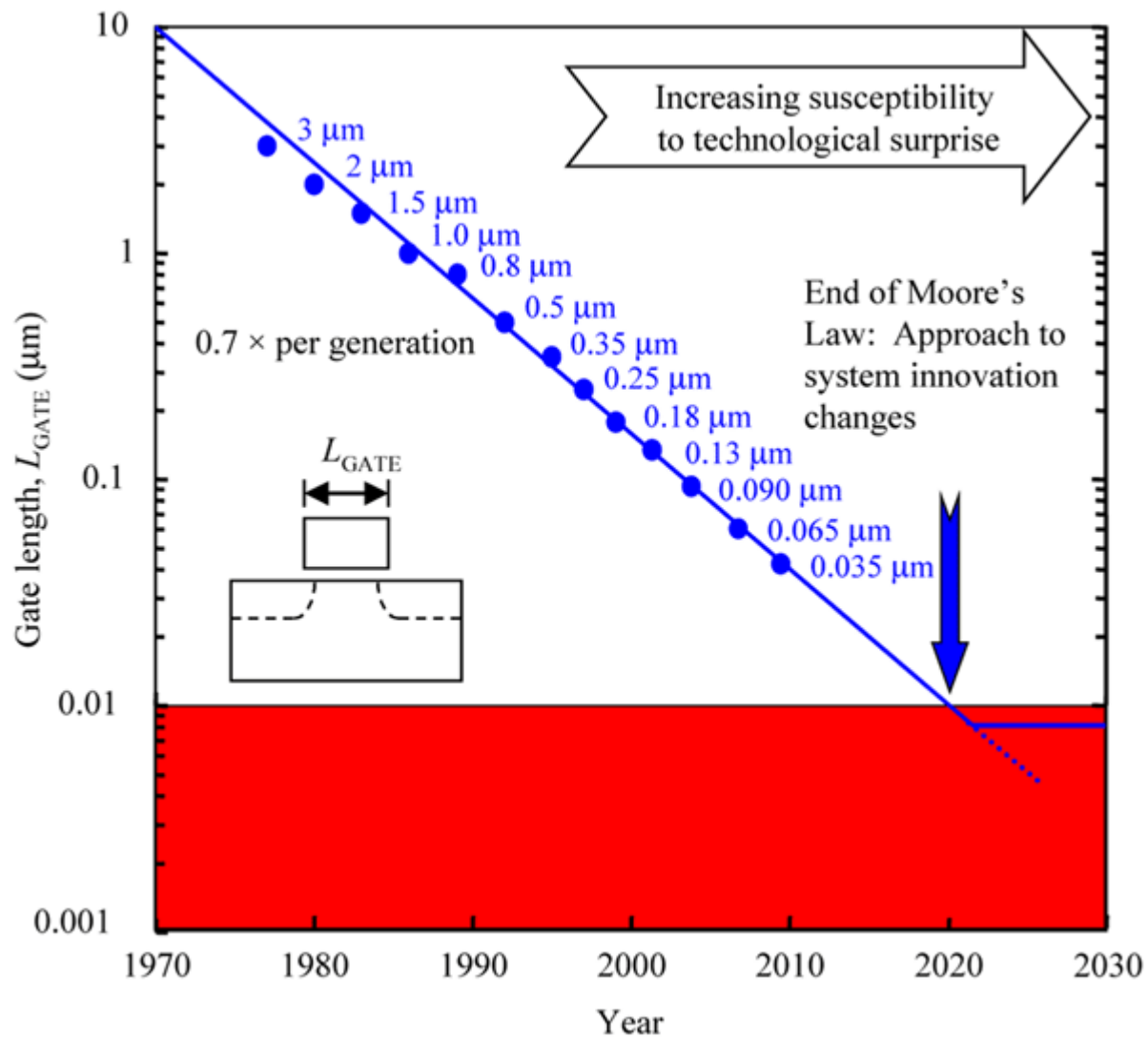


Legge di Moore



Regole di Scaling dei dispositivi MOSFET

Le regole di scaling determinano il modo in cui i parametri fondamentali del MOSFET devono essere variati per scalare le dimensioni.

Esse prendono il nome di **scaling a campo costante, a tensione costante** o generale a seconda del parametro che viene modificato.

Lo scopo è quello di migliorare la **densità di integrazione**, i **tempi di ritardo** e il **consumo di potenza**.

Vedremo che non sempre è possibile migliorare significativamente tutti questi parametri in contemporanea

Inoltre, **diminuire L su scale molto piccole (ben al di sotto dei 100 nm)** significa andare incontro a **problematiche di vario tipo**

Scaling a campo costante

La teoria dello scaling segue in generale tre regole:

1. **Riduzione di tutte le dimensioni verticali e orizzontali di un fattore S**
2. **Riduzione della tensione di soglia e di alimentazione di un fattore S**
3. **Incrementare il livello di drogaggio di un fattore S**

Con $S > 1$

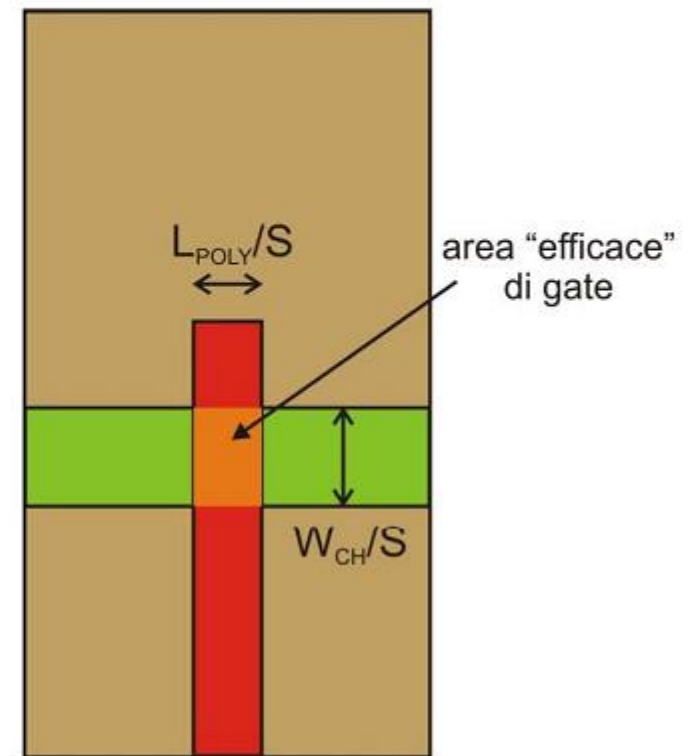
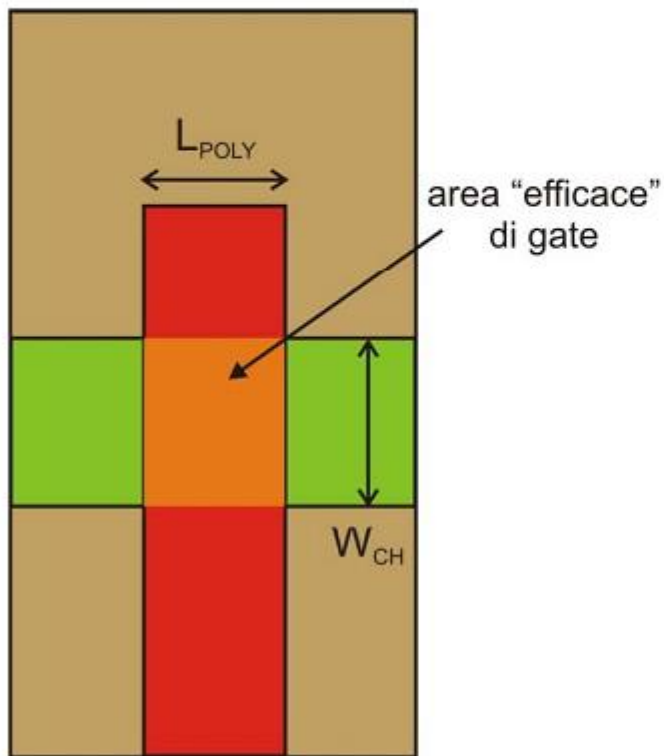
Dato che le dimensioni e il voltaggio scalano di pari passo, tutti i campi rimangono costanti

Constant field scaling

Scaling a campo costante

Definiamo le nuove grandezze, scalate in dimensioni di un fattore S , in questo modo:

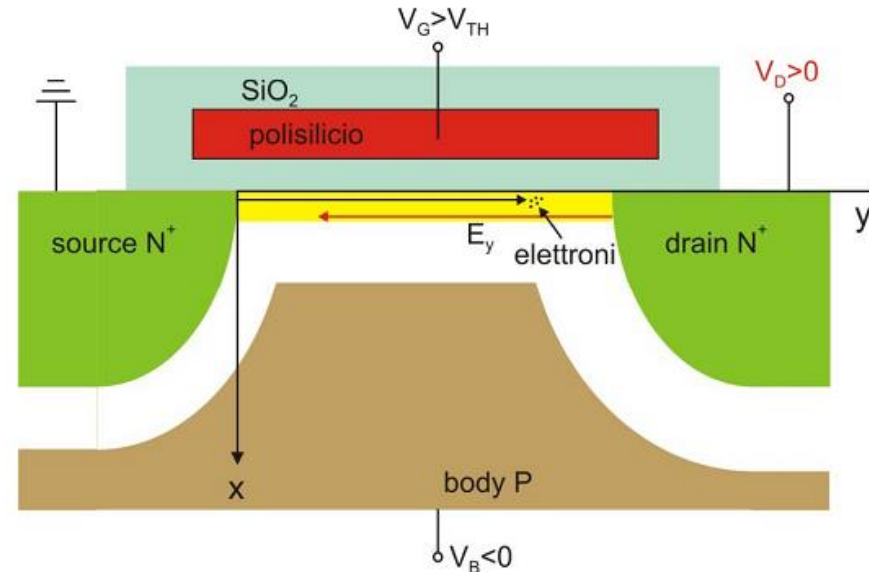
$$L' = \frac{L}{S} \quad W' = \frac{W}{S}$$



Scaling a campo costante

Nella strategia “a campo costante” **si interviene anche sulla tensione di alimentazione**, e la si scala di un fattore S rispetto al valore della generazione precedente (V_{DD})

$$V_{DD}' = \frac{V_{DD}}{S}$$



Questo comporta che tutte le tensioni di polarizzazione in gioco (limitate dalla tensione di V_{DD} , scalino “mediamente” di un fattore S):

$$V_{DS}' = \frac{V_{DS}}{S} \quad V_{GS}' = \frac{V_{GS}}{S} \quad V_{SB}' = \frac{V_{SB}}{S}$$

Scaling a campo costante

Cosa succede al campo orizzontale, E_y ?

Ci ricordiamo che il canale è formato ($V_{GS} > V_{TH}$) ed applichiamo una tensione tra drain e source $V_{DS} > 0$ V.

Il campo elettrico longitudinale in un canale di lunghezza L risulta essere dato approssimativamente da

$$E_y \approx \frac{V_{DS}}{L}$$

$$E_y' \approx \frac{V_{DS}'}{L'} = \frac{V_{DS}}{S} \frac{S}{L} = E_y$$

Lo scalamento non influisce sul campo orizzontale, campo costante!

Scaling a campo costante

Come varia la corrente di drain nell'ipotesi - per semplicità - di essere in regione di pinch-off?

Supponiamo inizialmente di **non eseguire alcuna variazione sulle grandezze tecnologiche** che influenzano la tensione di soglia e che la dipendenza di quest'ultima dalla d.d.p. tra source e body sia trascurabile.

La corrente di drain del dispositivo "scalato" risulta, in tal caso, data da

$$I_D' = K'(V_{GS}' - V_{TH})^2$$

Per quel che concerne la transconduttanza, supponendo di non variare lo spessore dell'ossido si ottiene invece:

$$K' = \frac{\mu_N C_{OX} W'}{2L'} = \frac{\mu_N C_{OX} W}{2L} = K$$

Scaling a campo costante

Da cui:

$$I_D' = K \left(\frac{V_{GS}}{S} - V_{TH} \right)^2$$

La corrente di uscita per cui viene ridotta dallo scalamento, a meno che non si decida di intervenire sul processo di fabbricazione con il fine di abbassare la tensione di soglia, ovvero:

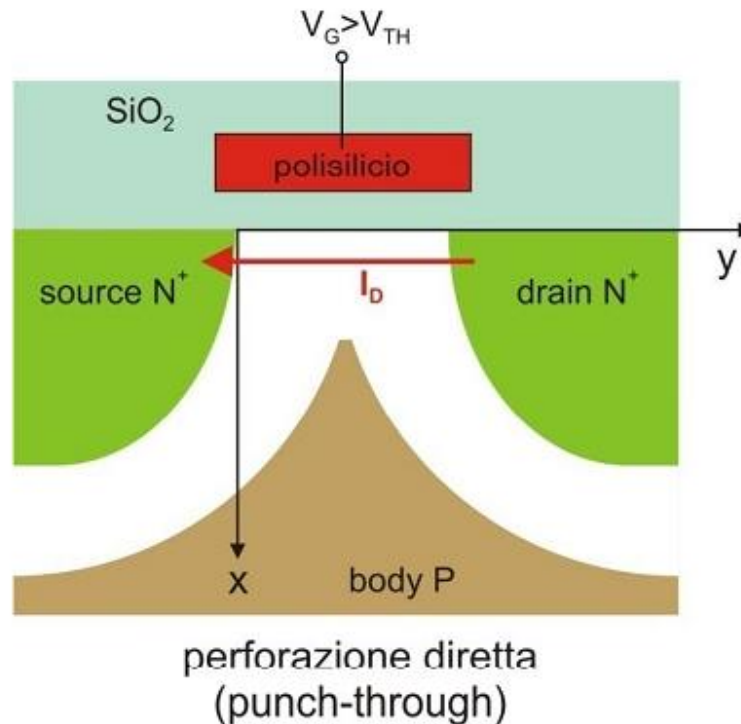
- Livello di drogaggio
- Spessore dell'ossido

$$V_T = \frac{\sqrt{2\varepsilon_S q N_A (2\psi_B)}}{C_o} + 2\psi_B + V_{FB}$$

Scaling a campo costante

Un modo per ridurre la tensione di soglia è sicuramente ridurre il drogaggio degli accettori N_A .

Il canale è corto, per cui **problemi associati al punch-through**



Giunzioni p-n+

W si estende sul lato meno drogato

La corrente non è più controllata dal gate!

Non si può diminuire il drogaggio del substrato

Scaling a campo costante

In genere si tende ad aumentarlo!

Proprio per evitare la perforazione diretta

La tensione di soglia aumenta!

La corrente diminuisce ulteriormente!

Inoltre, se il substrato è molto drogato:

- Riduzione tensione di rottura per valanga
- Aumento capacità di svuotamento

Il drogaggio viene aumentato SOLO nell'area di canale

Impiantazione localizzata

$$N_A' = S \times N_A$$

Scaling a campo costante

Dobbiamo ancora ridurre la tensione di soglia, come facciamo?

Scalamento dello spessore dell'ossido

$$d'_{ox} = \frac{d_{ox}}{S}$$

Da cui:

$$K' = \frac{\mu_N C_{OX} 'W'}{2L'} = \frac{\mu_N C_{OX} 'W}{2L} = \frac{\mu_N \epsilon_{OX} W}{2t_{OX} 'L} = \frac{S \mu_N \epsilon_{OX} W}{2t_{OX} L} = S \cdot K$$

Lo scalamento ci ha permesso di aumentare K

Cosa succede alla tensione di soglia in virtù di tutte queste modifiche

È molto complicato dare una risposta immediata!

Scaling a campo costante

Ci ricordiamo che

$$V_{TH} = V_{FB} + 2\psi_B + \gamma\sqrt{2\psi_B + V_{SB}}$$

E che

$$t_{OX}' = \frac{t_{OX}}{S} \quad N_A' = S \cdot N_A \quad V_{SB}' = \frac{V_{SB}}{S}$$

La tensione di soglia sarà maggiormente influenzata dall'effetto body, gli altri due termini variano poco

$$V_{TH} \approx \gamma\sqrt{V_{SB}}$$

$$V_{TH}' \approx \gamma'\sqrt{V_{SB}'}$$

Scaling a campo costante

Ci ricordiamo che

$$\gamma' = \frac{\sqrt{2\varepsilon_S q N_A'}}{C_{OX}'} = \frac{\sqrt{2\varepsilon_S q S N_A} t_{OX}}{\varepsilon_{OX} S} = \frac{\sqrt{2\varepsilon_S q N_A}}{C_{OX}} \frac{1}{\sqrt{S}} = \frac{\gamma}{\sqrt{S}}$$

Mentre

$$\sqrt{V_{SB}'} = \sqrt{\frac{V_{SB}}{S}} = \frac{\sqrt{V_{SB}}}{\sqrt{S}}$$

Di conseguenza:

$$V_{TH}' \approx \gamma' \sqrt{V_{SB}'} = \frac{\gamma}{\sqrt{S}} \frac{\sqrt{V_{SB}}}{\sqrt{S}} = \frac{\gamma \sqrt{V_{SB}}}{S} = \frac{V_{TH}}{S}$$

Anche la tensione di soglia scala con S

Scaling a campo costante

Come cambia l'espressione della corrente del dispositivo?

$$I_D' = K'(V_{GS}' - V_{TH}')^2 = S \cdot K \left(\frac{V_{GS}}{S} - \frac{V_{TH}}{S} \right)^2 = \frac{K}{S} (V_{GS} - V_{TH})^2 = \frac{I_D}{S}$$

La corrente viene comunque a diminuire in seguito allo scalamento

Ma questa variazione è meno importante rispetto a quella che abbiamo ottenuto senza variare drogaggio e spessore dell'ossido

Scaling a campo costante

Analizziamo ora le grandezze critiche del sistema, partendo dal campo verticale E_x

Ricordiamo che:

$$V_{DD}' = \frac{V_{DD}}{S} \quad L' = \frac{L}{S} \quad W' = \frac{W}{S} \quad N_A' = S \cdot N_A \quad t_{OX}' = \frac{t_{OX}}{S}$$

Il campo verticale, in prima approssimazione, vale:

$$E_x \propto E_{OX} = \frac{V_{GS} - 2\psi_B - V(y)}{t_{OX}} \approx \frac{V_{GS}}{t_{OX}}$$

$$E_x' \propto \frac{V_{GS}'}{t_{OX}'} = \frac{V_{GS}}{S} \frac{S}{t_{OX}} = \frac{V_{GS}}{t_{OX}} \Rightarrow E_x = E_x'$$

Lo scalamento ha mantenuto anche il campo verticale costante!

Scalamento a campo costante

Effetti dello scalamento

Supponiamo di considerare un inverter elementare simmetrico

L'area di gate sarà:

$$A_{GATE}' = W_N' L_N' + W_P' L_P' = \frac{W_N L_N}{S^2} + \frac{W_P L_P}{S^2} = \frac{A_{GATE}}{S^2}$$

Occupazione della porta logica minore!

Tempo di propagazione:

$$t_P = \frac{C}{I} V_{DD} = \frac{C}{K} \frac{V_{DD} / 2}{(V_{DD} - V_{TH})^2}$$

$$t_P' = \frac{C'}{K'} \frac{V_{DD}' / 2}{(V_{DD}' - V_{TH}')^2} = \frac{C}{S} \frac{1}{S \cdot K} \frac{V_{DD} / 2 S}{\left(\frac{V_{DD}}{S} - \frac{V_{TH}}{S} \right)^2} = \frac{t_P}{S}$$

Tempo di propagazione minore!

La velocità della porta logica aumenta!

Effetti dello scalamento

Frequenza di commutazione:

Immediata conseguenza dell'aumento di velocità è che anche la frequenza di commutazione aumenta di conseguenza di un fattore S

$$f' = S \cdot f$$

Potenza dissipata:

$$P_D' = f' C' (V_{DD}')^2 = S \cdot f \frac{C}{S} \left(\frac{V_{DD}}{S} \right)^2 = \frac{P_D}{S^2}$$

Diminuzione notevole della potenza dissipata

- Scalamento della tensione applicata
- Riduzione dell'area e quindi delle capacità parassite

Effetti dello scalamento

Cosa succede invece alla densità di potenza dissipata, ovvero alla potenza dissipata per unità di area?

Ora il numero di porte da integrare è aumentato in funzione dello scalamento del dispositivo!

Potenza dissipata per unità di area

$$P_D' = \frac{P_D'}{A_{GATE}'} = \frac{P_D}{S^2} \frac{S^2}{A_{GATE}} = P_D$$

Riassumendo:

- Lo scalamento **a campo costante** porta ad un miglioramento della maggior parte dei parametri di interesse
- Alcuni rimangono invariati

Effetti dello scalamento

Precisazioni importanti

Inoltre è inutile scalare le tensioni se lo standard è differente

La maggior parte dei dispositivi vengono comunque alimentati ad una tensione di alimentazione standard!

Inoltre, il ragionamento fatto fino ad ora avrebbe senso se tutte le tensioni scalassero contemporaneamente

Questo per esempio non avviene per le tensioni intrinseche

built in delle giunzioni pn, ampiezza banda proibita → grandezze non scalabili

Inoltre **scalare troppo** la tensione di soglia può far sì che diventi **complicato spegnere un MOS**, che può essere acceso in virtù del rumore sulle piste di silicio

Alternativa: Scalamento a tensione costante

In questo caso, si preferisce **mantenere la tensione di alimentazione costante**.

Vantaggio dal punto di vista della variazione degli standard

Anche questa tecnica presenta delle problematiche

Le dimensioni del MOSFET verranno comunque scalate (aumentare la possibilità di integrazione)

$$L' = \frac{L}{S} \quad W' = \frac{W}{S}$$

Ma le tensioni no:

$$V_{DS}' = V_{DS} \quad V_{GS}' = V_{GS} \quad V_{SB}' = V_{SB}$$

Alternativa: Scalamento a tensione costante

Il campo orizzontale risulterà amplificato

$$E_y' \approx \frac{V_{DS}'}{L'} = \frac{SV_{DS}}{L} = SE_y$$

In generale quando si opera questa scelta, **si decide comunque di diminuire lo spessore dell'ossido al fine di abbassare la tensione di soglia** (N:B: in tutte le tecniche di scaling si tende a diminuire lo spessore dell'ossido)

$$t_{OX}' = \frac{t_{OX}}{S}$$

Anche in questo caso si opera un **aumento del drogaggio localizzato alla regione di canale**, ma generalmente si aumenta **di un fattore S^2**

$$N_A' = S^2 N_A$$

Scalamento a tensione costante

Cosa succede al campo verticale?

$$E_x' \propto \frac{V_{GS}'}{t_{OX}'} = \frac{SV_{GS}}{t_{OX}} \Rightarrow E_x' = SE_x$$

Di conseguenza il campo verticale **aumenta di un fattore S**

L'aumento dei campi orizzontali e verticali induce però una serie di problematiche:

Rottura dell'ossido di gate

In genere la break down field dell'ossido è intorno ai 6 MV/cm

Scalamento a tensione costante

Cerchiamo ora di valutare l'effetto dello scalamento a tensione costante sugli altri parametri del dispositivo

$$K' = \frac{\mu_N C_{OX}' W'}{2L'} = \frac{\mu_N (S \cdot C_{OX}) W}{2L} = S \cdot K$$

Per quel che concerne la tensione di soglia dobbiamo ricordarci che anche in questo caso vengono comunque effettuate delle modifiche tecnologiche

$$t_{OX}' = \frac{t_{OX}}{S} \quad N_A' = S^2 N_A$$

Considerando che la tensione di body non viene alterata in seguito alle modifiche, si ottiene:

$$\gamma' = \frac{\sqrt{2\varepsilon_S q N_A'}}{C_{OX}'} = \frac{\sqrt{2\varepsilon_S q S^2 N_A}}{\varepsilon_{OX}} \frac{t_{OX}}{S} = \frac{\sqrt{2\varepsilon_S q N_A}}{C_{OX}} = \gamma$$

$$V_{TH}' \approx \gamma \sqrt{V_{SB}} = V_{TH}$$

Scalamento a tensione costante

Gli effetti della riduzione dello spessore dell'ossido t_{OX} di un fattore S e l'incremento del drogaggio di accettori N_A di un fattore S^2 (dovuto alla criticità della V_D non scalata) si controbilanciano e la tensione di soglia V_{TH} rimane la stessa dei dispositivi della precedente generazione.

$$I_D' = K'(V_{GS} - V_{TH})^2 = S \cdot K(V_{GS} - V_{TH})^2 = S \cdot I_D$$

A differenza di quanto osservato per lo scaling a campo costante, in questo caso la corrente viene amplificata di un fattore S

Cosa avviene per gli altri parametri significativi?

Scalamento a tensione costante

Area di gate

$$A_{GATE}' = W_N' L_N' + W_P' L_P' = \frac{W_N L_N}{S^2} + \frac{W_P L_P}{S^2} = \frac{A_{GATE}}{S^2}$$

Non abbiamo differenze rispetto al caso precedente, perché in questo parametro non entrano in gioco le tensioni

Tempo di propagazione

La capacità varia allo stesso modo, ma non scalano le tensioni!

$$C' = \frac{C}{S}$$

$$t_P' = \frac{C'}{K' (V_{DD} - V_{TH})^2} = \frac{C}{S} \frac{1}{S \cdot K} \frac{V_{DD} / 2}{(V_{DD} - V_{TH})^2} = \frac{t_P}{S^2}$$

Nel caso del full scaling, t_p diminuisce di un fattore S

Scalamento a tensione costante

In generale, se teniamo conto del fatto che

- Le capacità parassite scalano dello stesso fattore
- La corrente di uscita aumenta in questo caso

Con lo scaling a tensione costante si ottengono delle logiche più veloci

Vedremo però che il prezzo da pagare sarà un aumento della potenza dissipata

Scalamento a tensione costante

Frequenza di commutazione

Aumenta maggiormente rispetto al full scaling

$$f' = S^2 \cdot f$$

Potenza dissipata

$$P_D' = f' C' (V_{DD}')^2 = S^2 \cdot f \frac{C}{S} V_{DD}^2 = S \cdot P_D$$

Ricordiamo che nel caso del full scaling risultava ridotta di un fattore S^2 rispetto al caso del dispositivo non scalato!

Potenza dissipata per unità di area

$$p_D' = \frac{P_D'}{A_{GATE}'} = S \cdot P_D \frac{S^2}{A_{GATE}} = S^3 \cdot p_D$$

Notevole aumento!

Full scaling vs tensione costante

Nello “**scaling a campo costante**”, la riduzione della tensione di alimentazione V_{DD} **permette di:**

- **evitare incrementi nefasti del campo elettrico longitudinale e di quello verticale**
- **di ridurre la potenza dissipata** dall’invertitore elementare rispetto alla generazione precedente (la densità di potenza dissipata rimane immutata).
- **Tuttavia l’aumento della velocità operativa è limitato** (i circuiti di nuova generazione sono relativamente “lenti”)
- la tensione di soglia (ridotta di un fattore S) può diventare troppo bassa e si creano **problemi di compatibilità** (a causa della riduzione di V_{DD}) con i circuiti delle generazioni precedenti.

Full scaling vs tensione costante

Per applicazioni in cui **l'aumento di velocità** riveste importanza più elevata rispetto ai consumi, **conviene adottare lo “scaling a tensione costante”**;

Questo si paga con:

- **aumento significativo di potenza dissipata** per unità d'area nei circuiti “scalati”
- eventuali problemi connessi agli elevati campi elettrici nella regione di canale (elettroni caldi e degradazione della mobilità).

Quindi questa strategia porta a **circuiti veloci, ma con problemi di affidabilità e di consumi elevati.**