

Ottimizzazione Energy-Delay di un Buffer CMOS a Tre Stadi

Raffaele Petrolo

1 Obiettivo del Progetto

Il presente elaborato descrive la progettazione e l'ottimizzazione di un buffer CMOS a tre stadi per il pilotaggio di un carico capacitivo $C_L = 50 \cdot C_{in}$. L'obiettivo è individuare la frontiera di Pareto nello spazio Energy-Delay tramite calibrazione tecnologica, simulazione Monte Carlo e Sensitivity Analysis.

2 Caratterizzazione Tecnologica

In questa prima fase è stata eseguita la caratterizzazione dell'inverter di riferimento in tecnologia CMOS 100 nm. L'estrazione accurata dei parametri elettrici e temporali è propedeutica alla corretta calibrazione del modello analitico di ottimizzazione.

2.1 Bilanciamento dei ritardi

Il primo passo del progetto ha riguardato il dimensionamento della larghezza del canale del transistor PMOS (W_p) al fine di eguagliare i tempi di commutazione.

Fissata la taglia dell'NMOS al valore minimo consentito dalla tecnologia ($W_n = 0.12 \mu m$), è stata effettuata una simulazione parametrica su LTspice variando W_p . I dati estratti sono stati rielaborati per individuare il punto di pareggio tra il tempo di salita (t_{rise}) e il tempo di discesa (t_{fall}).

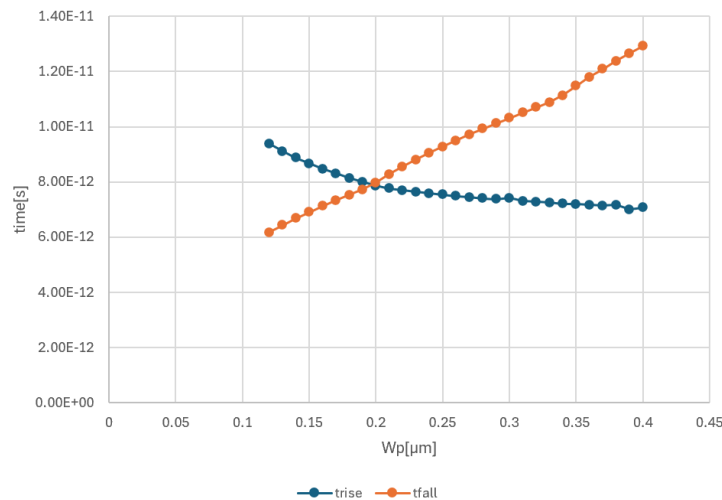


Figura 1: Andamento dei tempi di commutazione t_{rise} e t_{fall} in funzione di W_p .

Dall'analisi del grafico in Figura 1, si evince che per $W_p = 0.2 \mu m$ i due ritardi coincidono. Tale valore di larghezza definisce l'inverter unitario ($1\times$) che verrà utilizzato come riferimento per l'intera catena di buffer.

2.2 Estrazione dei parametri capacitivi (C_{in} e γ_E)

Per la calibrazione del modello energetico, è necessario determinare la capacità d'ingresso dell'inverter unitario (C_{in}) e la capacità parassita media di drain ($C_{d,avg}$). Il metodo utilizzato si basa sulla misura dell'energia dissipata dal generatore di alimentazione durante una commutazione completa, seguendo la relazione:

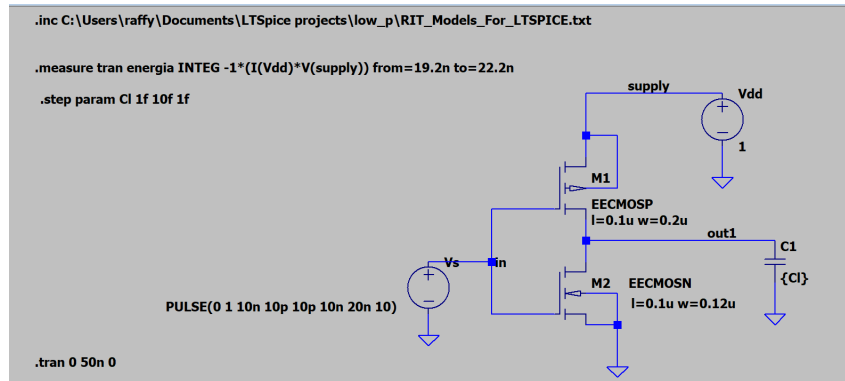
$$E_{diss} = C_{tot} \cdot V_{DD}^2 = (C_d + C_l) \cdot V_{DD}^2 \quad (1)$$

che per $V_{DD} = 1$ si riduce a

$$E_{diss} = C_d + C_l \quad (2)$$

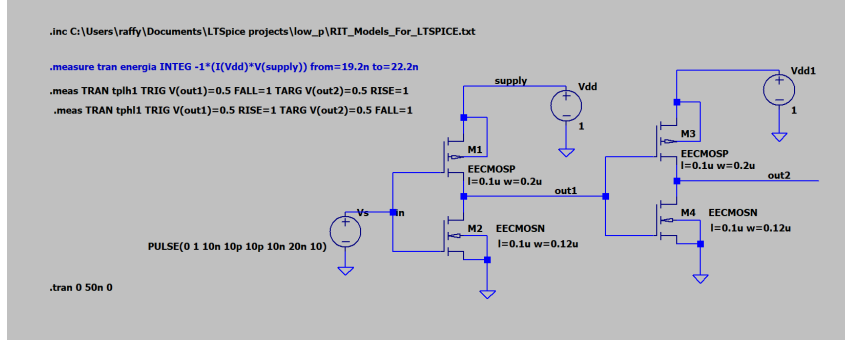
Il procedimento è stato articolato in due fasi di simulazione:

1. **Misura su carico noto:** È stata misurata l'energia dissipata dall'inverter mentre pilotava una capacità di carico nota $1fF < C_l < 10fF$. Sottraendo il contributo energetico del carico, è stata isolata la componente parassita media di drain dello stadio di driver pari a $C_{d,avg} = 0.602 fF$.



Cl	INTEG(-1*(FROM	TO	Cd	Cd_avg
1.00E-15	1.73E-15	1.92E-08	2.22E-08	7.30E-16	6.02E-16
2.00E-15	2.68E-15	1.92E-08	2.22E-08	6.76E-16	
3.00E-15	3.62E-15	1.92E-08	2.22E-08	6.24E-16	
4.00E-15	4.60E-15	1.92E-08	2.22E-08	5.97E-16	
5.00E-15	5.58E-15	1.92E-08	2.22E-08	5.82E-16	
6.00E-15	6.57E-15	1.92E-08	2.22E-08	5.72E-16	
7.00E-15	7.57E-15	1.92E-08	2.22E-08	5.65E-16	
8.00E-15	8.56E-15	1.92E-08	2.22E-08	5.61E-16	
9.00E-15	9.56E-15	1.92E-08	2.22E-08	5.58E-16	
1.00E-14	1.06E-14	1.92E-08	2.22E-08	5.55E-16	

2. **Misura con inverter di carico:** Sostituendo C_l con un inverter minimo identico, è stata misurata l'energia totale. Conoscendo già il contributo parassita del driver, si è ricavata per differenza la capacità di ingresso C'_{in} .



Dall'elaborazione dei dati sono stati ottenuti i seguenti parametri fondamentali:

- **Capacità d'ingresso:** $C_{in} = 0.73 \text{ fF}$
- **Fattore energetico γ_E :** Definito come il rapporto tra la capacità parassita e quella nominale d'ingresso:

$$\gamma_E = \frac{C_{d,avg}}{C_{in}} = 0.825 \quad (3)$$

2.3 Modello di ritardo e parametri temporali (τ_0 e γ_D)

Per completare la calibrazione, è necessario estrarre i parametri del modello di ritardo lineare di Elmore, espresso dalla relazione:

$$t_p = \tau_0 \left(1 + \frac{f}{\gamma_D} \right) \quad (4)$$

dove $f = C_{out}/C_{in}$ rappresenta il fan-out elettrico dello stadio. La procedura di estrazione ha previsto due simulazioni distinte del ritardo di propagazione medio ($t_p = \frac{t_{pLH} + t_{pHL}}{2}$):

1. **Misura a vuoto ($f = 0$):** È stato simulato l'inverter unitario senza alcun carico capacitivo collegato al nodo di uscita. In questa condizione, il ritardo misurato coincide con il ritardo intrinseco dovuto esclusivamente ai parassiti interni dello stadio.
 - **Risultato:** $\tau_0 = 7.92 \text{ ps}$.
2. **Misura con carico unitario ($f = 1$):** È stato simulato l'inverter mentre pilota un carico identico a se stesso (un inverter $1 \times$). Il ritardo risultante, t_{p1} , include sia il contributo intrinseco che quello del carico esterno.
 - **Risultato:** $t_{p1} = 20.6 \text{ ps}$.

Il coefficiente di sensibilità al ritardo γ_D è stato quindi ricavato analiticamente invertendo la formula del modello:

$$t_{p1} = \tau_0 + \frac{\tau_0}{\gamma_D} \implies \gamma_D = \frac{\tau_0}{t_{p1} - \tau_0} \quad (5)$$

Sostituendo i valori estratti:

$$\gamma_D = \frac{7.92}{20.6 - 7.92} \approx 0.627 \quad (6)$$

I parametri tecnologici finali che verranno utilizzati per la Sensitivity Analysis sono riassunti nella seguente tabella:

Parametro	Valore Estratto
Capacità d'ingresso C_{in}	0.73 fF
Fattore energetico γ_E	0.825
Ritardo intrinseco τ_0	7.92 ps
Coefficiente di delay γ_D	0.627

Tabella 1: Riepilogo dei parametri tecnologici.

3 Analisi Monte Carlo e Spazio di Progetto

Dopo aver caratterizzato l'inverter unitario, l'analisi è stata estesa al buffer a tre stadi per studiare il trade-off tra velocità e consumo energetico.

3.1 Configurazione della Simulazione Monte Carlo

Per generare una base dati significativa, è stata impostata una simulazione statistica di 10.000 run. Le taglie degli stadi intermedi sono state variate casualmente utilizzando la direttiva `mc(valore, tolleranza)` con una variazione del 75%:

- `.param X1 = {mc(4, 0.75)}` (Taglia del secondo stadio)
- `.param X2 = {mc(16, 0.75)}` (Taglia del terzo stadio)

3.2 Metodologia di misura dell'Energia

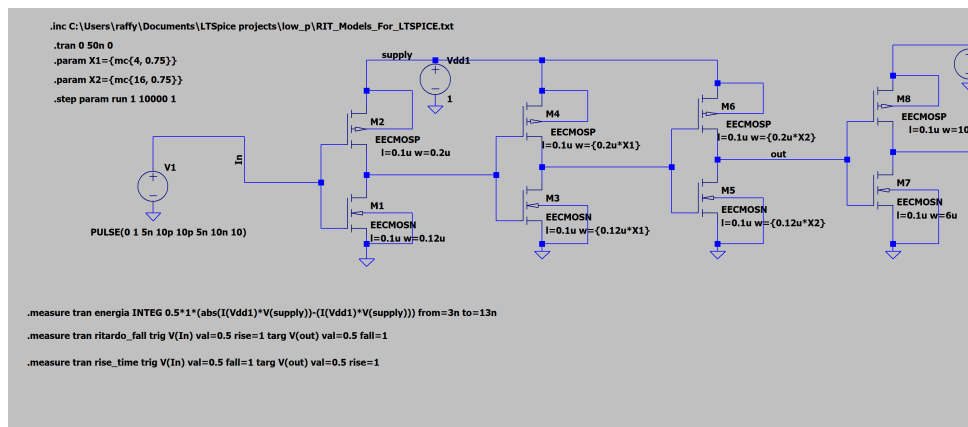
La misura dell'energia dissipata richiede particolare attenzione alla dinamica di commutazione degli stadi. La direttiva utilizzata in LTspice è:

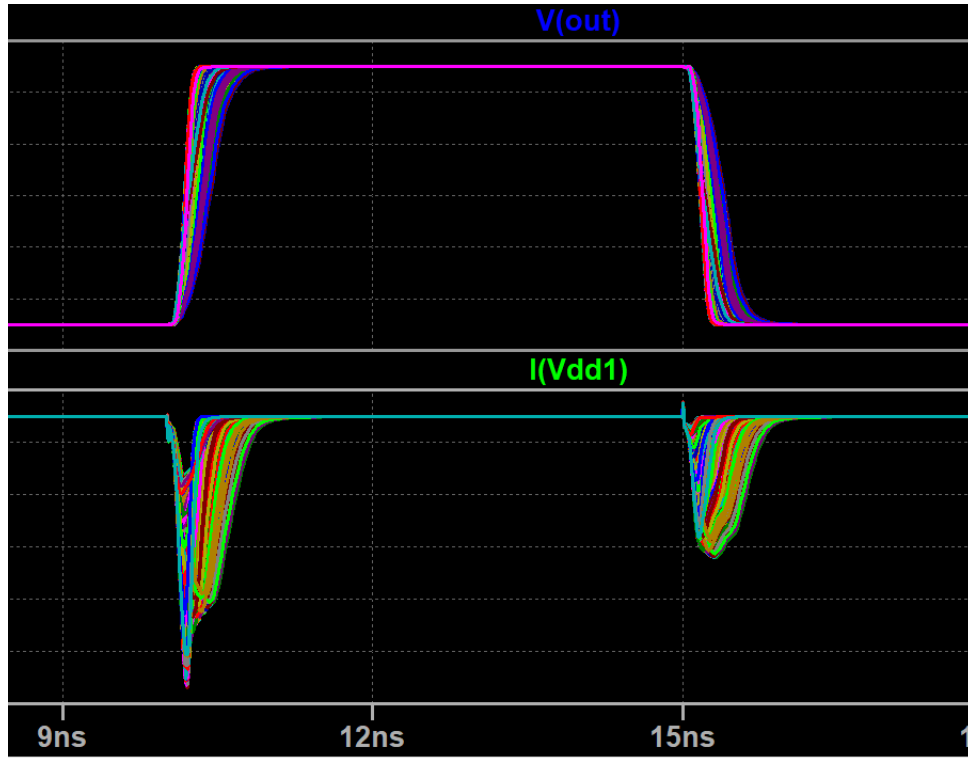
```
.measure tran energia INTEG 0.5*(abs(I(Vdd1)*V(supply))-(I(Vdd1)*V(supply)))  
from=3n to=13n
```

L'integrazione è stata eseguita su un intervallo temporale che comprende sia una transizione $0 \rightarrow 1$ che una $1 \rightarrow 0$ dell'ingresso. Questa scelta è dettata dalla struttura stessa del buffer a tre stadi:

- **Stadi 1 e 3:** Quando il primo inverter carica il suo nodo di uscita, anche il terzo stadio sta caricando il carico capacitivo C_L .
- **Stadio 2:** Opera in opposizione di fase rispetto ai precedenti, mentre il primo e il terzo stadio scaricano i rispettivi nodi, il secondo sta caricando la propria capacità di uscita.

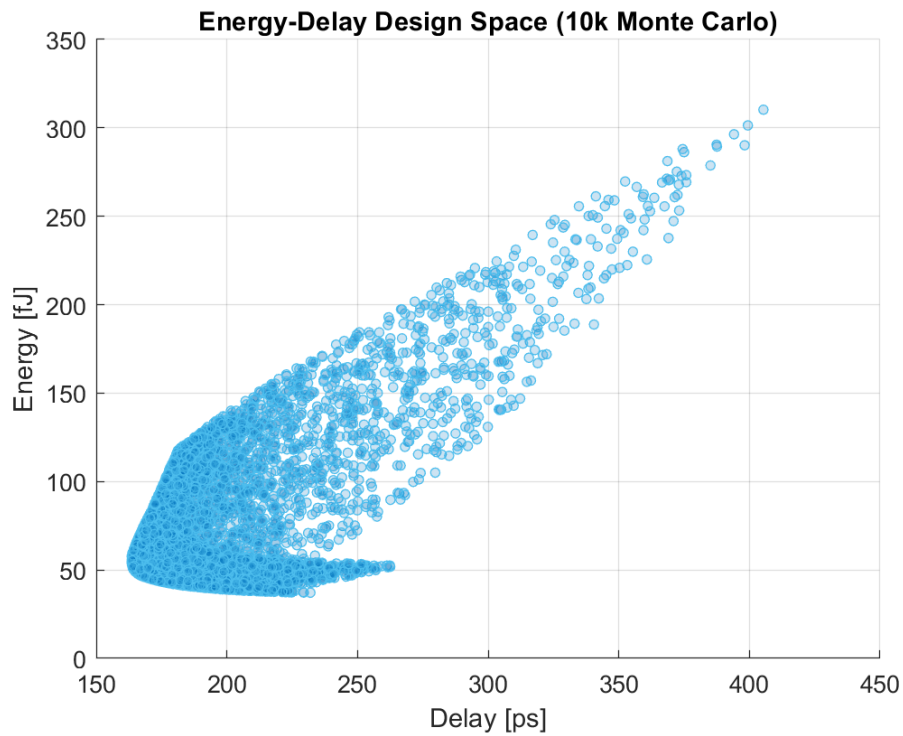
Per misurare correttamente l'energia totale è dunque necessario attendere il completamento di entrambi i fronti. Si osserva che sul fronte di discesa di V_{in} (che corrisponde al fronte di salita di V_{out} in figura) il picco di corrente assorbita è sensibilmente più alto rispetto al fronte successivo. Ciò accade perché il sistema deve caricare simultaneamente due capacità significative, di cui una è il carico esterno $C_L = 50 \cdot C_{in}$.





3.3 Analisi della Nuvola di Punti Energy-Delay

I risultati estratti dai 10.000 run sono stati processati in MATLAB per generare un plot nello spazio Energy-Delay.



Dalla figura si identifica chiaramente la **Frontiera di Pareto** come il limite inferiore sinistro della nuvola. Tale frontiera rappresenta il luogo dei punti a efficienza massima:

per ogni punto sulla curva, non è possibile ridurre il ritardo senza aumentare l'energia dissipata, definendo così il limite tecnologico del design ottimizzato.

4 Ottimizzazione e Validazione della Frontiera di Pareto

L'ultima fase del progetto riguarda la determinazione analitica delle taglie ottime X_1 e X_2 per minimizzare il consumo energetico fissato un vincolo di ritardo. Dal grafico della simulazione Monte Carlo, si osserva che l'intervallo di ritardo di interesse per i punti ottimali è compreso tra 150 ps e 250 ps.

4.1 Modello Analitico e Funzione di Ottimizzazione

Per trovare il miglior compromesso Energy-Delay, è stata utilizzata la funzione `fmincon` di MATLAB. Il problema è stato impostato definendo una funzione di vincolo `DelayConstraint` che impone il raggiungimento del ritardo target D_0 .

- **Funzione Obiettivo (Energia):**

$$E = V_{DD}^2 \cdot C_{min} [\gamma_E + (1 + \gamma_E)X_1 + (1 + \gamma_E)X_2 + 50] \quad (7)$$

- **Vincoli Non Lineari :** Il ritardo totale, espresso in picosecondi, è stato vincolato al valore D_0 tramite la somma dei contributi dei tre stadi:

$$\tau_0 \left(1 + \frac{X_1}{\gamma_D}\right) + \tau_0 \left(1 + \frac{X_2}{\gamma_D X_1}\right) + \tau_0 \left(1 + \frac{50}{\gamma_D X_2}\right) = D_0 \quad (8)$$

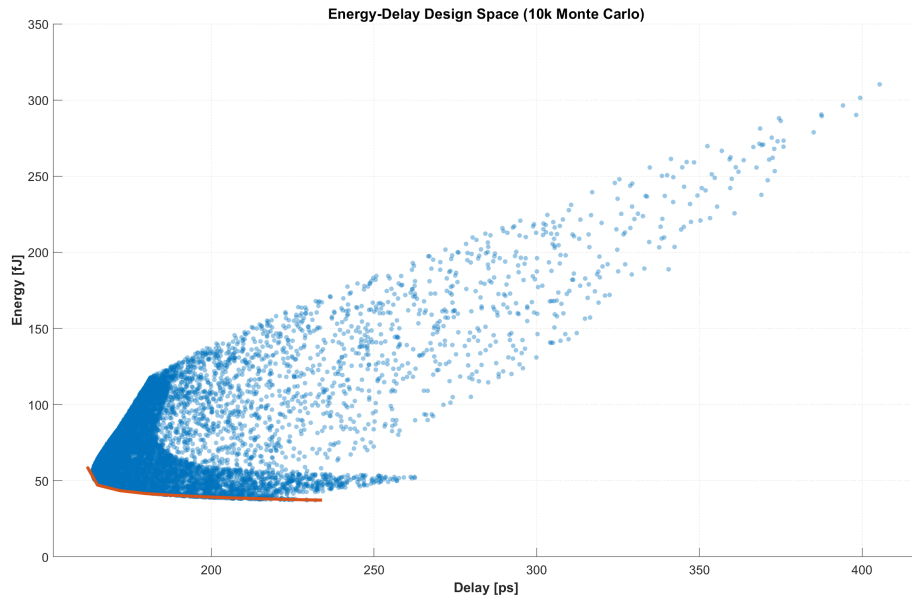
- **Vincoli di Disuguaglianza:** Per garantire la fattibilità fisica, si è imposto che la taglia del secondo stadio non sia inferiore a quella minima unitaria ($X_1 \geq 1$), tradotto nel codice come $1 - X_1 \leq 0$.

4.2 Risultati dell'Ottimizzazione e Validazione Spice

Di seguito sono riportati i risultati ottenuti dall'ottimizzazione in MATLAB (espressi come moltiplicatori dell'inverter minimo) e la relativa verifica effettuata in LTspice.

D_0 [ps]	X_1 (MATLAB)	X_2 (MATLAB)	Spice Delay [ps]	Energia [fJ]
150	3.69	13.60	162	58.5
160	3.68	13.60	162	58.5
170	2.70	9.05	165	47.1
180	2.21	7.29	172	43.4
190	1.90	6.29	180	41.5
200	1.67	5.60	189	40.3
210	1.49	5.07	198	39.4
220	1.35	4.66	206	38.7
230	1.23	4.32	216	38.1
240	1.12	4.03	225	37.6
250	1.04	3.78	234	37.2

Tabella 2: Confronto tra i parametri ottimi calcolati in MATLAB e i risultati della simulazione LTspice.



Plottando i valori ottenuti e interpolandoli (per via dei soli 10 campioni) si nota una eccellente corrispondenza tra il modello analitico e le simulazioni circuitali. E' presente una leggera discrepanza tra il target D_0 e il ritardo misurato in Spice (circa l'8% nel caso peggiore) imputabile sia al modello che a possibili approssimazioni effettuate da excel o da Spice stesso. Tuttavia, i punti ottimi si posizionano correttamente sulla frontiera di Pareto della nuvola Monte Carlo validando sia il modello che le tecniche utilizzate per la ricerca dei rispettivi parametri di fitting.