**“Progettazione Low Power”**

**2023-2024**

**Docente: Prof. Fabio FRUSTACI**

**<energy-delay optimization of a three-stages CMOS buffer>**

|  |  |
| --- | --- |
| **Data** | <18/10/2023> |
| **Documento** | Documento Finale – D3 |

|  |  |  |
| --- | --- | --- |
| **Nome e Cognome** | **Matricola** | **E-mail address** |
| Giorgio Ubbriaco | 247284 | bbrgrg00h11d086x@studenti.unical.it |

Sommario

[Abstract 1](#_Toc151028511)

[Stato dell’Arte 2](#_Toc151028512)

[Obiettivi 4](#_Toc151028513)

[Analisi e progettazione di un inverter CMOS minimo 5](#_Toc151028514)

[Analisi e progettazione di un buffer CMOS a tre stadi 6](#_Toc151028515)

[Conclusioni 7](#_Toc151028516)

# Vincoli di progetto

1. Il buffer deve avere tre stadi;
2. Il primo stadio è un inverter dimensionato minimo;
3. L’ultimo stadio ha come carico capacitivo un inverter dimensionato 50x l’inverter minimo;
4. Le variabili indipendenti che possono essere usate per l’ottimizzazione sono il rapporto d’aspetto del secondo e del terzo inverter rispetto alle dimensioni dell’inverter minimo;

# Task da compiere

1. Trovare il dimensionamento dell’inverter minimo;
2. Calibrare il modello di energia e ritardo del buffer sulla tecnologia utilizzata. In particolare, trovare le seguenti costanti: gamma\_delay, gamma\_energia, tau\_0, capacità di ingresso dell’inverter minimo;
3. Considerando il rapporto d’aspetto del secondo e del terzo inverter rispetto alle dimensioni dell’inverter minimo come variabili da poter settare nel processo di ottimizzazione, ricavare la curva di Pareto energy-delay in maniera empirica, cioè ricavando l’inviluppo dei punti di design nello spazio Energy-Delay ricavati tramite simulazione Monte Carlo.
4. Ricavare la stessa curva di Pareto utilizzando la metodologia della sensitivity analysis: attraverso un tool di ottimizzazione numerica e i modelli di energia e ritardo di cui sopra, ricavare, per un set opportuno di constraint di ritardo, i rapporti di aspetto che minimizzano la dissipazione di energia dinamica. In questa maniera, si otterrà, per ognuno dei constraint di delay utilizzati, una coppia di dimensionamenti;
5. Per ognuna delle coppie trovate al punto precedente, simulare il buffer con LTspice per ricavare i reali punti di design corrispondenti nello spazio Energy-Delay. L’insieme di tali punti costituisce la curva di Pareto ottima ricavata tramite la sensitivity analysis. Verificare che tale curva sia pressoché coincidente con la curva ricavata per via empirica.

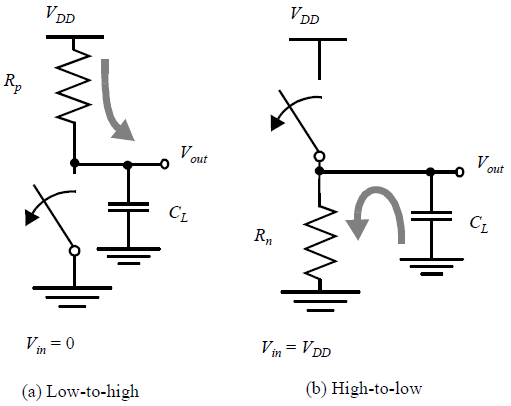
# Abstract

La parte iniziale del progetto ha previsto la progettazione di un'inverter chain (buffer) a 3 stadi: il primo stadio composto da un inverter di dimensioni minime, il secondo ed il terzo stadio composti da un inverter ciascuno e rispettivamente dimensionati S1 ed S2 volte l'inverter minimo. Nello specifico, tali parametri di dimensionamento sono stati ottenuti considerando il Metodo di Monte Carlo così da farci ottenere una certa randomicità durante le N\_RUNS simulazioni condotte. In uscita al buffer è stato considerato un carico rappresentato da un inverter dimensionato S\_Load volte l'inverter minimo. Nello specifico, è stato analizzato il comportamento del buffer considerando una certa tensione V\_IN, di tipologia pulse avente determinati parametri, una tensione V\_Supply per i 3 stadi di inverter ed una V\_Supply\_L per l'inverter di carico. Dopo aver effettuato delle simulazioni riguardo il circuito in questione, è stata calcolata l'energia associata all'inverter chain considerando le due transizioni possibili: 0->1 (da 13 ns a 18 ns) e 1->0 (da 18 ns a 22 ns). Inoltre, sono stati calcolati i ritardi di rise e di fall considerando rispettivamente fronti di salita e di discesa del segnale. Bisogna specificare che nelle analisi è stato considerato l'inverter minimo isolato così da poter calcolare la capacità totale (collegando l'inverter minimo dal carico), la capacità di uscita intrinseca (scollegando l'inverter minimo dal carico) e, infine, la capacità di ingresso (minima) ottenuta come differenza tra le due appena citate. Pertanto, dopo aver calcolato i parametri caratteristici del buffer, quali gamma\_e, gamma\_d, tau\_nom e c\_min, è stato possibile passare alla successiva fase del progetto: l'ottimizzazione dei fattori di dimensionati S1 ed S2 del circuito. Nello specifico, considerando un algoritmo di ottimizzazione non lineare e tenendo conto di una funzione obiettivo rappresentata dal modello dell'energia di un'inverter chain, e di constraint di uguaglianza e diseguaglianza ricavati dal modello del delay, è stato possibile ottenere valori ottimi considerando un delay range tra d\_max e d\_min a step di 10. Successivamente, i valori ottimi, ottenuti dall'algoritmo appena citato, sono stati utilizzati all'interno dello schematic buffer\_optimized per simulare nuovamente il circuito dell'inverter chain e ricavarne i parametri di delay ed energia. Pertanto, infine, è stato possibile effettuare un'analisi comparativa tra gli esperimenti condotti durante la parte iniziale del progetto, ottenuti tramite Metodo Monte Carlo, e la curva di Pareto ottima, ottenuta tramite processo di ottimizzazione e simulazione del circuito.

# Stato dell’Arte

L'invertitore è veramente il fulcro di tutti i progetti digitali. Una volta compreso chiaramente il suo funzionamento e le sue caratteristiche, la progettazione di strutture più complesse come porte NAND, sommatori, moltiplicatori e microprocessori diventa molto più semplice. Il comportamento elettrico di questi circuiti complessi può essere in gran parte dedotto extrapolando i risultati ottenuti dagli invertitori.

Immagine che contiene diagramma, Disegno tecnico, linea, Piano

Descrizione generata automaticamente

Nello specifico, l’analisi di questi circuiti prevede una comprensione dei costi, espressi in termini di complessità e area occupazionale, integrità e robustezza, espressa dal comportamento statico dello stesso, performance, determinata dalla risposta dinamica, e dall’efficienza energetica, stabilita dal consumo dell’energia e della potenza.

La capacità di ingresso di una porta logica, negli odierni circuiti integrati, ha un valore molto basso e il ritardo di propagazione di una porta CMOS che pilota una capacità di carico così limitata può essere al di sotto di . È comune il caso in cui le capacità associate a determinate architetture risultano essere più elevate. Ad esempio, le wordline di una RAM o di una ROM costituiscono un carico capacitivo notevole per circuiti di questo genere. Bisogna tenere presente che, il ritardo di una porta CMOS è proporzionale alla capacità di carico. Pertanto, considerando un invertitore ad area minima, si avrebbe un ritardo considerevolmente grande per pilotare capacità così elevate. Sarebbe logico pensare di aumentare le dimensioni del CMOS per diminuirne il delay associato ma questo comporterebbe un aumento della capacità di ingresso dello stesso. Quest’ultima conseguenza, inoltre, porterebbe ad aumento del ritardo di propagazione dello stadio che pilota l’invertitore. Quindi, quello che si dovrebbe fare è minimizzare il ritardo complessivo sfruttando qualche aspetto che non comporta il peggioramento di altri parametri circuitali.

Durante gli anni sono stati effettuati diverse ricerche a riguardo dimostrando che è possibile minimizzare il delay complessivo utilizzando una inverter chain (buffer), cioè diversi stadi invertitori, in modo da formare il cosiddetto buffer. Se si riesce ad ottimizzare il buffer evitando problemi di dimensionamento, si sta praticamente minimizzando l’energia di accesso alla memoria. Tanto è vero che, in questo caso, si parla di problemi di energia dinamica e l’obiettivo sarà la diminuzione della stessa. Un ipotetico approccio potrebbe essere quello di scegliere una tensione di soglia differente per ogni stadio di inverter. Pertanto, fissando un determinato sizing ma variando il parametro di soglia, ciò che comporta è l’ottimizzazioen dell’energia statica poiché la corrente di sottosoglia dipende dalla . Bisogna notare, inoltre, che, siccome questo buffer viene ripetuto per ogni array di celle di memoria, la corrente (potenza) di leakage che viene risparmiata dovrà essere moltiplicata per il numero di array di celle considerato per una determinata memoria. Quindi, non ha senso utilizzare come parametro di ottimizzazione la tensione di soglia poiché la potenza dinamica risulterebbe essere di gran lunga superiore alla potenza di leakage.

Nella progettazione circuitale viene solitamente considerata una tecnica, denominata logical effort, che permette di stimare il ritardo in un circuito CMOS. Nel caso di un inverter chain, ciò che si fa è di ottenere, per ogni stadio, lo stesso delay in modo che il ritardo totale sia pari alla somma dei singoli tale che ognuno sia pari al precedente e al successivo. Supponendo, però, che lo stadio di invertitori sia tale che ogni inverter sia dimensionato in maniera differente: il primo con dimensioni minime, il secondo con un sizing maggiore e così via fino all’ultimo inverter che presenta, ad esempio, un sizing 50 volte più grande di quello minimo. In questo caso, bisognerebbe considerare una tensione di soglia differente per ognuno ritornando, pertanto, al discorso fatto precedentemente. Infatti, ciò che succederebbe è che, considerando una tensione di soglia maggiore per gli inverter con sizing maggiore rispetto agli altri, il delay associato all’ultimo risulterebbe maggiore rispetto a quello con una tensione di soglia ed un sizing minore. Nello specifico, facendo riferimento ad una tensione di soglia di riferimento per un inverter a dimensioni medie, e, rispettivamente, ad una e una per l’inverter minimo e per l’inverter di dimensioni 50 volte il minimo, ciò che succederebbe e che i delay verrebbero dimensionati in maniera differente. In particolare, associando un delay per l’inverter a tensione di soglia , i delay del minimo e del ‘massimo’ saranno, rispettivamente, e . Questo significa che il delay totale dell’inverter chain non sarà cambiato ma ciò che è mutato è il ritardo per ogni singolo stadio. Però, utilizzando questo approccio, il leakage risulterà essere minore poiché anche se è aumentato nel primo stadio, considerando una tensione di soglia minore di quella standard, ne consegue che nello stadio dell’inverter ‘massimo’ sarà minore ma essendo che quest’ultimo è stato dimensionato molto più grande del minimo, ciò che è stato risparmiato a livello di leakage è di gran lunga maggiore di ciò che è stato peggiorato nello stadio dell’inverter minimo. Pertanto, si è ottenuto un leakage totale minimo a parità di delay.

Immagine che contiene testo, schermata, linea, Diagramma

Descrizione generata automaticamenteUna delle tecniche di ottimizzazione low power prevede di considerare il rapporto di aspetto di ogni stadio dell’inverter chain rispetto allo stadio dell’inverter minimo. Nello specifico, il parametro relativo allo stadio minimo viene posto pari a 1 poiché fa riferimento all’inverter in questione rispetto a quello minimo e, quindi, esso stesso. Pertanto, l’obiettivo sarà trovare una formulazione per l’energia e una per il delay dipendente dal rapporto di aspetto di ogni stadio della chain. Dunque, considerando il modello di energia e il vincolo associato al delay, è possibile risolvere un problema di ottimizzazione rispetto ai rapporti di aspetto:

supponendo che sia il rapporto di aspetto relativo all’inverter minimo. L’obiettivo sarà, quindi, trovare un insieme ottimo ( che rispetti i vincoli imposti. Ovviamente, considerando ciò è possibile ricavare una curva ottima, cioè l’insieme di tutte gli -esimi rapporti di aspetto ottimi , ottenendo la curva ottima associata al problema in questione.

# Obiettivi

# Analisi e progettazione di un inverter CMOS minimo

# Analisi e progettazione di un buffer CMOS a tre stadi

# Conclusioni