



AA 2010/2011

Università degli Studi di Padova

FACOLTÀ DI INGEGNERIA

Corso di Laurea Triennale in Ingegneria Elettronica

TESINA DI LAUREA TRIENNALE

**COMPORTAMENTO STOCASTICO DELL'INVERTER CMOS IN
REGIONE DI SOTTOSOGLIA**

Limiti del modello deterministico per l'analisi a bassissime tensioni di
alimentazione

CANDIDATO:

Giovanni Rovere
580805-IL

RELATORE:

Dr. Lorenzo Finesso

Nella vita reale non c'è alcun uomo medio.

A. Huxley

Abstract

La continua richiesta di circuiti integrati digitali a basso consumo spinge sempre più spesso i progettisti a ridurre le tensioni di alimentazione. Tale riduzione però comporta, inevitabilmente, anche un aumento della sensibilità al rumore. Per questo motivo è necessario raffinare gli attuali modelli deterministici tenendo conto degli effetti di natura aleatoria introdotti dal termine di rumore. In particolare, questo lavoro è focalizzato allo sviluppo di un modello dell'inverter CMOS per bassissime tensioni di alimentazione, la cui regione di funzionamento dei MOSFET è sempre quella in sottosoglia per qualsiasi tensione di ingresso applicata. Di tale dispositivo se ne è dapprima ricavato il modello stocastico e, successivamente, ne abbiamo simulato al computer il comportamento al variare dei parametri caratteristici. Grazie a queste simulazioni abbiamo ottenuto il limite inferiore di tensione per il quale l'inverter opera in maniera deterministica. Questo metodo e questi strumenti possono essere utilizzati come base per lo studio di circuiti CMOS più complessi il cui contributo dovuto al termine di rumore non è più una componente trascurabile.

Desidero ringraziare Mamma e Papà per il loro affetto e sostegno immenso, i familiari e gli amici più cari. Un pensiero speciale va a G.M., E.A. ed a L.E.N. a cui, nonostante tutto, devo moltissimo. Ringrazio infinitamente il Dr. L. Finesso per avermi guidato e supportato nella stesura di questo lavoro, per la grande disponibilità e cortesia dimostratemi, e per avermi dato sempre utili consigli. Ringrazio inoltre il Prof. A. Bevilacqua per il supporto nel Capitolo 2.

INDICE

1	Introduzione	7
2	Modellizzazione dell'inverter in regione di sottosoglia	11
2.1	Funzionamento del transistor nMOS in regione di sottosoglia	11
2.1.1	Modello nMOS in sottosoglia con tensioni V_{DS} elevate	13
2.1.2	Modello nMOS in sottosoglia con tensioni V_{DS} basse	14
2.2	Funzionamento dell'inverter CMOS per tensioni molto basse	15
2.2.1	Dimensionamento dei transistor	17
2.2.2	Equazione di uscita	21
3	Modellazione degli effetti non deterministici	27
3.1	Concetti fondamentali	27
3.1.1	Processi stocastici	27
3.1.2	Passeggiata aleatoria 1D (a random walk)	29
3.1.3	Processo di Wiener	30
3.2	Il rumore	35
3.2.1	Rumore termico	35
3.2.2	Rumore shot	36
3.2.3	Modello unificato	37
3.3	Il modello stocastico dell'inverter CMOS in regione di sottosoglia	37
4	Risoluzione ed analisi dei risultati	39
4.1	L'integrale stocastico e le SDE	39
4.1.1	Differenziabilità del processo di Wiener	39
4.1.2	Integrale stocastico di Itô	40
4.1.3	Equazioni differenziali stocastiche (SDE)	41
4.2	Soluzione Numerica	41

4.2.1	Analisi dei dati	42
4.3	Probabilità di errore ed informazione associata	44
5	Conclusioni	49
A	Codice MATLAB	51
A.1	PDF.m	51
A.2	funzionePDF.m	55
	Bibliografia	57

CAPITOLO

1

Introduzione

Sempre più spesso, oggigiorno, il concetto di elaborazione e memorizzazione di dati digitali riveste un ruolo fondamentale nella nostra società e nella nostra vita. Ciascuno di noi ha familiarità con dispositivi elettronici quali computer, telefoni cellulari, orologi e macchine fotografiche digitali. Questi oggetti stanno acquisendo tale importanza da essere indispensabili ed insostituibili per il lavoro, lo studio ed il tempo libero. È proprio per questo motivo che in questi anni la domanda di dispositivi ed applicazioni *wireless* e *low power consumption* è in continua crescita. In altre parole, si tende ad aumentare il più possibile la “portabilità” dei dispositivi.

Il concetto stesso di “portabilità” ed i suoi requisiti pongono però pesanti limiti ai parametri di progetto. Uno dei vincoli più importanti e delicati è sicuramente il consumo di energia. Questa è una caratteristica fondamentale per un dispositivo che vorremmo fosse il più portatile possibile, al fine di poter disporre di una maggiore autonomia e quindi libertà d’uso. Per soddisfare queste specifiche si può far ricorso all’utilizzo di batterie più grandi e performanti, ma ciò sarebbe in contrasto con l’ingombro ed il peso del dispositivo e quindi in contrasto con la “portabilità”. Un’alternativa più intelligente, al fine di aumentare l’autonomia, consiste nell’applicare tecniche o accorgimenti a livello circuitale che consentano di ridurre al minimo il consumo di energia.

Per studiare come ridurre il consumo di energia riportiamo brevemente la composizione delle potenze dissipate per il funzionamento di un circuito CMOS. Tale configurazione circuitale è largamente impiegata nei moderni circuiti digitali per l’elevato margine di sensibilità al rumore ed il basso consumo di potenza statica. La sua attenta analisi ci permette così di identificare dove si concentrano gli sprechi di potenza e da quali grandezze fisiche essi dipendano. Consideriamo un caso particolarmente semplice da studiare: l’invertitore digitale in

funzionamento sopra-soglia. Una volta compreso a fondo il funzionamento di tale invertitore è facile estendere l'analisi a qualsiasi altra porta logica CMOS.

La formula della potenza totale richiesta dall'inverter è costituita dalla somma di tre componenti:

$$P_{tot} = P_{stat} + P_{din} + P_{cc}$$

- $P_{stat} = V_{dd}I_{stat}$

Potenza statica: ovvero la potenza consumata dal circuito in assenza di commutazioni e con i valori di tensione e corrente stabilizzati. Essa è dovuta principalmente alla corrente di perdita I_{stat} delle giunzioni polarizzate inversamente *drain/source-substrato* e alla corrente di sottosoglia dei transistor.

- $P_{din} = V_{dd}^2 C_L f_{0 \rightarrow 1}$

Potenza dinamica: cioè la potenza impiegata per caricare e scaricare il condensatore C_L all'uscita del circuito. Esso può essere la capacità parassita del circuito analizzato, la capacità di ingresso di un circuito posto a valle dell'inverter, o come spesso accade, la somma di entrambi i contributi.

- $P_{cc} = V_{dd}I_{cc,max}t_s$

Potenza di corto circuito: è la potenza che viene dissipata per quella frazione di tempo t_s nella quale esiste un cammino diretto tra alimentazione e massa. Tale cammino si instaura quando la tensione di ingresso $V_i(t)$ del circuito commuta il proprio valore. In questo istante la tensione è sufficiente per accendere contemporaneamente entrambi i MOS.

Si osserva che la tensione V_{dd} è presente in ogni addendo della composizione della potenza totale P_{tot} . Nella componente dinamica P_{din} , in particolare, notiamo che essa dipende dalla tensione di alimentazione con relazione quadratica. Da queste considerazioni si evince che una possibile strategia per ridurre il consumo di potenza consiste nel diminuire la tensione di alimentazione V_{dd} . Diminuire tale tensione ha come secondo vantaggio la riduzione dell'intensità dei campi elettrici presenti all'interno del transistor MOS. Infatti, in virtù di questa riduzione, si possono scalare le dimensioni geometriche delle aree di silicio in sicurezza, senza correre il rischio di malfunzionamenti o peggio, di rotture dell'ossido di *gate*.

In particolare, esistono applicazioni nelle quali il consumo di energia è veramente un parametro critico che è necessario diminuire il più possibile anche a costo di sacrificare velocità e margini di rumore. Per questo tipo di applicazioni si può ricorrere a particolari configurazioni di circuiti CMOS alimentati a tensioni talmente basse da non essere nemmeno sufficienti per la formazione del canale conduttivo tra *drain* e *source* dei MOSFET. Questa particolare strategia si chiama funzionamento in sotto-soglia o, in inglese, *subthreshold*. Il principio di funzionamento consiste nello sfruttare la corrente di perdita presente nei dispositivi MOS. Essa infatti diventa rilevante quando si applicano al circuito tensioni di alimentazione e di controllo inferiori alla tensione di soglia. In pratica si riesce a costruire un inverter CMOS senza mai dover accendere i transistor. Ovviamente, come già menzionato, ciò si paga con un ridotto margine di immunità al rumore e con una maggior lentezza nella commutazione del

dispositivo. Ciò nonostante configurazioni di questo tipo oggi sono molto diffuse in circuiti come per esempio pacemaker, orologi e strumentazione da polso e dispositivi autoalimentati.

Si ha però che, quando le tensioni in gioco sono così piccole da poter essere comparate con le ampiezze del rumore intrinseco al dispositivo, la tensione di ingresso può fluttuare attorno alla tensione di soglia dell'inverter causando così ambiguità del valore logico da interpretare. Questi effetti di fatto limitano la possibilità di analizzare e studiare il circuito con i convenzionali strumenti matematici deterministici e richiedono l'uso della teoria dei fenomeni aleatori. Lo scopo di questo lavoro è proprio quello di analizzare dal punto di vista stocastico il comportamento dell'inverter CMOS per bassissime tensioni di alimentazione.

Questo lavoro è stato ispirato dall'articolo [1] sia come struttura sia come obiettivi, anche se notevolmente ampliato ed integrato. Invece, la sezione 2.2.1, i passaggi matematici che conducono dalla Formula 2.16 alla Formula 2.20, assieme al metodo per la soluzione dell'equazione differenziale 2.20 ed il codice in Appendice, sono originali.

CAPITOLO

2

Modellizzazione dell'inverter in regione di sottosoglia

2.1 Funzionamento del transistor nMOS in regione di sottosoglia

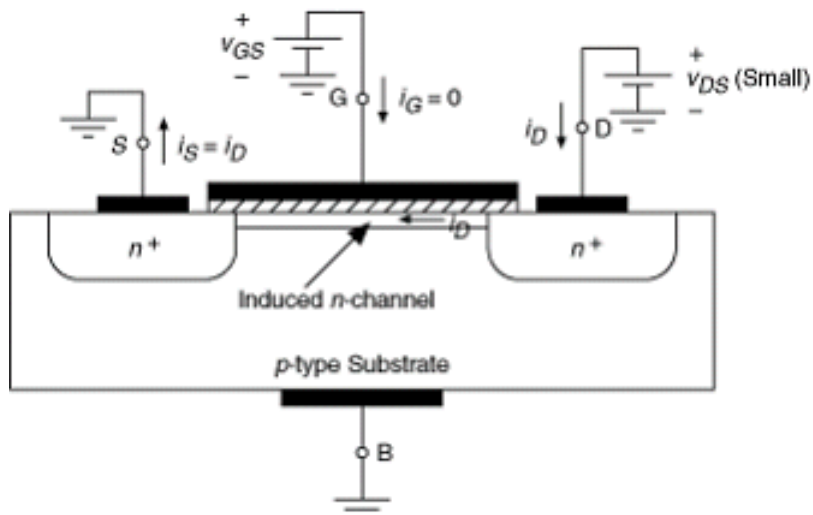


Figura 2.1: Sezione di un dispositivo Mosfet a canale n.

Indichiamo con V_{TH} la tensione di soglia del transistor, cioè la minima differenza di poten-

ziale applicabile tra *gate* e *source* del MOS per formare il canale. Normalmente un transistor nMOS è considerato acceso quando la tensione V_{GS} è superiore alla tensione di soglia V_{TH} (canale formato); mentre lo consideriamo spento quando essa è inferiore a V_{TH} (canale parzialmente o non formato). Con questo modello però presupponiamo che le regioni di funzionamento siano nettamente divise l'una dall'altra e che le transizioni tra tali regioni siano secche e repentine. Questo è ovviamente un modello alquanto impreciso ed utilizzabile solo per l'analisi circuitale del primo ordine e con un intervallo di tensioni di lavoro abbastanza elevato (comunque superiori a V_{TH} di modo da accendere il transistor). Come è logico aspettarsi però, nella realtà, si osserva che la transizione tra le due regioni non è netta. Se infatti applichiamo una tensione positiva V_{GS} prossima alla tensione di soglia (e quindi a cavallo tra le due regioni di funzionamento) faremo funzionare il transistor in una regione intermedia chiamata regione di *svuotamento* o di *inversione debole*. In tale condizione si osserva che nella zona del *substrato* (p) più prossima all'elettrodo di *gate* inizia a formarsi un sottile strato la cui densità di lacune si riduce progressivamente. Questa riduzione di cariche (da qui il nome *svuotamento*) continua fino ad annullarsi perché attratte dall'elettrodo di *substrato*. Possiamo quindi considerare questo strato non più drogato di tipo (p) ma bensì neutro. In queste condizioni l'nMOS non è acceso, ma nemmeno completamente spento. Esso diventa parzialmente conduttivo perché, anche se non è presente un canale conduttivo vero e proprio tra *source* e *drain*, una piccola quantità di elettroni ha sufficiente energia per entrare nel canale e dare origine alla corrente I_{DS} di sottosoglia. Tale corrente presenta un andamento esponenziale, del tutto simile alla corrente di un transistor bipolare. Possiamo infatti osservare che in assenza del canale conduttivo tra *drain* e *source* le regioni *source* ($n+$) - *substrato* (p) - *drain* ($n+$) formano effettivamente un transistor bipolare parassita.

Il transistor nMOS operante in regione di sottosoglia presenta in tutto tre contributi di corrente. La corrente di gate I_G che scorre tra *gate* e *source/drain*, la corrente di giunzione I_J che scorre dai *source/drain* nel *substrato* ed infine la sopra citata corrente di sottosoglia I_{DS} . A tensioni di esercizio sufficientemente basse la corrente I_{DS} è dominante e possiamo quindi trascurare le altre due componenti. Otteniamo così che la corrente che scorre in un transistor nMOS in regione di sottosoglia è regolata dalla ben nota equazione[2]:

$$I_{DS}(t) = I_0 \frac{W}{L} e^{\frac{k(V_{GS}(t)-V_{TH})}{V_T}} \left(1 - e^{-\frac{V_{DS}(t)}{V_T}} \right) \quad (2.1)$$

In tale formula I_0 e k sono parametri che dipendono dalla tecnologia a disposizione, $\frac{W}{L}$ è il rapporto tra le dimensioni larghezza e lunghezza del canale, mentre $V_T = \frac{k_b T}{q}$ è la tensione termica. Ricordiamo che k_b è la costante di Boltzmann, T è la temperatura in gradi Kelvin, mentre q è la carica dell'elettrone. A temperatura ambiente ($27^\circ\text{C} \approx 300\text{K}$) si ha che $V_T \approx 26\text{mV}$. La tensione di soglia effettiva V_{TH} del transistor è regolata dall'equazione:

$$V_{TH} = V_{TH0} - \lambda_{DS} V_{DS} - \lambda_{BS} V_{BS} \quad (2.2)$$

nella quale V_{TH0} è la tensione di soglia quando il transistor non è polarizzato, mentre λ_{DS} e λ_{BS} sono parametri tecnologici che indicano di quanto varia la tensione di soglia rispetto a variazioni delle tensioni V_{DS} e V_{BS} . Questa variazione della tensione di soglia è una

conseguenza degli effetti secondari come l'effetto body ed il DIBL. Da notare che essi diventano particolarmente significativi e sempre meno trascurabili man mano che le dimensioni del transistor si riducono.

Inserendo la Formula 2.2 nella 2.1 si ottiene per un transistor nMOS,

$$I_{DS}(t) = \beta_n e^{\frac{k_n V_{GS}(t)}{V_T}} \left[e^{\frac{k_n \lambda_{DS} V_{DS}(t)}{V_T}} \left(1 - e^{-\frac{V_{DS}(t)}{V_T}} \right) \right] \quad (2.3)$$

con β_n :

$$\beta_n = I_0 \frac{W_n}{L} e^{-\frac{k_n (V_{TH0} - \lambda_{BS} V_{BS})}{V_T}} \quad (2.4)$$

che rappresenta sostanzialmente la “forza” del transistor, cioè quanto esso è conduttivo. Questo concetto è fondamentale per determinare correttamente la tensione di commutazione dell'inverter, essa può essere opportunamente regolata scegliendo le dimensioni W dei MOS impiegati.

Riportiamo (da [3]) indicativamente i parametri tecnologici dei transistor nMOS/pMOS con tecnologia a 65-nm, $\frac{W}{L} = \frac{120}{80} nm$:

	k	$I_0 [A]$	$V_{TH0} [V]$	λ_{DS}	λ_{BS}
nMOS	1/1.39	6.65×10^{-5}	0.598	9.0×10^{-2}	9.9×10^{-2}
pMOS	1/1.27	5.95×10^{-6}	0.532	8.0×10^{-2}	1.1×10^{-1}

Tabella 2.1: Parametri nMOS/pMOS tecnologia da 65-nm e $\frac{W}{L} = \frac{120}{80} nm$

La Formula 2.3 è rigorosa ma, per l'analisi manuale di un circuito, essa risulta complessa e poco intuitiva. Per questo motivo per la spiegazione del funzionamento dell'inverter in regione di sottosoglia ci affideremo a dei modelli semplificati: uno per tensioni V_{DS} elevate (ma pure sempre inferiori alla tensione di soglia V_{TH}) ed uno per tensioni V_{DS} molto basse.

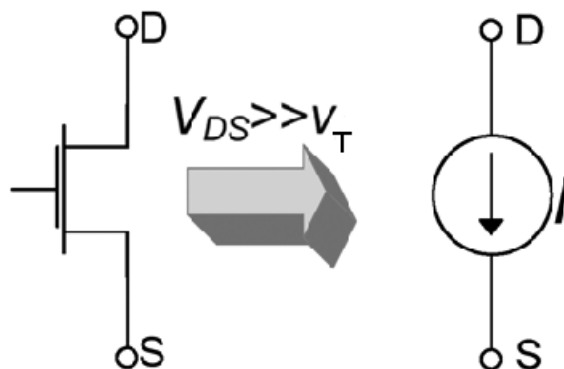
Innanzitutto notiamo che il termine $e^{\frac{k_n \lambda_{DS} V_{DS}(t)}{V_T}}$ della Formula 2.3 può essere approssimato a 1. Infatti, secondo i parametri della Tabella 2.1, $\frac{k_n \lambda_{DS} V_{DS}(t)}{V_T} \ll 1$ considerando che la V_{DS} è dell'ordine delle decine di millivolt.

2.1.1 Modello nMOS in sottosoglia con tensioni V_{DS} elevate

Supponiamo ora che V_{DS} abbia valori abbastanza elevati rispetto a V_T . Esempio $V_{DS} = 100mV \simeq 4V_T$. Osservando la Formula 2.3 notiamo che l'ultimo fattore $1 - e^{-\frac{V_{DS}}{V_T}}$ è circa 1, quindi possiamo eliminarlo dal prodotto. Otteniamo quindi la seguente semplificazione:

$$I_{DS}(t) \approx \beta_n e^{k_n \frac{V_{GS}(t)}{V_T}} \quad \text{valida per} \quad V_{TH} > V_{DS} \gg V_T \quad (2.5)$$

Da questa equazione si deduce che si può modellare il transistor come un generatore di corrente ideale controllato dalla tensione $V_{GS}(t)$. Tale rappresentazione è mostrata in Figura 2.2.


 Figura 2.2: Modello per V_{DS} elevate

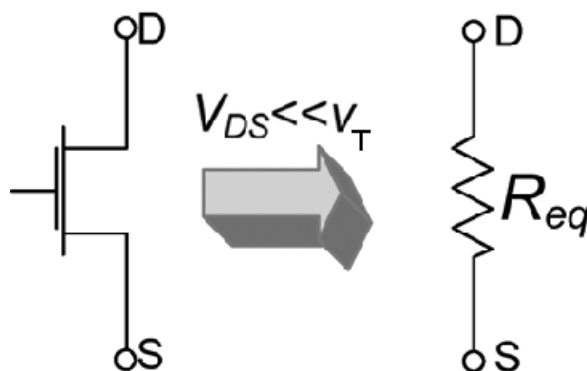
2.1.2 Modello nMOS in sottosoglia con tensioni V_{DS} basse

Supponiamo ora che V_{DS} sia circa uguale a zero, e quindi molto inferiore rispetto a V_T . Il termine $1 - e^{-\frac{V_{DS}}{V_T}}$ non è più trascurabile ma è approssimabile con la serie di Taylor. Al secondo ordine vale $1 - e^{-\frac{V_{DS}}{V_T}} \simeq \frac{V_{DS}}{V_T}$, otteniamo quindi:

$$I_{DS}(t) \approx \beta_n e^{k_n \frac{V_{GS}(t)}{V_T}} \frac{V_{DS}}{V_T} \quad \text{valida per} \quad V_{TH} > V_T \gg V_{DS} \quad (2.6)$$

Il transistor in questo caso si comporta come una resistenza, come mostrato nella Figura 2.3.

$$R_{eq} = \frac{V_{DS}}{I} = \frac{V_T}{\beta_n e^{k_n \frac{V_{GS}(t)}{V_T}}} \quad (2.7)$$


 Figura 2.3: Modello per V_{DS} basse

Si noti che non viene fornita una modellazione per tensioni V_{DS} intermedie. Questo di fatto non è un problema poiché in elettronica digitale le tensioni intermedie si verificano solo per un lasso di tempo limitato. Nel caso in analisi ciò accade quando la tensione di ingresso $V_i(t)$ commuta il proprio valore $H \rightarrow L$ oppure $L \rightarrow H$. Quindi, al fine di agevolare lo studio, possiamo considerare esclusivamente le zone di funzionamento sopra menzionate.

Tutte le considerazioni fatte nel Paragrafo 2.1 continuano a valere anche per i transistor di tipo pMOS. Bisogna però tenere conto che le tensioni in gioco (V_{GS} e V_{DS}) sono negative, mentre la corrente che scorre nel pMOS è per convenzione assunta positiva uscente dal terminale di *drain*.

2.2 Funzionamento dell'inverter CMOS per tensioni molto basse

L'inverter CMOS è uno dei circuiti più semplici che si possano realizzare secondo il paradigma CMOS. Esso è costituito da due soli transistor, un nMOS ed un pMOS connessi in serie. Tali transistor hanno in comune i due terminali di *gate* che fungono da ingresso della porta ed i due terminali di *drain*, il cui valore di tensione rappresenta il segnale di uscita. I terminali di *source* sono invece connessi uno a V_{dd} (pMOS) ed uno a massa (nMOS). Il circuito dell'inverter sotto esame è mostrato in Figura 2.4.

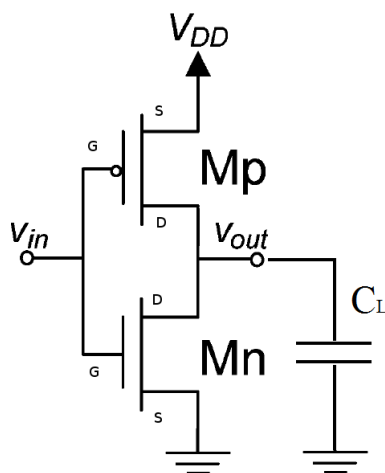


Figura 2.4: Schema circuitale dell'inverter CMOS

Da tale figura osserviamo che la V_{GS} dell'nMOS è pari alla tensione di ingresso $V_i(t)$ mentre la tensione V_{GS} del pMOS è pari a $V_i(t) - V_{dd}$. In maniera analoga si deduce che V_{DS} dell'nMOS è pari a $V_o(t)$ mentre V_{DS} del pMOS è $V_o(t) - V_{dd}$. Particolare attenzione va prestata al fatto che nei transistor pMOS le tensioni V_{GS} e V_{DS} sono di fatto negative. Quindi, la Formula 2.1 è esatta così com'è solo per gli nMOS mentre, per i pMOS, tale formula va leggermente modificata ricordandosi di prendere il modulo di V_{GS} e V_{DS} . Questo significa utilizzare $|V_{GS}| = -V_{GS} = V_{SG}$ e allo stesso modo $|V_{DS}| = -V_{DS} = V_{SD}$. Sempre per questione di convenzioni assumiamo positiva la corrente I_{DS} nell'nMOS e negativa la corrente I_{DS} del pMOS. Quindi, nei pMOS, la corrente scorre di fatto dal source al drain, che l'esatto opposto di quanto si verifica nell'nMOS.

Da queste considerazioni, dalla Figura 2.4 e dalla Formula 2.1 possiamo quindi scrivere le equazioni delle correnti che scorrono all'interno dei MOS dell'inverter:

$$I_n(t) = I_{0n} \frac{W_n}{L} e^{\frac{k_n V_i(t)}{V_T}} \left(1 - e^{-\frac{V_o(t)}{V_T}} \right) \quad (2.8)$$

$$I_p(t) = I_{0p} \frac{W_p}{L} e^{\frac{k_p (V_{dd} - V_i(t))}{V_T}} \left(1 - e^{-\frac{V_{dd} - V_o(t)}{V_T}} \right) \quad (2.9)$$

Da notare che il termine all'esponente V_{TH} , presente nell'equazione 2.1, è stato inglobato all'interno di I_{0n} e di I_{0p} poiché circa costante a meno degli effetti del secondo ordine. Cioè: $I_{0n} = I_0 e^{-\frac{k_n V_{TH}}{V_T}}$.

Per comprendere il funzionamento del circuito lo analizzeremo dapprima applicando al suo ingresso una tensione molto bassa e successivamente una tensione elevata. In questo modo avremo due modelli differenti del circuito sotto esame che ne agevolano la comprensione del funzionamento. Ovviamente supponiamo sempre che, sia l'alimentazione V_{dd} che la tensione di ingresso $V_i(t)$, siano entrambe inferiori alla tensione di soglia V_{TH} , di modo che l'inverter funzioni costantemente con i transistor "spenti" in regione sotto-soglia.

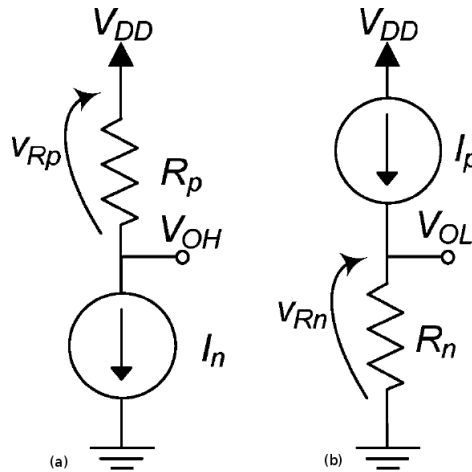


Figura 2.5: Modellazione dell'inverter CMOS per tensioni di ingresso basse (a) ed elevate (b)

Applicando una tensione bassa (circa zero) al terminale di ingresso dell'inverter, si ha che il modulo della V_{GSn} è basso, mentre il modulo della V_{GSp} , è elevato. In base a questa osservazione e dopo i ragionamenti fatti precedentemente, possiamo schematizzare i due MOS dell'inverter uno come un resistore ed uno come un generatore ideale di corrente. Considerazioni analoghe e duali si applicano al caso di una tensione di ingresso elevata. Il circuiti così ottenuti sono mostrati in Figura 2.5.

Studiamo ora le tensioni sul nodo di uscita per osservare se, e come, esse dipendano dalla tensione applicata all'ingresso. Per il circuito mostrato nella Figura 2.5(a) possiamo notare che la tensione di uscita dipende sostanzialmente dalla caduta di tensione presente sul resistore R_p che indichiamo con V_{rp} . Infatti la tensione ai capi di un generatore ideale di corrente non è nota a priori e non c'è modo di determinarla in maniera diretta senza considerazioni derivanti dalla topologia circuitale. A questo punto possiamo scrivere senza ulteriori indugi la seguente

formula per la tensione di uscita del circuito quando la tensione di ingresso è bassa.

$$V_{OH} = V_{dd} - \Delta V_{OH} \quad (2.10)$$

Dove, ΔV_{OH} è la caduta di tensione sulla resistenza R_p .

$$\Delta V_{OH} = R_p I_n = V_T \frac{\beta_n}{\beta_p} e^{-k_p \frac{V_{dd}}{V_T}} \quad (2.11)$$

Si può procedere analogamente anche per il circuito in Figura 2.5 (b). In questo caso si è supposta la tensione di ingresso abbastanza elevata per modellare l'nMOS come un resistore ed il pMOS come un generatore, ottenendo quindi un circuito duale al precedente. Anche in questo caso quindi la tensione di uscita è fondamentalmente dipendente dalla caduta di tensione sulla resistenza R_n che, ricordiamo, non è altro che il transistor nMOS. Si può quindi scrivere l'equazione:

$$V_{OL} = \Delta V_{OL} = R_n I_p = V_T \frac{\beta_p}{\beta_n} e^{-k_n \frac{V_{dd}}{V_T}} \quad (2.12)$$

Supposto che lo swing ammissibile di tensione sia compreso tra 0 e V_{dd} , dalle equazioni 2.11 e 2.12 si nota come la tensione di uscita si può considerare HIGH, se in ingresso abbiamo uno "0" logico, mentre è LOW, se in ingresso applichiamo un "1" logico. Il livello di tensione che permette di determinare se il segnale in ingresso è da interpretare come High oppure Low si chiama **tensione di soglia logica** V_M . Essa dipende da β_n e da β_p , cioè dalla conduttività di ciascun MOSFET. Si presti particolare attenzione alla differenza tra la tensione di soglia del MOSFET V_{TH0} e V_{TH} e la tensione di soglia logica dell'inverter V_M .

Ecco dimostrato, tramite la modellizzazione dei transistor in sottosoglia, che il comportamento di tale circuito è proprio quello di una porta NOT. Riportiamo in Tabella 2.2 la tavola di verità di questa semplicissima funzione logica.

IN	OUT
0	1
1	0

Tabella 2.2: Tabella di verità di un invertitore

2.2.1 Dimensionamento dei transistor

Nel design dei circuiti integrati digitali una specifica che viene spesso richiesta al progettista è di fare in modo che la tensione di soglia V_M sia pari a $V_{dd}/2$. Questa richiesta è ragionevole perché permette di avere il massimo margine di rumore bilanciato. Infatti se in ingresso all'inverter disponiamo di un segnale digitale ideale (High = V_{dd} e Low = 0) segue che l'ampiezza minima del rumore per far interpretare erroneamente l'ingresso dipende dalla soglia logica ed è pari al minimo tra $|V_M - 0|$ e $|V_M - V_{dd}|$. Essendo inoltre $0 < V_M < V_{dd}$, per massimizzare entrambi i margini di rumore è quindi necessario porre $V_M = \frac{V_{dd}}{2}$.

La tensione di soglia è un parametro che dipende essenzialmente dal rapporto tra le "forze" dei singoli transistor, cioè da β_n e da β_p . Se abbiamo un nMOS più conduttivo del

pMOS ($\beta_n > \beta_p$) otteniamo un invertitore la cui soglia logica avrà un valore inferiore a $\frac{V_{dd}}{2}$. Viceversa avere un pMOS più conduttivo rispetto all'nMOS ($\beta_n < \beta_p$) spinge la V_M ad un valore che si avvicina a V_{dd} . Dunque è chiaro che per imporre la tensione di soglia $V_M = \frac{V_{dd}}{2}$ è necessario agire sulle conduttività dei transistor in modo da avere $\beta_n = \beta_p$.

A livello pratico questo effetto si ottiene modificando la larghezza di canale W dei due MOSFET (vedere Formula 2.4), che è peraltro uno dei pochi parametri accessibili al progettista di circuiti integrati. Aumentare la larghezza di canale W permette quindi di avere un canale più largo e quindi di offrire un mezzo più conduttivo al flusso di cariche. L'aumento di W però, oltre ad un aumento della grandezza del transistor e quindi di un maggiore impiego di area di silicio, comporta anche un incremento della capacità parassita relativa ai terminali di *gate*, *source* e *drain*. Si evince quindi che la modifica di W è un'operazione estremamente critica e complessa, il cui corretto settaggio dipende pesantemente dall'applicazione e dal contesto per cui deve essere progettato il circuito. In genere, per i motivi sopra esposti, si tende a progettare circuiti integrati con transistor a larghezza di canale minima (imposta da vincoli tecnologici), per poi aumentare il W laddove sia strettamente necessario.

Per esempio, dai parametri mostrati in Tabella 2.1 si evince che, a parità di dimensioni $W_n = W_p$, il transistor nMOS è più conduttivo del transistor pMOS di circa un ordine di grandezza. Questa disparità è conseguenza diretta della diversa mobilità dei portatori di carica (elettroni - lacune) all'interno dei canali (n - p) dei transistor. Quindi, per ottenere una transcaratteristica dell'inverter bilanciata, si parte da pMOS ed nMOS a dimensione minima per poi aumentare opportunamente la dimensione W_p del pMOS.

La tensione di soglia logica V_M è definita come il punto sul grafico della transcaratteristica per cui la tensione di uscita $V_o(t)$ uguaglia la tensione in ingresso $V_i(t)$. L'espressione analitica di V_M si ottiene imponendo che, in tale punto, le correnti circolanti nei MOSFET siano uguali.

Quindi, staticamente devono valere le seguenti condizioni:

$$\begin{cases} V_i(t) = V_o(t) = V_M = \frac{V_{dd}}{2} = V_{DSn} = V_{SDp} \\ I_n(t) = I_p(t) \end{cases} \quad (2.13)$$

che possiamo condensare nella seguente equazione:¹

$$I_{0n} \frac{W_{min}}{L} e^{k_n \frac{V_{dd} - V_{THn}}{V_T}} \left(1 - e^{-\frac{V_{dd}}{2V_T}} \right) = I_{0p} \frac{\chi W_{min}}{L} e^{k_p \frac{V_{dd} - V_{THp}}{V_T}} \left(1 - e^{-\frac{V_{dd}}{2V_T}} \right) \quad (2.14)$$

da cui otteniamo:

$$\chi = \frac{I_{0n} e^{k_n \frac{V_{dd} - V_{THn}}{V_T}}}{I_{0p} e^{k_p \frac{V_{dd} - V_{THp}}{V_T}}} \quad (2.15)$$

Tali risultati sono riassunti nella Tabella 2.3 e nel Grafico 2.6.

¹Indicando con $W_{min} := W_{n_{min}} = W_{p_{min}}$ la dimensione minima della larghezza di canale si ha: $W_n = W_{min}$ e $W_p = \chi W_{min}$

	T = 300K	T = 400K
V_{dd} [V]	χ	χ
0.30	4.8	5.9
0.25	5.2	6.3
0.20	5.5	6.6
0.15	5.9	6.9
0.10	6.3	7.3
0.05	6.8	7.7
0.01	7.1	8.0
0.001	7.1	8.0

Tabella 2.3: Fattore moltiplicativo della larghezza di canale del pMOS (W_p) per alcuni valori di tensione V_{dd} e temperatura T

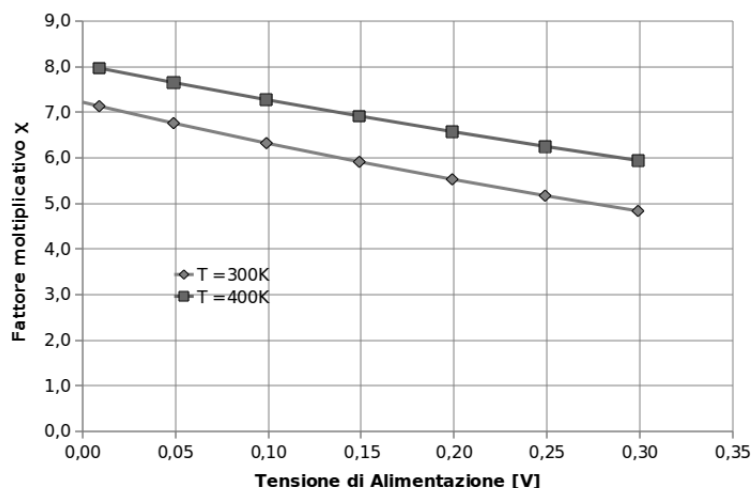


Figura 2.6: Grafico Excel χ necessario per ottenere $V_M = \frac{V_{dd}}{2}$

Le relazioni ingresso-uscita dell'inverter, per varie tensioni di alimentazione, sono mostrate in Figura 2.7. Tali funzioni sono state ottenute con il software di simulazione MATLAB (per il codice si veda l'Appendice) utilizzando come parametri tecnologici quelli della Tabella 2.1.

La tensione V_M è facilmente individuabile nel punto del grafico in cui le varie curve statiche e la retta bisettrice si intersecano (Figure 2.7 e 2.8). È inoltre particolarmente evidente come le curve ingresso-uscita perdano la loro forma a gradino man mano che si diminuisce la tensione di alimentazione. Questo fenomeno è dovuto all'andamento esponenziale delle correnti di perdita nei transistor in accordo con la diminuzione della tensione di ingresso $V_i(t)$ e quindi con la tensione di alimentazione V_{dd} .

Da notare che il valore della capacità $C_L = (C_{load} + \sum C_{parassite})$ posta all'uscita del circuito è ininfluenza per il tracciamento delle curve di Figura 2.7 e Figura 2.8. Questo perché tali curve rappresentano solo il comportamento statico dell'inverter, ovvero il comportamento dopo che tutte le tensioni e correnti nel circuito si sono assestate. Il valore del condensatore è invece un parametro importantissimo per il tracciamento delle curve dinamiche, le quali

mostrano il valore della tensione sul nodo di uscita del circuito in funzione del tempo. È immediato intuire che, più il valore di tale condensatore è elevato, maggiore sarà il tempo necessario a caricarlo, e quindi il tempo di risposta della tensione di uscita $V_o(t)$ del circuito.

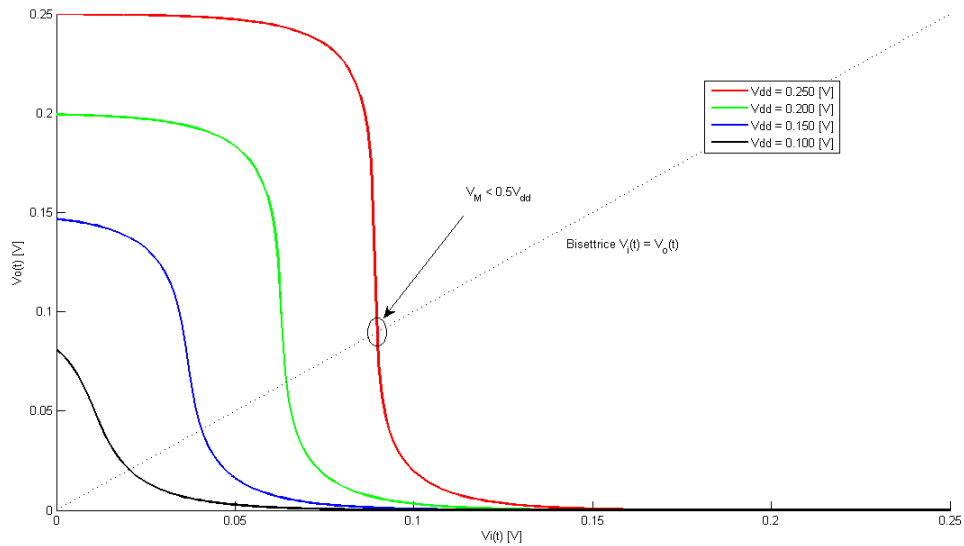


Figura 2.7: Caratteristiche statiche dell'inverter CMOS **non bilanciato** a basse tensioni di alimentazione

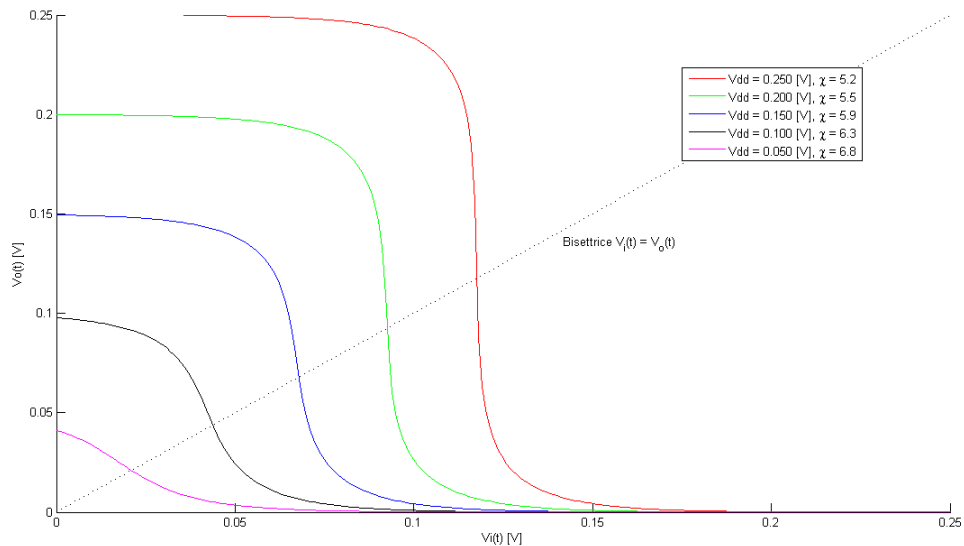


Figura 2.8: Caratteristiche statiche dell'inverter CMOS **bilanciate** (χ adattato $\forall V_{dd}$) a basse tensioni di alimentazione

2.2.2 Equazione di uscita

A partire dalle Formule 2.8 e 2.9 possiamo scrivere l'equazione della tensione di uscita del circuito $V_o(t)$. Tale equazione si ricava immediatamente dalla relazione tensione-corrente su un condensatore che è $i_c(t) = C\dot{v}_c(t)$. Nel circuito di Figura 2.4 la corrente che scorre all'interno della capacità è regolata dalla legge di Kirchhoff ed è pari alla differenza tra le correnti che scorrono all'interno dei due transistor MOS. Tale relazione è indicata dalla Formula 2.16

$$\boxed{\frac{dV_o(t)}{dt} = \frac{I_p(t) - I_n(t)}{C_L}} \quad (2.16)$$

Inserendo nella Formula 2.16 le espressioni delle correnti 2.8 e 2.9 otteniamo²:

$$\frac{I_p(t) - I_n(t)}{C_L} = I_{0p} \frac{W}{LC_L} e^{\frac{k_p(V_{dd}-V_i(t))}{V_T}} \left(1 - e^{-\frac{V_{dd}-V_o(t)}{V_T}}\right) - I_{0n} \frac{W}{LC_L} e^{\frac{k_n V_i(t)}{V_T}} \left(1 - e^{-\frac{V_o(t)}{V_T}}\right) =$$

$$\frac{W}{LC_L} \left[\underbrace{I_{0p} e^{k_p \frac{V_{dd}}{V_T}} e^{-k_p \frac{V_i(t)}{V_T}}}_{\mathcal{A}} \underbrace{- I_{0n} e^{k_n \frac{V_i(t)}{V_T}}}_{\mathcal{B}} \underbrace{- I_{0p} e^{k_p \frac{V_{dd}}{V_T}} e^{-k_p \frac{V_i(t)}{V_T}} e^{\frac{V_o(t)}{V_T}} e^{-\frac{V_{dd}}{V_T}}}_{\mathcal{C}} + \underbrace{I_{0n} e^{k_n \frac{V_i(t)}{V_T}} e^{-\frac{V_o(t)}{V_T}}}_{\mathcal{D}} \right]$$

$$\mathcal{A}: I_{0p} e^{k_p \frac{V_{dd}}{V_T}} e^{-k_p \frac{V_i(t)}{V_T}} = e^{\frac{1}{2} \left[2 \ln I_{0p} + 2k_p \frac{V_{dd}}{V_T} - 2k_p \frac{V_i(t)}{V_T} \right]} =$$

$$e^{\frac{1}{2} \left[\ln(I_{0p} I_{0n}) + k_p \frac{V_{dd}}{V_T} + (k_n - k_p) \frac{V_i(t)}{V_T} \right]} \times e^{\frac{1}{2} \left[\ln \frac{I_{0p}}{I_{0n}} + k_p \frac{V_{dd}}{V_T} - (k_n + k_p) \frac{V_i(t)}{V_T} \right]}$$

$$\underbrace{\hspace{10em}}_{\mathcal{A1}} \times \underbrace{\hspace{10em}}_{\mathcal{A2}}$$

$$\mathcal{B}: -I_{0n} e^{k_n \frac{V_i(t)}{V_T}} = -e^{\frac{1}{2} \left[2 \ln I_{0n} + 2k_n \frac{V_i(t)}{V_T} \right]} =$$

$$e^{\frac{1}{2} \left[\ln(I_{0p} I_{0n}) + k_p \frac{V_{dd}}{V_T} + (k_n - k_p) \frac{V_i(t)}{V_T} \right]} \times \left(-e^{-\frac{1}{2} \left[\ln \frac{I_{0p}}{I_{0n}} + k_p \frac{V_{dd}}{V_T} - (k_n + k_p) \frac{V_i(t)}{V_T} \right]} \right)$$

$$\underbrace{\hspace{10em}}_{\mathcal{B1} = \mathcal{A1}} \times \underbrace{\hspace{10em}}_{\mathcal{B2}}$$

$$\mathcal{C}: -I_{0p} e^{k_p \frac{V_{dd}}{V_T}} e^{-k_p \frac{V_i(t)}{V_T}} e^{\frac{V_o(t)}{V_T}} e^{-\frac{V_{dd}}{V_T}} = -e^{\frac{1}{2} \left[2 \ln I_{0p} + 2(k_p - 1) \frac{V_{dd}}{V_T} - 2k_p \frac{V_i(t)}{V_T} \right]} e^{\frac{V_o(t)}{V_T}} =$$

$$e^{\frac{1}{2} \left[\ln(I_{0p} I_{0n}) + (k_p - 1) \frac{V_{dd}}{V_T} + (k_n - k_p) \frac{V_i(t)}{V_T} \right]} \times \left(-e^{\frac{V_o(t)}{V_T}} e^{-\frac{1}{2} \left[\ln \frac{I_{0n}}{I_{0p}} + (1 - k_p) \frac{V_{dd}}{V_T} + (k_n + k_p) \frac{V_i(t)}{V_T} \right]} \right)$$

$$\underbrace{\hspace{10em}}_{\mathcal{C1}} \times \underbrace{\hspace{10em}}_{\mathcal{C2}}$$

²D'ora in avanti consideriamo $W_n = W_p = W_{min} = W$ per non appesantire la notazione. Nel momento in cui sarà necessario modificare la "forza" dei transistor agendo sulla loro larghezza sarà sufficiente moltiplicare il termine I_{0n} oppure I_{0p} per l'opportuno fattore moltiplicativo χ . Si presti attenzione al fatto che I_{0n} e I_{0p} non sono parametri accessibili al progettista, però dal punto di vista matematico le due operazioni sono equivalenti.

$$\mathcal{D} : I_{0n} e^{k_n \frac{V_i(t)}{V_T}} e^{-\frac{V_o(t)}{V_T}} = e^{\frac{1}{2} [2 \ln I_{0n} + 2k_n \frac{V_i(t)}{V_T}]} e^{-\frac{V_o(t)}{V_T}} =$$

$$\underbrace{e^{\frac{1}{2} [\ln(I_{0p} I_{0n}) + (k_p - 1) \frac{V_{dd}}{V_T} + (k_n - k_p) \frac{V_i(t)}{V_T}]} }_{\mathcal{D}1 = \mathcal{C}1} \times \underbrace{e^{-\frac{V_o(t)}{V_T}} e^{\frac{1}{2} [\ln \frac{I_{0n}}{I_{0p}} + (1 - k_p) \frac{V_{dd}}{V_T} + (k_n + k_p) \frac{V_i(t)}{V_T}]} }_{\mathcal{D}2}$$

Da $\frac{W}{LC_L} [\mathcal{A} + \mathcal{B}]$ otteniamo

$\frac{W}{LC_L} [\mathcal{A}1 (\mathcal{A}2 + \mathcal{B}2)] = \frac{2W}{LC_L} \mathcal{A}1 \sinh \left(\frac{1}{2} \left[\ln \frac{I_{0p}}{I_{0n}} + k_p \frac{V_{dd}}{V_T} - (k_n + k_p) \frac{V_i(t)}{V_T} \right] \right)$. Che per comodità riscriviamo:

$$a(t) = \frac{2W}{LC_L} e^{\frac{1}{2} [\ln(I_{0p} I_{0n}) + k_p \frac{V_{dd}}{V_T} + (k_n - k_p) \frac{V_i(t)}{V_T}]} \sinh \left(\frac{1}{2} \left[\ln \frac{I_{0p}}{I_{0n}} + k_p \frac{V_{dd}}{V_T} - (k_n + k_p) \frac{V_i(t)}{V_T} \right] \right) \quad (2.17)$$

Mentre da $\frac{W}{LC_L} [\mathcal{C} + \mathcal{D}]$ otteniamo

$$\frac{W}{LC_L} [\mathcal{C}1 (\mathcal{C}2 + \mathcal{D}2)] = \frac{2W}{LC_L} \mathcal{C}1 \sinh \left(-\frac{V_o(t)}{V_T} + \frac{1}{2} \left[\ln \frac{I_{0n}}{I_{0p}} + (1 - k_p) \frac{V_{dd}}{V_T} + (k_n + k_p) \frac{V_i(t)}{V_T} \right] \right)$$

Quindi, rinominando $\frac{2W}{LC_L} \mathcal{C}1 = b(t)$ si ha:

$$b(t) = \frac{2W}{LC_L} e^{\frac{1}{2} [\ln(I_{0p} I_{0n}) + (k_p - 1) \frac{V_{dd}}{V_T} + (k_n - k_p) \frac{V_i(t)}{V_T}]} \quad (2.18)$$

e ponendo

$$c(t) = \frac{V_T}{2} \left[\ln \frac{I_{0n}}{I_{0p}} + (1 - k_p) \frac{V_{dd}}{V_T} + (k_p + k_n) \frac{V_i(t)}{V_T} \right] \quad (2.19)$$

possiamo scrivere l'equazione che regola la tensione sul nodo di uscita

$$\boxed{\frac{dV_o(t)}{dt} = a(t) + b(t) \sinh \left(\frac{c(t) - V_o(t)}{V_T} \right)} \quad (2.20)$$

Si noti che $a(t)$, $b(t)$ e $c(t)$ sono delle funzioni che dipendono dalla tensione di ingresso $V_i(t)$, ma non sono dipendenti dalla tensione di uscita $V_o(t)$. Quindi abbiamo che la 2.20 è un'equazione differenziale non lineare del primo ordine.

Per risolvere tale ODE analiticamente possiamo fare alcune considerazioni e semplificazioni. Innanzitutto ipotizziamo che la tensione di ingresso $V_i(t)$ abbia un andamento ideale, cioè sia costante (valore *HIGH* o *LOW*) a meno di alcuni punti (istanti temporali) in cui si ha il cambio del valore logico. In tali punti si avrà una discontinuità. Questa supposizione ovviamente è un'approssimazione poiché in realtà la pendenza in questi punti è molto elevata,

ma non infinita. Comunque questa ipotesi semplifica di molto l'analisi senza una significativa perdita di precisione. A noi infatti interessa soltanto l'andamento della tensione di uscita, una volta che il livello logico di ingresso ha commutato il proprio valore. Quindi le funzioni del tempo $a(t)$, $b(t)$ e $c(t)$ diventano di fatto delle costanti perché, per quanto detto sopra, esse dipendono solamente dalla variabile $V_i(t)$ che abbiamo assunto essere costante subito dopo la transizione di livello.

Riscriviamo l'equazione 2.20 esplicitando la dipendenza dei vari termini da $V_o(t)$ e cambiando la notazione.

$$\begin{aligned} \dot{V}_o(t) &= a(t) + \frac{b(t)}{2} \left[e^{\frac{c(t)}{V_T}} e^{-\frac{V_o(t)}{V_T}} - e^{\frac{V_o(t)}{V_T}} e^{-\frac{c(t)}{V_T}} \right] = \\ &= A + B e^{-DV_o(t)} + C e^{DV_o(t)} \end{aligned} \quad (2.21)$$

dove i valori delle costanti sono:

$$\begin{aligned} \bullet A &= a(t) & \bullet C &= -\frac{b(t)}{2} e^{-\frac{c(t)}{V_T}} \\ \bullet B &= \frac{b(t)}{2} e^{\frac{c(t)}{V_T}} & \bullet D &= V_T^{-1} \end{aligned}$$

Ora nella Equazione 2.21 effettuiamo una sostituzione ponendo:

$$u(t) = e^{DV_o(t)} \implies \dot{u}(t) = D e^{DV_o(t)} \dot{V}_o(t) \quad (2.22)$$

e dopo alcuni passaggi algebrici otteniamo la seguente equazione differenziale non lineare del primo ordine:

$$\begin{aligned} \dot{u}(t) &= D e^{DV_o(t)} (C e^{DV_o(t)} + A + B e^{-DV_o(t)}) = \\ &= CD u(t)^2 + AD u(t) + BD \end{aligned} \quad (2.23)$$

Questa scrittura è la celebre equazione di Riccati che prende nome dai matematici italiani Jacopo e Vincenzo Riccati i quali, assieme ai Bernoulli, per primi ne determinarono le soluzioni.

Per risolvere l'equazione di Riccati si possono seguire diverse strade. Una di queste consiste nell'effettuare un'altra sostituzione al fine di ricondurci ad una equazione differenziale la cui risoluzione è più semplice.

Poniamo quindi

$$u(t) = -\frac{\dot{y}(t)}{CD y(t)} \implies \dot{u}(t) = -\frac{1}{CD} \left(\frac{\ddot{y}(t)y(t) - \dot{y}(t)^2}{y(t)^2} \right) = CD u(t)^2 - \frac{\ddot{y}(t)}{CD y(t)} \quad (2.24)$$

che sostituita nella Riccati ci permette di ottenere la seguente equazione differenziale.

$$\ddot{y}(t) - AD \dot{y}(t) + BCD^2 y(t) = 0 \quad (2.25)$$

Ci siamo ricondotti ad una equazione differenziale lineare del secondo ordine, omogenea ed a coefficienti costanti. È noto che la soluzione della 2.25 è una funzione della forma:

$$y(t) = c_1 e^{\gamma_1 t} + c_2 e^{\gamma_2 t} \quad (2.26)$$

con c_1 e c_2 parametri da determinare in base alla condizione iniziale e con γ_1 e γ_2 radici della equazione caratteristica 2.27 ottenuta dalla 2.25,

$$z^2 - ADz + BCD^2 = 0 \quad (2.27)$$

che fornisce:

$$\gamma_{1,2} = \frac{AD \pm \sqrt{A^2D^2 - 4BCD^2}}{2} = \frac{a(t) \pm \sqrt{a(t)^2 + b(t)^2}}{2V_T} \quad (2.28)$$

Osservazione: L'integrale (soluzione) di una generica equazione differenziale ordinaria di ordine n

$$F(x, y(x), y'(x), y''(x), \dots, y(x)^{(n)}) = 0 \quad (2.29)$$

è una funzione $\phi(t)$ che soddisfa le condizioni:

$$\phi(t) \in C^n(I) \text{ e } F(x, \phi(x), \phi'(x), \phi''(x), \dots, \phi(x)^{(n)}) = 0, \forall x \in I \subset \mathbb{R} \quad (2.30)$$

Solitamente i metodi risolutivi delle equazioni differenziali ci forniscono una famiglia di funzioni che soddisfano le condizioni 2.30. Nel nostro caso la soluzione 2.26 presenta due parametri c_1 e c_2 al cui variare si ottengono infinite soluzioni, cioè tutte le possibili combinazioni lineari dei due esponenziali.

Quindi, per determinare una sola tra tutte le infinite soluzioni dell'equazione differenziale, si mette a sistema tale famiglia di soluzioni con le condizioni iniziali del sistema. Si tratta del ben noto problema di Cauchy ed esiste un teorema che ci assicura, sotto opportune ipotesi, l'esistenza di un'unica funzione che soddisfi tutte le condizioni iniziali.

Qui però si presenta un inconveniente. Difatti il numero di condizioni iniziali da porre a sistema è pari all'ordine dell'equazione differenziale. L'equazione 2.20 è del primo ordine, quindi ci aspettiamo abbia come sola condizione iniziale $V_o(0)$, che rappresenta la tensione e quindi la carica presente nel condensatore all'istante di tempo $t = 0$. Da considerazioni fisiche sappiamo che tale variabile di stato è, e deve essere, unica poiché il condensatore è l'unico elemento presente nel nostro circuito in grado di accumulare energia. Nel passare dalla 2.20 alla 2.25 però l'ordine dell'equazione aumenta da 1 a 2. Quindi, dal punto di vista matematico sembrerebbe necessario imporre due condizioni iniziali, anche se ciò è in contrasto con le considerazioni fisiche esposte poco sopra. Dobbiamo quindi mettere d'accordo la matematica, che richiede due condizioni iniziali, con la fisica che ne mette a disposizione soltanto una. In realtà adesso mostreremo che le due condizioni iniziali richieste dalla Formula 2.25 non sono tra loro indipendenti. Dalle sostituzioni 2.22 e 2.24 valutate all'istante $t = 0$ possiamo scrivere:

$$e^{DV_o(0)} = -\frac{\dot{y}(0)}{CDy(0)} \quad (2.31)$$

da cui si vede che, dato $V_o(0)$, non siamo interessati ai valori di $\dot{y}(0)$ e $y(0)$ ma solo al loro rapporto. Quindi, a partire dall'unica condizione iniziale a nostra disposizione $V_o(0)$, possiamo ottenere le altre due $y(0)$ e $\dot{y}(0)$ semplicemente fissandone una arbitrariamente e ricavando l'altra di conseguenza. Fissiamo per esempio arbitrariamente $\dot{y}(0) = 1$, dalla Formula 2.31 ricaviamo immediatamente:

$$y(0) = -\frac{1}{DCe^{DV_o(0)}} \quad (2.32)$$

Fatta questa necessaria osservazione possiamo finalmente ricavare i parametri c_1 e c_2 :

$$\begin{cases} y(0) & = c_1 + c_2 \\ \dot{y}(0) & = 1 = c_1\gamma_1 + c_2\gamma_2 \end{cases} \Rightarrow \begin{cases} c_1 = \frac{V_T \left(1 + \frac{-a(t) + \sqrt{a(t)^2 + b(t)^2}}{V_o(0) - c(t)} \right)}{\sqrt{a(t)^2 + b(t)^2}} \\ c_2 = \frac{V_T \left(-1 + \frac{a(t) + \sqrt{a(t)^2 + b(t)^2}}{V_o(0) - c(t)} \right)}{\sqrt{a(t)^2 + b(t)^2}} \end{cases} \quad (2.33)$$

A questo punto da $y(t)$, soluzione dell'equazione differenziale 2.25, e dalla trasformazione 2.31 (considerando ovviamente non più gli istanti iniziali ma le funzioni nel tempo) otteniamo la soluzione dell'equazione differenziale iniziale 2.20:

$$V_o(t) = V_T \log \left(-\frac{\dot{y}(t)}{CDy(t)} \right) = V_T \log \left(\frac{V_T}{C} \right) + V_T \log \left(-\frac{\dot{y}(t)}{y(t)} \right) = c(t) + V_T \log \left(\frac{2V_T \dot{y}(t)}{b(t) y(t)} \right) \quad (2.34)$$

$$\text{in cui l'argomento del logaritmo è: } \frac{2V_T \dot{y}(t)}{b(t) y(t)} = \frac{2V_T c_1 \gamma_1 e^{\gamma_1 t} + c_2 \gamma_2 e^{\gamma_2 t}}{b(t) c_1 e^{\gamma_1 t} + c_2 e^{\gamma_2 t}} =$$

$$\text{(moltiplicando numeratore e denominatore per } e^{-\frac{a+\sqrt{a^2+b^2}}{2V_T}t} \sqrt{a^2+b^2} \text{)}$$

$$= \frac{\left(1 + \frac{-a+\sqrt{a^2+b^2}}{\frac{V_0(0)-c}{be \frac{V_T}{V_T}}} \right) (a + \sqrt{a^2+b^2}) e^{\frac{\sqrt{a^2+b^2}}{V_T}t} + \left(-1 + \frac{a+\sqrt{a^2+b^2}}{\frac{V_0(0)-c}{be \frac{V_T}{V_T}}} \right) (a - \sqrt{a^2+b^2})}{b \left(1 + \frac{-a+\sqrt{a^2+b^2}}{\frac{V_0(0)-c}{be \frac{V_T}{V_T}}} \right) e^{\frac{\sqrt{a^2+b^2}}{V_T}t} + b \left(-1 + \frac{a+\sqrt{a^2+b^2}}{\frac{V_0(0)-c}{be \frac{V_T}{V_T}}} \right)} \quad (2.35)$$

La tensione sul nodo di uscita allo stato stazionario (steady-state), esauriti i transienti di carica e scarica, è ricavata facendo il limite al tempo infinito dell'equazione 2.34.

$$V_{ss}(t) = \lim_{t \rightarrow +\infty} v_o(t) = c + V_T \log \left(\frac{a + \sqrt{a^2+b^2}}{b} \right) \quad (2.36)$$

Tali andamenti sono mostrati nelle Figure 2.9 e 2.10.

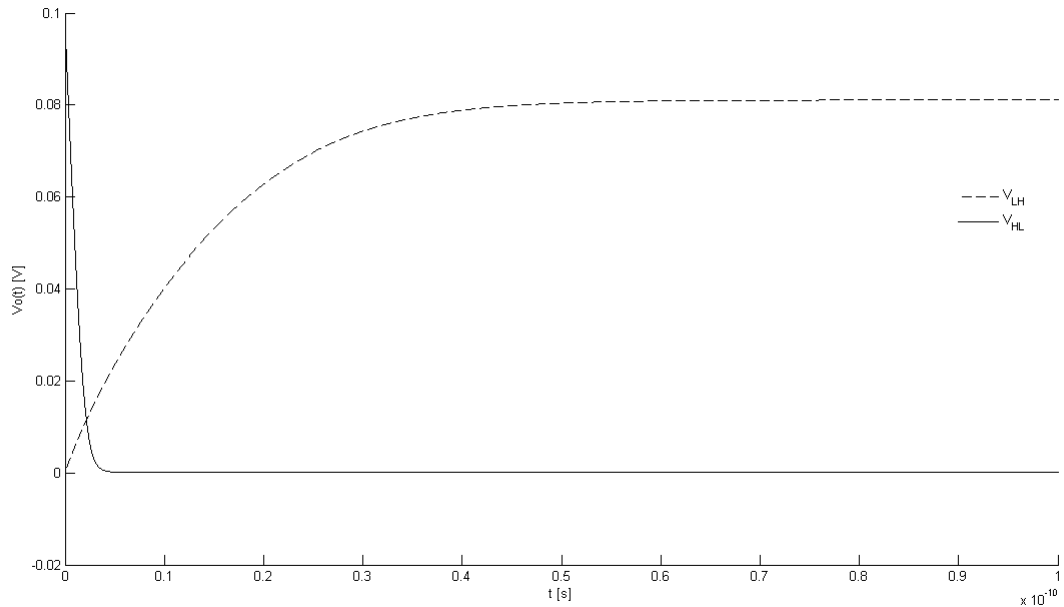


Figura 2.9: Andamento temporale ($V_{dd} = 100mV$) dell'inverter CMOS nelle transizioni H-L ed L-H a temperatura $T = 300K$.

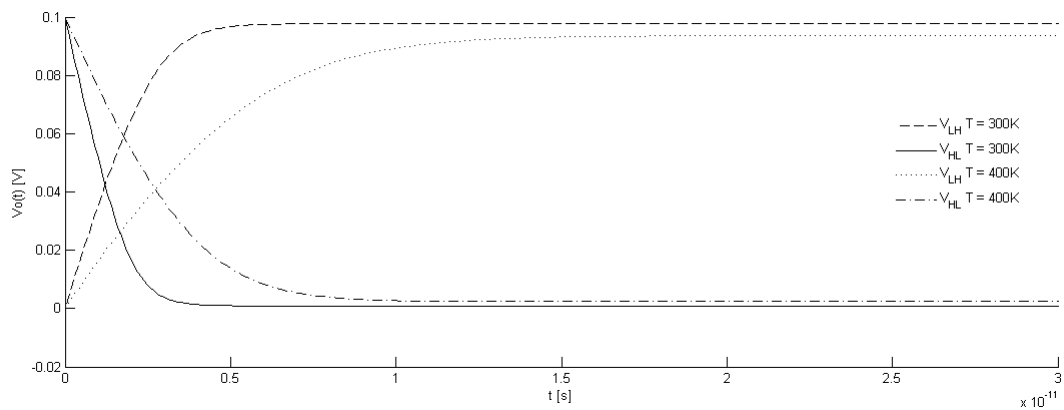


Figura 2.10: Andamento temporale ($V_{dd} = 100mV$) dell'inverter CMOS bilanciato incrementando la larghezza di canale W_p del pMOS di un fattore moltiplicativo $\chi = 6.3$. Curve a temperatura $T = 300K$ e a $T = 400K$.

CAPITOLO

3

Modellazione degli effetti non deterministici

3.1 Concetti fondamentali

Prima di studiare come modellizzare matematicamente gli effetti fisici non deterministici presenti nell'inverter CMOS, riassumiamo i concetti base su cui si fondano gli strumenti matematici che ci apprestiamo ad utilizzare.

3.1.1 Processi stocastici

Definizione 3.1 (Processo stocastico a tempo e valori continui). Sia Ω uno spazio campionario ed \mathbb{R} lo spazio dei numeri reali. Si definisce processo stocastico una applicazione a due variabili a valori reali così definita:

$$X : (\Omega, \mathbb{R}) \longrightarrow \mathbb{R} \quad (3.1)$$

Un semplice esempio di processo aleatorio definito su uno spazio campionario discreto (lancio di un dado) è il seguente. Consideriamo la funzione $X(\omega_i, t) = \sin(2\pi\omega_i f_0 t)$ in cui, oltre alla dipendenza dal tempo $t \in \mathbb{R}$, la sinusoide dipende dalla effettiva realizzazione $\omega_i \in \Omega$. Se, per esempio, consideriamo come spazio campionario Ω i sei possibili numeri rappresentati sulle facce di un dado equilibrato, otteniamo che $\omega_1, \omega_2, \dots, \omega_6 = 1, 2, \dots, 6$, assunti ciascuno con probabilità pari a $\frac{1}{6}$. Quindi, il processo aleatorio così costruito avrà come realizzazioni delle funzioni sinusoidali la cui frequenza sarà data dal prodotto tra la frequenza base (f_0) con un numero non noto a priori. Tale cifra è il numero "scelto aleatoriamente" tramite il lancio di un dado a sei facce. L'insieme di funzioni dipendenti dalla realizzazione ω_i e dal tempo t è quindi ciò che chiamiamo **processo aleatorio**. Da notare che in tutta la trattazione di questa tesi la variabile reale del processo stocastico sarà il tempo e verrà indicata con il

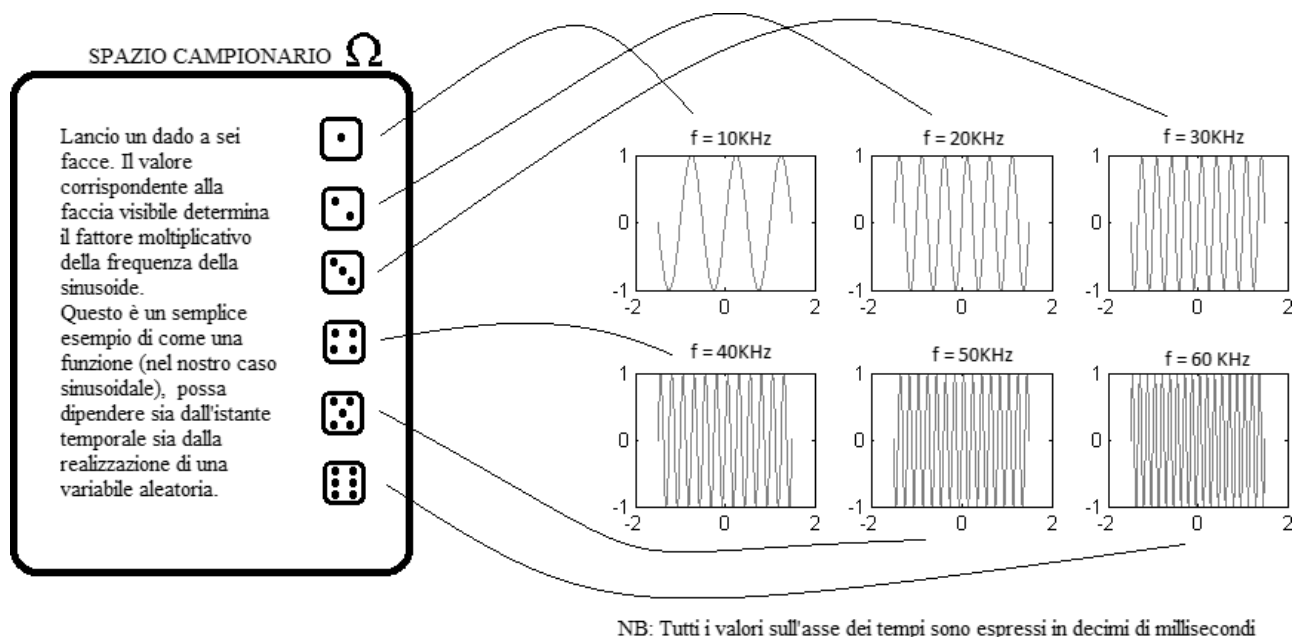


Figura 3.1: Rappresentazione dell'esempio esplicativo sui processi stocastici

simbolo t . Inoltre, appena non ci sarà il pericolo di confusione, si preferirà indicare il processo aleatorio semplicemente con il simbolo $X(t)$ o X_t al posto di $X(\omega_i, t)$, omettendo di fatto la dipendenza implicita dalla realizzazione ω_i .

Un processo stocastico ha quattro possibili interpretazioni differenti. Si faccia attenzione al differente uso tra X e x che rispecchia le convenzioni usuali per indicare variabili e funzioni.

- $X(\cdot, \cdot)$ Processo aleatorio, cioè famiglia di funzioni deterministiche dipendenti dal tempo t e dalla particolare realizzazione di un esperimento.
- $x(\omega_i, \cdot)$ Realizzazione del processo stocastico o funzione campione. È una funzione deterministica "estratta" a caso in base all'esito di un esperimento.
- $X(\cdot, t_i)$ Variabile aleatoria. Preso un processo stocastico e, osservando ciascuna realizzazione nello stesso istante di tempo t_i , otteniamo una serie di numeri dipendenti dall'esito dell'esperimento.
- $x(\omega_i, t_i)$ Numero reale. Scelta una realizzazione ed un istante di tempo si ottiene un numero reale.

Per descrivere esattamente l'andamento di un processo stocastico è necessario conoscere l'andamento del processo per ogni realizzazione, oltre alla distribuzione statistica dell'esperimento definito su Ω . Cioè, in pratica, bisognerebbe conoscere la densità di probabilità congiunta per tutti gli ordini e per tutti gli istanti di tempo. Anche se ottenere questo grado di conoscenza è impossibile nelle applicazioni pratiche, spesso è sufficiente conoscere le statistiche descrittive del processo aleatorio per poterlo studiare ed utilizzare con efficacia in molti ambiti di interesse.

Riportiamo brevemente le formule per ricavare alcuni parametri statistici del primo e secondo ordine utili per la comprensione della tesi.

Sia $f_x(x, t)$ la funzione di densità di probabilità della variabile aleatoria $X_t := X(\cdot, t)$ di un processo aleatorio X . Definiamo come **valore medio** del processo X la funzione $m_X(t)$ generalmente dipendente dal tempo:

$$m_X(t) = E[X(t)] = \int x f_X(x, t) dx \quad (3.2)$$

Tale espressione equivale a calcolare il valore atteso di ogni variabile aleatoria (estrapolata dal processo) in ogni istante t .

Definiamo invece la **varianza**:

$$\sigma_X^2(t) = \text{var}[X(t)] = \int (x - m_X(t))^2 f_X(x, t) dx \quad (3.3)$$

Funzione di **autocovarianza**:

$$\mathbf{c}_X(t_1, t_2) = E [[X(t_1) - m_{X_{t_1}}][X(t_2) - m_{X_{t_2}}]] \quad (3.4)$$

Funzione di **autocorrelazione**:

$$\mathbf{r}_X(t_1, t_2) = E [X(t_1)X(t_2)] \quad (3.5)$$

È immediato notare che, se il valore medio del processo aleatorio è nullo, allora $m_X(t) = 0$ e quindi si ha che funzioni di autocovarianza $\mathbf{c}_X(t_1, t_2)$ e autocorrelazione $\mathbf{r}_X(t_1, t_2)$ coincidono.

La funzione di autocorrelazione $\mathbf{r}_X(t_1, t_2)$, intuitivamente, si può pensare come una funzione che descrive la somiglianza tra le variabili aleatorie $X(t_1)$ e $X(t_2)$.

3.1.2 Passeggiata aleatoria 1D (a random walk)

La passeggiata casuale descrive la posizione di una particella che si muove, a tempo discreto $t = nT$ con $n \in \mathbb{N}$ e T il quanto temporale, sulla retta reale. All'istante iniziale $n = 0$ la particella si trova nell'origine. Negli istanti successivi, $n > 0$, la particella si sposta istantaneamente di $+1$ o di -1 , rispetto alla posizione occupata all'istante $n - 1$, a seconda del risultato del lancio di una moneta equilibrata. I lanci della moneta si suppongono indipendenti. Indicando con X_n la variabile aleatoria (V.A.) binaria ($0 = \text{Croce}$, $1 = \text{Testa}$) che rappresenta il risultato del lancio n -esimo della moneta, la posizione della particella sulla retta all'istante n è descritta dalle variabili aleatorie $S_0 = 0$ e, per $n \geq 1$,

$$S_n = \sum_{i=1}^n X_i - \left(n - \sum_{i=1}^n X_i \right) = 2 \sum_{i=1}^n X_i - n$$

È immediato ricavare, usando l'indipendenza delle V.A. X_n ,

$$E[S_n] = 0, \quad \text{var}(S_n) = n$$

In Figura 3.2 sono tracciate alcune traiettorie di una passeggiata casuale, simulate in MATLAB. È interessante osservare che, data l'indipendenza delle v.a. X_n , il processo S_n è ad incrementi indipendenti e stazionari. L'indipendenza degli incrementi significa che per $n_1 \leq n_2 \leq n_3 \leq n_4$

$$S_{n_4} - S_{n_3} = \sum_{i=n_3+1}^{n_4} X_i, \quad S_{n_2} - S_{n_1} = \sum_{i=n_1+1}^{n_2} X_i,$$

sono indipendenti, essendo infatti funzioni di v.a. X_i di indici diversi. La stazionarietà degli incrementi significa che, per ogni k , le distribuzioni di $S_{n+k} - S_k$ e di $S_n - S_0 = S_n$ sono identiche, essendo infatti entrambi gli incrementi somme di n variabili i.i.d.

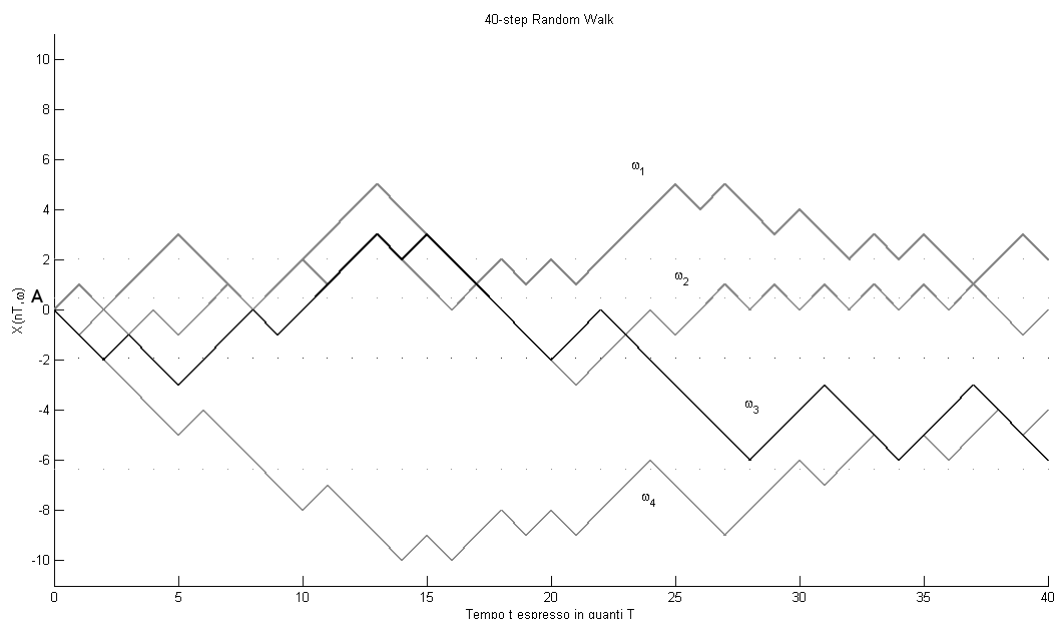


Figura 3.2: Simulazione MATLAB di 4 realizzazioni di una random walk con 40 passi

3.1.3 Processo di Wiener

Il processo di Wiener, che avrà un ruolo fondamentale nel seguito della tesi, descrive la posizione di una particella che si muove, a tempo continuo, sulla retta reale, partendo dall'origine all'istante $t = 0$. Forniamo qui la derivazione del processo di Wiener come limite di una passeggiata casuale. Allo scopo, per $t > 0$ fissato, si definiscano n molto grande e Δt molto piccolo tali che $t = n\Delta t$. Il processo di Wiener all'istante t è il limite (in distribuzione)

$$W(t) = \lim_{n \rightarrow \infty} S_n \Delta x$$

dove S_n è una passeggiata casuale come sopra. $W(t)$ si può quindi interpretare come limite di una passeggiata casuale che "salta" agli istanti $k\Delta t$, $k = 1, 2, \dots, n$, effettuando passi di ampiezza $\pm\Delta x$ negli istanti di salto, a seconda del risultato del lancio della moneta. Come vedremo immediatamente le quantità Δx e Δt non possono essere scelte in modo arbitrario se si desidera che il processo $W(t)$ abbia varianza non nulla e finita.

Osservazione: La varianza di una V.A o di un processo stocastico, di norma, soddisfa la seguente disuguaglianza: $0 < \text{var}(\cdot) < \infty$. Ciò è particolarmente importante nella nostra trattazione poiché avere una varianza nulla equivarrebbe ad avere una variabile aleatoria triviale di valore costante centrata nel proprio valore medio. Avere una varianza infinita implica invece la trattazione di segnali la cui potenza non è limitata. Tale condizione è da evitare accuratamente sia per problemi di tipo matematico, molti teoremi (es: il CLT) non sono più validi, sia per considerazioni fisiche. Infatti avere un segnale a potenza infinita, nel nostro caso, richiederebbe un generatore di tensione anch'esso di potenza infinita. Ciò è chiaramente irrealizzabile nella pratica, quindi questi due sono casi limite da evitare opportunamente.

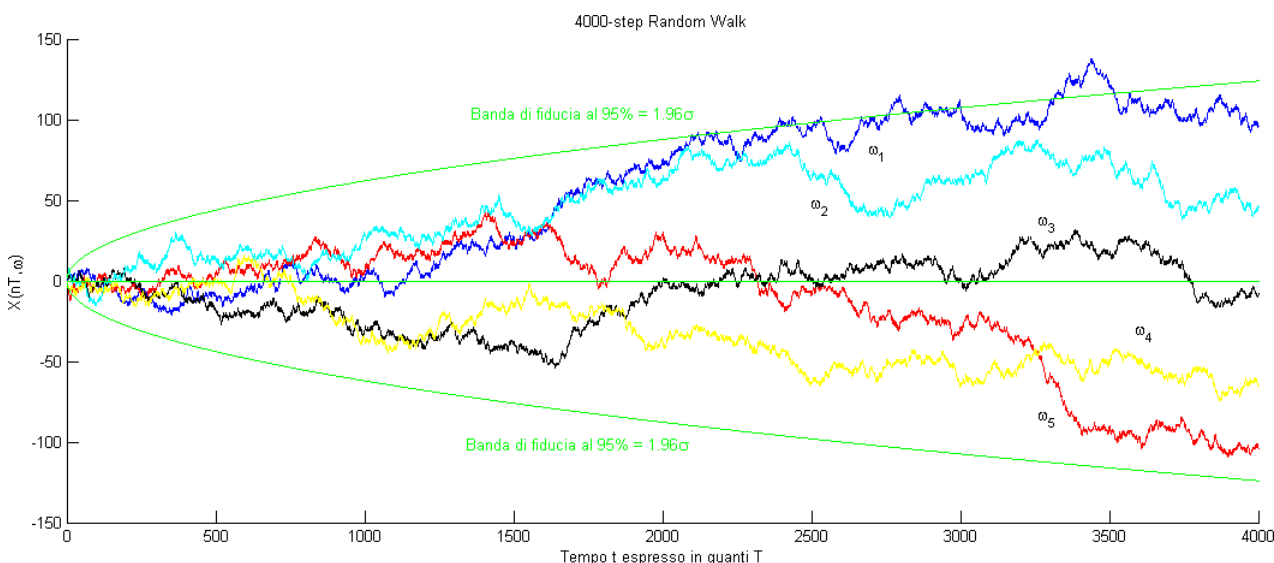


Figura 3.3: Simulazione MATLAB di alcune realizzazioni di un processo di Wiener con banda di fiducia al 95 %. Ciò significa che le varie realizzazioni del processo aleatorio hanno un valore che è compreso entro $\pm 1.96\sigma$ con probabilità $p = 0.95$.

In virtù della definizione il processo di Wiener eredita dalla passeggiata casuale S_n la media nulla, $E[W(t)] = 0$ per ogni $t \geq 0$, inoltre gli incrementi di $W(t)$ sono indipendenti e stazionari. Grazie a queste proprietà, scrivendo

$$W(t+s) = [W(t+s) - W(s)] + [W(s) - W(0)],$$

possiamo ricavare un'equazione funzionale per la varianza di $W(t)$. Sia $V(t) := \text{var}(W(t))$ allora:

$$\begin{aligned} V(t+s) &= E(W(t+s)^2) = E((W(t+s) - W(s)) + (W(s) - W(0)))^2 \\ &= E([W(t+s) - W(s)]^2) + E([W(s) - W(0)]^2) \\ &= E(W(t)^2) + E(W(s)^2) = V(t) + V(s). \end{aligned}$$

La soluzione generale dell'equazione funzionale

$$V(t+s) = V(t) + V(s)$$

è chiaramente $V(t) = ct$. Poichè però $V(t)$ è una varianza, il vincolo di positività impone $c > 0$ (si noti che $t > 0$ per costruzione: il processo di Wiener è definito per $t > 0$). Per convenzione la costante si indica σ^2 e quindi

$$V(t) = \sigma^2 t$$

Il valore di σ^2 deve essere in accordo con quello che si ottiene dalla definizione iniziale del processo di Wiener. Poichè $\text{var}(S_n) = n$ la varianza $V(t)$ si può anche esprimere come

$$\begin{aligned} V(t) &= \text{var}(W(t)) = \lim_{n \rightarrow \infty} \text{var}(S_n \Delta x) \\ &= \lim_{n \rightarrow \infty} n(\Delta x)^2 = \lim_{n \rightarrow \infty} \frac{(\Delta x)^2}{\Delta t} t, \end{aligned}$$

e poichè $V(t) = \sigma^2 t$ si deve scegliere Δx tale che

$$\frac{(\Delta x)^2}{\Delta t} = \sigma^2.$$

Per il processo di Wiener standard $W(t)$ si sceglie $(\Delta x)^2 = \Delta t$, si avrà quindi $\sigma^2 = 1$.

Il teorema del limite centrale ci permette di desumere la distribuzione del processo $W(t)$. Infatti

$$\frac{S_n}{\sqrt{n}} \Rightarrow N(0, 1),$$

dove \Rightarrow denota la convergenza in distribuzione. Dalla definizione di $W(t)$, e per quanto detto sopra, si ricava quindi che

$$\begin{aligned} W(t) &= \lim_{n \rightarrow \infty} \frac{S_n}{\sqrt{n}} \sqrt{n} \Delta x \\ &\Rightarrow N(0, 1) \frac{\Delta x}{\sqrt{\Delta t}} \sqrt{t} \\ &= N(0, 1) \sqrt{t} \\ &= N(0, t) \end{aligned}$$

Il processo di Wiener è spesso conosciuto anche con il nome di **Moto Browniano**. Tale nome è dovuto al fatto che il processo di Wiener descrive molto bene il moto disordinato di particelle presenti in sospensioni fluide, fenomeno osservato sperimentalmente nel 1828 dal naturalista R. Brown, e studiato matematicamente nel '900 da Einstein e Bachelier. Brown osservò che il polline, in una sospensione acquosa, si muoveva disordinatamente disperdendosi nel liquido. La causa di questo moto inaspettato è legata al fatto che la particella di polline è soggetta ad un elevatissimo numero di urti continui con le molecole di fluido in cui essa è immersa. Ciascun urto è però caratterizzato da un vettore forza (intensità ed angolazione) non deterministico. La sommatoria di queste forze aleatorie determina così lo spostamento casuale delle particelle di polline.

Dunque, il processo di Wiener precedentemente costruito presenta 3 fondamentali **proprietà**:

1. $W(0) = 0$ Se così non fosse ci troveremmo in presenza di una “componente continua” deterministica sommata al processo di Wiener. Tale componente è facilmente isolabile al fine di ricondurci ad un processo con valore iniziale pari a zero.
2. $W(t)$ è continuo quasi certamente.
3. Il processo $W(t)$ presenta incrementi indipendenti e stazionari ovvero, se $0 \leq s_1 \leq t_1 \leq s_2 \leq t_2$, allora $W(t_1) - W(s_1)$ e $W(t_2) - W(s_2)$ sono V.A. indipendenti in cui

$$W(t_2) - W(t_1) \sim N(0, t_2 - t_1)$$

per il C.L.T.

Da tali proprietà possiamo derivare almeno altre due utili **conseguenze**:

1. Per ogni istante t^* , $W(t^*)$ è una V.A. che ha distribuzione $\mathcal{N}(0, t^*)$
2. $E[W(t)] = 0$ Facile da dedurre in base alla affermazione precedente.

Studiamo ora la funzione di autocorrelazione:

$$\mathbf{c}_W(t_1, t_2) = E[W(t_1)W(t_2)] \tag{3.6}$$

$$t_1 > t_2 > 0$$

$$\begin{aligned} \mathbf{c}_W(t_1, t_2) &= E[W(t_1)W(t_2)] = E[(W(t_1) - W(t_2) + W(t_2))(W(t_2) - W(0))] = \\ &=^{\#} E[(W(t_1) - W(t_2))(W(t_2) - W(0)) + (W(t_2)W(t_2) - W(t_2)W(0))] = \\ &=^* 0 + E[W(t_2)^2] = \text{var}(W(t_2)) = t_2 \end{aligned} \tag{3.7}$$

(#,*) tali uguaglianze sono giustificate dalla *proprietà 3*, dalla *utile conseguenza 1* del processo di Wiener e dalla linearità del valore atteso $E[\cdot]$ per gli eventi indipendenti.

Per simmetria, se $0 < t_1 < t_2$, scambiando i ruoli di t_1 e t_2 otteniamo:

$$\mathbf{c}_W(t_1, t_2) = \text{var}(W(t_1)) = t_1 \tag{3.8}$$

Quindi possiamo riscrivere la funzione autocovarianza che, come già sottolineato, nel nostro caso è uguale alla funzione aturocorrelazione, come:

$$\mathbf{c}_W(t_1, t_2) = \min(t_1, t_2) \tag{3.9}$$

Per meglio comprenderne l'andamento possiamo aiutarci con la sua rappresentazione in 3D mostrata in Figura 3.4.

Forniamo ora l'enunciato di un importante teorema che si rivelerà essere fondamentale al fine di legare il modello matematico con il concetto fisico di rumore.

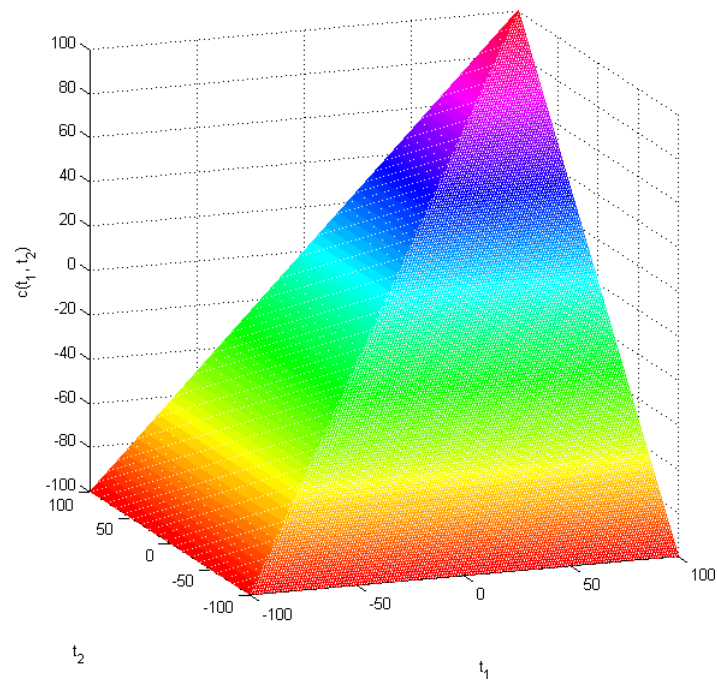


Figura 3.4: Grafico MATLAB della funzione di autocorrelazione $\mathbf{c}_W(t_1, t_2)$ per il processo di Wiener

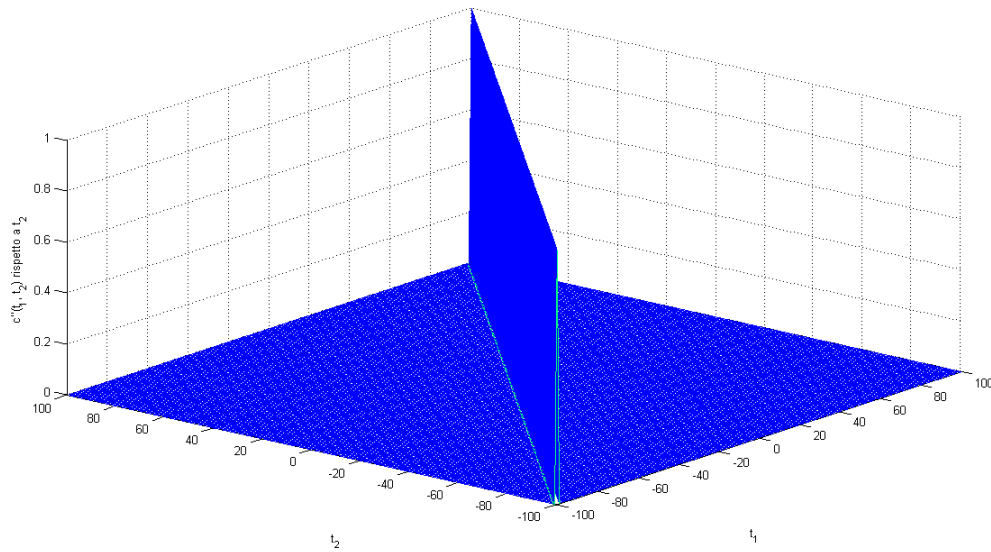


Figura 3.5: Grafico MATLAB della derivata seconda della funzione correlazione $\frac{\partial^2 \mathbf{c}_W(t_1, t_2)}{\partial t_1 \partial t_2}$

Teorema 3.1. Sia dato un processo stocastico $W(t)$ la cui autocovarianza è $\mathbf{c}_W(t, s)$. Allora il processo aleatorio $\eta(t)$ costruito derivando il processo di partenza (cioè: $\eta(t) = \frac{dW(t)}{dt}$) ha autocovarianza

$$\mathbf{c}_\eta(t, s) = \frac{\partial^2 \mathbf{c}_W(t, s)}{\partial t \partial s} \quad (3.10)$$

Procedendo formalmente, ricaviamo la covarianza della derivata del processo di Wiener. Calcoliamo quindi le derivate parziali della funzione correlazione rispetto alle variabili t_1 e t_2 che, non dimentichiamo, non sono altro che istanti temporali di una realizzazione del processo $W(t)$.

$$\frac{\partial \mathbf{c}_W(t_1, t_2)}{\partial t_1} = \begin{cases} 1 & \text{se } t_1 > t_2 \\ 0 & \text{se } t_1 < t_2 \end{cases}, \quad \frac{\partial^2 \mathbf{c}_W(t_1, t_2)}{\partial t_1 \partial t_2} = \delta(t_2 - t_1) \quad (3.11)$$

Questa funzione è rappresentata graficamente nella Figura 3.5.

3.2 Il rumore

Il rumore, in generale, è una componente indesiderata e quasi sempre aleatoria che si sovrappone al segnale utile modificandone l'andamento temporale e le caratteristiche spettrali. Tale distorsione altera l'informazione associata al segnale originale, a volte in maniera talmente pesante e distruttiva da non permettere di estrarre l'informazione desiderata. Il rumore, nei circuiti elettronici, si manifesta come variazioni di tensione o di corrente dovute alle caratteristiche fisiche intrinseche alla materia di cui sono composti. Tali fluttuazioni sono l'effetto macroscopico di un insieme elevatissimo di alterazioni e movimenti microscopici. Per la natura aleatoria del rumore, dal punto di vista matematico si fa largo uso dei processi stocastici per la sua modellizzazione.

Esistono (almeno quattro) differenti "sorgenti" di rumore all'interno di un dispositivo elettronico. In questo capitolo analizzeremo, per brevità, esclusivamente il rumore termico ed il rumore shot, che in genere sono le sorgenti di rumore più rilevanti in dispositivi elettronici attivi quali sono i MOSFET.

3.2.1 Rumore termico

Scoperto sperimentalmente da Johnson ed inquadrato matematicamente da Nyquist alla fine degli anni 20, il rumore termico è una delle sorgenti di rumore più diffuse e rilevanti nel mondo dei dispositivi elettronici. Esso è dovuto all'agitazione termica dei portatori di carica (nel nostro caso trattasi di elettroni) dei materiali impiegati. Tali elettroni infatti, non trovandosi alla temperatura di zero assoluto $T = 0K$, si muovono, vibrano, attorno al loro punto di equilibrio a causa della loro energia cinetica $E_c = \frac{3}{2}k_bT$ non nulla. Per avere riscontro pratico delle ripercussioni che ha questo fenomeno si immagina di prendere un resistore di valore $R > 0$, posto ad una temperatura costante $T > 0$ e di misurare ai suoi capi la d.d.p. con un voltmetro sufficientemente sensibile. Osserveremo che la tensione misurata dallo strumento avrà come valore medio zero Volt ma che, con l'evolversi del tempo, la tensione misurata varierà in maniera aleatoria. Questa piccola variazione di tensione è dovuta al movimento non ordinato degli elettroni liberi che, essendo tantissimi e muovendosi rapidamente, si concentrano istantaneamente ed aleatoriamente ad uno dei capi del resistore. Tale fenomeno è anche termodinamicamente giustificabile considerando che, così come un elemento resistivo è in grado di trasformare potenza elettrica in calore, allora è possibile che avvenga anche il processo contrario, cioè che il calore si trasformi in potenza elettrica (e quindi in tensione ai capi del resistore).

Teoricamente, in una delle pietre miliari della ricerca scientifica, Nyquist ha dimostrato che ai capi di un resistore di resistenza R , posto alla temperatura $T > 0$ e mantenuto isolato elettricamente, è presente un valore di tensione di media nulla, e di potenza spettrale (PSD - Power Spectral Density) costante pari a:

$$\mathcal{P}_{w_thm}(f) = 2k_bTR \quad (3.12)$$

con al solito $k_b =$ costante di Boltzmann. Tale relazione è un'approssimazione, in accordo con le osservazioni sperimentali, per frequenze di utilizzo inferiori al THz. Indicando con Δf la banda di osservazione (solo frequenze positive), ed essendo il valore medio m_V della tensione nullo, la varianza σ_V^2 equivale di fatto alla potenza statistica M_V . (Per approfondimenti si faccia riferimento a [6] Cap.1.7).

Da tali considerazioni possiamo quindi ricavare la varianza del rumore:

$$\int_{-\infty}^{\infty} \mathcal{P}_{w_thm}(f)df = M_V = 0 + \sigma_V^2 = 4k_bTR\Delta f \quad [V^2] \quad (3.13)$$

Otteniamo che la distribuzione del rumore termico in un componente è una Normale di media nulla e varianza σ_V^2 . Si ponga attenzione alla dipendenza lineare dalla temperatura T e dalla resistenza R , l'unico modo per ridurre quindi l'effetto di tale rumore è di abbassare la temperatura il più possibile. Inoltre si noti che tale rumore è una caratteristica intrinseca di tutti i componenti dissipativi (cioè che hanno resistenza R diversa da zero).

3.2.2 Rumore shot

Il rumore shot è un tipo di rumore di natura corpuscolare scoperto da W. Schottky nel 1918. In elettronica tale fenomeno è dovuto ad alcuni elettroni che dispongono di una energia abbastanza elevata da poter superare le barriere di potenziale come, per esempio, quelle presenti nelle giunzioni $p-n$. Questo piccolo gruppo di elettroni ad alta energia si ha solamente in presenza di un flusso di corrente all'interno del dispositivo elettronico ed è dipendente dall'intensità di tale flusso. La dipendenza non è però lineare ed il fenomeno del rumore shot è particolarmente rilevante per basse intensità di correnti.

Possiamo immaginare il rumore shot all'interno di un dispositivo MOSFET come un insieme di elettroni ad elevata energia che, ad intervalli ed istanti aleatori, riescano a superare le barriere di potenziale e quindi a fluire dal *source* al *drain* (o viceversa in accordo con i valori delle barriere potenziali) del transistor. Ciascuno di questi elettroni, che arriva ad uno dei due terminali, contribuisce alla corrente di rumore con un quanto di carica e quindi di corrente. Il movimento di questi elettroni viene modellato come un processo di Poisson. Cioè, ogni evento discreto "arrivo di un elettrone al terminale (sia esso *source* o *drain*)" si verifica in maniera aleatoria ed indipendente dagli altri arrivi. Questa modellazione è giustificata dal fatto che il numero di eventi che si verificano tra l'istante $N(t_1)$ e l'istante $N(t_2)$ ha una distribuzione di probabilità discreta che segue la legge di Poisson $\sim P(\lambda)$. Con λ indichiamo il parametro di tale processo di Poisson, cioè il **rate** dell'evento "arrivo di un elettrone al terminale".

Omettendo i passaggi matematici riportiamo da [5] la PSD del rumore shot che è pari a:

$$\mathcal{P}_{w_shot}(f) = qI \quad (3.14)$$

3.2.3 Modello unificato

In realtà, anche se tali sorgenti di rumore sono state analizzate e presentate come differenti fenomeni fisici, possono essere unificate in un unico modello di rumore bianco. La giustificazione di questa semplificazione è stata introdotta da un gruppo di ricerca condotta da C.A. Mead ed è illustrata con tutti i dettagli matematici nell'articolo[7]. Per non dilungarci omettiamo tale dimostrazione ma, la spiegazione fondamentale risiede nel fatto che l'energia degli elettroni che sono in grado di superare la barriera di potenziale è legata anch'essa alla temperatura e quindi rumore termico e rumore shot, accadendo simultaneamente, si possono identificare come un'unica sorgente di rumore governata dalla temperatura.

Ricordiamo che la trasformata di Fourier della autocorrelazione è la PSD:

$$\mathcal{P}(f) = \mathcal{F}[\mathbf{r}(t_1, t_2)] \Rightarrow \mathbf{r}(t_1, t_2) = \mathcal{F}^{-1}[\mathcal{P}(f)] \quad (3.15)$$

Quindi, se utilizziamo come PSD l'espressione riportata in 3.14 otteniamo:

$$\mathcal{P}_w(f) = qI(f) = q(I_p(f) + I_n(f)) \implies \mathbf{r}(t_1, t_2) = \mathcal{F}^{-1}[\mathcal{P}_w(f)] = q(I_p(t) + I_n(t))\delta(t) \quad (3.16)$$

Si nota che tale risultato è proprio la derivata seconda dell'autocorrelazione di un processo di Wiener (Figura 3.5) quindi, in base al Teorema 3.1, si può finalmente affermare che, dato un circuito elettronico, il rumore termico e shot (che abbiamo già detto essere due facce della stessa medaglia), si possono modellare come un processo che è la derivata del processo di Wiener.

3.3 Il modello stocastico dell'inverter CMOS in regione di sottosoglia

A questo punto si è capito che l'Equazione 2.16 che regola la tensione di uscita del circuito deve essere modificata al fine di poter tenere conto degli effetti non deterministici. È quindi sufficiente aggiungere una componente aleatoria $\xi(t)$ che rappresenta le piccole variazioni di corrente nel condensatore dovute al rumore.

$$\frac{dV_o(t)}{dt} = \frac{I_p(t) - I_n(t) + \xi(t)}{C_L} \quad (3.17)$$

Tale termine non è altro che un processo aleatorio ottenuto derivando il processo di Wiener,

$$\xi(t) = \sqrt{\mathbf{c}_w} \frac{dW(t)}{dt} \quad (3.18)$$

in cui l'elemento radice quadrata dell'autocorrelazione è semplicemente un termine di normalizzazione.

Ora, ponendo $\Gamma(t) = \frac{q(I_p(t)+I_n(t))}{C_L^2}$, otteniamo dalla Formula 2.20 e dalla 3.18 l'equazione differenziale che regola la tensione di uscita dell'inverter CMOS considerando l'effetto del rumore. Si ricordi che $a(t), b(t), c(t)$ e $d(t)$ sono gli stessi simboli utilizzati all'inizio dell'esposizione.

$$\boxed{dV_o(t) = \left(a(t) + b(t) \sinh \left(\frac{c(t) - V_o(t)}{V_T} \right) \right) dt + \sqrt{\Gamma(t)} dW(t)} \quad (3.19)$$

Questa nuova equazione differenziale non lineare è concettualmente molto differente dall'Equazione 2.20 per via del processo stocastico $dW(t)$. Tale componente non deterministica ha infatti introdotto un ente matematico difficilmente manipolabile con gli strumenti messi a disposizione dall'analisi matematica. Per poter fornire una interpretazione di tale **Equazione Differenziale Stocastica (SDE)** è necessario ridefinire opportunamente il concetto di integrale per questo contesto.

CAPITOLO

4

Risoluzione ed analisi dei risultati

4.1 L'integrale stocastico e le SDE

4.1.1 Differenziabilità del processo di Wiener

Consideriamo un lasso di tempo dt infinitesimo. Se campioniamo il processo di Wiener in due istanti di tempo che differiscono di dt e ne facciamo la differenza otteniamo:

$$W_{t+dt} - W_t = \delta W_t = \text{(per la proprietà 3)} = \mathcal{N}(0, dt) = \mathcal{N}(0, 1)\sqrt{dt} \quad (4.1)$$

Quindi, dividendo ambo i membri per dt otteniamo la scrittura canonica utilizzata per rappresentare il differenziale di una funzione (in tale circostanza si tratta di un processo aleatorio).

$$\frac{W_{t+dt} - W_t}{dt} = \frac{\mathcal{N}(0, 1)}{\sqrt{dt}} \quad (4.2)$$

Si noti che, ragionando con gli infinitesimi, mentre dt al denominatore tende a zero, lo stesso non si può dire per il numeratore che si comporta come una V.A. $\sim \mathcal{N}(0, 1)$ indipendentemente da dt . Il processo di Wiener è dunque a variazione non limitata cioè, intuitivamente e per una dimensione, possiamo pensare che la distanza che percorre un punto sul grafico della funzione, **non** tende a zero man mano che diminuiamo l'intervallo di osservazione dt . Possiamo quindi concludere affermando che il processo di Wiener è un processo continuo, ma non differenziabile secondo l'usuale accezione della matematica classica. È però possibile definire il differenziale di tale processo in senso stocastico.

4.1.2 Integrale stocastico di It \bar{o}

Dato il solito processo di Wiener W_t e data $f(s)$ una funzione continua e sufficientemente regolare chiamiamo integrale stocastico la seguente scrittura:

$$X_t = \int_{t_0}^t f(s) dW_s \quad (4.3)$$

Tale oggetto, come già ribadito, non è trattabile secondo le definizioni convenzionali di integrale, per esempio secondo le definizioni di integrale di Riemann o di Riemann-Stieltjes. Infatti, secondo la definizione di Riemann-Stieltjes, e sostituendo il processo di Wiener $W(\cdot)$ con una funzione deterministica $F(\cdot)$ avremmo l'integrale I così definito:

$$I = \lim \sum_{k=1}^n f(a_k) [F(t_k) - F(t_{k-1})] \quad (4.4)$$

con a_k un punto qualsiasi contenuto nell'intervallo di estremi $[t_{k-1}, t_k]$. Normalmente, raffinando la partizione, accadrebbe che il valore dell'integrale I convergerebbe ad un valore ben preciso, indipendentemente dal dove si scelga a_k nell'intervallo $[t_{k-1}, t_k]$. Ciò accade per funzioni normali, per esempio per funzioni deterministiche sufficientemente regolari. It \bar{o} ha però dimostrato che se la funzione $F(\cdot)$ è un processo aleatorio allora raffinando la partizione di I otteniamo valori differenti a seconda del valore di a effettivamente scelto. Quindi la definizione di integrale di Riemann-Stieltjes (che altro non è che una generalizzazione dell'integrale di Riemann) è priva di senso in ambito stocastico.

It \bar{o} ha fornito una propria definizione di integrale stocastico scegliendo di valutare la funzione integranda $f(s)$ nell'estremo inferiore di ogni intervallo $[t_{k-1}, t_k]$. Quindi la definizione di integrale stocastico secondo It \bar{o} è la seguente:

$$I_{It\bar{o}} = \lim \sum_{k=1}^n f(t_{k-1}) [F(t_k) - F(t_{k-1})] \quad (4.5)$$

tale scelta, anche se apparentemente può sembrare arbitraria, è supportata da considerazioni analitiche. In particolare l'integrale di It \bar{o} converge in media quadratica per ogni funzione $f(\cdot)$ sufficientemente regolare.

Questa nuova definizione di integrale soddisfa alcune importanti proprietà:

1. **linearità** $\int_{t_0}^t (af_1(t) + bf_2(t)) dW_t = a \int_{t_0}^t f_1(t) dW_t + b \int_{t_0}^t f_2(t) dW_t$

Tale proprietà è conseguenza della definizione di Integrale di It \bar{o} (Formula 4.5) e dalla proprietà di linearità della sommatoria e del limite.

2. **valore atteso** $E \left[\int_{t_0}^t f(t) dW_t \right] = 0$ Anche in questo caso questa proprietà è diretta conseguenza della proprietà 3 del processo di Wiener e della linearità della sommatoria e valore atteso.

3. **covarianza** $E \left[\left(\int_{t_0}^t f_1(t) dW_t \right) \left(\int_{t_0}^t f_2(t) dW_t \right) \right] = \int_{t_0}^t f_1(s) f_2(s) ds$

Tale dimostrazione non è banale, se ne può trovare una traccia in [9].

4. **varianza** $E \left[\left(\int_{t_0}^t f(t) dW_t \right)^2 \right] = \int_{t_0}^t f(s)^2 ds$

Dimostrazione immediata dalla 3.

Differenziale stocastico

A questo punto, data la definizione di Integrale Stocastico è possibile ridefinire il concetto di Differenziale Stocastico. Se X_t è un processo aleatorio ottenuto dall'integrale stocastico mostrato nella Formula 4.3 allora possiamo scrivere il suo differenziale come:

$$\begin{cases} dX_t = f(t) dW_t \\ X_{t_0} = X_0 \end{cases} \quad (4.6)$$

Nella cui espressione la condizione iniziale X_{t_0} deve essere specificata separatamente. Ricordiamo che tale espressione ha senso solo in ambito stocastico. È altrimenti priva di significato, in accordo con la non differenziabilità del processo di Wiener.

4.1.3 Equazioni differenziali stocastiche (SDE)

Si dice Equazione Differenziale Stocastica una scrittura del tipo:

$$dX_t = F(t, X_t) dt + G(t, X_t) dW_t, \quad X_{t_0} = c \quad (4.7)$$

Con $F(t, X_t) : \mathbb{R} \times \mathbb{R}^k \rightarrow \mathbb{R}^d$, $G(t, X_t) : \mathbb{R} \times \mathbb{R}^k \rightarrow \mathbb{R}^{d \times m}$. Tale equazione è una equazione differenziale che presenta un o più termini aggiuntivi di natura aleatoria (processo stocastico). È facile individuare nella Formula 3.19 una equazione differenziale stocastica.

4.2 Soluzione Numerica

Sfortunatamente molte volte le SDE non possono essere risolte per via analitica, la loro soluzione è perciò ottenibile esclusivamente per via numerica. Siccome si tratta di approssimazioni, la bontà di tali risultati è strettamente dipendente dal metodo utilizzato dal software di simulazione. Uno dei metodi più semplici è il cosiddetto **Metodo di Eulero in Avanti (o Esplicito)**. Esso consiste nel discretizzare ad intervalli regolari di ampiezza h , il dominio del tempo t , in modo da avere un vettore temporale $t = [t_0, t_1, t_2, \dots, t_n]$ equispaziato. Vale quindi la relazione seguente: $t_{n^*} = t_0 + hn^*$, con $n \in \mathbb{N}$. A questo punto si approssima la soluzione della SDE con l'equazione della tangente della stessa, valutata ad intervalli di passo h . Il risultato così ottenuto è una una linea spezzata ed è facile rendersi conto che più si infittiscono gli intervallini h , minore sarà l'errore tra la funzione soluzione della SDE e la sua approssimazione.

Mentre dal punto di vista matematico non si presentano difficoltà nel pensare ad h come un infinitesimo, nella realtà, tale astrazione ha dei pesanti vincoli per via della rappresentazione

finita dei numeri all'interno degli elaboratori. I moderni calcolatori, avendo in genere a disposizione 64bit Floating Point per ogni numero, non sono in grado di rappresentare numeri arbitrariamente grandi o arbitrariamente piccoli senza incorrere in approssimazioni talvolta inaccettabili.

Data quindi una generica SDE nella forma della 4.7 possiamo approssimarne la soluzione X_t , per ogni intervallo $[t_{k-1}; t_k]$, con la sua funzione derivata.

$$\begin{aligned}
 X_{t_1} - X_{t_0} &= F(t_0, X_{t_0})(t_1 - t_0) + G(t_0, X_{t_0})\mathcal{N}(0, 1)\sqrt{t_1 - t_0} \\
 X_{t_2} - X_{t_1} &= F(t_1, X_{t_1})(t_2 - t_1) + G(t_1, X_{t_1})\mathcal{N}(0, 1)\sqrt{t_2 - t_1} \\
 &\dots \quad \dots \quad \dots \quad \dots \quad \dots \quad \dots \quad \dots \quad \dots \\
 X_{t_{hn}} - X_{t_{h(n-1)}} &= F(t_n, X_{t_n})(t_n - t_{n-1}) + G(t_n, X_{t_n})\mathcal{N}(0, 1)\sqrt{t_n - t_{n-1}}
 \end{aligned} \tag{4.8}$$

Da queste equazioni possiamo ricorsivamente ottenere la formula approssimata, indicata con \hat{X}_t , della soluzione della SDE:

$$\boxed{\hat{X}_{t_{hn}} = \hat{X}_{t_{h(n-1)}} + F(t_n, X_{t_n})h + G(t_n, X_{t_n})\mathcal{N}(0, 1)\sqrt{h}} \tag{4.9}$$

Qui abbiamo sfruttato il fatto che $t_{hn^*} - t_{h(n^*-1)} = h$ e che, ricordando le proprietà del processo di Wiener, otteniamo che il suo differenziale $dW_t \sim \mathcal{N}(0, h) = \mathcal{N}(0, 1)\sqrt{h}$ per la nota proprietà della varianza. Si evince dunque che, dato il valore assunto nell'istante iniziale $X_{t_0} = X_{iniziale}$, l'Equazione 4.9 permette di calcolare la funzione a tempo discreto \hat{X}_t che approssima X_t soluzione della 4.7.

4.2.1 Analisi dei dati

Una volta ottenute le curve temporali della tensione di uscita dell'inverter è importante dare una interpretazione ai risultati dal punto di vista quantitativo. Cioè, in pratica, bisogna capire in che situazioni il rumore implica che i comportamenti aleatori non siano più trascurabili. Il modo più semplice per sintetizzare i risultati consiste nel ricavare la PDF (Probability Density Function) della tensione di uscita allo steady state. Cioè, siamo interessati a capire come si comporta probabilisticamente la $V_o(t)$ dell'inverter una volta terminati i transitori di carica/scarica.

Per raggiungere questo obiettivo è necessario prima di tutto raccogliere una gran quantità di dati. Maggiore sarà il numero dei campioni, maggiore sarà l'accuratezza e la precisione ricavata. Nel nostro caso abbiamo raccolto campioni di 250 realizzazioni per una finestra temporale pari a $7 \cdot 10^{-11}$ [s] (ovviamente si trascurano i primi $3 \cdot 10^{-11}$ [s] di transitorio perché la tensione non è ancora stabile) con passo di quantizzazione pari a $2 \cdot 10^{-14}$ [s]. Disponiamo quindi di $250 \cdot 3500 = 875000$ campioni per ogni transizione (H->L oppure L->H). La tecnica per ottenere la PDF consiste nel creare un arrayX di $NSTEP$ intervalli di tensione (compresi tra 0V e V_{dd}) suddiviso ad intervalli regolari di ampiezza $STEP = \frac{V_{dd}-0}{NSTEP}$, ed un'altro arrayY di zeri di dimensione $NSTEP$. C'è quindi una relazione univoca tra gli indici di arrayX e gli indici di arrayY visto che gli array hanno la stessa dimensione. A questo punto, per ciascuno dei nostri 875000 campioni a disposizione, bisogna valutare a quale intervallo di tensione di

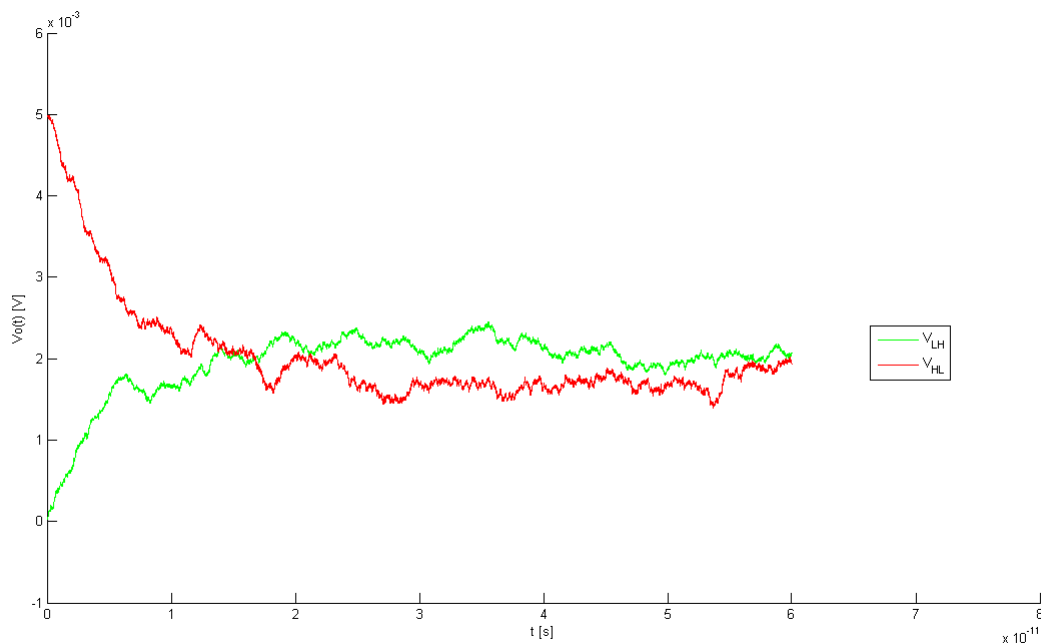


Figura 4.1: Andamento temporale (Euler, $n = 50000$ campioni) dell'inverter CMOS bilanciato ($\chi = 6.3$) in presenza di rumore con $v_{DD} = 5mV$. Nel grafico vengono mostrate solamente due realizzazioni.

arrayX esso appartiene per poi incrementare il rispettivo elemento dell'arrayY. In pratica si realizza una funzione definita da arrayX ad arrayY che conta, tramite arrayY, il numero di campioni che giacciono nell'intervallo di tensione i -esimo individuato dall'arrayX.

Infittendo a sufficienza gli NSTEP, per esempio utilizzando $NSTEP = 1000$, otteniamo una curva di densità che ben rappresenta la distribuzione dei campioni acquisiti. Si nota che tale curva ricorda molto una curva gaussiana. Ciò è ragionevole e lo si può intuire analiticamente studiando la relazione mostrata nella Formula 3.19. Per apprezzare graficamente queste considerazioni è utile sovrapporre, alla PDF ricavata, una curva gaussiana ottenuta estrapolando media e varianza dai dati acquisiti, cioè sfruttando il concetto statistico di massima verosimiglianza. Valgono quindi le seguenti relazioni:

$$\begin{cases} m_x = \frac{1}{n} \sum_{i=1}^n X_i \\ \sigma_x^2 = \frac{1}{n} \sum_{i=1}^n (X_i)^2 - (m_x)^2 \end{cases} \quad (4.10)$$

con n numero dei dati acquisiti (nel nostro caso $n = 875000$) e con X_i il valore di tensione del singolo dato.

I risultati ottenuti dalle simulazioni sono riassunti nelle Figure 4.3 e 4.4.

4.3 Probabilità di errore ed informazione associata

Si è visto dalla Figura 4.3 come la tensione di uscita dell'inverter non sia ideale ($V_{HIGH} \neq V_{dd}$ e $V_{LOW} \neq 0V$). Ciò accade perché a tensioni di alimentazione particolarmente basse la transcaratteristica dell'inverter si degrada scostandosi sempre più dall'idealità. Per definire quindi la tensione di soglia logica in uscita prendiamo il valore di tensione $V_{threshold}$ ottenuto facendo la media tra tutti i valori della tensione di uscita $V_o(t)$ allo steady-state. Tale tensione si collocherà graficamente a metà strada tra la tensione V_{HIGH} e la tensione V_{LOW} in uscita allo steady-state.

In particolari situazioni, per esempio quando la tensione di alimentazione è estremamente bassa o per funzionamento ad alte temperature, osserviamo che le curve di uscita H->L e L->H allo steady state tendono a sovrapporsi (Figura 4.4). In tale situazione è facile che la tensione di uscita si trovi ad attraversare più volte la tensione di soglia logica $V_{threshold}$ rendendo difficile, se non impossibile, l'interpretazione del valore logico da associare $V_o(t)$. In questi particolari casi di ambiguità l'inverter CMOS assume un comportamento di fatto aleatorio e non deterministico. È quindi utile ed opportuno cercare di quantificare la probabilità di errore, cioè, la probabilità che la tensione di uscita dell'inverter non rispetti la tabella di verità mostrata in Figura 2.2. In pratica bisogna determinare quanto è probabile avere $V_o(t) > V_{threshold}$ se in ingresso applico $V_i(t) = V_{dd}$, e quanto è probabile avere $V_o(t) < V_{threshold}$ se in ingresso applico $V_i(t) = 0V$. Per ricavare questa informazione ci serviamo delle PDF ottenute empiricamente e mostrate nelle Figure 4.3 e 4.4.

Definita quindi $P_{HIGH(LOW)}$ la PDF della tensione di uscita V_o allo steady-state applicando

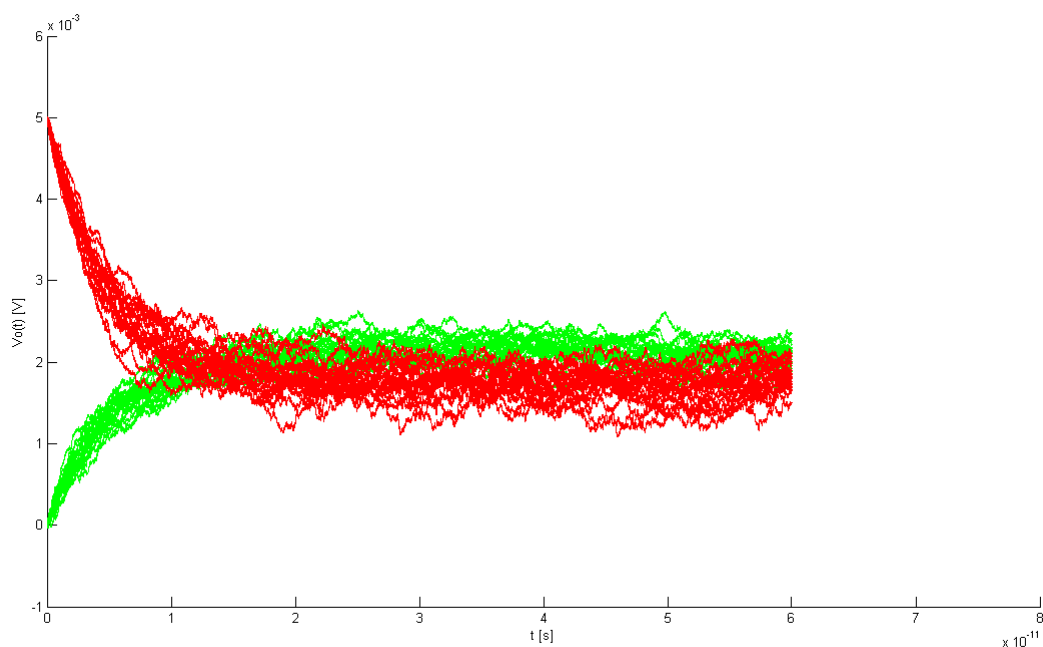


Figura 4.2: Andamento temporale (Euler, $n = 50000$ campioni) dell'inverter CMOS bilanciato ($\chi = 6.3$) in presenza di rumore con $v_{DD} = 5mV$. Il grafico rappresenta 20 realizzazioni per ciascuna transizione.

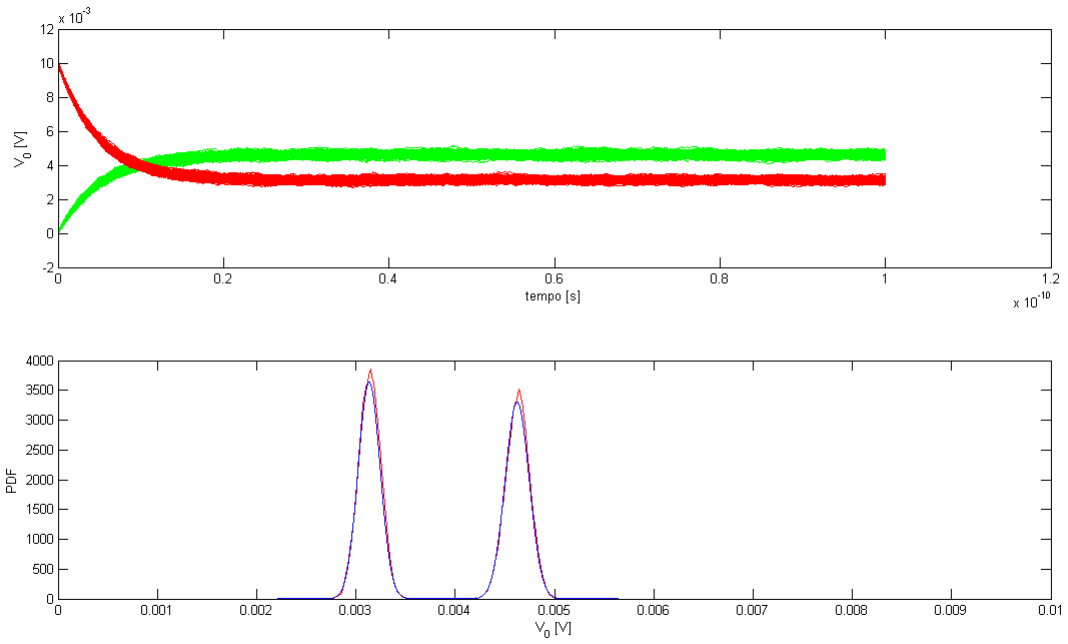


Figura 4.3: $V_{dd} = 10mV$. Andamento temporale della tensione di uscita(grafico sopra). PDF di $V_o(t)$ allo steady state, (grafico sotto).

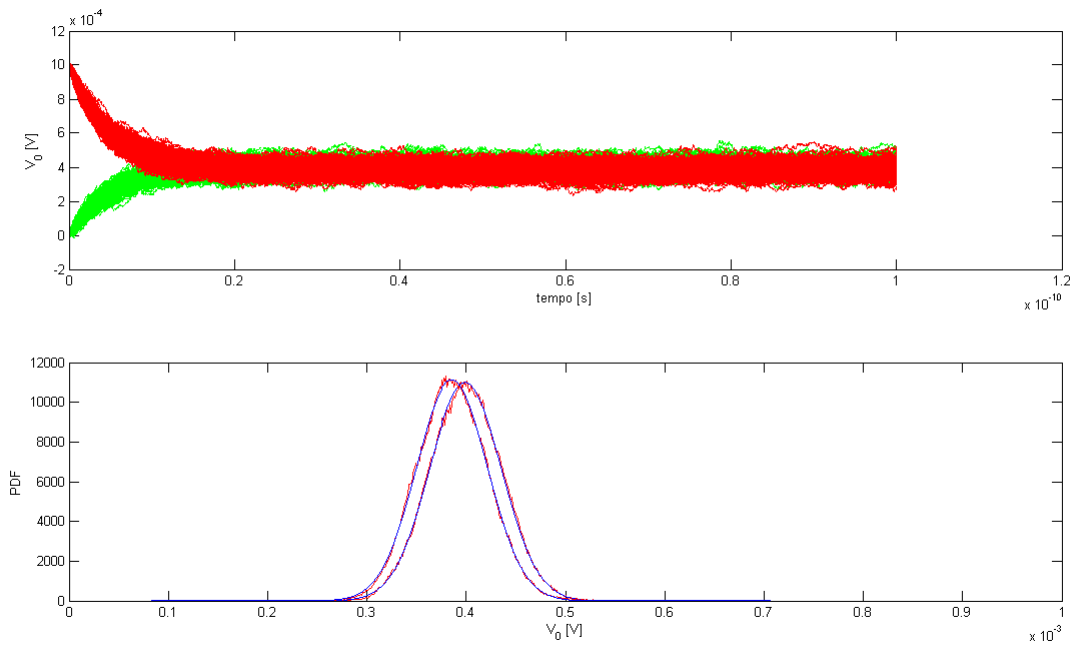


Figura 4.4: $V_{dd} = 1mV$. Andamento temporale della tensione di uscita(grafico sopra). PDF di $V_o(t)$ allo steady state, (grafico sotto).

in ingresso rispettivamente $V_i = V_{dd}(0V)$ otteniamo:

$$p_{errore} = \frac{1}{2} \left(\int_{-\infty}^{V_{threshold}} P_{LOW}(V) dV + \int_{V_{threshold}}^{+\infty} P_{HIGH}(V) dV \right) \quad (4.11)$$

Si noti che la probabilità di errore più sfavorevole è quando $p_{errore} = 0.5$. Ciò è giustificato dal fatto che, potendo l'uscita V_o assumere soltanto 2 valori, avere una $p_{errore} = 0.5$ implica che non si sa se il valore di tensione in uscita sia corretto oppure no. Quindi in questo caso abbiamo il massimo dell'incertezza anche se, statisticamente, “solo” il 50 per cento delle volte l'inverter non rispetterà la Tabella 2.2. Se avessimo avuto invece una $p_{errore} = 1$ (matematicamente possibile dalla Formula 4.11), avremmo avuto un inverter che, nonostante non rispetti mai la funzione 2.2, è comunque un circuito perfettamente deterministico visto che ne implementava la funzione negata \overline{NOT} .

Siccome è possibile modellare l'inverter CMOS in regione di sottosoglia come un canale binario rumoroso [6], è interessante definire l'informazione mutua tra ingresso ed uscita dell'inverter. Ciò quantificare quanto l'inverter “trasferisca” l'informazione logica dall'ingresso all'uscita.

$$I_{information} [bits] = 1 + p_{errore} \log_2(p_{errore}) + (1 - p_{errore}) \log_2(1 - p_{errore}) \quad (4.12)$$

I risultati sono riassunti in Figura 4.5 ottenuta simulando il comportamento dell'inverter per tensioni di alimentazioni comprese tra il decimo ed il centinaio di millivolt. È immediato osservare come al diminuire dell'alimentazione aumenti la probabilità di errore ed inevitabilmente l'informazione associata al nodo di uscita cali fino a diventare nulla. In pratica, per $V_{dd} < 10mV$, otteniamo un dispositivo la cui V_o non ha alcun legame deterministico con la tensione V_i applicata in ingresso.

Il codice MATLAB che realizza i grafici in Figura 4.3, 4.4 e 4.5 è riportato nell'appendice A.1 ed A.2.

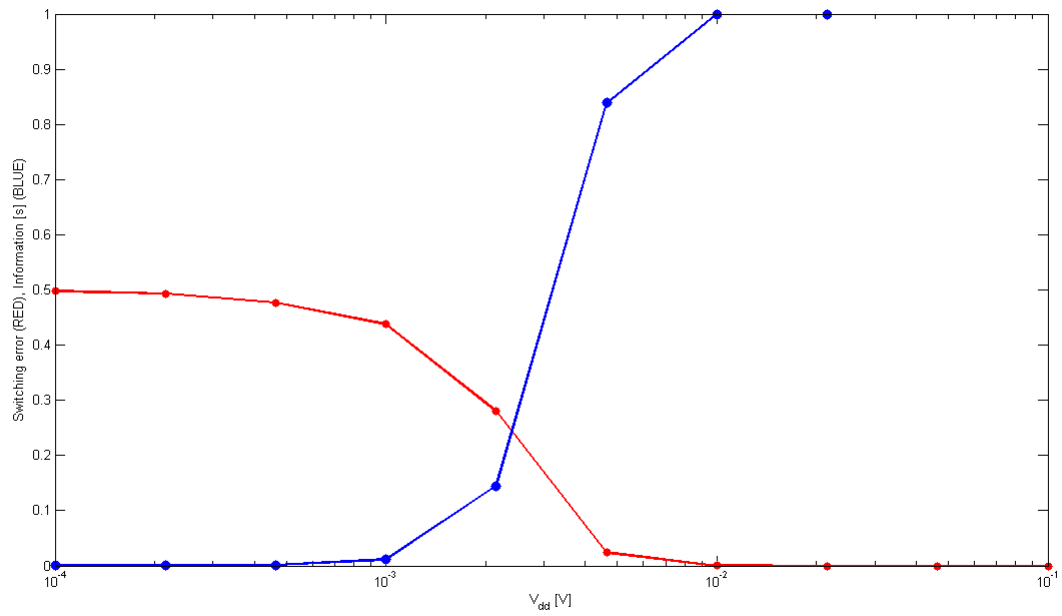


Figura 4.5: Grafico che mostra la probabilità di switch indesiderata (RED) e l'informazione associata all'uscita dell'inverter CMOS in sottosoglia (BLUE)

CAPITOLO

5

Conclusioni

Al fine di ridurre la potenza richiesta per il funzionamento dell'inverter CMOS, abbiamo sfruttato l'idea di abbassare la tensione di alimentazione, ed abbiamo mostrato che questa è di fatto una valida soluzione. Inoltre, siccome normalmente i circuiti digitali hanno tensioni di alimentazione attorno alle unità di volt, c'è un discreto margine di diminuzione di tale tensione e quindi di riduzione di potenza. Tale alimentazione può essere infatti ridotta fino ad un minimo di $V_{dd} \simeq 10mV$, soglia per cui il comportamento del dispositivo è ancora quello proprio di un invertitore. Bisogna porre particolare attenzione invece se ci si spinge a tensioni di alimentazioni inferiori alla decina di millivolt. In questa particolare configurazione, infatti, otteniamo un dispositivo il cui comportamento non è più deterministico. In particolare la tensione di uscita non è più funzione della tensione di ingresso, ma è pesantemente influenzata dal rumore termico/shot intrinsecamente presente nel dispositivo. Questa componente di rumore, ovviamente, è sempre presente anche per tensioni di alimentazioni superiori, ma diventa rilevante solo se V_{dd} è comparabile con l'ampiezza del rumore. In particolare si è anche mostrato come la temperatura influenzi negativamente il determinismo dell'inverter. Un aumento della temperatura comporta infatti un aumento dell'agitazione termica e quindi incrementa l'ampiezza del rumore.

Quindi è evidente come l'invertitore CMOS possa essere impiegato anche in applicazioni la cui alimentazione è molto bassa, senza scostarsi troppo dal modello ideale. In definitiva in applicazioni *low power consumption*, come per esempio orologi digitali o dispositivi elettromedicali, la riduzione della tensione V_{dd} può essere una soluzione applicabile ed efficace per ridurre il consumo di potenza.

Da notare però che il limite di $10mV$ è molto labile. Infatti per rispettare tale vincolo non solo è necessario avere a disposizione un generatore di tensione estremamente stabile, ma è anche indispensabile tenere sotto controllo la temperatura a cui il dispositivo opera.

Mentre in elettronica digitale il determinismo dei componenti è un requisito imprescindibile, si può sfruttare l'effetto indesiderato dell'aleatorietà della tensione di uscita per altri scopi. Per esempio si possono creare dei circuiti generatori di numeri casuali che sfruttano il rumore termico come sorgente generatrice. Tali dispositivi sono molto utili perché permettono di generare dei numeri casuali veri e propri, al contrario di ciò che accade nei computer che utilizzano algoritmi deterministici e quindi prevedibili.

APPENDICE

A

Codice MATLAB

A.1 PDF.m

```
1
3 %%%%%%%%%%% PARAMETRI %%%%%%%%%%%
5 clear
6 chi = 7.2;
7 ip = chi*5.95*10^-6;
8 in = 6.65*10^-5;
9 kp = 1/1.27;
10 kn = 1/1.39;
11 W = 120 * 10^-9;           % Larghezza canale MOS
12 L = 80 * 10^-9;           % Lunghezza canale MOS
13
14 C_L = 30*10^-15;           % Condensatore Load = 30fF
15 q = 1.60217653 * 10^-19;   % Modulo della carica dell'elettrone
16
17 %%%%%%%%%%% PARAMETRI %%%%%%%%%%%
18
19 v_dd = [10^-4 10^(-11/3) 10^(-10/3) 10^-3 10^(-8/3) 10^(-7/3)
20         10^-2 10^(-5/3) 10^(-4/3) 10^-1]';
21
22 p_errore = zeros(length(v_dd),1); % Inizializzo a zero il vettore della p. di errore
23 information = zeros(length(v_dd),1); % Inizializzo a zero il vettore dell'informazione mutua
24
25 for contatore=1:length(v_dd); % Eseguo il ciclo tante volte quanto e' la dimensione di v_dd
26     vdd = v_dd(contatore);
27
28     vt = 0.026;           % Thermal Noise. 27C = 300K => 26mV; 125C = 400K => 34,5mV
29     T = 10^-10;           % Finestra di visualizzazione
30     N = 5000;             % Numero di step da prendere nella finestra di visualizzazione
31     h=T/N;                % Quanto temporale
32     NSTEP = 1000;         % Numero step arrayX ed arrayY, "scatoline" della PDF
33     threshold = 7*10^-11; % Soglia temporale da cui iniziare a valutare i campioni
```

```

35 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%% Curava in VERDE L -> H %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
37
38 % Calcolo delle costanti utilizzate nella SDE
39 voo = 0; % Condizione iniziale condensatore
vi = voo; % Tensione applicata all'ingresso
41
42 a = (2*W)/(L*C_L) * exp(0.5*(log(ip*in) + kp*(vdd)/(vt) + ((kn-kp)/vt)*(vi))) *
43 sinh(0.5*(log(ip/in) + kp*(vdd)/(vt) - (kp+kn)*(vi)/(vt)));
45
46 aa = (2*W)/(L*C_L) * exp(0.5*(log(ip*in) + kp*(vdd)/(vt) + ((kn-kp)/vt)*(vi))) *
47 cosh(0.5*(log(ip/in) + kp*(vdd)/(vt) - (kp+kn)*(vi)/(vt)));
48
49 b = (2*W)/(L*C_L) * exp(0.5*(log(ip*in) + (kp-1)*(vdd)/(vt) + (kn-kp)*(vi)/(vt)));
c = (vt)/(2) * log((in)/(ip)) + (1-kp)*vdd/(2) + (kp+kn)*vi/(2);
51 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
53 step = vdd / NSTEP;
array = [0 : step :vdd]';
55 densita1 = zeros(NSTEP+1, 1); % Inizializzo il vettore densita' V_o = HIGH
57
58 media = 0; % Inizializzo la media
59 media2 = 0; % Inizializzo il secondo momento
campioni = 0;
61
62 J = 250; % Numero di simulazioni da effettuare per ogni curva (L->H);
for k=1:J
65
66 t=zeros(N,1); % Inizializzo vettore dei tempi
67 z=zeros(N,1); % Inizializzo vettore della tensione di uscita
69
70 t(1) = 0; % Fisso l'origine dei tempi
z(1) = voo; % Il primo valore della tensione di uscita e' pari alla condiz. iniziale
71
72 for i=1:N % Approssimo la soluz. della SDE con N tangenti
73
74 t(i+1) = t(i) + h; % Asse dei tempi. -> t(i)++;
75
76 % Soluzione della SDE
77 z(i+1) = z(i) + (a+b*sinh((c-z(i))/vt))*h +
sqrt(q*(aa-b*cosh((c-z(i))/vt))/C_L)*randn*sqrt(h);
79
80 % Per i conti statistici utilizzo solo i valori di tensione V_o(t) con t>threshold = 7*10^-11.
% Devo infatti essere sicuro che i transistori si siano esauriti completamente.
83
84 if (t(i) > threshold)
85
86 campioni = campioni + 1;
87 media = media + z(i);
88 media2 = media2 + (z(i))^2;
89 % richiama la funzione funzione_PDF (APPENDICE A.2)
densita1 = funzione_PDF(z(i), array, densita1);
91 end
93
94 end;
95
96 hold on;
97 figure(contatore); % Crea un oggetto figura per ogni tensione dell'array v_dd
99 % Stampa nel grafico superiore la curva della tensione di uscita V_o
101 subplot(2,1,1),plot(t,z, 'g');

```

```

103 end;
105 media = media / campioni; % Estrapolo la media
107 var = (media2 / campioni) - (media ^ 2) % Estrapolo la varianza
109
110 max = densital(round(media / step));
111
112 % Calcolo la densita' normalizzata. Di modo che l'area sia unitaria
113 densital_n = densital / (max*sqrt(var*2*pi));
115 subplot(2,1,2), plot(array, densital_n, 'r');
116 hold on
117
118 % Equazione di una normale con media e var. estrapolate dai dati
119 gaussianal = (1/sqrt(var*2*pi))*exp((array-media).^2/(-2*var));
121 subplot(2,1,2), plot(array, gaussianal, 'b');
123
124 soglia = media;
125
126 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%% Curava in ROSSA H -> L %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
127
128 % Calcolo delle costanti utilizzate nella SDE
129 voo = vdd; % Condizione iniziale condensatore
131 vi = voo; % Tensione applicata all'ingresso
132
133 a = (2*W)/(L*C_L) * exp(0.5*(log(ip*in) + kp*(vdd)/(vt) + ((kn-kp)/vt)*(vi))) *
134 sinh(0.5*(log(ip/in) + kp*(vdd)/(vt) - (kp+kn)*(vi)/(vt)));
135
136 aa = (2*W)/(L*C_L) * exp(0.5*(log(ip*in) + kp*(vdd)/(vt) + ((kn-kp)/vt)*(vi))) *
137 cosh(0.5*(log(ip/in) + kp*(vdd)/(vt) - (kp+kn)*(vi)/(vt)));
138
139 b = (2*W)/(L*C_L) * exp(0.5*(log(ip*in) + (kp-1)*(vdd)/(vt) + (kn-kp)*(vi)/(vt)));
140 c = (vt)/(2) * log((in)/(ip)) + (1-kp)*vdd/(2) + (kp+kn)*vi/(2);
141
142 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
143 densita2 = zeros(NSTEP+1, 1); % Inizializzo il vettore densita' V_o = LOW
144
145 media = 0; % Inizializzo la media
147 media2 = 0; % Inizializzo il secondo momento
148 campioni = 0;
149
150 for k=1:J
151
152 t=zeros(N,1); % Inizializzo vettore dei tempi
153 z=zeros(N,1); % Inizializzo vettore della tensione di uscita
154
155 t(1) = 0; % Fisso l'origine dei tempi
157 z(1) = voo; % Il primo valore della tensione di uscita e' pari alla condiz. iniziale
158
159 for i=1:N % Approssimo la soluz. della SDE con N tangenti
160
161 t(i+1) = t(i) + h; % Asse dei tempi. -> t(i)++;
162
163 % Soluzione della SDE
164 z(i+1) = z(i) + (a+b*sinh((c-z(i))/vt))*h +
165 sqrt(q*(aa-b*cosh((c-z(i))/vt))/C_L)*randn*sqrt(h);
166
167 % Per i conti statistici utilizzo solo i valori di tensione V_o(t) con t>threshold = 7*10^-11.
169 % Devo infatti essere sicuro che i transistori si siano esauriti completamente.

```

```

171     if (t(i) > threshold)
172         campioni = campioni + 1;
173         media = media + z(i);
174         media2 = media2 + (z(i))^2;
175         % richiama la funzione funzione_PDF.m
176         densita2 = funzione_PDF(z(i), array, densita2);
177     end
178
179     end;
180
181     hold on;
182     subplot(2,1,1), plot(t, z, 'r');
183     xlabel('tempo [s]');
184     ylabel('V_0 [V]');
185     % Rinomina l'asse X
186     % Rinomina l'asse Y
187 end;
188
189
190 media = media / campioni;
191 var = (media2 / campioni) - (media ^ 2);
192 % Estrapolo la media
193 % Estrapolo la varianza
194
195 max = densita2(round(media / step));
196 % Calcolo la densita' normalizzata. Di modo che l'area sia unitaria
197
198 densita2_n = densita2 / (max*sqrt(var*2*pi));
199 subplot(2,1,2), plot(array, densita2_n, 'r');
200 hold on
201 % Equazione di una normale con media e var. estrapolate dai dati
202
203 gaussiana2 = (1/sqrt(var*2*pi))*exp((array-media).^2/(-2*var));
204 subplot(2,1,2), plot(array, gaussiana2, 'b');
205 xlabel('V_0 [V]');
206 ylabel('PDF');
207
208 soglia = (soglia + media)/2;
209 % Calcolo la tensione di soglia in base ai campioni in uscita
210
211 %%%%%%%%%%% P_ errore %%%%%%%%%%%
212
213 % Codice che implementa e traccia il grafico (al variare di Vdd) delle Formule 4.11 e 4.12
214 % Valutazione integrale PDF dato Vin = LOW
215 intLOW = 0;
216 V_M = soglia;
217 Vm = round((NSTEP * V_M / vdd));
218 for i=1:Vm
219     intLOW = intLOW + gaussiana1(i)* step;
220 end
221
222 % Valutazione integrale PDF dato Vin = HIGH
223 intHIGH = 0;
224 for i=Vm:NSTEP
225     intHIGH = intHIGH + gaussiana2(i)* step;
226 end
227
228 p_errore(contatore) = 0.5*(intLOW + intHIGH);
229 information(contatore) = 1 + p_errore(contatore)*log2(p_errore(contatore)) +
230     (1-p_errore(contatore))*log2(1-p_errore(contatore));
231
232 end % END del ciclo FOR iniziale (quello della selezione di Vdd)
233
234
235 % Stampa il grafico di Figura 4.5
236
237 figure(length(v_dd)+1);
238 semilogx(v_dd, p_errore, '+-r', v_dd, information, '*-b');

```

```
239 xlabel('V_{dd} [V]');  
ylabel('Switching error (RED), Information [s] (BLUE)');
```

A.2 funzionePDF.m

```
1 % Tale funzione serve per confrontare il valore di tensione di z con i valori  
3 % di tensione contenuti nel vettore array. Una volta individuato l'indice di array  
5 % il cui intervallo contiene z, si incrementa il corrispondente elemento nel vettore densita.  
7 % In questo modo, come spiegato nel par. 4.2.1, si realizza una funzione univoca  
9 % tra il vettore array (arrayX) ed il vettore densita (arrayY).  
11 % Il vettore array si occupa di stabilire in quale intervallo di tensione giace z.  
13 % Il vettore densita si occupa invece di contare le occorrenze di z in un dato intervallo  
15 % di tensione.  
17  
19 function [densita] = funzione_PDF(z, array, densita)  
21 for i = 1:length(array)  
23     if (array(i) >= z)  
25         densita(i) = densita(i) + 1;  
27         break;  
29     end  
31 end
```


BIBLIOGRAFIA

- [1] P.Xu, P.Abshire, "*Stochastic Behavior of a CMOS Inverter*" 14th IEEE International Conference on Electronics, Circuits and Systems, 11-14 Dec. 2007, Marrakech, pp. 94 - 97, 2007.
- [2] J.M.Rabaey, A. Chandrakasan, B. Nikolic, "*Circuiti Integrati Digitali*", Pearson Education Italia, Seconda edizione, 2005.
- [3] M.Alioto, "*Understanding DC Behavior of Subthreshold CMOS Logic Through Closed-Form Analysis*", IEEE Transactions on Circuits and Systems I°, vol.57, n.7, pp. 1597, 19/17/2010.
- [4] A.D.Polyanin, V.F.Zaitsev, "*Handbook of Exact Solutions for Ordinary Differential Equations*", Second Edition, Chapman and Hall/CRC Press, Boca Raton, 2003.
- [5] A.Papoulis, S.U.Pillai, "*Probability, Random Variables and Stochastic Processes*", McGraw-Hill International, Fourth Edition, 2002.
- [6] N.Benvenuto, R.Corvaja, T.Erseghe, N.Laurenti, "*Communication Systems - Fundamentals and Design Methods*", Wiley, 2007.
- [7] R.Sarpeshkar, T.Delbruck, C.A.Mead, "*White Noise in MOS Transistors and Resistors*", IEEE Circuits and Devices Magazine, vol.9, n.6, pp. 23-29, 1993.
- [8] R.Matterazzo, "*Il problema della stima nelle Equazioni Differenziali Stocastiche: un approccio operativo*", Tesi di Laurea in Economia Aziendale, A.A. 1998/1999, Università Ca' Foscari di Venezia, Relatore: Pietro Mantovan.
- [9] L.Arnold, "*Stochastic Differential Equations*", Wiley-Interscience, 1974.
- [10] R.Sarpeshkar, "*Ultra Low Power Bioelectronics: Fundamentals, Biomedical Applications, and Bio-inspired Systems*", Cambridge University Press, 2010.