



**UNIVERSITÀ  
DEGLI STUDI  
DI PADOVA**



**DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE**

**CORSO DI LAUREA IN  
INGEGNERIA INFORMATICA**

**Modellizzazione e funzionamento della  
Resistive Random Access Memory (RRAM)**

**Relatore: Prof. Enrico Zanoni**

**Laureando: Marco Tomaiuoli  
1224262**

**ANNO ACCADEMICO 2022 –2023**

**27 settembre 2023**



# Indice

<b>Capitolo 1: Introduzione alla memoria RRAM</b> .....	<b>4</b>
1.0 Memorie e dispositivi a semiconduttore .....	4
1.1 Panoramica sulle nuove memorie .....	6
1.2 RRAM funzionamento .....	7
1.3 Modalità di commutazione .....	8
<b>Capitolo 2: Caratteristiche e modellazione della memoria RRAM</b> ...	<b>10</b>
2.0 Meccanismo chimico-fisico .....	10
2.1 Visualizzazione del filamento conduttivo .....	12
2.2 Modello per la simulazione spice .....	13
<b>Capitolo 3: Organizzazione delle celle di memoria RRAM</b> .....	<b>19</b>
3.0 Architettura 1T1R .....	19
3.1 Lettura/Scrittura nella architettura 1T1R .....	21
3.2 Architettura cross-point .....	23
3.3 Lettura/Scrittura nella architettura cross-point .....	25
3.4 Dispositivi di selezione .....	28
<b>Capitolo 4: Computazione in memoria</b> .....	<b>32</b>
4.0 Introduzione .....	32
4.1 Operazioni logiche .....	33
4.2 Logica switch + RRAM .....	35
4.3 Applicazioni nelle reti neurali .....	38

# CAPITOLO 1: Introduzione alla memoria RRAM

## 1.0 MEMORIE E DISPOSITIVI A SEMICONDUCTORE

Un dispositivo a semiconduttore è un componente elettronico che si basa sulle proprietà elettroniche di un materiale semiconduttore (principalmente silicio), la cui conduttività<sup>1</sup> si trova tra quella dei conduttori<sup>2</sup> e degli isolanti<sup>3</sup>. I materiali semiconduttori sono utilizzati in quanto il loro comportamento può essere facilmente manipolato mediante l'aggiunta di impurità, nota come drogaggio<sup>4</sup>. Il dispositivo a semiconduttore più comune è il MOSFET.

Il MOSFET è un transistor a effetto di campo<sup>5</sup> che ha un gate isolato a cui viene applicata una tensione che determina la conduttività del dispositivo. Questa capacità di modificare la conduttività attraverso la quantità di tensione applicata può essere utilizzata per amplificare o commutare segnali elettronici.

Un floating-gate MOSFET è un tipo di transistor MOSFET a semiconduttore di ossido di metallo in cui il gate è isolato elettricamente. Lo stato 0 e 1 di un bit dipende dal fatto che il floating gate sia caricato o meno. Quando gli elettroni sono presenti sul floating gate, la corrente non può fluire attraverso il transistor e lo stato del bit è 0, quando invece gli elettroni vengono rimossi dal floating gate, la corrente può fluire e lo stato del bit è 1.

Nella figura 1.0.1 viene riportata la struttura del MOSFET, mentre nella figura 1.0.2 viene riportata la struttura di un floating-gate MOSFET.

figura 1.0.1

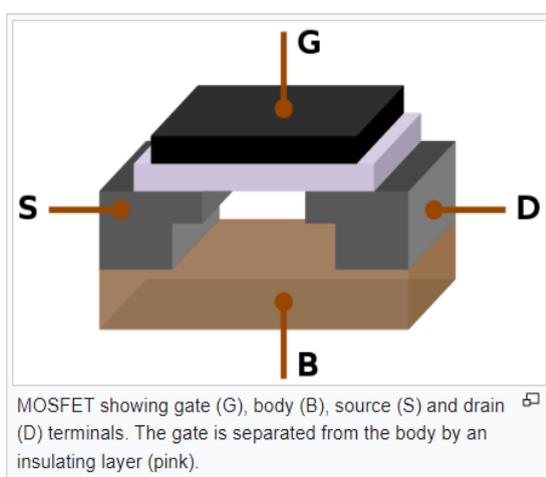
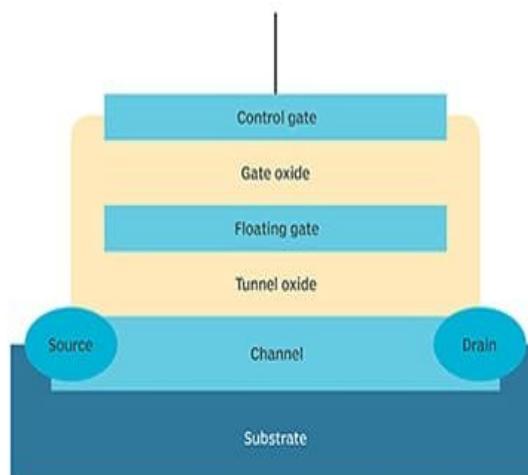


figura 1.0.2



<sup>1</sup> misura della capacità di un materiale a condurre corrente elettrica

<sup>2</sup> materiali nei quali gli elettroni si muovono liberamente (metalli, acqua)

<sup>3</sup> materiali nei quali gli elettroni sono fortemente legati ai nuclei degli atomi (plastica, vetro)

<sup>4</sup> introduzione di impurità in un semiconduttore allo scopo di modularne le proprietà elettriche

<sup>5</sup> transistor che utilizza un campo elettrico per controllare il flusso di corrente in un semiconduttore

La memoria a semiconduttore è un dispositivo elettronico utilizzato per l'archiviazione di dati digitali. In generale si riferisce alla memoria MOS, in cui i dati vengono archiviati all'interno di transistor MOSFET. I due tipi principali di memoria a semiconduttore volatile sono la RAM statica (SRAM) e la RAM dinamica (DRAM), mentre tra le principali memorie non volatili che utilizzano questa caratteristica troviamo la memoria FLASH.

Nella memoria FLASH le celle del chip flash sono disposte in una griglia che ha un transistor a ciascuna intersezione. Ogni transistor ha due porte: una è nota come porta flottante e l'altra è chiamata porta di controllo. Le due porte sono separate l'una dall'altra da un sottile materiale dielettrico denominato strato di ossido. Poiché il gate flottante è isolato elettricamente dallo strato di ossido, tutti gli elettroni posizionati su di esso vengono intrappolati lì. Questo è ciò che rende la memoria flash non volatile.

La memoria DRAM è un tipo di memoria a semiconduttore ad accesso casuale che memorizza ogni bit in una cella di memoria, solitamente costituita da un condensatore e un transistor, entrambi basati sulla tecnologia MOS (semiconduttore di ossido di metallo). Il condensatore può essere carico o scarico, questi due stati rappresentano i due valori dei bit (0 e 1). La carica dei condensatori si perde gradualmente, per ovviare a questo problema la DRAM richiede un circuito di aggiornamento della memoria esterna che riscrive periodicamente i dati nei condensatori riportandoli alla loro carica principale. Quando viene tolta l'alimentazione comunque i dati vengono persi rapidamente.

La memoria statica ad accesso casuale (SRAM) è un tipo di memoria ad accesso casuale<sup>6</sup> che utilizza circuiti latching (flip-flop<sup>7</sup>) per memorizzare ogni bit. Il termine statico differenzia la SRAM dalla DRAM per il fatto che la SRAM non deve essere periodicamente aggiornata. La SRAM è più veloce e più costosa della DRAM, infatti viene in genere utilizzata per la cache e i registri interni della CPU, mentre la DRAM è utilizzata come memoria principale del computer. Poiché la SRAM richiede più transistor per essere implementata, è meno densa e più costosa della DRAM e ha un consumo energetico maggiore durante l'accesso in lettura e scrittura.

---

<sup>6</sup> accesso ad un elemento della memoria in tempo costante indipendentemente dalla dimensione della memoria stessa

<sup>7</sup> è un circuito sequenziale utilizzato come dispositivo di memoria elementare

## 1.1 PANORAMICA SULLE NUOVE MEMORIE

Le memorie tradizionali sono basate sul meccanismo dell'accumulo di carica per memorizzare le informazioni all'interno delle celle della memoria. La SRAM immagazzina la carica nei nodi di stoccaggio degli inverter ad accoppiamento incrociato, la DRAM immagazzina la carica nel condensatore della cella di memoria e la FLASH immagazzina la carica nel gate del transistor.

Negli ultimi decenni sono in fase di ricerca e di sviluppo delle nuove memorie non basate sull'accumulo di carica.

Tra le emergenti NVM (memorie non volatili) risultano molto promettenti la MRAM (Magnetoresistive Random Access Memory) e la RRAM (Resistive Random Access Memory).

Queste memorie sono dispositivi a due terminali non volatili che modificano il loro stato passando dallo stato ad alta resistenza (HRS) a quello a bassa resistenza (LRS).

La MRAM si basa sulla differenza di resistenza tra la configurazione parallela (LRS) e antiparallela (HRS) di due strati ferromagnetici separati da un sottile strato isolante, mentre la RRAM si basa sulla formazione (LRS) e sulla rottura (HRS) di filamenti conduttivi tra due elettrodi separati da un sottile strato di ossido.

In questo contesto, le memorie RRAM spiccano per le loro proprietà quali l'eccellente scalabilità, la facilità di fabbricazione e la notevole resistenza.

tabella 1.1.1: caratteristiche delle memorie tradizionali ed emergenti

	Mainstream Memories				Emerging Memories		
	SRAM	DRAM	FLASH		STT-MRAM	PCRAM	RRAM
			NOR	NAND			
Cell Area	>100F <sup>2</sup>	6F <sup>2</sup>	10F <sup>2</sup>	<4F <sup>2</sup> (3D)	6~20F <sup>2</sup>	4~20F <sup>2</sup>	<4F <sup>2</sup> if 3D
Multi-bit	1	1	2	3	1	2	2
Voltage	<1V	<1V	>10V	>10V	<2V	<3V	<3V
Read Time	~1ns	~10ns	~50ns	~10μs	<10ns	<10ns	<10ns
Write Time	~1ns	~10ns	10μs-1ms	100μs-1ms	<5ns	~50ns	<10ns
Retention	N/A	~64ms	>10y	>10y	>10y	>10y	>10y
Endurance	>1E16	>1E16	>1E5	>1E4	>1E15	>1E9	>1E6~1E12
Write Energy (J/bit)	~fJ	~10fJ	100pJ	~10fJ	~0.1pJ	~10pJ	~0.1 pJ

F: feature size of the lithography, and the energy estimation is on the cell-level (not the array-level)

## 1.2 RRAM FUNZIONAMENTO

Il principio di funzionamento della memoria RRAM si basa sull'utilizzo di un materiale multistrato capace, quando viene applicata una certa tensione, di cambiare la propria resistenza elettrica. Tali cambiamenti vengono utilizzati per codificare i singoli bit (0 e 1) e memorizzare i dati.

La transizione dallo stato ad alta resistenza (HRS) allo stato a bassa resistenza (LRS) viene chiamato SET, mentre l'inverso (LRS - HRS) viene detto RESET.

Applicando al materiale la tensione di SET o di RESET vengono modificate le caratteristiche di resistenza elettrica, formando così un elemento a comportamento commutativo (bassa resistenza risulta acceso, mentre ad alta resistenza risulta spento) adatto a realizzare celle di memoria.

Ad oggi sono stati individuati numerosi ossidi binari, di cui la maggior parte sono ossidi di metalli di transizione<sup>8</sup>, che mostrano un comportamento di commutazione resistivo, uno su tutti l'ossido di afnio ( $\text{HfO}_2$ ).

Infatti i dispositivi RRAM a base di ossido di afnio ( $\text{HfO}_2$ ) hanno mostrato caratteristiche molto promettenti: tempi di commutazione brevi ( $\sim 1\text{ns}$ ), buona resistenza ( $>10^{12}$ ) e conservazione, compatibilità con il processo di fabbricazione CMOS e una struttura molto semplice.

I materiali adatti per lo stato resistivo di ossido di commutazione e gli elettrodi vengono riportati nella tavola periodica sottostante (figura 1.2.1).

Nella figura invece 1.2.2 viene mostrata la struttura base della memoria RRAM, un sottile strato di ossido<sup>9</sup> inserito tra due elettrodi<sup>10</sup>.

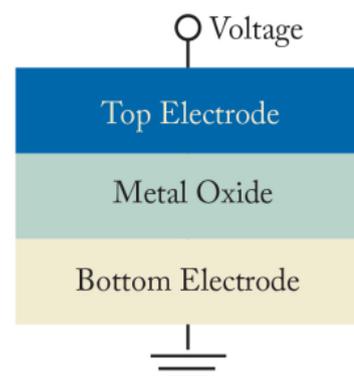
figura 1.2.1

The Periodic Table of the Elements

Legend:

- Yellow box: corresponding binary oxide that exhibits bistable resistance switching
- Blue box: metal that is used for electrode

figura 1.2.2



<sup>8</sup> sono gli elementi del gruppo "d" della tavola periodica.

<sup>9</sup> è un composto chimico binario che si ottiene dalla reazione dell'ossigeno con un altro elemento.

<sup>10</sup> conduttore che trasporta corrente o che crea un campo elettrico in un mezzo.

### 1.3 MODALITA' DI COMMUTAZIONE

Solitamente le memorie RRAM nuove hanno una resistenza molto elevata e quindi è necessario applicare ad esse una grande tensione al primo ciclo affinché si innestino i comportamenti di commutazione per i cicli successivi. Questo è chiamato processo di “forming”.

Le modalità di commutazione della RRAM possono essere classificate in due tipologie: unipolare e bipolare.

La modalità di commutazione è unipolare quando le tensioni SET e RESET hanno la stessa polarità, è bipolare se presenta tensioni opposte.

Nel caso della modalità unipolare la direzione della commutazione dipende dal valore della tensione applicata e non dalla sua polarità, il SET ed il RESET possono avvenire con la stessa polarità. Nel caso della modalità bipolare, invece, la direzione della commutazione dipende dalla polarità della tensione applicata, quindi il SET si può verificare solo ad una polarità, mentre il RESET può avvenire solo con la polarità inversa.

Per entrambe le modalità di commutazione, per evitare una permanente rottura dielettrica nel processo di “forming”, si utilizza una corrente di conformità<sup>11</sup>, solitamente fornita da un dispositivo di selezione della cella (transistor, diodo, o un resistore in serie) on-chip<sup>12</sup>.

Solitamente la modalità unipolare è ottenuta attraverso un metallo nobile come il platino (Pt), utilizzato come elettrodo superiore ed inferiore. Sostituendo uno dei due elettrodi con un materiale ossidabile, come ad esempio il titanio (Ti), si forma la modalità bipolare.

Generalmente, la commutazione unipolare richiede una corrente di ripristino maggiore rispetto alla commutazione bipolare e mostra anche una maggiore variabilità, pertanto oggi la ricerca e lo sviluppo di RRAM si concentrano maggiormente sulla modalità di commutazione bipolare.

Nelle figure 1.3.1 e 1.3.2 vengono mostrati i diagrammi V-I delle due modalità.

figura 1.3.1 (unipolare)

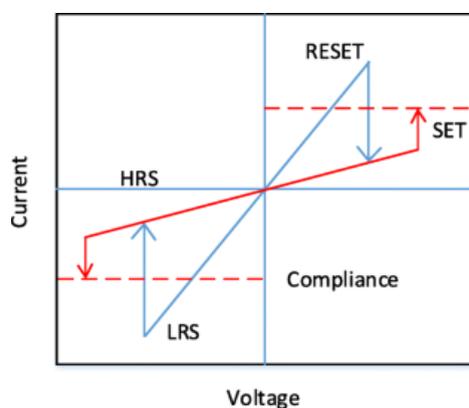
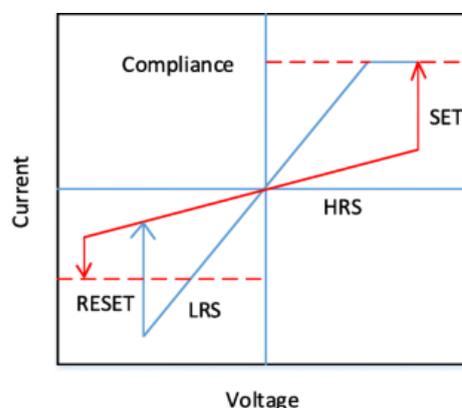


figura 1.3.2 (bipolare)

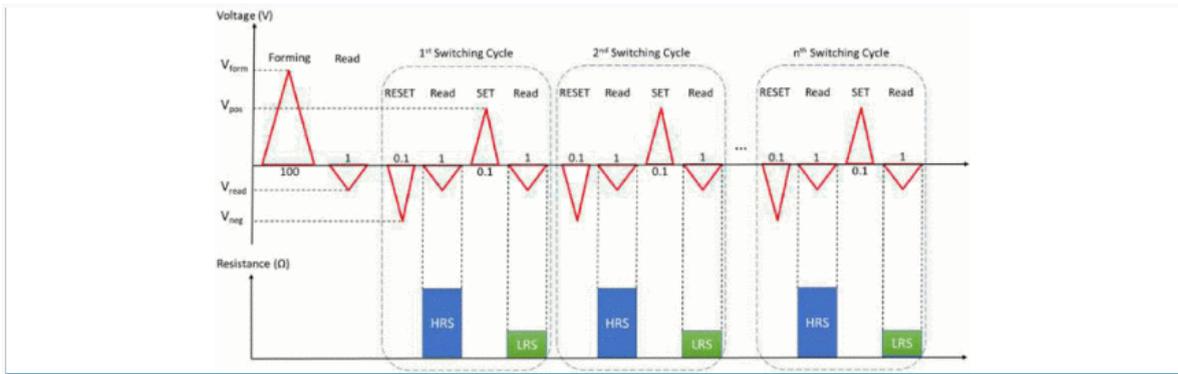


<sup>11</sup> è la corrente massima che può essere applicata attraverso una cella elettrochimica.

<sup>12</sup> è un circuito integrato che in un solo chip contiene un intero sistema.

Nella figura 1.3.3 viene riportato lo schema del tipico ciclo di commutazione ad impulsi dei dispositivi RRAM. Un impulso di lettura di  $-0.2\text{ V}$  è stato applicato dopo ogni impulso SET/RESET per leggere i valori di LRS e HRS, mentre la massima corrente di funzionamento utilizzata durante il “forming” delle celle RRAM è stata controllata dalla modulazione della tensione di gate del transistor NMOS integrato.

figura 1.3.3



Nella figura 1.3.4 invece viene mostrato il processo fisico del fenomeno della commutazione resistiva. Inizialmente lo strato incontaminato di  $\text{HfO}_2$  contiene pochi atomi di ossigeno vacanti<sup>13</sup>.

Successivamente, dopo aver applicato la tensione di “forming”, vengono create sufficienti vacanze di ossigeno che formano un filamento conduttivo che collega l’elettrodo superiore (TE) con l’elettrodo inferiore (BE); il dispositivo è considerato acceso (LRS). Per spegnerlo, invece, si deve applicare una tensione di RESET che provoca la rottura parziale del filamento conduttivo.

L’operazione di SET risulta molto simile al “forming”, poichè riporta il dispositivo da HRS a LRS nuovamente, ma richiede una tensione minore in quanto, dopo il “forming”, parte del filamento conduttivo rimane ancora nel livello di commutazione.

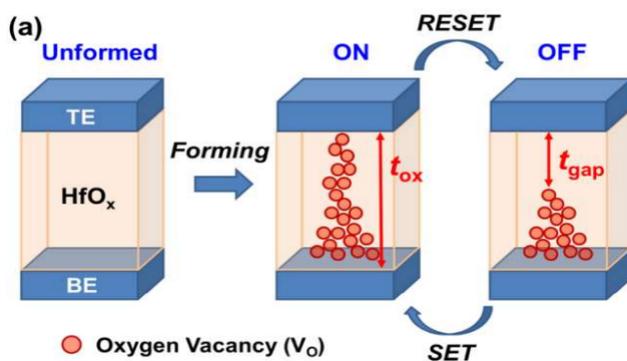


figura 1.3.4

<sup>13</sup> è una perdita di atomo di ossigeno dalla rispettiva posizione nel reticolo cristallino

## **CAPITOLO 2:**

### **Caratteristiche e modellazione della memoria RRAM**

#### ***2.0 MECCANISMO CHIMICO-FISICO***

In generale la commutazione resistiva è associata alla generazione di vacanza di ossigeno ( $V_o$ ) e migrazione di ioni di ossigeno ( $O_2^-$ ) per formare uno o più filamenti conduttivi tra i due elettrodi. Questo processo è solitamente accompagnato da reazioni elettrochimiche, quindi viene anche indicato come effetto redox<sup>14</sup> (riduzione/ossidazione).

Il processo di formatura per i campioni nuovi è simile ad una rottura del dielettrico. Inizialmente la  $V_o$  è bassa. Sotto il controllo di un elevato campo elettrico ( $>10$  MV/cm), gli atomi di ossigeno vengono separati dal reticolo<sup>15</sup> muovendosi verso l'anodo<sup>16</sup> e, quindi, formano  $O_2^-$ , mentre le  $V_o$  rimangono nello strato di ossido.

Nel frattempo la carenza localizzata di ossigeno porta alla formazione del filamento conduttivo (CF) e il dispositivo RRAM si trova in modalità LRS.

Durante il processo di ripristino l' $O_2^-$  migra verso l'ossido per ricombinarsi con il  $V_o$ .

Nella commutazione unipolare il riscaldamento, dovuto dall'effetto Joule da parte della corrente, attiva termicamente la diffusione di  $O_2$ .

Per la commutazione bipolare, lo strato interfacciale può presentare una significativa barriera e, quindi, la pura diffusione termica risulta insufficiente conseguentemente la migrazione di  $O_2^-$  viene aiutata da un campo elettrico inverso.

Tuttavia in entrambi i casi il CF è parzialmente rotto, e le regioni con poche  $V_o$  formano un tunnel per gli elettroni. Ora la RRAM si trova nello stato HRS.

Il residuo del CF con la regione ricca di  $V_o$  viene chiamato "elettrodo virtuale".

---

<sup>14</sup> reazione chimica in cui si ha un passaggio di elettroni da una specie chimica ad un'altra

<sup>15</sup> struttura formata da atomi legati tra loro attraverso l'interazione elettrostatica

<sup>16</sup> elettrodo positivo

Nella figura 2.0.1 viene mostrato il meccanismo di commutazione del filamento conduttivo.

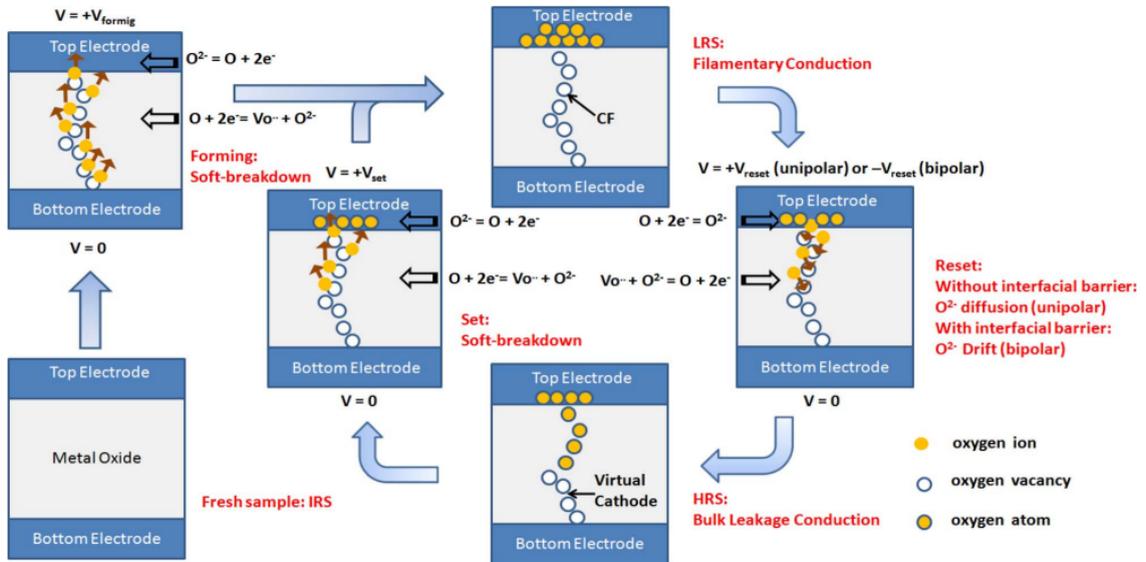


figura 2.0.1

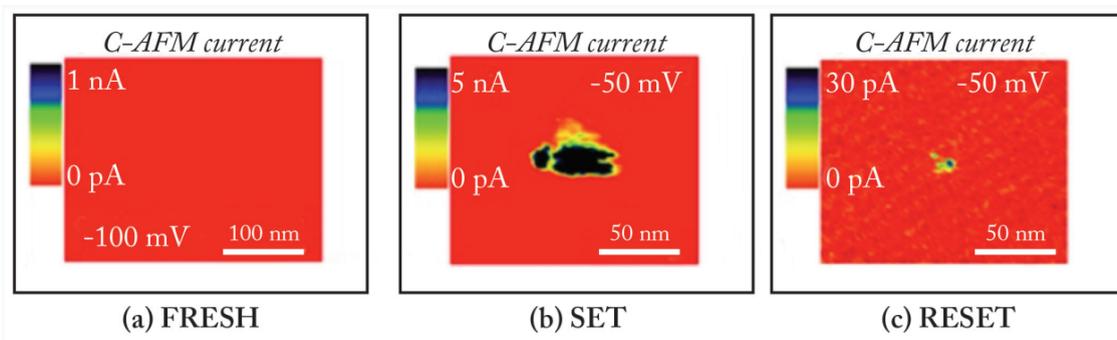
## 2.1 VISUALIZZAZIONE DEL FILAMENTO CONDUTTIVO

La microscopia conduttiva a forza atomica<sup>17</sup> (C-AFM) è uno dei metodi per osservare il CF formato sotto gli elettrodi. Per farlo è stata sviluppata una tecnica che consiste nel rimuovere l'elettrodo dopo le normali operazioni di SET-RESET. Per minimizzare i cambiamenti del CF durante la preparazione del campione, l'elettrodo viene rimosso fisicamente mediante la forza di taglio esercitata durante scansioni ripetute ad alta pressione effettuate attraverso una punta AFM<sup>18</sup> sullo strato dell'elettrodo.

Nella figura 2.1.1 viene mostrata l'immagine ottenuta attraverso C-AFM di una RRAM basata su HfO in stato nuovo, in LRS e in HRS dopo aver rispettivamente rimosso l'elettrodo.

Inizialmente nella cella nuova non è presente nessuna corrente, nella fase di SET invece si forma un filamento da un diametro che varia da 30 a 50 nm e la corrente che lo attraversa è dell'ordine di 1 nA. Nella fase di RESET, invece, viene mostrato il filamento residuo che ha un diametro nettamente più piccolo e varia da 5 a 10 nm e la corrente di dispersione si riduce nell'intervallo dei pA.

figura 2.1.1



<sup>17</sup> modalità della microscopia che misura simultaneamente la topografia del materiale e il flusso di corrente elettrica nel punto di contatto della punta con la superficie del campione

<sup>18</sup> è un tipo di microscopia a scansione di sonda (SPM) ad altissima risoluzione, con una risoluzione dimostrata dell'ordine di frazioni di un nanometro

## 2.2 MODELLO PER LA SIMULAZIONE SPICE

Il modello della RRAM si basa concettualmente sui sistemi memristivi<sup>19</sup> caratterizzati dalle due equazioni sotto riportate.

$$y(t) = h(s, x) * t \quad \text{eq(1)}$$

$$\frac{ds}{dt} = f(s, x) \quad \text{eq(2)}$$

La prima equazione descrive la relazione tra un input variabile nel tempo ( $x$ ) e il suo corrispondente output ( $y$ );  $h$  è una funzione di trasferimento generalizzata che dipende sia dall'input che da una variabile di stato<sup>20</sup> ( $s$ ) che è solitamente un vettore che rappresenta un insieme di condizioni interne del dispositivo.

Per i dispositivi resistivi l'eq(1) è rappresentata dall'equazione I-V:

$$i(t) = g(s, v) * v(t) \quad \text{eq(3)}$$

dove  $v(t)$  rappresenta la tensione arbitraria d'ingresso al dispositivo,  $i(t)$  è la corrente in uscita all'istante  $t$ ,  $g(s, v)$  è la conduttanza generalizzata ed  $s$  è la variabile di stato. Per la RRAM, la variabile di stato  $s$ , normalmente rappresenta la lunghezza o la larghezza della regione conduttiva (CF), mentre la funzione  $f(s, x)$  è spesso una relazione complessa non lineare che cattura i processi fisici che si verificano durante la trasformazione dello stato.

Nella figura 2.2.1 viene mostrata una possibile variabile di stato.

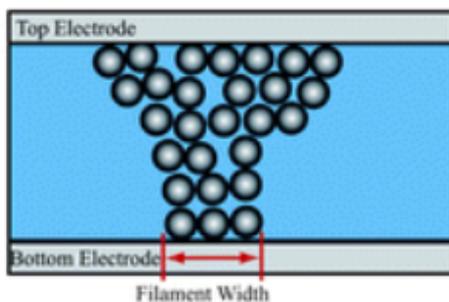


figura 2.2.1

<sup>19</sup>sistema che comprende un circuito, di più componenti convenzionali, che imita le proprietà chiave del componente memristor ideale.

Il memristor ideale è un componente elettrico a due terminali non lineare che mette in relazione la carica elettrica e il collegamento del flusso magnetico

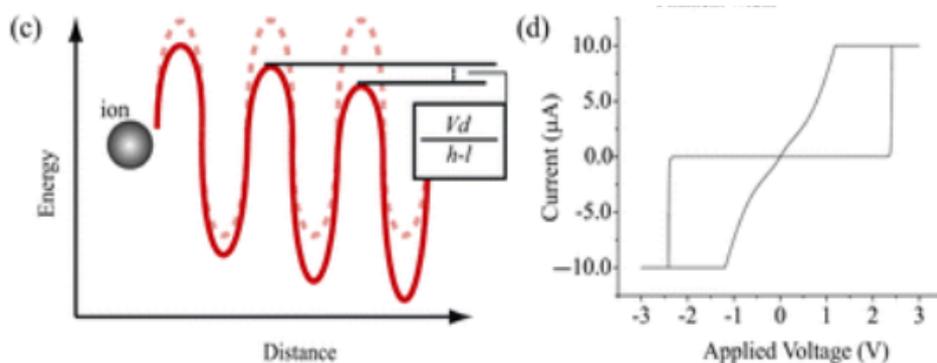
<sup>20</sup> variabile matematica che descrive lo stato di un sistema

Nella maggior parte dei dispositivi RRAM, le transizioni ON/OFF sono causate dal movimento e dalla redistribuzione degli ioni che vanno a formare il filamento conduttivo. Con l'applicazione di una polarizzazione positiva all'elettrodo superiore, mantenendo l'elettrodo inferiore collegato a terra, è possibile avviare la crescita del CF all'interno del materiale isolante. Si presume che il corpo del filamento sia metallico e con bassa resistenza, cosicché, per modellare la crescita del filamento, dobbiamo solo considerare il moto dello ione che si presume muoversi in una sola dimensione, parallela al campo elettrico applicato. Il tasso di crescita del filamento è quindi determinato dalla velocità di deriva<sup>21</sup> dello ione, che può essere derivata calcolando quanto tempo impiega lo ione a saltare oltre una barriera di energia:

$$\frac{ds}{dt} = 2dv * e\left[\left(\frac{-qU}{kT}\right) * \sinh * \left(\frac{qvd}{2kT(h-s)}\right)\right] \quad \text{eq(4)}$$

dove d è la distanza del salto, v la frequenza con cui si verifica un salto, U è il potenziale di attivazione, k è la costante di Boltzmann, T è temperatura in Kelvin, q la carica dell'elettrone, h è lo spessore complessivo di dielettrico e s la lunghezza del filamento.

figura 2.2.2



Nella figura 2.2.2(c) viene mostrata una barriera di energia in un campo elettrico applicato. L'abbassamento della barriera, visto da uno ione in movimento risulta essere  $\frac{qvd}{2(h-s)}$ . Nella figura 2.2.2(d) vengono mostrati i risultati ottenuti dalla simulazione SPICE delle caratteristiche I-V di una cella RRAM con codice elencato nella tabella 2.2.3 e corrente applicata di 10uA.

<sup>21</sup> la media delle componenti delle velocità degli elettroni nella direzione e nel verso della corrente.

Tabella 2.2.3

```

*Parameters:
*area is device area
*h is the film thickness
*Ua is ion barrier height
*Po is electron barrier height
*n_l_init is the initial filament length
*beta is a parameter equal to the attempt frequency multiplied by the hopping distance
*kappa is a paramter that determines the rate dependence on applied voltage
*****
.subckt memristor 2 1 params:
+ area=1e-18 Ua=0.87 h=5e-9 Po=1 n_l_init=0n beta=10.56 kappa=1903.84

.param k=1.38e-23          ;Boltzmann Constant
.param q=1.6021e-19       ;Charge on Electron
.param m=.31*9.1e-31      ;Effective Mass in a-Si
.param plank=6.62606e-34 ;Plank's Constant
.param T 300              ;Operating Temperature in K
*****
*Current Output:
.func logistic(x) {1/(1+exp(-x**3))}

.param alpha=1.4e19
.param F0=2*alpha*sqrt(q*Po**3)/3
.func F(z)=abs(V(1,2))/(h-V(n_l))
.func high_bias(z) 4*3.14*m*q**2/(plank**3*alpha**2*Po)*F(0)**2*exp(-F0/F(0))

.func c1_l(z) alpha*(h-V(n_l))/(2*sqrt(Po*q))
.func low_bias(z) 8*3.14*m*q*(k*T)**2/plank**3*3.14/(c1_l(0)*k*T*sin(3.14*c1_l(0)*k*T))*exp(-alpha*(h-V(n_l))*sqrt(q*Po))*sinh(alpha*(h-V(n_l))*abs(V(1,2))/4*sqrt(q*Po))

Gcurr 1 2 value={sgn(V(1,2))*area*((1-logistic(abs(V(1,2))-Po))*low_bias(0)+logistic(abs(V(1,2))-Po)*high_bias(0))}
*****
*State variables:
Cpvar n_l 0 1

*Initial Conditions:
.ic V(n_l) = n_l_init

*Supplementary Functions:
.func ion_flow(z) { beta*exp(-Ua*q/(k*T))*sinh(V(1,2)*kappa/T) };
.func prev_underflow(z) u( sgn(V(n_l)-1e-12)+sgn(ion_flow(0))+1)

*State Variable Rate Equation:
Gpos 0 n_l value={prev_underflow(0)*ion_flow(0)}

.ends memristor

```

Seguendo i ragionamenti soprastanti, in un dispositivo RRAM in cui il filamento conduttivo non è collegato agli elettrodi, la resistenza sarà dominata da quella tra la punta del filamento e l'elettrodo opposto, con una distanza che dovrebbe essere dell'ordine di pochi nanometri. Per queste distanze si presume che la corrente sia dominata dal tunneling<sup>22</sup> e ne consegue la seguente formula:

$$I = A \frac{4q\pi m(kT)^2}{h^3} * e^{-b1} * \frac{1}{(c1kT)^2} * \frac{\pi c1kT}{\sin(\pi c1kT)} * (1 - e^{(c1qV)}) \quad \text{eq(5)}$$

dove:

$$\frac{2a(h-s)\sqrt{q}}{3V} (\phi^{\frac{3}{2}} - (\phi - V)^{\frac{3}{2}}) \quad \text{se } V < \phi$$

b1 =

$$\frac{2a(h-s)\sqrt{q}}{3V} \phi^{\frac{3}{2}} \quad \text{se } V > \phi$$

e

$$\frac{a(h-s)}{V\sqrt{q}} (\phi^{\frac{1}{2}} - (\phi - V)^{\frac{1}{2}}) \quad \text{se } V < \phi$$

c1 =

$$\frac{a(h-s)}{V\sqrt{q}} \phi^{\frac{1}{2}} \quad \text{se } V > \phi$$

A è l'area del filamento, m è la massa effettiva dell'elettrone, h0 è la costante di Planck,

$\phi$  è l'altezza della barriera a polarizzazione zero e  $a = \frac{2\sqrt{2m}}{h_0}$ .

---

<sup>22</sup> si verifica quando gli elettroni si muovono attraverso una barriera che classicamente non dovrebbero essere in grado di attraversare

Per ridurre la complessità computazionale ai fini della modellazione, l'espressione è stata semplificata:

$$I = Aq \frac{8\pi^2 m}{ho^3} * \frac{kT}{c1\sin(\pi c1kT)} \quad se V < \phi \quad eq(6A)$$

$$I = A \frac{4\pi q^2 m}{ho^3 a^2 \phi} * \left(\frac{V}{h-s}\right)^2 * e^{-\frac{2a\sqrt{q(h-s)\phi^{3/2}}}{3V}} \quad se V > \phi \quad eq(6B)$$

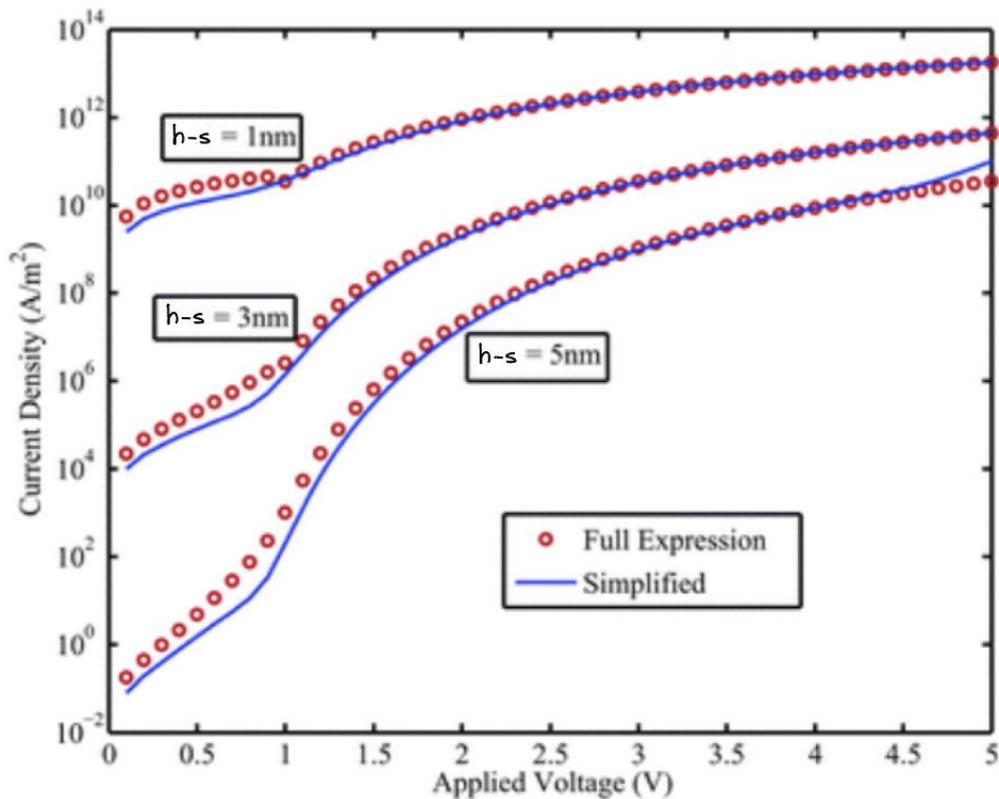


figura 2.2.4

Nella figura 2.2.4 viene mostrato il confronto dell'espressione della corrente di tunneling completa (eq(5)) e la funzione semplificata (eq(6)).

Risolviendo, quindi, l'eq(6), per un dispositivo basato sulla crescita dalla lunghezza del filamento conduttivo, possiamo prevedere completamente il comportamento del dispositivo RRAM. Il software SPICE non consente però variabili mutabili interne, come l'eq(4). Per ovviare a questo, viene creato un sottocircuito in cui la variabile di stato è rappresentata come tensione flottante al nodo tra un condensatore ideale e un generatore di corrente come mostrato in figura 2.2.5. Quindi, il generatore di corrente produce una corrente controllata dal segnale esterno  $V_{in}$  e dalle variabili di stato interne alla RRAM ( $V_n$ ,  $V_{n\_length}$ ,  $V_{n\_width}$ ). L'altro componente del sottocircuito, il resistore variabile<sup>23</sup>, si occupa della corrente attraverso la cella RRAM.

Schema del sottocircuito RRAM

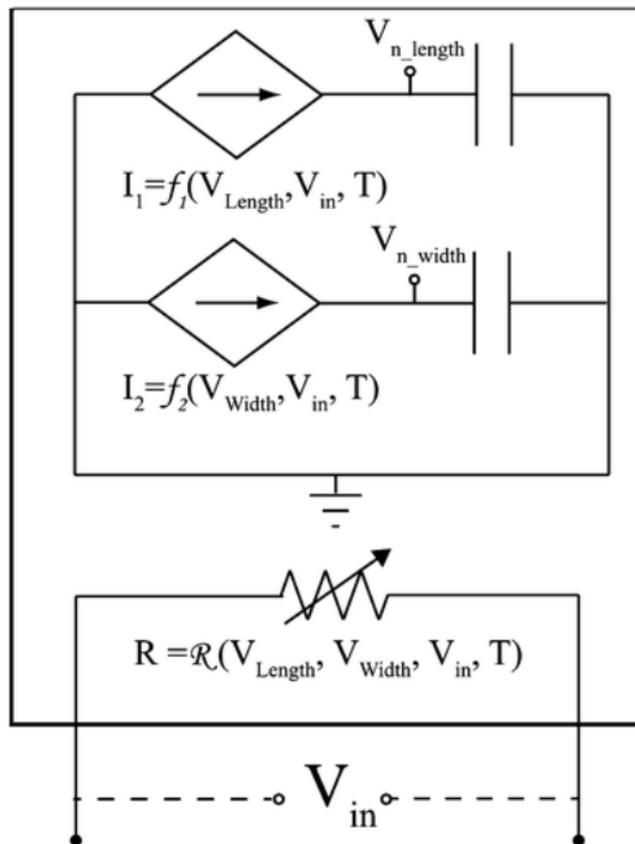


figura 2.2.5

<sup>23</sup> resistore con il quale per mezzo di conduttore mobile strisciante sull'elemento resistivo è possibile variare il valore della resistenza.

## **CAPITOLO 3:**

### **Organizzazione delle celle di memoria RRAM**

#### ***3.0 ARCHITETTURA 1T1R***

Una delle architetture di memoria più comuni per la RRAM è la 1T1R (1 transistor e 1 resistore), dove ogni cella della memoria è composta da un transistor di selezione<sup>24</sup> in serie con un resistore, come mostrato nella figura 3.0.1. Mediante l'utilizzo del transistor di selezione è possibile isolare la cella di memoria selezionata dalle altre.

La memoria RRAM è organizzata in una matrice di celle disposte in riga (Word Line) e colonne (Bit Line) e delle linee di collegamento che vengono utilizzate per connettere i componenti all'interno della memoria.

Ogni cella di memoria corrisponde ad un punto di intersezione tra una riga e una colonna. Nella figura 3.0.2 è riportato lo schema 3D della memoria RRAM.

Le linee di collegamento principali sono: la Source Line (SL), la Word Line (WL) e la Bit Line (BL).

La Source Line (SL) è una delle linee di collegamento collegata al source del transistor di selezione all'interno della cella di memoria, è utilizzata per fornire il potenziale al source del transistor necessario per il suo corretto funzionamento.

La Word Line (WL) è una delle linee di collegamento collegata al gate del transistor di selezione, è responsabile di selezionare e abilitare specifiche celle di memoria nella matrice, consente di eseguire operazioni di lettura e scrittura solamente sulle celle desiderate.

La Bit Line (BL) è una delle linee di collegamento collegata all'elettrodo superiore della cella RRAM, è responsabile di leggere il valore memorizzato nella cella di memoria durante le operazioni di lettura e scrittura.

---

<sup>24</sup> Per leggere i dati in una specifica cella di memoria, viene applicata una tensione al transistor di selezione della determinata cella, questo funge da "interruttore" e crea un percorso di conduzione tra l'elemento di memoria RRAM e il circuito di lettura esterno.

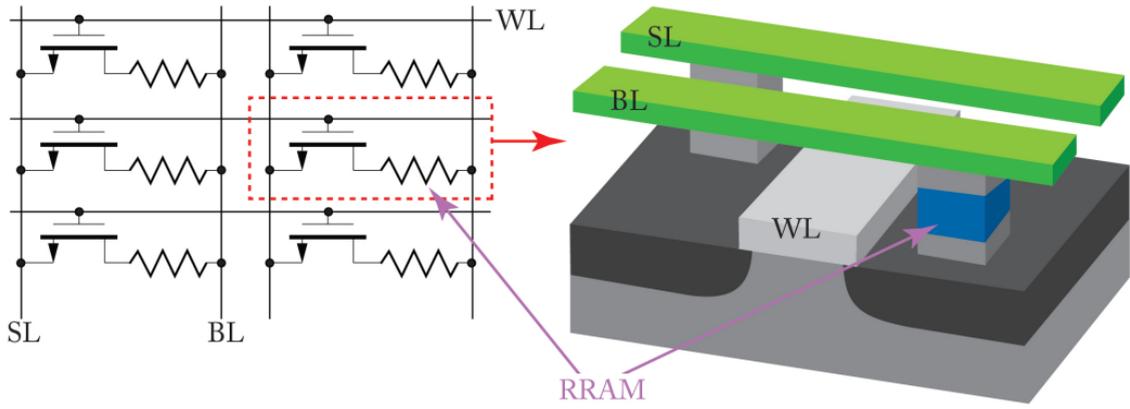


figura 3.0.1

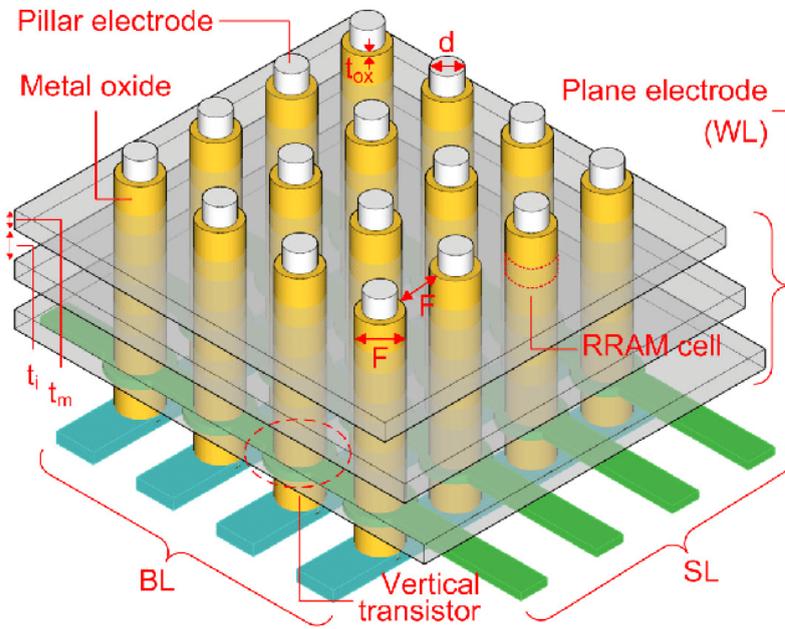


figura 3.0.2

### **3.1 LETTURA/SCRITTURA NELLA ARCHITETTURA 1T1R**

Durante l'operazione di *lettura*, la WL viene attivata e portata ad un livello di tensione specifico (valore che corrisponde alla riga della cella di memoria desiderata). L'attivazione della WL consente di collegare il gate del transistor di selezione all'interno di ogni cella di memoria nella riga selezionata.

Questo processo attiva il transistor di selezione consentendo il passaggio di corrente dalla BL alla SL attraverso la resistenza. In questo modo la corrente, che scorre attraverso la cella, può essere rilevata sulla BL e può essere letto il valore (0 o 1) memorizzato nella cella.

Durante l'operazione di lettura la SL, insieme alla WL, è coinvolta nell'attivazione del transistor per consentire il passaggio di corrente attraverso la cella.

Nell'operazione di *scrittura*, la WL gioca un ruolo simile al processo di lettura. Viene selezionata la riga di interesse attivando la WL corrispondente. Successivamente viene attivato il transistor di selezione, così si accede alla cella desiderata, aprendo un canale di corrente tra BL e SL. Per scrivere il nuovo valore viene applicata una tensione appropriata alla BL. Questa tensione modifica la resistenza dello strato resistivo all'interno della cella RRAM, modificando così il valore (0, 1) presente nella cella.

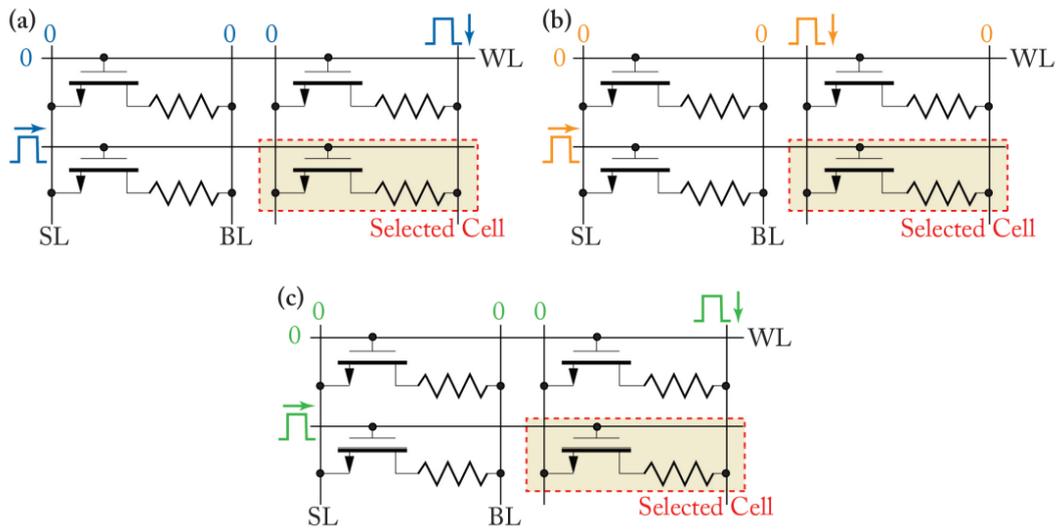


figura 3.1.1 (a) schema nella modalità SET, (b) schema nella modalità RESET, (c) schema della lettura di una cella di memoria RRAM

Nella figura 3.1.1 viene mostrato il tipico schema di lettura/scrittura per la memoria RRAM con architettura 1T1R.

Per l'operazione di SET viene applicata la tensione WL per accendere il transistor della cella selezionata e la tensione di scrittura è applicata alla BL della cella selezionata, mentre SL risulta essere a massa.

Per l'operazione di RESET viene sempre applicata la tensione WL per accendere il transistor di selezione della cella selezionata, mentre la tensione di scrittura viene applicata alla SL, mentre la BL viene collegata a massa per invertire la corrente.

Per leggere i dati dalla memoria, come per l'operazione di SET, una tensione WL viene applicata per accendere il transistor di selezione della cella selezionata e una tensione di lettura viene applicata a BL, mentre SL è collegato a massa.

Essendo spenti i transistor delle celle non desiderate, non ci sono problemi di interferenza e si può accedere ad ogni cella in modo random ed indipendente.

Più bit possono essere scritti/letti in parallelo nella stessa riga attivando più colonne.

Tipicamente la tensione di RESET è maggiore della tensione di SET, perché parte della tensione di ripristino viene dissipata attraverso il materiale resistivo della cella RRAM.

A causa della differenza tra queste due tensioni di WL le operazioni di SET e RESET non possono avvenire simultaneamente nella stessa riga.

### **3.2 ARCHITETTURA CROSS-POINT**

Un'altra architettura di memoria molto comune è quella cross-point. Questa architettura è stata progettata per migliorare la densità di memorizzazione e consentire l'accesso simultaneo a molteplici celle di memoria. Nella matrice cross-point, le celle della memoria RRAM, sono disposte in una griglia composta da righe e colonne perpendicolari tra di loro, ogni punto di intersezione rappresenta una singola cella di memoria. L'architettura cross-point supporta celle di dimensione  $4 F^{2(25)}$  ciò comporta una maggiore densità rispetto all'architettura 1T1R, la quale in media arrivava a  $12 F^2$  e al massimo poteva essere ridotta a  $6 F^2$ .

Tipicamente viene aggiunto un dispositivo selettore in serie alla cella RRAM presso ciascun punto di intersezione per interrompere la corrente di percorso "sneak" delle celle non selezionate, come mostrato in figura 3.2.1.

Lo scopo del dispositivo selettore è quello di controllare l'accesso delle celle di memoria all'interno della matrice e di prevenire il flusso indesiderato di corrente attraverso le celle non selezionate (sneak path current<sup>26</sup>), come mostrato nella figura 3.2.2.

Nelle architetture cross-point, in cui WL e BL si intersecano per formare le celle di memoria, si presenta un problema di interferenza noto come "sneak path problem". Quando si attiva una specifica WL e BL per selezionare una cella di memoria, la tensione applicata può influenzare anche le celle adiacenti, causando un flusso di corrente anche in esse. Ciò può portare a letture e scritture non corrette.

Il dispositivo selettore risolve questo problema agendo come "porta di accesso" tra BL e la cella di memoria. Si trova in serie con la cella RRAM, e quando la cella non è selezionata, il dispositivo selettore interrompe il flusso di corrente, al contrario quando la cella è selezionata il dispositivo selettore si attiva e permette il passaggio di corrente.

---

<sup>25</sup> F è un fattore di scala che rappresenta l'unità base dell'area della cella.

<sup>26</sup> Si verifica quando la tensione applicata per accedere a una cella selezionata genera una corrente indesiderata che attraversa anche celle non selezionate.

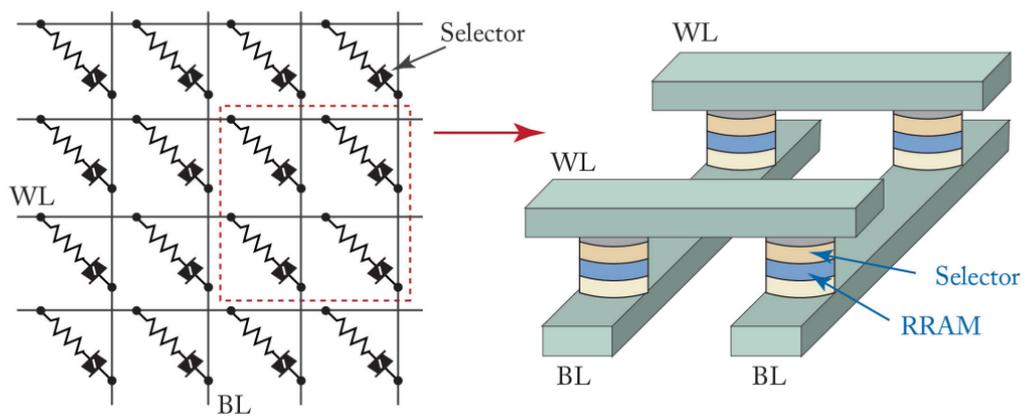


figura 3.2.1 schema dell'architettura cross-point con il selettore in serie alla RRAM per ogni cella

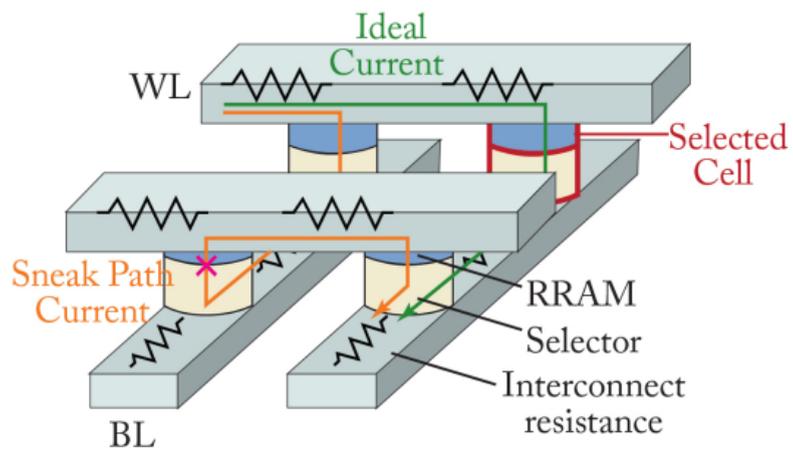


figura 3.2.2 schema dello "sneak path problem"

### 3.3 LETTURA/SCRITTURA NELLA ARCHITETTURA CROSS-POINT

I due schemi lettura/scrittura che vengono utilizzati per l'architettura cross-point sono: la V/2 e la V/3. Figura 3.3.1.

Nello schema V/2, per l'operazione di SET, la WL e la BL della cella selezionata sono polarizzate rispettivamente alla tensione di scrittura  $V_w$  e a massa. Per l'operazione di RESET, invece, le due polarizzazioni sono invertite per la commutazione bipolare.

In entrambe le operazioni tutte le celle non selezionate hanno rispettivamente WL e BL impostato a  $V_w/2$ . Pertanto solo la cella selezionata vede una tensione  $V_w$  completa, mentre le altre celle nella stessa riga (WL) o nella stessa colonna (BL) vedono metà  $V_w$ . Infine le celle non selezionate posizionate in diverse righe e colonne vedono una tensione pari a zero.

Nello schema V/3, per l'operazione di SET, la WL e la BL della cella selezionata sono polarizzate rispettivamente alla tensione di scrittura  $V_w$  e a massa. Per l'operazione di RESET, invece, si invertono le due polarizzazioni, come nella configurazione V/2.

Per le celle non selezionate vengono impostati  $WL = \frac{1}{3} V_w$  e  $BL = \frac{2}{3} V_w$  per l'operazione di SET e  $WL = \frac{2}{3} V_w$  e  $BL = \frac{1}{3} V_w$  per l'operazione di RESET.

Lo schema V/2 consuma meno energia dello schema V/3, in quanto le celle non selezionate nello schema V/2 vedono zero tensione a differenza di  $\frac{1}{3} V_w$  nel caso dello schema V/3. D'altro canto lo schema V/3 ha una migliore resistenza ai disturbi di scrittura dello schema V/2.

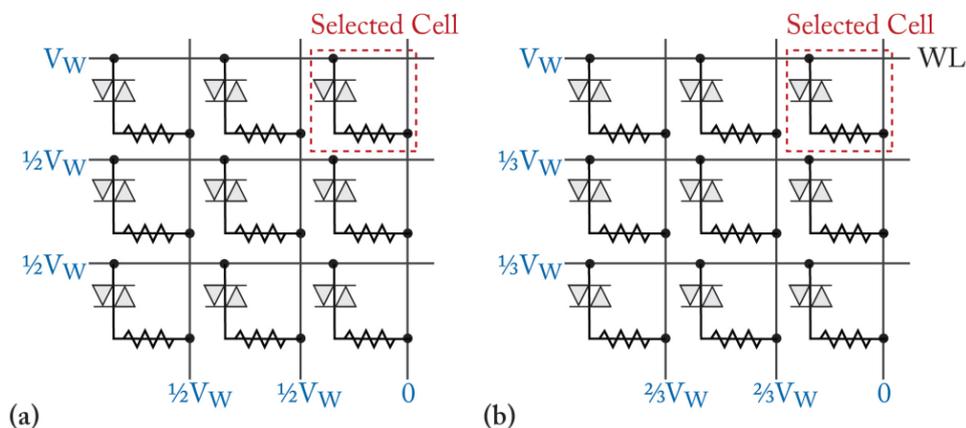


figura 3.3.1 a) schema scrittura V/2 b) schema scrittura V/3

Diversamente da come avviene nell'architettura 1T1R, dove la corrente che attraversa la BL o la SL è indipendente per colonne diverse, nell'architettura cross-point la corrente è comune a tutte le colonne della WL. Questo implica il fatto che le operazioni di SET e RESET non possono avvenire simultaneamente.

Nell'architettura cross-point, è necessario dividere il processo di *scrittura* in due parti quando sono selezionate più celle nella stessa riga.

Il primo metodo separa le operazioni di SET e RESET (come mostrato nella figura 3.3.2). Per esempio per scrivere "0101", prima scriviamo "x1x1", dove x indica che la riga e la colonna della cella corrispondente mantengono i loro stati originali e dopo viene scritto "0x0x".

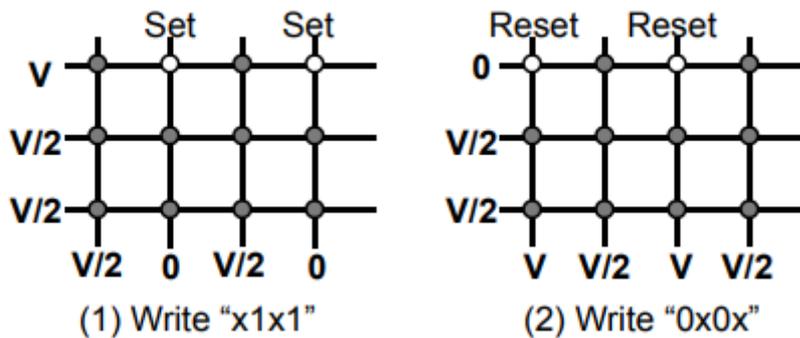


figura 3.3.2

Il secondo metodo elimina tutte le celle della riga selezionata prima dell'operazione di RESET. Per esempio per scrivere "0101" prima facciamo un SET su tutte e 4 le celle desiderate e quindi scriviamo "1111" e poi scriviamo "0x0x". Figura 3.3.3

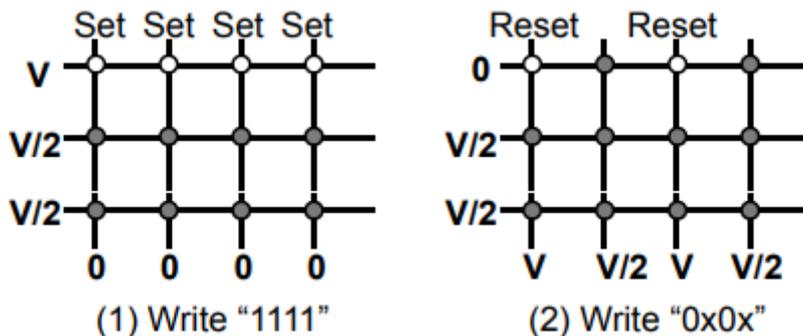


figura 3.3.3

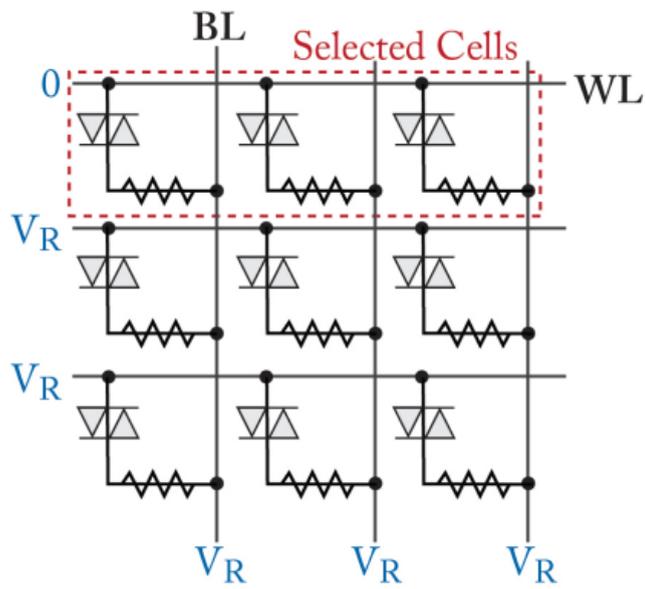


figura 3.3.4

Nella figura 3.3.4 viene mostrato lo schema di *lettura* per una intera riga nell'architettura cross-point. Tutte le colonne sono polarizzate alla tensione di lettura  $V_{read}$ , la riga selezionata è polarizzata a massa e tutte le altre righe non selezionate vengono polarizzate anch'esse a  $V_{read}$ . Perciò solo le celle della riga selezionate vedono una differenza di potenziale, mentre le altre celle non selezionate vedono zero.

### 3.4 DISPOSITIVI DI SELEZIONE

L'architettura cross-point della RRAM presenta, come accennato precedentemente, due problemi principali: l'"IR drop problem" e lo "sneak path problem". Per risolvere i problemi è stato studiato un dispositivo selettore collegato in serie alla RRAM (1S1R)<sup>27</sup>. Il sistema 1S1R viene classificato in base alle caratteristiche del selettore collegato. La figura 3.4.1 mostra tre tipi diversi di sistemi 1S1R. Questi tre sistemi hanno una caratteristica in comune: la corrente in regime di bassa tensione viene soppressa dal dispositivo selettore.

Quando il dispositivo selettore permette il flusso di una corrente diretta<sup>28</sup> elevata per entrambe le polarità di tensione positiva e negativa, la corrente di commutazione del sistema 1S1R viene soppressa solo nel regime di bassa tensione, come mostrato nella figura 3.4.1a. Questo tipo di sistema è chiamato non lineare.

D'altra parte, un sistema 1S1R è chiamato "rectifying cell" quando il dispositivo selettore permette una corrente diretta elevata per una sola polarità di tensione. Questo tipo di cella è ulteriormente suddivisa in celle rectifying-SET e rectifying-RESET, come mostrato nella figura 3.4.1b,c.

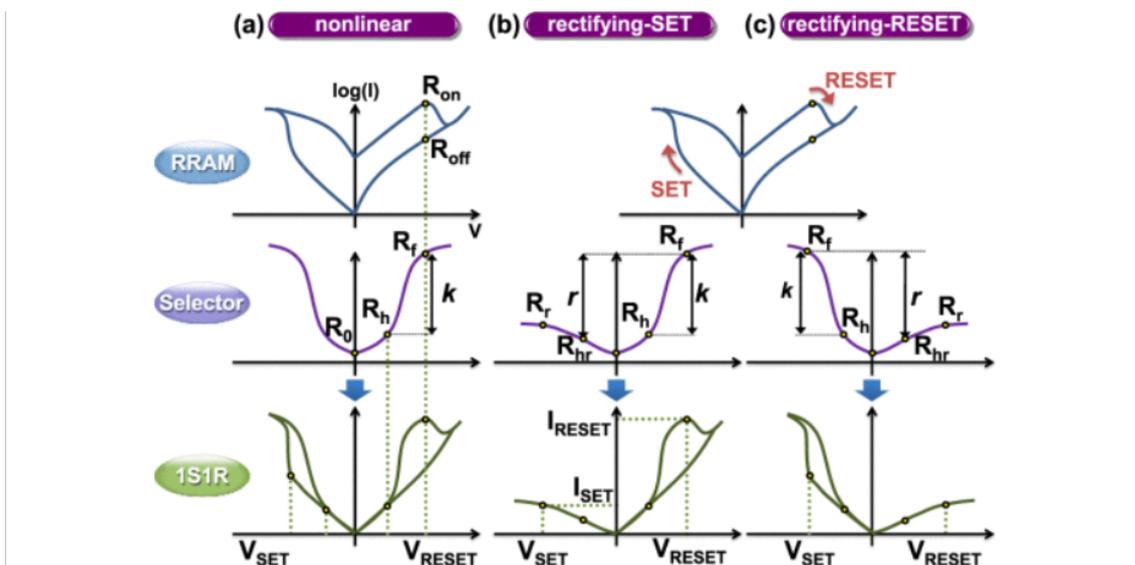


figura 3.4.1 rappresentazione delle 3 diverse tipologie di sistemi 1S1R, a) non lineare, b) rectifying-SET, c) rectifying-RESET

<sup>27</sup> sistema un selettore-una RRAM.

<sup>28</sup> corrente significativamente elevata che scorre attraverso un dispositivo inoltrato.

Il problema dell' *IR drop* diventa significativo quando la larghezza della WL e della BL viene ridotta a meno di 50 nm, in quanto la resistività dell'interconnessione<sup>29</sup> aumenta a causa di fenomeni come lo scattering degli elettroni<sup>30</sup> sulla superficie dei fili.

Questo causa una significativa caduta di tensione sul filo quando viene applicata una corrente che lo percorre, poiché la resistenza del filo è maggiore. Nel caso in cui le matrici di memoria diventano molto elevate, come nel caso di una matrice 1024 x 1204, la cella più lontana sperimenta una resistenza di interconnessione notevole. Se la resistenza della cella RRAM è paragonabile a questa resistenza di interconnessione, una porzione della tensione di scrittura cadrà sul filo anziché sulla cella RRAM.

Per garantire un'operazione di scrittura corretta, la tensione di scrittura deve essere aumentata rispetto alla tensione di commutazione della cella RRAM, per compensare all' *IR drop*.

Tuttavia la tensione di scrittura non può essere aumentata troppo perché  $V_w/2$ , nel caso del modello  $V/2$ , non deve "disturbare" la resistenza delle celle vicino.

Il problema dello *sneak path* è associato a quello dell' *IR drop*. Prendendo come esempio sempre il modello  $V/2$ , le celle semi selezionate lungo WL e BL conducono una corrente e formano *sneak path* durante l'operazione di scrittura. Il problema del percorso nascosto degrada anche il margine di lettura durante l'operazione di lettura: se la cella da leggere si trova nello stato HRS, con celle non selezionate circostanti in LRS, la corrente dello *sneak path* può fluire attraverso le celle circostanti in LRS e defluire dal percorso di lettura, causando una corrente maggiore rispetto alla corrente effettiva della cella nello stato HRS. Quando si legge una cella nello stato HRS, l'obiettivo è rilevare una bassa corrente, poiché l'elemento è in uno stato di alta resistenza, tuttavia a causa dello *sneak path*, la corrente che fluisce attraverso le celle circostanti nello stato LRS contribuisce alla corrente di lettura, generando una nuova corrente maggiore di quella effettiva della cella HRS.

---

<sup>29</sup> Si riferisce alla resistività elettrica dei fili di interconnessione.

<sup>30</sup> Lo scattering degli elettroni è un fenomeno fisico che si verifica quando gli elettroni in un materiale vengono deviati dalla loro traiettoria a causa delle interazioni con altre particelle.

Per sopprimere in parte gli sneak paths, è consigliato aggiungere un selettore con una forte non linearità corrente-tensione alla cella di memoria RRAM. Il dispositivo selettore può essere un diodo (1D1R) o un selettore bidirezionale come nell'architettura 1S1R.

Per la modalità di commutazione unipolare, il diodo p-n è il dispositivo selettore più comune, solitamente un diodo p-n ossido, in quanto offre una maggiore flessibilità. Se il materiale ossido è carente di ossigeno, con una sufficiente quantità di vacanze di ossigeno, si dice di tipo-n, altrimenti se il materiale ossido è carente di metallo, con una sufficiente quantità di vacanze di ossigeno, si dice di tipo-p. La combinazione di un ossido di tipo n e un ossido di tipo p genera un diodo p-n. Figura 3.4.2.

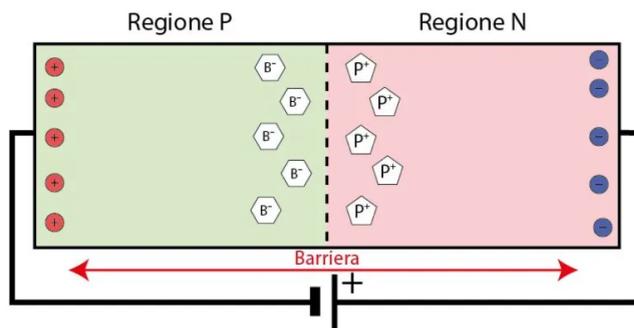


figura 3.4.2

Per la modalità di commutazione bipolare è richiesto un dispositivo non lineare bi-direzionale, come per esempio il selettore bidirezionale Ni/TiO<sub>2</sub>/Ni, integrato con la RRAM bipolare Ni/HfO<sub>x</sub>/Pt. Le caratteristiche corrente-tensione altamente non lineari sono realizzate dall'emissione Schottky<sup>31</sup> sulle barriere Ni/TiO<sub>2</sub>. Il collegamento in serie con un dispositivo di memoria resistivo HfO<sub>2</sub>, mostra una commutazione resistiva bipolare riproducibile.

<sup>31</sup> Scarica di elettroni dalla superficie di un materiale riscaldato mediante l'applicazione di un campo elettrico che riduce il valore dell'energia necessaria per l'emissione di elettroni.

Nella tabella 3.4.3 vengono comparati diversi dispositivi selettori secondo: intervalli di tensione, rapporto di non linearità e corrente.

Type	Stack	Voltage Range	Current Drivability	Rectify Ratio
p-n diode	p-CuO/n-InZnO [84]	-2V~+2V	$3 \times 10^4$ A/cm <sup>2</sup>	$3 \times 10^4$
Schottky diode	Pt/TiO <sub>2</sub> /Ti/Pt [85]	-2V~+2V	$3 \times 10^5$ A/cm <sup>2</sup>	$2.4 \times 10^6$
Bi-directional selector	Ni/TiO <sub>2</sub> /Ni [86]	-4V~+4V	$10^5$ A/cm <sup>2</sup>	$10^3$
Bi-directional selector	Pt/TaO <sub>x</sub> /TiO <sub>2</sub> /TaO <sub>x</sub> /Pt [87]	-2.5V~+2.5V	$3.2 \times 10^7$ A/cm <sup>2</sup>	$10^4$
Bi-directional selector	MIEC [90]	-1.6V~+1.6V	$50 \times 10^6$ A/cm <sup>2</sup>	$10^4$
Threshold selector	Pt/VO <sub>2</sub> /Pt [92]	-0.5V~+0.5V	$6 \times 10^6$ A/cm <sup>2</sup>	50
Threshold selector	TiN/NbO <sub>2</sub> /W [94]	-1V~+1V	$10 \times 10^6$ A/cm <sup>2</sup>	50
Threshold selector	FAST [95]	-1V~+1V	$5 \times 10^6$ A/cm <sup>2</sup>	$10^7$

tabella 3.4.3

# CAPITOLO 4: Computazione in memoria

## 4.0 INTRODUZIONE

L'architettura di von Neumann è caratterizzata dal fatto che le unità di elaborazione e memorizzazione sono indipendenti e sono collegate tramite bus. Tuttavia, l'architettura tradizionale ha riscontrato alcuni problemi: da un lato, la migrazione dei dati attraverso il bus consuma molta energia, d'altra parte la differenza nella velocità di elaborazione dei dati è diventata notevolmente maggiore della velocità di elaborazione della memoria. La mancata corrispondenza tra processore e memoria comporta un grande spreco di risorse. Per risolvere questi problemi è stato sviluppato il concetto di computazione in memoria.

La computazione in memoria (CIM) è un paradigma informatico in cui i dati sono archiviati ed elaborati direttamente nella memoria principale del sistema, anziché essere trasferiti avanti ed indietro tra la memoria e l'unità di elaborazione centrale (CPU).

La memoria, quindi, non viene utilizzata unicamente per memorizzare i dati, ma anche per eseguire i calcoli su di essi. Con una tale dualità memoria-calcolo, il calcolo in memoria basato su RRAM consente operazioni aritmetiche altamente parallele ed efficienti dal punto di vista energetico. Tra i diversi supporti di archiviazione, la memoria RRAM è un candidato promettente per raggiungere il pieno potenziale del CIM per le reti neurali profonde, grazie alla sua alta densità, non volatilità e compatibilità con la tecnologia CMOS.

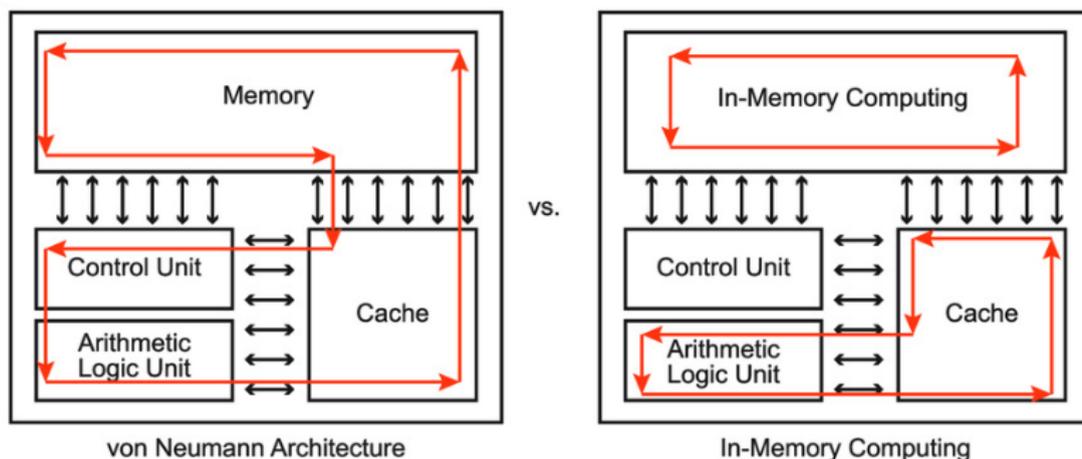


figura 4.0.1 Schema dell'architettura di Von Neumann (sinistra) rispetto al calcolo in memoria (destra)

## 4.1 OPERAZIONI LOGICHE

Nel 2008 è stato preparato per la prima volta un modello fisico di RRAM che implementa le operazioni logiche IMPLY<sup>32</sup>(figura 4.1.0).

Le operazioni IMPLY e NOT possono formare un insieme logico completo<sup>33</sup>, tuttavia questa combinazione potrebbe risultare inefficiente in alcune situazioni. Ad esempio l'uso dell'implicazione logica potrebbe sovrascrivere i valori di input iniziali o potrebbe richiedere più passaggi rispetto alle più dirette porte AND e OR. Si è passati, quindi, ad una struttura costituita da tre celle RRAM e un resistore, dove lo stato ad alta resistenza rappresenta l'1 logico e lo stato a bassa resistenza lo 0. P e Q rappresentano le variabili d'ingresso, mentre S la variabile d'uscita (Figura 4.1.1).

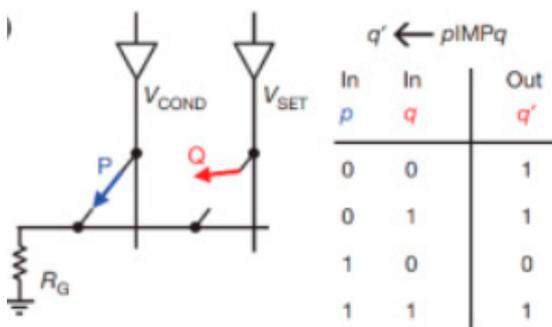


Figura 4.1.0

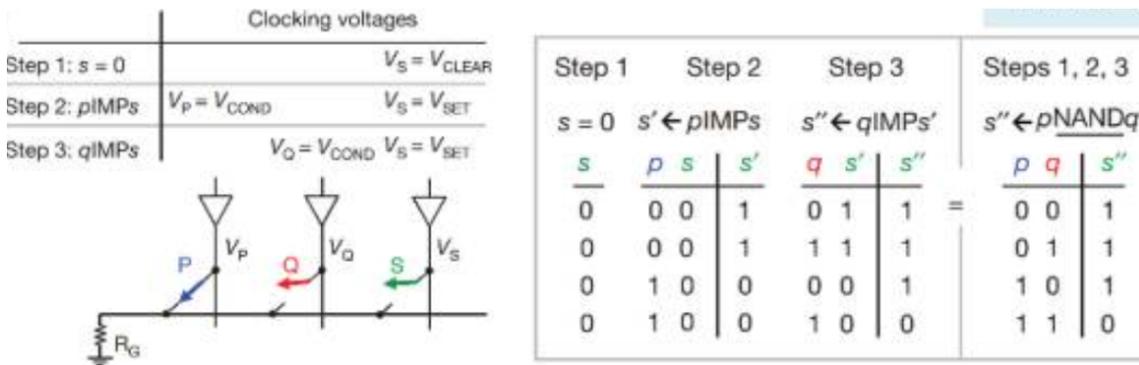
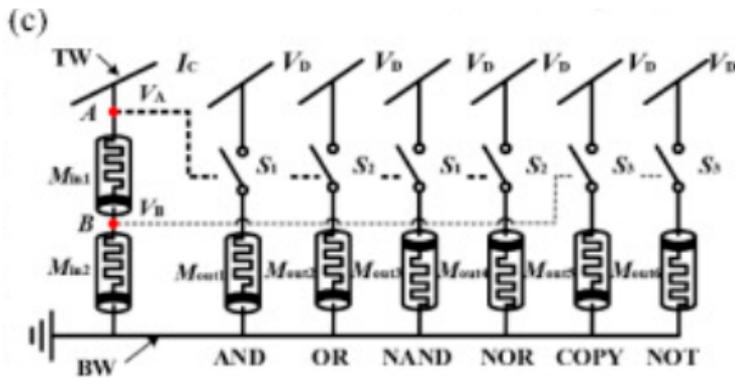


Figura 4.1.1

<sup>32</sup> implicazione logica

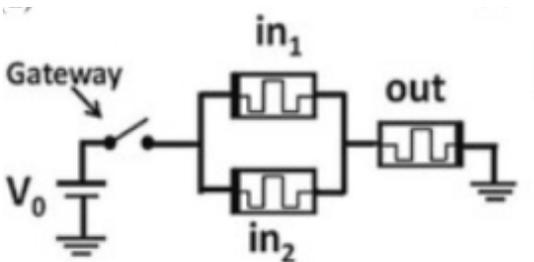
<sup>33</sup> usando solo queste due operazioni, è possibile eseguire qualsiasi operazione logica che coinvolge due variabili booleane e ottenere tutti i possibili risultati.

Combinando più resistori logici in un singolo circuito per implementare operazioni logiche, come mostrato nella figura 4.1.2, si offre la possibilità di implementare architetture avanzate. Nell'esempio sottostante vengono implementate le operazioni AND, OR, NOR, NAND, COPY, NOT. In ciascuna porta logica, il circuito utilizza RRAM come stati di ingresso ed uscita, consentendo così operazioni logiche basate sugli stati. A differenza di altri metodi esistenti, questo approccio genera un circuito logico di stato versatile che può eseguire contemporaneamente molteplici operazioni logiche di stato.



4.1.2

Sotto il controllo della tensione applicata, le cinque funzioni logiche essenziali possono essere ottenute collegando le RRAM in serie o in parallelo. Questo approccio progettuale, consente di memorizzare i risultati dei calcoli logici in una cella separata, evitando il problema della sovrascrittura dei dati derivante dall'operazione logica IMPLY. Figura 4.1.3



4.1.3

Utilizzando le porte RRAM e la struttura ad cross-point è possibile realizzare operazioni logiche complesse. Ad esempio quello mostrato nella figura 4.1.4 dove viene eseguita la funzione X OR (Y NOR Z) in un unico passaggio con tre interruttori memristivi.

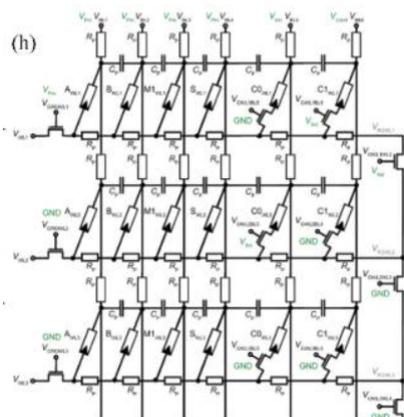


Figura 4.1.4

## 4.2 LOGICA SWITCH + RRAM

Nella figura 4.2.0 viene mostrato uno schema RRAM + Switch. In questo caso all'interruttore viene assegnato uno stato logico S, mentre alla tensione applicata alla RRAM viene assegnato uno stato logico V. Entrambi gli stati possono essere utilizzati come segnali logici di ingresso. Se l'interruttore è aperto il valore di S è 1, se è chiuso il valore di S è 0. In un sistema associato ad una caratteristica unipolare, il valore di V è definito come 1 quando la tensione applicata al dispositivo è maggiore o uguale alla tensione impostata, in quanto trasforma il dispositivo come LRS. Se la tensione applicata al dispositivo è maggiore della tensione di ripristino ed inferiore alla tensione impostata, il valore di V è 0 perché rende lo stato del dispositivo come HRS. Allo stesso modo, in un sistema con caratteristiche bipolari, assegniamo 1 al valore di V se la tensione applicata al dispositivo è maggiore della tensione impostata ed assegniamo 0 a V se la tensione è inferiore alla tensione di ripristino.

Quattro possibili combinazioni di interruttore e tensione applicata sono SV = 00, 01, 10 e 11. SV = 10, 11 sono stati aperti dall'interruttore, quindi è stata applicata tensione zero quando è stato applicato l'impulso. Pertanto, SV=10, 11 non può causare il cambiamento di stato. Solo SV=00, 01 può portare a modificare lo stato del dispositivo. Gli stati logici 0 e 1 possono essere assegnati al dispositivo se è rispettivamente in HRS e LRS.

Utilizzando il dispositivo RRAM Ni/SiN/Si con caratteristiche sia unipolari che bipolari, implementiamo funzioni booleane per ciascuno. Tutti gli operatori tranne NOR e XNOR possono essere rappresentati da questo sistema combinato.

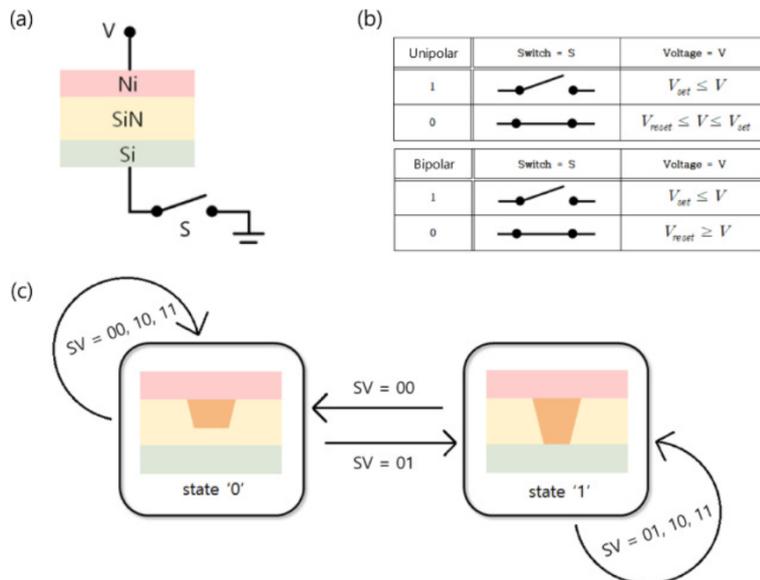


Figura 4.2.0

Questo sistema può essere espresso matematicamente dall'equazione:

$$Z = (S \text{ OR } V) \times Z0 + (S \text{ RNIMP } V) \times (\text{NOT } Z0) \quad (\text{Figura 4.2.1})$$

Dove lo stato corrente è Z e lo stato precedente Z0.

14 delle 16 funzioni booleane possono essere implementate se il sistema viene ripetuto tre volte o meno. Tutte le funzioni sono mostrate nella tabella 4.2.2

Esempio NAND:

Inizialmente, un'uscita rappresentata come 'y' inizia con 0 e quando viene immesso un ingresso SV=01, y diventa 1. Inoltre, dove l'ingresso è SV=00 y diventa 0. Si vede che quando S=1, lo stato del dispositivo non cambia affatto. Se applichiamo questo sistema a un particolare operatore, ad esempio, nella prima fase del funzionamento NAND, assumiamo che lo stato originale sia 0. SV = 00 viene applicato in modo che lo stato logico del dispositivo rimanga 0. Nel secondo passaggio, SV = X1. Se X = 1, il dispositivo mantiene lo stato 0, se X=0 diventa 1. Nell'ultimo passaggio viene applicato SV = Y1. Se lo stato precedente del dispositivo era dichiarato 1, mantiene lo stato 1 indipendentemente da Y, se lo stato precedente del dispositivo era dichiarato 0, mantiene lo stato 0 quando Y=1 o passa allo stato 1 quando Y=0.

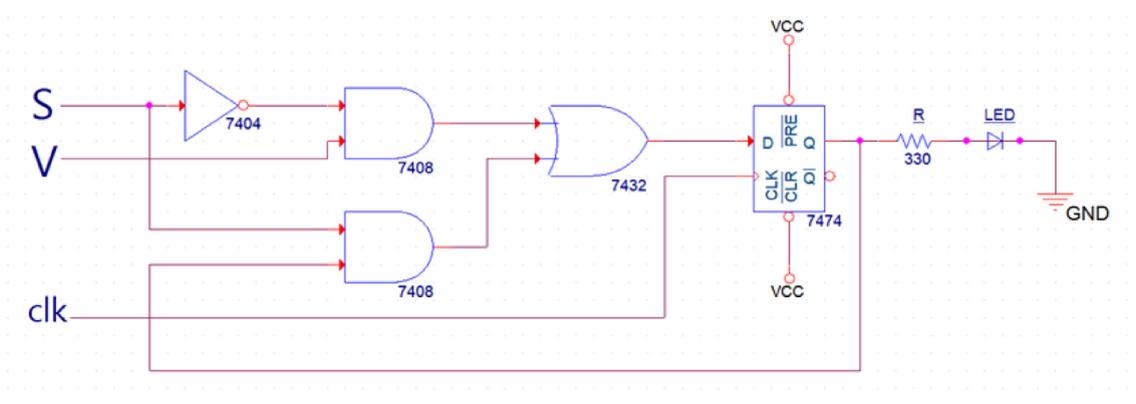


Figura 4.2.1

Boolean Function	input		step 1			step 2			step 3			output	
	X	Y	S	V	SV	S	V	SV	S	V	SV	state of device	
True	0	0	0	1	01							1	[b]
	0	1	0	1	01							1	[b]
	1	0	0	1	01							1	[b]
	1	1	0	1	01							1	[b]
False	0	0	0	0	00							0	[a]
	0	1	0	0	00							0	[a]
	1	0	0	0	00							0	[a]
	1	1	0	0	00							0	[a]
X	0	0	0	X	00							0	[a]
	0	1	0	X	00							0	[a]
	1	0	0	X	01							1	[b]
	1	1	0	X	01							1	[b]
Y	0	0	0	Y	00							0	[a]
	0	1	0	Y	01							1	[b]
	1	0	0	Y	00							0	[a]
	1	1	0	Y	01							1	[b]
NOT X	0	0	0	0	00	X	1	01				1	[f]
	0	1	0	0	00	X	1	01				1	[f]
	1	0	0	0	00	X	1	11				0	[d]
	1	1	0	0	00	X	1	11				0	[d]
NOT Y	0	0	0	0	00	Y	1	01				1	[f]
	0	1	0	0	00	Y	1	11				0	[d]
	1	0	0	0	00	Y	1	01				1	[f]
	1	1	0	0	00	Y	1	11				0	[d]
X AND Y	0	0	0	X	00	Y	0	00				0	[c]
	0	1	0	X	00	Y	0	10				0	[d]
	1	0	0	X	01	Y	0	00				0	[h]
	1	1	0	X	01	Y	0	10				1	[e]
X OR Y	0	0	0	1	01	X	Y	00				0	[h]
	0	1	0	1	01	X	Y	01				1	[g]
	1	0	0	1	01	X	Y	10				1	[e]
	1	1	0	1	01	X	Y	11				1	[e]
X XOR Y	0	0	0	0	00	X	Y	00	Y	X	00	0	[i]
	0	1	0	0	00	X	Y	01	Y	X	10	1	[k]
	1	0	0	0	00	X	Y	10	Y	X	01	1	[l]
	1	1	0	0	00	X	Y	11	Y	X	11	0	[j]
X NAND Y	0	0	0	0	00	X	1	01	Y	1	01	1	[m]
	0	1	0	0	00	X	1	01	Y	1	11	1	[k]
	1	0	0	0	00	X	1	11	Y	1	01	1	[l]
	1	1	0	0	00	X	1	11	Y	1	11	0	[j]
X NIMP Y	0	0	0	0	00	Y	X	00				0	[c]
	0	1	0	0	00	Y	X	10				0	[d]
	1	0	0	0	00	Y	X	01				1	[f]
	1	1	0	0	00	Y	X	11				0	[d]
X RIMP Y	0	0	0	X	00	Y	1	01				1	[f]
	0	1	0	X	00	Y	1	11				0	[d]
	1	0	0	X	01	Y	1	01				1	[g]
	1	1	0	X	01	Y	1	11				1	[e]
X RNIMP Y	0	0	0	0	00	X	Y	00				0	[c]
	0	1	0	0	00	X	Y	01				1	[f]
	1	0	0	0	00	X	Y	10				0	[d]
	1	1	0	0	00	X	Y	11				0	[d]
X IMP Y	0	0	0	Y	00	X	1	01				1	[f]
	0	1	0	Y	01	X	1	01				1	[g]
	1	0	0	Y	00	X	1	11				0	[d]
	1	1	0	Y	01	X	1	11				1	[e]

Tabella 4.2.2

### 4.3 APPLICAZIONI NELLE RETI NEURALI

Negli ultimi anni, la RRAM, è stata ampiamente utilizzata nelle reti neurali. La produzione continua di grandi quantità di dati consente di addestrare e gestire con successo reti neurali artificiali. Tuttavia, per ottenere un'elevata accuratezza, le reti neurali di solito richiedono un gran numero di parametri.

La struttura tradizionale della memoria e dei punti di calcolo fa sì che venga consumato molto tempo ed energia, ma si prevede che l'utilizzo di RRAM per costruire reti neurali risolva questo problema.

Un singolo dispositivo RRAM può essere utilizzato come sinapsi e integra contemporaneamente l'elaborazione e l'archiviazione, eliminando quindi la trasmissione del dato.

Nella figura 4.3.0 vengono mostrate due matrici 20 x 20 di RRAM integrate a componenti CMOS per implementare una rete neurale MLP. Nella figura 4.3.1 viene mostrata la rete composta da: 16 neuroni di input, 10 neuroni nell'hidden layer e 4 neuroni di output. La rete è in grado di classificare modelli in bianco e nero di 4 x 4 pixel, come mostrato nella figura 4.3.2.

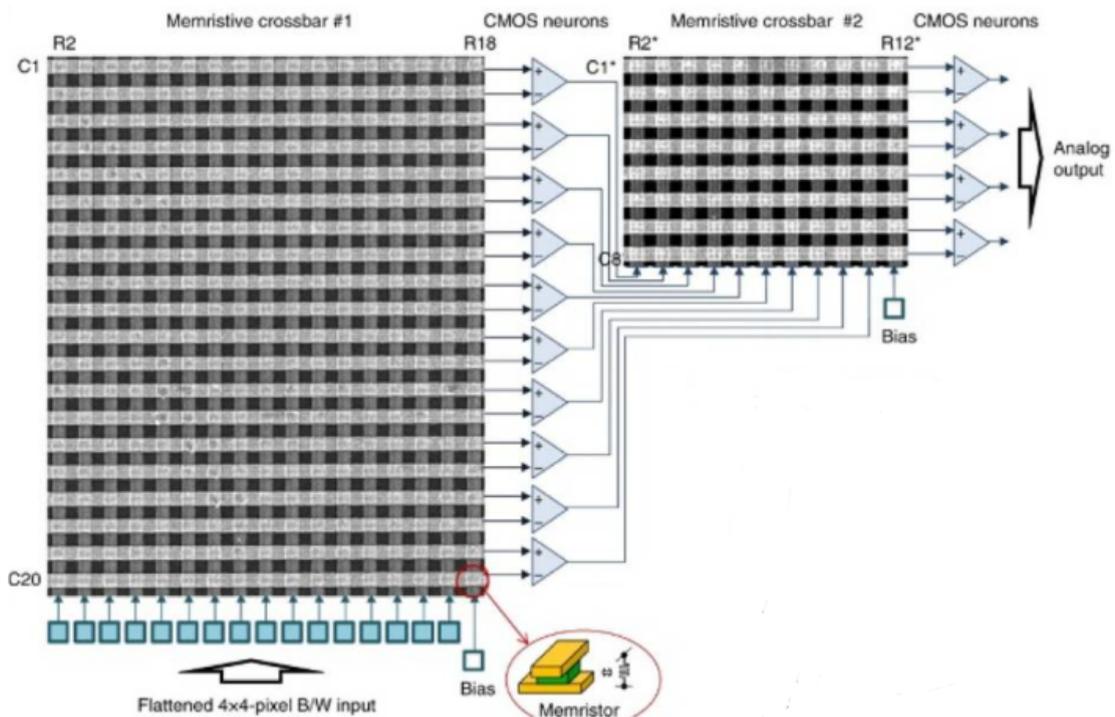


Figura 4.3.0

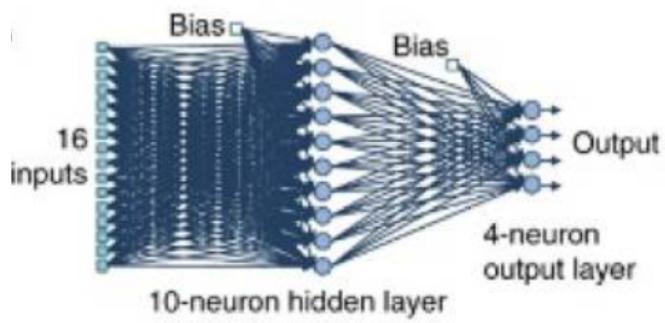


Figura 4.3.1

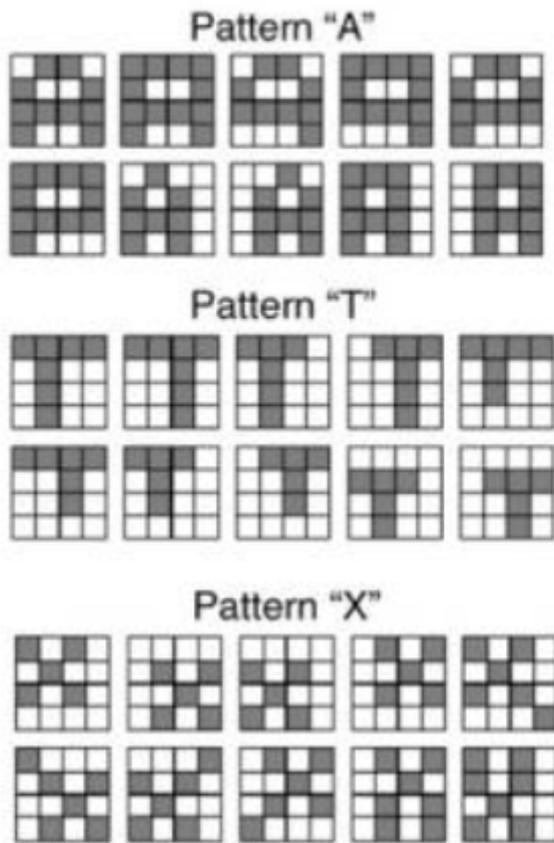


Figura 4.3.2



# Bibliografia

- [1] "Resistive Random Access Memory (RRAM), From Devices to Array Architectures". Shimeng Yu.
- [2] "Resistive random access memory: doping technology, pulse characterization and scalability". Liang Zhao.
- [3] "In-Memory Learning With Analog Resistive Switching Memory: A Review" and Perspective. Yue Xi, Bin Gao, Jianshi Tang, An Chen, Meng-Fang Chan, Xiaobo Sharon Hu, Jan Van Der Spiegel, He Qian and Huaqiang wu.
- [4] Wikipedia: <https://en.wikipedia.org/wiki/MOSFET>.
- [5] "Floating gate transistor (FGT)". Kinza Yasar.  
<https://www.techtarget.com/searchstorage/definition/floating-gate>.
- [6] H. . -S. P. Wong et al., "Metal–Oxide RRAM," in Proceedings of the IEEE, vol. 100, no. 6, pp. 1951-1970, June 2012, doi: 10.1109/JPROC.2012.2190369.
- [7] "Impact of vertical RRAM device characteristics on 3D cross-point array design". Pai-Yu Chen, Shimeng Yu.
- [8] S. Kim, J. Zhou and W. D. Lu, "Crossbar RRAM Arrays: Selector Device Requirements During Write Operation," in IEEE Transactions on Electron Devices, vol. 61, no. 8, pp. 2820-2826, Aug. 2014, doi: 10.1109/TED.2014.2327514.
- [9] "Paleotecnologia V: La giunzione p-n". Fabrizio Bianchini.
- [10] J. -J. Huang, Y. -M. Tseng, C. -W. Hsu and T. -H. Hou, "Bipolar Nonlinear Ni/TiO/Ni Selector for 1S1R Crossbar Array Applications," in IEEE Electron Device Letters, vol. 32, no. 10, pp. 1427-1429, Oct. 2011, doi: 10.1109/LED.2011.2161601.
- [11] "Logic-in-memory application of CMOS compatible silicon nitride memristor". Dahyne Kim, Sunghum Kim, Sungjum Kim.
- [12] C. -W. S. Yeh and S. S. Wong, "Compact One-Transistor-N-RRAM Array Architecture for Advanced CMOS Technology," in IEEE Journal of Solid-State Circuits, vol. 50, no. 5, pp. 1299-1309, May 2015, doi: 10.1109/JSSC.2015.2402217.
- [13] Y. -D. Chen and A. Chin, "An Offset Readout Current Sensing Scheme for One-Resistor RRAM-Based Cross-Point Array," in IEEE Electron Device Letters, vol. 40, no. 2, pp. 208-211, Feb. 2019, doi: 10.1109/LED.2018.2886552.
- [14] "Design Implications of Memristor-Based RRAM Cross-Point Structures". Cong Xu, Xiangyu Dong, Norman P. Jouppi, Yuan Xie.
- [15] Y. He, Y. Huang, J. Yue, W. Sun, L. Zhang and Y. Liu, "C-RRAM: A Fully Input Parallel Charge-Domain RRAM-based Computing-in-Memory Design with High Tolerance for RRAM Variations," 2022 IEEE International Symposium on Circuits and Systems (ISCAS), Austin, TX, USA, 2022, pp. 3279-3283, doi: 10.1109/ISCAS48785.2022.9937513.

[16] M. Uhlmann et al., "One-Transistor-Multiple-RRAM Cells for Energy-Efficient In-Memory Computing," 2023 21st IEEE Interregional NEWCAS Conference (NEWCAS), Edinburgh, United Kingdom, 2023, pp. 1-5, doi: 10.1109/NEWCAS57931.2023.10198073.

[17] Z. Wang et al., "Digital-Assisted Analog In-Memory Computing with RRAM Devices," 2023 International VLSI Symposium on Technology, Systems and Applications (VLSI-TSA/VLSI-DAT), HsinChu, Taiwan, 2023, pp. 1-4, doi: 10.1109/VLSI-TSA/VLSI-DAT57221.2023.10134272.

[18] Z. Yang and L. Wei, "Logic Circuit and Memory Design for In-Memory Computing Applications using Bipolar RRAMs," 2019 IEEE International Symposium on Circuits and Systems (ISCAS), Sapporo, Japan, 2019, pp. 1-5, doi: 10.1109/ISCAS.2019.8702555.