

Università della Calabria

Dipartimento di ingegneria informatica, modellistica,
elettronica e sistemistica



Corso di studio in
Ingegneria Elettronica

“Ottimizzazione energy-delay di un buffer CMOS a tre stadi”

Studente

Giuseppe Pirilli 237909

Anno Accademico 2023/2024

Indice

Introduzione.....	3
1 Struttura del buffer.....	4
2 Sizing inverter minimo.....	4
3 Calibrazione modelli di energia e ritardo del buffer	6
4 Simulazione Monte Carlo	8
4.1 Primo caso: carico x50 volte l'inverter minimo.....	9
4.2 Secondo caso: carico x100 volte l'inverter minimo.....	10
5 Sensitivity analysis	11
5.1 Primo caso: carico x50 volte l'inverter minimo.....	12
5.2 Secondo caso: carico x100 volte l'inverter minimo.....	13
6 Curva di Pareto ottima: metodo Monte Carlo vs sensitivity analysis	14
7 Conclusioni.....	16

Indice figure

Figura 1.1 - Schematic Buffer con inverter di carico variabile in termini di dimensioni	4
Figura 2.1 - Schematic inverter minimo in assenza di carico.....	4
Figura 2.2 - Rise delay e fall delay al variare di Wp	5
Figura 2.3 - Corrente dissipata dall'inverter minimo senza carico in un ciclo di carica-scarica	5
Figura 3.1 - Cascata d'inverter	6
Figura 3.2 - Schematic inverter minimo con inverter di carico.....	7
Figura 3.3 - Corrente dissipata dall'inverter minimo con inverter di carico variabile in un ciclo di carica-scarica.....	8
Figura 4.1 – Simulazione Monte Carlo: schematic buffer con inverter di carico x50 volte l'inverter minimo	9
Figura 4.2 – Simulazione Monte Carlo: curva di Pareto ottima empirica (carico x50)	9
Figura 4.3 – Simulazione Monte Carlo: corrente dissipata dal buffer in un ciclo di carica-scarica al variare del rapporto d'aspetto del secondo e del terzo inverter (carico x50).....	9
Figura 4.4 – Simulazione Monte Carlo: schematic buffer con inverter di carico x100 volte l'inverter minimo	10
Figura 4.5 – Simulazione Monte Carlo: curva di Pareto ottima empirica (carico x100)	10
Figura 4.6 – Simulazione Monte Carlo: corrente dissipata dal buffer in un ciclo di carica-scarica al variare del rapporto d'aspetto del secondo e del terzo inverter (carico x100).....	10
Figura 5.1 - Sensitivity analysis: schematic buffer con inverter di carico x50 volte l'inverter minimo	12
Figura 5.2 - Sensitivity analysis: corrente dissipata dal buffer in un ciclo di carica-scarica al variare del rapporto d'aspetto del secondo e del terzo inverter (carico x50).....	12
Figura 5.3 - Sensitivity analysis: Curva di Pareto ottima (carico x50)	13
Figura 5.4 - Sensitivity analysis: schematic buffer con inverter di carico x100 volte l'inverter minimo	13
Figura 5.5 - Sensitivity analysis: corrente dissipata dal buffer in un ciclo di carica-scarica al variare del rapporto d'aspetto del secondo e del terzo inverter (carico x100).....	13
Figura 5.6 - Sensitivity analysis: Curva di Pareto ottima (carico x100)	14
Figura 6.1 - Metodo Monte Carlo vs sensitivity analysis: curva di Pareto ottima (carico x50)	14
Figura 6.2 - Metodo Monte Carlo vs sensitivity analysis: curva di Pareto ottima (carico x100)	15
Figura 7.1 - Curva di Pareto ottima con inverter di carico dimensionato x50 e x100 volte l'inverter minimo	16

Introduzione

Il consumo di energia è diventato il principale problema di progettazione nei circuiti integrati CMOS di oggi. Per cercare di contrastare questo problema, sono state sviluppate svariate tecniche di progettazione “low power” che consentono di ottenere il miglior compromesso tra prestazioni ed efficienza energetica.

In questo elaborato lo studio viene rivolto alla progettazione e all’analisi di un buffer CMOS a tre stadi attraverso l’utilizzo del tool LTspice, con l’obiettivo di trovare il miglior compromesso energy-delay al variare del rapporto d’aspetto e, quindi delle dimensioni, del secondo e del terzo inverter, tramite simulazione Monte Carlo, rispetto alle dimensioni del primo inverter (dimensionato minimo) nei casi in cui l’ultimo stadio abbia come carico capacitivo un inverter dimensionato x50 e x100 volte l’inverter minimo, derivando la cosiddetta curva di Pareto ottima empirica, cioè l’involuppo dei punti nello spazio energy-delay ottenuti in seguito alla simulazione. La stessa curva di Pareto ottima viene ricavata utilizzando la metodologia della sensitivity analysis: attraverso un tool di ottimizzazione numerica in MATLAB e i modelli di energia e ritardo del buffer calibrati sulla tecnologia impiegata, in seguito al calcolo dei parametri caratteristici τ_{nom} , γ_d , γ_e e $C_{g,min}$, vengono ricavati, per un set opportuno di constraint di ritardo, i rapporti d’aspetto del secondo e del terzo inverter che minimizzano il consumo di energia. Questi, vengono utilizzati per simulare il buffer e ricavare i reali punti corrispondenti nello spazio energy-delay. L’insieme di questi punti costituisce la curva di Pareto ottima ricavata tramite la sensitivity analysis.

I risultati numerici vengono rappresentati graficamente attraverso l’utilizzo del tool MATLAB.

1 Struttura del buffer

Il buffer realizzato per questo elaborato è composto da tre inverter in cascata, dei quali il primo inverter viene dimensionato minimo, mentre gli altri due inverter sono caratterizzati da un rapporto d'aspetto variabile rispetto alle dimensioni dell'inverter minimo. L'ottimizzazione energy-delay viene eseguita per i casi in cui sul nodo d'uscita del buffer, anziché applicare un carico puramente capacitivo di valore noto, viene utilizzato come carico capacitivo un inverter dimensionato, rispettivamente, x50 e x100 volte l'inverter minimo.

La tensione di alimentazione, uguale per tutti e tre gli inverter, viene mantenuta fissa.

I transistor che compongono il circuito vengono modellati utilizzando un modello abbastanza semplice. Più è accurato il modello, maggiore sarà l'insieme degli effetti del secondo ordine di cui questo tiene conto.

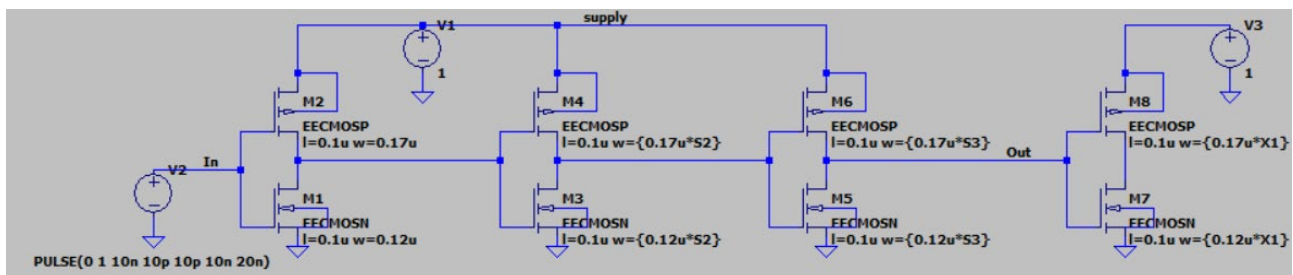


Figura 1.1 - Schematic Buffer con inverter di carico variabile in termini di dimensioni

2 Sizing inverter minimo

Nel dimensionare correttamente l'inverter minimo è necessario trovare il giusto rapporto d'aspetto tra il PMOS di pull-up e l'NMOS di pull-down, per il quale la conducibilità dei due transistor sia pressoché uguale. In altre parole, il tempo di rise (transizione L-H) deve essere pressoché uguale al tempo di fall (transizione H-L).

Fissata $L = 0.1 \mu m$ per entrambi i transistor e $W_{n,min} = 0.12 \mu m$, viene fatta variare $W_{p,min}$ in un certo range di valori.

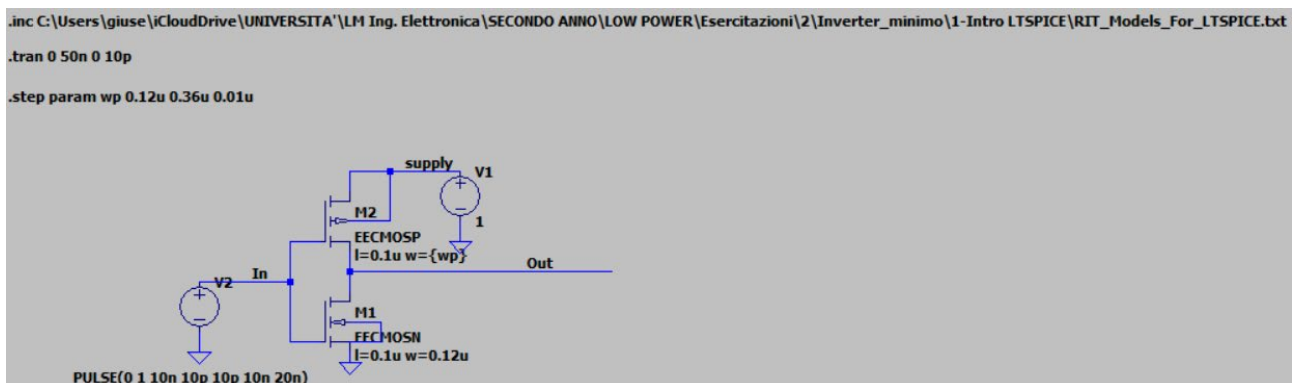


Figura 2.1 - Schematic inverter minimo in assenza di carico

Nella **fig. 2.2**, è possibile osservare che $\tau_{rise} \cong \tau_{fall}$ quando $W_{p,min} = 0.17 \mu m$.
In particolare, si ha che:

$$\tau_{rise} = 8.28371 \text{ ps}, \quad \tau_{fall} = 8.23174 \text{ ps}$$

Ciò equivale ad un rapporto d'aspetto pari a:

$$S = \frac{W_{p,min}}{W_{n,min}} = 1.4$$

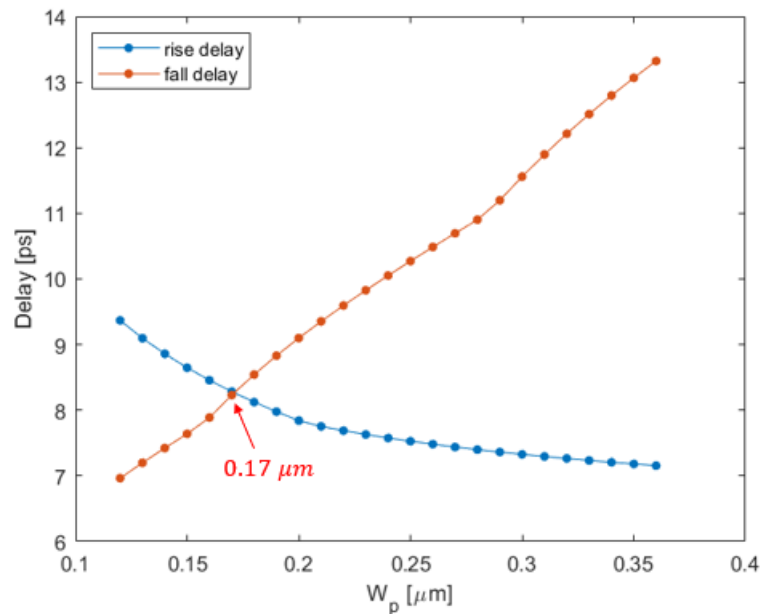


Figura 2.2 - Rise delay e fall delay al variare di W_p

Le forme d'onda riportate nella **fig. 2.3** mostrano il comportamento dell'inverter nelle due differenti fasi di commutazione (H-L e L-H).



Figura 2.3 - Corrente dissipata dall'inverter minimo senza carico in un ciclo di carica-scarica

Gli spike di tensione sono dovuti agli effetti capacitivi, mentre per quanto riguarda la corrente, è possibile distinguere due picchi:

- Picco positivo in fase di scarica (trascurabile), è riferito alla corrente entrante sul nodo d'uscita ed è dovuto all'accoppiamento capacitivo tra gate e source per il quale, durante il transitorio, viene fornita corrente al generatore di alimentazione.

- Picco negativo in fase di carica, è riferito alla corrente uscente dal nodo d'uscita ed è quella che effettivamente contribuisce alla carica della capacità di carico e, conseguentemente, al consumo di energia in un ciclo di carica-scarica.

3 Calibrazione modelli di energia e ritardo del buffer

In presenza di una cascata di inverter, come illustrato nella **fig. 3.1**, i modelli di energia e ritardo dell'inverter i-esimo sono definiti dalle seguenti relazioni:

$$E_i = V_{DD}^2 C_{L,i} = V_{DD}^2 (C_{int,i} + C_{g,i+1}) = V_{DD}^2 (\gamma_e C_{g,i} + S_{i+1} C_{g,min}) = V_{DD}^2 C_{g,min} (\gamma_e S_i + S_{i+1})$$

$$\tau_i = \tau_{nom} \left(1 + \frac{C_{g,i+1}}{\gamma_d C_{g,i}} \right) = \tau_{nom} \left(1 + \frac{S_{i+1}}{\gamma_d S_i} \right)$$

Con,

$$C_{g,i} = S_i C_{g,min}$$

$$C_{int,i} = \gamma_e C_{g,i}$$

Dove,

- $C_{g,i}$ rappresenta la capacità d'ingresso o di gate dell'inverter i-esimo;
- $C_{int,i}$ rappresenta la capacità intrinseca o di drain dell'inverter i-esimo;
- $C_{g,min}$ rappresenta la capacità d'ingresso o di gate dell'inverter minimo;
- S_i rappresenta il rapporto d'aspetto dell'inverter i-esimo;
- γ_e e γ_d rappresentano parametri caratteristici della tecnologia impiegata;
- τ_{nom} rappresenta il ritardo dell'inverter minimo in assenza di carico (ritardo intrinseco).

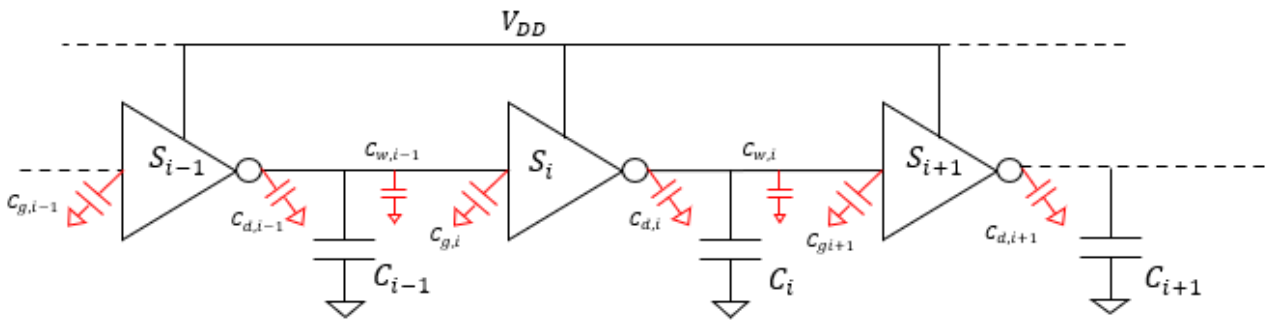


Figura 3.1 - Cascata d'inverter

Per un buffer a tre stadi si ha:

$$\begin{aligned} E_{tot} &= V_{DD}^2 C_1 + V_{DD}^2 C_2 + V_{DD}^2 C_3 = \\ &= V_{DD}^2 (\gamma_e C_{g,min} + S_2 C_{g,min}) + V_{DD}^2 (\gamma_e S_2 C_{g,min} + S_3 C_{g,min}) \\ &\quad + V_{DD}^2 (\gamma_e S_3 C_{g,min} + C_L) \end{aligned}$$

dove $C_L = K C_{g,min}$, con K pari a 50 e 100. Dunque, la relazione precedente può essere riscritta come:

$$E_{tot} = V_{DD}^2 C_{g,min} (\gamma_e + S_2 + \gamma_e S_2 + S_3 + \gamma_e S_3 + K)$$

I tre contributi che costituiscono E_{tot} sono relativi ad un ciclo di carica-scarica (transizione L-H-L), in particolare, due di questi contributi impattano sul consumo di energia nella transizione L-H, mentre il contributo restante impatta sul consumo di energia nella transizione H-L. Questo potrà essere apprezzato meglio in seguito alla simulazione Monte Carlo.

Allo stesso modo, il ritardo è definito come:

$$\tau_{tot} = \tau_{nom} \left(1 + \frac{S_2}{\gamma_d S_1}\right) + \tau_{nom} \left(1 + \frac{S_3}{\gamma_d S_2}\right) + \tau_{nom} \left(1 + \frac{K}{\gamma_d S_3}\right)$$

I modelli di energia e ritardo del buffer vengono calibrati sulla tecnologia impiegata calcolando i parametri τ_{nom} , γ_d , γ_e e $C_{g,min}$. Questi vengono determinati in riferimento all'inverter minimo.

Nota l'energia consumata dall'inverter minimo in assenza di carico è possibile calcolare la capacità intrinseca di quest'ultimo:

$$C_{int,min} = \frac{E_{min}}{V_{DD}^2} = \frac{0.512651 \cdot 10^{-15}}{1} = 0.512651 \text{ fF}$$

All'inverter minimo viene aggiunto come carico un altro inverter (vedi **fig. 3.2**) di cui viene fatto variare il rapporto d'aspetto rispetto alle dimensioni dell'inverter minimo, in modo da poter determinare la capacità d'ingresso media di quest'ultimo nota l'energia media consumata:

$$E_{min} = C_L V_{DD}^2 = (C_{int,min} + C_{g,inv2}) V_{DD}^2 = (C_{int,min} + S_2 C_{g,min}) V_{DD}^2$$

$$\Rightarrow C_{g,min} = \frac{E_{min}}{S_2 V_{DD}^2} - \frac{C_{int,min}}{S_2}$$

$$C_{g,min,avg} = 0.54915 \text{ fF}$$

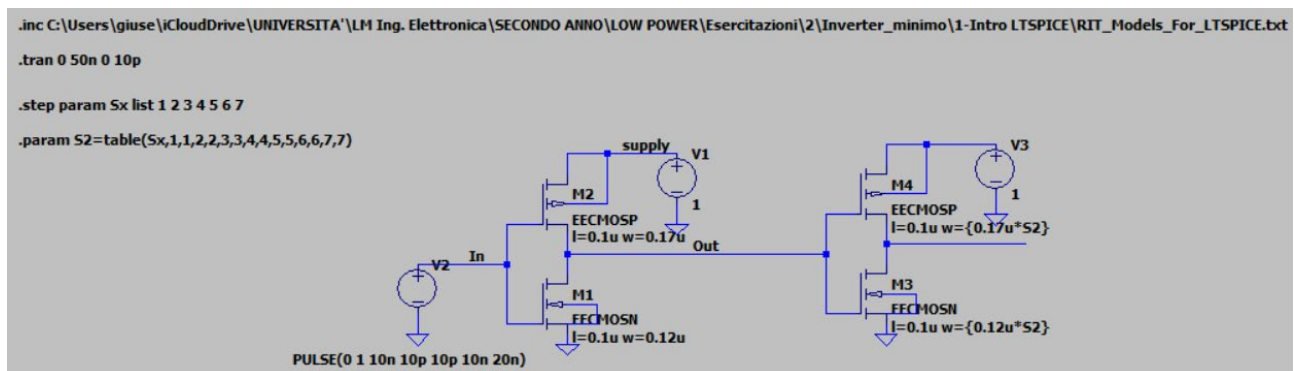


Figura 3.2 - Schematic inverter minimo con inverter di carico

Nella **fig. 3.3** è riportato il comportamento dell'inverter minimo nelle due differenti fasi di commutazione al variare delle dimensioni del carico.

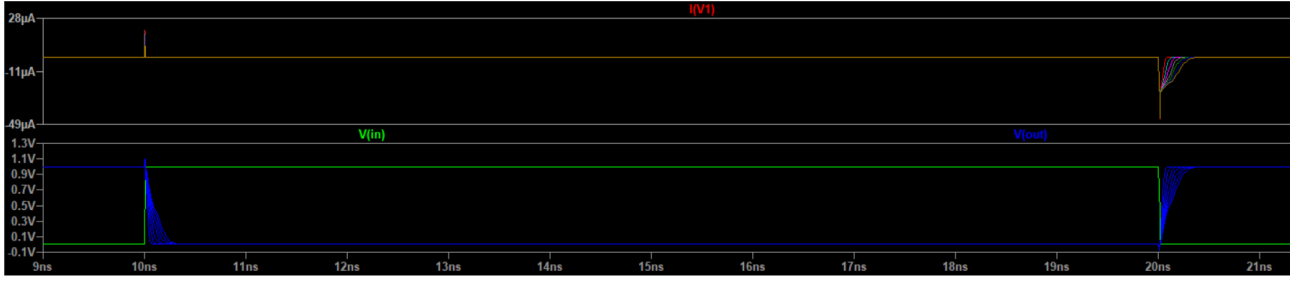


Figura 3.3 - Corrente dissipata dall'inverter minimo con inverter di carico variabile in un ciclo di carica-scarica

A questo punto, è possibile determinare il valore medio di γ_e :

$$\gamma_{e,avg} = \frac{C_{int,min}}{C_{g,min,avg}} = 0.93$$

Il ritardo intrinseco dell'inverter minimo è già noto ed è pari a:

$$\tau_{nom} = 8.28371 \text{ ps}$$

Il ritardo dell'inverter minimo, τ_{min} , viene determinato allo stesso modo della capacità d'ingresso, ovvero, applicando come carico un inverter di dimensioni variabili. In particolare, viene preso in considerazione il valor medio del ritardo:

$$\tau_{min,avg} = 57.993 \text{ ps}$$

A questo punto, è possibile determinare il valore medio di γ_d :

$$\tau_{min} = \tau_{nom} \left(1 + \frac{C_{g,inv2}}{\gamma_d C_{g,min}} \right) = \tau_{nom} \left(1 + \frac{S_2}{\gamma_d} \right)$$

$$\Rightarrow \gamma_d = \frac{S_2 \tau_{nom}}{(\tau_{min} - \tau_{nom})}$$

$$\gamma_{d,avg} = 0.67$$

4 Simulazione Monte Carlo

Tramite la simulazione Monte Carlo vengono fatti variare i rapporti d'aspetto del secondo e del terzo inverter rispetto alle dimensioni dell'inverter minimo, nei casi in cui l'inverter di carico è dimensionato, rispettivamente, x50 e x100 volte l'inverter minimo, derivando la curva di Pareto ottima empirica dall'involuppo dei punti energy-delay. Questa rappresenta il miglior compromesso energy-delay.

4.1 Primo caso: carico x50 volte l'inverter minimo

Nella **fig. 4.1** viene illustrato il caso in cui l'inverter di carico viene dimensionato x50 volte l'inverter minimo.

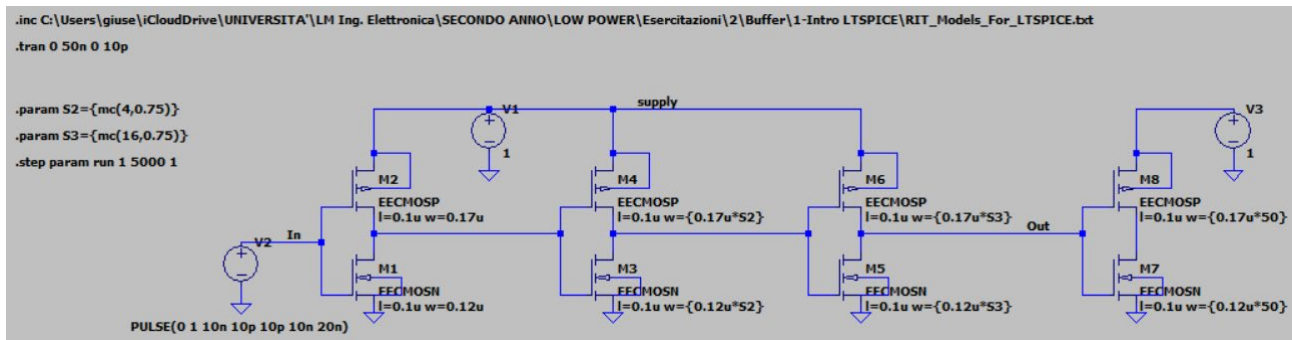


Figura 4.1 – Simulazione Monte Carlo: schematic buffer con inverter di carico x50 volte l'inverter minimo

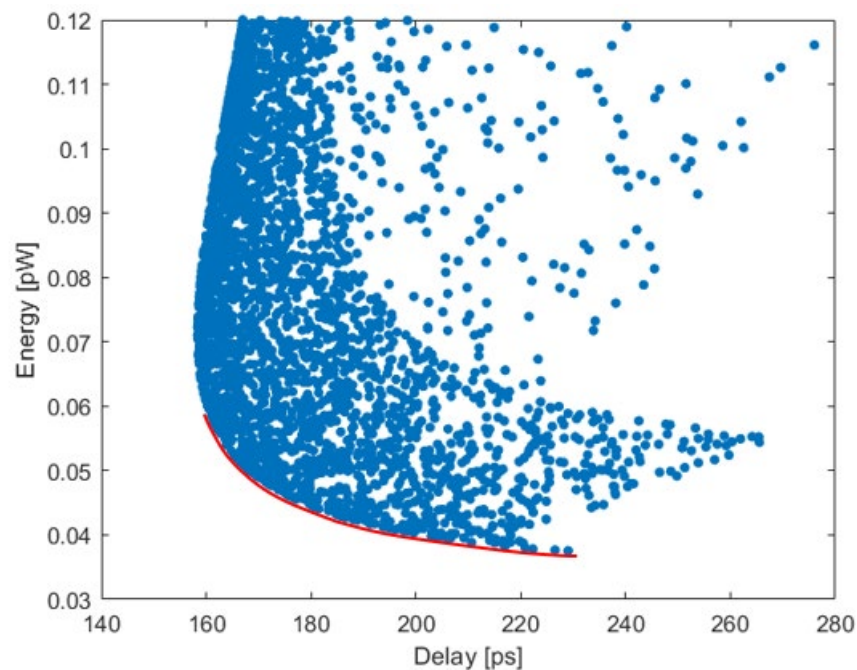


Figura 4.2 – Simulazione Monte Carlo: curva di Pareto ottima empirica (carico x50)



Figura 4.3 – Simulazione Monte Carlo: corrente dissipata dal buffer in un ciclo di carica-scarica al variare del rapporto d'aspetto del secondo e del terzo inverter (carico x50)

4.2 Secondo caso: carico x100 volte l'inverter minimo

Nella **fig. 4.4** viene illustrato il caso in cui l'inverter di carico viene dimensionato x100 volte l'inverter minimo.

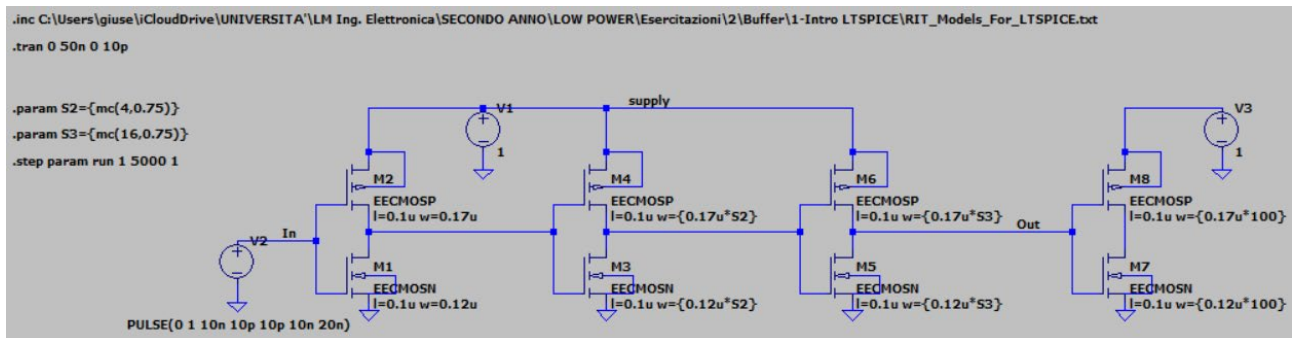


Figura 4.4 – Simulazione Monte Carlo: schematic buffer con inverter di carico x100 volte l'inverter minimo

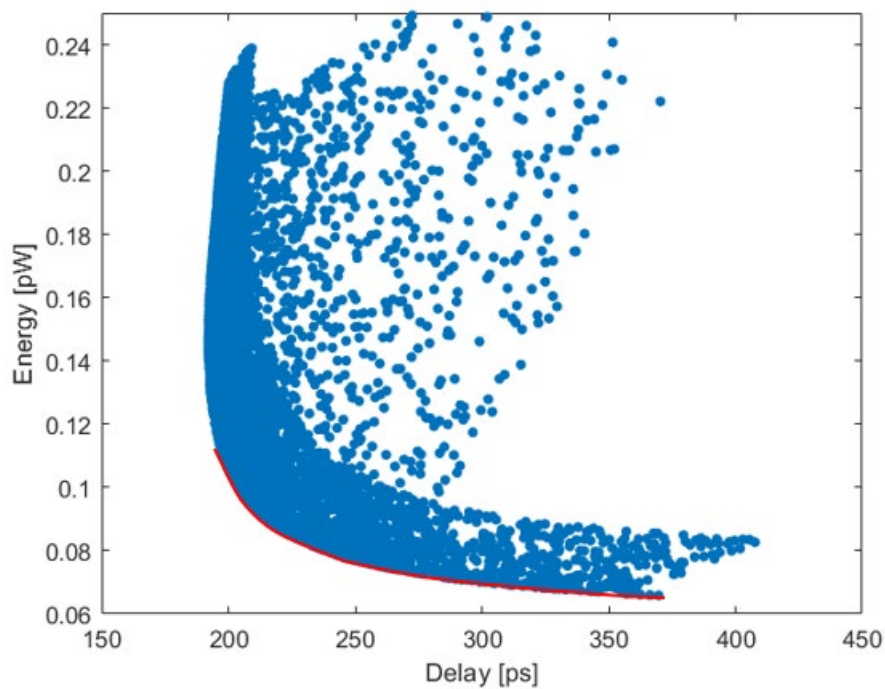


Figura 4.5 – Simulazione Monte Carlo: curva di Pareto ottima empirica (carico x100)

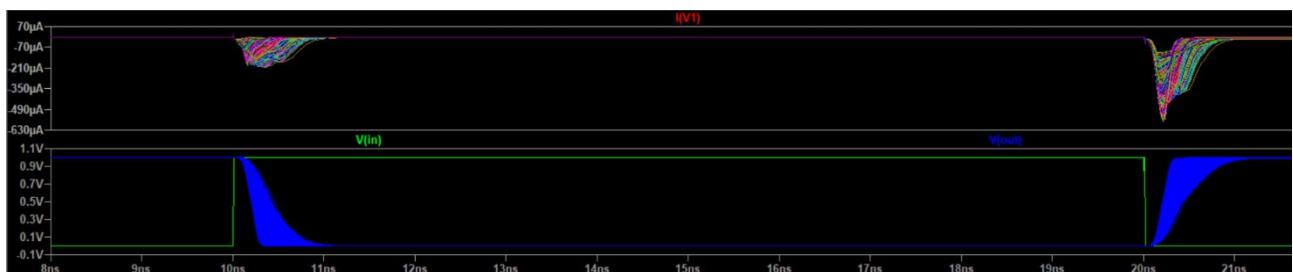


Figura 4.6 – Simulazione Monte Carlo: corrente dissipata dal buffer in un ciclo di carica-scarica al variare del rapporto d'aspetto del secondo e del terzo inverter (carico x100)

Dalle **fig. 4.2** e dalla **fig. 4.5** risulta evidente che, aumentando le dimensioni dell'inverter di carico e, conseguentemente la sua capacità di gate, si ha un incremento del consumo energetico a parità di velocità.

Le **fig. 4.3** e **4.6** mostrano che il maggior consumo di energia si verifica durante la commutazione L-H dell'uscita del buffer. Ciò è dovuto alla carica della capacità sul nodo d'uscita del primo e del terzo inverter, mentre nella commutazione H-L si verifica solo la carica della capacità del secondo inverter. Questi rappresentano i tre contributi di energia consumata ricavati precedentemente con il modello dell'energia.

5 Sensitivity analysis

I modelli di energia e ritardo del buffer calibrati nella **sez. 3** vengono utilizzati per ricavare, attraverso l'utilizzo di un tool di ottimizzazione in MATLAB, per un opportuno set di constraint di ritardo, i valori dei rapporti d'aspetto del secondo e del terzo inverter che minimizzano il consumo di energia nei due differenti casi (inverter di carico dimensionato x50 e x100 volte l'inverter minimo).

Inverter di carico x50			Inverter di carico x100		
D_0 [ps]	S_{inv2}	S_{inv3}	D_0 [ps]	S_{inv2}	S_{inv3}
-	-	-	360	1.1803	4.5227
-	-	-	350	1.1943	4.7177
-	-	-	340	1.2555	4.9000
-	-	-	330	1.2833	5.1490
-	-	-	320	1.3291	5.3975
-	-	-	310	1.3850	5.6696
-	-	-	300	1.4518	5.9698
-	-	-	290	1.5698	6.3041
-	-	-	280	1.6686	6.6812
-	-	-	270	1.7831	7.1125
-	-	-	260	1.9166	7.6136
-	-	-	250	1.4337	6.2274
220	1.3614	4.4728	240	1.5005	6.4577
215	1.4430	4.6688	230	2.5092	9.7705

210	1.4584	4.9286	220	2.8050	11.0139
205	1.5793	5.1312	210	3.1848	12.9346
200	1.6495	5.3272	200	3.9039	16.7312
195	1.7419	5.5756	-	-	-
190	1.8771	6.0526	-	-	-
185	1.8070	5.9402	-	-	-
180	2.1582	6.9844	-	-	-
175	2.3322	7.6302	-	-	-
170	2.5916	8.5409	-	-	-
165	2.9426	9.9901	-	-	-
160	3.6741	13.4065	-	-	-

Questi valori vengono utilizzati per simulare il buffer e ricavare i reali punti corrispondenti nello spazio energy-delay come riportato nelle **sotto-sez. 5.1** e **5.2**. L'insieme di questi punti costituisce la curva di Pareto ottima ricavata tramite la sensitivity analysis.

5.1 Primo caso: carico x50 volte l'inverter minimo

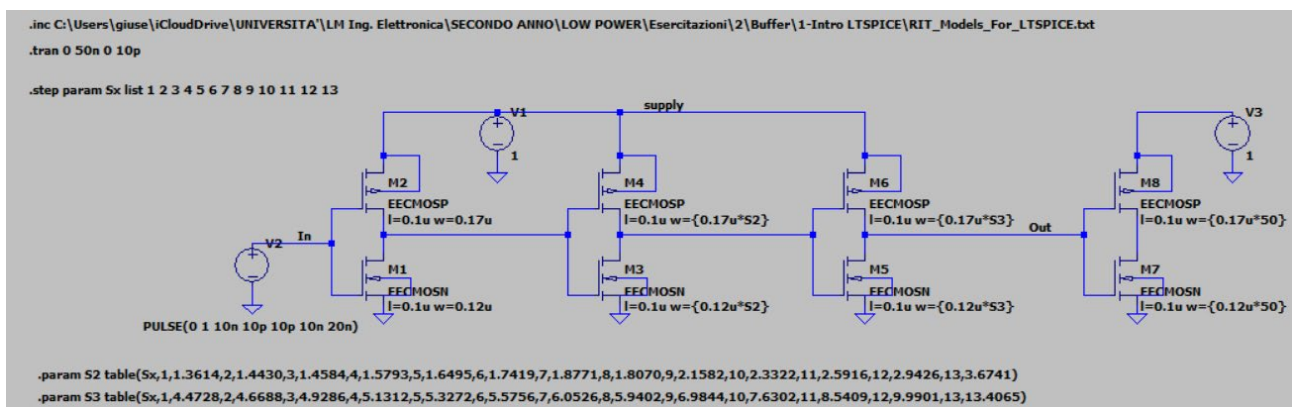


Figura 5.1 - Sensitivity analysis: schematic buffer con inverter di carico x50 volte l'inverter minimo

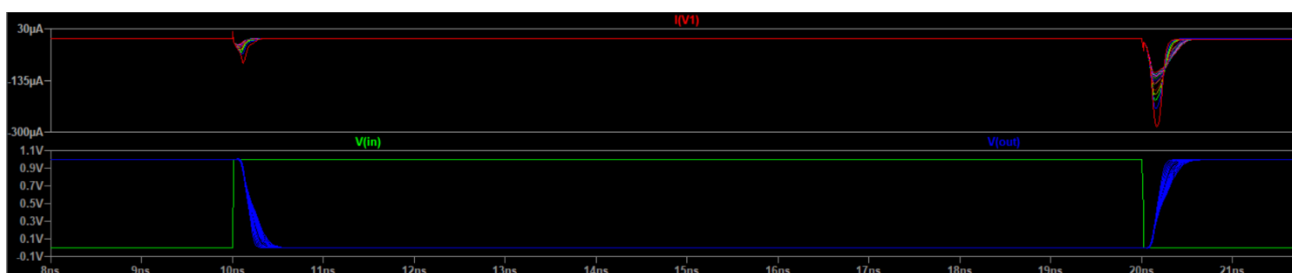


Figura 5.2 - Sensitivity analysis: corrente dissipata dal buffer in un ciclo di carica-scarica al variare del rapporto d'aspetto del secondo e del terzo inverter (carico x50)

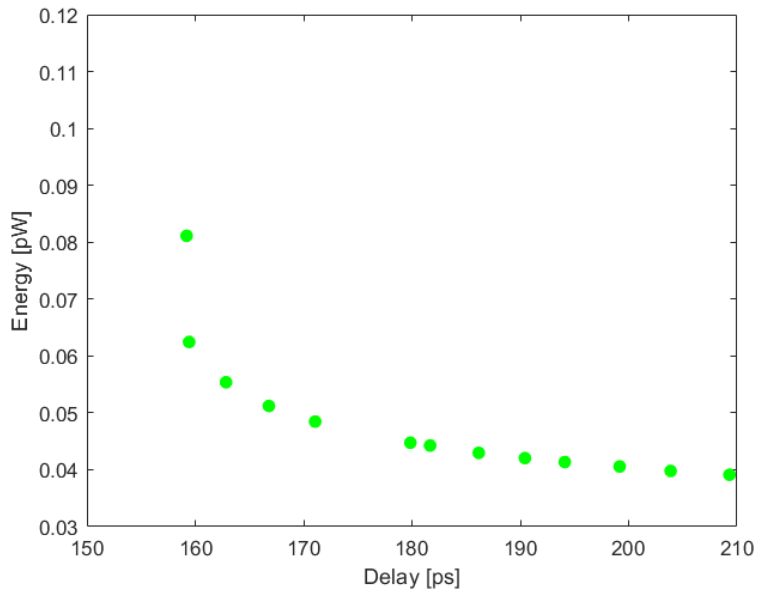


Figura 5.3 - Sensitivity analysis: Curva di Pareto ottima (carico x50)

5.2 Secondo caso: carico x100 volte l'inverter minimo

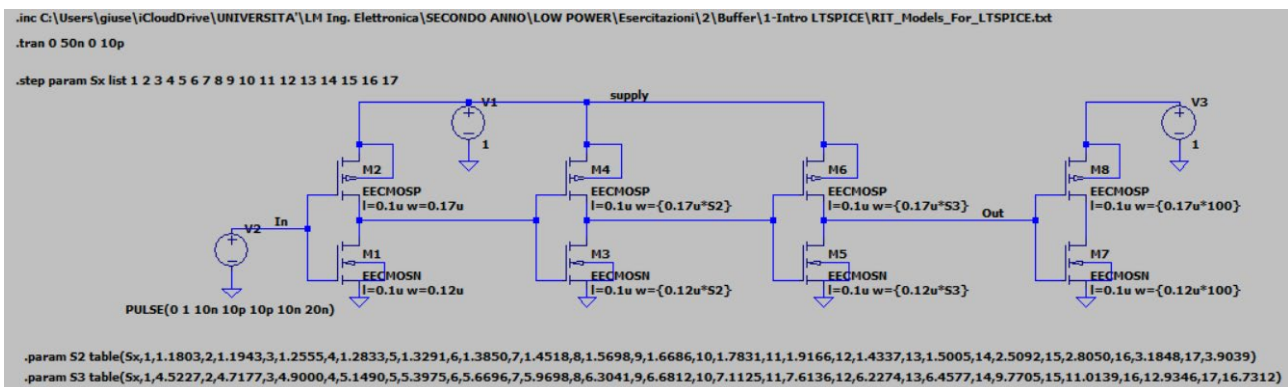


Figura 5.4 - Sensitivity analysis: schematic buffer con inverter di carico x100 volte l'inverter minimo

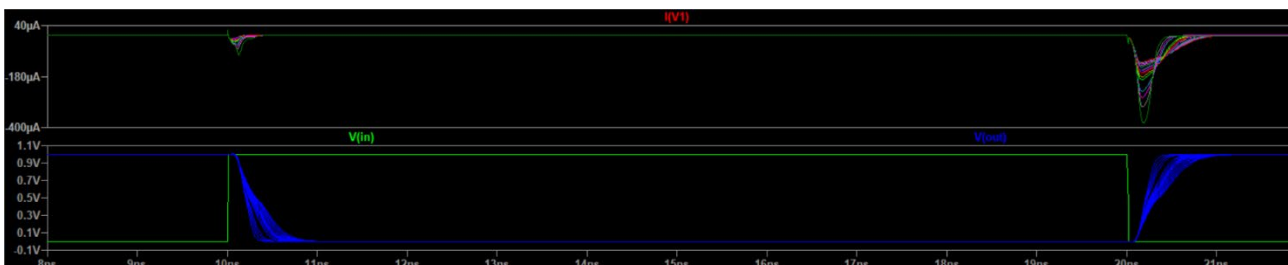


Figura 5.5 - Sensitivity analysis: corrente dissipata dal buffer in un ciclo di carica-scarica al variare del rapporto d'aspetto del secondo e del terzo inverter (carico x100)

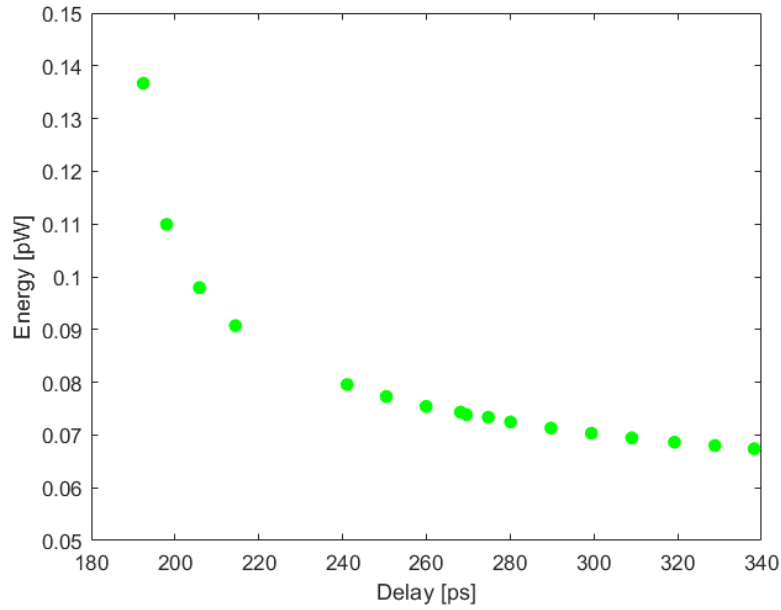


Figura 5.6 - Sensitivity analysis: Curva di Pareto ottima (carico x100)

6 Curva di Pareto ottima: metodo Monte Carlo vs sensitivity analysis

La curva di Pareto ottima ottenuta con la metodologia della sensitivity analysis viene confrontata con la curva di Pareto ottima empirica ottenuta dalla simulazione Monte Carlo. I risultati sono illustrati nelle **fig. 6.1** e **6.2** per i casi in cui l'inverter di carico applicato sul nodo d'uscita del buffer sia dimensionato, rispettivamente, x50 e x100 volte l'inverter minimo. Dai grafici è possibile osservare come entrambe i metodi di analisi convergono al medesimo risultato.

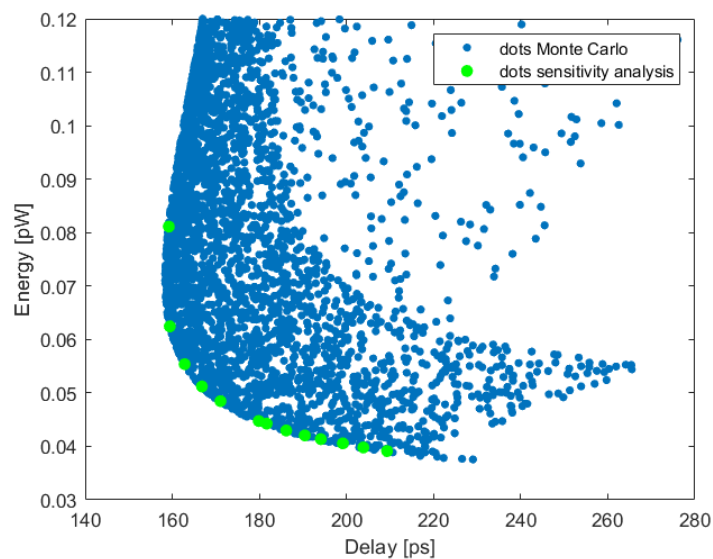


Figura 6.1 - Metodo Monte Carlo vs sensitivity analysis: curva di Pareto ottima (carico x50)

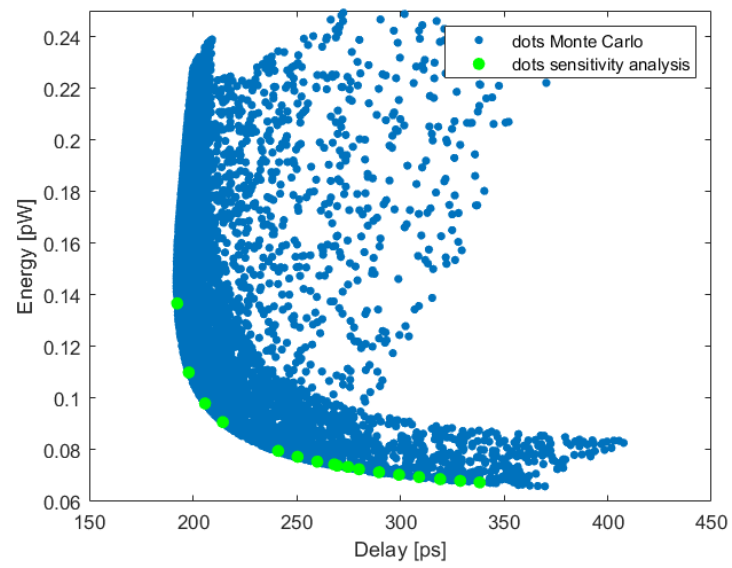


Figura 6.2 - Metodo Monte Carlo vs sensitivity analysis: curva di Pareto ottima (carico x100)

7 Conclusioni

Agire sulle dimensioni degli inverter è una tra le tante soluzioni per minimizzare il consumo di energia per un dato constraint di ritardo, ottenendo il miglior compromesso energy-delay. Ovviamente, utilizzare inverter più prestazionali, in termini di velocità, correttamente dimensionati, comporta un maggior consumo di energia. Viceversa, se vengono ridotte le dimensioni. Un'altra osservazione importante riguarda l'inverter di carico applicato sul nodo d'uscita del buffer: aumentare le dimensioni dell'inverter di carico rispetto all'inverter minimo e, quindi la sua capacità di gate, comporta una traslazione positiva della curva di Pareto ottima, ovvero, un peggioramento del compromesso energy-delay (vedi **fig 7.1**).

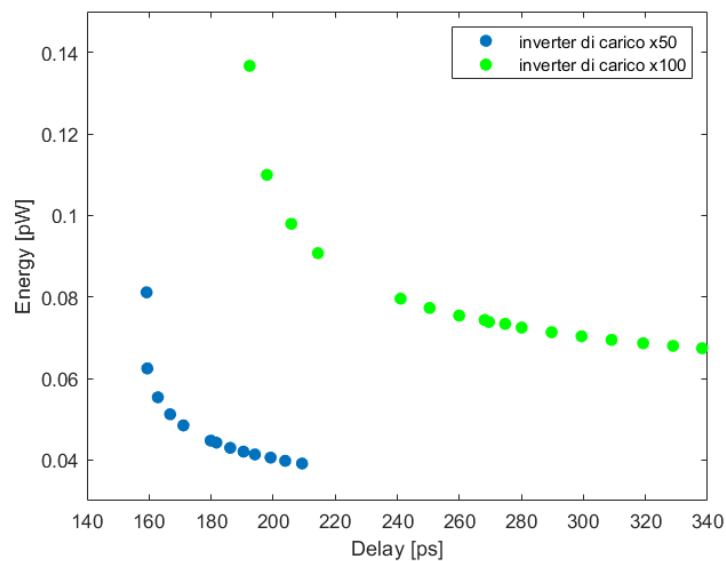


Figura 7.1 - Curva di Pareto ottima con inverter di carico dimensionato x50 e x100 volte l'inverter minimo