il collettore. Per evitare problemi con i segni, seguiamo la convenzione per la quale le correnti sono positive se entrano nel dispositivo [si vedano le freccine riportate in Fig. 2(b)]. Notate che, secondo questa convenzione, le correnti di emettitore e collettore hanno segni opposti: infatti nelle condizioni di Fig. 2(b) si ha $I_E < 0$ (entrano elettroni nell'emettitore, corrispondenti a una corrente che esce dall'emettitore stesso) e $I_C > 0$ (escono elettroni dal collettore, e quindi la corrente entra nel dispositivo).

L'effetto transistor è riassunto in questa semplice equazione:

$$I_C = -\alpha_F I_E \ , \tag{1}$$

dove il coefficiente α_F ha un valore prossimo all'unità: a seconda del tipo di transistor, esso è infatti tipicamente compreso tra 0.95 e 0.999. Questo significa che quasi la totalità della corrente (fatta di elettroni) che entra e si muove nell'emettitore preferisce farsi raccogliere, e quindi uscire, dal collettore piuttosto che ricombinare nella base e dare origine a una corrente di base (fatta di lacune). Dunque il funzionamento della giunzione emettitore-base ha caratteristiche molto diverse rispetto a quelle di un diodo ordinario.

Come già affermato (dovrebbe essere chiaro tenendo conto della descrizione costruttiva data sopra), la serie delle due giunzioni *npn* che formano il transistor *non è affatto simmetrica*: in altre parole, emettitore e collettore non possono scambiarsi di ruolo. Infatti se si prova a "invertire" il sistema, cioè a polarizzare inversamente la giunzione EB e direttamente quella CB [si tratta, in sostanza, di scambiare i segni dei due generatori di d.d.p. che compaiono in Fig. 2(b)], si ottiene che l'effetto transistor, seppur ancora presente, è nettamente meno importante. In particolare in queste condizioni si ha $I_E = -\alpha_R I_C$, con α_R tipicamente dell'ordine di 0.5, o inferiore. Notate il pedice "R" che sta per "reverse" così come il pedice "F" usato prima stava per "forward". Dunque è ancora possibile avere trasferimento di corrente attraverso la base, ma questo trasferimento è nettamente meno efficiente e il transistor non è in grado di operare efficacemente come dispositivo attivo. Il motivo principale della asimmetria è nel drogaggio della regione di emettitore, che è molto maggiore anche di quello della regione di collettore [il drogaggio n ad alta densità è indicato con n^+ in Fig. 2(b)].

B. Amplificazione di corrente

Torniamo alla configurazione di Fig.2(b). Le tre correnti che entrano nel dispositivo attraverso emettitore, base e collettore devono dare somma nulla. Infatti all'interno del transistor non esiste alcuna sorgente, o pozzo, di cariche. Quindi il transistor può essere considerato come un nodo al quale applicare una delle cosiddette leggi di Kirchoff. Si ha allora

$$I_E + I_B + I_C = 0 \ . \tag{2}$$

Utilizzando la relazione di Eq. 1 e rimaneggiando si ottiene facilmente

$$I_B = -(I_C + I_E) = (\frac{1}{\alpha_F} - 1) I_C = \frac{1-\alpha_F}{\alpha_F} I_C \ , \tag{3}$$

ovvero

$$I_C = \frac{\alpha_F}{1-\alpha_F} I_B = \beta_F I_B \ . \tag{4}$$

Questa equazione dimostra due aspetti fondamentali:

1. nelle condizioni che stiamo considerando, cioè quando il transistor si trova ad operare in regime attivo (polarizzazione diretta per la giunzione EB, inversa per la CB), la corrente di collettore dipende linearmente dalla corrente di base e, in prima approssimazione, *solamente* da questa.

2. Poichè $\alpha_F \sim 1$, il coefficiente β_F, talvolta indicato come h_{FE} e noto come *guadagno in corrente continua* del transistor, assume valori molto grandi, tipicamente compresi fra 50 e 1000: *la corrente di collettore è quindi amplificata rispetto a quella di base.*

In questa affermazione c'è molto del carattere "attivo" del dispositivo transistor: la corrente di base controlla la corrente di collettore, cioè modificando l'una si modifica l'altra. In condizioni opportune questo tipo di legame può essere sfruttato per ottenere un'amplificazione, cioè un guadagno in termini di correnti.

È bene osservare che il controllo che è in grado di operare il transistor *BJT* richiede di manipolare delle correnti. Come sapete, far passare delle correnti all'interno

di un dispositivo dà inevitabilmente luogo a dissipazione di potenza (la resistenza non è mai nulla), e la dissipazione è sempre mal vista, sia perché comporta dispendio di energia, sia perché provoca, in un modo o nell'altro, surriscaldamento. Se un chip di quelli che sono all'interno di un qualsiasi dispositivo elettronico attuale, che contengono miliardi di dispositivi di "controllo" più o meno indipendenti tra loro, fosse realizzato con la tecnologia *BJT* occorrerebbero delle ventole gigantesche e ogni ora di facebook costerebbe uno stonfo in bolletta dell'ENEL... Questo è uno dei principali motivi per cui la tecnologia *BJT* è stata abbandonata a favore di quella a effetto di campo (*MOS-FET*), che implica correnti di controllo trascurabili, nelle applicazioni su larga scala.

IV. EQUAZIONI DI EBERS-MOLL

La semplicissima Eq. 4 riporta tutto quello che è necessario sapere quando si interpreta la maggior parte delle applicazioni di un transistor. Tuttavia vale la pena di soffermarsi su un modello che è in grado di chiarire un po' meglio alcuni aspetti del funzionamento. Tale modello, detto di Ebers-Moll, fa uso di equazioni che sono formalmente simili a quelle del modello di Shockley per le giunzioni bipolari in semiconduttori.

Se non ci fosse l'effetto transistor, cioè se il nostro dispositivo fosse completamente interpretabile come una coppia di diodi "back to back", allora le correnti di emettitore e collettore andrebbero scritte come

$$I_E = -I_{0E}(\exp(\frac{V_{BE}}{\eta V_T}) - 1) \tag{5}$$

$$I_C = -I_{0C}(\exp(\frac{V_{BC}}{\eta V_T}) - 1)\ , \tag{6}$$

dove I_{0E} e I_{0C} sono le correnti di saturazione inversa per le due giunzioni EB e BC, $V_T = k_B T/e$ è la differenza di potenziale termica (vale circa 26 mV a temperatura ambiente) ed η è il coefficiente dovuto alle caratteristiche specifiche delle giunzioni (vale circa 2 se si tratta di giunzioni fatte di silicio). Riconoscete facilmente la forma delle equazioni di Shockley, qui adattate alla presenza di due distinte giunzioni. Notate anche che il segno negativo usato è dovuto alla convenzione che abbiamo dichiarato di seguire prima, per la quale le correnti sono positive se entrano nel dispositivo. Infatti tutte e due le giunzioni considerate, se polarizzate direttamente (V_{BE} e V_{BC} entrambi positivi e maggiori di V_{thr}), sostengono un flusso di elettroni entrante, che corrisponde a una corrente uscente, da cui il segno negativo.

Per l'effetto transistor queste due equazioni non sono sufficienti a descrivere il comportamento del dispositivo. Infatti è necessario aggiungere due termini, ottenendo le equazioni di Ebers-Moll:

$$I_E = -I_{0E}(\exp(\frac{V_{BE}}{\eta V_T}) - 1) + \alpha_R I_{0C}(\exp(\frac{V_{BC}}{\eta V_T}) - 1) \tag{7}$$

$$I_C = -I_{0C}(\exp(\frac{V_{BC}}{\eta V_T}) - 1) + \alpha_F I_{0E}(\exp(\frac{V_{BE}}{\eta V_T}) - 1) \tag{8}$$

dove i coefficienti α_F e α_R sono quelli, già introdotti, che mescolano la corrente di collettore con quella di emettitore, e viceversa, a causa dell'effetto transistor. Notate che, sulla base delle considerazioni svolte prima, i segni sono corretti.

Quando il transistor lavora in regime attivo, cioè quando la giunzione EB è polarizzata direttamente e quella CB inversamente, alcuni dei termini appena scritti diventano trascurabili. Infatti in queste condizioni $V_{BE} > V_{thr}$ (tecnicamente vuol dire maggiore di 0.6-0.7 V), quindi i termini che contengono questa tensione sopravvivono (e pure bene, tanto da prevalere sul -1). Invece si ha $V_{BC} < V_{thr}$, per cui i termini corrispondenti possono essere trascurati. Per meglio dire, essi danno luogo alla debole corrente di saturazione inversa, che, nelle giunzioni di silicio, vale nA, o anche meno.

In altre parole, in regime attivo si ha

$$I_E \simeq -I_{0E}\exp(\frac{V_{BE}}{\eta V_T}) \tag{9}$$

$$I_C \simeq \alpha_F I_{0E}\exp(\frac{V_{BE}}{\eta V_T}) = -\alpha_F I_E\ , \tag{10}$$

che equivale sostanzialmente a quanto riportato nell'Eq. 1. Quindi questo modello conduce ancora, almeno in prima approssimazione, a una corrente di collettore che è proporzionale a quella di emettitore, con un fattore di proporzionalità prossimo all'unità, quando il transistor si trova a operare in regime attivo.

Le equazioni di Ebers-Moll permettono poi di capire un po' meglio cosa succede quando il transistor è in interdizione (tutte e due le giunzioni sono polarizzate inversamente) o in saturazione (tutte e due polarizzate direttamente).

In interdizione tutti i termini con l'esponenziale sono trascurabili rispetto al -1; quindi le correnti di emettitore e collettore corrispondono a somme di quelle di saturazione inversa, che sono generalmente così piccole da poter essere approssimate con lo zero. Allora *in interdizione non ci sono correnti* di emettitore e collettore (ovvero, esse sono trascurabili).

In saturazione tutti i termini con l'esponenziale prevalgono sul -1. Dunque in questo caso *circolano correnti sia per la base che per il collettore*, e la corrente di collettore è più intensa di un fattore β_F rispetto a quella di base. È allora evidente che la corrente di base I_B è in grado di controllare la corrente di collettore I_C, cioè agendo su I_B si può realizzare uno switch che commuta il valore di I_C. Questa è la modalità con cui normalmente si usano i transistor nei circuiti digitali, dove switch acceso o spento significa in genere che viene realizzato uno stato logico "1" o "0" (o viceversa). Osservate che, quando un transistor BJT viene usato come switch, esso non lavora più nel regime lineare, ma, appunto, in saturazione. Dunque il controllo che la corrente di base esercita su quella di collettore non implica più una proporzionalità diretta fra queste due grandezze. Anzi, la corrente di base può variare (entro certi limiti) e la corrente di collettore restare pressoché inalterata al suo valore di saturazione (il

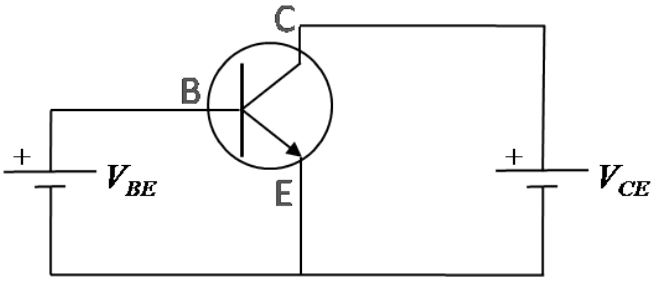

Figura 3. Schema circuitale della configurazione a emettitore comune con polarizzazione diretta della giunzione base-emettitore e inversa di quella collettore-base.

massimo possibile per la configurazione circuitale scelta). Notate anche che, se $V_{BE} \approx V_{BC}$, i due termini della corrente di collettore nell'Eq. 7 tendono a elidersi, essendo praticamente uguali e di segno opposto.

V. CONFIGURAZIONE A EMETTITORE COMUNE E CURVE CARATTERISTICHE

Come già discusso, nell'impiego pratico di un transistor occorre riferire i potenziali a uno degli elettrodi. La Fig. 2(b) e gran parte dei discorsi fatti finora hanno avuto a che fare, almeno implicitamente, alla configurazione in cui l'elettrodo di base è quello di riferimento, cioè alla configurazione *a base comune*. Questo ci ha consentito di descrivere l'effetto transistor e di ottenere il guadagno in corrente β_F, che, tuttavia, sono caratteristiche del dispositivo che permangono anche quando esso viene montato in una diversa configurazione.

Quella di maggiore interesse, sia pratico che concettuale, è probabilmente la configurazione *a emettitore comune* rappresentata schematicamente in Fig. 3. Poiché il riferimento è la linea dell'emettitore, le tensioni rilevanti sono qui V_{BE} e V_{CE}.

Come già anticipato, la complessità del dispositivo e del suo comportamento rendono non immediato individuare delle rappresentazioni grafiche (curve caratteristiche) che permettano di capire al primo colpo il comportamento di un transistor. Infatti le grandezze che entrano in gioco sono tante, almeno le tre tensioni V_{BE}, V_{BC}, V_{CE} e le tre correnti I_E, I_B, I_C. I legami fra queste grandezze non sono sempre ovvi.

Una delle curve caratteristiche, quella che generalmente viene chiamata *caratteristica di trasferimento di ingresso*, riguarda l'andamento della corrente di base I_B in funzione della tensione V_{BE}. Questa curva possiamo facilmente indovinarla: infatti essa, riguardando sostanzialmente la sola giunzione BE, avrà l'andamento tipico alla Shockley, cioè sarà di tipo esponenziale, come rappresentato in Fig. 4(a): sotto soglia non si avrà praticamente passaggio di corrente (a parte l'effetto della debole corrente di saturazione inversa), sopra soglia si avrà una crescita esponenziale della corrente in funzione della tensione come in un diodo, con l'unica, ma importante, differenza che i valori (la scala) della corrente saranno qui ben minori che nel caso del diodo (μA invece di mA) a causa del fatto che il flusso di elettroni viene raccolto dal collettore invece di fluire nell'elettrodo di base. È anche chiaro che, superata la soglia, V_{BE} tenderà a rimanere costante a prescindere dal valore di I_B, o, se preferite, I_B potrà aumentare anche di parecchio senza apprezzabili aumenti di V_{BE}. Questo significa andamento esponenziale.

Un altro tipo di rappresentazione che è abbastanza frequentemente usata, almeno per transistor montati a emettitore comune, è quella che riporta I_C in funzione di V_{CE}: fate bene attenzione al fatto che stavolta la differenza di potenziale non è riferita a una giunzione, ma, se volete, alla serie di due giunzioni. Inoltre notate che questo tipo di rappresentazione non è univoco, nel senso che la dipendenza che si vuole rappresentare non ha un solo andamento. Essa infatti dipende da altri parametri, in particolare dalla corrente di base I_B. Dunque si ottiene una *famiglia di curve* in cui si riporta l'andamento di I_C in funzione di V_{CE} e ogni curva si riferisce a un dato valore di I_B [vedi Fig. 4(b)]. In genere ci si riferisce a questi grafici come quelli delle *caratteristiche di trasferimento di uscita*.

Vediamo di indovinare anche l'andamento di queste curve. Sappiamo già che nel regime attivo I_C dipende solamente da I_B, almeno in prima approssimazione, e che questa dipendenza è lineare e avviene attraverso un fattore di guadagno, β_F, che è dell'ordine delle centinaia (valore tipico, ovviamente cambia in funzione del modello di transistor e anche delle specifiche peculiarità di fabbricazione). Allora nel regime attivo queste curve devono essere tante linee orizzontali (non c'è dipendenza da V_{CE}, cioè dalla grandezza graficata lungo l'asse orizzontale) parallele fra loro e il rapporto tra il valore di I_B corrispondente a una certa curva e quello di I_C riportato sull'asse verticale deve essere dell'ordine di grandezza di β_F.

A. Effetto Early

Nella realtà, come si osserva anche in Fig. 4(b), c'è una *piccola* dipendenza residua di I_C da V_{CE} che fa sì che i tratti orizzontali siano inclinati a formare delle rette con un piccolo coefficiente angolare positivo, almeno nell'intervallo di parametri di interesse pratico. Spesso si fa risalire questa circostanza al cosiddetto *effetto Early*. In sostanza, aumentando V_{CE} i campi elettrici all'interno del transistor, cioè nella regione delle giunzioni, vengono modificati e in particolare viene aumentato lo spessore della giunzione base-collettore, che è polarizzata inversamente. In queste condizioni viene aumentato il coefficiente α_F, e quindi il guadagno β_F, cioè il flusso di elettroni viene raccolto più efficacemente dal collettore. La corrente di collettore è allora approssimata dalla

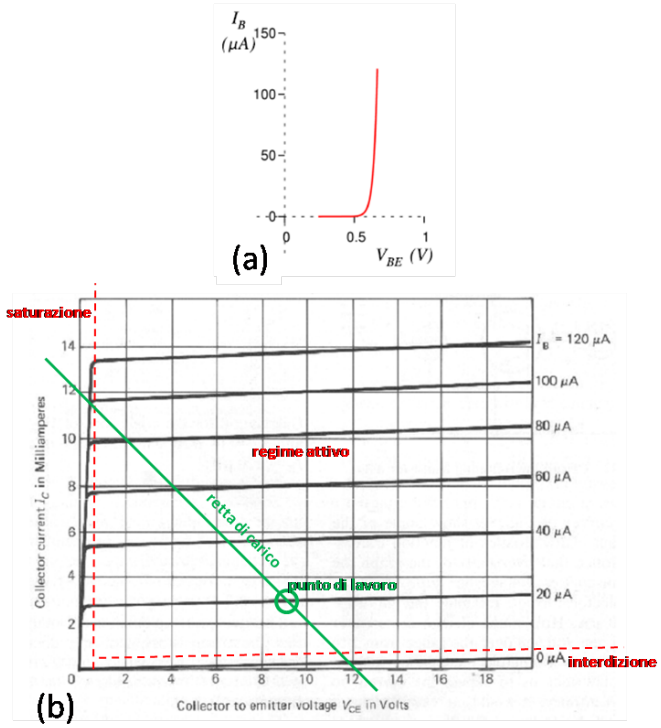

Figura 4. Curva caratteristica di trasferimento diretto (a) e di uscita (b) per un transistor *npn*. Su queste ultime sono indicate grossolanamente le zone corrispondenti ai vari regimi di funzionamento del transistor e sono sovrapposti retta di carico e punto di lavoro, individuati supponendo come esempio $V_0 = 12$ V,$R_C = 1$ kohm, $I_B = 10\mu$A.

funzione

$$I_C \sim \beta_F I_B (1 + \frac{V_{CE}}{V_{Early}}) \, , \tag{11}$$

dove il *potenziale di Early* V_{Early} dipende dalla costruzione del transistor (vale tipicamente un centinaio di V, per cui l'effetto di cui si sta parlando è poco pronunciato).

B. Retta di carico e punto di lavoro

Sulla base di quanto abbiamo stabilito, le condizioni di lavoro del transistor dipendono dalle tensioni applicate alle due giunzioni, le quali determinano anche le correnti che scorrono nel dispositivo e, in definitiva, il *punto di lavoro* del transistor. Normalmente il regime operativo del transistor viene realizzato usando un solo generatore di differenza di potenziale continua, invece dei due generatori indipendenti che abbiamo ipotizzato finora (vedi Fig. 3, ma anche Fig. 2).

Uno schema possibile, super-semplificato (e per questo, in realtà, poco comune, discuteremo poi le ragioni per cui esso non è adeguato) è quello di Fig. 5: un opportuno collegamento di resistori consente di ottenere la corretta polarizzazione della giunzione base-emettitore, diretta, e della giunzione collettore-base, inversa. La corrente I_B

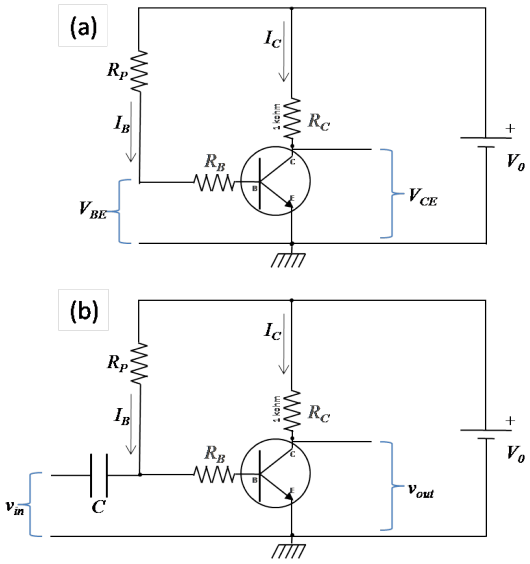

Figura 5. Esempio di semplice circuito di polarizzazione del transistor a emettitore comune.

scorre attraverso la serie dei resistori R_P e R_B, la corrente I_C scorre attraverso il resistore R_C. Dimensionando opportunamente R_P si varia la d.d.p. V_{BE} e dunque si determina l'eventuale funzionamento del transistor in regime attivo.

Per determinare il punto di lavoro del transistor occorre tenere conto dei valori di V_0 e di R_C. Esaminiamo la "maglia del collettore", costituita dal generatore V_0 (supposto ideale), dalla resistenza R_C e dal transistor. Notiamo che in questa maglia passa la maggior parte della corrente che viene fornita dal generatore: infatti nella "maglia di base" scorre normalmente una corrente molto bassa (β_F-volte minore di quella che scorre nel collettore). L'equazione della maglia recita

$$V_0 = V_{CE} + R_C I_C \, , \tag{12}$$

che, sul piano V_{CE}, I_C, dà luogo a una retta, che si chiama *retta di carico*, la quale vincola I_C a V_{CE}. Le intercette di questa retta con gli assi sono ovviamente V_0 e V_0/R_C, per cui la retta può essere disegnata in modo immediato conoscendo i valori rilevanti.

Con denominazioni differenti e una diversa forma della curva caratteristica, la retta di carico e il ruolo che essa ha per determinare in via grafica il punto di lavoro erano stati già esaminati nel caso del diodo, dove la curva caratteristica era solo una e rappresentava l'equazione di Shockley. Qui, invece, le curva caratteristiche I_C vs V_{CE} sono infinite (una famiglia di curve), dipendendo dal valore di I_B. Il punto di lavoro del transistor sarà allora determinato dall'intersezione fra la retta di carico e la curva caratteristica che corrisponde al valore di I_B che di fatto entra nella base nelle condizioni specifiche di funzionamento del transistor.

C. Interdizione e saturazione

Vediamo ora di individuare nei grafici di Fig. 4 i regimi di interdizione e di saturazione. In interdizione la giunzione BE è polarizzata inversamente, cioè $V_{BE} < V_{thr}$, per cui $I_B \approx 0$. Dunque il regime di interdizione corrisponde alle curve con $I_B \sim 0$ e in queste condizioni, giustamente, $I_C \approx 0$ a prescindere dal valore di V_{CE}.

A saturazione entrambe le giunzioni sono polarizzate direttamente. Essendo il transistor considerato di tipo *npn*, affinché la giunzione collettore-base sia polarizzata direttamente occorre che il potenziale al collettore sia minore di quello alla base (entrambi i potenziali sono riferiti alla stessa linea di massa, a cui è collegato l'emettitore). D'altra parte il potenziale di base deve essere superiore, o dell'ordine, di V_{thr} se si vuole che anche la giunzione emettitore-base sia polarizzata direttamente. Quindi in queste condizioni V_{CE} deve essere piccola; tenendo conto che nella pratica polarizzare direttamente la giunzione BE significa porre $V_{BE} \gtrapprox V_{thr}$ [vedi anche Fig. 4(a), dove si intuisce come V_{BE} vari di poco al di sopra di V_{thr} per "ragionevoli" variazioni di I_B], occorre $V_{CE} < V_{thr}$. Nel grafico di Fig. 4(b) il regime di saturazione corrisponde alla parte "di sinistra", dove V_{CE} è minore di 0.6-0.7 V.

Può essere utile a questo punto mettere in chiaro una precisazione che, in realtà, è stata data per sottointesa visto il carattere di ovvietà. È evidente, credo per tutti, che le nozioni di saturazione e interdizione devono essere prese con una certa "flessibilità". Infatti esse hanno a che fare con le condizioni di polarizzazione, diretta o inversa, delle giunzioni. Dal punto di vista modellistico è abbastanza facile definire la polarizzazione di una giunzione come diretta o inversa. Dal punto di vista pratico le situazioni sono più confuse. Infatti, se è ragionevole affermare che una giunzione è polarizzata direttamente quando la differenza di potenziale applicata ha il segno "giusto" e supera il valore di soglia, il valore di soglia non è ben definito. Di conseguenza, spesso si assume che il regime di operazione di un transistor sia in saturazione o in interdizione quando esso si avvicina alla saturazione o all'interdizione, cioè quando il regime è *all'incirca* quello di saturazione o interdizione. Ricordatevene quando sarete di fronte a situazioni pratiche.

VI. COMPORTAMENTO A PICCOLI SEGNALI

Come sottolineato in precedenza, far passare le condizioni di funzionamento un transistor *BJT* da interdizione e saturazione può essere utile per realizzare degli switch (per esempio, la corrente di base controlla il passaggio della corrente di collettore da zero a un certo valore). Tuttavia è di maggiore interesse pratico e concettuale esaminare il comportamento del transistor come *amplificatore*, che si realizza operando nel *regime attivo*, cioè lontano da interdizione e saturazione.

Prestate la massima attenzione a quanto abbiamo appena affermato: per usare il transistor come amplificatore ci si deve preoccupare di fornire le giuste tensioni di *polarizzazione* al dispositivo. In altre parole, il dispositivo compie la sua funzione attiva solo se è "correttamente alimentato". Quindi ogni circuito amplificatore dovrà provvedere delle *tensioni continue* che realizzino le polarizzazioni richieste. Focalizzando la nostra attenzione sulla configurazione a emettitore comune, questo vuol dire che dovremo in ogni caso preoccuparci di avere le tensioni V_{BE} e V_{CE} richieste per accedere al regime attivo e di conseguenza avere anche le correnti I_C e I_B, sempre *continue*, che scorrono nel transistor.

A questo punto dovrebbero risultare chiari un paio di punti che sono tipici quando si impiega un amplificatore:

- quello che normalmente si vuole amplificare non è la tensione, o corrente, continua di polarizzazione, ma si vuole invece che l'amplificazione avvenga a carico di *segnali alternati* eventualmente sovrapposti a quelli di polarizzazione;
- di conseguenza, in un circuito amplificatore a transistor è sempre presente una parte che serve per creare le giuste tensioni (e correnti) di polarizzazione, cioè un *circuito di polarizzazione*;
- infine, tenendo conto del fatto che segnali oscillanti da amplificare e tensioni di polarizzazione sono, almeno in parte, sovrapposti, e che è necessario che il transistor operi sempre nel regime attivo, è ovvio che esistono dei limiti nell'*ampiezza* del segnale che deve essere amplificato, che cioè va all'ingresso dell'amplificatore: un'ampiezza eccessiva potrebbe infatti portare il transistor fuori dal regime attivo. Di conseguenza si parla spesso di amplificazione di *piccoli* segnali;
- infine, c'è una conseguenza "tipografica" del fatto che ci si riferisce a piccoli segnali: le grandezze che caratterizzano il comportamento in queste condizioni sono in genere scritte in *caratteri minuscoli*, come già visto per la determinazione della resistenza dinamica del diodo.

La configurazione a emettitore comune si presta ad amplificare piccoli segnali oscillanti che vengono inviati alla base. Quindi la base funge da *ingresso* per tali segnali (potenziali riferiti alla linea dell'emettitore) e l'*uscita* si ritrova al collettore (anche qui potenziali riferiti alla linea dell'emettitore, che pertanto viene normalmente collegato a terra). La risposta ai piccoli segnali oscillanti può essere diversa rispetto al comportamento in continua trattato finora. La caratteristica di amplificazione in corrente, che si indica con β_f, o anche con h_{fe} (notate i caratteri minuscoli dei pedici), è definita come segue:

$$\beta_f = \frac{\partial i_c}{\partial i_b} \, |_{V_{CE}} \, , \tag{13}$$

dove i_c e i_b sono le correnti (oscillanti) che scorrono in collettore e base a causa della presenza del piccolo segnale

oscillante in ingresso, e il simbolo $|_{V_{CE}}$ indica che il valore considerato si riferisce a un determinato valore della tensione V_{CE} (continua), cioè a determinate condizioni di polarizzazione. Tipicamente, anche β_f vale alcune centinaia, cioè è dello stesso ordine di grandezza di β_F, almeno nei casi di interesse pratico.

Misurare β_f è generalmente difficile, poiché le intensità di corrente alternata sono difficilmente leggibili (l'oscilloscopio vede differenze di potenziale, non correnti). È possibile stimare la grandezza facendo riferimento a misure in continua: $\beta_f \approx \Delta I_C/\Delta I_B$, dove i simboli Δ indicano differenze tra valori continui ottenuti con polarizzazioni leggermente diverse.

A. Amplificatore di tensione

Un transistor montato in configurazione a emettitore comune è in grado di amplificare delle piccole correnti oscillanti, con un guadagno β_f. L'interesse principale della configurazione è che essa può servire anche da *amplificatore di tensione* per piccoli segnali oscillanti. A questo scopo il segnale alternato viene inviato alla base, cioè sovrapposto alla tensione continua V_{BE}, per cui l'ingresso è tra la base e terra, come rappresentato in Fig. 5(b). Notate che nello schema è stato inserito un condensatore C, che serve a "disaccoppiare" l'ingresso rispetto alla tensione continua di polarizzazione. Infatti si vuole che questa tensione serva a polarizzare la giunzione base-emettitore, e quindi occorre evitare che fluisca corrente nel circuito che produce il segnale oscillante: il condensatore blocca le correnti continue e fa passare le componenti alternate.

L'uscita dell'amplificatore, invece, è tra collettore e terra.

Dette v_{in} e v_{out} le *ampiezze* dei segnali in ingresso e in uscita, si definisce il *guadagno in tensione* come $A_v = v_{out}/v_{in}$. Facendo riferimento alle curve caratteristiche di Fig. 4, si può intuire come l'applicazione del segnale oscillante v_{in} induca una modulazione della corrente di base, cioè una modulazione oscillante i_b sovrapposta alla I_B continua di polarizzazione. Il transistor è vincolato a funzionare in punti di lavoro determinati dall'intersezione tra retta di carico e curve a diverse I_B. La modulazione fa sì che il transistor sposti (in modo oscillante) il proprio punto di lavoro: di conseguenza appare una modulazione nella corrente di collettore, $i_c = \beta_f i_b$.

Cerchiamo ora un collegamento tra queste correnti oscillanti e le tensioni v_{in} e v_{out}. Il segnale v_{in} è applicato alla serie della resistenza di base R_B e della *giunzione* base-emettitore. Questa giunzione è polarizzata direttamente, per cui essa viene vista come una resistenza finita. Dato che il comportamento che stiamo esaminando riguarda piccoli segnali oscillanti, la resistenza della giunzione base-emettitore può essere approssimata con la *resistenza dinamica* r_b della giunzione stessa, determinata per il punto di lavoro che corrisponde alla I_B continua di polarizzazione.

Quando abbiamo studiato la resistenza dinamica della giunzione di un diodo bipolare, abbiamo definito la resistenza dinamica come $r_d = (\partial I/\partial V|_{I=Iq})^{-1}$ e abbiamo trovato che essa, per una giunzione che segue l'equazione di Shockley, può essere espressa come $r_d \approx \eta V_T/I_q$, essendo I_q la corrente di lavoro. Lo stesso approccio può essere applicato in questo caso, dato che la giunzione base-emettitore (polarizzata direttamente) è in questo ambito molto simile a quella di un diodo. Questo ci permette di approssimare la resistenza dinamica della base come $r_b = \eta V_T/I_B$. Tenendo conto del collegamento in serie tra le resistenze, possiamo scrivere $v_{in} = (R_B + r_b)i_b = (R_B + \eta V_T/I_B)i_b$.

Per quanto riguarda il legame fra v_{out} e i_c possiamo fare riferimento all'equazione della retta di carico (Eq. 12), notando però che in questo caso siamo interessati solo ai segnali oscillanti, ovvero possiamo trascurare le tensioni continue (che invece servono per la polarizzazione del transistor). Se nell'Eq. 12 indichiamo con v_{out} la componente oscillante del segnale al collettore, legata alla corrente oscillante i_c, togliendo per il motivo appena detto il termine V_0, otteniamo $v_{out} = -R_C i_c$. Il segno meno di questa relazione può essere compreso anche notando che, quando la corrente di collettore aumenta, allora c'è una maggiore caduta di tensione attraverso la resistenza R_C, e quindi la tensione v_{out} diminuisce.

Mettendo tutto insieme, possiamo scrivere:

$$A_v = \frac{v_{out}}{v_{in}} = -\frac{R_C}{R_B + r_b}\frac{i_c}{i_b} = -\frac{R_C}{R_B + \eta V_T/I_B}\beta_f \ . \quad (14)$$

Ritroviamo ancora un segno meno, che sta a significare che il segnale oscillante in uscita è *sfasato* (di π) rispetto a quello in ingresso. Questa equazione dimostra che c'è amplificazione di tensione e che il guadagno dipende non solo da β_f, ma anche dal rapporto $R_C/(R_B + r_b)$. Per aumentare il guadagno si cerca in genere di usare alte resistenze di collettore e basse resistenze di base. Ciò comporta una importante particolarità della configurazione a emettitore comune, interessante per chi deve progettare circuiti e dunque occuparsi dell'accoppiamento fra diversi "stadi", cioè diverse parti del circuito stesso: *l'impedenza di ingresso è bassa, quella di uscita è alta.*

Notate che in certe condizioni, in particolare per alte correnti di lavoro I_B, r_b è trascurabile rispetto a R_B e pertanto $A_v \simeq (R_C/R_B)\beta_f$. Nel caso dell'esperienza svolta in laboratorio questa approssimazione non è generalmente valida. Nell'esperienza che ho svolto io, la polarizzazione era tale da produrre $I_B \approx 12.5\mu$A, a cui corrispondeva $r_b \approx 4.2$ kohm. Poiché nel mio caso era $\beta_f = 208$, tenendo conto che $R_B = 560$ ohm e $R_C = 1$ kohm, si otteneva un valore atteso del guadagno in tensione $A_{v,att} \approx 44$. Dalle misure delle ampiezze (picco-picco) dei segnali in ingresso e in uscita avevo $v_{in} \approx 21$ mV e $v_{out} \approx 940$ mV, per un guadagno misurato $A_v \approx 45$, in ottimo accordo con l'aspettativa (l'ottimo accordo è stabilito sulla base delle incertezze di misura, che qui non sono state riportate per semplicità).

Infine, osservate che la presenza di un guadagno in corrente e di un guadagno in tensione implicano un *guadagno di potenza*, cioè la configurazione esaminata si comporta anche da amplificatore di potenza. È facile verificare che il guadagno in potenza, proporzionale al prodotto $\beta_f A_v$, va con il quadrato di β_f.

B. Stabilità della polarizzazione

Nei voluminosi tomi studiati dai tecnici che progettano circuiti ci sono ampi capitoli dedicati alle tecniche di polarizzazione, cioè alla discussione dei metodi e delle relative configurazioni circuitali che meglio permettono di determinare la polarizzazione del transistor.

Come abbiamo già notato, normalmente un circuito ha a disposizione un solo generatore di differenza di potenziale, per cui tensioni e correnti di polarizzazione devono essere ottenute partizionando in modo opportuno la tensione e la corrente fornita da questo unico generatore attraverso una rete di resistori. L'esempio riportato in Fig. 5, che usa soltanto le due resistenze R_B, R_C, pur essendo semplicissimo e dunque valido dal punto di vista didattico, non è certamente quello che fornisce le migliori performance. Infatti esso può dare facilmente luogo a condizioni di polarizzazione che non sono stabili nel tempo e che possono addirittura condurre al danneggiamento del transistor.

Infatti il circuito funziona solo se c'è un continuo passaggio di corrente I_C nel transistor. Tale corrente non è generalmente trascurabile (nell'esperienza svolta da me valeva diversi mA), essendo β_F-volte maggiore di I_B. Essa può quindi provocare una consistente dissipazione di potenza che può condurre al surriscaldamento del transistor. Se aprite un (vecchio) apparecchio elettronico, vedrete che spesso i transistor sono montati su apposite alette di raffreddamento, proprio per limitare i problemi di surriscaldamento.

Nel nostro circuito questi problemi possono diventare molto seri. Infatti all'aumentare della temperatura il materiale semiconduttore (silicio) di cui è composto il transistor tende a diminuire la propria resistività, facendo aumentare, in una sorta di processo a valanga, la corrente I_C. Al di sopra di un certo limite, riportato nei datasheets (per il transistor usato nell'esperienza questo limite è piuttosto alto, la massima I_C vale 500 mA), la corrente può diventare tale da distruggere il transistor.

Per limitare questo problema, e anche per rendere meglio definito il punto di lavoro del transistor e per ottenere migliori proprietà di linearità dell'amplificazione, cioè per ridurre i fenomeni di *distorsione* della forma d'onda in uscita rispetto a quella in ingresso, esistono numerose varianti di circuiti di polarizzazione. Essi normalmente prevedono la presenza di una resistenza R_E tra emettitore e terra. Il ruolo di questa resistenza può essere capito in modo semplice. Tale resistenza è percorsa dalla corrente di emettitore, I_E, che, nelle condizioni ordinarie di funzionamento, è circa uguale alla corrente di collettore I_C (avevamo visto che $I_C = \alpha_F I_E$, con $\alpha_F \sim 1$). All'aumentare della corrente di collettore aumenta anche quella di emettitore e quindi aumenta la caduta di potenziale ai capi di R_E. L'emettitore si viene allora a trovare a un potenziale positivo diverso da quello di terra. Nel fare questo, la differenza di potenziale V_{BE} tende a diminuire, e quindi la corrente I_B, la quale determina il punto di lavoro, diminuisce. Di conseguenza I_C tende a diminuire, in un processo che, in un certo senso, "auto-stabilizza" le condizioni di funzionamento del transistor evitando conseguenze pericolose e rendendo meglio definito il punto di lavoro.

Un processo di questo tipo può essere considerato come una "retro-azione" (*feedback*). Come vedrete nel prosieguo della vostra carriera, applicare un *feedback* a un dispositivo "attivo" può essere di grandissima utilità per ottenere delle funzionalità particolari. Tra queste c'è anche quella che consente di convertire l'amplificatore in un oscillatore: provate a pensare un pochino a come questo si possa realizzare, cioè in quali condizioni un amplificatore può avere tendenza ad auto-oscillare.