

POLITECNICO DI MILANO

Scuola di Ingegneria Industriale e dell'Informazione

Corso di Laurea Magistrale in Ingegneria Elettronica



**MODELLISTICA DI DISPOSITIVI A SWITCHING RESISTIVO PER
APPLICAZIONI IN RETI NEUROMORFICHE**

Relatore: Prof. Daniele IELMINI

Correlatore: Ing. Stefano AMBROGIO

Tesi di Laurea Magistrale di:

Valerio Milo

Matricola: 786272

Anno Accademico 2014-2015

Indice

Indice	i
Elenco delle figure	iii
Introduzione	xiii
1 Introduzione alle memorie emergenti	1
1.1 Memorie non volatili	1
1.2 Gerarchia delle memorie	2
1.3 Memorie non volatili a switching resistivo	5
1.3.1 Memorie a cambiamento di fase	5
1.3.2 Memorie Spin-Transfer-Torque (STT-MRAM)	7
1.3.3 Memorie RRAM	10
1.4 Introduzione alle reti neuromorfiche	12
2 Modello della cella RRAM	15
2.1 Memorie resistive a switching bipolare	15
2.2 Modello filamentare	16
2.3 Modello di switching a gap variabile	23
2.3.1 Reset	24
2.3.2 Set	28
3 Modellistica dell'overshoot capacitivo	31
3.1 Overshoot capacitivo	31
4 Modello Simulink della struttura 1T1R	39
4.1 Implementazione del circuito 1T1R	39
4.2 Analisi del circuito simulato	41

4.3	Implementazione in Simulink della plasticità	50
5	Apprendimento di pattern mediante sinapsi RRAM a switching stocastico	55
5.1	Sinapsi 1T1R	55
5.2	Apprendimento di pattern con STDP deterministica	60
5.3	Caratterizzazione dello switching stocastico	61
5.4	Apprendimento di pattern con sinapsi stocastiche	65
6	Applicazioni di reti neuromorfiche a RRAM	71
6.1	Apprendimento di più pattern mediante sinapsi inibitorie	71
6.2	Apprendimento di pattern a colori	73
6.3	STDP negativa	74
7	Modellistica di un generatore di numeri casuali a RRAM	77
7.1	RNG a 2 RRAM in parallelo	77
7.2	RNG a 2 RRAM in serie	79
	Conclusioni	85
	Bibliografia	87
	Ringraziamenti	97

Elenco delle figure

1.1	Gerarchia delle memorie in un calcolatore. Sono descritte le memorie fisiche, dai registri vicini alla CPU fino ai tre livelli di cache, la memoria principale, la memoria secondaria e quella terziaria. E' mostrata anche la velocità di accesso ad ogni singola memoria insieme al numero di cicli di CPU necessari per un accesso in memoria e alla tipica dimensione, in byte, della memoria. Riprodotta da [4].	3
1.2	Andamento qualitativo del costo in funzione delle prestazioni per differenti memorie. La categoria di memorie detta Storage Class Memory (SCM) si pone a metà tra le memorie costose e ad alte prestazioni, come le SRAM e le DRAM, e quelle a basse prestazioni, ma aventi elevata densità come le Flash e gli hard disk, (a). Evoluzione negli anni delle differenti classi di memorie, con le future SCM che renderanno meno definito il confine tra l'area delle memorie di storage e quella delle logiche, (b) [6].	4
1.3	Classificazione delle memorie resistive. Riprodotta da [2].	6
1.4	Tabella delle prestazioni dei differenti tipi di memoria. I nomi eDRAM e eFLASH indicano rispettivamente le DRAM embedded e le FLASH embedded). Gli svantaggi della tecnologia sono segnati in grassetto. Riprodotta da [3].	7
1.5	Struttura di una PCM. Tra i due elettrodi detti rispettivamente top e bottom electrode sono interposti un GST cristallino e un isolante in cui un heater consente il passaggio di una corrente e il riscaldamento locale. Lo stato a bassa resistenza corrisponde alla regione completamente cristallina, (a), mentre lo stato di reset mostra la caratteristica regione amorfa a forma di cupola, (b).	8
1.6	Principio di funzionamento delle memorie STT. Il primo disegno spiega la commutazione dallo stato parallelo allo stato anti-parallelo dovuta agli elettroni aventi spin opposti rispetto alla magnetizzazione del layer fisso. Al contrario, il secondo disegno rappresenta lo switching dallo stato anti-parallelo allo stato parallelo. Tratto da [17].	9

1.7	Pittura fisica del meccanismo di switching delle RRAM. La transizione di set porta alla formazione di un filamento conduttivo, mentre la transizione di reset attiva la rottura del filamento. Questi due fenomeni si verificano per tensioni applicate di polarità opposta.	11
2.1	Struttura metallo-isolante-metallo [8].	16
2.2	Illustrazione schematica della transizione di set. I difetti sono addensati al TE, (a), come conseguenza di un precedente reset della cella. In (b) si osserva la nucleazione di un sottile filamento conduttivo, il cui spessore nel tempo tende a crescere, (c), finchè la cella non raggiunge il definitivo stato di set, (d) [29].	17
2.3	Rappresentazione semplificata delle buche di potenziale nell'ossido. In assenza di una tensione applicata, gli ioni vedono delle buche di altezza pari a E_{A0} , (1). Applicando invece una tensione, l'altezza della barriera di potenziale si riduce di un fattore αqV , favorendo così la migrazione, (2).	18
2.4	Struttura 1T1R. La tensione di alimentazione V_A si ripartisce tra la cella e il transistor.	20
2.5	Caratteristica I-V simulata con il modello filamentare. $I_C \simeq 50 \mu A$, $V_{set} = V_{reset} \simeq 0.4 V$.	20
2.6	Andamento calcolato di V_A (in rosso) e della tensione V ai capi della cella (in blu) in funzione del tempo. Dopo un'iniziale crescita a rampa, V si stabilizza alla tensione V_C , che è uguale a 0.3 V.	21
2.7	Andamento misurato e calcolato di R_{set} , (a), e I_{reset} , (b), in funzione di I_C . Sono riportati e confrontati i dati relativi a molti tipi di dispositivi e materiali come le RRAM unipolari in NiO e le RRAM bipolari in HfO_x , TiO_x e HfO_x/ZrO_x . Si nota che le simulazioni sono in pieno accordo con i dati sperimentali e consistenti con le formule delle Eqs.(2.11) e (2.15) [29].	22
2.8	Caratteristiche I-V del reset misurate e simulate per due differenti condizioni iniziali di set. La curva blu è stata ottenuta da una resistenza iniziale del dispositivo pari a $R = 1k\Omega$, mentre quella rossa si riferisce ad uno stato resistivo iniziale di 0.4 k Ω . Inoltre, lungo una delle curve calcolate, sono segnati dei punti, chiamati A, B, C, D, che corrispondono a particolari stati in cui si trova la cella durante la transizione di reset [37].	23
2.9	Rappresentazione dei profili della concentrazione dei difetti ionizzati n_D , (a), di temperatura, (b), e della tensione, (c), calcolati nei punti A, B, C, D delle curve mostrate in Fig. 2.8 [37].	25

2.10	Mappe 2D della concentrazione dei difetti ionizzati n_D , (a), di temperatura, (b), e della tensione, (c), calcolate nei punti A,B,C,D di Fig 2.8 [37].	26
2.11	Illustrazione schematica della transizione di reset secondo il modello a gap variabile. Il reset della cella resistiva è dovuto all'apertura di un gap Δ a partire dal centro del filamento continuo laddove il profilo di temperatura presenta il proprio massimo. Da (a) dove viene rappresentato il CF continuo (stato di set completo), mediante la migrazione ionica verso il TE, inizia ad aprirsi un gap, (b), la cui lunghezza cresce gradualmente, (c), finchè non viene raggiunto lo stato finale di reset, (d) [77].	27
2.12	Illustrazione schematica del set secondo il modello a gap variabile. In (a) è rappresentato un filamento metallico continuo risultato di un precedente set. Applicando una tensione negativa al TE si esegue il reset della cella che comporta l'apertura di un gap nel CF, (b). Dopo l'operazione di reset, cambiando la polarità della tensione fornita al TE, viene eseguito un set del dispositivo che porta alla nucleazione, (c), e alla successiva crescita di diametro, (d), di un CF all'interno dell'ossido [77].	28
2.13	Caratteristica I-V sperimentale, in rosso, e relativa simulazione, in nero, generata a partire dal modello di switching a gap variabile.	29
2.14	Andamenti calcolati di V , ϕ , R , I e T_{max} in funzione del tempo. Inizialmente, all'aumentare della tensione applicata, (a), non accade nulla nel CF, (b), in quanto la temperatura raggiunta da quest'ultimo, (c), è ancora troppo bassa. La resistenza, (d), diminuisce in modo non lineare a causa del campo elettrico presente nel gap. Raggiunta la soglia $T_{crit} = 800$ K, il diametro del CF ϕ subisce una crescita molto brusca a causa dell'attivazione di una reazione positiva. Quando I eguaglia I_C si verifica una riduzione della tensione ai capi della cella con seguente stabilizzazione del diametro del CF.	30
3.1	Struttura 1T1R con una capacità parassita che insiste sul nodo di drain del MOS.	32
3.2	Andamento nel tempo della tensione applicata al TE.	32
3.3	Rappresentazione dell'andamento della corrente della RRAM durante il processo di set per differenti valori di V_{set} . All'aumentare della tensione di set si osservano spike di corrente sempre più alti che descrivono il peso crescente della capacità nel funzionamento della struttura 1T1R. Si è assunto $I_C = 20\mu A$ e $C_{par} = 50$ fF.	33
3.4	Andamento dell'ampiezza dello spike di corrente legato alla capacità parassita al variare di V_{set} . Si osserva una crescita molto repentina.	34

3.5	Andamenti, ottenuti rispettivamente dal modello analitico e dalla formula, di R_{LRS} al variare di C_{par} , per $V_{set} = 1.8\text{ V}$, 2 V e 2.2 V . Si osserva che solo per bassi valori di C_{par} l'accordo tra modello e formula viene a mancare. L'andamento delle curve ricavate dal modello non si discosta molto dalla power law descritta dalla formula, la quale prevede una pendenza pari a $-\frac{2}{3}$	35
3.6	Andamento di R_{LRS} in funzione di V_{set} per C_{par} che varia tra 10 fF e 600 fF . All'aumentare della capacità parassita, la resistenza dello stato LRS tenderà ad assumere a valori progressivamente più bassi.	36
3.7	Confronto tra le misure sperimentali e gli andamenti simulati di R_{LRS} in funzione di V_{set} per tre differenti valori di I_C . Inoltre viene riportato il profilo di R_{LRS} ottenuto mediante la formula analitica, il quale, per $V_{set} < 1.5\text{ V}$, si discosta dai dati e dal modello poichè non si è più nel regime di funzionamento controllato dalla capacità parassita.	37
3.8	Confronto tra l'andamento di V_{set} ottenuto al variare di C_{par} , dal modello e quello ricavato dalla formula (3.7) per tre valori fissati di R_{LRS} . Si osserva un ottimo accordo tra i risultati del modello e quelli ottenuti dalla formula analitica.	38
4.1	Schema circuitale della struttura 1T1R progettato in ambiente Simulink.	40
4.2	Curve I_{DS} - V_{DS} misurate (in blu) con le corrispondenti simulazioni (in rosso) per $V_{bulk} = 0\text{ V}$, (a), e $V_{bulk} = -2\text{ V}$, (b).	41
4.3	Rappresentazione su scala lineare, (a), e semilogaritmica, (b), delle curve I_{DS} - V_{GS} misurate/calcolate per $V_{bulk} = 0\text{ V}$ e $V_{bulk} = -2\text{ V}$	41
4.4	Interfaccia grafica Simulink attraverso cui è possibile fissare i parametri della simulazione circuitale.	42
4.5	Andamento nel tempo della tensione V_{top} applicata al TE della struttura durante la simulazione. Il segnale consiste nella successione di due impulsi triangolari aventi ampiezza di picco rispettivamente pari a -1.7 V e 2.2 V . L'applicazione di questo segnale attiva una sequenza reset-set della cella resistiva.	43
4.6	Profilo temporale della tensione V_G applicata al gate del transistor MOS. Fino a metà simulazione, la tensione di gate è uguale a 5 V per minimizzare la resistenza del MOS. All'istante $t = 1\text{ ms}$ inizia la fase di set e la tensione di gate commuta a 2.94 V , valore che sarà mantenuto fino alla fine della simulazione.	44
4.7	Caratteristica I-V sperimentale della struttura 1T1R, in rosso, e relativa simulazione, in blu, generata mediante il modello Simulink. Si può constatare un buon accordo tra la curva simulata e i dati sperimentali.	45

4.8	Andamento nel tempo della corrente che attraversa la cella. Durante il reset, a causa dell'incremento di resistenza, la corrente I simulata diminuisce sino ad annullarsi per $V_{top} = 0$ V. In seguito all'attivazione della transizione di set, I aumenta molto rapidamente fino a raggiungere la corrente di compliance I_C , ovvero il valore per cui la crescita si arresta a causa della limitazione imposta dal transistor in serie.	46
4.9	Evoluzione temporale della tensione che cade ai capi della cella. Dato che durante il reset R_{MOS} è trascurabile rispetto a R , nella prima parte di simulazione V segue l'andamento di V_{top} . Con l'attivazione della successiva transizione di set, la tensione V subisce una riduzione che tenderà a rallentare la crescita del filamento conduttivo fino ad arrestarla, ponendo così fine al processo di set.	47
4.10	Simulazione dell'andamento nel tempo della resistenza del dispositivo.	48
4.11	Rappresentazione della crescita repentina del diametro del sub-filamento conduttivo ϕ_G che si forma nel gap durante il processo di set.	49
4.12	Rappresentazione compatta dello schema circuitale impiegato per simulare la plasticità.	50
4.13	Schema circuitale con cui è stato generato l'impulso applicato al gate della struttura 1T1R.	51
4.14	Interfacce grafiche delle impostazioni associate ai tre generatori impulsati utilizzati per ottenere la forma d'onda del segnale di gate.	51
4.15	Forma d'onda della tensione V_G	52
4.16	Rappresentazione dello schema a blocchi impiegato per implementare la forma d'onda dell'impulso generato da un post-neurone durante un evento di fire.	52
4.17	Rappresentazione dell'impulso di tensione applicato al top electrode della struttura. A questo segnale sarà applicato un ritardo variabile Δt in modo da simulare l'andamento della variazione relativa di conduttanza della sinapsi in funzione dello stato iniziale e del ritardo stesso.	53
4.18	Esempio di curva STDP simulata attraverso il modello Simulink. Per $\Delta t < 0$ si nota una diminuzione del peso sinaptico, mentre per $\Delta t \geq 0$ si ha un incremento della conduttanza della sinapsi.	54

5.1	Rappresentazione schematica della sinapsi 1T1R che connette il PRE-neurone al POST-neurone, (a), e dei classici impulsi V_G e V_{TE} applicati rispettivamente al gate e al TE, (b). Un impulso di tensione V_G induce il passaggio di una corrente che verrà poi integrata dal POST-neurone, con la possibilità di attivare la generazione di un impulso di fire. Nell'istante in cui si verifica un evento di fire, la tensione V_{TE} può indurre il potenziamento della sinapsi ($\Delta t \geq 0$) o la depressione della stessa ($\Delta t < 0$), eseguendo in questo modo il protocollo STDP.	56
5.2	Rappresentazione delle caratteristiche STDP, ricavate dai dati, (a), e dalle simulazioni, (b), che descrivono la variazione di conduttanza R_0/R in funzione del ritardo Δt definito in Fig. 5.1 (b). Le simulazioni sono state generate attraverso un modello Simulink. La variazione di conduttanza è stata misurata/calcolata all'aumentare della resistenza iniziale R_0	57
5.3	Rappresentazione delle curve STDP simulate come mappa 3D di R_0/R in funzione di Δt e R_0 . Si osserva che per alti (bassi) valori di resistenza iniziale R_0 si attiva il potenziamento (depressione) della sinapsi.	58
5.4	Schema circuitale di una rete neuromorfica a 2 layer adottata per dimostrare l'apprendimento di pattern, (a), pattern visivo d'ingresso, (b), rappresentazione AER (address-event representation) degli eventi di fire innescati dal pattern e dal rumore in ingresso, (d), e rappresentazione dell'andamento del potenziale interno del POST-neurone che genera impulsi di fire in seguito al raggiungimento della soglia interna, (e).	59
5.5	Simulazioni dell'andamento dei pesi sinaptici associati al pattern (in rosso) e al background (in blu) nel caso di sinapsi a 1 cella con $V_{TE-} = -1.6$ V e set deterministico. Le figure in alto mostrano le matrici 8x8 delle conduttanze delle sinapsi ad istanti fissati. Si osserva che l'apprendimento del pattern è molto inefficiente.	60
5.6	Sequenza degli impulsi di tensione applicati per eseguire gli esperimenti di random set mostrati in Fig. 5.7.	62
5.7	Distribuzioni cumulative della resistenza del dispositivo misurate e calcolate dopo l'operazione di random set eseguita all'aumentare di V_A , (a), e le tipiche caratteristiche I-V associate agli stati A, B e C, (b). A causa dello switching stocastico solo un certo numero di celle effettuano la transizione di set, (A), mentre le restanti o rimangono nello stato HRS, (B), o, in un numero minore di casi, effettuano una transizione di set parziale, (C).	63

5.8	Distribuzioni cumulative della resistenza della cella, misurate e calcolate dopo l'operazione di random-reset al crescere di V_{stop} , (a), andamento dei valori mediani corrispondenti, (b), e della deviazione standard della resistenza R del dispositivo, (c). Inoltre sono riportate le simulazioni dell'andamento dei valori mediani, (b), ricavato dal modello analitico, e della deviazione standard, (c), ricavato dal modello semi-empirico di variabilità.	64
5.9	Misure e simulazioni della probabilità di random-set in funzione di V_A e di random-reset al variare di V_{stop}	65
5.10	Andamento simulato dei pesi $1/R$ delle sinapsi associate al pattern (in rosso) e di quelle riferite al background (in ciano) per due differenti strutture di sinapsi, ovvero la sinapsi a 1 cella stocastica, (a), e quella a 4 celle stocastiche in parallelo, (b). In entrambi i casi $V_{TE+} = 1.3$ V. Le figure poste in alto descrivono i pesi sinaptici ad istanti stabiliti. Si osserva che l'efficienza di apprendimento del sistema migliora all'aumentare del numero di celle per sinapsi.	66
5.11	Mappe di colore calcolate che descrivono l'efficienza di apprendimento della rete in funzione di V_{TE-} e V_{TE+} nei casi di 1 cella per sinapsi, (a), 2 celle per sinapsi, (b), e 4 celle per sinapsi, (c). L'efficienza di apprendimento migliora all'aumentare del numero di celle per sinapsi a causa della riduzione del tasso di fire spuri e della maggiore incisività della media eseguita sul rumore.	67
5.12	Andamento dell'efficienza di apprendimento del sistema neuromorfico in funzione di V_{TE+} , (a), e del numero di celle per singola sinapsi, (b).	68
6.1	Rappresentazione della rete neuromorfica con sinapsi inibitorie impiegata nell'apprendimento multi-pattern, (a), e rappresentazione dei pesi sinaptici dopo 7000 epoche di apprendimento, (b). La presenza delle sinapsi inibitorie impedisce ai neuroni d'uscita di specializzarsi sullo stesso pattern.	72
6.2	Evoluzione nel tempo del processo di apprendimento di un pattern a colori. Attraverso la relazione indicata in (a), la tensione V_G modula la corrente di compliance I_C e quindi la massima conduttanza ottenibile. In (b) è rappresentato il pattern a colori di dimensione pari a 133x100 pixel e in (c) viene mostrato l'aggiornamento dei pesi sinaptici dopo rispettivamente 1, 4, 8 e 40 epoche di apprendimento. Ogni sinapsi conta tre celle e ciascuna di esse tiene conto dell'intensità delle tre componenti di colore fondamentali, ovvero rosso, verde e blu.	73

6.3	Andamento della frequenza media dell'attività dei neuroni d'ingresso (in blu) e d'uscita (in rosso) al variare del valore della soglia in corrente dei post-neuroni nei casi di 1 cella per sinapsi, (a), 2 celle per sinapsi, (b), e 4 celle per sinapsi, (c). All'aumentare della soglia, la corrente che deve essere integrata per attivare un evento di fire in uscita aumenta e quindi di conseguenza il numero di eventi in uscita tende a diminuire.	74
7.1	Schema circuitale di un RNG a 2 RRAM in parallelo, (a), e rappresentazione dell'evoluzione nel tempo di V_P , V_Q e V_{out} , (b).	78
7.2	Misure e simulazioni delle distribuzioni cumulative di resistenza delle celle P e Q dopo le operazioni di set e reset, (a), e delle distribuzioni cumulative delle tensioni V_{out} e V_{out2} , (b).	78
7.3	Misure sperimentali di V_{out} e V_{out2} su 10^3 cicli, (a), e andamento delle distribuzioni misurate e calcolate di V_{out} e V_{out2} , (b). Grazie alla variabilità dello stato HRS per alte V_{stop} , V_{out} mostra una distribuzione uniforme tra $-V_{max}$ e V_{max} , che è poi rigenerata attraverso il comparatore nella distribuzione bimodale 0/1 associata a V_{out2}	79
7.4	Schema circuitale di un RNG a 2 RRAM in serie, (a), e sequenza delle tensioni V_P , V_Q e V_{out} , (b).	80
7.5	Misure e simulazioni delle distribuzioni cumulative di resistenza di P e Q prima e dopo l'operazione di random reset. Come conseguenza dell'operazione di random reset, P e Q commutano in uno stato HRS solo nel 50% del numero totale di cicli eseguiti.	80
7.6	Rappresentazione schematica della curva I-V che descrive la transizione di reset della cella P, (a), e rappresentazione dell'evoluzione nel tempo delle cadute di tensione misurate ai capi delle due celle durante la transizione di reset della cella P, (b). . . .	81
7.7	Plot di correlazione delle resistenze di P e Q dopo l'operazione di set e reset. Dopo il set, le misure di resistenza di entrambe le celle mostrano valori tutti prossimi a 10 k Ω , mentre dopo il random reset o resetta Q, con P che si mantiene a bassa resistenza, o viceversa.	82
7.8	Distribuzioni misurate e calcolate di V_{out} , indicate rispettivamente in rosso e in nero, e di V_{out2} , rappresentate in blu e in ciano. Grazie al comparatore presente nel circuito, dopo un ciclo RNG si passa dalla distribuzione bimodale di V_{out} alla distribuzione 0/1 associata a V_{out2}	83

7.9 Andamento delle misure di V_{out} e V_{out2} su 10^3 cicli, (a), e rappresentazione delle distribuzioni ottenute dai dati sperimentali e dalle simulazioni di V_{out} e V_{out2} , (b). Grazie al random reset di P o Q, la tensione V_{out} mostra una distribuzione bimodale che viene poi rigenerata mediante un comparatore nella distribuzione 0/1 associata a V_{out2} 83

Introduzione

In questi ultimi anni si è assistito ad un enorme incremento della densità delle memorie realizzate con le tecnologie tradizionali. Tuttavia l'incessante domanda di dispositivi portatili sempre più economici e a basso consumo ha portato la ricerca a scontrarsi con i limiti della tecnologia attuale. La tecnologia Flash, che nel corso degli anni ha assunto il ruolo di leader nel campo delle memorie, è stata soggetta ad un impressionante sviluppo che l'ha condotta fino al nodo dei 16 nm e alla realizzazione di architetture 3D. A causa dei limiti di scaling imposti dalla fisica, la ricerca ha concentrato enormi sforzi sullo sviluppo di nuovi dispositivi in grado di unire ai vantaggi delle memorie Flash una maggiore scalabilità, un minore consumo e un'alta velocità di accesso, superando in questo modo le problematiche che affliggono le memorie a floating gate.

Allo stesso tempo è diventata sempre più indispensabile l'introduzione di una nuova classe di memorie, detta Storage Class Memory (SCM), in grado di vantare allo stesso tempo le caratteristiche proprie delle memorie a stato solido, come l'elevata velocità di accesso e l'alta robustezza, e quelle che caratterizzano gli hard disk come la grande capacità di immagazzinamento e il basso costo.

A questo proposito, le memorie a switching resistivo basate su ossidi metallici si sono imposte come principale candidato a ricoprire il ruolo di memorie non volatili di nuova generazione. In particolare, alcune proprietà come la semplicità della struttura a due terminali, il basso consumo di potenza e la grande velocità delle operazioni di lettura e programmazione fanno delle memorie a switching resistivo una valida alternativa alle memorie Flash e un ottimo candidato come SCM.

Il principale limite delle logiche CMOS attuali è rappresentato dall'elevata variabilità statistica che affligge i dispositivi odierni. Tuttavia esistono applicazioni, come la computazione neuromorfica, in grado di tollerare e sfruttare a proprio vantaggio le fluttuazioni statistiche dei dispositivi, emulando così il funzionamento delle reti neuronali degli esseri umani. In questo contesto è stato affrontato lo studio del comportamento delle sinapsi, le

quali sono state implementate mediante memorie resistive a switching stocastico.

In questa tesi di laurea magistrale verrà presentato un modello di sinapsi basato su una struttura 1 transistore/1 resistore (1T1R). Queste connessioni inter-neuronali, implementate mediante memorie resistive, consentono sia la trasmissione passiva degli impulsi che l'aggiornamento del peso secondo il protocollo di spike-timing-dependent-plasticity (STDP). Inoltre sarà simulata una rete neuronale a 2 layer, basata sulle caratteristiche STDP misurate sperimentalmente, mediante la quale si dimostrerà l'apprendimento e riconoscimento di pattern visivi assumendo il set stocastico delle memorie resistive.

Qui di seguito vengono riportati i principali argomenti trattati nella tesi.

Capitolo 1. Dopo una breve descrizione del panorama delle memorie non volatili tradizionali, con particolare accento posto sui limiti della tecnologia Flash e sui possibili sviluppi da essi derivanti, saranno introdotte nuove classi di memorie cosiddette emergenti, tra cui le memorie a switching resistivo basate su ossidi metallici (RRAM), di cui verranno analizzati il principio di funzionamento, la struttura e le prospettive di scaling.

Capitolo 2. In questo capitolo sono descritti due importanti modelli analitici per le memorie resistive in configurazione 1T1R basati rispettivamente sulla formazione/rottura di un filamento conduttivo e sull'apertura/chiusura di una barriera di ossido all'interno di un filamento preesistente.

Capitolo 3. In questo capitolo viene discusso l'impatto della capacità parassita della struttura 1T1R sulla transizione di set della cella resistiva.

Capitolo 4. In questo capitolo viene presentato un modello circuitale dell'architettura 1T1R implementato in Simulink. Dopo una breve descrizione dei blocchi circuitali utilizzati per realizzarlo, si procede con l'analisi dei principali parametri fisici che caratterizzano il circuito simulato. Infine questo modello viene sfruttato per implementare una sinapsi 1T1R e calcolare le caratteristiche STDP.

Capitolo 5. Questo capitolo è dedicato all'apprendimento e riconoscimento di pattern

visivi da parte di reti neuromorfiche che utilizzano RRAM come sinapsi artificiali. In particolare vengono confrontati due approcci che si distinguono per l'implementazione o meno del set stocastico. Sarà dimostrato che sfruttando la variabilità intrinseca della cella resistiva si riesce ad ottenere un'efficienza di apprendimento maggiore.

Capitolo 6. In questo capitolo si approfondisce lo studio dell'apprendimento di pattern visivi, allargando le applicazioni oggetto di studio all'apprendimento selettivo di più pattern mediante una rete neuromorfica dotata di sinapsi inibitorie tra i neuroni del layer d'uscita e all'apprendimento di pattern a colori. Infine è brevemente discussa l'implementazione 'negativa' del protocollo STDP.

Capitolo 7. In questo capitolo vengono mostrati due innovativi blocchi circuitali per la generazione fisica di numeri casuali che impiegano rispettivamente due memorie resistive connesse o in parallelo o in serie.

Capitolo 1

Introduzione alle memorie emergenti

L'elettronica moderna necessita di memorie veloci e ad alta densità per poter soddisfare varie applicazioni della vita quotidiana. Questa crescente domanda ha portato ad un enorme sviluppo delle memorie a stato solido e in particolare della tecnologia Flash. Ciononostante, i limiti intrinseci di questi dispositivi hanno spinto le industrie verso nuovi prototipi di memorie emergenti per superare questo problema. Questo capitolo, dopo una breve introduzione sul panorama attuale, è dedicato alla descrizione dei principali candidati al ruolo di memorie non volatili del futuro.

1.1 Memorie non volatili

Negli ultimi decenni, la crescente domanda di dispositivi elettronici ha condotto l'industria delle memorie a radicali evoluzioni e miglioramenti. La necessità di calcolatori ad alte prestazioni, dispositivi portatili e applicazioni compatte ha generato il crescente bisogno di memorie non volatili ad alta densità e sempre più veloci. La tecnologia attualmente dominante sul mercato, con la sua presenza nei cellulari, tablet e laptop, è la tecnologia Flash. L'enorme crescita delle memorie non volatili è stata una conseguenza della legge di Moore, il quale aveva previsto che negli anni si sarebbe verificato un incremento esponenziale del numero di dispositivi integrati nei chip al fine di ottenere memorie più economiche e a maggiore densità. Per questo motivo negli ultimi trent'anni il costo di hard disk e memorie Flash NAND si è estremamente ridotto. Se nel 1990 una Flash NAND costava 100.000 \$

per gigabyte [1], la stessa memoria attualmente costa circa 0.3 \$ (Ref. Amazon Italy). Ciò significa che, ogni 15 mesi, la densità d'integrazione è raddoppiata con un dimezzamento dei costi [1]. L'aumento della densità dei dispositivi è stato possibile grazie ai grandi sforzi dell'industria e della ricerca che hanno portato a notevoli progressi nel campo della litografia, essenziali per la miniaturizzazione (scaling) dei dispositivi, del processo tecnologico e dell'incremento delle dimensioni del wafer di Silicio da 150 mm (1987) agli attuali 300 mm. La densità di memoria è cresciuta anche grazie alla possibilità di memorizzare più bit per cella (cella multilivello o MLC) con la prospettiva nel prossimo futuro di adottare architetture tridimensionali. Le strutture di memoria 3D sono attualmente oggetto di intensa ricerca da parte delle principali industrie di memorie (Micron, Samsung) proprio perchè consentirebbero un brusco incremento della densità delle memorie. La tecnologia leader nel campo delle memorie non volatili è la tecnologia Flash. In questi dispositivi ogni bit è memorizzato in un transistor a doppio gate, ovvero un transistor che ha un control gate e un floating gate interposto tra due ossidi. Il principio di funzionamento delle Flash è il seguente: la presenza o assenza di cariche elettroniche nel floating gate determina una variazione della tensione di soglia V_T del transistor. Quando gli elettroni sono intrappolati nel floating gate, si verifica un aumento della soglia che rende più difficile l'accensione del transistor. L'operazione di lettura è eseguita misurando la corrente che attraversa il MOS. Se V_T è alta, non può scorrere corrente nel transistor e ciò corrisponde a un certo stato logico, mentre se la soglia è bassa, il MOS è attraversato da una corrente non nulla e si ottiene lo stato logico opposto. Le due architetture caratteristiche delle memorie Flash sono la NAND e la NOR [1, 2]. Attualmente la struttura NAND, realizzata mediante transistori in serie, è diventata l'architettura dominante a scapito di quella di tipo NOR, dove i MOSFET sono invece connessi in parallelo. Per programmare e cancellare una memoria NAND, gli elettroni devono attraversare l'ossido interposto tra il canale del MOSFET e il floating gate per tunneling Fowler-Nordheim [1]. Questa è un'operazione abbastanza lenta, circa 100 μ s, che richiede tensioni molto alte di circa 20 V e limita l'affidabilità della memoria a circa 10^5 cicli di programmazione-cancellazione [3]. Comunque questa tecnologia è ormai molto matura e ha una posizione dominante nel mercato delle memorie non volatili.

1.2 Gerarchia delle memorie

Un calcolatore è caratterizzato da molti tipi di memorie che si differenziano a seconda dello scopo e della vicinanza in termini di operazioni con la CPU [4]. In Fig. 1.1 è mostrata la

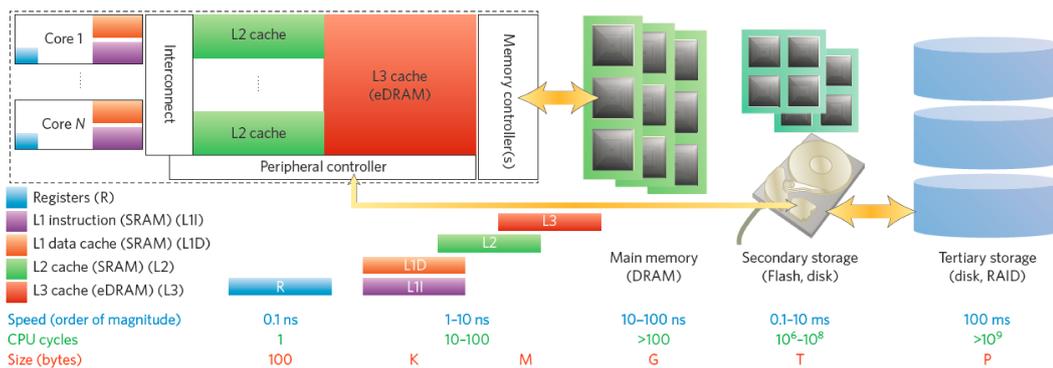


Figura 1.1: Gerarchia delle memorie in un computer. Sono descritte le memorie fisiche, dai registri vicini alla CPU fino ai tre livelli di cache, la memoria principale, la memoria secondaria e quella terziaria. E' mostrata anche la velocità di accesso ad ogni singola memoria insieme al numero di cicli di CPU necessari per un accesso in memoria e alla tipica dimensione, in byte, della memoria. Riprodotta da [4].

struttura base di una memoria presente all'interno di un computer. La CPU legge i dati e memorizza i risultati delle operazioni nei primi tre livelli L1, L2 e L3 della cache. Queste memorie presentano una velocità di lettura/scrittura molto elevata, dell'ordine dei nanosecondi, il che significa che per ogni accesso in memoria, la CPU esegue alcune decine di cicli [4]. I livelli L1 e L2 della cache sono realizzati mediante memorie SRAM (RAM statiche) le quali sono caratterizzate da una velocità di accesso estremamente elevata, intorno o sotto il 1 ns, ma hanno il grande svantaggio di essere molto costose e di occupare molta area. Una SRAM è costituita da 6 transistori di dimensione pari a $100 F^2$, con F che rappresenta la minima dimensione riproducibile attraverso la litografia. Il livello L3 è invece implementato mediante DRAM (RAM dinamiche) le quali presentano tempi di accesso leggermente superiori di circa 10-30 ns [3] rispetto alle SRAM, ma consentono di raggiungere densità d'integrazione maggiori con costi inferiori, dato che sono realizzate semplicemente con un transistor e un condensatore. Sia le SRAM che le DRAM sono memorie volatili, ovvero memorie in cui l'informazione memorizzata viene persa con lo spegnimento dell'alimentazione.

Dopo le DRAM, i dati vengono memorizzati in una sezione di memoria secondaria che è composta da memorie Flash o Hard Disk. Queste memorie sono caratterizzate da tempi di accesso lunghi, dell'ordine di $1\mu s$ - 1 ms per le Flash, o addirittura più lunghi per gli Hard Disk Drives (HDD) [1, 4]. I principali vantaggi che le contraddistinguono sono l'elevata densità, che comporta un minore costo per singolo bit, e la non volatilità, che consente la

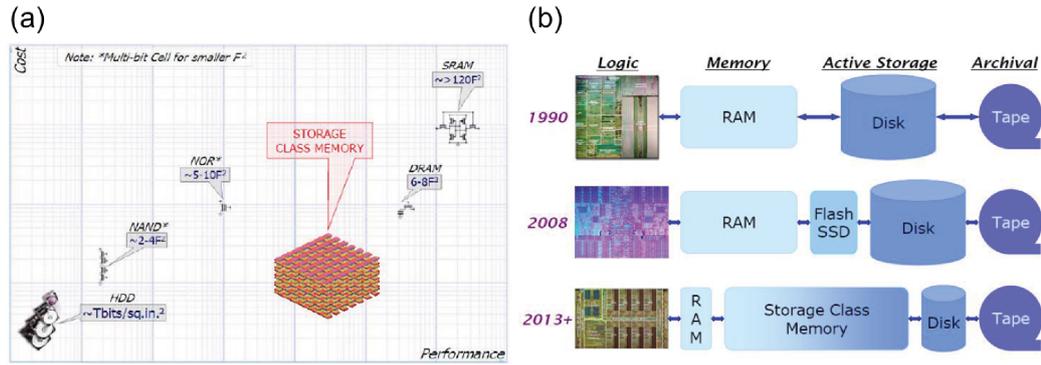


Figura 1.2: Andamento qualitativo del costo in funzione delle prestazioni per differenti memorie. La categoria di memorie detta Storage Class Memory (SCM) si pone a metà tra le memorie costose e ad alte prestazioni, come le SRAM e le DRAM, e quelle a basse prestazioni, ma aventi elevata densità come le Flash e gli hard disk, (a). Evoluzione negli anni delle differenti classi di memorie, con le future SCM che renderanno meno definito il confine tra l'area delle memorie di storage e quella delle logiche, (b) [6].

memorizzazione di enormi quantità (terabit) d'informazione. Tuttavia la differenza di ordini di grandezza rispetto ai tempi di accesso delle DRAM e delle SRAM genera problematiche di difficile soluzione in termini di efficienza di elaborazione [4, 5]. A conferma di ciò, il principale limite per lo sviluppo futuro di calcolatori ad elevate prestazioni è rappresentato non dall'efficienza dei processori, ma dall'accesso ai dati troppo lento e quindi si rende necessario un nuovo tipo di architettura di memoria [6, 5, 1].

Come rappresentato in Fig. 1.2 (a), l'attuale panorama industriale delle memorie è diviso tra memorie cache caratterizzate da alto costo, bassa densità e alte prestazioni (SRAM e DRAM) ed elementi di memoria aventi alte prestazioni, elevata densità e basso costo (Flash e dischi magnetici). Per colmare il gap tra queste due classi deve essere inventato un nuovo tipo di memoria [6, 1]. Per superare questo problema IBM ha proposto il concetto di Storage Class Memory (SCM) [5]. Questo tipo di memoria dovrebbe essere in grado di prendere il meglio da entrambe le categorie, ovvero alte velocità di lettura/scrittura, cioè tempi di accesso inferiori a 100 ns per poter sostituire almeno parzialmente le DRAM nelle cache, riducendone così il costo e di conseguenza aumentando la densità di memoria, e il basso costo per bit unito all'alta densità d'integrazione tipica delle Flash. In Fig. 1.2 (b) è mostrata l'evoluzione della gerarchia delle memorie durante gli ultimi decenni [6]. Da un'architettura iniziale dove il confine tra le memorie finalizzate alle logiche e quelle impiegate per l'immagazzinamento di dati era chiaro, la tendenza negli anni è stata quella

di ridurre questo gap con l'introduzione delle SCM per consentire entrambe le operazioni, rivoluzionando totalmente l'architettura di calcolo con il raggiungimento di brevi tempi di accesso a grandi quantità di dati [4]. A questo proposito, le memorie resistive sono molto interessanti come future SCM [7, 1].

1.3 Memorie non volatili a switching resistivo

La maggior parte delle memorie in commercio sono basate sull'intrappolamento e rilascio di carica. Ad esempio, in una SRAM troviamo due invertitori che possono assumere due stati stabili e ciò si ottiene muovendo elettroni, ovvero carica elettrica [4]. Una DRAM è costituita da un transistor, che funge da selettore, e da un condensatore che immagazzina la carica. Le memorie Flash sono basate sull'intrappolamento di carica nel floating gate che genera una variazione della tensione di soglia del transistor. Pertanto l'informazione viene memorizzata muovendo elettroni [4]. Le memorie emergenti sono invece basate su un meccanismo completamente differente che prevede la memorizzazione dei valori logici '0' e '1' sotto forma di differenti configurazioni di atomi che si traducono in stati resistivi distinti [8, 2, 7, 4, 9]. In linea di principio ogni fenomeno fisico che porta ad uno switch resistivo può essere utilizzato come memoria resistiva. Una classificazione generale è stata realizzata da Waser [2] ed è riportata in Fig. 1.3. Tra tutte le memorie proposte, quelle che hanno guadagnato grande interesse per le notevoli proprietà sono le memorie a cambiamento di fase (Phase change memory, PCM), le memorie magnetoresistive, le memorie elettrochimiche, dette anche RAM a ponte conduttivo (Conductive bridge RAM, CBRAM) e le memorie RAM a switching resistivo (Resistive RAM, RRAM). La tabella in Fig. 1.4 mostra indicatori molto utili per confrontare le differenti tipologie di memoria [3].

1.3.1 Memorie a cambiamento di fase

Le memorie a cambiamento di fase (PCM) sono basate sulle differenti proprietà elettriche che caratterizzano lo stato cristallino e lo stato amorfo in alcuni materiali e sulla capacità di switching reversibile tra questi due stati attraverso l'applicazione di impulsi di tensione [2]. Al giorno d'oggi, le PCM sono realizzate attraverso una lega composta da Germanio (Ge), Antimonio (Sb) e Tellurio (Te), tipicamente indicata con la sigla GST [2]. Per passare dallo stato cristallino, caratterizzato da una bassa resistività, allo stato amorfo, che mostra un'alta resistività, deve essere applicata una tensione elevata in modo da riscaldare il GST al di sopra della temperatura di fusione. Dopo il raggiungimento della fusione del GST, la tensione

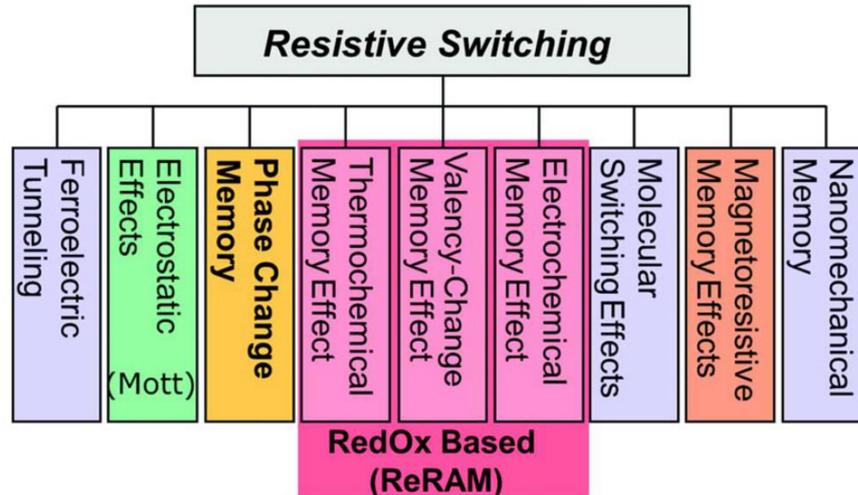


Figura 1.3: Classificazione delle memorie resistive. Riprodotto da [2].

applicata viene improvvisamente annullata e quindi il materiale si raffredda immediatamente nello stato amorfo. Per commutare in modo reversibile dallo stato cristallino allo stato amorfo viene applicato un impulso di tensione che produce una crescita della temperatura sopra la temperatura di cristallizzazione, ma al di sotto di quella di fusione. Ciò comporta la riorganizzazione degli atomi in una struttura cristallina.

Il dispositivo può essere impiegato in differenti architetture, tra cui la più comune è la cosiddetta architettura 'mushroom' riportata in Fig. 1.5 [2]. In questa struttura, la corrente passa attraverso l'heater, ovvero un canale che, in virtù del suo sottile diametro, permette di riscaldare in modo efficiente il GST. In Fig. 1.5 (a) è mostrata una PCM nello stato cristallino, mentre in Fig. 1.5 (b) è rappresentata una PCM nello stato amorfo, dove nel GST vicino all'heater appare la classica regione amorfa a forma di cupola.

Le PCM sono una delle tipologie di memorie emergenti più mature in assoluto e quindi sono considerate serie candidate per la realizzazione di SCM [1, 6, 2]. I vantaggi di questa tecnologia sono la semplicità della struttura a due terminali, l'ampia finestra resistiva ($R_{OFF}/R_{ON} \approx 10^3$), che potrebbe consentire l'implementazione di celle multilivello, l'elevata velocità di programmazione (la cristallizzazione si verifica in poche centinaia di nanosecondi mentre l'amorfezza avviene su una scala di decine di nanosecondi), basse tensioni (1-2 V) e un'alta resistenza alla ciclatura fino a 10^{12} cicli [10]. Inoltre è stata dimostrata una grande scalabilità attraverso l'utilizzo di elettrodi realizzati mediante nanotubi in carbonio [11].

Table 1 Comparison of key features of existing and emerging memories.									
	SRAM	eDRAM	DRAM	eFlash (NOR)	Flash (NAND)	FeRAM	PCM	STT-MRAM	RRAM
Endurance (cycles)	Unlimited	Unlimited	Unlimited	10^5	10^5	10^{14}	10^9	Unlimited	10^9
Read/write access time (ns)	<1	1-2	30	10/10 ³	100/10 ⁶	30	10/100	2-30	1-100
Density	Low (six transistors)	Medium	Medium	Medium	High (multiple bits per cell)	Low (limited scalability)	High (multiple bits per cell)	Medium	High (multiple bits per cell)
Write power	Medium	Medium	Medium	High	High	Medium	Medium	Medium	Medium
Standby power	High	Medium	Medium	Low	Low	Low	Low	Low	Low
Other	Volatile	Volatile. Refresh power and time needed	Volatile. Refresh power and time needed	High voltage required	High voltage required	Destructive readout	Operating T<125°C	Low read signal	Complex mechanism

Significant disadvantages are marked in bold. Estimates for emerging memories are based on expectations for functioning chips, not demonstrations of individual bits.

Figura 1.4: Tabella delle prestazioni dei differenti tipi di memoria. I nomi eDRAM e eFLASH indicano rispettivamente le DRAM embedded e le FLASH embedded). Gli svantaggi della tecnologia sono segnati in grassetto. Riprodotto da [3].

Tuttavia, alcune problematiche devono essere ancora affrontate per consentire la produzione industriale. Nello specifico, risulta complicato ridurre la corrente di reset, ovvero la corrente necessaria per fondere il GST (circa 20-30 μA), ai valori caratteristici della tecnologia CMOS. Infatti, per fondere il GST, è necessaria un'alta temperatura e quindi un'alta densità di corrente. Inoltre, al fine di ottenere una maggiore densità di corrente non è possibile ridurre le dimensioni dei layer principali in quanto essi presentano già dimensioni sub-litografiche. Le PCM, in quanto dispositivo passivo a due terminali, richiedono un selettore per ogni singola cella. I MOSFET scalati non consentono il passaggio di una corrente tale da attivare l'operazione di reset e quindi le principali alternative sono rappresentate dai diodi in polisilicio [12] o da interruttori ovonici a soglia [13], ovvero dispositivi caratterizzati da uno switching a soglia, ma non a transizione di fase. Infine l'instabilità degli stati resistivi limita il funzionamento multilivello delle celle. In particolare, lo stato amorfo è uno stato metastabile che, per tempi lunghi, mostra un incremento di resistenza, dovuto all'annichilazione dei difetti responsabili del passaggio di corrente, seguita da una riduzione dovuta alla cristallizzazione [14, 15, 16].

1.3.2 Memorie Spin-Transfer-Torque (STT-MRAM)

Le memorie magnetiche, come ad esempio gli hard disk, esistono ormai da molti decenni e rappresentano la tecnologia più utilizzata per le memorie ad alta densità [1, 6]. Il principio fisico in base al quale vengono memorizzati i bit d'informazione '0' e '1', è abbastanza

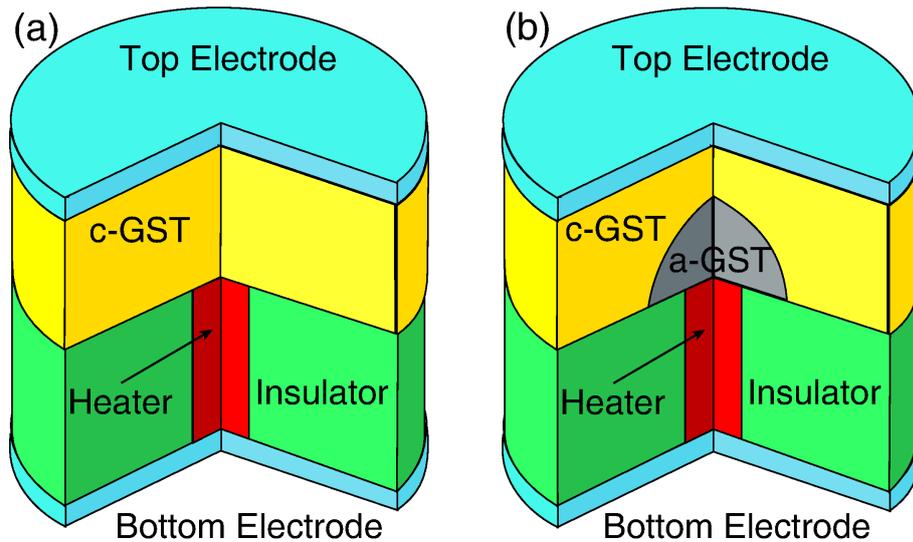


Figura 1.5: Struttura di una PCM. Tra i due elettrodi detti rispettivamente top e bottom electrode sono interposti un GST cristallino e un isolante in cui un heater consente il passaggio di una corrente e il riscaldamento locale. Lo stato a bassa resistenza corrisponde alla regione completamente cristallina, (a), mentre lo stato di reset mostra la caratteristica regione amorfa a forma di cupola, (b).

semplice. La magnetizzazione locale di un disco ferromagnetico può essere modificata per memorizzare l'informazione e lo stato dello spin viene rivelato attraverso una testina che può leggere e scrivere lo stato magnetico. La velocità di lettura/scrittura è molto bassa, dell'ordine dei kbit/s, a causa dell'utilizzo di parti meccaniche che devono spostare la testina per raggiungere la posizione dei dati nel disco.

Negli ultimi anni c'è stata un'intensa ricerca nel campo delle memorie magnetiche, denominate spin transfer torque magnetic RAM (STT-MRAM). Il principio base è la memorizzazione dell'informazione attraverso la magnetizzazione di materiali ferroelettrici, come per gli hard disk, ma leggendo e scrivendo non mediante parti meccaniche, ma applicando impulsi elettrici, in modo da ottenere un'elevata velocità di switching. Come mostrato in Fig. 1.6, la struttura fondamentale, detta giunzione a tunnel magnetico (Magnetic tunnel junction, MTJ), è costituita da un metallo ferromagnetico, un isolante sottile e un altro metallo ferromagnetico. I due layer ferromagnetici sono tipicamente realizzati con leghe come il CoFeB [3]. Uno dei due è detto layer fisso in quanto la magnetizzazione non può essere modificata e quindi per questo motivo è considerato il layer di riferimento [17, 3]. L'altro layer ferromagnetico presenta invece una magnetizzazione che può commutare tra

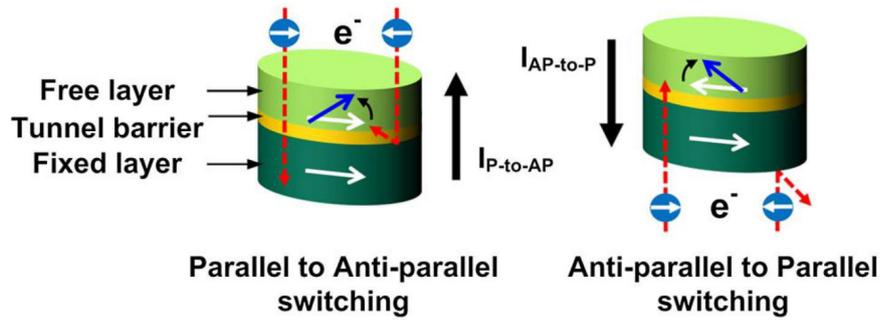


Figura 1.6: Principio di funzionamento delle memorie STT. Il primo disegno spiega la commutazione dallo stato parallelo allo stato anti-parallelo dovuta agli elettroni aventi spin opposti rispetto alla magnetizzazione del layer fisso. Al contrario, il secondo disegno rappresenta lo switching dallo stato anti-parallelo allo stato parallelo. Tratto da [17].

parallela e anti-parallela, ovvero può avere la stessa direzione o quella opposta rispetto alla magnetizzazione del layer fisso. Per commutare la magnetizzazione di un layer anti-parallelo in quella del layer di riferimento deve essere attivato il passaggio di una corrente dal lato di quello anti-parallelo (in alto in Fig. 1.6) [17]. Essendo la corrente non polarizzata, gli spin degli elettroni mostrano orientamenti casuali. Quando gli elettroni urtano il layer fisso, quelli con spin anti-parallelo, a differenza degli altri, vengono respinti verso il layer fisso, generando una torsione sugli elettroni del layer libero. In Fig. 1.6 [3, 17] si osserva, che se la corrente è abbastanza alta, la torsione applicata è tale da commutare in modo permanente la magnetizzazione del layer ferromagnetico libero. Al contrario, per passare da una magnetizzazione anti-parallela ad una parallela, il layer fisso dovrà essere attraversato da una corrente non polarizzata (in basso in Fig. 1.6). In seguito al passaggio della corrente, gli spin degli elettroni risultano orientati parallelamente alla magnetizzazione del layer fisso. Dopodichè, una corrente polarizzata fluisce attraverso il layer libero esercitando una torsione sul vettore magnetizzazione del layer che risulterà in definitiva orientato parallelamente a quello del layer fisso.

I due stati sono elettricamente distinguibili poichè lo stato parallelo mostra una resistenza più bassa rispetto a quella dello stato anti-parallelo. Uno dei principali problemi delle STT-MRAM è la finestra resistiva molto stretta, circa un fattore due, tra gli stati parallelo e anti-parallelo. Un altro limite è la stabilità termica della magnetizzazione del layer libero che può influenzare negativamente la ritenzione dei dati. Il tempo di switching τ da uno stato all'altro può essere modellizzato con la formula Arrhenius [17, 4, 3]

$$\tau = \tau_0 e^{\frac{E}{kT}} \quad (1.1)$$

dove E è la barriera degli stati, k è la costante di Boltzmann, T è la temperatura e τ_0 è il tempo medio del processo, che è circa pari a 1 ns. Molti sforzi sono stati compiuti per aumentare l'altezza di barriera. Tuttavia, aumentare eccessivamente E può condurre ad un aumento del tempo di switching e quindi ne segue che è più conveniente un trade-off.

Infine è stata dimostrata la possibilità di scendere al di sotto del nodo 20 nm [3]. Questa proprietà, che si aggiunge all'eccellente resistenza alla ciclatura (endurance), che risulta quasi illimitata, e all'elevata velocità di switching (circa 10 ns), rende queste memorie magnetiche in grado di sostituire adeguatamente le DRAM. Tutte queste premesse hanno permesso alle STT-MRAM di candidarsi come possibili SCM.

1.3.3 Memorie RRAM

Un'altra famiglia di memorie resistive potenzialmente in grado di soddisfare i requisiti delle SCM sono le memorie RRAM [18, 8, 19, 20, 7]. Queste memorie sono interessanti per molti motivi, tra i quali spicca la grande facilità di fabbricazione, in quanto sono costituite esclusivamente da due elettrodi metallici (generalmente in Platino (Pt) o Nitruro di Titanio (TiN)) e da un isolante interposto tra essi. Inoltre, le RRAM possono essere realizzate durante il back-end del processo CMOS, ottenendo grandi vantaggi in termini di costo e occupazione di area [7, 9].

Il dispositivo può assumere due stati, ovvero uno stato ad alta resistenza (high-resistive state, HRS) e uno stato a bassa resistenza (low-resistive state) [8]. La transizione dallo stato alto resistivo allo stato basso resistivo è denominata set mentre quella opposta è detta reset. I primi fenomeni di switching resistivo osservati risalgono al 1960 [21]. A quel tempo lo switching era possibile soltanto applicando tensioni molto elevate, circa 100 V, a causa dell'impossibilità di realizzare ossidi abbastanza sottili. Tuttavia, negli ultimi quindici anni, i progressi del processo tecnologico hanno permesso la crescita di layer di spessore inferiore a 100 nm, rendendo così possibile lo switching a tensioni molto più basse (al di sotto di 5 V) e la realizzazione di queste memorie innovative [7]. Tra i numerosi materiali caratterizzati da switching resistivo, i più importanti sono gli ossidi metallici binari. Inizialmente sono stati studiati soprattutto materiali come l'ossido di nickel (NiO) [22, 23, 24, 25] o il biossido di titanio (TiO₂) [26, 27]. Successivamente sono stati sfruttati altri materiali come l'ossido di Afnio, HfO₂ [19, 28, 29], l'ossido di alluminio, Al₂O₃ [30] e l'ossido di tantalio, Ta₂O₅

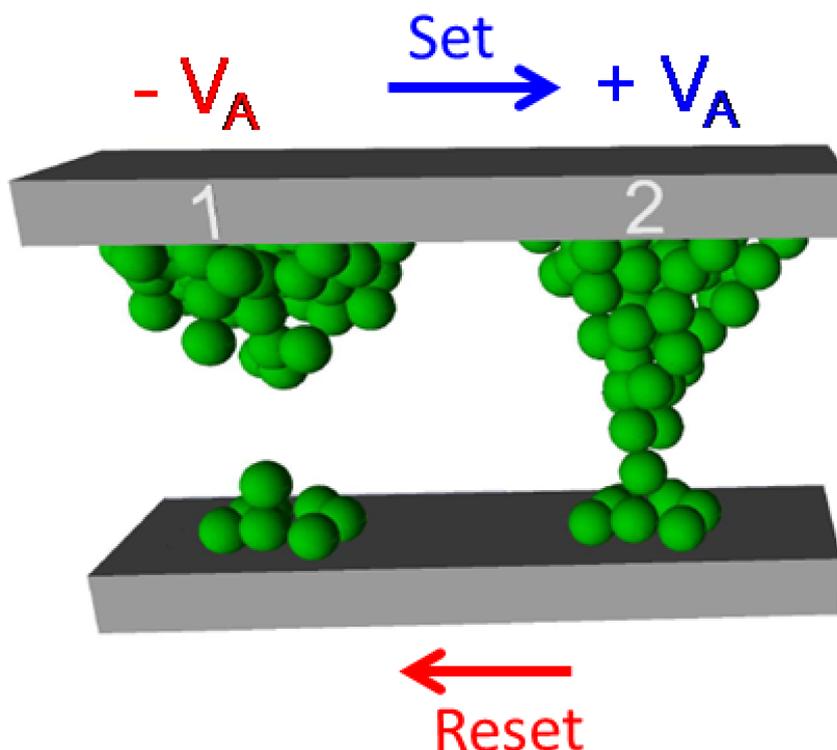


Figura 1.7: Pittura fisica del meccanismo di switching delle RRAM. La transizione di set porta alla formazione di un filamento conduttivo, mentre la transizione di reset attiva la rottura del filamento. Questi due fenomeni si verificano per tensioni applicate di polarità opposta.

[31, 32, 33, 34], i quali possono essere fabbricati facilmente in quanto composti da elementi già impiegati nel processo CMOS [2].

Il principio fisico dello switching è ancora oggetto di dibattito tra i ricercatori [7]. In generale, esistono due tipi di switching resistivo: lo switching unipolare e quello bipolare [8]. Nelle memorie a switching unipolare le due transizioni di set e reset sono attivate mediante l'applicazione di impulsi della stessa polarità, mentre in quelle bipolari attraverso impulsi di tensione di polarità opposta. Attualmente, sia la ricerca che l'industria si sono concentrate soprattutto sulle memorie a switching bipolare in quanto forniscono prestazioni superiori rispetto a quelle unipolari [2]. La prima operazione eseguita sulle RRAM bipolari è detta forming [35]. Attraverso l'applicazione di una tensione, ad esempio positiva, il forming causa un breakdown non distruttivo dell'isolante che, a sua volta, porta alla formazione

di una regione localizzata in cui la concentrazione di ossigeno è ridotta con conseguente incremento della densità di vacanze o difetti. La crescita della corrente che attraversa il dispositivo è bloccata da un limitatore esterno, come ad esempio un transistor, ad un valore massimo denominato corrente di compliance, o I_C [29]. Lo stato finale è caratterizzato da un filamento conduttivo (CF) che connette i due elettrodi [36]. A questo punto, applicando all'elettrodo superiore del dispositivo una tensione negativa, la cella può essere commutata nello stato HRS (operazione di reset). Questa transizione si verifica in quanto la tensione negativa applicata genera un campo elettrico che trascina i difetti ionizzati verso l'elettrodo superiore, causando così la rottura del filamento conduttivo [37]. Per commutare nuovamente nello stato a bassa resistenza basta applicare una tensione positiva che ripristini il CF.

Un approccio simile, ma comunque differente, è quello adottato dalle memorie resistive a ponte conduttivo (CBRAM). In questi dispositivi, il top electrode (TE) è un elettrodo attivo, ovvero è realizzato con materiali metallici come il rame (Cu) o l'argento (Ag), i quali, durante il set, penetrano nell'ossido (tipicamente un calcogenuro) formando in questo modo un CF tra i due terminali [38, 39], mentre durante il reset si ritirano verso il TE. L'elettrodo inferiore è invece realizzato con un materiale inerte.

Sia le RRAM che le CBRAM sono caratterizzate da tempi di switching molto brevi dell'ordine dei nanosecondi [28]. Questi due tipi di memorie resistive assicurano un funzionamento a basso consumo in quanto le tensioni richieste non sono superiori a 1-2 V e le correnti sono comprese in un range di 20-30 μA [40]. Tuttavia, quando si riduce la corrente di compliance, si manifestano numerose problematiche relative all'affidabilità e variabilità (instabilità del filamento conduttivo) [41], alla comparsa del rumore RTN (random telegraph noise) [42, 43], a effetti di stress [44], alla ritenzione dei dati memorizzati [30, 45] e alla degradazione dell'endurance.

1.4 Introduzione alle reti neuromorfiche

L'implementazione di sistemi in grado di emulare il funzionamento del cervello umano rappresenta una delle principali sfide della ricerca nel campo delle tecnologie per l'informazione e la comunicazione. Il cervello degli esseri umani è capace di eseguire attività di estrema complessità come, ad esempio, l'apprendimento e riconoscimento di pattern visivi/uditivi e l'adattamento all'arrivo di nuove informazioni. Per affrontare questa grande sfida, i sistemi neuronali artificiali devono essere costituiti da un numero di neuroni e sinapsi simile a quello del sistema nervoso umano, il quale è caratterizzato da circa 10^{12} neuroni e 10^{15} sinapsi [46].

Ovviamente, un sistema così complesso può essere implementato unicamente attraverso il miglioramento delle tecniche di fabbricazione (integrazione 3D) e la realizzazione di blocchi compatti in grado di simulare i neuroni e le sinapsi.

In particolare, dato il numero estremamente elevato di connessioni inter-neuronali, nella progettazione di reti neuromorfiche l'attenzione è stata focalizzata principalmente sulla miniaturizzazione e riduzione di complessità delle sinapsi elettroniche. Pertanto, per la realizzazione di sinapsi ultra-compatte sono stati proposti switch resistivi di dimensioni nanometriche detti memristori [47, 48, 49]. I memristori, rispetto ai dispositivi basati su silicio come le memorie a floating gate [50] e le SRAM [51], mostrano grandi vantaggi come il comportamento intrinsecamente analogico, la semplice struttura a due terminali e la scalabilità dell'area occupata e della potenza dissipata. Nello specifico sono stati proposti differenti tipi di memorie come le PCM [52, 53, 54], gli switch organici [55], gli interruttori calcogenuri [56, 57] e le memorie a switching resistivo (RRAM) [58, 59, 60, 61, 62]. Tra questi, le memorie RRAM assicurano molti vantaggi come lo switching analogico, il comportamento non volatile, la compatibilità con la tecnologia CMOS e la scalabilità della potenza dissipata [7].

I memristori garantiscono sia la regolabilità elettrica delle conduttanze delle sinapsi che la capacità di rispondere con un'eventuale variazione del peso agli impulsi generati dal pre-neurone e dal post-neurone. Per realizzare questo funzionamento in passato è stato proposto un approccio detto TDM (Time division multiplexing) in cui l'evoluzione degli eventi di comunicazione, potenziamento e depressione era regolata in modo totalmente sincrono. Ciò tuttavia risulta troppo ideale rispetto all'effettivo funzionamento biologico del cervello umano in quanto quest'ultimo è costituito da sinapsi che vengono potenziate e depresse in modo asincrono secondo il protocollo spike-timing-dependent-plasticity (STDP) [63]. Inoltre, una scansione temporale degli eventi totalmente sincrona risulta impossibile da realizzare per reti neuromorfiche molto estese [64]. Recentemente è stato presentato un approccio completamente asincrono per l'apprendimento/comunicazione delle sinapsi elettroniche, in cui vengono adottati neuroni leaky-integrate-and-fire (LIF) [64, 65].

Talvolta lo switching analogico è difficile da ottenere in quanto, a causa della riduzione delle dimensioni e delle correnti, si ha un aumento della variabilità statistica dei parametri dei dispositivi [58, 66] che richiede l'impiego di algoritmi di controllo e di circuiti ad essi associati [67]. Le fluttuazioni statistiche sono tipicamente dannose per il funzionamento digitale di una memoria, ma possono essere tollerate in alcune applicazioni come la generazione di numeri casuali [68, 69, 70, 71] e la computazione neuromorfica [72, 73, 74]. Le reti neuromorfiche riescono inoltre a sfruttare a proprio vantaggio le variazioni stocastiche che contribuiscono

al normale funzionamento delle reti neuronali degli animali e degli esseri umani [75].

In definitiva, per progettare sistemi neuromorfici avanzati, è fondamentale sviluppare sinapsi artificiali molto compatte impiegando i moderni dispositivi RRAM in regime di switching stocastico.

Capitolo 2

Modello della cella RRAM

Questo capitolo è dedicato alla modellizzazione dei processi di set e reset che caratterizzano il funzionamento delle memorie a switching resistivo. Alla descrizione del classico modello filamentare seguirà quella del più avanzato modello a gap variabile.

2.1 Memorie resistive a switching bipolare

Le memorie non volatili a switching resistivo sono caratterizzate da una struttura metallo-isolante-metallo (MIM).

L'isolante è un ossido metallico (tipicamente ossido di Hafnio (HfO_x), ossido di Titanio (TiO_x) o ossido di Nickel (NiO)) mentre i due elettrodi, detti rispettivamente bottom electrode (BE) e top electrode (TE), sono layer metallici realizzati con materiali relativamente inerti come Nitrato di Titanio (TiN), Tungsteno (W) o Platino (Pt). Inizialmente questa struttura mostra un'elevata resistenza a causa della presenza dello strato isolante, il cui spessore può variare tra 3 e 20 nm.

Per attivare il meccanismo di switching resistivo, la cella viene sottoposta ad una tensione tale da indurre il breakdown dielettrico dello strato isolante. Questo processo è detto comunemente forming ed è controllato limitando la massima corrente che attraversa la cella, mediante l'aggiunta in serie di un transistor MOS, in modo da evitare il breakdown irreversibile del dielettrico e permettere la formazione di un filamento conduttivo (conductive filament, CF) di dimensione controllabile. Il filamento mostra un'elevata conducibilità elettrica locale in virtù della presenza di un eccesso di impurezze metalliche (ioni Hf se si adotta come isolante l' HfO_x) o di vacanze di ossigeno (V_o). Quindi, dopo il forming, data l'alta conducibilità elettrica del filamento, la cella mostra una bassa resistenza. Questa condizione

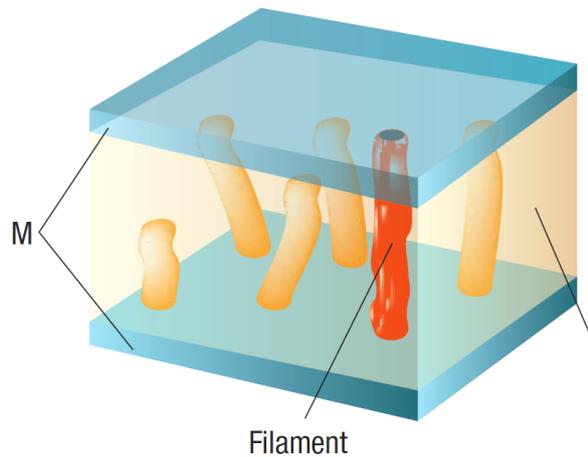


Figura 2.1: Struttura metallo-isolante-metallo [8].

di bassa resistenza della cella è detta anche stato di set (low-resistive state, LRS). Una volta formato, è possibile indurre elettricamente la dissoluzione del filamento mediante un'operazione detta reset. L'annichilazione del filamento, costituito da un eccesso di ioni Afnio o equivalentemente di vacanze di ossigeno, determinerà il passaggio della cella RRAM in uno stato ad alta resistenza (high-resistive state, HRS) detto anche stato di reset. Dato che la cella di Afnio presenta un comportamento bipolare, la transizione di set può essere attivata applicando una tensione, per esempio positiva, al top electrode del dispositivo, mentre per quella di reset si applica una tensione di polarità opposta. Per comprendere il meccanismo di switching bipolare si ricorre al fenomeno di migrazione ionica, in base al quale i difetti ionizzati, muovendosi sotto l'azione del campo elettrico, cambiano la forma e la conducibilità del filamento conduttivo.

2.2 Modello filamentare

Nel modello filamentare le transizioni di set e reset si traducono rispettivamente nella formazione e dissoluzione di un filamento conduttivo costituito da ioni metallici o, equivalentemente, da vacanze di ossigeno la cui migrazione, guidata dal campo elettrico, risulta accelerata dalla crescita locale di temperatura dovuta all'effetto Joule [29, 76].

In Fig. 2.2 è descritta la cinetica del processo di set di una RRAM bipolare dove la crescita del CF è governata dalla migrazione ionica attivata termicamente. Inizialmente i

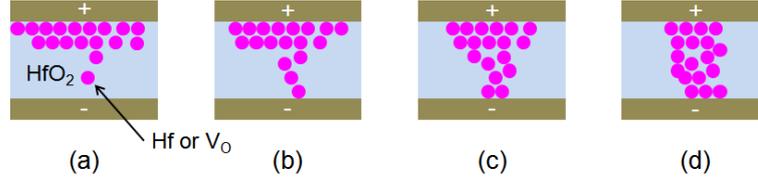


Figura 2.2: Illustrazione schematica della transizione di set. I difetti sono addensati al TE, (a), come conseguenza di un precedente reset della cella. In (b) si osserva la nucleazione di un sottile filamento conduttivo, il cui spessore nel tempo tende a crescere, (c), finchè la cella non raggiunge il definitivo stato di set, (d) [29].

difetti sono addensati al TE per cui la cella resistiva si trova in uno stato di reset, come indicato in Fig. 2.2 (a). In Fig. 2.2 (b) viene rappresentato il successivo processo di nucleazione del CF. Si ritiene che si formi un canale conduttivo avente uno spessore di pochi atomi in seguito alla migrazione degli ioni lungo un cammino preferenziale ad alta concentrazione di difetti. Questo percorso è caratterizzato da un addensamento di campo elettrico tale da generare un abbassamento della barriera delle buche di potenziale che gli ioni incontrano lungo il loro spostamento. In Fig. 2.2 (c) viene rappresentata la crescita dello spessore del CF che termina con il raggiungimento dello stato finale di set, come mostrato in Fig. 2.2 (d). Questa successione di processi è possibile a patto che al TE sia presente una riserva di ioni metallici.

La crescita del CF durante il set può essere modellizzata mediante una semplice equazione analitica in cui lo stato del filamento è indicato da un unico parametro, ovvero il diametro ϕ . In questo modello, il diametro del filamento cresce nel tempo con un tasso definito dalla seguente equazione:

$$\frac{d\phi}{dt} = A e^{-\frac{E_A}{kT}} \quad (2.1)$$

dove A è una costante pre-esponenziale, k è la costante di Boltzmann, T è la temperatura del filamento ed E_A è l'energia di attivazione del processo di migrazione ionica definita dall'espressione

$$E_A = E_{A0} - \alpha qV \quad (2.2)$$

dove E_{A0} rappresenta l'energia di attivazione del processo di migrazione ionica attraverso buche di potenziale successive in assenza di campo elettrico, α è il coefficiente di abbassamento di barriera, q è la carica elementare e V è la tensione ai capi della cella.

In Fig. 2.3 (1) è mostrata la successione delle buche di potenziale a cui sono sottoposti gli ioni metallici in assenza di campo elettrico applicato. Il termine αqV , indicato in

Fig. 2.3 (2), descrive l'abbassamento della barriera di potenziale in presenza di un campo elettrico applicato. Tale riduzione dell'altezza di barriera comporta che gli ioni si muovano sotto l'azione del campo elettrico con una maggiore mobilità.

La temperatura del filamento risente in modo significativo del riscaldamento per effetto Joule a causa dell'alta densità di corrente che attraversa il dispositivo. Per calcolare la temperatura del CF si risolve l'equazione di Fourier in regime stazionario

$$k_{th} \frac{d^2 T}{dz^2} + J^2 \rho = 0 \quad (2.3)$$

dove z è la coordinata spaziale lungo il filamento, k_{th} è la conducibilità termica dell'Afnio, pari a 23 W/mK, J è la densità di corrente e ρ rappresenta la resistività del metallo nel CF [29].

Imponendo come condizioni al contorno $T(0) = T(L) = T_0 = 300$ K, dato che gli elettrodi, a distanza L l'uno dall'altro, non si scaldano, ma rimangono a temperatura ambiente, la soluzione dell'equazione è un profilo spaziale di temperatura parabolico avente il vertice in corrispondenza del centro del filamento. Il valore del massimo di temperatura del CF è dato dalla seguente espressione:

$$T_{max} = T_0 + \frac{J^2 \rho L^2}{8k_{th}} \quad (2.4)$$

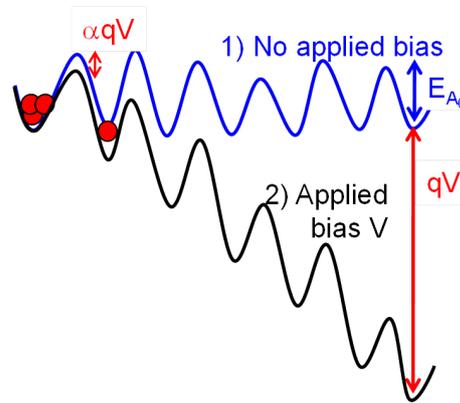


Figura 2.3: Rappresentazione semplificata delle buche di potenziale nell'ossido. In assenza di una tensione applicata, gli ioni vedono delle buche di altezza pari a E_{A_0} , (1). Applicando invece una tensione, l'altezza della barriera di potenziale si riduce di un fattore αqV , favorendo così la migrazione, (2).

Assumendo per semplicità un filamento cilindrico avente sezione di area

$$A_{CF} = \frac{\pi\phi^2}{4} \quad (2.5)$$

la resistenza del filamento è uguale a

$$R = \frac{\rho L}{\frac{\pi}{4}\phi^2} \quad (2.6)$$

Se esprimiamo poi la densità di corrente in funzione della tensione V ai capi della cella

$$J = \frac{V}{\rho L} \quad (2.7)$$

e sostituiamo (2.7) in (2.4), otteniamo

$$T_{max} = T_0 + \frac{V^2}{8\rho k_{th}} \quad (2.8)$$

ovvero una relazione che stabilisce che la temperatura dipende unicamente dalla tensione applicata alla cella e non dal diametro del filamento. Infine, sostituendo (2.8) e (2.2) in (2.1) si ricava la seguente equazione

$$\frac{d\phi}{dt} = Ae^{-\frac{E_{A0} - \alpha q V}{k(T_0 + \frac{V^2}{8k_{th}\rho})}} \quad (2.9)$$

che, risolta numericamente, restituisce l'evoluzione del CF durante il set.

Questa relazione conferma che la crescita del filamento conduttivo è governata unicamente dalla tensione applicata alla cella.

Mediante questo modello è inoltre possibile descrivere analiticamente anche il processo di dissoluzione del CF, a condizione di aggiungere un segno meno davanti al fattore pre-esponenziale A .

In Fig. 2.4 è rappresentata la tipica configurazione in cui viene impiegata una RRAM ovvero la struttura 1T1R. Il transistor MOS, connesso in serie alla cella resistiva, permette di controllare la crescita del CF durante il set in quanto impone un limite alla corrente che può attraversare la cella.

Si consideri ora la caratteristica corrente-tensione mostrata in Fig. 2.5, ottenuta applicando alla cella 1T1R un impulso triangolare di tensione positivo seguito da uno negativo. Si osserva che V_{set} e V_{reset} sono uguali in modulo a conferma del fatto che sia la crescita che la rottura del CF sono processi governati dalla temperatura e quindi dalla tensione, indipendentemente dal diametro del filamento.

Analizziamo ora le due transizioni, partendo dal set. La cella si trova inizialmente in uno stato alto resistivo e ammettiamo che la resistenza R di quest'ultima sia molto maggiore rispetto alla resistenza R_{MOS} del transistor. Al top electrode viene applicata una rampa di

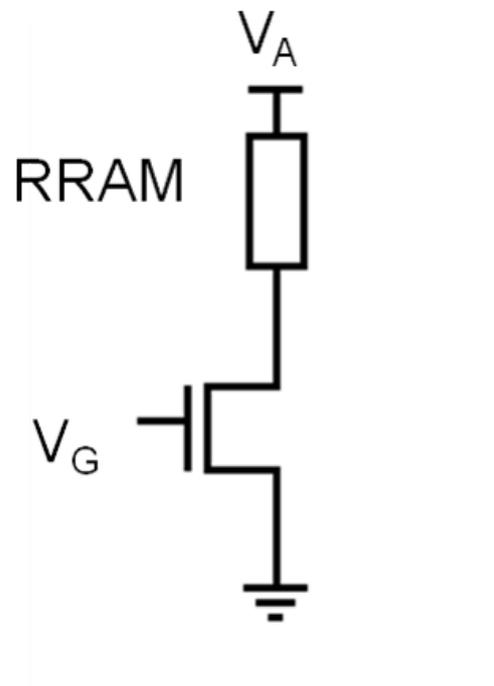


Figura 2.4: Struttura 1T1R. La tensione di alimentazione V_A si ripartisce tra la cella e il transistore.

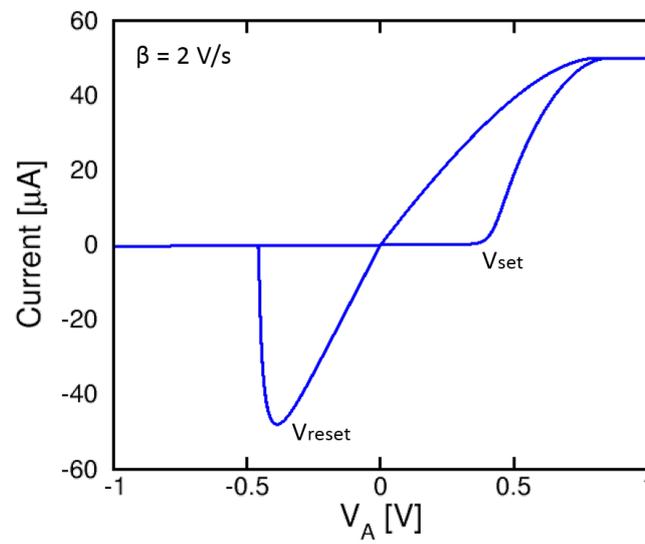


Figura 2.5: Caratteristica I-V simulata con il modello filamentare. $I_C \simeq 50 \mu\text{A}$, $V_{set} = |V_{reset}| \simeq 0.4$ V.

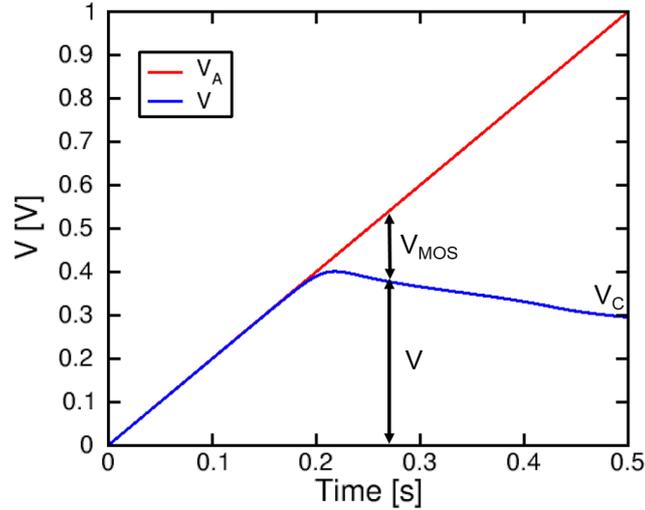


Figura 2.6: Andamento calcolato di V_A (in rosso) e della tensione V ai capi della cella (in blu) in funzione del tempo. Dopo un'iniziale crescita a rampa, V si stabilizza alla tensione V_C , che è uguale a 0.3 V.

tensione positiva V_A e tale tensione si ripartisce tra la cella resistiva e il transistor secondo la seguente relazione:

$$V = V_A \frac{R}{R + R_{MOS}} \quad (2.10)$$

All'aumentare di V_A , il diametro del CF ϕ cresce e di conseguenza, essendo $R \propto \phi^{-2}$, la resistenza della cella diminuisce. La riduzione di R si riflette nella diminuzione della tensione V ai capi della cella, che, di conseguenza, determina un abbassamento di temperatura nel CF. Ne segue pertanto che la crescita del CF rallenta fino ad arrestarsi raggiungendo così lo stato di set completo.

E' importante ora soffermarsi sulla dipendenza della resistenza R dalla corrente di compliance. Questa dipendenza può essere interpretata attraverso una reazione negativa dovuta al meccanismo di compliance. Quando V raggiunge V_{set} e quindi ha inizio la transizione di set, la crescita del CF determina un incremento della corrente finchè non viene eguagliata I_C . Raggiunta I_C , dato che un'ulteriore crescita della corrente è impedita dal meccanismo di limitazione dovuto al transistor connesso in serie, la crescita del CF viene accomodata da una riduzione della tensione V ai capi della cella, detta snapback, che si esaurisce quando V eguaglia V_C . La resistenza finale di set della cella si ottiene quindi attraverso la seguente

equazione:

$$R = \frac{V_C}{I_C} \quad (2.11)$$

Quest'ultima equazione fissa una relazione di proporzionalità inversa tra I_C e la resistenza finale dello stato di set, come mostrato in Fig 2.7 (a).

Raggiunto lo stato di set completo, per iniziare il processo di reset applichiamo al TE una rampa di tensione negativa. Assumiamo il transistor polarizzato con una tensione V_G alta in modo che R_{MOS} sia trascurabile ($\sim 1 \text{ k}\Omega$). In Fig. 2.5 si osserva che la tensione per cui ha inizio la transizione di reset è V_{reset} .

Dato che

$$I_{reset} = \frac{V_{reset}}{R} \quad (2.12)$$

$$I_C = \frac{V_C}{R} \quad (2.13)$$

$$V_{reset} \simeq V_C \quad (2.14)$$

si ottiene:

$$I_{reset} \simeq I_C \quad (2.15)$$

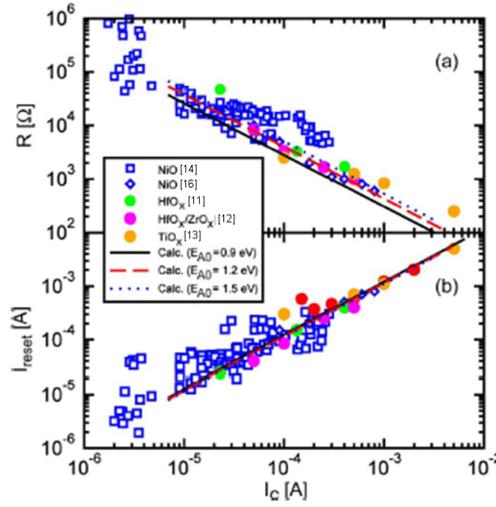


Figura 2.7: Andamento misurato e calcolato di R_{set} , (a), e I_{reset} , (b), in funzione di I_C . Sono riportati e confrontati i dati relativi a molti tipi di dispositivi e materiali come le RRAM unipolari in NiO e le RRAM bipolari in HfO_x , TiO_x e $\text{HfO}_x/\text{ZrO}_x$. Si nota che le simulazioni sono in pieno accordo con i dati sperimentali e consistenti con le formule delle Eqs.(2.11) e (2.15) [29].

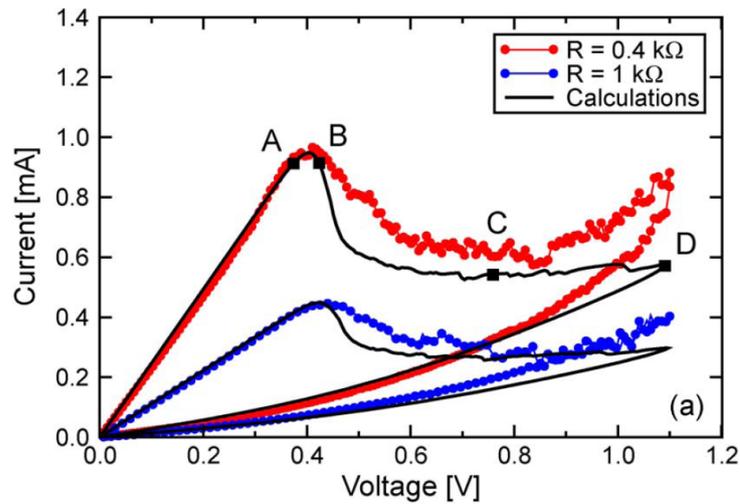


Figura 2.8: Caratteristiche I-V del reset misurate e simulate per due differenti condizioni iniziali di set. La curva blu è stata ottenuta da una resistenza iniziale del dispositivo pari a $R = 1\text{k}\Omega$, mentre quella rossa si riferisce ad uno stato resistivo iniziale di $0.4\text{k}\Omega$. Inoltre, lungo una delle curve calcolate, sono segnati dei punti, chiamati A, B, C, D, che corrispondono a particolari stati in cui si trova la cella durante la transizione di reset [37].

In conclusione, l'equivalenza tra V_{reset} e V_C e quindi tra I_{reset} e I_C è dovuta al fatto che l'evoluzione del filamento è governata unicamente dalla tensione, indipendentemente dalla sua polarità.

2.3 Modello di switching a gap variabile

Il modello di switching a gap variabile offre una nuova interpretazione dei processi di set e reset di una RRAM bipolare in termini di chiusura e apertura di una barriera di ossido nel filamento metallico attivata da effetti di feedback termici. In base a questo modello, la transizione di set risulta caratterizzata da una reazione positiva, attivata dalla temperatura, che accelera la chiusura della barriera di ossido, mentre quella di reset è controllata da una reazione negativa che regola la graduale apertura di tale barriera.

2.3.1 Reset

In Fig. 2.8 sono riportati i risultati sperimentali della transizione di reset con le relative simulazioni ottenute mediante un modello numerico realizzato con il simulatore COMSOL [37]. Questo modello studia i processi di set e reset che caratterizzano il funzionamento di una cella RRAM in termini di trasporto per deriva e diffusione degli ioni metallici o delle vacanze di ossigeno. Il filamento conduttivo viene interpretato come una regione avente una concentrazione di difetti ionizzati superiore rispetto a ciò che lo circonda. L'applicazione di una tensione attiva un flusso di ioni metallici \mathbf{j}_D che, secondo l'equazione di deriva e diffusione, è uguale a:

$$\mathbf{j}_D = -D\nabla n_D + \mu n_D \mathbf{F} \quad (2.16)$$

dove D , espresso in $[\text{cm}^2\text{s}^{-1}]$, rappresenta la diffusività ionica, μ $[\text{cm}^2\text{V}^{-1}\text{s}^{-1}]$ è la mobilità degli ioni Afnio e \mathbf{F} è il campo elettrico. La diffusione dei difetti viene trattata come un processo Arrhenius, per cui la diffusività, che permette di controllare la mobilità degli ioni attraverso la relazione di Einstein $D = \mu kT/q$, può essere espressa come $D = D_0 e^{-E_A/kT}$, dove D_0 è un prefattore ed E_A è l'energia di attivazione del processo. L'equazione di deriva e diffusione è risolta numericamente mettendo quest'ultima a sistema con l'equazione di Poisson per la densità di corrente \mathbf{j} ,

$$\nabla \cdot \mathbf{j} = 0 \quad (2.17)$$

la legge di Ohm, in cui la conducibilità aumenta linearmente con la concentrazione ionica,

$$\mathbf{j} = \sigma \mathbf{F} \quad (2.18)$$

e l'equazione di Fourier per il calore, nel caso stazionario

$$\nabla \cdot (k_{th} \nabla T) = -\mathbf{j} \cdot \mathbf{F} \quad (2.19)$$

dove k_{th} rappresenta la conducibilità termica, anch'essa linearmente crescente con la densità ionica n_D . La risoluzione di questo sistema viene eseguita in una geometria 3D che può essere ridotta a 2D in virtù della simmetria cilindrica, dato che assumiamo un filamento di forma cilindrica.

Le curve mostrate in Fig. 2.8 sono state ricavate a partire da differenti stati iniziali di set ottenuti mediante la scelta di due diverse correnti di compliance. La sequenza dei punti A, B, C, D, rappresentati sulla curva simulata a corrente maggiore, indica l'evoluzione temporale del processo di reset. Infine si osserva che V_{reset} , ovvero la tensione per cui ha inizio la transizione, rimane costante a 0.4 V, indipendentemente dal valore iniziale di resistenza di set, confermando ciò che si vede dai dati sperimentali.

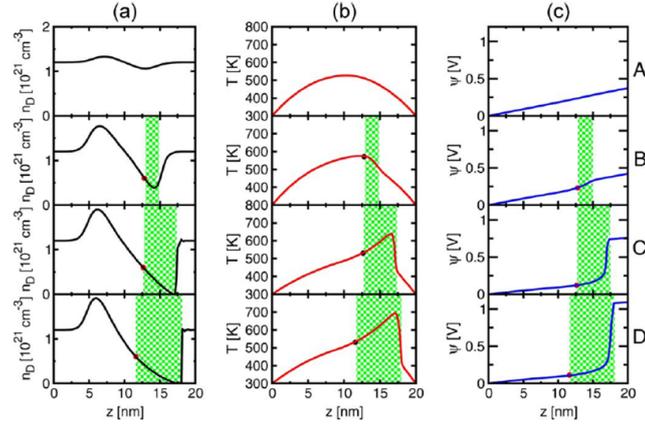


Figura 2.9: Rappresentazione dei profili della concentrazione dei difetti ionizzati n_D , (a), di temperatura, (b), e della tensione, (c), calcolati nei punti A, B, C, D delle curve mostrate in Fig. 2.8 [37].

In Fig. 2.9 (b) e 2.10 (b) sono rappresentati i profili temporali, rispettivamente 1D e 2D, di temperatura durante la transizione di reset. Come mostrato nelle Figs. 2.9 (a) e 2.10 (a), la minore conducibilità termica ed elettrica dovuta allo svuotamento di ioni metallici comporta una crescita del campo elettrico e della temperatura all'interno del gap. Con la crescita dello spessore del gap, il picco di temperatura sarà sempre più localizzato in corrispondenza dell'estremità del gap prossima al BE, dove, come si osserva nelle Figs. 2.9 (c) e 2.10 (c), si ha anche il picco del profilo di potenziale e quindi di campo elettrico.

In Fig. 2.11 viene rappresentato schematicamente il processo di reset dove la migrazione dei difetti ionizzati, guidata dal campo elettrico e accelerata dalla temperatura in direzione del TE polarizzato negativamente, è responsabile della graduale apertura di una regione svuotata lungo il CF. La condizione iniziale (a) consiste in un stato di set completo che fisicamente si traduce in un filamento metallico continuo, di cui, per semplicità, assumiamo forma cilindrica. Come già discusso per il set del modello filamentare, risolvendo l'equazione di Fourier unidimensionale in regime stazionario si ottiene un profilo parabolico $T(z)$, il cui massimo è situato a metà del filamento. Quando T_{max} raggiunge la temperatura critica per l'attivazione della migrazione ionica, inizia ad aprirsi un gap proprio al centro del CF dove si ha il picco di temperatura e quindi la massima accelerazione del moto di deriva degli

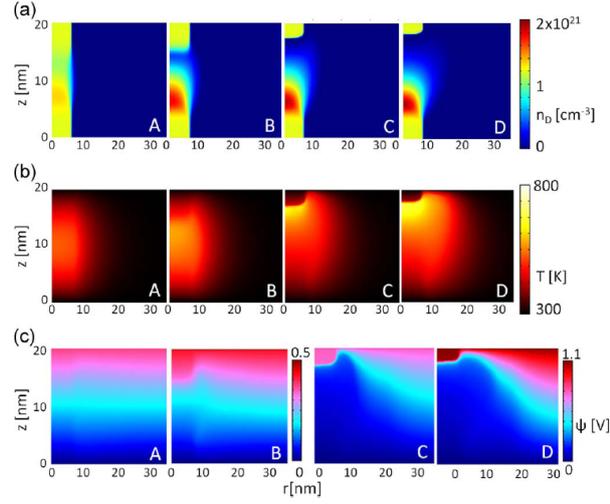


Figura 2.10: Mappe 2D della concentrazione dei difetti ionizzati n_D , (a), di temperatura, (b), e della tensione, (c), calcolate nei punti A,B,C,D di Fig 2.8 [37].

ioni, (b). Il processo di reset porta pertanto alla rottura del CF in due estremità separate da un gap di lunghezza Δ , (c). All'aumentare della tensione negativa al TE si registra una progressiva crescita della lunghezza del gap, (d), dovuta alla maggiore intensità con cui il campo elettrico, a cui si aggiunge l'effetto della temperatura, trascina i difetti ionizzati verso il TE, erodendo così lo spessore dell'estremità inferiore del filamento. Essendo il gap una regione svuotata di ioni metallici, l'incremento di spessore del gap si traduce in un aumento di resistenza che culmina con il raggiungimento di uno stato alto resistivo. Per descrivere analiticamente il reset, il tasso di crescita della lunghezza del gap può essere definito mediante la seguente relazione Arrhenius:

$$\frac{d\Delta}{dt} = A e^{-\frac{E_A}{kT(z_2)}} \quad (2.20)$$

dove A è una costante pre-esponenziale, k è la costante di Boltzmann, T è la temperatura del filamento, z_2 è la coordinata spaziale associata all'estremità inferiore del gap ed E_A è l'energia di attivazione del processo di migrazione ionica, espressa come

$$E_A = E_{A0} - \alpha q V_{gap} \quad (2.21)$$

dove E_{A0} è l'energia di attivazione del processo di migrazione ionica in assenza di campo elettrico, α è il coefficiente di abbassamento di barriera, q è la carica elementare e V_{gap} è la tensione che cade ai capi del gap.

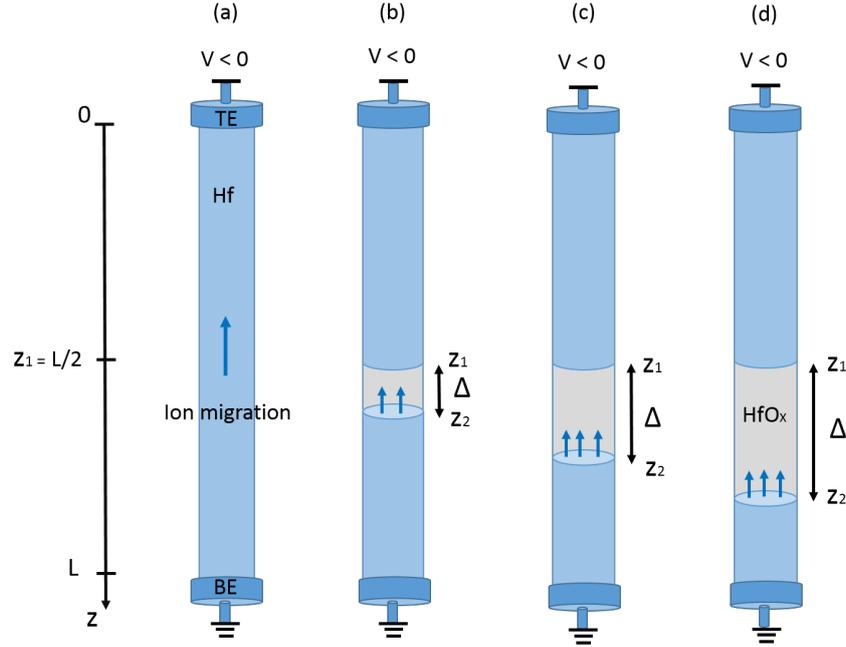


Figura 2.11: Illustrazione schematica della transizione di reset secondo il modello a gap variabile. Il reset della cella resistiva è dovuto all'apertura di un gap Δ a partire dal centro del filamento continuo laddove il profilo di temperatura presenta il proprio massimo. Da (a) dove viene rappresentato il CF continuo (stato di set completo), mediante la migrazione ionica verso il TE, inizia ad aprirsi un gap, (b), la cui lunghezza cresce gradualmente, (c), finchè non viene raggiunto lo stato finale di reset, (d) [77].

Per determinare il profilo di temperatura della cella durante il reset basta risolvere, con opportune condizioni al contorno, l'equazione di Fourier, nelle tre regioni (estimità metallica superiore, gap ed estimità metallica inferiore) evidenziate in Fig 2.11 (b - c - d). Nel calcolo si assume che i due frammenti di CF abbiano uguale resistività ρ_m e conducibilità termica $k_{th,m}$, mentre la resistività della regione svuotata ρ_{gap} è espressa come

$$\rho_{gap} = \frac{\rho_{ox}}{1 + \gamma F} \quad (2.22)$$

dove γ è una costante pari a circa 55 nm/V , F è il campo elettrico locale e ρ_{ox} è la resistività del gap in condizione di campo elettrico nullo. Quest'ultima equazione tiene conto della non linearità della conduzione nello stato di reset.

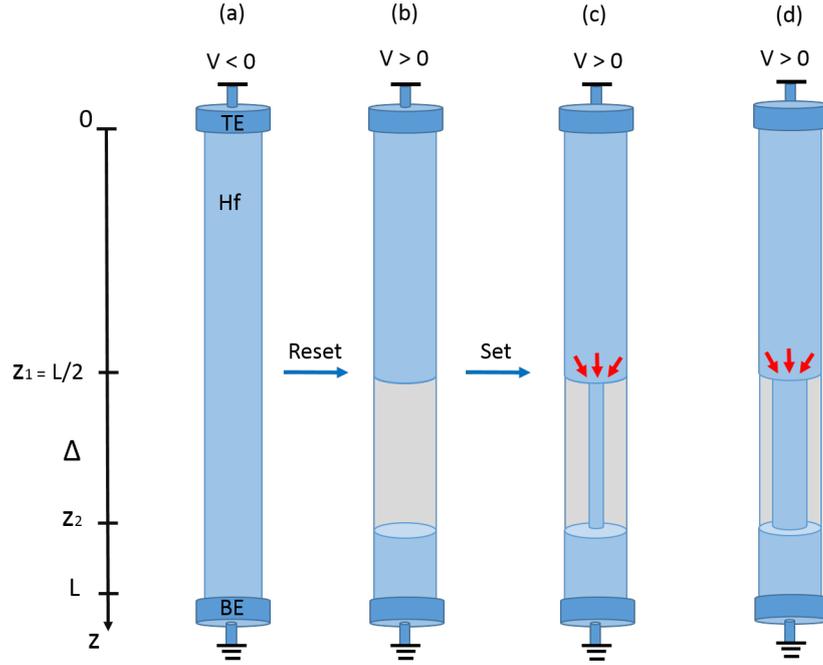


Figura 2.12: Illustrazione schematica del set secondo il modello a gap variabile. In (a) è rappresentato un filamento metallico continuo risultato di un precedente set. Applicando una tensione negativa al TE si esegue il reset della cella che comporta l'apertura di un gap nel CF, (b). Dopo l'operazione di reset, cambiando la polarità della tensione fornita al TE, viene eseguito un set del dispositivo che porta alla nucleazione, (c), e alla successiva crescita di diametro, (d), di un CF all'interno dell'ossido [77].

2.3.2 Set

Per descrivere il processo di set della cella resistiva è opportuno fare riferimento alla rappresentazione schematica riportata in Fig. 2.12. In Fig. 2.12 (a) è rappresentato lo stato iniziale della cella che prevede la presenza di un filamento metallico continuo. Applicando al TE una tensione negativa si ottiene il reset del dispositivo e quindi l'apertura di un gap nel CF, (b). A questo punto, per $V > 0$, si verifica la transizione nello stato a bassa resistenza, detta set, che si traduce nella formazione, (c), seguita dalla crescita del diametro, (d), di un filamento conduttivo all'interno dell'ossido.

Analiticamente, la transizione di set può essere descritta dalla seguente equazione che regola la velocità di chiusura della barriera:

$$\frac{d\Delta}{dt} = Ae^{-\frac{E_A}{kT(z_1)}} \quad (2.23)$$

dove Δ rappresenta lo spessore del gap.

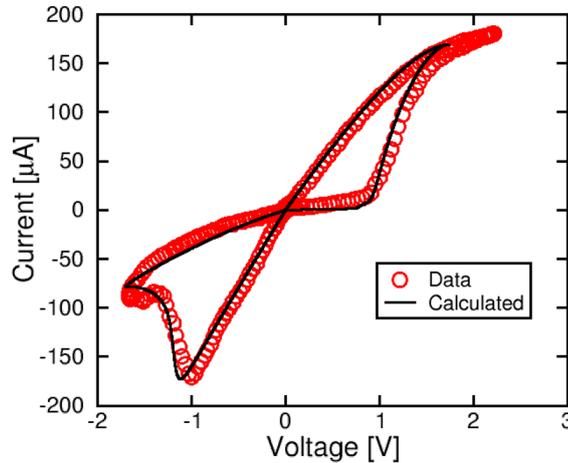


Figura 2.13: Caratteristica I-V sperimentale, in rosso, e relativa simulazione, in nero, generata a partire dal modello di switching a gap variabile.

Si considerino ora le caratteristiche I-V rappresentate in Fig. 2.13 . Si osserva immediatamente che sia la curva sperimentale che quella simulata evidenziano una transizione di set brusca. Questo andamento può essere spiegato in base a un effetto di reazione positiva attivato termicamente.

In Fig. 2.14 viene rappresentata l'evoluzione temporale dei parametri fisici d'interesse per il processo di set. L'applicazione di una rampa di tensione positiva, rappresentata in Fig. 2.14 (a) in rosso, determina un incremento della temperatura, ma finchè $T < T_{crit}$, il processo di migrazione ionica non parte. Ciò significa che il diametro ϕ del CF nel gap, (b), resta nullo e la resistenza della cella si mantiene elevata. All'aumentare della tensione V_A , la resistenza della cella, (c), inizia progressivamente a ridursi a causa della non linearità dovuta alla crescita del campo elettrico all'interno dell'ossido che riduce la resistività di quest'ultimo. Questo effetto spiega la diminuzione non lineare della resistenza che di conseguenza determina la crescita, anch'essa non lineare, della corrente, (d). All'istante $t \approx 0.3$ s, la temperatura raggiunge il valore di soglia $T_{crit} \approx 800$ K attivando la reazione positiva (positive feedback, PF) che governa la crescita del CF. Si osserva infatti una brusca crescita del diametro del CF e quindi della corrente che, a causa del meccanismo di compliance realizzato dal transistor in serie, si blocca a $I_C = 1$ mA. In seguito all'arresto della corrente a I_C si ha che l'ulteriore crescita del diametro del filamento sarà accomodata da un abbassamento della tensione ai capi della cella. Questa riduzione innesca a sua volta una reazione negativa

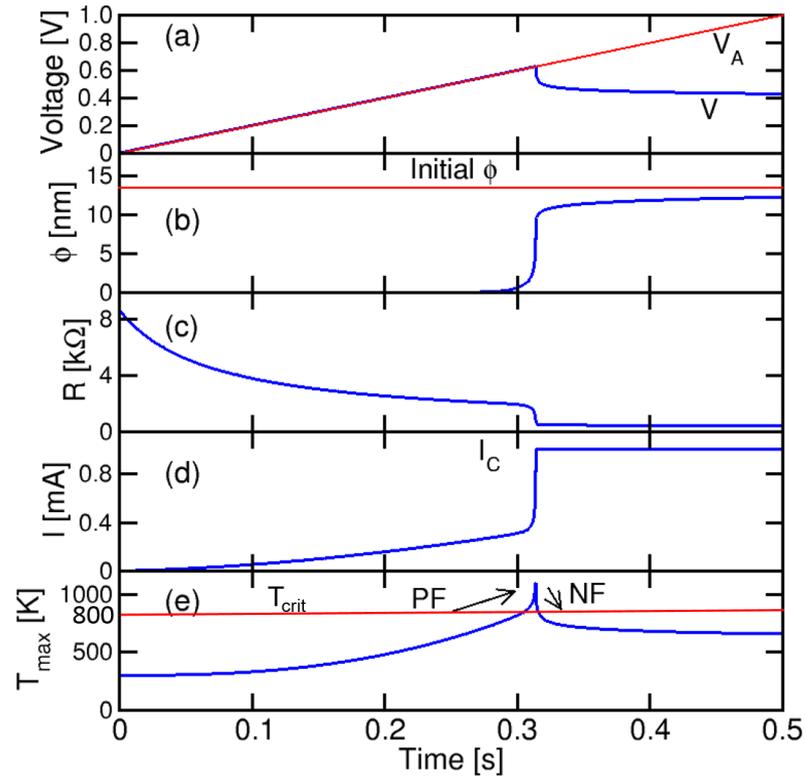


Figura 2.14: Andamenti calcolati di V , ϕ , R , I e T_{max} in funzione del tempo. Inizialmente, all'aumentare della tensione applicata, (a), non accade nulla nel CF, (b), in quanto la temperatura raggiunta da quest'ultimo, (e), è ancora troppo bassa. La resistenza, (c), diminuisce in modo non lineare a causa del campo elettrico presente nel gap. Raggiunta la soglia $T_{crit} = 800$ K, il diametro del CF ϕ subisce una crescita molto brusca a causa dell'attivazione di una reazione positiva. Quando I eguaglia I_C si verifica una riduzione della tensione ai capi della cella con seguente stabilizzazione del diametro del CF.

(negative feedback, NF) che porta rapidamente alla stabilizzazione di ϕ a circa 12 nm.

Capitolo 3

Modellistica dell'overshoot capacitivo

Questo capitolo è dedicato alla descrizione dell'impatto della capacità parassita sul funzionamento della struttura 1T1R e, in particolare, sulle principali caratteristiche del processo di set della cella resistiva quali la tensione di set e la resistenza dello stato basso resistivo.

3.1 Overshoot capacitivo

Si consideri la seguente struttura riportata in Fig. 3.1. Tale struttura ripropone la tipica configurazione 1T1R con l'aggiunta di una capacità C_{par} che ingloba tutti i termini capacitivi parassiti che insistono sul nodo intermedio.

Per comprendere l'impatto di C_{par} sul funzionamento dell'architettura 1T1R occorre rivisitare il processo di set del dispositivo.

A questo proposito si assuma di applicare al top electrode (TE) della struttura una tensione la cui forma d'onda è rappresentata in Fig. 3.2. Come già discusso nel capitolo precedente, la transizione di set ha inizio quando $V_A = V_{set}$, ovvero quando la temperatura del punto d'iniezione dei difetti ionizzati supera la temperatura critica di attivazione della migrazione ionica verso il BE.

Per $V_A > V_{set}$, si registra una brusca crescita della corrente e di conseguenza del diametro del filamento conduttivo (conductive filament, CF) finchè la corrente non eguaglia I_C , ovvero la massima corrente erogabile dal transistor MOS.

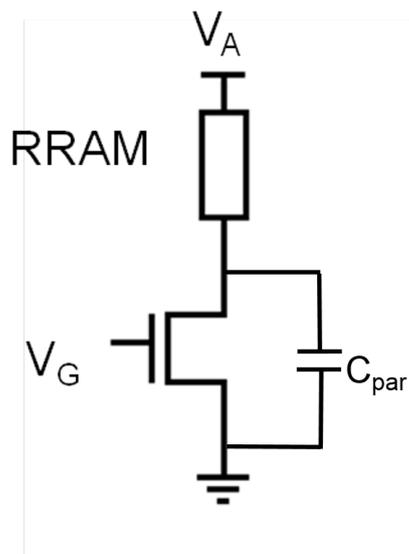


Figura 3.1: Struttura 1T1R con una capacità parassita che insiste sul nodo di drain del MOS.

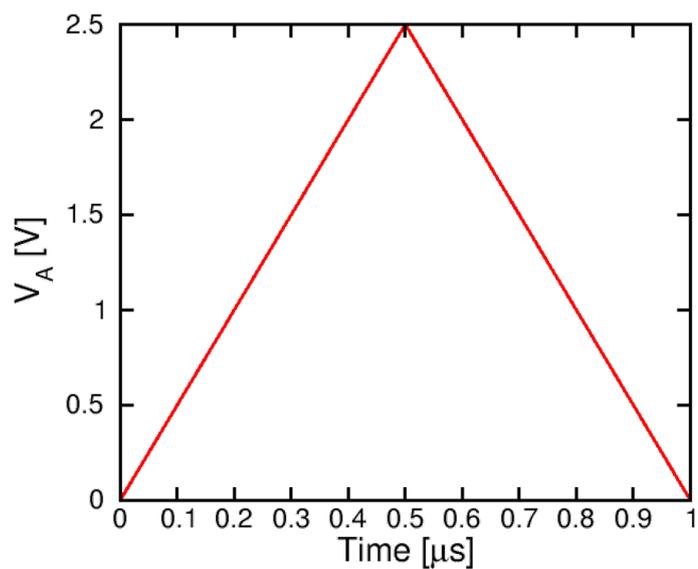


Figura 3.2: Andamento nel tempo della tensione applicata al TE.

Dato che la corrente non può superare I_C , l'ulteriore crescita del CF sostenuta dalla rampa positiva di tensione applicata al TE, con conseguente riduzione di resistenza della cella, dovrebbe essere immediatamente accomodata da una diminuzione della tensione V ai capi

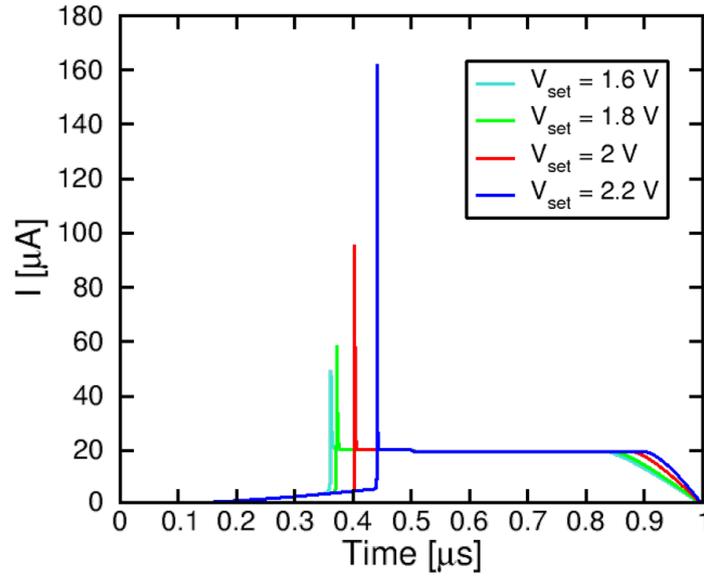


Figura 3.3: Rappresentazione dell'andamento della corrente della RRAM durante il processo di set per differenti valori di V_{set} . All'aumentare della tensione di set si osservano spike di corrente sempre più alti che descrivono il peso crescente della capacità nel funzionamento della struttura 1T1R. Si è assunto $I_C = 20\mu A$ e $C_{par} = 50$ fF.

della cella, ma la capacità parassita non lo permette. A causa del ritardo con cui C_{par} inizia a scaricarsi, in questo intervallo di tempo, la tensione V anziché diminuire resta bloccata, causando il passaggio nella cella resistiva di una corrente superiore a I_C . Ciò comporta una crescita aggiuntiva di ϕ_{CF} e quindi un'ulteriore riduzione della resistenza. Tuttavia, con la scarica di C_{par} , V inizia progressivamente a diminuire, bloccando definitivamente la crescita del CF.

A conferma di questa descrizione è mostrato l'andamento nel tempo della corrente che attraversa la cella resistiva al variare della tensione di set.

In Fig. 3.3 si osserva che in corrispondenza della transizione di set, mostrata per differenti valori di V_{set} , la corrente subisce una brusca variazione caratterizzata da uno spike di ampiezza crescente che rappresenta il contributo di corrente legato alla capacità parassita. In particolare, si nota che se da un lato per $V_{set} = 1.6$ V l'ampiezza dello spike di corrente risulta non molto elevata e quindi la struttura 1T1R opera in regime controllato dalla corrente di compliance, dall'altro, per valori di V_{set} crescenti, si assiste invece ad una repentina crescita dell'ampiezza dello spike di corrente, che porta la struttura in un regime di funzio-

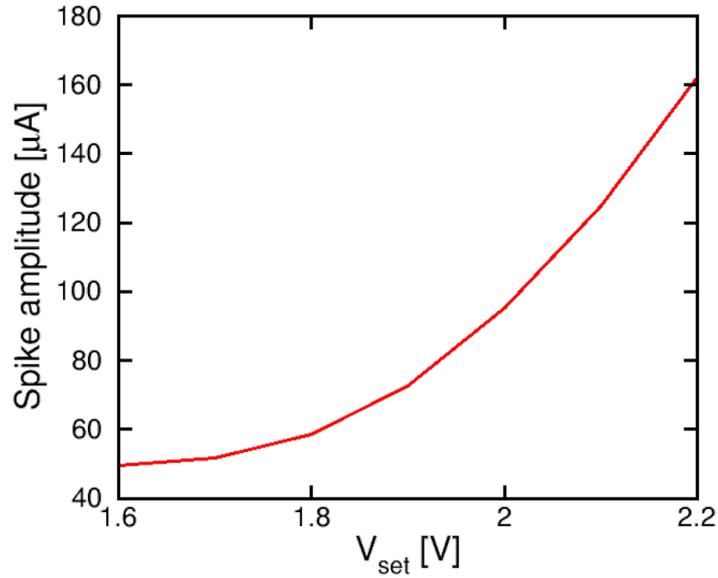


Figura 3.4: Andamento dell'ampiezza dello spike di corrente legato alla capacità parassita al variare di V_{set} . Si osserva una crescita molto repentina.

namento differente controllato da C_{par} . A questo proposito, in Fig. 3.4, viene rappresentato l'andamento dell'ampiezza dello spike di corrente al variare di V_{set} , dal quale si evince che per tensioni di set progressivamente più alte si ottiene un brusco incremento della corrente che attraversa la cella durante il set.

In definitiva, l'effetto di overshoot capacitivo, a causa della brusca crescita della corrente al di sopra di I_C durante il set, comporta la diminuzione della resistenza di set della cella ad un valore $R_{fin} < R_{LRS}$ dipendente dalla cinetica del processo di crescita del CF, dal valore di C_{par} e dalla tensione applicata.

A partire dal modello della memoria resistiva a switching bipolare è possibile determinare una formula analitica in grado di descrivere R_{LRS} in funzione di C_{par} e V_{set} .

Si consideri l'equazione che regola il tasso di crescita del filamento conduttivo durante il set:

$$\frac{d\phi}{dt} = Ae^{-\frac{E_A}{k(T_0 + \frac{V^2}{8k_{th}\rho})}} \quad (3.1)$$

Essendo molto breve l'intervallo in cui si verifica la transizione di set, se integriamo nel

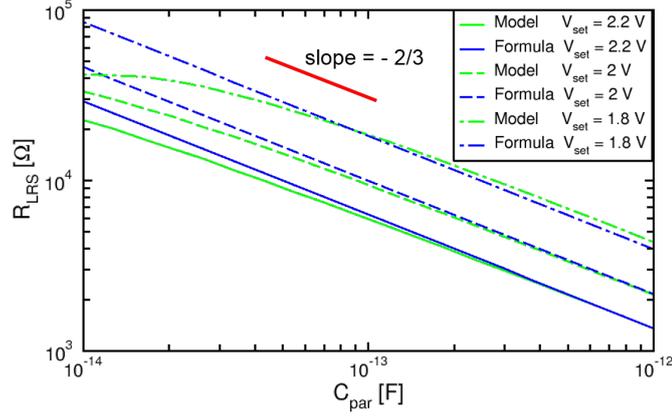


Figura 3.5: Andamenti, ottenuti rispettivamente dal modello analitico e dalla formula, di R_{LRS} al variare di C_{par} , per $V_{set} = 1.8$ V, 2 V e 2.2 V. Si osserva che solo per bassi valori di C_{par} l'accordo tra modello e formula viene a mancare. L'andamento delle curve ricavate dal modello non si discosta molto dalla power law descritta dalla formula, la quale prevede una pendenza pari a $-\frac{2}{3}$.

tempo tra 0 e t_{set} , V si mantiene circa costante e si ottiene:

$$\phi_{set} = At_{set} e^{-\frac{EA}{k(T_0 + \frac{V_{set}^2}{8k_{th}\rho})}} \quad (3.2)$$

Sostituendo in (3.2) $t_{set} = RC_{par}$, si ricava la seguente espressione:

$$\phi_{set} = ARC_{par} e^{-\frac{EA}{k(T_0 + \frac{V_{set}^2}{8k_{th}\rho})}} \quad (3.3)$$

Dato che $R = \frac{\rho L}{4\phi_{set}^2}$, la (3.3) restituisce la seguente espressione di ϕ_{set}

$$\phi_{set} = \left(\frac{4A\rho LC_{par}}{\pi}\right)^{\frac{1}{3}} e^{-\frac{EA}{3k(T_0 + \frac{V_{set}^2}{8k_{th}\rho})}} \quad (3.4)$$

Infine, sostituendo (3.4) in (3.5)

$$R_{LRS} = \frac{\rho L}{\frac{\pi}{4}\phi_{set}^2} \quad (3.5)$$

otteniamo l'espressione finale che lega R_{LRS} a C_{par} e V_{set}

$$R_{LRS} = \left(\frac{4\rho L}{\pi A^2}\right)^{\frac{1}{3}} C_{par}^{-\frac{2}{3}} e^{-\frac{2EA}{3k(T_0 + \frac{V_{set}^2}{8k_{th}\rho})}} \quad (3.6)$$

Nota l'equazione (3.6), è immediato ricavare V_{set} in funzione di R_{LRS} e C_{par}

$$V_{set} = \sqrt{\frac{\frac{2EA}{3k\vartheta}}{\ln\left(\frac{R_{LRS}C_{par}^{\frac{2}{3}}}{\left(\frac{4\rho L}{\pi A^2}\right)^{\frac{1}{3}}}\right)} - \frac{T_0}{\vartheta}} \quad (3.7)$$

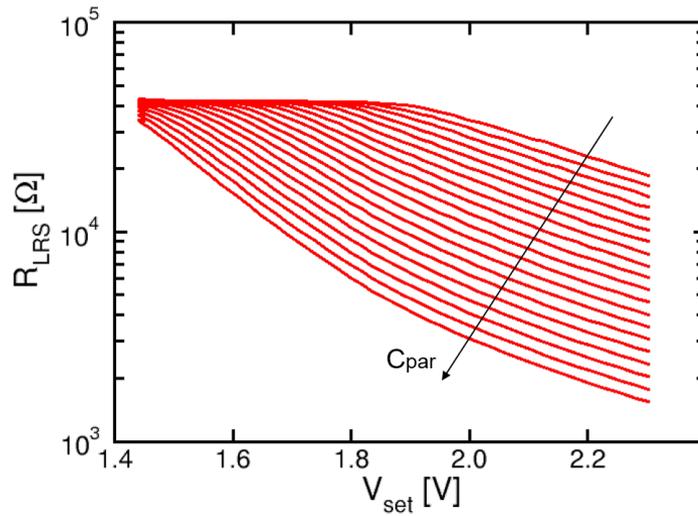


Figura 3.6: Andamento di R_{LRS} in funzione di V_{set} per C_{par} che varia tra 10 fF e 600 fF. All'aumentare della capacità parassita, la resistenza dello stato LRS tenderà ad assumere a valori progressivamente più bassi.

dove $\vartheta = \frac{1}{8\rho k_{th}}$.

In Fig. 3.5 viene mostrato un confronto su scala logaritmica, per tre differenti valori di V_{set} , tra l'andamento di R_{LRS} in funzione di C_{par} ricavato dal modello analitico e quello ottenuto invece mediante la formula. I profili rappresentati nei tre casi mettono in evidenza una power law con pendenza pari a $-\frac{2}{3}$ e si nota inoltre che l'accordo tra i risultati del modello e quelli della formula, ottimo per valori elevati di capacità, viene a mancare per bassi valori dato che si esce dal regime di funzionamento controllato da C_{par} .

La Fig. 3.6 mostra l'andamento della resistenza dello stato LRS, calcolata mediante il modello analitico, in funzione di V_{set} al crescere della C_{par} . Fissato un valore di tensione V_{set} , per valori crescenti di C_{par} si osserva una progressiva riduzione della resistenza di set legata proprio all'effetto di overshoot capacitivo.

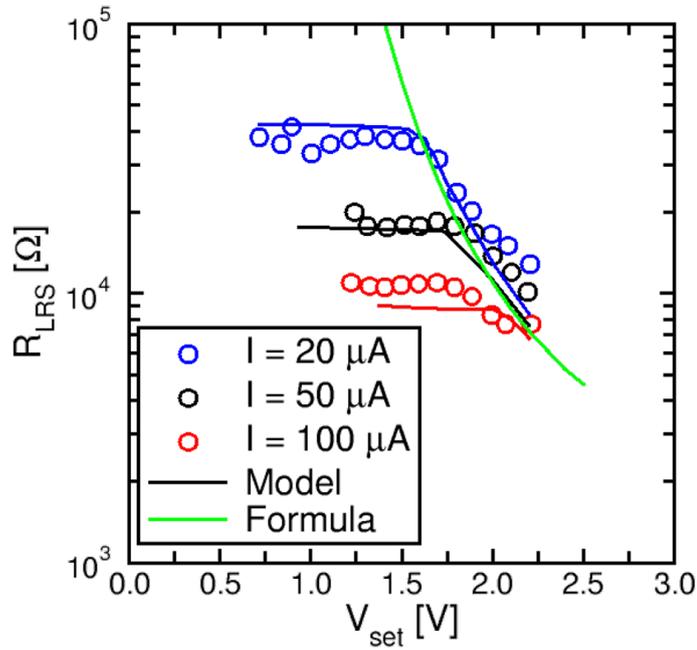


Figura 3.7: Confronto tra le misure sperimentali e gli andamenti simulati di R_{LRS} in funzione di V_{set} per tre differenti valori di I_C . Inoltre viene riportato il profilo di R_{LRS} ottenuto mediante la formula analitica, il quale, per $V_{set} < 1.5$ V, si discosta dai dati e dal modello poiché non si è più nel regime di funzionamento controllato dalla capacità parassita.

In Fig. 3.7 sono riportati i dati sperimentali, con le relative simulazioni ricavate dal modello analitico, della resistenza associata allo stato LRS al variare di V_{set} per $I_C = 20\mu A$, $50\mu A$ e $100\mu A$. Inoltre si osserva l'andamento ricavato dalla formula, il quale è in grado di interpolare molto bene le misure finché non si esce dal regime controllato dalla capacità parassita.

Infine, in Fig. 3.8 sono mostrati i profili di V_{set} in funzione di C_{par} , ottenuti dal modello analitico e dalla formula per $R_{LRS} = 10$ k Ω , 20 k Ω e 30 k Ω . L'ottimo accordo che si evince nei tre casi tra modello analitico e formula conferma quindi la correttezza della descrizione attribuita al fenomeno di overshoot capacitivo.

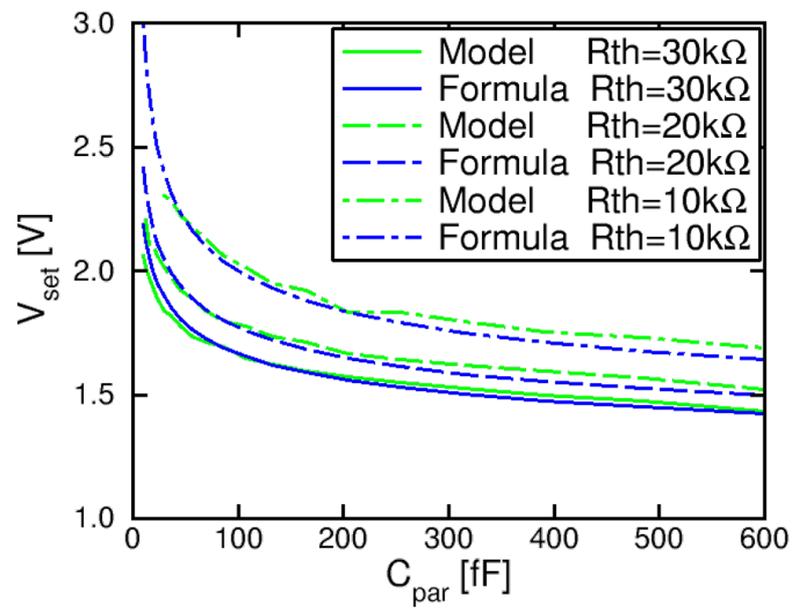


Figura 3.8: Confronto tra l'andamento di V_{set} ottenuto al variare di C_{par} , dal modello e quello ricavato dalla formula (3.7) per tre valori fissati di R_{LRS} . Si osserva un ottimo accordo tra i risultati del modello e quelli ottenuti dalla formula analitica.

Capitolo 4

Modello Simulink della struttura 1T1R

Questo capitolo è dedicato all'implementazione di un modello circuitale in grado di simulare il funzionamento della struttura 1T1R. Il circuito, progettato in ambiente Simulink, sarà prima analizzato e poi impiegato in un modello più complesso con l'obiettivo di simulare una rete neuromorfica e in particolare la plasticità delle sinapsi reali.

4.1 Implementazione del circuito 1T1R

Lo schema riportato in Fig. 4.1 mostra l'implementazione circuitale di una RRAM in configurazione 1T1R realizzata mediante il simulatore Simulink.

A livello di layout l'architettura 1T1R si ottiene collegando un transistor NMOS, avente il source connesso a massa, in serie ad un resistore variabile il cui valore di resistenza è generato attraverso un blocco Matlab Function che implementa il modello analitico della cella RRAM descritto nel secondo capitolo [77].

Questo blocco, denominato Modello RRAM, genera i valori dei principali parametri fisici che regolano il funzionamento della cella, i quali, memorizzati mediante un blocco Memory, saranno utilizzati, insieme al valore successivo della tensione V ai capi del resistore e del passo temporale, come nuovi ingressi per calcolare il valore che assumerà la resistenza all'istante successivo. In questo schema circuitale la resistenza viene quindi calcolata e aggiornata in modo iterativo sfruttando un anello reazionato.

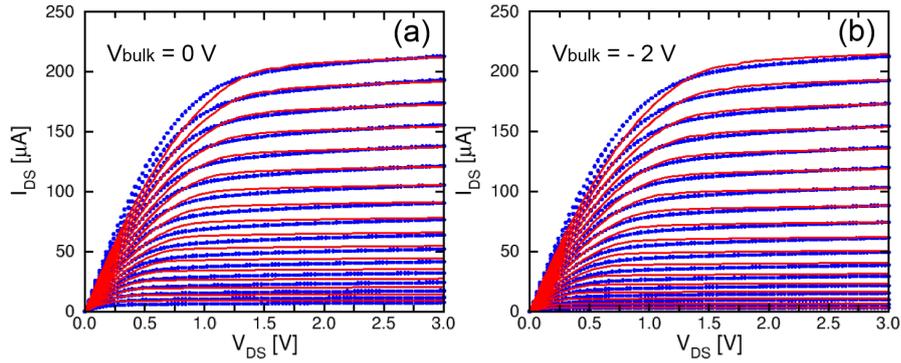


Figura 4.2: Curve I_{DS} - V_{DS} misurate (in blu) con le corrispondenti simulazioni (in rosso) per $V_{bulk} = 0$ V, (a), e $V_{bulk} = -2$ V, (b).

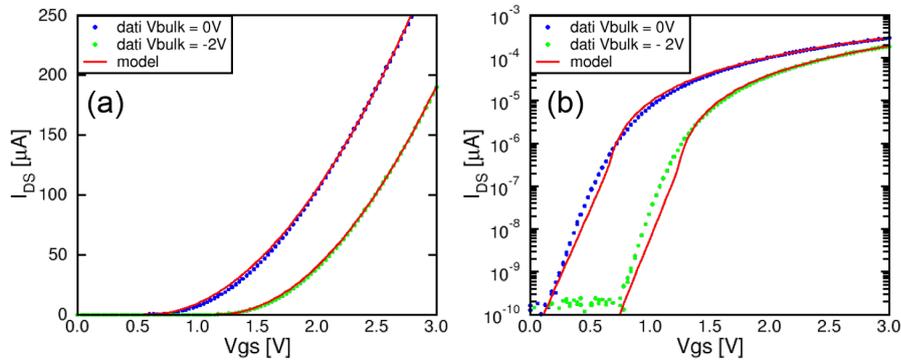


Figura 4.3: Rappresentazione su scala lineare, (a), e semilogaritmica, (b), delle curve I_{DS} - V_{GS} misurate/calcolate per $V_{bulk} = 0$ V e $V_{bulk} = -2$ V.

4.2 Analisi del circuito simulato

Conclusa l'operazione di connessione degli elementi circuitali si procede con la scelta delle impostazioni della simulazione. Tra queste le più importanti sono la durata e i parametri del solver. A tal proposito, in Fig. 4.4 è mostrata l'interfaccia grafica offerta dal simulatore mediante la quale è possibile fissare questi parametri.

Dopo aver stabilito l'istante iniziale e finale della simulazione, il passo successivo consiste nella scelta delle caratteristiche del solver. Il solver è il metodo matematico impiegato dal simulatore per risolvere numericamente le equazioni differenziali che regolano il funzionamento dei blocchi appartenenti al circuito implementato. I metodi di risoluzione del circuito possono essere a passo fisso o a passo variabile. Una volta stabilita la categoria, si può

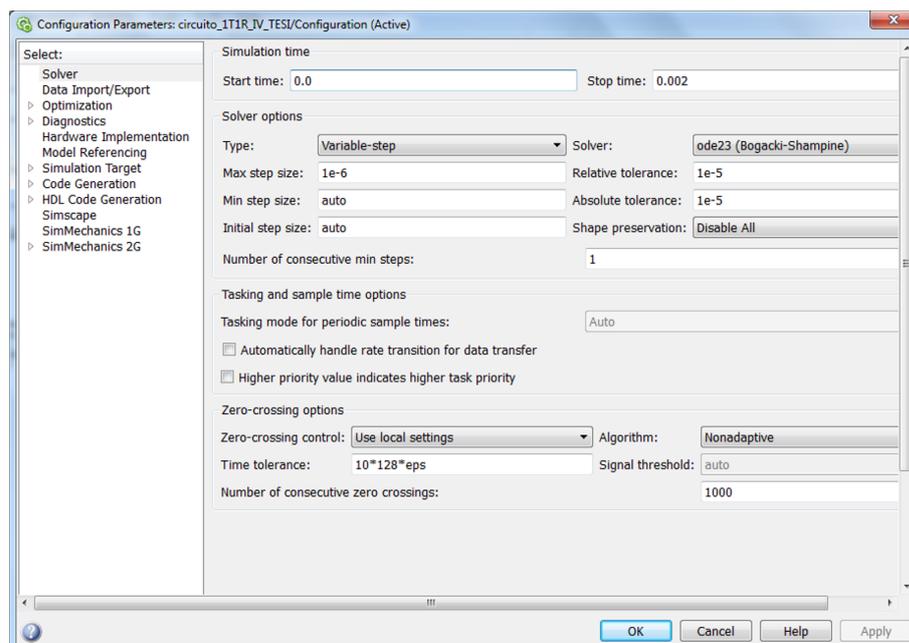


Figura 4.4: Interfaccia grafica Simulink attraverso cui è possibile fissare i parametri della simulazione circuitale.

quindi scegliere, a seconda delle caratteristiche del circuito, il particolare solver da utilizzare e i valori dei parametri ad esso associati come il massimo passo temporale e i valori di tolleranza relativa e assoluta. Per quanto concerne la simulazione discussa in questa sezione la scelta è ricaduta sul metodo a step variabile `ode23` poichè consente di realizzare una simulazione caratterizzata da un passo d'integrazione adattativo e quindi più veloce rispetto ad una avente passo costante. Questa simulazione è stata effettuata applicando al top electrode della struttura un impulso triangolare negativo di tensione seguito da un impulso avente la medesima forma, ma di polarità opposta. Come mostrato in Fig. 4.5 i due impulsi, aventi la medesima durata di 1 ms, hanno ampiezze di picco rispettivamente pari a -1.7 V e 2.2 V.

Al gate dell'NMOS, come riportato in Fig. 4.6, è stata invece applicata una tensione costante pari a 5 V fino a 1 ms, istante in corrispondenza del quale commuta a 2.94 V. La polarizzazione del transistor è stata poi ultimata fissando il potenziale del substrato a -2 V.

L'applicazione di queste tensioni si traduce nell'attivazione di una transizione di reset seguita da una di set. La prima conferma arriva dalla caratteristica corrente-tensione tracciata dal simulatore e mostrata in Fig. 4.7.

Inizialmente si osserva che la rampa negativa di tensione applicata al TE induce una variazione lineare della corrente. Questo accade perchè la resistenza della cella, corrispon-

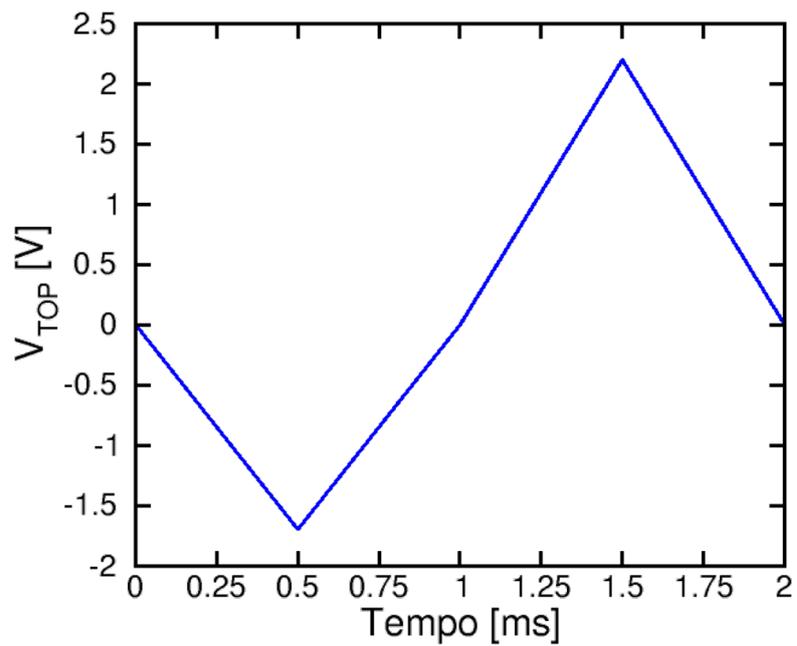


Figura 4.5: Andamento nel tempo della tensione V_{top} applicata al TE della struttura durante la simulazione. Il segnale consiste nella successione di due impulsi triangolari aventi ampiezza di picco rispettivamente pari a -1.7 V e 2.2 V. L'applicazione di questo segnale attiva una sequenza reset-set della cella resistiva.

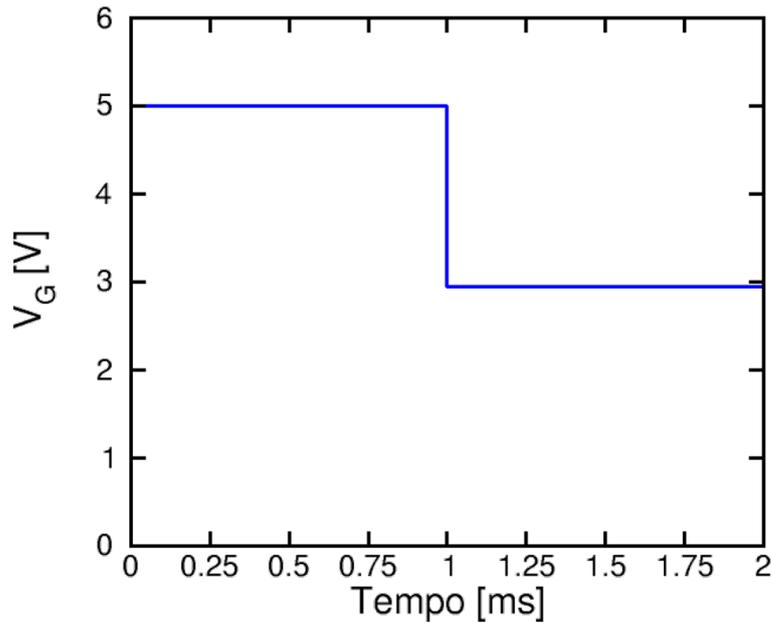


Figura 4.6: Profilo temporale della tensione V_G applicata al gate del transistor MOS. Fino a metà simulazione, la tensione di gate è uguale a 5 V per minimizzare la resistenza del MOS. All'istante $t = 1$ ms inizia la fase di set e la tensione di gate commuta a 2.94 V, valore che sarà mantenuto fino alla fine della simulazione.

dente al reciproco della pendenza della curva I-V, si mantiene costante finché la tensione applicata non raggiunge il valore $V_{reset} \approx -1.1$ V. Non appena la rampa negativa di tensione eguaglia V_{reset} s'innesca il processo di reset e la resistenza della cella inizia gradualmente ad aumentare sino a raggiungere il valore massimo quando $V_{top} = V_{stop} = -1.7$ V. Il dispositivo giunge quindi in uno stato ad alta resistenza (HRS) e vi rimane fino all'esaurimento dell'impulso triangolare negativo. Applicando poi una rampa di tensione positiva si nota che la corrente, dopo un'iniziale crescita lineare dovuta alla resistenza costante della cella, subisce una brusca variazione. Questo cambio di andamento si verifica quando la tensione applicata raggiunge il valore necessario ad attivare la transizione di set, in questo caso circa pari a 1.1 V. In seguito a questo rapido incremento della corrente, dovuto alla natura autoaccelerata del processo di set, la pendenza della curva aumenta notevolmente e quindi di conseguenza si ha una forte riduzione della resistenza della cella. Il processo di set si conclude perciò con il raggiungimento da parte del dispositivo RRAM di uno stato a bassa resistenza (LRS) che rimarrà immutato fino al termine dell'impulso triangolare positivo.

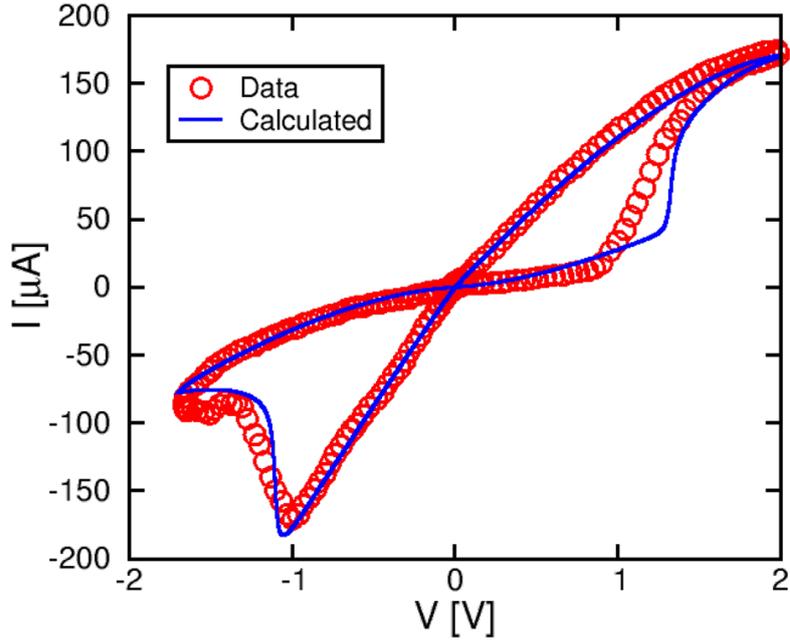


Figura 4.7: Caratteristica I-V sperimentale della struttura 1T1R, in rosso, e relativa simulazione, in blu, generata mediante il modello Simulink. Si può constatare un buon accordo tra la curva simulata e i dati sperimentali.

Analizziamo ora gli andamenti temporali dei principali parametri che caratterizzano la cella 1T1R partendo dalla corrente.

Nelle prime fasi, come mostrato in Fig. 4.8, la corrente segue l'andamento lineare di V_{top} in quanto la resistenza della cella si mantiene costante. Giunti all'istante $t \approx 0.3$ ms, la tensione applicata al TE raggiunge il valore di soglia V_{reset} , attivando così la transizione di reset. Fisicamente il processo di reset comporta l'apertura di un gap svuotato di ioni metallici all'interno del filamento conduttivo iniziale. La formazione di un gap sempre più esteso causa una progressiva riduzione della temperatura e di conseguenza della corrente in quanto diventa sempre più debole la forza che sostiene la migrazione ionica. Con la riduzione della corrente si ha un graduale incremento della resistenza che porta la cella in uno stato HRS. Terminato l'impulso negativo al TE, la corrente inizia ad aumentare finché all'istante $t \approx 1.3$ ms si nota un brusco incremento. Questo particolare andamento è dovuto alla crescita autoaccelerata dello spessore del CF, sostenuta dal campo elettrico e dalla temperatura, che ha inizio all'istante $t \approx 1.3$ ms, ovvero, come mostrato in Fig. 4.5, quando la tensione al TE raggiunge la soglia $V_{set} = 1.1$ V. La crescita della corrente prosegue fino

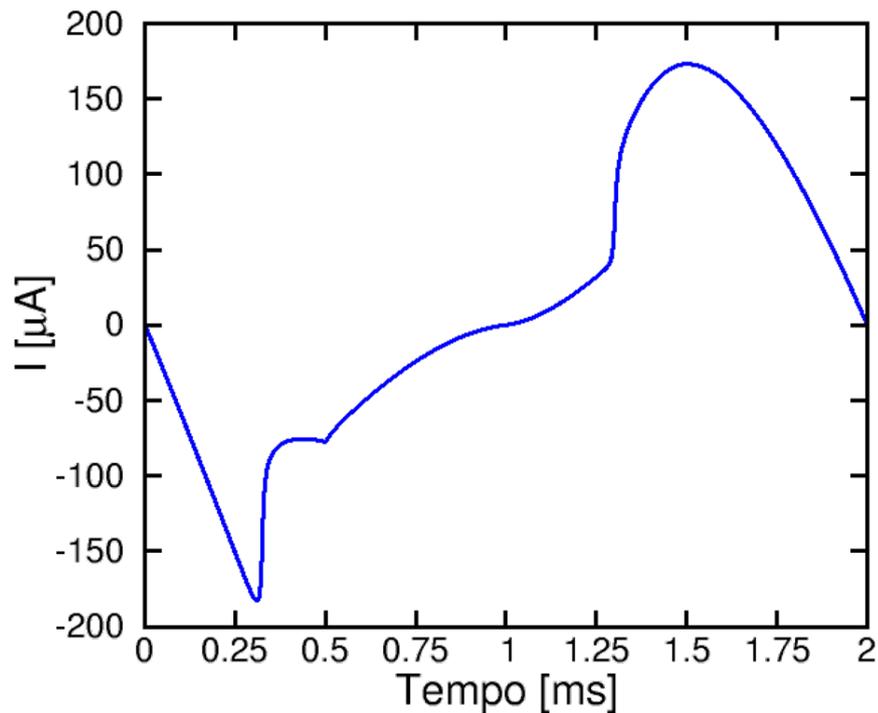


Figura 4.8: Andamento nel tempo della corrente che attraversa la cella. Durante il reset, a causa dell'incremento di resistenza, la corrente I simulata diminuisce sino ad annullarsi per $V_{top} = 0$ V. In seguito all'attivazione della transizione di set, I aumenta molto rapidamente fino a raggiungere la corrente di compliance I_C , ovvero il valore per cui la crescita si arresta a causa della limitazione imposta dal transistorore in serie.

a $I_C = 170 \mu\text{A}$, ovvero la massima corrente erogabile dal transistorore. Eguagliata I_C , la corrente resta inizialmente circa costante per poi diminuire a causa della rampa di tensione con cui si esaurisce il segnale di TE.

Si consideri adesso l'andamento della tensione V rappresentato in Fig. 4.9. La tensione che si sviluppa ai capi della resistenza è data dalla partizione di V_{top} tra la cella resistiva e il transistorore. Dato che durante il periodo di reset $V_G = 5$ V, il transistorore mostra un'elevata conducibilità e di conseguenza la sua resistenza è molto bassa. Questo significa che quasi tutta la tensione V_{top} cade ai capi del dispositivo e quindi nella prima metà della simulazione l'andamento di V non si discosterà molto da quello della tensione di TE. Durante l'impulso di set ai capi della cella cade una tensione che cresce linearmente nel tempo finchè, in corrispondenza dell'istante di attivazione del set, essa inizia a diminuire. Questo

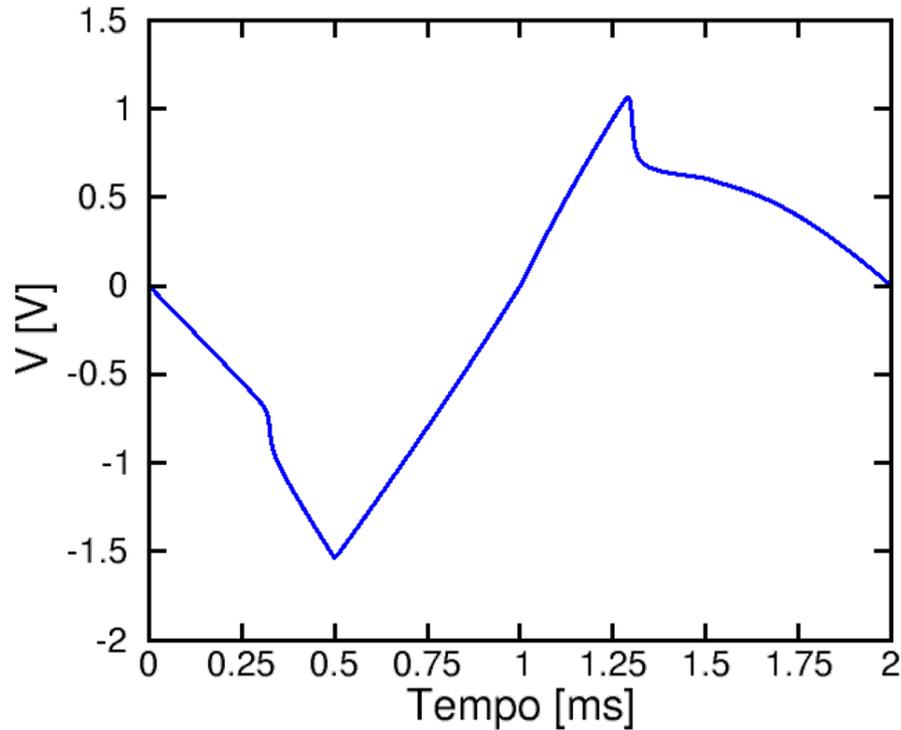


Figura 4.9: Evoluzione temporale della tensione che cade ai capi della cella. Dato che durante il reset R_{MOS} è trascurabile rispetto a R , nella prima parte di simulazione V segue l'andamento di V_{top} . Con l'attivazione della successiva transizione di set, la tensione V subisce una riduzione che tenderà a rallentare la crescita del filamento conduttivo fino ad arrestarla, ponendo così fine al processo di set.

comportamento, denominato snapback, si verifica in quanto la riduzione di resistenza, legata alla crescita del CF, avviene a corrente limitata dal transistor. Raggiunta I_C , la corrente che attraversa la cella si blocca e la diminuzione di resistenza porta di conseguenza ad un abbassamento della tensione ai capi del dispositivo che poi si annulla con la fine dell'impulso di TE.

Dopo aver esaminato i profili nel tempo della tensione V ai capi della cella e della corrente è importante visualizzare l'andamento della resistenza del dispositivo. In Fig. 4.10 si osserva che la resistenza della cella inizia ad aumentare gradualmente dall'istante $t \approx 0.3$ ms, ovvero quando la tensione applicata raggiunge il valore necessario per attivare il processo di migrazione degli ioni verso il TE che porta alla formazione di un gap nel CF. La crescita di

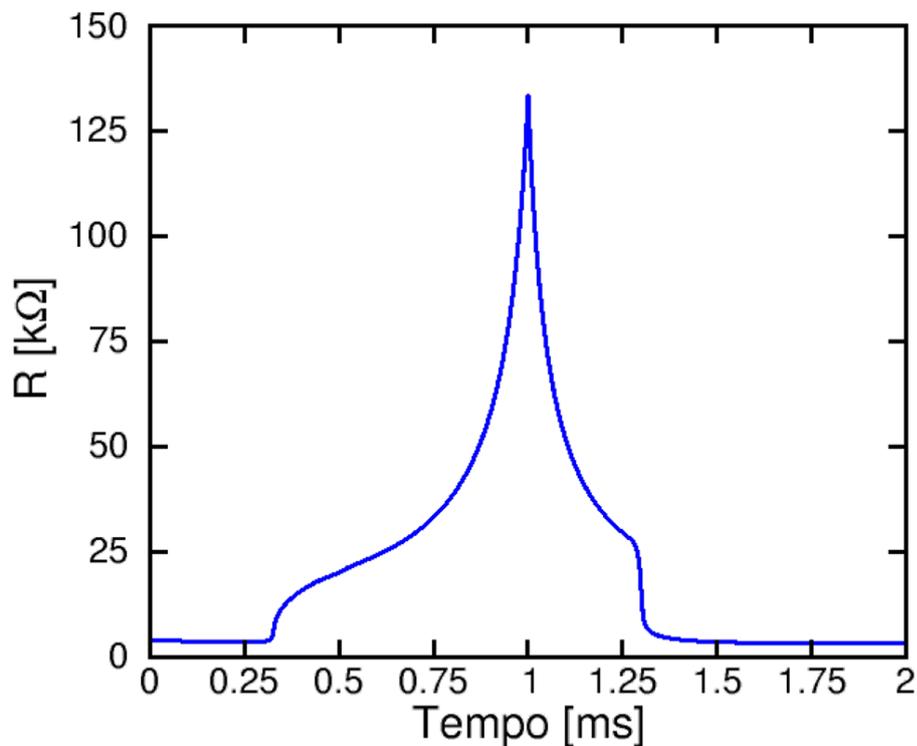


Figura 4.10: Simulazione dell'andamento nel tempo della resistenza del dispositivo.

resistenza si conclude con il raggiungimento da parte della cella resistiva di uno stato ad alta resistenza circa pari a 140 $k\Omega$. Conclusa la transizione di reset, la resistenza del dispositivo subisce una riduzione non lineare, dovuta alla diminuzione della resistività indotta dal campo elettrico nel gap, finchè, giunti all'istante $t \approx 1.3$ ms, ha inizio il processo di set. In questo istante, la resistenza subisce una brusca diminuzione da 30 $k\Omega$ a circa 6 $k\Omega$ causata dalla repentina crescita del diametro del CF sostenuta dal campo elettrico e accelerata dalla temperatura. Il processo di set termina pertanto con il raggiungimento da parte della cella di uno stato a bassa resistenza (~ 4 $k\Omega$) che resterà immutato fino alla conclusione della simulazione.

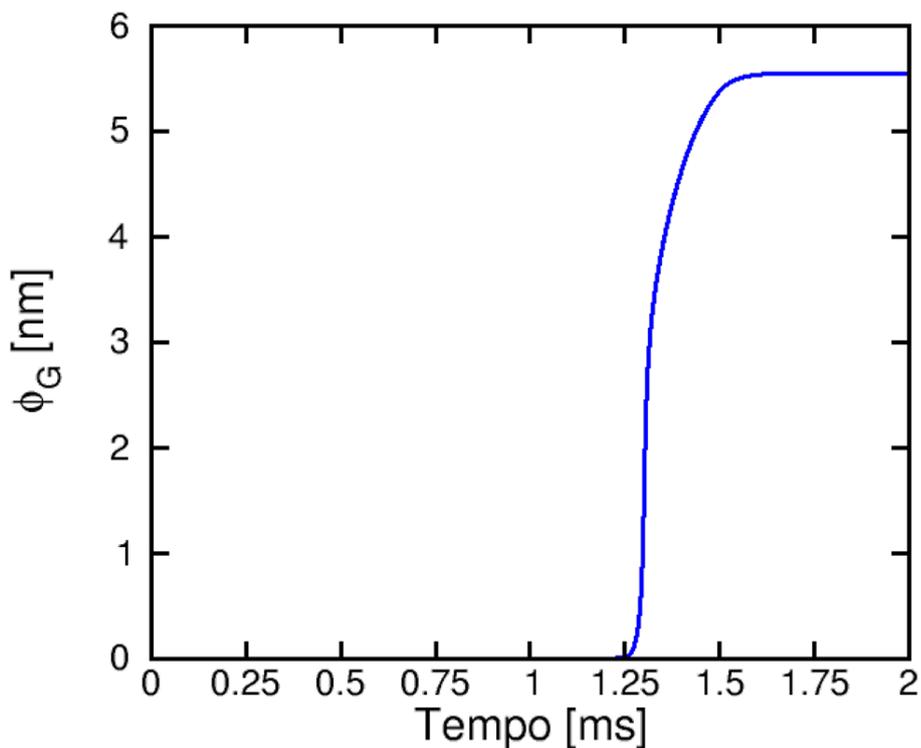


Figura 4.11: Rappresentazione della crescita repentina del diametro del sub-filamento conduttivo ϕ_G che si forma nel gap durante il processo di set.

L'ultimo parametro fisico interessante da analizzare è il diametro del filamento conduttivo ϕ_G che si forma all'interno del gap durante il set. Come riportato in Fig. 4.11, durante il reset il diametro ϕ_G è ovviamente nullo in quanto si sta formando il gap svuotato di ioni metallici. Concluso il reset, anche all'inizio della fase di set resta trascurabile finché, superato l'istante di attivazione della transizione di set, inizia un'improvvisa crescita, limitata soltanto dalla corrente I_C , fino a circa 5.55 nm. Questo significa che alla fine della fase di set il filamento conduttivo è nuovamente continuo.

In definitiva, i risultati ottenuti mediante la simulazione realizzata con Simulink sono consistenti sia con i risultati ottenuti dal modello analitico [77], descritti nel capitolo 2, che con i dati sperimentali, come mostrato in Fig. 4.7.

4.3 Implementazione in Simulink della plasticità

Lo schema di riferimento di una semplice rete neuronale artificiale è basato sulla connessione tra i neuroni del layer d'ingresso e un neurone d'uscita mediante sinapsi elettroniche a conduttanza variabile.

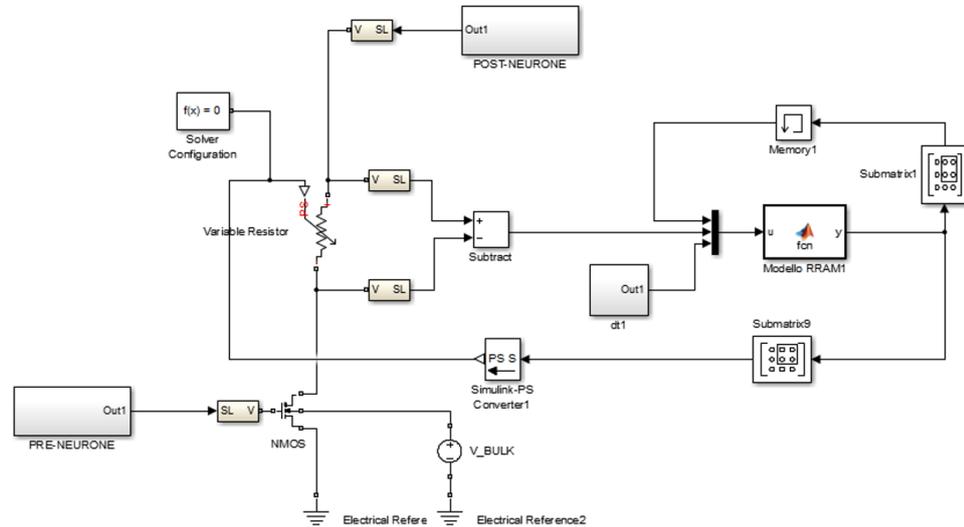


Figura 4.12: Rappresentazione compatta dello schema circuitale impiegato per simulare la plasticità.

Per semplicità, nel circuito riportato in Fig. 4.12 è rappresentato un unico blocco pre-neurone connesso a una cella 1T1R, identica a quella descritta in Fig. 4.1, che funge da sinapsi. Si osserva inoltre la presenza di un altro blocco, collegato al TE della cella, avente la funzione di generare la forma d'onda del tipico impulso di fire del post-neurone.

Analizziamo pertanto la struttura di questi due sottosistemi cominciando da quello che genera l'impulso pre-sinaptico. La struttura mostrata in Fig. 4.13 presenta 3 generatori di impulsi rettangolari e due blocchi sommatori mediante i quali viene realizzata la sovrapposizione dei tre segnali.

L'effettivo impulso di gate è generato dal blocco Pulse Generator, mentre i restanti due generatori hanno la funzione di implementare due impulsi di lettura necessari per calcolare successivamente la variazione relativa di conduttanza della sinapsi. A questo proposito, in Figs. 4.14 e 4.15 sono riportate le impostazioni scelte per i tre generatori e il segnale risultante applicato al gate del transistor.

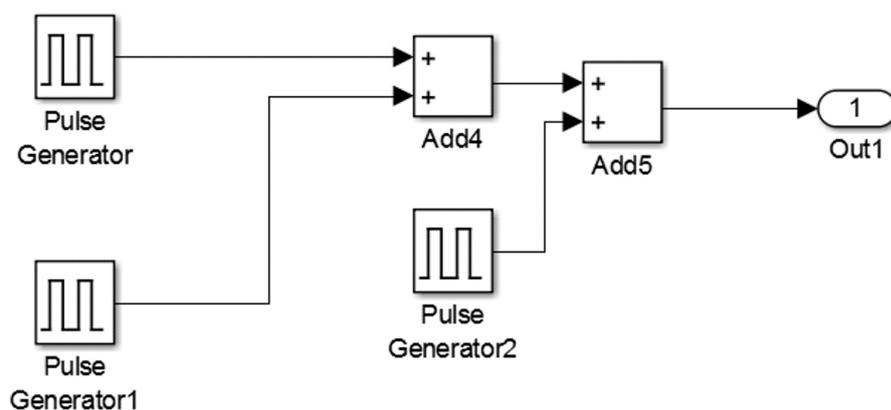


Figura 4.13: Schema circuitale con cui è stato generato l'impulso applicato al gate della struttura 1T1R.

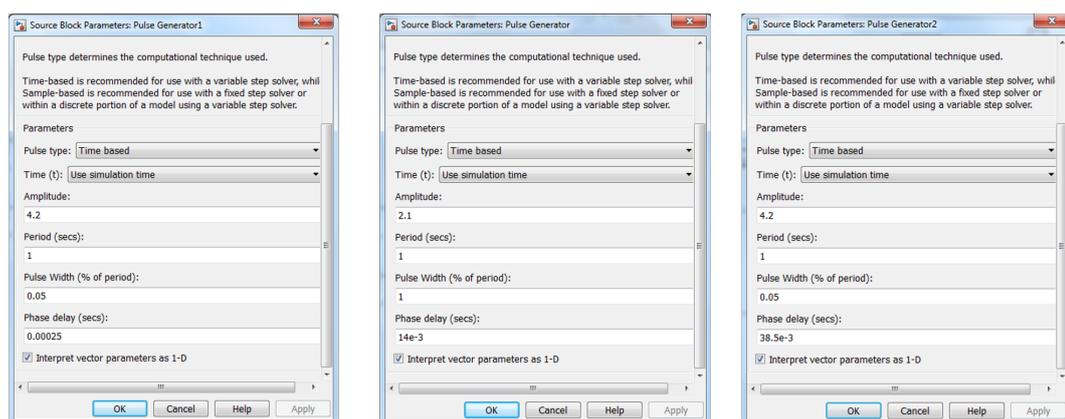


Figura 4.14: Interfacce grafiche delle impostazioni associate ai tre generatori impulsati utilizzati per ottenere la forma d'onda del segnale di gate.

In Fig 4.16 è mostrato lo schema a blocchi con cui è stato generato l'impulso di TE. Gli elementi principali sono i blocchi condizionali in quanto, mediante opportune condizioni sugli istanti di tempo, consentono di costruire l'impulso che sarà poi applicato al TE della cella. La forma d'onda risultante da questo schema circuitale è rappresentata in Fig 4.17. Si tratta di un segnale che prevede un impulso positivo di ampiezza 2.5 V e durata 1 ms seguito, dopo 9 ms, da un altro impulso rettangolare della medesima durata, ma di ampiezza

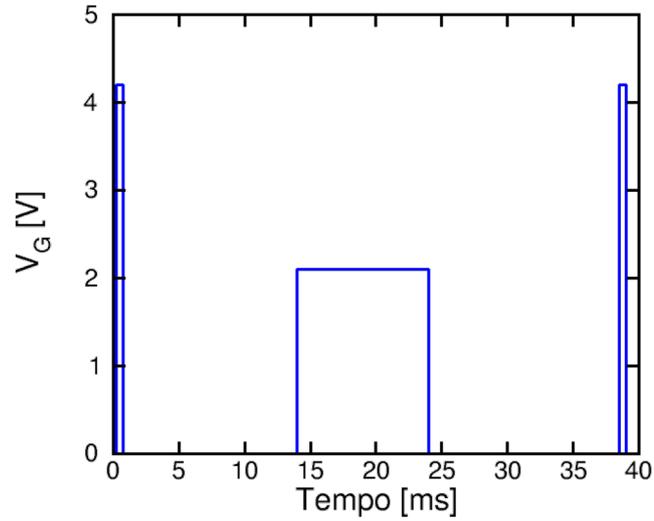


Figura 4.15: Forma d'onda della tensione V_G .

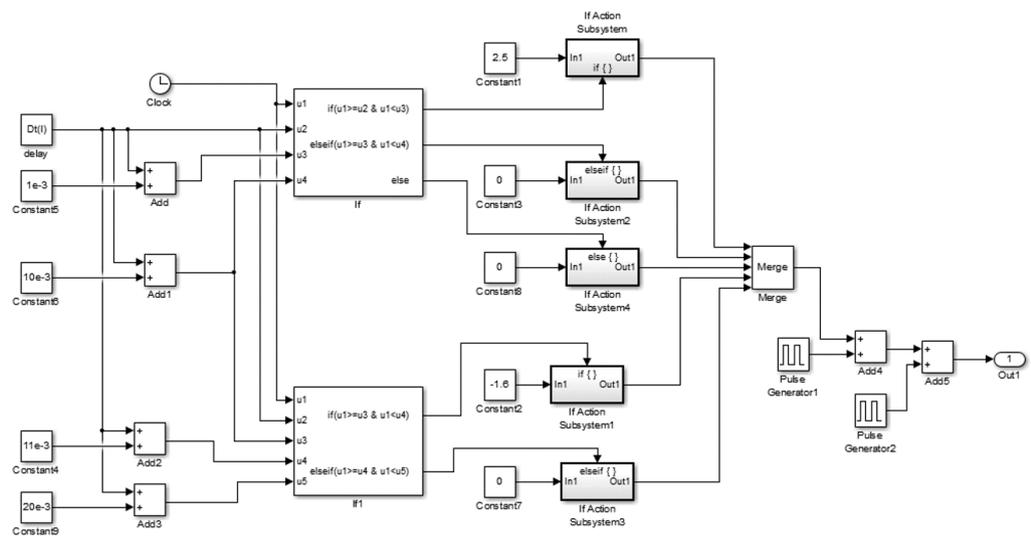


Figura 4.16: Rappresentazione dello schema a blocchi impiegato per implementare la forma d'onda dell'impulso generato da un post-neurone durante un evento di fire.

pari a -1.6 V.

Nel circuito di Fig 4.16 sono inoltre presenti due generatori di impulsi che, come nel caso della struttura del pre-neurone, generano i due impulsi di lettura necessari per calcolare i valori di conduttanza della cella all'inizio e alla fine della simulazione. Infine si nota un

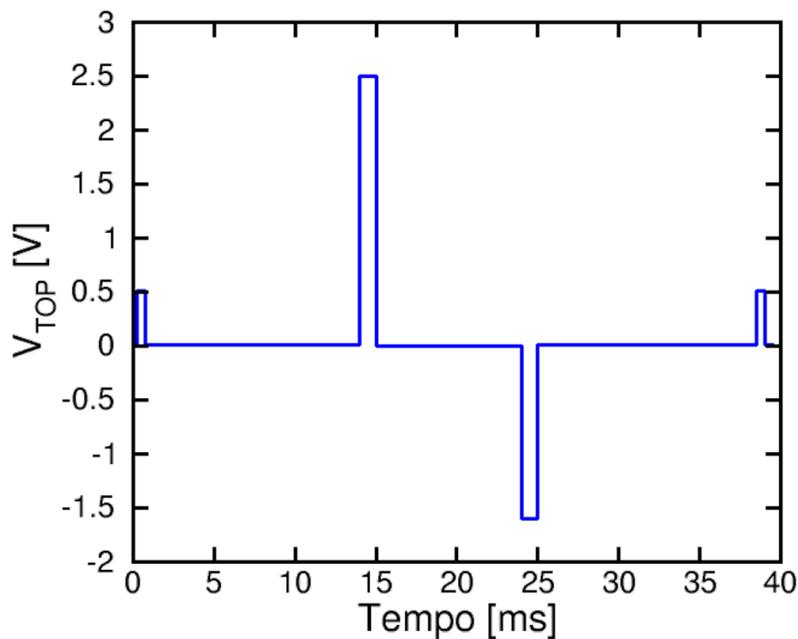


Figura 4.17: Rappresentazione dell'impulso di tensione applicato al top electrode della struttura. A questo segnale sarà applicato un ritardo variabile Δt in modo da simulare l'andamento della variazione relativa di conduttanza della sinapsi in funzione dello stato iniziale e del ritardo stesso.

blocco Constant denominato delay a cui viene assegnato il valore del ritardo temporale da applicare al segnale di TE per calcolare le curve STDP.

In Fig 4.18 è riportata una curva STDP calcolata al variare di Δt partendo da uno stato iniziale $R_0 = 125 \text{ k}\Omega$. Questa curva descrive la variazione relativa di conduttanza della sinapsi al variare del ritardo Δt tra l'inizio dell'impulso generato dal pre-neurone e l'inizio dell'impulso generato dal post-neurone. Si osserva che per $\Delta t < 0$ si ha la depressione della sinapsi, mentre per $\Delta t \geq 0$ si verifica invece il potenziamento del peso sinaptico.

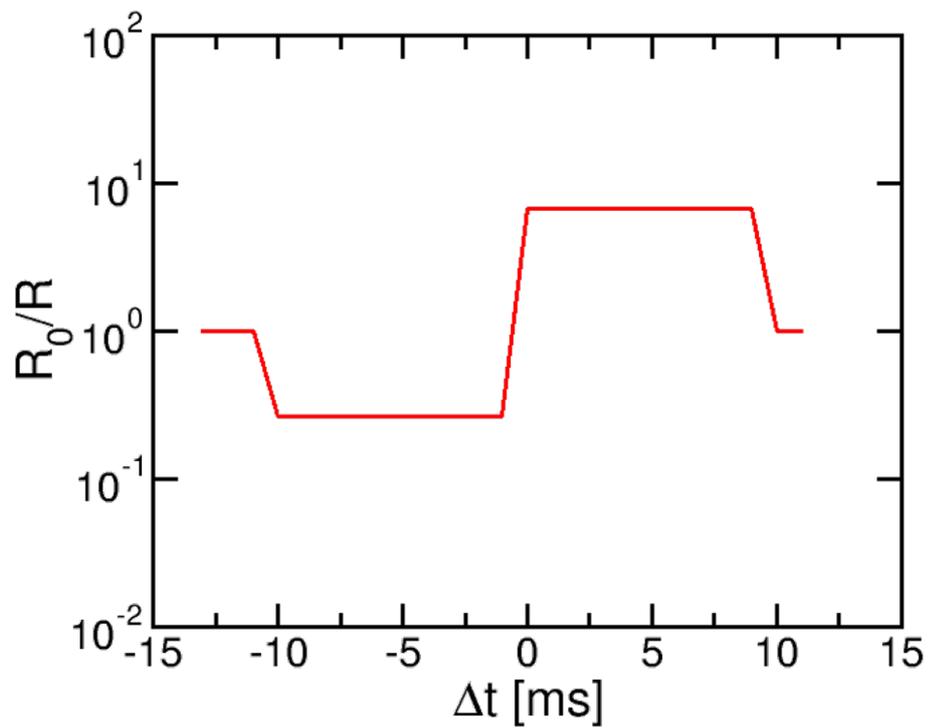


Figura 4.18: Esempio di curva STDP simulata attraverso il modello Simulink. Per $\Delta t < 0$ si nota una diminuzione del peso sinaptico, mentre per $\Delta t \geq 0$ si ha un incremento della conduttanza della sinapsi.

Capitolo 5

Apprendimento di pattern mediante sinapsi RRAM a switching stocastico

Questo capitolo è dedicato alla presentazione di una nuova architettura di sinapsi denominata 1T1R (1 transistore/1 resistore) e al confronto tra lo switching deterministico e quello stocastico. Si dimostrerà che lo switching stocastico, sfruttando la variabilità intrinseca del dispositivo RRAM, può essere molto vantaggioso per implementare l'apprendimento di pattern.

5.1 Sinapsi 1T1R

Come riportato in Fig. 5.1 (a), la struttura 1T1R può essere impiegata come sinapsi artificiale. Si nota che l'impulso generato dal PRE-neurone viene applicato al gate del MOS, mentre la tensione V_{TE} è controllata dal POST-neurone e assume tipicamente valori positivi non molto elevati (0.2 V). Inoltre, ogni impulso generato dal PRE-neurone attiva una corrente che è inversamente proporzionale alla resistenza del circuito 1T1R.

La durata t_P dell'impulso di gate adottato negli esperimenti è 1 ms, mentre la corrente di compliance, regolata opportunamente mediante la tensione V_G , è pari a $50 \mu\text{A}$. Inoltre è stata utilizzata una tensione $V_{stop} = -1.5 \text{ V}$.

La corrente che fluisce nella sinapsi 1T1R viene raccolta sul nodo di terra virtuale del POST-neurone e sommata alle correnti delle altre sinapsi. Quando la corrente integrata

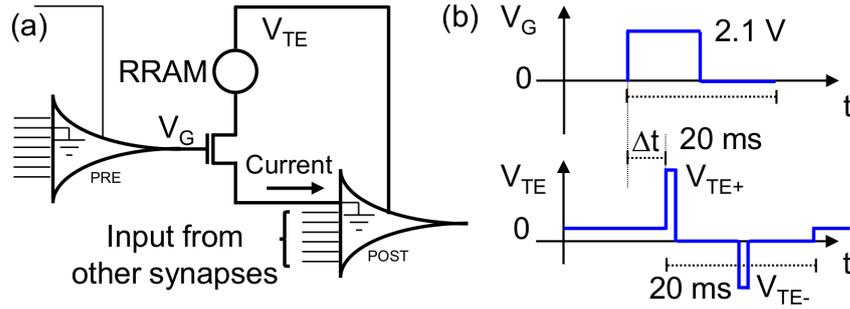


Figura 5.1: Rappresentazione schematica della sinapsi 1T1R che connette il PRE-neurone al POST-neurone, (a), e dei classici impulsi V_G e V_{TE} applicati rispettivamente al gate e al TE, (b). Un impulso di tensione V_G induce il passaggio di una corrente che verrà poi integrata dal POST-neurone, con la possibilità di attivare la generazione di un impulso di fire. Nell'istante in cui si verifica un evento di fire, la tensione V_{TE} può indurre il potenziamento della sinapsi ($\Delta t \geq 0$) o la depressione della stessa ($\Delta t < 0$), eseguendo in questo modo il protocollo STDP.

totale supera una soglia che rappresenta il potenziale di membrana del POST-neurone, quest'ultimo genera un impulso in accordo allo schema integrazione seguita da sparo (integrate and fire) che lo caratterizza [78]. Oltre a sparare un impulso verso i layer successivi, il POST-neurone genera un impulso, avente la forma indicata in Fig. 5.1 (b), anche verso il top electrode (TE) della sinapsi 1T1R. L'andamento della tensione applicata al TE prevede un primo impulso positivo di durata 1 ms seguito, dopo 9 ms, da un impulso negativo, avente la medesima durata del precedente, che, a sua volta, è seguito da un altro periodo di 9 ms in cui la tensione di TE si mantiene nulla. Negli esperimenti effettuati, l'impulso V_G applicato al gate del transistor prevede un primo periodo di 10 ms in cui la tensione è pari a 2.1 V seguito da un altro intervallo sempre di 10 ms in cui la tensione è 0 V. L'ampiezza dell'impulso è scelta in modo da ottenere $I_C = 50 \mu A$, la quale risulta abbastanza bassa da garantire un consumo di potenza ridotto nelle transizioni di set e reset. In fase di comunicazione la tensione di TE è mantenuta costante a 0.2 V, mentre le ampiezze degli impulsi durante il periodo di fire sono rispettivamente $V_{TE+} = +2.5$ V e $V_{TE-} = -1.6$ V. Nel caso di STDP stocastica, queste due ampiezze vengono modulate in modo da regolare la probabilità delle transizioni di set e reset. I valori scelti per le tensioni V_{TE+} e V_{TE-} , molto più alti rispetto alla tensione di TE durante la fase di comunicazione ($V_{TE} < V_{set}$), consentono di attivare il protocollo STDP in base al valore del ritardo tra gli impulsi di gate e di TE. Il ritardo Δt

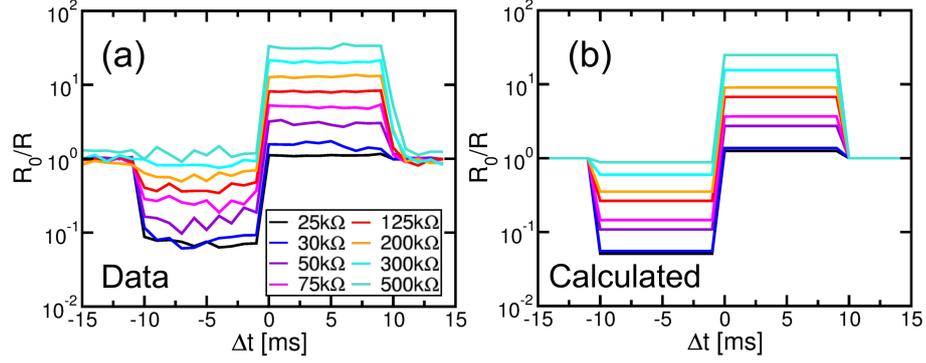


Figura 5.2: Rappresentazione delle caratteristiche STDP, ricavate dai dati, (a), e dalle simulazioni, (b), che descrivono la variazione di conduttanza R_0/R in funzione del ritardo Δt definito in Fig. 5.1 (b). Le simulazioni sono state generate attraverso un modello Simulink. La variazione di conduttanza è stata misurata/calcolata all'aumentare della resistenza iniziale R_0 .

è definito nel seguente modo

$$\Delta t = t_{post} - t_{pre} \quad (5.1)$$

dove, come mostrato in Fig. 5.1 (b), t_{pre} e t_{post} sono misurati in corrispondenza dell'istante iniziale degli impulsi del PRE-neurone e del POST-neurone e il segno di Δt stabilisce la variazione di resistenza della cella. Per $\Delta t \geq 0$, l'impulso positivo V_{TE+} è sovrapposto all'impulso di gate e quindi si attiva una transizione di set che porta al potenziamento sinaptico. Al contrario, per $\Delta t < 0$, è l'impulso negativo di ampiezza V_{TE-} che risulta sovrapposto a V_G , attivando la transizione di reset e quindi la depressione della sinapsi [74]. Si osserva inoltre che la tensione di 0.2 V, applicata al TE quando non si verifica alcun evento di fire, risulta nettamente minore rispetto a V_{set} , impedendo in questo modo ogni variazione di conduttanza del dispositivo.

Per testare la sinapsi 1T1R fin qui descritta, sono stati applicati gli impulsi V_G e V_{TE} , mostrati in Fig. 5.1 (b) alla cella 1T1R variando il ritardo relativo Δt e la resistenza iniziale R_0 con l'obiettivo di ottenere le curve STDP.

Per ogni coppia di tensioni di gate/TE applicate è stato misurato il nuovo valore di resistenza raggiunto dalla cella RRAM.

In Fig. 5.2 (a) viene rappresentato il rapporto R_0/R , ovvero la variazione di resistenza rispetto al valore iniziale R_0 in funzione del ritardo Δt tra gli impulsi. Le curve descrivono la variazione relativa di resistenza del dispositivo, che è stata inizializzata con un'operazione di partial reset condotta variando V_{stop} [35], e si osserva che il range della resistenza finale risulta limitato tra 25 k Ω e 500 k Ω . Inoltre le caratteristiche STDP mostrano potenziamento

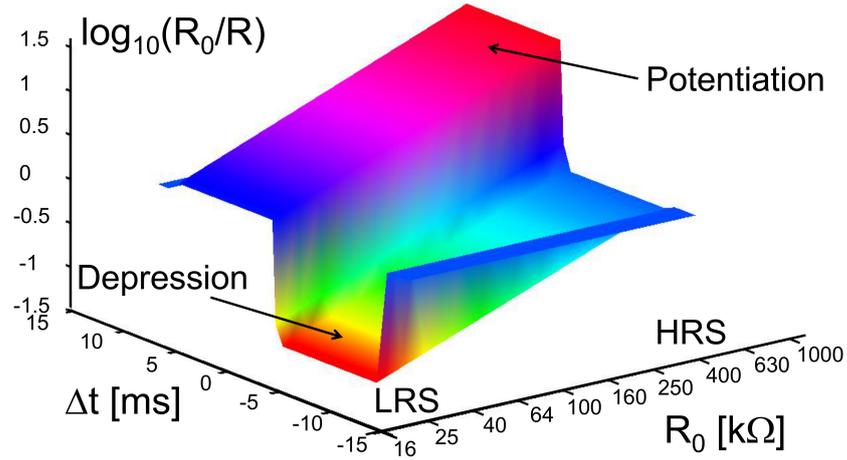


Figura 5.3: Rappresentazione delle curve STDP simulate come mappa 3D di R_0/R in funzione di Δt e R_0 . Si osserva che per alti (bassi) valori di resistenza iniziale R_0 si attiva il potenziamento (depressione) della sinapsi.

e depressione per ritardi Δt rispettivamente positivi e negativi. Come già detto in precedenza [74], la variazione dei pesi sinaptici dipende anche dallo stato iniziale di resistenza della cella e quindi nessun potenziamento è possibile se lo stato iniziale è uno stato LRS ($R_0 = 25 k\Omega$ in Fig. 5.2 (a)) in quanto la resistenza raggiunta con la transizione di set dipende dalla dimensione del CF che è controllata dalla corrente di compliance I_C . Poichè nello schema circuitale di Fig. 5.1 (b) è stata applicata una tensione V_G costante, non si osserva alcuna variazione del massimo diametro raggiunto dal filamento e di conseguenza non può verificarsi alcun potenziamento di uno stato LRS.

In modo analogo non è possibile alcuna depressione sinaptica se lo stato iniziale è uno stato HRS ($R_0 = 500 k\Omega$ in Fig. 5.2 (a)) poichè la tensione V_{stop} che controlla la resistenza dello stato HRS [79, 35] è tenuta costante nell'impulso di TE. Una simile dipendenza dallo stato iniziale è stata riscontrata anche nei sistemi biologici dove la conduttanza delle sinapsi non può superare i due valori estremi [80].

Al contrario, gli stati di resistenza intermedi possono raggiungere sia lo stato LRS che lo stato HRS.

In ogni caso, le caratteristiche STDP mostrano R_0/R costante sia per $\Delta t < 0$ che per $\Delta t \geq 0$, come conseguenza dell'andamento costante delle tensioni V_{TE+} , V_{TE-} e V_G . Si vedrà che questa semplificazione non limita la capacità di apprendimento, aggiornamento e riconoscimento di una rete neuromorfica.

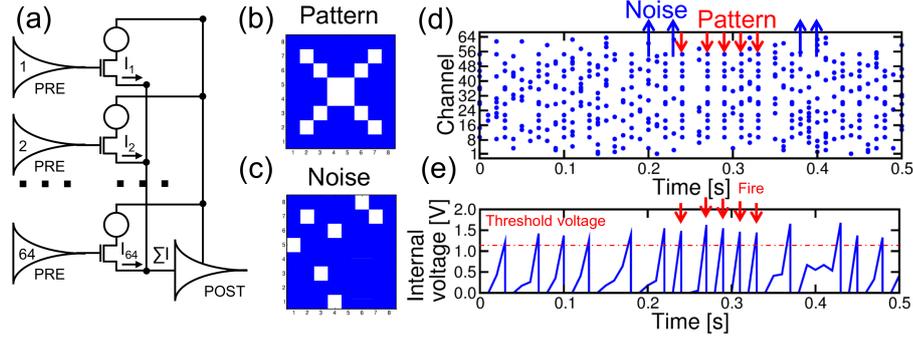


Figura 5.4: Schema circuitale di una rete neuromorfica a 2 layer adottata per dimostrare l'apprendimento di pattern, (a), pattern visivo d'ingresso, (b), rappresentazione AER (address-event representation) degli eventi di fire innescati dal pattern e dal rumore in ingresso, (d), e rappresentazione dell'andamento del potenziale interno del POST-neurone che genera impulsi di fire in seguito al raggiungimento della soglia interna, (e).

Le curve STDP sono state generate attraverso un modello circuitale Simulink in grado di simulare l'architettura 1T1R.

Nella struttura 1T1R simulata la cella resistiva è stata implementata a partire dal modello analitico [77]. Nelle simulazioni sono stati adottati gli stessi impulsi mostrati in Fig. 5.1 (b) e utilizzati in Fig. 5.2 (a), variando Δt e R_0 come in Fig. 5.2 (a).

In Fig. 5.2 sono riportate le caratteristiche STDP simulate e si nota un ottimo accordo con i dati. La Fig. 5.3 mostra sempre le curve STDP, ma come mappa di colore 3D in cui il $\log_{10}(R_0/R)$ riportato lungo l'asse z è rappresentato in funzione di R_0 (asse x) e Δt (asse y). Il potenziamento (long-term potentiation, LTP) si verifica per $\Delta t \geq 0$ e risulta più pronunciato al crescere di R_0 , mentre la depressione (long-term depression, LTD) si verifica per valori negativi del ritardo Δt e diventa più pronunciata per valori di R_0 sempre più bassi. Inoltre è importante osservare che la massima variazione di conduttanza durante il potenziamento è di circa un fattore 20, mentre la massima variazione di conduttanza durante la depressione è di circa un fattore 1/20. Questo significa che la finestra resistiva tra lo stato HRS ($\approx 500 \text{ k}\Omega$) e quello LRS ($\approx 25 \text{ k}\Omega$) è molto piccola.

La sinapsi, pertanto, assume un comportamento bistabile in quanto, a partire da un qualsiasi stato intermedio di resistenza, potrà raggiungere esclusivamente o lo stato HRS, in corrispondenza di un reset, o lo stato LRS, in corrispondenza di una transizione di set. Questo comportamento, che contrasta con quello analogico osservato nelle sinapsi reali [63, 81, 80], deriva dall'andamento brusco delle transizioni di set e di reset e dai valori abbastanza

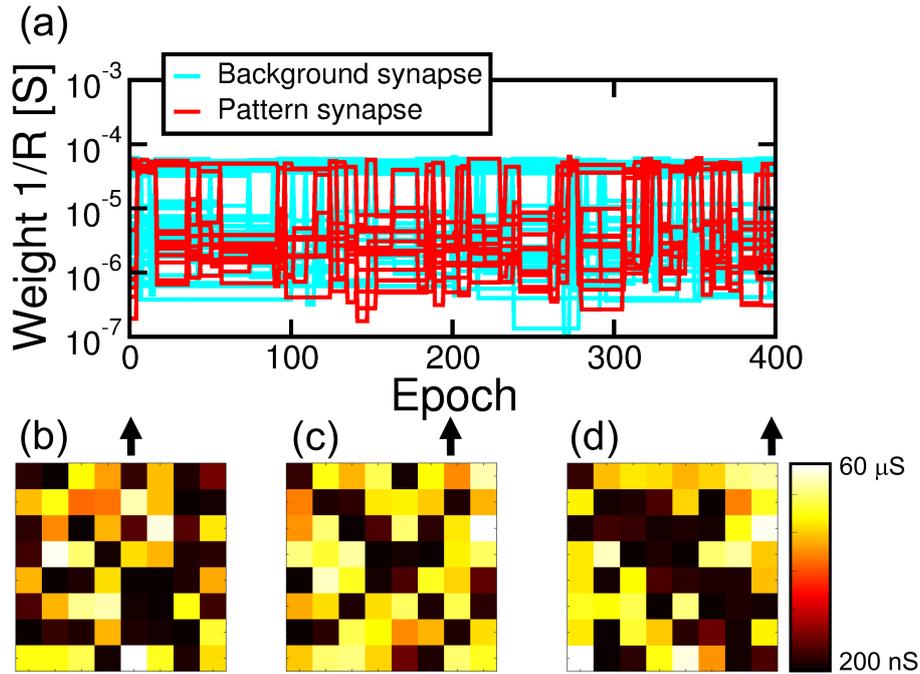


Figura 5.5: Simulazioni dell'andamento dei pesi sinaptici associati al pattern (in rosso) e al background (in blu) nel caso di sinapsi a 1 cella con $V_{TE-} = -1.6$ V e set deterministico. Le figure in alto mostrano le matrici 8x8 delle conduttanze delle sinapsi ad istanti fissati. Si osserva che l'apprendimento del pattern è molto inefficiente.

alti delle tensioni V_{TE+} e V_{TE-} .

5.2 Apprendimento di pattern con STDP deterministica

Per dimostrare la funzionalità della sinapsi 1T1R nell'apprendimento di pattern visivi è stata simulata una rete neuromorfica a 2 layer in cui 64 sinapsi connettono il primo layer, composto da 64 PRE-neuroni, ad un singolo POST-neurone [74]. Come rappresentato schematicamente in Fig. 5.4 (a), il primo layer funge da retina che genera impulsi in corrispondenza della presentazione in ingresso di un pattern visivo, che riproduce una X (Fig. 5.4 (b)), alternato a rumore casuale avente una densità media di 7 pixel illuminati (Fig. 5.4 (c)). Tutte le correnti sinaptiche attivate dagli impulsi d'ingresso vengono raccolte sul nodo di terra virtuale del POST-neurone, il quale è modellizzato come un integratore con scarica avente costante di tempo $\tau = 30$ ms (leaky & fire integrator, LIF).

Ciascuna epoca di apprendimento, durante la quale viene applicato in ingresso o il pattern o un rumore casuale, rispettivamente con probabilità del 20 % e dell'80 %, dura 10 ms.

In Fig. 5.4 (d) è mostrata l'evoluzione nel tempo degli ingressi per ognuno dei 64 canali del primo layer. Questo tipo di rappresentazione, che evidenzia chiaramente l'alternanza in ingresso tra pattern e rumore, è detta address-event representation (AER). La Fig. 5.4 (e) mostra l'andamento nel tempo del potenziale interno del POST-neurone, ovvero il potenziale d'uscita del leaky-integrator, che corrisponde al potenziale di membrana dei neuroni biologici [80]. All'aumentare della corrente integrata, il potenziale interno cresce finché, superata una certa soglia, il POST-neurone genera un impulso di fire e il suo potenziale si scarica. La Fig. 5.4 descrive il comportamento della rete neuromorfica una volta conclusa la fase di apprendimento, ovvero quando l'applicazione di un pattern X in ingresso induce in uscita un fire, rivelando così la capacità del neurone, ormai specializzato, di riconoscere il pattern proprio come accade nel cervello umano.

In Fig. 5.5 (a) è mostrata una simulazione dell'evoluzione nel tempo, indicato in funzione del numero di epoche (singolo evento temporale), delle conduttanze $1/R$ delle 64 sinapsi. Le linee rosse e azzurre corrispondono rispettivamente ai pesi delle sinapsi associate al pattern e al background. Inoltre, nelle Figs. 5.5 (b), (c) e (d), sono riportati come matrici 8×8 i pesi delle 64 sinapsi ad istanti fissati corrispondenti a 0, 200 e 400 epoche di apprendimento. Inizialmente, tutti i pesi sono stati distribuiti in modo uniforme tra lo stato HRS e quello LRS. I pesi mostrano un andamento oscillatorio poichè sia quelli associati al pattern che quelli del background commutano senza riuscire ad apprendere il pattern. Questo è dovuto al comportamento deterministico dello switching che induce transizioni alto-basso o basso-alto dei pesi sinaptici in risposta ad un qualsiasi impulso generato, che esso sia indotto dal pattern o dal rumore. Ne segue pertanto che il rumore, generalmente necessario per sostenere la depressione delle sinapsi associate al background, induce fire spuri che, al contrario, potenziano i pesi del background e deprimono quelli del pattern.

5.3 Caratterizzazione dello switching stocastico

Per ottenere l'apprendimento e il riconoscimento del pattern si è passati dallo switching deterministico a quello stocastico. La variabilità intrinseca che caratterizza la tensione V_{set} consente alle memorie resistive di commutare nello stato di set con una probabilità finita regolando opportunamente la tensione $V_A \simeq V_{set}$. Per eseguire l'operazione di random-set della cella RRAM è stata applicata una tensione V_{TE} la cui forma d'onda è rappresentata

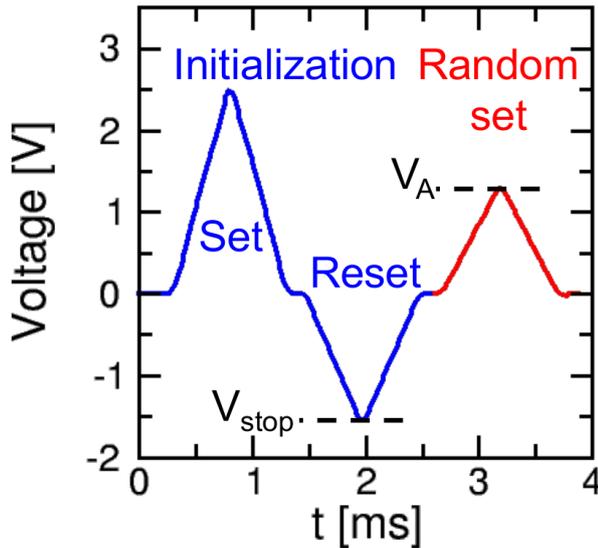


Figura 5.6: Sequenza degli impulsi di tensione applicati per eseguire gli esperimenti di random set mostrati in Fig. 5.7.

in Fig. 5.6. L'andamento di V_{TE} prevede l'inizializzazione del dispositivo attraverso l'applicazione di impulsi di set e reset, che preparano quindi il dispositivo in uno stato di reset, seguita dall'operazione di random-set, che attiva la transizione di set solo in un certo numero di cicli in quanto la tensione applicata è vicina a V_{set} [71]. In Fig. 5.7 (a) sono mostrate le distribuzioni di resistenza misurate. Terminata l'inizializzazione, la resistenza R è associata ad uno stato HRS completo, indicato dalla distribuzione log-normale di colore ciano. Poi, dopo il random-set, la resistenza commuta nello stato LRS con una certa probabilità che aumenta al crescere dell'ampiezza V_A . Ne segue pertanto che l'operazione di random-set genera una distribuzione di resistenza bimodale che include sia lo stato di set completo, associato alla curva I-V (A) di Fig. 5.7 (b), che lo stato HRS iniziale, associato alla curva I-V (C) di Fig. 5.7 (b) e quindi alla condizione di impossibilità di switching per $V_A < V_{set}$ (B). Infine si osservano degli stati intermedi (C) che sono raggiunti mediante transizioni di partial-set attivate per $V_A \approx V_{set}$. Come si nota dalla curva I-V (C) in Fig. 5.7 (b), in questi casi particolari il dispositivo commuta, ma la caduta di tensione ai suoi capi crolla immediatamente impedendo in questo modo la crescita di un filamento conduttivo corrispondente ad uno stato LRS completo. Per la precisione, durante l'operazione di random-set sono stati applicati impulsi di ampiezza crescente, ovvero $V_A = 1.05$ V, 1.3 V, 1.4 V e 1.6 V.

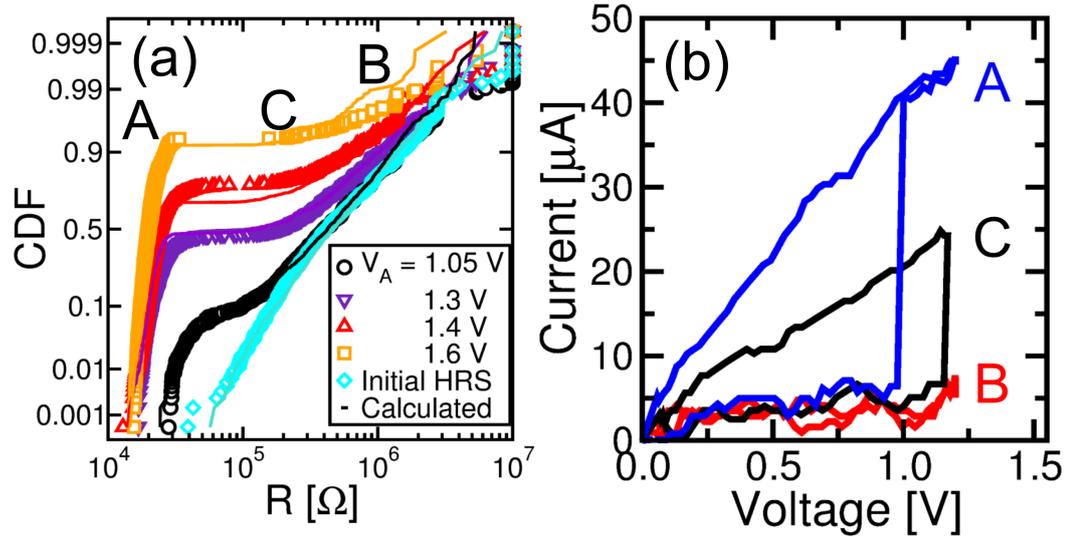


Figura 5.7: Distribuzioni cumulative della resistenza del dispositivo misurate e calcolate dopo l'operazione di random set eseguita all'aumentare di V_A , (a), e le tipiche caratteristiche I-V associate agli stati A, B e C, (b). A causa dello switching stocastico solo un certo numero di celle effettuano la transizione di set, (A), mentre le restanti o rimangono nello stato HRS, (B), o, in un numero minore di casi, effettuano una transizione di set parziale, (C).

Il processo di random-reset è stato attuato eseguendo una sequenza che prevede l'inizializzazione della cella RRAM, mediante un impulso di set completo, seguita dall'operazione di random-reset, che consiste nell'applicazione di una tensione negativa di ampiezza V_{stop} crescente in modulo. In Fig. 5.8 (a) sono mostrate le distribuzioni misurate in seguito al processo di random-reset ottenute applicando tensioni $V_{stop} = -0.7$ V, -1 V, -1.1 V, -1.2 V, -1.3 V, -1.6 V.

Per $V_{stop} = -0.7$ V, la resistenza del dispositivo non commuta e quindi mostra una distribuzione LRS. Successivamente, all'aumentare di V_{stop} , si ha un incremento della probabilità di switching e quindi si osserva una crescita della coda HRS. In Fig. 5.8 (b) sono rappresentati i valori mediani di resistenza misurati per i differenti valori di V_{stop} ed estratti dalle distribuzioni di Fig. 5.8 (a). Si osserva che da 20 k Ω , valore mediano corrispondente allo stato LRS completo, i valori ricavati tendono ad aumentare al crescere in modulo di V_{stop} . La curva calcolata è stata ottenuta utilizzando il modello Simulink descritto nel precedente capitolo e si nota un ottimo accordo con i dati sperimentali. La Fig. 5.8 (c) mostra l'andamento della deviazione standard log-normale ottenuto fittando proprio con distribuzioni log-normali le code HRS associate ad ognuno dei valori assunti da V_{stop} . La curva simulata è

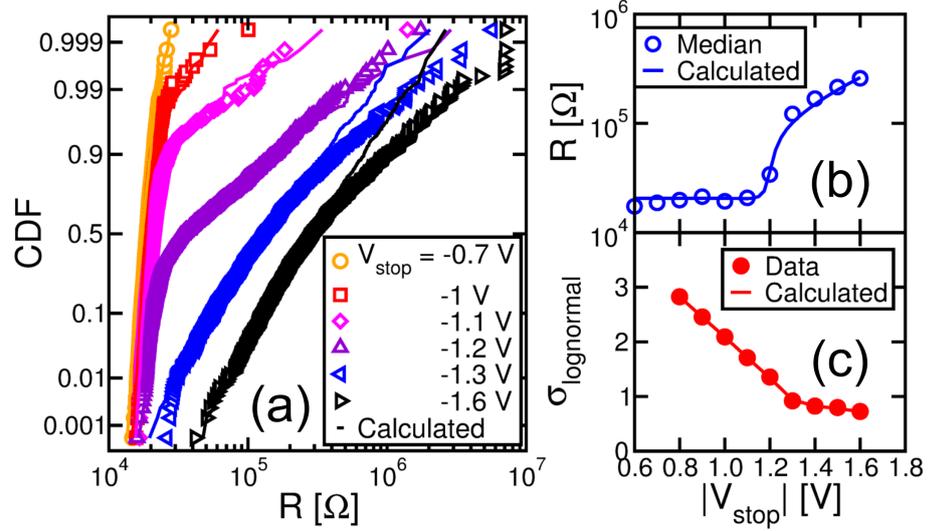


Figura 5.8: Distribuzioni cumulative della resistenza della cella, misurate e calcolate dopo l'operazione di random-reset al crescere di V_{stop} , (a), andamento dei valori mediani corrispondenti, (b), e della deviazione standard della resistenza R del dispositivo, (c). Inoltre sono riportate le simulazioni dell'andamento dei valori mediani, (b), ricavato dal modello analitico, e della deviazione standard, (c), ricavato dal modello semi-empirico di variabilità.

composta da due linee ed è stata ricavata fittando linearmente i dati [66]. Le curve mostrate nelle Figs. 5.8 (c - d) sono state poi utilizzate nelle simulazioni di apprendimento.

In Fig. 5.9 sono rappresentate le probabilità di random-set e random-reset, ricavate dalle Figs. 5.7 e 5.8, in funzione della tensione applicata, ovvero V_A per il set e $|V_{stop}|$ per il reset. La probabilità di set è stata ricavata leggendo il numero di celle delle distribuzioni di Fig. 5.7 (a) che hanno una resistenza $R < 80$ kΩ. La probabilità di reset è stata invece ottenuta leggendo il numero di celle associate alle distribuzioni di Fig. 5.8 (a) aventi $R > 80$ kΩ. Inoltre, mentre il profilo simulato della probabilità di reset è stato ottenuto a partire dal modello empirico mostrato in Fig. 5.8 (c), l'andamento della probabilità di set P_{set} è stato ottenuto fittando i dati sperimentali con la funzione

$$P_{set} = \frac{1 + \operatorname{erf} \frac{V_A - \mu}{\sqrt{2}\sigma}}{2} \quad (5.2)$$

dove $\mu = 1.31$ V e $\sigma = 0.2$ V.

Infine, è importante sottolineare che le distribuzioni simulate mediante il modello analitico/stocastico della cella resistiva [66, 77], riportate anch'esse in Fig. 5.7 (a) e Fig. 5.8 (a), risultano molto consistenti con le misure sperimentali.

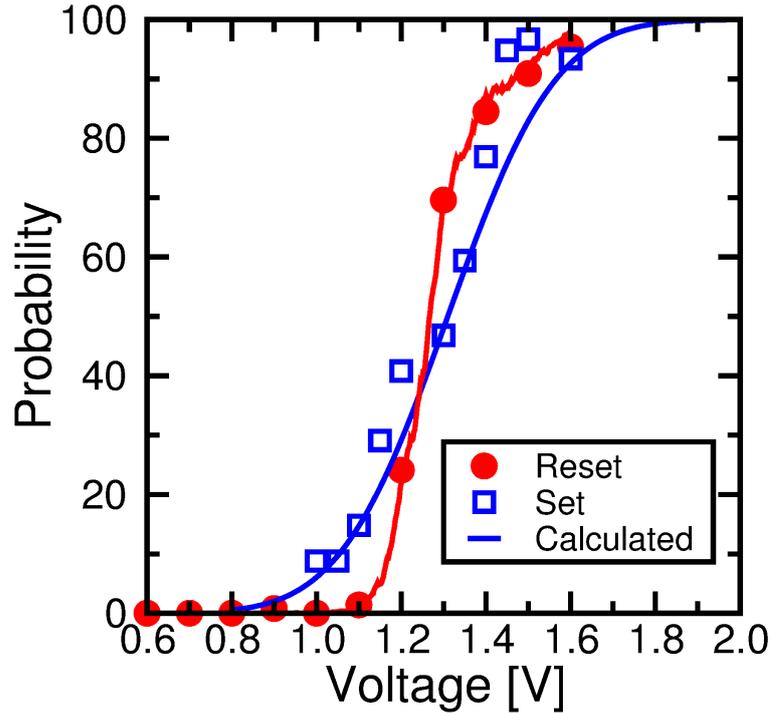


Figura 5.9: Misure e simulazioni della probabilità di random-set in funzione di V_A e di random-reset al variare di V_{stop} .

5.4 Apprendimento di pattern con sinapsi stocastiche

In Fig. 5.10 (a) è descritto il processo di apprendimento di una rete neuromorfica avente le sinapsi caratterizzate da set stocastico e reset deterministico ($V_{TE+} = 1.3$ V e $V_{TE-} = -1.6$ V).

Dalla Fig. 5.9 si osserva che la tensione V_{TE+} attiva la transizione di set con probabilità circa pari a 0.5. L'evoluzione dei pesi sinaptici parte da una condizione assolutamente casuale e si conclude con l'apprendimento del pattern. Questo è evidente perchè dopo circa 60 epoche le sinapsi del pattern risultano potenziate, mentre quelle di background depresse. Il pregio della stocasticità è la possibilità di ridurre il numero di eventi di switching, consentendo così al sistema di essere più indipendente dal rumore e più sensibile al pattern. Inoltre, quando le sinapsi sono potenziate, l'applicazione in ingresso del pattern innesca un flusso di corrente elevato che attiva un evento di fire. Questo impulso, generato dal POST-neurone, potenzia ulteriormente le sinapsi corrispondenti al pattern, se esse non sono già nello stato LRS, e deprime le sinapsi del background, attivate dal rumore che appare in ingresso dopo l'istante di fire ($\Delta t < 0$). S'instaura pertanto un meccanismo di reazione positiva che aumenta la conduttanza delle sinapsi del pattern e deprime quelle del background. Tuttavia, le

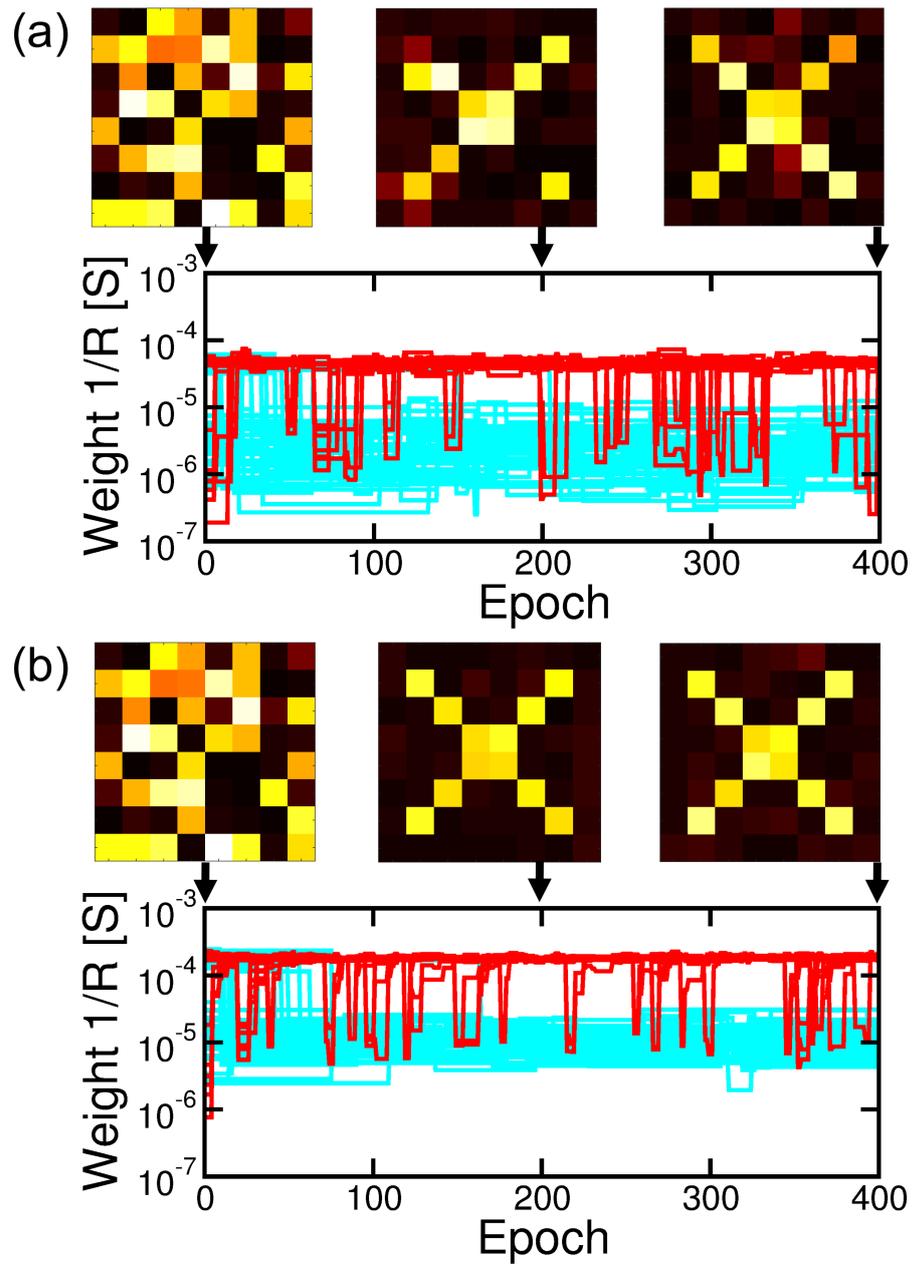


Figura 5.10: Andamento simulato dei pesi $1/R$ delle sinapsi associate al pattern (in rosso) e di quelle riferite al background (in ciano) per due differenti strutture di sinapsi, ovvero la sinapsi a 1 cella stocastica, (a), e quella a 4 celle stocastiche in parallelo, (b). In entrambi i casi $V_{TE+} = 1.3$ V. Le figure poste in alto descrivono i pesi sinaptici ad istanti stabiliti. Si osserva che l'efficienza di apprendimento del sistema migliora all'aumentare del numero di celle per sinapsi.

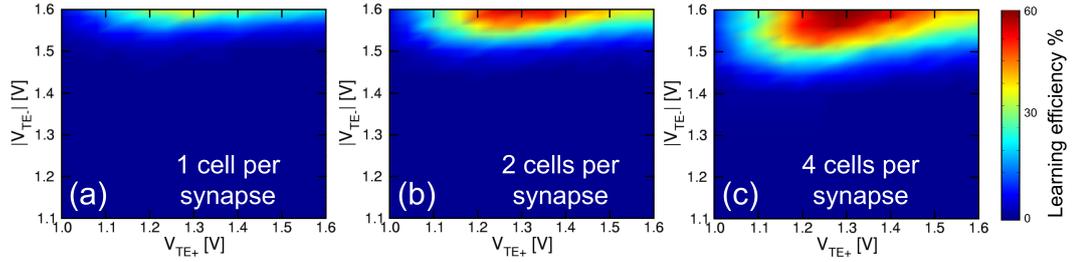


Figura 5.11: Mappe di colore calcolate che descrivono l'efficienza di apprendimento della rete in funzione di V_{TE-} e V_{TE+} nei casi di 1 cella per sinapsi, (a), 2 celle per sinapsi, (b), e 4 celle per sinapsi, (c). L'efficienza di apprendimento migliora all'aumentare del numero di celle per sinapsi a causa della riduzione del tasso di fire spuri e della maggiore incisività della media eseguita sul rumore.

sinapsi associate al pattern possono perdere lo stato LRS, iniziando ad oscillare a causa dell'integrazione del rumore. Il motivo è che il contributo di corrente del background non è trascurabile in quanto la finestra resistiva del dispositivo è soltanto di un fattore 20. A questo si aggiunge la grande dispersione dello stato HRS, come evidente dalle distribuzioni rappresentate nelle Figs. 5.7 (a) e 5.8 (a), che induce il passaggio di elevate correnti di leakage nelle sinapsi di background, innescando in questo modo la generazione di fire spuri.

Una soluzione a questo problema è rappresentata dall'implementazione di sinapsi aventi più celle resistive in parallelo. Si dimostra infatti che adottando sinapsi aventi più RRAM in parallelo, che possono essere interpretate come molteplici canali biologici [82], l'efficienza del processo di apprendimento migliora. In Fig. 5.10 (b) viene mostrata l'evoluzione nel tempo del processo di apprendimento di una rete neuromorfica che impiega sinapsi con 4 RRAM in parallelo. Si nota che lo stato iniziale delle 64 sinapsi è identico a quello considerato nelle Figs. 5.5 e 5.10 (a) e che, dopo circa 80 epoche, l'apprendimento del pattern può considerarsi concluso. In corrispondenza dell'istante di apprendimento inizia ad aprirsi una finestra che indica il rapporto tra i pesi sinaptici associati al pattern e quelli associati al background. In particolare, durante il processo di apprendimento, si osservano meno oscillazioni. La riduzione delle oscillazioni è dovuta da un lato alla minore dispersione della distribuzione HRS delle sinapsi di background, che comporta un forte abbassamento del numero di impulsi indesiderati emessi dal POST-neurone, e dall'altro al comportamento analogico delle sinapsi durante il potenziamento dovuto allo switching stocastico. Infatti, grazie al random switching, le sinapsi realizzate con più celle resistive in parallelo possono commutare tra più stati intermedi ottenuti dal parallelo tra gli stati HRS e LRS. Questo significa che per poten-

ziare completamente una sinapsi saranno necessari più eventi di potenziamento consecutivi e ciò rende automaticamente meno probabile il potenziamento indesiderato del background attivato dal rumore in ingresso.

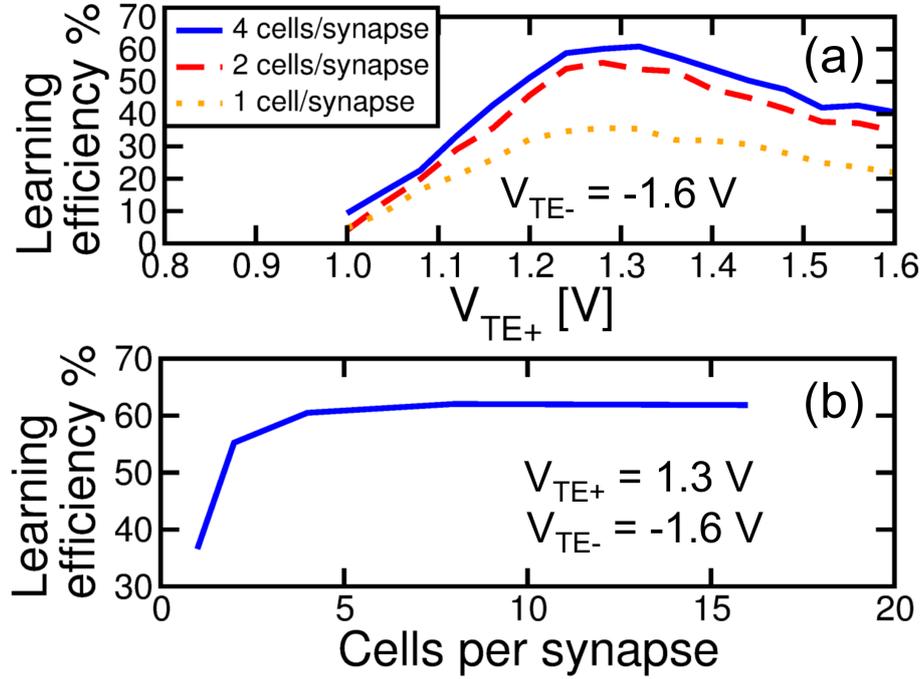


Figura 5.12: Andamento dell'efficienza di apprendimento del sistema neuromorfico in funzione di V_{TE+} , (a), e del numero di celle per singola sinapsi, (b).

Si consideri ora la Fig. 5.11 che mostra le mappe dell'efficienza di apprendimento P_{learn} della rete neuromorfica per sinapsi aventi rispettivamente 1, 2 e 4 celle stocastiche in parallelo. Per efficienza di apprendimento s'intende la capacità del sistema di avere entro 800 epoche le sinapsi del pattern completamente potenziate e, contemporaneamente, le sinapsi del background totalmente depresse, con una tolleranza di al più tre sinapsi del pattern non potenziate e tre sinapsi di background non depresse. In tutti e tre i casi il picco di apprendimento si ha per $|V_{TE-}| = 1.6$ V e $V_{TE+} \simeq 1.3$ V. In Fig. 5.11 (a) è rappresentata l'efficienza P_{learn} di una rete neuromorfica avente sinapsi a 1 cella. Il massimo di efficienza ottenuto è molto basso a causa degli eventi spuri di fire indotti dal rumore ed è circa pari al 40%. Nel caso di sinapsi a 2 celle, la mappa in Fig. 5.11 (b) mostra un risultato migliore, ma il massimo ottenuto si aggira intorno al 52%. Infine, come mostra la Fig. 5.11 (c), il massimo di efficienza ottenuto nel caso di sinapsi a 4 celle raggiunge il 62%. Si osserva inoltre che all'aumentare del numero di celle in parallelo la regione in cui l'efficienza di apprendimento

è alta tende ad allargarsi rendendo la rete neuromorfica meno sensibile alle variazioni delle tensioni V_{TE-} e V_{TE+} e quindi più robusta alle fluttuazioni dei parametri di switching delle RRAM.

La Fig. 5.12 (a) mostra l'andamento dell'efficienza di apprendimento della rete neuromorfica per $V_{TE-} = -1.6$ V al variare di V_{TE+} nei casi di 1, 2 e 4 celle per sinapsi. Si nota che da $V_{TE+} \approx 1$ V, insufficiente per innescare il set delle celle resistive, P_{learn} inizia a crescere fino ad assumere il valore massimo per $V_{TE+} = 1.3$ V. Raggiunto il picco, comincia progressivamente a ridursi in quanto l'impatto favorevole della stocasticità dello switching va esaurendosi a vantaggio del comportamento deterministico che induce una maggiore dipendenza dal rumore. Come già riscontrato precedentemente, l'efficienza diventa più alta all'aumentare del numero di celle per sinapsi.

Per comprendere meglio questa dipendenza, in Fig. 5.12 (b) viene rappresentato l'andamento del massimo di efficienza, ottenuto per $V_{TE-} = -1.6$ V e $V_{TE+} = 1.3$ V, all'aumentare del numero di RRAM per singola sinapsi. Si osserva che, dopo una repentina crescita dovuta all'incisività dell'operazione di media sul rumore e al comportamento analogico delle sinapsi durante la transizione di set, la curva inizia a saturare. La saturazione dell'efficienza di apprendimento per un numero di celle superiore a 4 è dovuta al fatto che, nonostante la media operata sul rumore dal parallelo delle celle resistive, i contributi relativi di background non vengono ridotti in quanto la finestra resistiva delle singole RRAM resta costante indipendentemente dal numero di celle per sinapsi. Ne segue pertanto che il tasso di spike spuri attivati dal rumore non può essere ridotto al di sotto di un limite fissato dal rumore in ingresso.

Capitolo 6

Applicazioni di reti neuromorfiche a RRAM

In questo capitolo prosegue lo studio dell'apprendimento di pattern. In particolare, l'attenzione sarà focalizzata sull'impiego di sinapsi inibitorie per l'apprendimento selettivo di più pattern e sull'apprendimento di pattern a colori. Infine verrà discussa l'utilità dell'implementazione dell'STDP negativa.

6.1 Apprendimento di più pattern mediante sinapsi inibitorie

Si consideri la rete neuromorfica rappresentata in Fig. 6.1 (a). Il layer-retina d'ingresso è costituito da 3600 neuroni, ciascuno dei quali connesso mediante sinapsi ai 4 neuroni d'uscita. Ogni neurone del secondo layer è connesso agli altri tre attraverso sinapsi inibitorie, ovvero sinapsi aventi peso fissato, ed è in grado di specializzarsi selettivamente su diversi pattern. Le sinapsi inibitorie possono essere modellizzate come resistori aventi resistenza fissata che connettono neuroni CMOS.

In seguito ad un evento di fire di uno dei neuroni d'uscita, l'architettura prevede una scarica parziale del potenziale interno dei post-neuroni restanti, determinando per questi ultimi l'impossibilità di generare impulsi di fire in risposta al pattern che ha attivato il primo fire. Quindi, quando un neurone si specializza su un dato pattern, i restanti, a causa dell'inibizione, possono apprendere esclusivamente i pattern non ancora imparati. Quando poi si specializza un secondo neurone, gli ultimi due possono imparare soltanto i due pattern

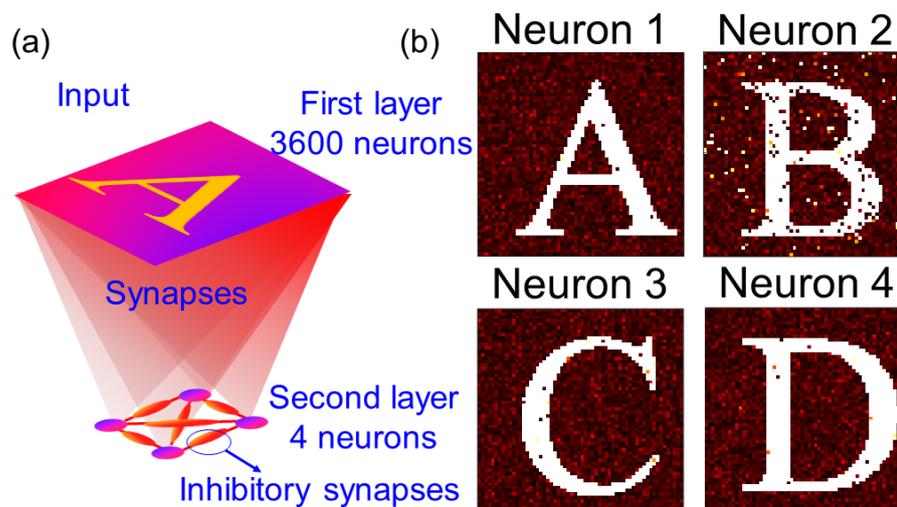


Figura 6.1: Rappresentazione della rete neuromorfica con sinapsi inibitorie impiegata nell'apprendimento multi-pattern, (a), e rappresentazione dei pesi sinaptici dopo 7000 epoche di apprendimento, (b). La presenza delle sinapsi inibitorie impedisce ai neuroni d'uscita di specializzarsi sullo stesso pattern.

restanti e si procede in questo modo finchè ogni neurone non si sarà specializzato su un pattern. Per questo motivo i 4 neuroni d'uscita non possono specializzarsi sullo stesso pattern e quindi consentono l'apprendimento di più immagini in parallelo.

Per dimostrare il funzionamento dell'architettura sopra descritta è stata eseguita una simulazione in cui al layer d'ingresso vengono presentati con una probabilità del 5% 4 pattern con un alto grado di sovrapposizione che rappresentano rispettivamente le lettere 'A', 'B', 'C', 'D'. La sovrapposizione tra i pattern rende la simulazione più complessa poichè, a causa dei contributi di corrente dei pixel in comune, risulterà più difficile per i neuroni specializzarsi sul pattern corretto e distinguere i diversi pattern. La Fig. 6.1 (b) mostra il risultato del processo di apprendimento della rete dopo 7000 epoche. Si osserva che ciascun neurone del secondo layer è riuscito a specializzarsi su uno dei pattern forniti in ingresso alla rete, dimostrando così il corretto funzionamento della rete neuromorfica a 2 layer con sinapsi inibitorie.

6.2 Apprendimento di pattern a colori

Nonostante finora sia stato trattato unicamente l'apprendimento di pattern in bianco e nero, è possibile dimostrare la centralità della sinapsi 1T1R anche nel processo di apprendimento di pattern a colori. Questa nuova applicazione richiede la modulazione dell'ampiezza degli impulsi di tensione V_G , generati dai pre-neuroni della retina, in base all'intensità luminosa dei pixel del pattern in ingresso. Pixel molto luminosi generano impulsi di ampiezza V_G elevata che a loro volta, applicati ai gate dei MOS delle sinapsi 1T1R, attivano correnti molto alte. Ciò si traduce in un aumento della corrente di compliance del transistor e quindi nella riduzione di resistenza dello stato LRS nel caso di set della cella. Pixel poco luminosi generano invece impulsi di gate bassi, dai quali si ottiene una bassa I_C e di conseguenza stati LRS a resistenza più alta.

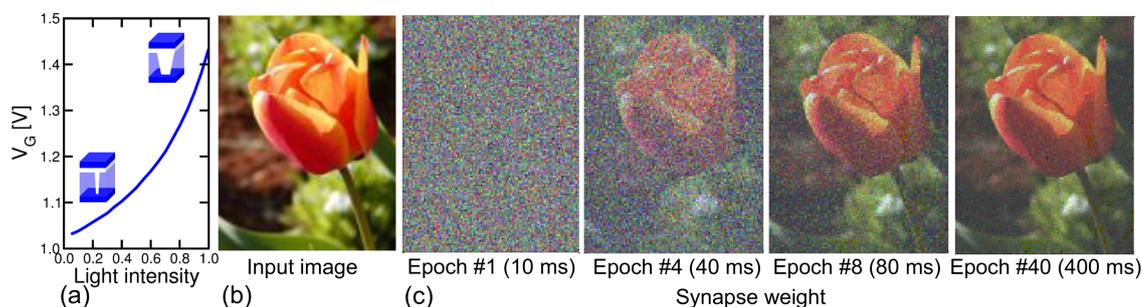


Figura 6.2: Evoluzione nel tempo del processo di apprendimento di un pattern a colori. Attraverso la relazione indicata in (a), la tensione V_G modula la corrente di compliance I_C e quindi la massima conduttanza ottenibile. In (b) è rappresentato il pattern a colori di dimensione pari a 133x100 pixel e in (c) viene mostrato l'aggiornamento dei pesi sinaptici dopo rispettivamente 1, 4, 8 e 40 epoche di apprendimento. Ogni sinapsi conta tre celle e ciascuna di esse tiene conto dell'intensità delle tre componenti di colore fondamentali, ovvero rosso, verde e blu.

La Fig. 6.2 (a) mostra la relazione adottata nelle simulazioni tra V_G e l'intensità luminosa. Ad una bassa luminosità del pixel corrisponde una bassa tensione V_G e quindi uno stato LRS a resistenza maggiore. Viceversa, ad un'alta luminosità viene associata un'elevata V_G e quindi uno stato LRS a resistenza minore, come indicato dal grande spessore del filamento conduttivo associato a questa situazione. Per ottenere l'apprendimento di un pattern a colori è stata scelta una codifica del colore basata sul protocollo RGB. La rete neuromorfica adottata prevede 3 layer d'ingresso dove ciascuno di essi è sensibile a uno dei tre colori. Questi tre layer d'ingresso sono connessi mediante sinapsi 1T1R a 3 post-neuroni d'uscita che costituiscono il layer finale della struttura. In Fig. 6.2 (b) viene mostrata l'immagine che

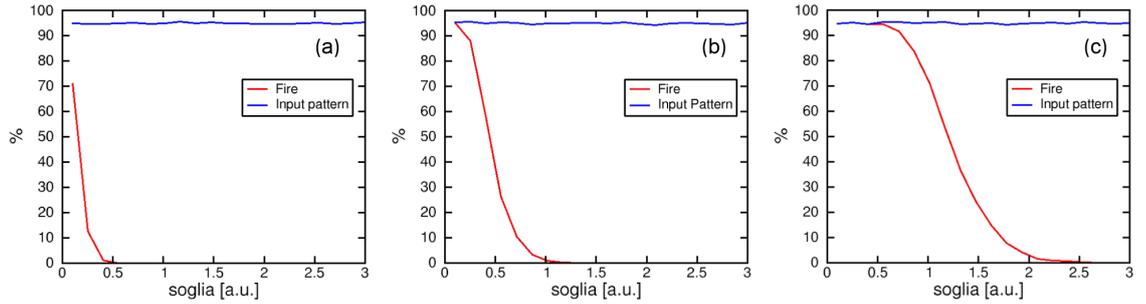


Figura 6.3: Andamento della frequenza media dell'attività dei neuroni d'ingresso (in blu) e d'uscita (in rosso) al variare del valore della soglia in corrente dei post-neuroni nei casi di 1 cella per sinapsi, (a), 2 celle per sinapsi, (b), e 4 celle per sinapsi, (c). All'aumentare della soglia, la corrente che deve essere integrata per attivare un evento di fire in uscita aumenta e quindi di conseguenza il numero di eventi in uscita tende a diminuire.

è stata fornita alla retina d'ingresso. La dimensione dell'immagine, che è identica a quella della retina, è pari a 133x100 pixel. Infine, la Fig. 6.2 (c) mostra l'evoluzione nel tempo dei pesi sinaptici in una simulazione avente una durata di 40 epoche. I colori riprodotti in figura sono stati ottenuti leggendo i pesi delle sinapsi mediante il protocollo RGB. Il risultato ottenuto dimostra che una rete neuromorfica di questo tipo è in grado di apprendere pattern a colori e quindi di conseguenza anche immagini catturate dal mondo reale.

6.3 STDP negativa

Un'ultima applicazione in cui è stata impiegata la sinapsi 1T1R consiste nell'implementazione dell'STDP negativa. A differenza dell'STDP tradizionale, l'STDP negativa prevede il potenziamento delle sinapsi per valori negativi del ritardo Δt tra l'inizio dell'impulso pre-sinaptico e l'inizio dell'impulso post-sinaptico e la depressione per valori positivi di Δt . Implementando l'STDP negativa si realizza un filtro che modula la frequenza dell'immagine, migliorando così le condizioni per cui l'apprendimento è più efficiente. Nelle simulazioni abbiamo adottato una rete neuromorfica a 2 layer dove sia il layer d'ingresso, che riceve il pattern X, che quello d'uscita contano 64 neuroni. L'obiettivo è determinare, al variare della soglia interna dei neuroni del secondo layer, in che percentuale di casi durante la simulazione si ha il pattern in ingresso e il pattern in uscita. A tal proposito, in Fig. 6.3, è mostrato il confronto tra l'andamento della frequenza media con cui il pattern X è applicato sul layer d'ingresso, fissata a un valore costante durante tutta la simulazione, e l'andamento della

frequenza media con cui i neuroni d'uscita generano impulsi di fire al variare della soglia interna dei post-neuroni nei casi di 1, 2 e 4 celle stocastiche per sinapsi. In ciascuno dei tre casi si osserva che il numero medio di eventi di fire in uscita diminuisce all'aumentare della soglia e ciò è dovuto al fatto che maggiore è il livello di soglia che la corrente integrata dal post-neurone deve superare per generare un impulso di fire, minore è la probabilità di eventi in uscita. Inoltre si nota che all'aumentare del numero di celle per sinapsi la pendenza della curva, che descrive la diminuzione degli eventi di fire al crescere della soglia, tende progressivamente a diminuire, consentendo quindi di regolare in modo più accurato, rispetto ai casi di sinapsi a 1 o 2 celle stocastiche in parallelo, la frequenza dell'immagine trasmessa al secondo layer della rete al fine di migliorarne la capacità di apprendimento.

In conclusione, l'STDP negativa impedisce l'apprendimento del pattern attraverso una diminuzione della sua frequenza in uscita, mentre l'STDP positiva consente alla rete di apprendere il pattern senza ridurne la frequenza.

Capitolo 7

Modellistica di un generatore di numeri casuali a RRAM

Le memorie non volatili a switching resistivo sono caratterizzate da un'elevata velocità di switching, basso consumo di energia, ma anche da un'elevata variabilità. Tuttavia proprio l'elevata stocasticità delle RRAM si rivela molto utile in alcune particolari applicazioni come ad esempio la generazione fisica di numeri casuali. In questo capitolo verrà discussa l'implementazione e il funzionamento di due distinti blocchi generatori di numeri casuali (Random Number Generator, RNG) che sfruttano proprio la variabilità intrinseca dello switching resistivo.

7.1 RNG a 2 RRAM in parallelo

L'implementazione di un generatore di numeri casuali rappresenta un problema di notevole complessità a causa della difficoltà di riprodurre fenomeni totalmente casuali. Una possibile soluzione a questo problema è fornita dall'alta variabilità statistica e dall'elevata casualità intrinseca di alcuni parametri fisici, come la resistenza e le tensioni di switch, delle memorie resistive. A tal proposito, in questo capitolo saranno trattate due differenti soluzioni circuitali che prevedono l'impiego di RRAM per realizzare un vero RNG.

Lo schema in Fig. 7.1 (a) mostra la struttura di un generatore di numeri casuali a RRAM in cui le due celle resistive P e Q sono connesse in parallelo e il comparatore (CMP) viene impiegato per la rigenerazione della tensione V_{out} .

La Fig. 7.1 (b) descrive le 3 fasi del ciclo di funzionamento del circuito RNG in esame.

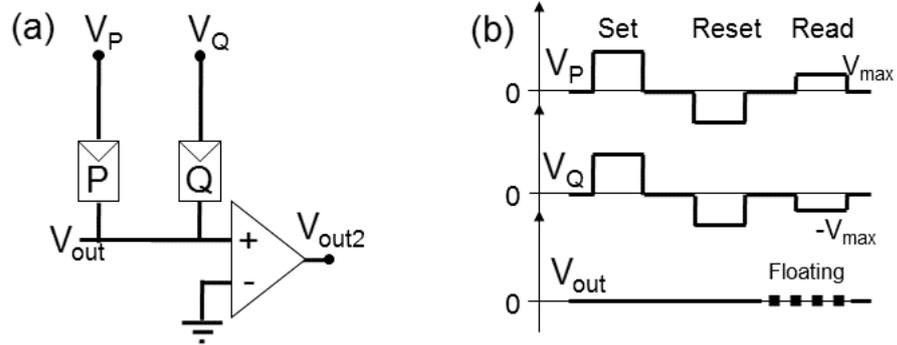


Figura 7.1: Schema circuitale di un RNG a 2 RRAM in parallelo, (a), e rappresentazione dell'evoluzione nel tempo di V_P , V_Q e V_{out} , (b).

La prima fase prevede l'applicazione di un impulso positivo di set ad entrambe le celle, con V_{out} fissa a 0 V. Questa operazione attiva la commutazione delle due celle in uno stato a bassa resistenza (LRS). Nella seconda fase sia P che Q vengono sottoposte ad un impulso di reset e quindi commutano in uno stato ad alta resistenza (HRS).

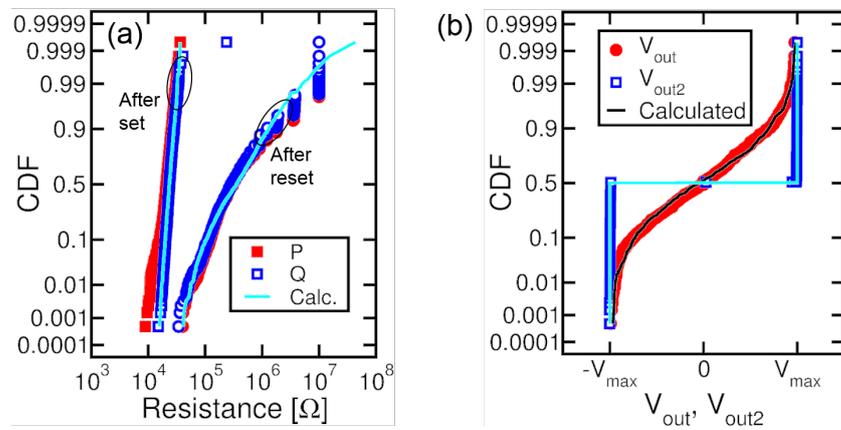


Figura 7.2: Misure e simulazioni delle distribuzioni cumulative di resistenza delle celle P e Q dopo le operazioni di set e reset, (a), e delle distribuzioni cumulative delle tensioni V_{out} e V_{out2} , (b).

In Fig. 7.2 (a) sono rappresentate le distribuzioni cumulative sperimentali di resistenza delle due celle rispettivamente dopo il set e il reset con le rispettive simulazioni, che sono state calcolate fittando i dati con distribuzioni log-normali.

Infine, l'ultima fase del ciclo consiste nell'applicazione alle due celle di impulsi di lettura ($V_P = V_{max}$ e $V_Q = -V_{max}$), con il nodo a tensione V_{out} non polarizzato.

Concluso il ciclo, a causa della variabilità dell'operazione di reset, la tensione V_{out} mostra una distribuzione uniforme tra $-V_{max}$ e V_{max} , che viene poi rigenerata, mediante un comparatore, in una distribuzione bimodale 0/1, come confermato in Fig. 7.3 (a - b).

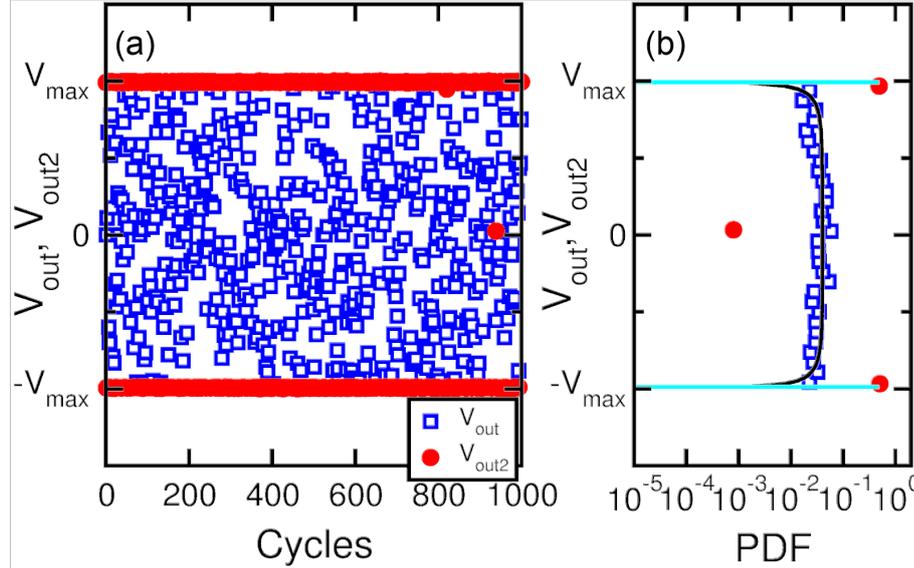


Figura 7.3: Misure sperimentali di V_{out} e V_{out2} su 10^3 cicli, (a), e andamento delle distribuzioni misurate e calcolate di V_{out} e V_{out2} , (b). Grazie alla variabilità dello stato HRS per alte V_{stop} , V_{out} mostra una distribuzione uniforme tra $-V_{max}$ e V_{max} , che è poi rigenerata attraverso il comparatore nella distribuzione bimodale 0/1 associata a V_{out2} .

7.2 RNG a 2 RRAM in serie

Il blocco circuitale descritto in Fig. 7.4 (a), a differenza dello schema precedentemente discusso, prevede l'impiego di due celle RRAM connesse in serie anziché in parallelo.

Il ciclo di funzionamento di questo circuito si articola in 4 fasi e in Fig. 7.4 (b) viene mostrato l'andamento delle tensioni V_P , V_Q e V_{out} .

Nella prima fase la cella P viene sottoposta ad un impulso di set con V_{out} e V_Q fisse a 0 V, mentre durante la seconda fase a settare è la cella Q, con V_P e V_{out} a massa. La terza fase del ciclo prevede l'applicazione di un impulso che innesca il random reset di entrambe le celle con il nodo intermedio lasciato privo di polarizzazione.

Infine, il ciclo si conclude con la lettura della tensione d'uscita V_{out} .

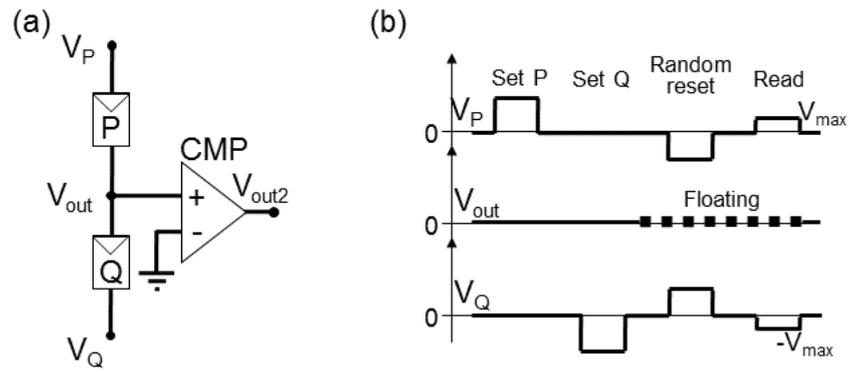


Figura 7.4: Schema circuitale di un RNG a 2 RRAM in serie, (a), e sequenza delle tensioni V_P , V_Q e V_{out} , (b).

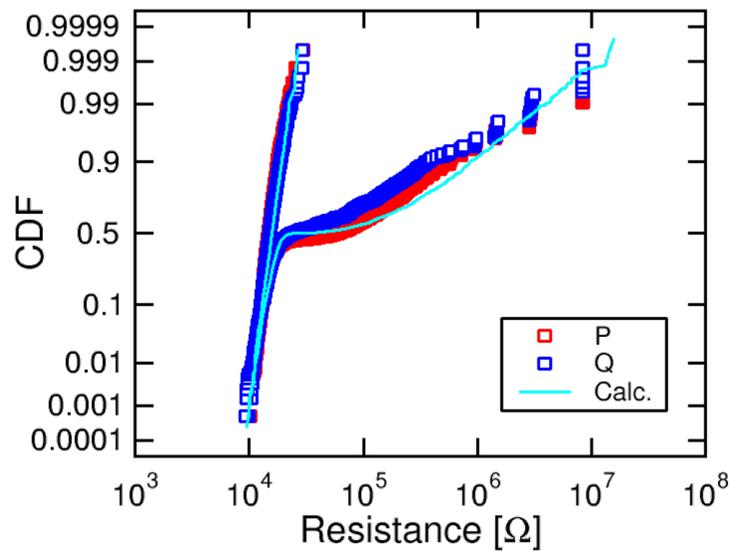


Figura 7.5: Misure e simulazioni delle distribuzioni cumulative di resistenza di P e Q prima e dopo l'operazione di random reset. Come conseguenza dell'operazione di random reset, P e Q commutano in uno stato HRS solo nel 50% del numero totale di cicli eseguiti.

In Fig. 7.5 sono mostrate le distribuzioni misurate della resistenza di P (in rosso) e di Q (in blu) dopo il set e il random reset con le relative simulazioni (in ciano). A partire dalla distribuzione log-normale calcolata per l'operazione di set, le celle commutano in uno stato

ad alta resistenza con una probabilità del 50%. Il risultato dello switching stocastico è quindi la generazione di una distribuzione finale di resistenza che consiste nella combinazione tra il 50% di celle associate allo stato LRS e il 50% di celle sottoposte alla commutazione nello stato HRS.

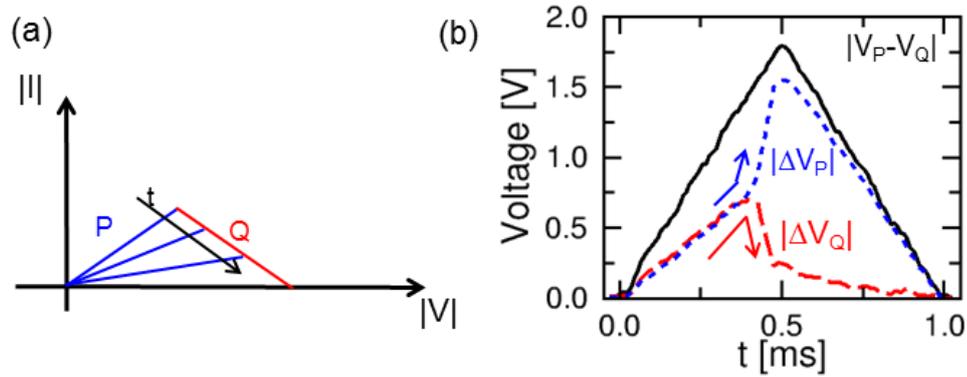


Figura 7.6: Rappresentazione schematica della curva I-V che descrive la transizione di reset della cella P, (a), e rappresentazione dell'evoluzione nel tempo delle cadute di tensione misurate ai capi delle due celle durante la transizione di reset della cella P, (b).

A causa della variabilità intrinseca che caratterizza l'operazione di reset, una cella (P in Fig. 7.6 (a)) inizierà la transizione di reset prima dell'altra vedendo così crescere nel tempo la caduta di tensione ai suoi capi, come mostrato in Fig. 7.6 (a - b). Di conseguenza, la caduta ai capi della cella che non resetta per prima sarà progressivamente sempre più piccola e quindi tale da impedirne il reset. Pertanto, come si evince dalle Figs. 7.6 e 7.7, durante ogni ciclo RNG resetta solo ed esclusivamente una cella.

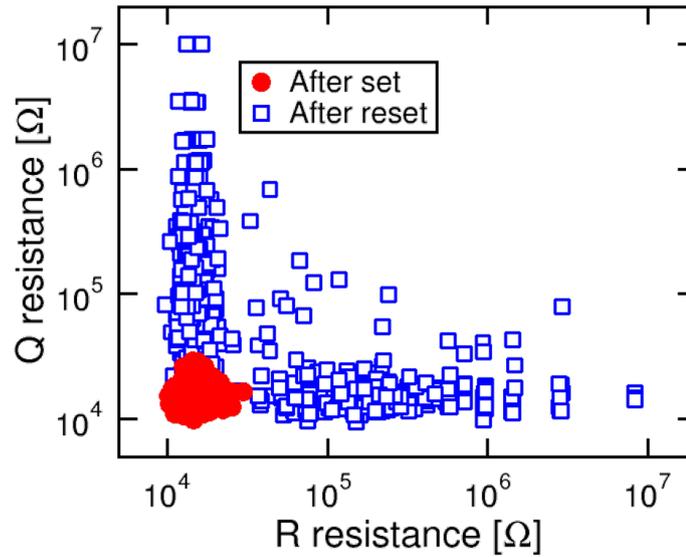


Figura 7.7: Plot di correlazione delle resistenze di P e Q dopo l'operazione di set e reset. Dopo il set, le misure di resistenza di entrambe le celle mostrano valori tutti prossimi a 10 kΩ, mentre dopo il random reset o resetta Q, con P che si mantiene a bassa resistenza, o viceversa.

Come mostrato in Fig. 7.8 e Fig. 7.9 (a - b), alla fine del ciclo la tensione V_{out} assumerà una distribuzione bimodale con valori prossimi a V_{max} e $-V_{max}$, conseguenza del fatto che la quasi totalità dell'intera tensione applicata al circuito cade ai capi di un'unica cella. Infine si nota che attraverso il comparatore (CMP) connesso al nodo di uscita, la distribuzione cumulativa di V_{out} viene squadrata in una distribuzione ideale 0/1.

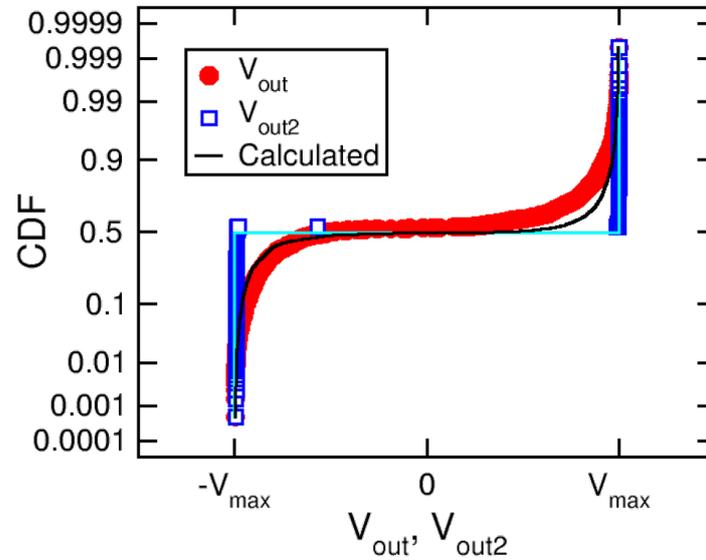


Figura 7.8: Distribuzioni misurate e calcolate di V_{out} , indicate rispettivamente in rosso e in nero, e di V_{out2} , rappresentate in blu e in ciano. Grazie al comparatore presente nel circuito, dopo un ciclo RNG si passa dalla distribuzione bimodale di V_{out} alla distribuzione 0/1 associata a V_{out2} .

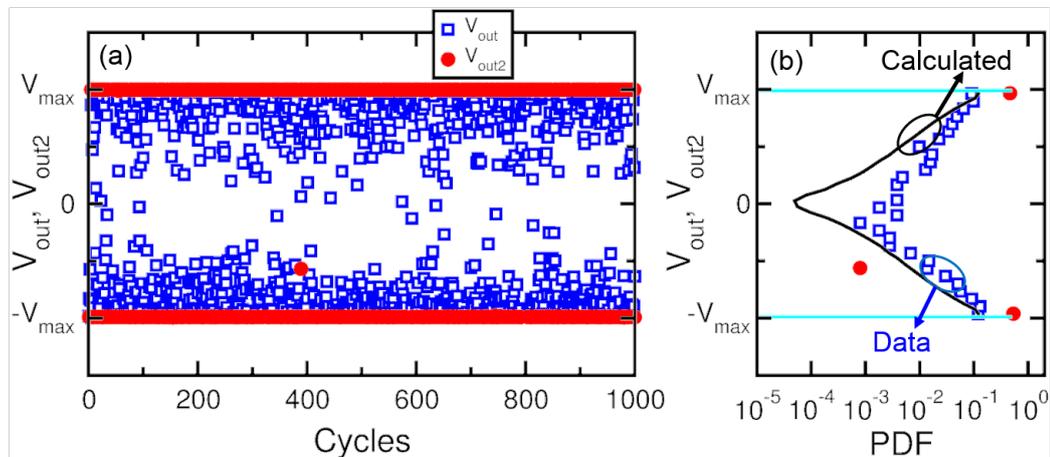


Figura 7.9: Andamento delle misure di V_{out} e V_{out2} su 10^3 cicli, (a), e rappresentazione delle distribuzioni ottenute dai dati sperimentali e dalle simulazioni di V_{out} e V_{out2} , (b). Grazie al random reset di P o Q, la tensione V_{out} mostra una distribuzione bimodale che viene poi rigenerata mediante un comparatore nella distribuzione 0/1 associata a V_{out2} .

Conclusioni

A conclusione di questo lavoro di tesi è opportuno elencare i principali risultati ottenuti.

In primo luogo è stato analizzato accuratamente l'impatto della capacità parassita della struttura 1T1R su alcune importanti caratteristiche della cella RRAM come la corrente, la resistenza associata allo stato basso resistivo e la tensione di set. L'interpretazione di questo effetto, denominato overshoot capacitivo, è stata suffragata in modo evidente dalle simulazioni effettuate mediante il modello analitico e mediante una formula ricavata analiticamente a partire dall'equazione che regola il processo di set della cella. Nel complesso, l'accordo ottenuto tra i dati sperimentali e le curve calcolate risulta molto soddisfacente.

In secondo luogo è stato presentato un modello circuitale dell'architettura 1T1R realizzato attraverso il simulatore Simulink, che, a partire dalla caratterizzazione elettrica del transistor MOS e dall'implementazione della cella RRAM mediante il modello analitico, ha consentito da un lato di confermare la correttezza della descrizione fisica fornita dal modello analitico e dall'altro di simulare le curve di plasticità di una sinapsi realizzata attraverso l'innovativo schema 1T1R.

Dopodichè l'attenzione è stata concentrata sulla simulazione dell'apprendimento di pattern visivi da parte di reti neuromorfiche aventi come elemento fondamentale la sinapsi 1T1R. Questo studio, eseguito a partire dalle distribuzioni di resistenza ricavate dalle misure sperimentali, ha messo in luce la possibilità di sfruttare l'elevata variabilità delle RRAM per ottenere reti neurali artificiali in grado di apprendere e riconoscere i pattern mostrati in ingresso. Infatti, se da un lato l'impiego di sinapsi basate su memorie resistive a switching deterministico non ha prodotto risultati in termini di capacità di apprendimento della rete, dall'altro, l'utilizzo di connessioni inter-neuronali costituite da più celle a switching stocastico, da cui trae origine il comportamento analogico delle sinapsi, ha consentito il raggiungimento di un'apprezzabile efficienza di apprendimento.

Il passo successivo è stato quello di estendere l'apprendimento a più pattern tramite l'introduzione del concetto di sinapsi inibitoria, la quale ha la funzione di permettere a più

neuroni del secondo layer di specializzarsi su più pattern. A questa applicazione ha fatto seguito l'apprendimento di pattern a colori, basato su una scala di grigi generata a partire dalla modulazione della tensione applicata al gate del transistor della cella 1T1R, e lo studio dell'STDP negativa con cui è possibile implementare un modulatore della frequenza dell'immagine trasmessa ai layer successivi della rete.

In ultima analisi è stato affrontato il difficile problema della generazione fisica di numeri puramente casuali. Le soluzioni circuitali presentate consentono di realizzare un vero generatore casuale grazie, ancora una volta, all'alta variabilità e casualità che caratterizzano lo switching delle memorie resistive.

Bibliografia

- [1] H.-S. P. Wong and S. Salahuddin. Memory leads the way to better computing. *Nature Nanotechnology*, 10:191–194, 2015.
- [2] C. Lam. Storage class memory: Opportunities and challenges. *Workshop on Innovative Memory Technologies*, 2009.
- [3] D.J. Wouters, R. Waser, and M. Wuttig. Phase-change and redox-based resistive switching memories. *Proc. IEEE*, 103(8):1274–1288, 2015.
- [4] A.D. Kent and D.C. Worledge. A new spin on magnetic memories. *Nature Nanotechnology*, 10:187–191, 2015.
- [5] J. Kim, A. Paul, P.A. Crowell, S.J. Koester, S.S. Sapatnekar, J.-P. Wang, and C.H. Kim. Spin-based computing: Device concepts, current status, and a case study on a high-performance microprocessor. *Proc. IEEE*, 103(1):106–130, 2015.
- [6] Rainer Waser and Masakazu Aono. Nanoionics-based resistive switching memories. *Nature Materials*, 6:833–840, 2007.
- [7] D. Ielmini. Modeling the universal set/reset characteristics of bipolar RRAM by field- and temperature-driven filament growth. *IEEE Trans. Electron Devices*, 58(12):4309–4317, 2011.
- [8] S. Larentis, F. Nardi, S. Balatti, D. C. Gilmer, and D. Ielmini. Resistive switching by voltage-driven ion migration in bipolar RRAM - Part II: Modeling. *IEEE Trans. Electron Devices*, 59(9):2468–2475, 2012.
- [9] S. Ambrogio, S. Balatti, D.C. Gilmer, and D. Ielmini. Analytical modeling of oxide-based bipolar resistive memories and complementary resistive switches. *IEEE Trans. Electron Devices*, 61:2378–2386, 2014.

-
- [10] G.W. Burr, B.N. Kurdi, J.C. Scott, C.H. Lam, K. Gopalakrishnan, and R.S. Shenoy. Overview of candidate device technologies for storage-class memory. *IBM J. Res. Dev.*, 52(4/5):449–464, 2008.
- [11] R.F. Freitas and W.W. Wilcke. Storage-class memory: The next storage system technology. *IBM J. Res. Dev.*, 52(4/5):439–447, 2008.
- [12] H.-S. Philip Wong, H.-Y. Lee, S. Yu, Y.-S. Chen, Y. Wu, P.-S. Chen, B. Lee, F. T. Chen, and M.-J. Tsai. Metal-oxide RRAM. *Proc. IEEE*, 100(6):1951–1970, 2012.
- [13] J. Joshua Yang, Dmitri B. Strukov, and Duncan R. Stewart. Memristive devices for computing. *Nature Nanotechnology*, 8:13–24, 2013.
- [14] S. Lai. Current status of the phase change memory and its future. *IEDM Tech. Dig.*, pages 255–258, 2003.
- [15] J. Liang, R.G.D. Jeyasingh, H.-Y. Chen, and H.-S.P. Wong. A 1.4 μA reset current phase change memory cell with integrated carbon nanotube electrodes for cross-point memory application. *Proc. Symp. VLSI*, pages 100–101, 2011.
- [16] Y. Sasago, M. Kinoshita, T. Morikawa, K. Kurotsuchi, S. Hanzawa, T. Mine, A. Shima, Y. Fujisaki, H. Kume, H. Moriya, N. Takaura, and K. Torii. Cross-point phase change memory with 4F^2 cell size driven by low-contact-resistivity poly-Si diode. *Proc. Symp. VLSI*, pages 24–25, 2009.
- [17] D.C. Kau, S. Tang, I.V. Karpov, R. Dodge, B. Klehn, J.A. Kalb, J. Strand, A. Diaz, N. Leung, J. Wu, S. Lee, T. Langtry, K.-W. Chang, C. Papagianni, J. Lee, J. Hirst, S. Erra, E. Flores, N. Righos, H. Castro, and G. Spadini. A stackable cross point phase change memory. *IEDM Tech. Dig.*, pages 571–574, 2009.
- [18] N. Ciocchini, M. Cassinerio, D. Fugazza, and D. Ielmini. Modeling of Threshold-Voltage drift in Phase-Change Memory (PCM) devices. *IEEE Trans. Electron Devices*, 59(11):3084–3090, 2012.
- [19] M. Boniardi and D. Ielmini. Physical origin of the resistance drift in amorphous phase-change materials. *Appl. Phys. Lett.*, 98:243506, 2011.
- [20] N. Ciocchini, M. Cassinerio, D. Fugazza, and D. Ielmini. Evidence for Non-Arrhenius kinetics of crystallization in phase change memory devices. *IEEE Trans. Electron Devices*, 60(11):3767–3774, 2013.
-

-
- [21] I.G. Baek, M.S. Lee, S. Seo, M.J. Lee, D.H. Seo, D.-S. Suh, J.C. Park, S.O. Park, H.S. Kim, I.K. Yoo, U.-I. Chung, and J.T. Moon. Highly scalable nonvolatile resistive memory using simple binary oxide driven by asymmetric unipolar voltage pulses. *IEDM Tech. Dig.*, pages 587–590, 2004.
- [22] H. Y. Lee, P. S. Chen, T. Y. Wu, Y. S. Chen, C. C. Wang, P. J. Tzeng, C. H. Lin, F. Chen, C. H. Lien, and M.-J. Tsai. Low power and high speed bipolar switching with a thin reactive Ti buffer layer in robust HfO₂ based RRAM. *IEDM Tech. Dig.*, pages 297–300, 2008.
- [23] R. Waser, R. Dittmann, G. Staikov, and K. Szot. Redox-based resistive switching memories - nanoionic mechanisms, prospects, and challenges. *Adv. Mater.*, 21:2632, 2009.
- [24] T.W. Hickmott. Low-frequency negative resistance in thin anodic oxide films. *J. Appl. Phys.*, 33:2669, 1962.
- [25] D. Ielmini, F. Nardi, and C. Cagli. Physical models of size-dependent nanofilament formation and rupture in NiO resistive switching memories. *Nanotechnology*, 22(25):254022, 2011.
- [26] D. Ielmini, F. Nardi, C. Cagli, and A. L. Lacaita. Size-dependent retention time in NiO-based resistive-switching memories. *IEEE Electron Device Lett.*, 31(4):353–355, 2010.
- [27] K. Kinoshita, K. Tsunoda, Y. Sato, H. Noshiro, S. Yagaki, M. Aoki, and Y. Sugiyama. Reduction in the reset current in a resistive random access memory consisting of NiO_x brought about by reducing a parasitic capacitance. *Appl. Phys. Lett.*, 93(3):033506, 2008.
- [28] F. Nardi, D. Ielmini, C. Cagli, S. Spiga, M. Fanciulli, L. Goux, and D. J. Wouters. Control of filament size and reduction of reset current below 10 μ A in NiO resistance switching memories. *Solid State Electronics*, 58:42–47, 2011.
- [29] H. Chung-Wei, I-Ting Wang, Chun-Li Lo, Ming-Chung Chiang, Wen-Yueh Jang, Chen-Hsi Lin, and Tuo-Hung Hou. Self-rectifying bipolar TaO_x/TiO₂ RRAM with superior endurance over 10¹² cycles for 3D high-density storage-class memory. *Proc. Symp. VLSI*, pages T166–T167, 2013.
-

-
- [30] W.-G. Kim and S.-W. Rhee. Effect of the top electrode material on the resistive switching of TiO₂ thin film. *Microelectron. Eng.*, 87:98–103, 2010.
- [31] H.Y. Lee, Y.-S. Chen, P.S. Chen, P.Y. Gu, Y.Y. Hsu, S.M. Wang, W.H. Liu, C.H. Tsai, S.S. Sheu, P.-C. Chiang, W.P. Lin, C.-H. Lin, W.-S. Chen, F.T. Chen, C.H. Lien, and M. Tsai. Evidence and solution of over-RESET problem for HfO_x based resistive memory with sub-ns switching speed and high endurance. *IEDM Tech. Dig.*, pages 460–463, 2010.
- [32] J. Guy, G. Molas, E. Vianello, F. Longnos, S. Blanc, C. Carabasse, M. Bernard, J.F. Nodin, A. Toffoli, J. Cluzel, P. Blaise, P. Dorion, O. Cueto, H. Grampeix, E. Souchier, T. Cabout, P. Brianceau, V. Balan, A. Roule, S. Maitrejean, L. Perniola, and B. De Salvo. Investigation of the physical mechanisms governing data-retention in down to 10 nm nano-trench Al₂O₃/CuTeGe conductive bridge RAM (CBRAM). *IEDM Tech. Dig.*, pages 30.2.1–30.2.4, 2013.
- [33] J.J. Yang, M.-X. Zhang, J.P. Strachan, F. Miao, M.D. Pickett, R.D. Kelley, G. Medeiros-Ribeiro, and R.S. Williams. High switching endurance in TaO_x memristive devices. *Appl. Phys. Lett.*, 97:232102, 2010.
- [34] M.-J. Lee, C.B. Lee, D. Lee, S.R. Lee, M. Chang, J.H. Hur, Y.-B. Kim, C.-J. Kim, D.H. Seo, S. Seo, U.-I. Chung, I.-K. Yoo, and K. Kim. A fast, high-endurance and scalable non-volatile memory device made from asymmetric Ta₂O_{5-x}=TaO_{2-x} bilayer structures. *Nature Materials*, 10:625–630, 2011.
- [35] C. Chen, L. Goux, A. Fantini, A. Redolfi, S. Clima, R. Degraeve, Y. Chen, G. Groeseneken, and M. Jurczak. Understanding the impact of programming pulses and electrode materials on the endurance properties of scaled Ta₂O₅ RRAM cells. *IEDM Tech. Dig.*, pages 355–358, 2014.
- [36] T. Liu, M. Verma, Y. Kang, and M. Orlowski. Volatile resistive switching in Cu/TaOx/ δ -Cu/Pt devices. *Appl. Phys. Lett.*, (101):073510, 2012.
- [37] F. Nardi, S. Larentis, S. Balatti, D. C. Gilmer, and D. Ielmini. Resistive switching by voltage-driven ion migration in bipolar RRAM - Part I: Experimental study. *IEEE Trans. Electron Devices*, 59(9):2461–2467, 2012.
- [38] Y. Yang, P. Gao, S. Gaba, T. Chang, X. Pan, and W. Lu. Observation of conducting filament growth in nanoscale resistive memories. *Nature communications*, 3(732):1–8, 2012.
-

-
- [39] C. Gopalan, Y. Ma, T. Gallo, J. Wang, E. Runnion, J. Saenz, F. Koushan, P. Blanchard, and S. Hollmer. Demonstration of conductive bridging random access memory (CBRAM) in logic CMOS process. *Solid-State Electr.*, 58:54–61, 2011.
- [40] I. Valov, R. Waser, J.R. Jameson, and M.N. Kozicki. Electrochemical metallization memories—fundamentals, applications, prospects. *Nanotechnology*, 22:254003, 2011.
- [41] A. Calderoni, S. Sills, and N. Ramaswamy. Performance comparison of O-based and Cu-based ReRAM for high-density applications. *IMW Proc.*, pages 1–4, 2014.
- [42] F. T. Chen, H.-Y. Lee, Y.-S. Chen, S. Z. Rahaman, C.-H. Tsai, K.-H. Tsai, T.-Y. Wu, W.-S. Chen, P.-Y. Gu, Y.-D. Lin, S.-S. Sheu, M.-J. Tsai, L.-H. Lee, T.-K. Ku, and P.-S. Chen. Resistance instabilities in a filament-based resistive memory. *Proc. IRPS*, pages 5E.1.1–5E.1.7, 2013.
- [43] S. Ambrogio, S. Balatti, A. Cubeta, A. Calderoni, N. Ramaswamy, and D. Ielmini. Statistical fluctuations in HfO_x resistive-switching memory (RRAM): Part II - Random telegraph noise. *IEEE Trans. Electron Devices*, 61(8):2920–2927, 2014.
- [44] Francesco M. Puglisi, Paolo Pavan, Andrea Padovani, Luca Larcher, and Gennadi Bersuker. RTS noise characterization of HfO_x RRAM in high resistive state. *Solid State Electronics*, 84:160–166, 2013.
- [45] S. Ambrogio, S. Balatti, S. Choi, and D. Ielmini. Impact of the mechanical stress on switching characteristics of electrochemical resistive memory. *Adv. Mat.*, 26(23):3885, 2014.
- [46] Yang Yin Chen, R. Degraeve, B. Govoreanu, S. Clima, L. Goux, A. Fantini, G.S. Kar, D.J. Wouters, G. Groeseneken, and M. Jurczak. Postcycling LRS retention analysis in HfO_2/Hf RRAM 1T1R device. *IEEE Electron Device Lett.*, 34(5):626–628, 2013.
- [47] Bipin Rajendran, Yong Liu, Jae-Sun Seo, Kailash Gopalakrishnan, Leland Chang, Daniel J. Friedman, and Mark B. Ritter. Specifications of nanoscale devices and circuits for neuromorphic computational systems. *IEEE Transactions on Electron Devices*, 60(1):246–253, 2013.
- [48] K.K Likharev, A. Mayr, I. Muckra, and O. T’urel. Crossnets - high-performance neuromorphic architectures for CMOL circuits. *Ann. N.Y. Acad. Sci.*, (1006):146–163, 2003.
-

-
- [49] G.S. Snider. Spike-timing-dependent learning in memristive nanodevices. *IEEE International Symposium on Nanoscale Architectures*, pages 85 – 92, 2008.
- [50] Sung Hyun Jo, Ting Chang, Idongesit Ebong, Bhavitavya B. Bhadviya, Pinaki Mazumder, and Wei Lu. Nanoscale memristor device as synapse in neuromorphic systems. *Nano Letters*, 10:1297–1301, 2010.
- [51] C.J. Diorio, P.E. Hasler, C.A. Mead, and B.A. Minch. A single-transistor silicon synapse. *IEEE Trans. Electron Devices*, 43:1972, 1996.
- [52] G. Indiveri, E. Chicca, and R. Douglas. A VLSI array of low-power spiking neurons and bistable synapses with Spike-Timing Dependent Plasticity. *IEEE Trans. Neural Networks*, 17:211, 2006.
- [53] D. Kuzum, R.G.D. Jeyasingh, B. Lee, and H.-S.P. Wong. Nanoelectronic programmable synapses based on phase change materials for brain-inspired computing. *Nano Lett.*, (12):2179, 2012.
- [54] C.D. Wright, Y. Liu, K.I. Kohary, M.M. Aziz, and R.J. Hicken. Arithmetic and biologically-inspired computing using phase-change materials. *Adv. Mater.*, (23):3408, 2011.
- [55] O. Bichler, M. Suri, D. Querlioz, D. Vuillaume, B. DeSalvo, and C. Gamrat. Visual pattern extraction using energy-efficient 2-PCM synapse neuromorphic architecture. *IEEE Trans. Electron Devices*, 59:2206, 2012.
- [56] O. Bichler, W. Zhao, F. Alibart, S. Pleutin, D. Vuillaume, and C. Gamrat. Functional model of a nanoparticle organic memory transistor for use as a spiking synapse. *IEEE Trans. Electron Devices*, 57:3115, 2010.
- [57] Takeo Ohno, Tsuyoshi Hasegawa, Tohru Tsuruoka, Kazuya Terabe, James K. Gimzewski, and Masakazu Aono. Short-term plasticity and long-term potentiation mimicked in single inorganic synapses. *Nature Materials*, 10:591–595, 2011.
- [58] Manan Suri, Damien Querlioz, Olivier Bichler, Giorgio Palma, Elisa Vianello, Dominique Vuillaume, Christian Gamrat, and Barbara DeSalvo. Bio-inspired stochastic computing using binary CBRAM synapses. *IEEE Transactions on Electron Devices*, 60(7):2402–2409, 2013.
-

-
- [59] S. Yu, Y. Wu, R. Jeyasingh, D. Kuzum, and H.-S.P. Wong. An electronic synapse device based on metal oxide resistive switching memory for neuromorphic computation. *IEEE Trans. Electron Devices*, 58:2729, 2011.
- [60] K. Seo, I. Kim, S. Jung, M. Jo, S. Park, J. Park, J. Shin, K.P. Biju, J. Kong, K. Lee, B. Lee, and H. Hwang. Analog memory and spike-timing-dependent plasticity characteristics of a nanoscale titanium oxide bilayer resistive switching device. *Nanotechnology*, 22:254023, 2011.
- [61] S. Park, H. Kim, M. Choo, J. Noh, A. Sheri, S. Jung, K. Seo, J. Park, S. Kim, W. Lee, J. Shin, D. Lee, G. Choi, J. Woo, E. Cha, J. Jang, C. Park, M. Jeon, B. Lee, and H. Hwang. RRAM-based synapse for neuromorphic system with pattern recognition function. *IEDM Tech. Dig.*, pages 231–234, 2012.
- [62] Stefano Ambrogio, Simone Balatti, Federico Nardi, Stefano Facchinetti, and Daniele Ielmini. Spike-timing dependent plasticity in a transistor-selected resistive switching memory. *Nanotechnology*, 24:384012, 2013.
- [63] S. Yu, B. Gao, Z. Fang, H. Yu, J. Kang, and H.-S.P. Wong. A low energy oxide-based electronic synaptic device for neuromorphic visual systems with tolerance to device variation. *Adv. Mater.*, (25):1774, 2013.
- [64] Guo-Qiang Bi and Mu-Ming Poo. Synaptic modifications in cultured hippocampal neurons: Dependence on spike timing, synaptic strength, and postsynaptic cell type. *Journal of Neuroscience*, 18(24):10464–10472, 1998.
- [65] C.Z. Ramos, L.A. Camuñas-Mesa, J.A. Pérez-Carrasco, T. Masquelier, T. Serrano-Gotarredona, and B. Linares-Barranco. On spike-timing-dependent-plasticity, memristive devices, and building a self-learning visual cortex. *Frontiers in Neuroscience*, 5(26):1–22, 2011.
- [66] Teresa Serrano-Gotarredona, Timothée Masquelier, Themistoklis Prodromakis, Giacomo Indiveri, and Bernabé Linares-Barranco. STDP and STDP variations with memristors for spiking neuromorphic learning systems. *Frontiers in Neuroscience*, 7(2):1–15, 2013.
- [67] S. Ambrogio, S. Balatti, A. Cubeta, A. Calderoni, N. Ramaswamy, and D. Ielmini. Statistical fluctuations in HfO_x resistive-switching memory (RRAM): Part I - Set/Reset variability. *IEEE Trans. Electron Devices*, 61(8):2912–2919, 2014.
-

-
- [68] T. y. Liu, T. H. Yan, R. Scheuerlein, Y. Chen, J. K. Lee, G. Balakrishnan, G. Yee, H. Zhang, A. Yap, J. Ouyang, T. Sasaki, A. Al-Shamma, C. Chen, M. Gupta, G. Hilton, A. Kathuria, V. Lai, M. Matsumoto, A. Nigam, A. Pai, J. Pakhale, C. H. Siau, X. Wu, Y. Yin, N. Nagel, Y. Tanaka, M. Higashitani, T. Minvielle, C. Gorla, T. Tsukamoto, T. Yamaguchi, M. Okajima, T. Okamura, S. Takase, H. Inoue, and L. Fasoli. A 130-mm² 2-layer 32-Gb ReRAM memory device in 24-nm technology. *Solid-State Circuits, IEEE Journal of*, 49(1):140–153, 2014.
- [69] S. Gaba, P. Sheridan, J. Zhou, S. Choi, and W. Lu. Stochastic memristive devices for computing and neuromorphic applications. *Nanoscale*, 5(13):5872–5878, 2013.
- [70] C.-Y. Huang, W. C. Shen, Y.-H. Tseng, Y.-C. King, and C.-J. Lin. A Contact-Resistive Random-Access-Memory-Based true random number generator. *IEEE Electron Device Lett.*, 33:1108, 2012.
- [71] W.H. Choi, L.V. Yang, J. Kim, A. Deshpande, G. Kang, J.-P. Wang, and C.H. Kim. A magnetic tunnel junction based true random number generator with conditional perturb and real-time output probability tracking. *IEDM Tech. Dig.*, pages 12.5.1–12.5.4, 2014.
- [72] S. Balatti, S. Ambrogio, Z.Q. Wang, and D. Ielmini. True random number generation by variability of resistive switching in oxide-based devices. *Emerging and Selected Topics in Circuits and Systems, IEEE Journal on*, 5(2):214–221, 2015.
- [73] G. Indiveri and S. Liu. Memory and information processing in neuromorphic systems. *Proc. IEEE*, 103(8):1379–1397, 2015.
- [74] M. Suri, O. Bichler, D. Querlioz, and G. Palma. CBRAM devices as binary synapses for low-power stochastic neuromorphic systems: Auditory (cochlea) and visual (retina) cognitive processing applications. *IEDM Tech. Dig.*, pages 10.3.1 – 10.3.4, 2012.
- [75] Z.-Q. Wang, S. Ambrogio, S. Balatti, and D. Ielmini. A 2-transistor/1-resistor artificial synapse capable of communication and stochastic learning for neuromorphic systems. *Frontiers in Neuroscience*, 8(438), 2015.
- [76] C.F. Stevens. Quantal release of neurotransmitter and long-term potentiation. *Cell Suppl.*, 72:55–63, 1993.
- [77] D. Ielmini. Filamentary-switching model in RRAM for time, energy and scaling projections. *IEDM Tech. Dig.*, page 409, 2011.
-

- [78] E. Chicca et al. Neuromorphic electronic circuits for building autonomous cognitive systems. *Proc. IEEE*, 102(9):1367, 2014.
- [79] S. Balatti, S. Ambrogio, Z.-Q. Wang, S. Sills, A. Calderoni, N. Ramaswamy, and D. Ielmini. Voltage-controlled cycling endurance of HfO_x -based resistive-switching memory (RRAM). *IEEE Trans. Electron Devices*, (accepted), 2015.
- [80] P. J. Sjöström, G. G. Turrigiano, and S. B. Nelson. Rate, timing, and cooperativity jointly determine cortical synaptic plasticity. *Neuron*, 32:1149–1164, 2001.
- [81] G.M. Wittenberg and S. S.-H. Wang. Malleability of spike-timing-dependent plasticity at the CA3-CA1 synapse. *J. Neurosci.*, (26):6610–6617, 2006.
- [82] Giacomo Indiveri, Bernabé Linares-Barranco, Robert Legenstein, George Deligeorgis, and Themistoklis Prodromakis. Integration of nanoscale memristor synapses in neuromorphic computing architectures. *Nanotechnology*, 24:384010, 2013.

Ringraziamenti

Giunti alla fine della tesi, è doveroso da parte mia ringraziare tutti coloro che in questi anni mi hanno aiutato a raggiungere questo importante traguardo.

Innanzitutto ringrazio il Prof. Daniele Ielmini per la possibilità concessami di lavorare in un gruppo di ricerca di alto livello e per i continui spunti volti a rendere il lavoro di ricerca sempre più stimolante e originale.

In secondo luogo ringrazio l'Ing. Stefano Ambrogio per l'attenzione e la pazienza che ha mostrato nei miei confronti in tutti questi mesi, le quali sono state decisive per il completamento di questa tesi e per la crescita della mia conoscenza nel campo della fisica dei dispositivi elettronici. Lo ringrazio anche e soprattutto perchè si è rivelato un buon amico con cui ho trascorso tante piacevoli giornate.

Ringrazio i dottorandi Davide, Mario, Simone, Nicola e Halid per l'ottima accoglienza ricevuta al mio ingresso in laboratorio e per la gentilezza con cui mi hanno trattato in questo anno di tesi.

Ringrazio inoltre Diego, Riccardo, Alessandro T., Gianluca, Luca e Alessandro B. per le belle giornate trascorse in dipartimento nonostante l'intenso lavoro da svolgere.

Un grande ringraziamento va agli amici conosciuti in questi anni di università Pietro T., Pietro P., Michele, Stefano S., Federico, Giacomo, Dario e Nicola con i quali ho vissuto bei momenti che mi porterò sempre dentro e agli amici di una vita Michele e Giuseppe per la loro costante presenza in questi anni.

Ringrazio la Prof.ssa Gritti per avermi trasmesso quando ero un ragazzino il piacere per lo studio delle discipline fisico-matematiche e il Prof. Corbanese per aver sostenuto durante gli anni della scuola media superiore la mia crescita intellettuale con preziosi consigli e insegnamenti.

Infine un ringraziamento sentito va alla mia famiglia che mi ha continuamente incoraggiato e sostenuto nei momenti difficili che si sono presentati in questi anni.