

MA0748 - FISICA PER I DISPOSITIVI IOT

Lorenzo Santi

AA 2022/23 – Lezione 15 16/05/2023

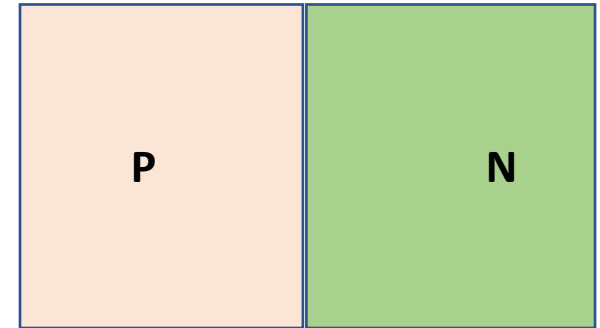
Argomenti della lezione di oggi

- Il transistor
 - Il transistor a giunzione bipolare
 - Il transistor a giunzione ad effetto di campo (JFET)
 - Il transistor MOSFET

Il transistor

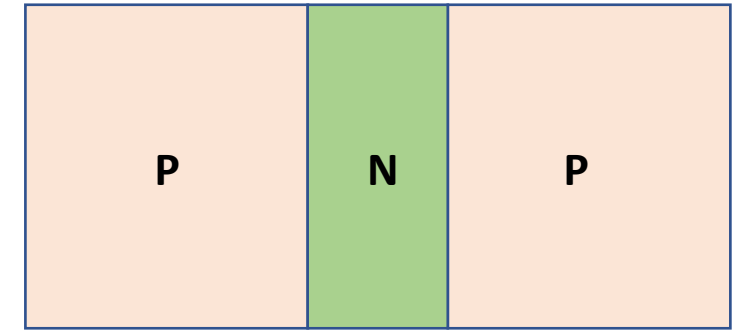
Abbiamo visto come in un diodo a giunzione bipolare il materiale semiconduttore sia drogato in maniera opposta in due regioni, collegate all'esterno con due terminali (anodo e catodo).

A parità di altre condizioni, le caratteristiche della conduzione elettrica in esso dipendono (in maniera non lineare) essenzialmente dalla tensione con cui viene alimentati tali terminali.



Una diversa configurazione di un dispositivo a semiconduttore è quella con tre regioni a diverso drogaggio, di tipo P e di tipo N: ad esempio è riportato in figura il caso di un dispositivo PNP.

Il dispositivo così realizzato si chiama **transistor** (dal termine inglese **transfer of resistance**).



Dotato di tre (o quattro, a seconda dei tipi) di terminali, il transistor ha la caratteristica che la sua resistenza elettrica (o in alternativa la sua conducibilità elettrica, che ne è il reciproco) tra due di tali terminali può essere controllata mediante una opportuna tensione applicata su un'altra delle coppie possibili di terminali.

In questo modo il segnale (in uscita) associato alla corrente circolante tra la prima coppia di terminali può essere controllato da un segnale in tensione (in ingresso) applicato tra i terminali della seconda coppia.

Se la potenza del segnale in uscita è maggiore di quella del segnale in ingresso, viene a realizzarsi un processo di amplificazione del segnale.

I transistor si suddividono in due categorie principali, a seconda del modo con cui viene a generarsi questa amplificazione

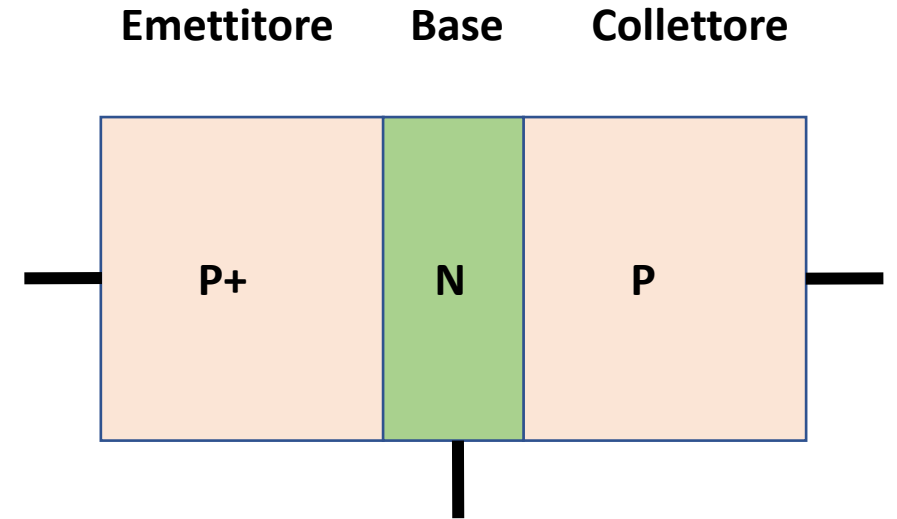
- Transistor a giunzione bipolare
- Transistor ad effetto di campo.

Transistor a giunzione bipolare

Nel transistor bipolare le parti esterne, dello stesso tipo, sono drogate con una percentuale di drogante almeno un ordine di grandezza diversa. Il **sito drogato più pesante** viene chiamato **emettitore**, l'altro **collettore**.

La zona centrale, chiamata **base**, è **drogata in maniera opposta** ed è relativamente sottile.

In figura, illustrante un transistor PNP, l'emettitore è la zona con drogaggio pesante P+, la base quella di tipo N ed il collettore la zona con drogaggio leggero P.

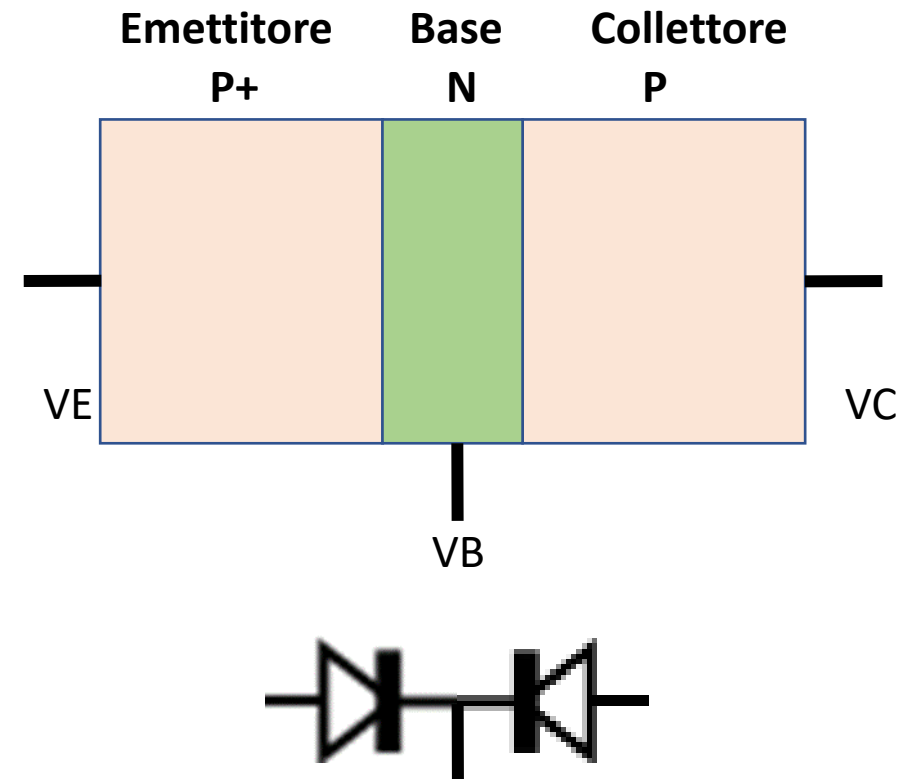


Se le due giunzioni P+N e NP sono alimentate in polarizzazione inversa (oppure non hanno alimentazione), il dispositivo equivale a due diodi messi testa a testa e quindi non conduce corrente misurabile.

Questo modo di funzionamento del transistor viene chiamato **regione di funzionamento di interdizione** e viene realizzata quando i potenziali elettrici sono tali che

$$V_E < V_B > V_C$$

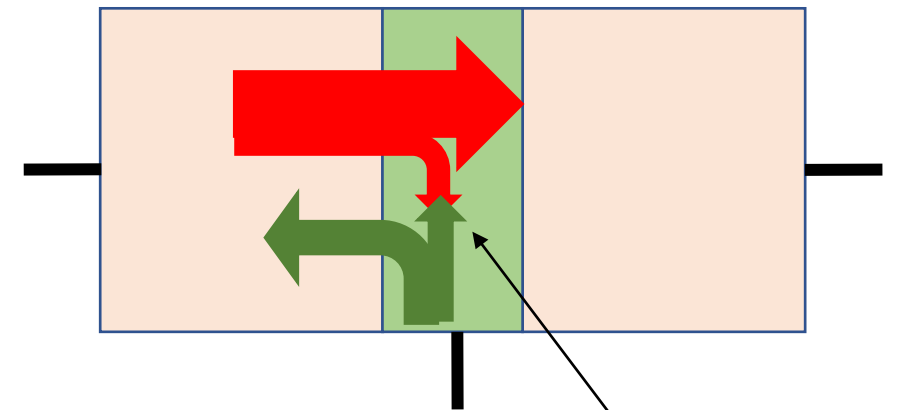
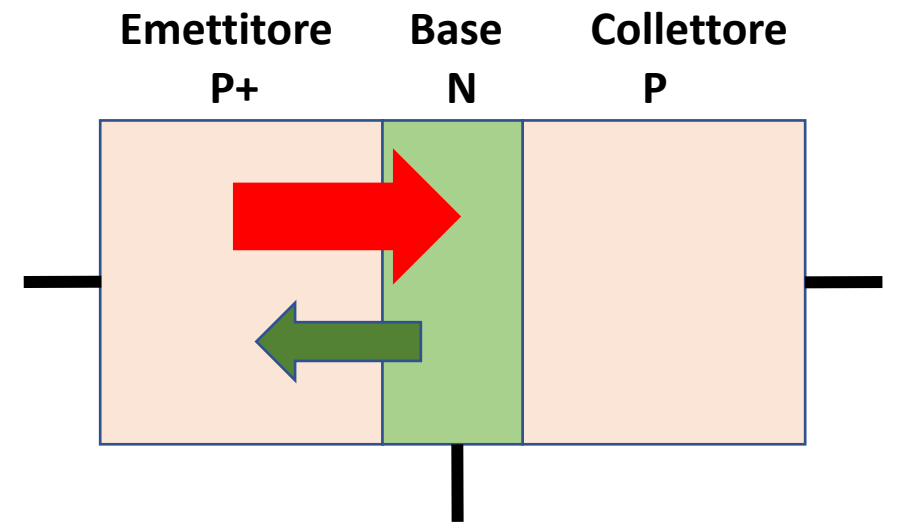
In questo modo le tensioni tra emettitore e base ($V_E - V_B$) e tra collettore e base ($V_C - V_B$) sono negative e quindi le polarizzazioni sono inverse.



Se la tensione applicata tra l'emettitore e la base è tale da polarizzare direttamente la giunzione P+N, incomincia a circolare corrente tra tali due regioni.

Ci sarà una diffusione di lacune da P+ a N e una di elettroni da N a P+. Poiché la densità di portatori di carica in P+ è maggiore, il termine di corrente dovuta alla diffusione di lacune è dominante.

Le lacune diffuse nella base possono ricombinarsi con degli elettroni, scomparendo. Se però la base è sufficientemente sottile, una frazione significativa di lacune sopravvive e giunge alla giunzione NP.



Ricombinazione

Quello che succede alla giunzione NP dipende dalla tensione tra base e collettore.

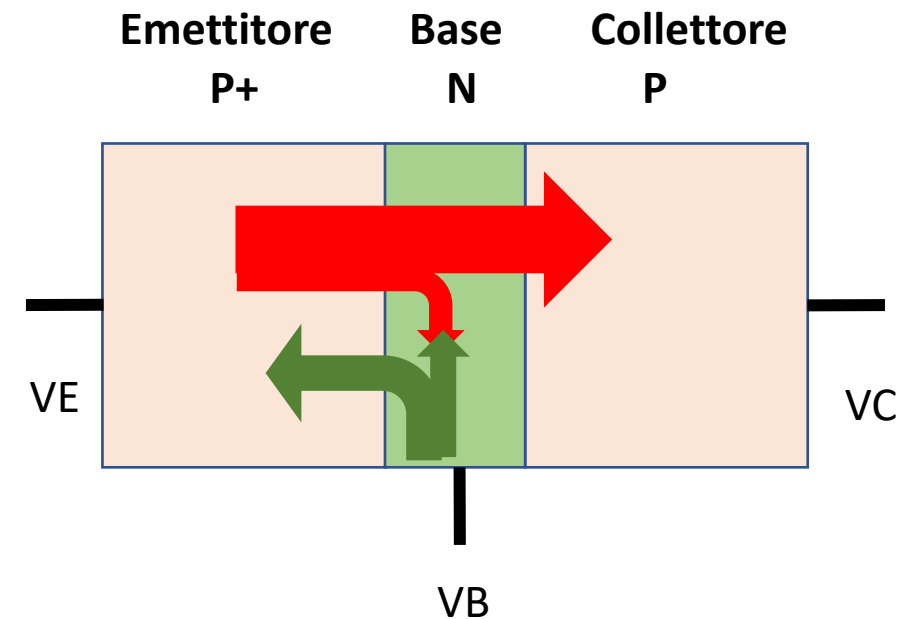
Se la giunzione è polarizzata inversamente e quindi

$$V_C - V_B < 0 \rightarrow V_B > V_C$$

Il flusso di lacune da P ad N e quello degli elettroni da N a P è interdetto, mentre il flusso delle lacune provenienti da P+ attraversa la giunzione, facilitato dalla tensione alla giunzione.

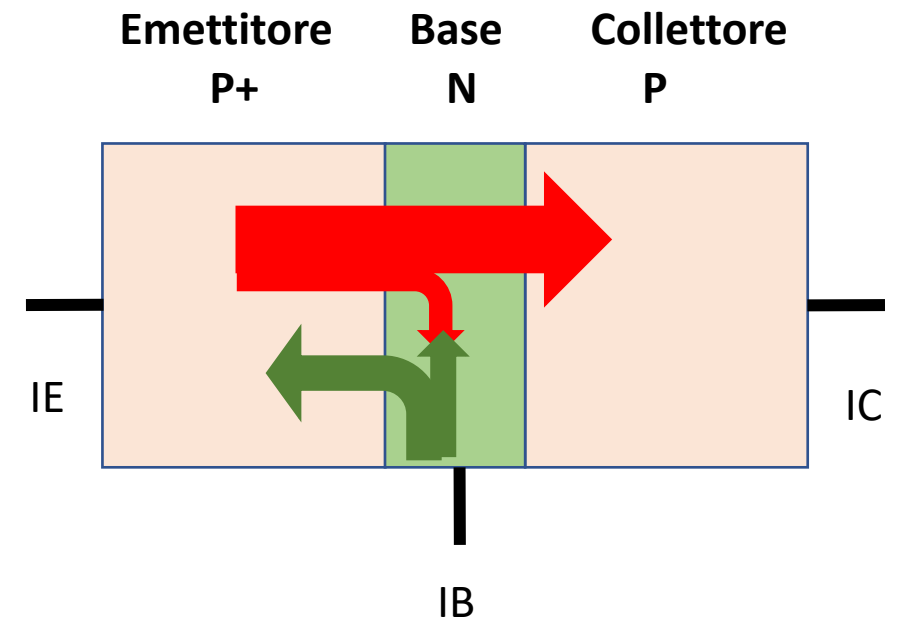
Questa **regione di funzionamento** viene chiamata **attiva diretta** ed è caratterizzata dalla condizione sulle tensioni (ricordiamo che la giunzione P+N è polarizzata diretta)

$$V_E > V_B > V_C$$



Le relazioni tra le tre correnti che abbiamo ai terminali del transistor possono essere descritte da due parametri

$$\begin{aligned} - \alpha_F &= \frac{|I_C|}{|I_E|} \\ - \beta_F &= \frac{|I_C|}{|I_B|} \end{aligned}$$



A causa del pesante drogaggio della zona P+, alla giunzione P+N la corrente dovuta alla diffusione degli elettroni è trascurabile rispetto a quella delle lacune in direzione inversa e quindi il parametro α_F misura essenzialmente la frazione di lacune che sopravvivono al fenomeno di ricombinazione nella regione N.

α_F è tanto maggiore tanto più sottile e meno drogato lo strato N (diminuisce la probabilità che una lacuna nella sua diffusione incontri un elettrone della regione N).

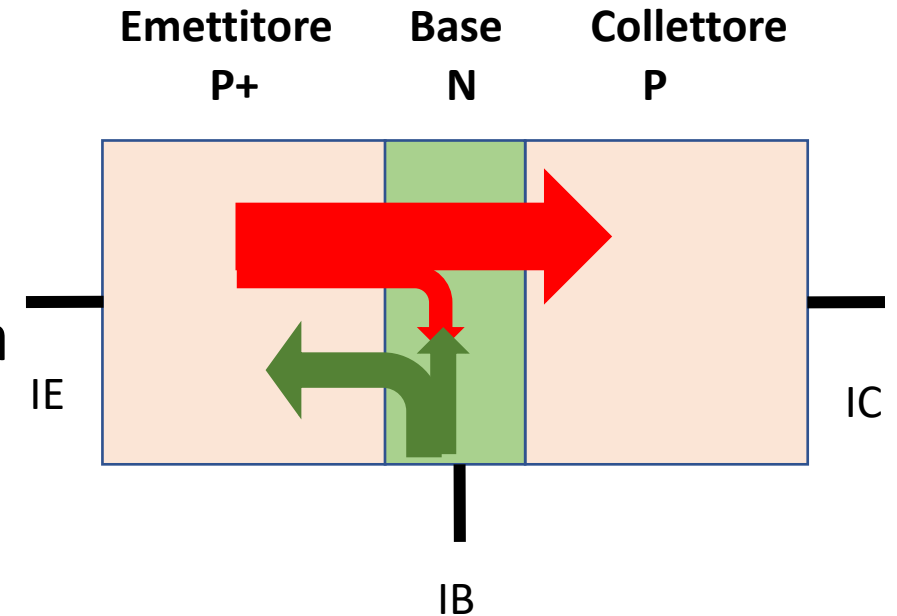
Per una ottimizzazione del dispositivo, α_F deve approssimare il valore unitario (valori commerciali tipici sono tra 0.98 e 0.998).

Il parametro

$$\beta_F = \frac{|I_C|}{|I_B|}$$

invece misura quello che è il guadagno tra la corrente che viene iniettata nella base e quella che esce dal collettore. Questo parametro dipende essenzialmente dal drogaggio delle tre zone del semiconduttore.

β_F nei dispositivi commerciali ha valori tipicamente superiori a 50. Ciò rende il transistor in regione attiva diretta un buon amplificatore di segnali (ad esempio, un dispositivo che amplifica il segnale in corrente in ingresso alla base).



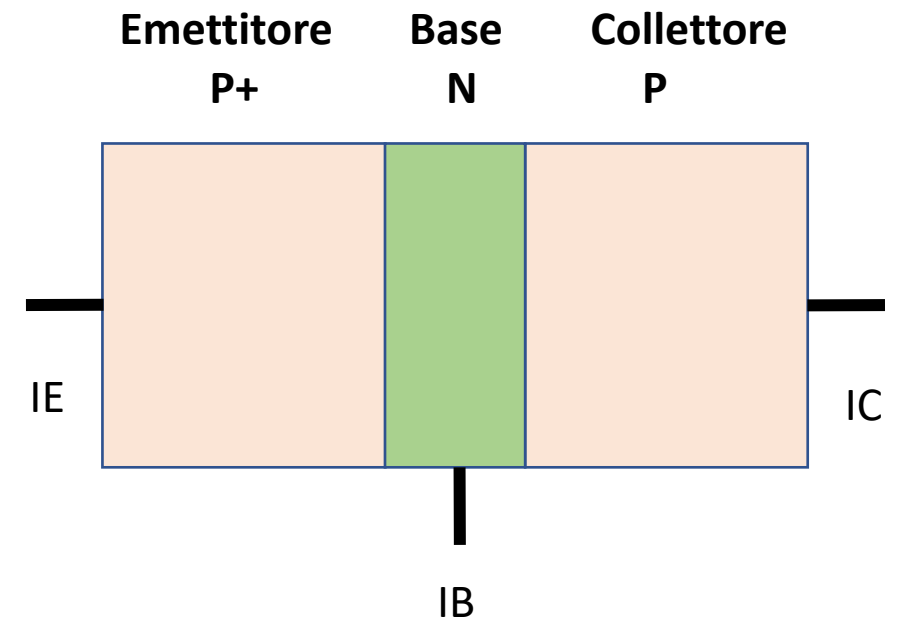
Esistono due ulteriori regioni di funzionamento

- A **saturazione**, quando sia la giunzione P+N che quella PN sono in polarizzazione diretta ($V_E > V_B < V_C$). In questo caso la corrente circolante è massima (satura), per qualunque valore di V_B (al di sotto di una certa soglia).

- **Attiva inversa**, quando $V_E < V_B < V_C$. In questo caso la giunzione PN è in polarizzazione diretta, mentre quella P+N è inversa.

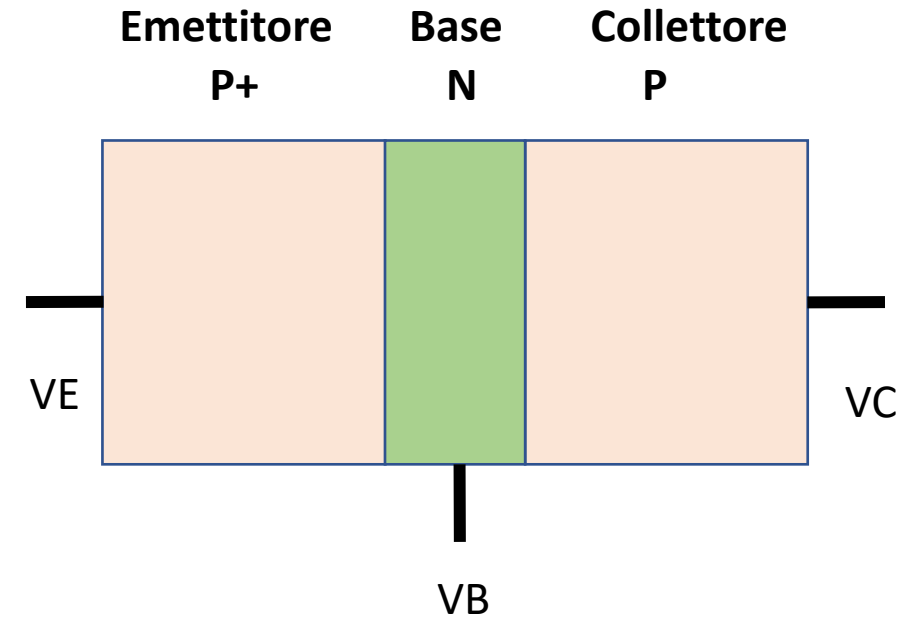
Il transistor funziona come in regione attiva diretta, solo che i ruoli tra emettitore e collettore sono invertiti.

Poiché il drogaggio della zona P è molto minore di quella P+ si ha che le correnti circolanti sono molto inferiori a quelle in regione attiva diretta, rendendo questa regione di funzionamento di scarsa importanza ed uso.



Riassumendo le quattro regioni di funzionamento

- Attiva diretta ($V_E > V_B > V_C$). Il transistor funziona come un amplificatore, con un segnale in corrente in uscita aumentato di un fattore β_F rispetto a quello in ingresso
- Attiva inversa ($V_E < V_B < V_C$). Non utilizzata nella pratica.
- Interdizione ($V_E < V_B > V_C$). Non vi è passaggio di corrente.
- Saturazione ($V_E > V_B < V_C$). La corrente circolante è massima.



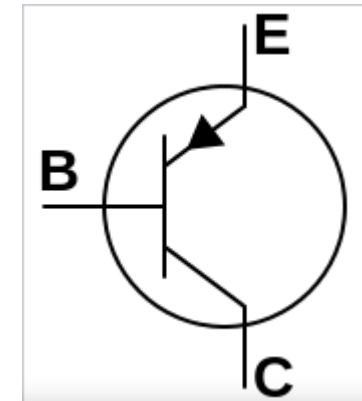
Nelle ultime due modalità il transistor funziona come un **interruttore elettronico** (switch): a seconda del valore di V_B il ramo in cui si trovano emettitore e collettore risulta essere aperto o chiuso.

È questo ad esempio il modo in cui in VISIR vengono implementati i circuiti ideati sulla breadboard: i vari componenti elettrici vengono connessi nella configurazione richiesta, attivando e disattivando gli switch che li attivano.

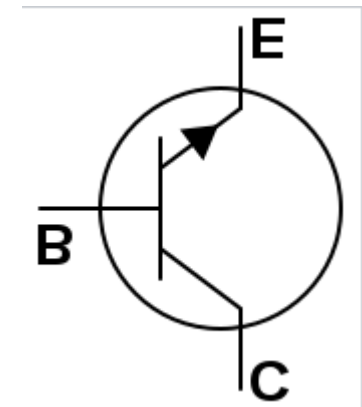
Il simbolo circuitale del transistor a giunzione PNP è quello rappresentato a lato. L'emettitore viene associato con una freccia, che indica il verso di polarizzazione attiva del diodo emettitore-base formato dalla giunzione P+N

In maniera perfettamente analoga vengono realizzati transistor NPN. L'emettitore è ancora la regione più pesantemente drogata (N+). Quanto abbiamo visto per il transistor PNP vale anche per il NPN, a patto di scambiare nelle descrizioni elettroni con lacune e invertire il segno di tensioni e correnti ai terminali.

Il simbolo circuitale associato mostra come sia il verso di polarizzazione attiva del diodo emettitore-base formato dalla giunzione N+P



Transistor PNP

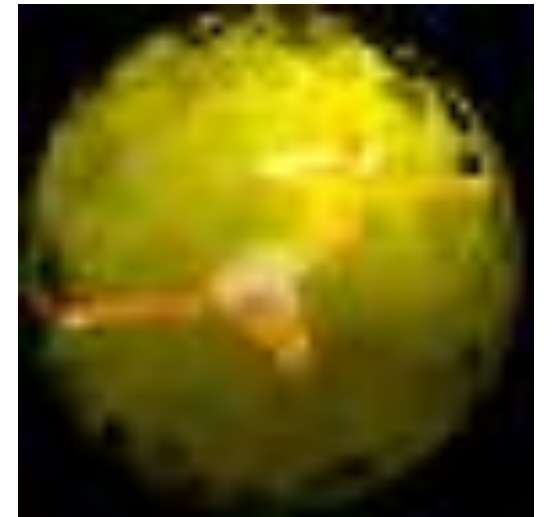


Transistor NPN

Continuiamo la visione di uno dei filmati,
per la parte riguardante i transistor NPN

Semiconductors - Physics inside Transistors and
Diodes

<https://youtu.be/hrpPKCDLRN0?t=601>

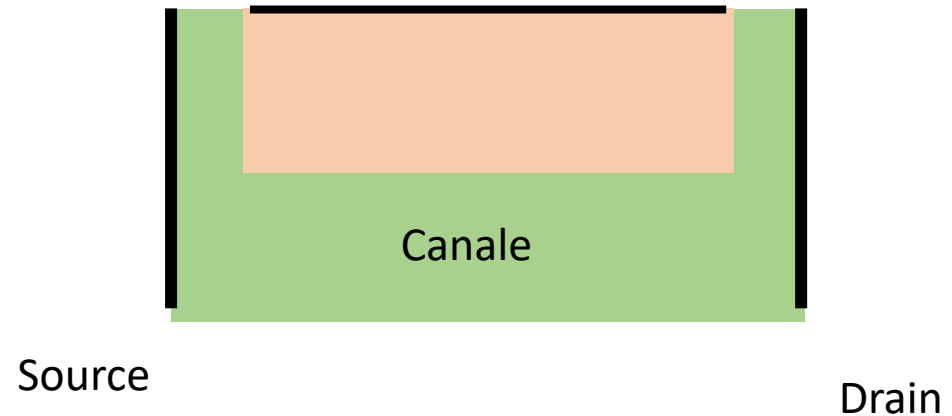


Il transistor a giunzione ad effetto di campo (JFET)

Un transistor di tipo diverso da quello a giunzione bipolare è il cosiddetto transistor a giunzione ad effetto di campo (**Junction Field Effect Transistor**).

In figura è mostrato un esempio schematizzato di un FET in cui due terminali (source e drain) sono connessi elettricamente da un semiconduttore drogato N (**canale n**).

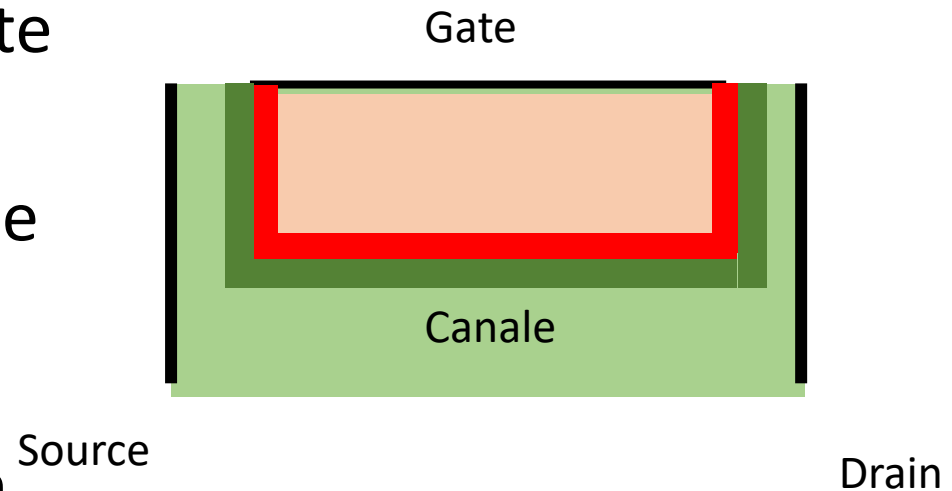
La conduzione lungo il canale dipende dalla densità dei portatori maggioritari (in questo caso elettroni) e quindi dal drogaggio del materiale.



La conduzione elettrica dipende però anche dalla larghezza della sezione trasversa del canale e questa è definita da una zona drogata in maniera opposta (in questo caso P) chiamate **gate**.

La giunzione NP che si ha così crea una regione di svuotamento, priva di portatori di carica, restringendo di fatto il canale.

Il potenziali elettrici di gate e di source e drain sono poi tali che la giunzione sia polarizzata inversamente, in maniera tale da rendere trascurabile la corrente verso il gate e significativa solo quella attraverso il canale.



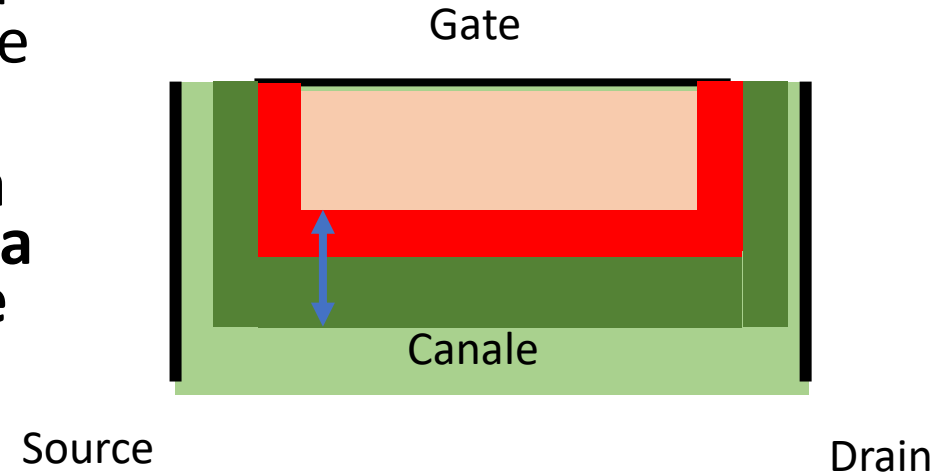
La regione di svuotamento può essere modificata in spessore, variando la tensione in polarizzazione inversa tra il gate e gli altri due terminali.

Questo è un effetto che abbiamo trascurato nella descrizione del diodo a giunzione: **aumentando la tensione in polarizzazione inversa alla giunzione i portatori di maggioranza nelle due regioni drogate vengono «allontanati» dalla giunzione, aumentando lo spessore della regione di svuotamento.**

In questo modo, regolando la tensione del gate rispetto ai terminali di source e drain, si può regolare la conducibilità elettrica del canale: applicando una tensione tra source e drain, si ha un flusso di corrente modulato dal potenziale elettrico sul gate.

Si ha così per il transistor JFET le stesse modalità di utilizzo di quello a giunzione PN:

- **Amplificatore di segnale** (il segnale in ingresso è quello al gate, quello in uscita è rappresentato dalla corrente tra source e drain)
- **Switch** (le due configurazioni limiti, con il canale completamente aperto oppure ridotto ad una larghezza trascurabile, corrispondono agli stati chiuso ed aperto di un interruttore elettronico)

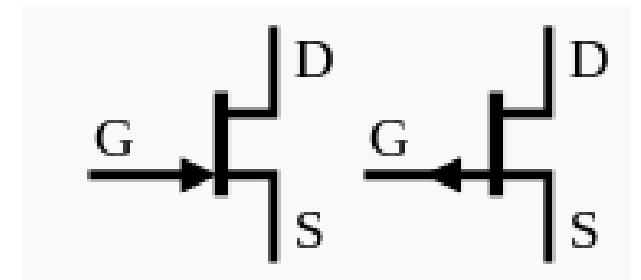


Il simbolo circuitale del JFET è riportato accanto, nelle due versioni:

- Canale n, con gate drogato P
- Canale p, con gate drogato N

Le frecce in questo caso indicano il tipo di giunzione realizzato dal terminale di gate:

- Gate drogato P, con giunzione PN da sinistra a destra
- Gate drogato N, con giunzione PN da destra a sinistra



Canale n

Canale p

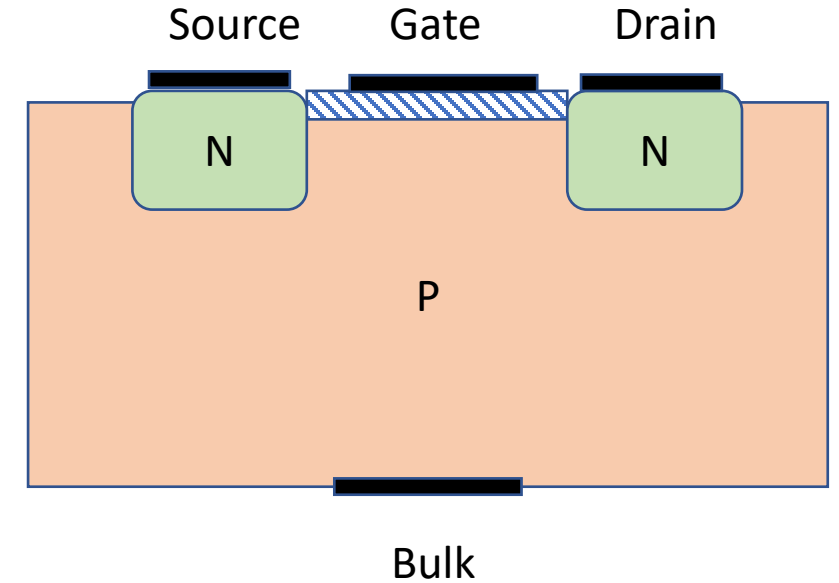
Confrontando il funzionamento del transistor JFET con quello a giunzione bipolare, possiamo osservare come

- Nel JFET il flusso di corrente è limitato a quello del canale conduttivo, coinvolgendo solo i portatori di maggioranza di tale zona. Nel transistor bipolare invece, la base contribuisce ai flussi di corrente, con portatori di carica di segno opposto a quelli che fluiscono da emettitore a collettore (è questa la ragione per cui viene chiamato bipolare, coinvolgendo correnti sia di elettroni che di lacune, anche se in proporzioni diverse).
- Poiché nel transistor bipolare ci sono correnti che fluiscono attraverso la base, l'inserzione della base in un circuito che fornisca il segnale in ingresso al transistor ne modifica le caratteristiche. Questo non avviene per il JFET, poiché essendo completamente trascurabile la corrente del gate, la sua inserzione in un circuito non modifica i flussi di corrente in esso.

Il transistor MOSFET

Il transistor MOSFET (Metal Oxide Semiconductor FET) è un transistor ad effetto di campo tipicamente realizzato in silicio che ha due differenze sostanziali rispetto al JFET

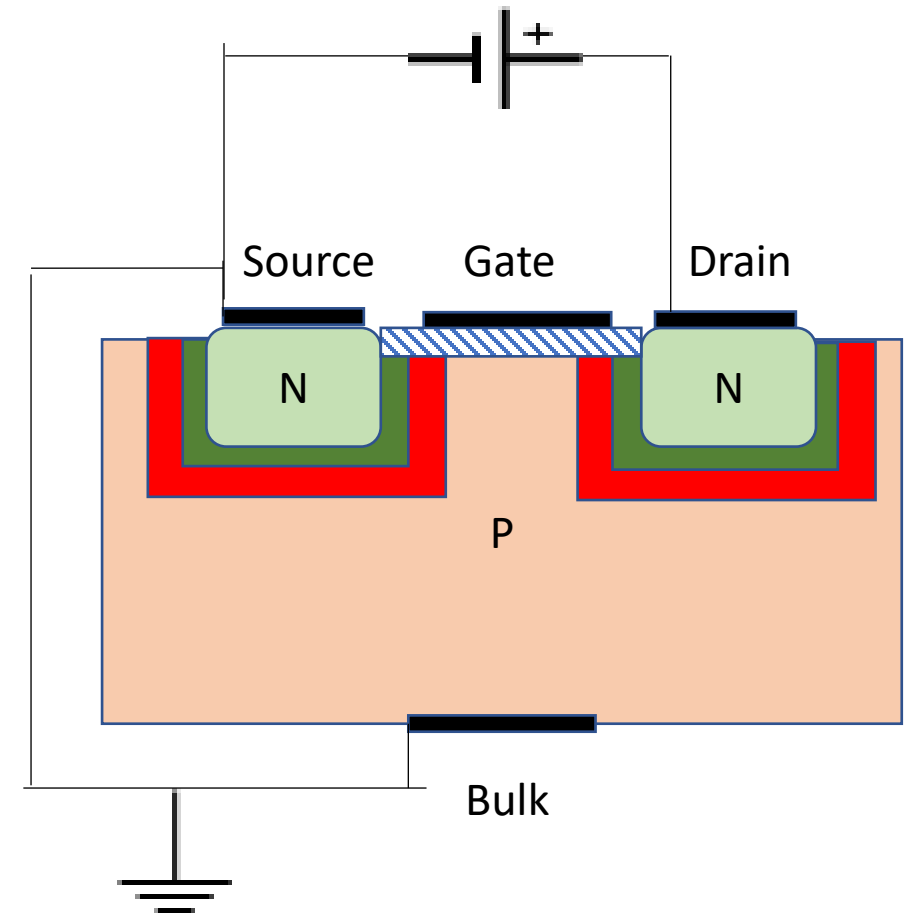
- 1) Il terminale metallico del gate è separato dal resto del dispositivo da uno strato di ossido di silicio, che funziona come isolante elettrico
- 2) Source e drain sono separati da una zona di semiconduttore drogato in maniera opposta (**substrato** o **bulk**) e senza alimentazione non c'è un canale conduttivo tra i due terminali S e D.



In figura è mostrato lo schema di un MOSFET con source e drain drogati N e bulk drogato P

Source e bulk sono normalmente cortocircuitati a massa, mentre al drain è applicata una tensione tale da polarizzare inversamente la giunzione con il bulk.

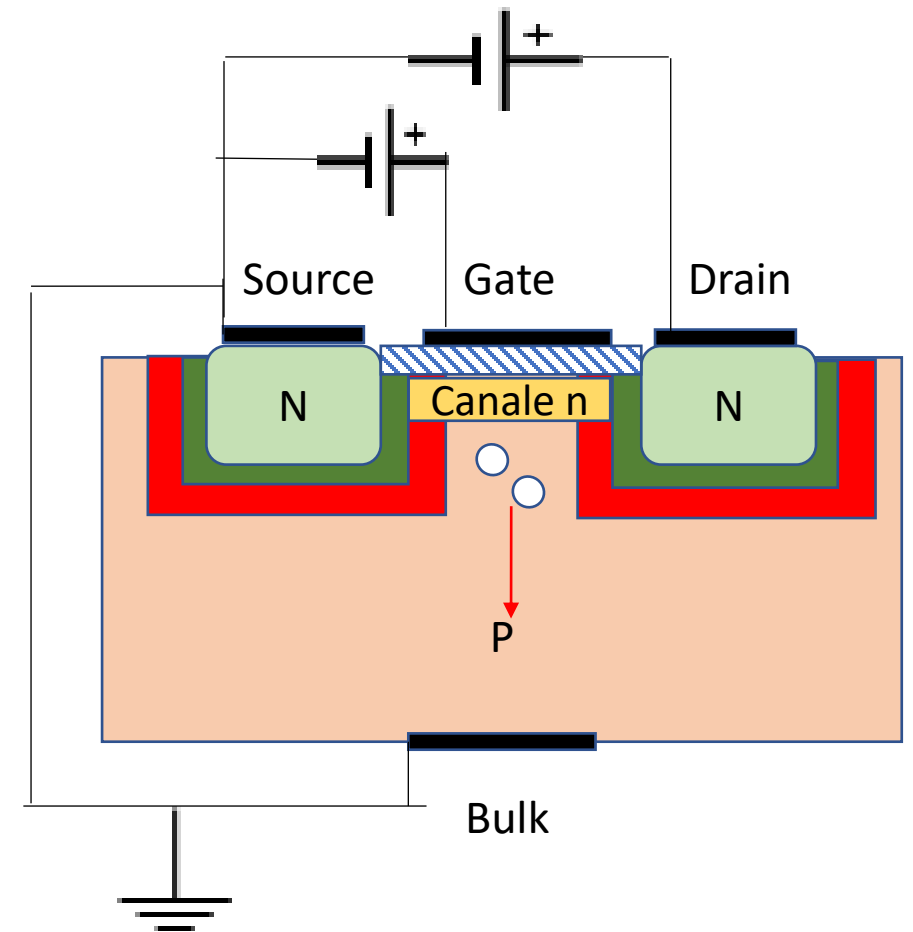
In questo modo source e drain sono isolati dal bulk mediante le zone di svuotamento delle due giunzioni.



Se sul gate non è applicata tensione, source e drain sono isolati elettricamente tra di loro e non può circolare corrente tra di loro.

La situazione cambia se applichiamo al gate una tensione tale da respingere i portatori di maggioranza (nell'esempio di figura, lacune) del bulk dai pressi della zona a contatto con lo strato di isolante.

Respingere lacune significa però attrarre degli elettroni: questi tipicamente provengono dalle zone drogate di tipo N, cioè source e/o drain.

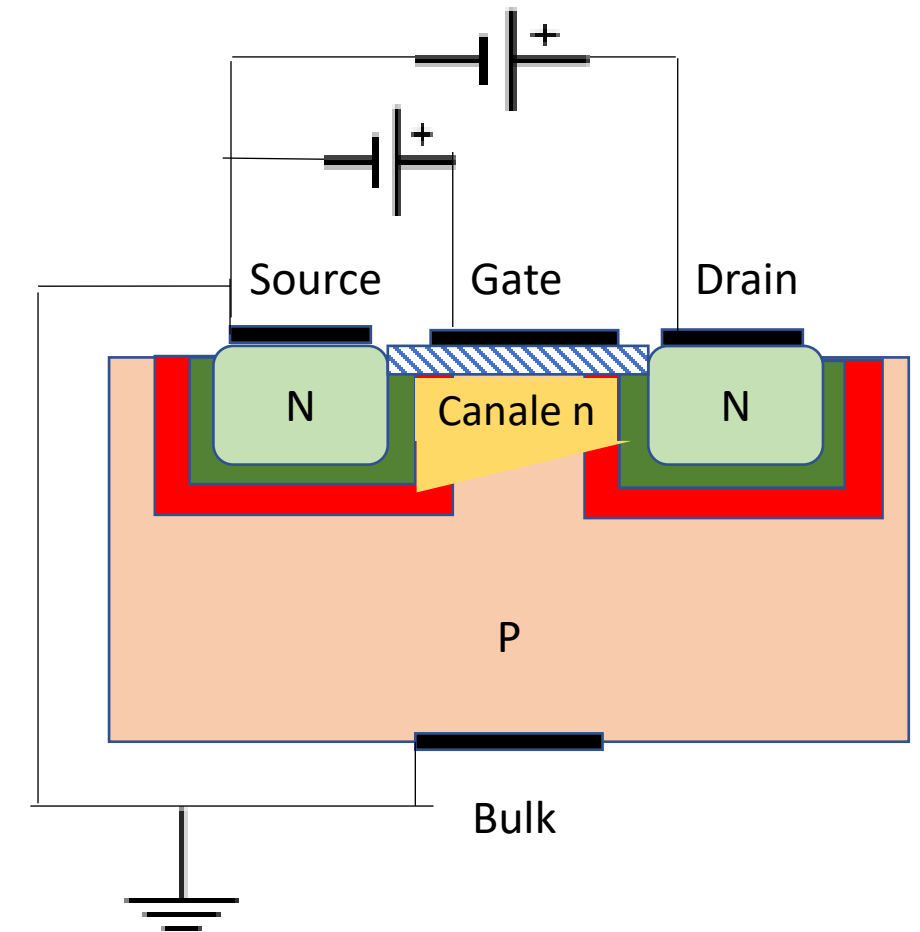


Per un valore di tensione al gate superiore ad un certo valore di soglia, il flusso di elettroni da source/drain alla zona sottostante lo strato isolante è tale da connettere le due zone a drogaggio N.

Si viene così a creare un canale (di tipo n) in cui gli elettroni possono spostarsi tra source e drain: dato il tipo di alimentazione, il loro moto sarà da source a drain e quindi la corrente sarà positiva nel verso opposto

Aumentando la tensione sul gate, aumenta lo spessore del canale n, aumentando quindi la corrente che attraversa il canale.

La correlazione corrente source-drain vs tensione su gate è con buona approssimazione lineare, fino ad un certo valore di tensione di gate.

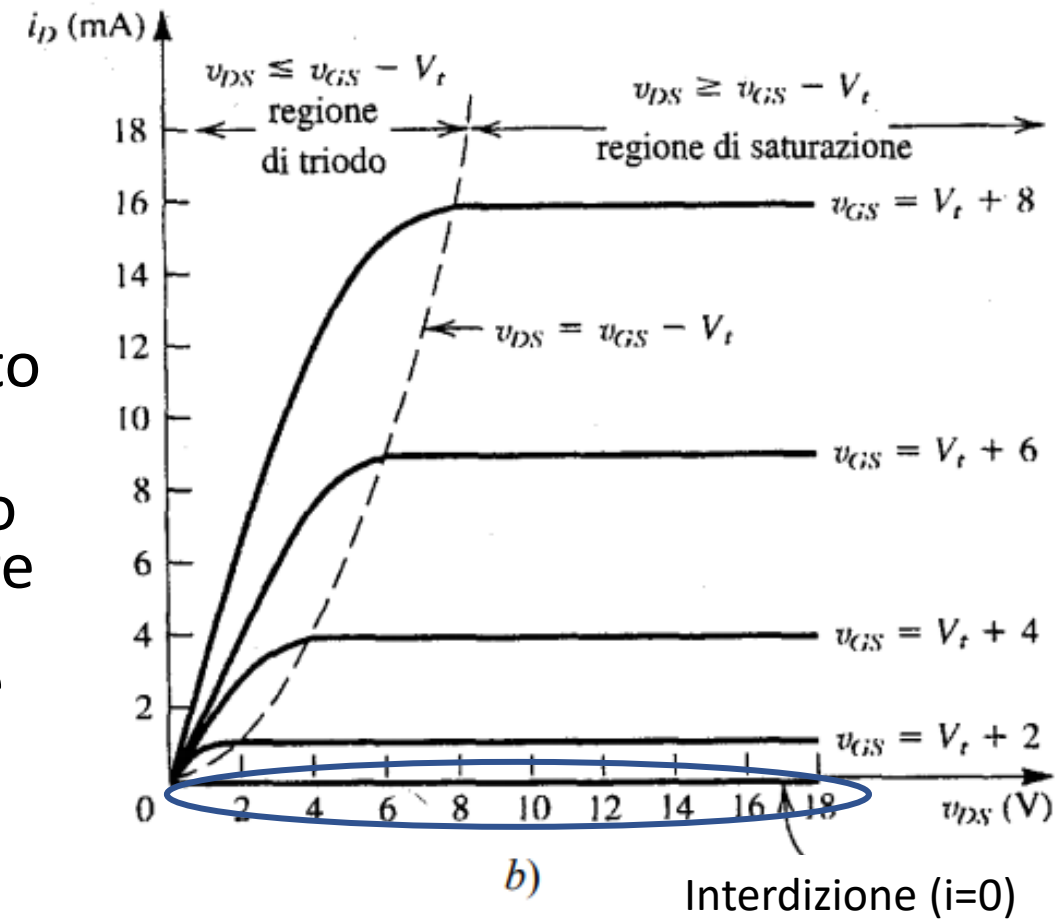


Al di sopra di tale valore, poiché la tensione tra gate e drain diminuisce all'aumentare di quella sul gate, il canale nei pressi del drain diminuisce di spessore (mentre aumenta nei pressi del source).

Si è in un regime chiamato di saturazione.

Abbiamo così tre modi di funzionamento del MOSFET

- In **interdizione**, quando la tensione V_{SG} sul gate (rispetto al source) non è sufficiente a creare un canale tra source e drain. In questo caso la corrente i_D in uscita al drain è nulla.
- Come **triode**, quando il canale risulta aperto ed aumenta la sua larghezza con l'aumentare della tensione sul gate. La corrente i_D è proporzionale alla tensione V_{SD} tra source e drain e la resistenza elettrica (il reciproco delle pendenze dei tratti rettilinei) dipende dalla tensione di gate V_{SG}
- In **saturation**, quando il canale tende a restringersi nei pressi del drain, mantenendo la corrente costante al variare della tensione source drain (l'aumento di tensione V_{SD} è compensato da un aumento della resistenza elettrica, dovuta al restringimento della sezione del canale). In questo caso la corrente i_D dipende praticamente solo dalla tensione sul gate V_{SG} .



Di nuovo, come per i precedenti tipi di transistor, per il MOSFET le due modalità di interdizione e di saturazione sono quelle utili alla realizzazione di circuiti che implementano

- Switch elettronici: a seconda che la tensione di gate realizza una condizione di interdizione o di saturazione, il canale che unisce source con drain è chiuso (switch off) oppure permette la conduzione di corrente (switch on).
- Amplificatori: in modalità di saturazione, la corrente che circola nel canale dipende esclusivamente dalla tensione di gate, realizzando così una amplificazione del segnale come corrente uscente dal drain.