

# POLITECNICO DI MILANO

Scuola di Ingegneria dell'Informazione

Corso di Laurea Specialistica in Ingegneria Elettronica



## MODELLI FISICI DI SWITCHING IN MEMORIE PCM: TENSIONE DI SOGLIA E TRANSIZIONE DI FASE

Relatore: Chiar.mo Prof. Daniele IELMINI

Correlatore: Dott. Ing. Davide FUGAZZA

Tesi di Laurea Magistrale di:

Nicola CIOCCHINI

Matricola: 745851

Anno Accademico 2010–2011

*A mio padre Giancarlo e a mia madre Fausta*

# Indice

<b>Indice</b>	<b>ii</b>
<b>Introduzione</b>	<b>v</b>
<b>1 Le Memorie non Volatili (NVM)</b>	<b>1</b>
1.1 Le memorie non volatili e la tecnologia Flash . . . . .	2
1.2 Le tecnologie emergenti . . . . .	5
1.3 La tecnologia PCM . . . . .	7
1.3.1 Principio di funzionamento delle PCM . . . . .	8
1.3.2 <i>Switching</i> a soglia . . . . .	11
1.3.3 Cristallizzazione indotta da campo elettrico . . . . .	12
1.3.4 Struttura della cella . . . . .	13
1.3.5 Dispositivo di selezione . . . . .	15
1.3.6 Affidabilità della PCM . . . . .	16
<b>2 Apparato Sperimentale</b>	<b>19</b>
2.1 Strumenti di misura . . . . .	20
2.2 Struttura dei dispositivi sperimentali . . . . .	20
2.2.1 La struttura 1R . . . . .	21
2.2.2 La struttura 1T1R . . . . .	22
2.3 Struttura 1T1R: caratterizzazione del selettore in temperatura .	23
2.4 Apparato di misura a temperatura ambiente . . . . .	25

2.4.1	Metodologia di misura . . . . .	25
2.4.2	Ottimizzazione dell'apparato a temperatura ambiente . .	27
2.5	Apparato di misura a temperature criogeniche . . . . .	29
2.5.1	Principi di funzionamento del criostato . . . . .	30
2.5.2	Metodologia di misura . . . . .	31
2.5.3	Un nuovo <i>set-up</i> per la struttura 1T1R . . . . .	32
2.6	Caratterizzazione preliminare del nuovo apparato . . . . .	38
<b>3</b>	<b>Modello di <i>drift</i> della tensione di soglia</b>	<b>40</b>
3.1	Lo <i>switching</i> a soglia e i fenomeni di <i>drift</i> . . . . .	41
3.1.1	Lo <i>switching</i> a soglia ( <i>threshold switching</i> ) . . . . .	41
3.1.2	Dipendenza dallo stato programmato . . . . .	43
3.1.3	Dipendenza dal tempo dopo programmazione ( <i>drift</i> ) . .	44
3.2	Modelli di conduzione nella fase amorfa di GST . . . . .	45
3.2.1	Diagramma a bande nel GST amorfo . . . . .	45
3.2.2	Conduzione ad alto campo . . . . .	47
3.2.3	Resistenza di basso campo . . . . .	51
3.3	Modello di <i>switching</i> a soglia . . . . .	51
3.3.1	Il criterio di potenza di soglia costante . . . . .	51
3.3.2	Il rilassamento strutturale e il <i>drift</i> dell'energia di attiva- zione . . . . .	55
3.4	Risultati sperimentali e previsioni teoriche . . . . .	59
3.4.1	Il <i>drift</i> della resistenza . . . . .	59
3.4.2	Applicazione della regola di Meyer-Neldel al calcolo della corrente . . . . .	59
3.4.3	Simulazione del <i>drift</i> della tensione di soglia . . . . .	61
<b>4</b>	<b>Modello di cristallizzazione indotta da campo elettrico</b>	<b>66</b>
4.1	Meccanismi di cristallizzazione . . . . .	67

4.1.1	Fenomeno di nucleazione . . . . .	67
4.1.2	Fenomeno di crescita . . . . .	74
4.1.3	Dipendenza dalla temperatura . . . . .	76
4.2	Cristallizzazione indotta da <i>annealing</i> . . . . .	79
4.3	Cristallizzazione indotta da impulso elettrico nel GST . . . . .	81
4.3.1	Risultati di caratterizzazione . . . . .	81
4.3.2	Modello di <i>Joule heating</i> a resistenza termica costante .	82
4.3.3	Modello di <i>Joule heating</i> a resistenza termica variabile .	87
	<b>Conclusioni</b>	<b>93</b>
	<b>Bibliografia</b>	<b>96</b>

# Introduzione

Questa Tesi di Laurea Specialistica affronta i problemi dello *switching* e del *drift* in memorie a cambiamento di fase (*phase change memory*, PCM). Il lavoro ha comportato sia un'attività di carattere sperimentale, volta alla caratterizzazione elettrica di dispositivi PCM, sia un'attività di modellistica fisica dei fenomeni di *switching* e *drift*. Il lavoro è stato svolto presso il Laboratorio di Micro e Nanoelettronica del Dipartimento di Elettronica e Informazione del Politecnico di Milano (NanoLab), in collaborazione con il gruppo delle memorie a cambiamento di fase di Micron, Agrate Brianza.

I dispositivi PCM si basano sulla proprietà di alcune leghe calcogenure (tra cui  $Ge_2Sb_2Te_5$ , detto GST) di modificare reversibilmente la loro fase, passando da una fase amorfa, caratterizzata da alta resistenza elettrica e alta costante dielettrica (stato logico di *reset*), ad una fase policristallina, che presenta invece bassa resistenza elettrica e bassa costante dielettrica (stato logico di *set*). La transizione tra un fase e l'altra è ottenibile mediante l'applicazione di un impulso elettrico o ottico. I materiali a cambiamento di fase sono stati a lungo e con successo impiegati come materiale attivo in dischi di storage ottico (CD-ROM, DVD e Blue Ray Disc). Nonostante un forte interesse nello sfruttamento degli effetti di *switching* di fase in dispositivi elettronici fin dagli anni '60 del secolo scorso, solo recentemente la fattibilità tecnologica di tali dispositivi si è concretizzata. La spinta verso l'esplorazione di concetti di storage basati su materiali a cambiamento di fase è stata inoltre rafforzata dai

sempre crescenti problemi di *scaling* delle memorie non volatili.

La tecnologia ad oggi dominante nel settore delle memorie non volatili (NVM) è infatti la tecnologia *Flash*, la cui scalabilità ha permesso di seguire il trend della legge di Moore negli ultimi decenni. Tuttavia, diversi autori [1, 2] concordano oggi sul fatto che il limite di scalabilità di questa tecnologia sia oramai prossimo. Tra le possibili soluzioni alternative, la tecnologia PCM ricopre sicuramente un ruolo di protagonista, essendo caratterizzata da basse tensioni operative, robustezza alla ciclatura molto elevata (fino a  $10^9$  cicli), alta velocità di *switching* tra gli stati logici, ritenzione del dato competitiva con la tecnologia *Flash*, e ampia finestra resistiva tra lo stato di *set* e quello di *reset*. Inoltre, la maturità della tecnologia di processo dovuta al successo dello storage ottico rende oggi possibile il passaggio da dispositivo di ricerca in laboratorio a vera e propria realtà di mercato, in grado di assolvere a nuove funzioni e generare nuove applicazioni.

La Tesi è così strutturata. Nel Capitolo 1, dopo un introduzione sul mondo delle memorie non volatili, si riassumono sinteticamente i principi di funzionamento, i pregi e le prospettive dei dispositivi di memoria PCM, mentre il Capitolo 2 descrive il lavoro di sviluppo e miglioramento dell'apparato sperimentale attuato per consentire misure in criostato affidabili e ripetibili. Nel Capitolo 3 è affrontato un fenomeno fondamentale che coinvolge la fase amorfa dei materiali calcogenuri utilizzati come materiali attivi, lo *switching* a soglia, che consiste in un marcato e repentino aumento di conduttività elettrica, cui corrisponde una regione a resistenza differenziale negativa nella curva IV. La tensione per cui si osserva la transizione tra un regime di conduzione e l'altro è detta tensione di soglia  $V_T$ . Tale tensione riveste un ruolo fondamentale per le applicazioni di memoria in quanto delimita il confine tra le tensioni utili alla lettura (minori di  $V_T$ ) e quelle utili alla scrittura (maggiori di  $V_T$ ), definendo perciò il range delle tensioni operative del dispositivo. Il lavoro svolto si è

soffermato in particolare sul fenomeno di variazione spontanea della tensione di soglia nel tempo dopo la programmazione della cella nello stato di *reset*, fenomeno cui si dà il nome di *drift*. Il *drift* della tensione di soglia è stato misurato su un intervallo temporale di almeno dieci decenni, da  $100ns$  a  $1000s$ . È stato elaborato un modello fisico per la previsione dell'evoluzione temporale della soglia, che tiene conto del rilassamento strutturale della fase amorfa, dei meccanismi di conduzione sotto-soglia per *hopping* da stati trappola nel *Band Gap* (sia Poole che Poole-Frenkel) e dello *switching* indotto da incremento di energia dei portatori per effetto degli alti campi nella fase amorfa. Il modello ha consentito di ottenere previsioni in linea con i dati sperimentali, pertanto si presta ad un futuro sviluppo che permetta di comprendere l'influenza dello *scaling* e della geometria di cella sul valore della tensione di soglia, sulla sua variabilità statistica e sull'entità del *drift* ad essa associato.

Il Capitolo 4 presenta un'originale metodologia sperimentale per la caratterizzazione della dinamica di cristallizzazione da fase amorfa indotta dall'applicazione di impulsi elettrici. Mentre la fase cristallina risulta essere stabile, la fase amorfa è per sua natura intrinsecamente metastabile ed evolve spontaneamente nel tempo, determinando una variazione delle proprietà elettriche dello stato programmato, oltre che costituire un fattore limitante la ritenzione del dato nei dispositivi PCM. In analogia con le misure accelerate in temperatura, le quali prevedono l'esposizione del campione nello stato programmato a lunghe sessioni di *annealing* in forno, l'applicazione di impulsi elettrici di ampiezza controllata permette di indurre localmente un riscaldamento (*Joule heating*) all'interno del materiale, determinando la cristallizzazione su scale temporali estremamente brevi dell'ordine delle decine di nanosecondi. L'elaborazione di un modello di *Joule heating* analitico ha permesso di dedurre la temperatura locale nell'elemento attivo, permettendo di riportare i tempi di cristallizzazione su un grafico Arrhenius. Il confronto di tali risultati con i dati ottenuti negli



esperimenti di ritenzione accelerati in temperatura ha mostrato una diversa energia di attivazione per la cristallizzazione. Nonostante la sua semplicità, la metodologia descritta permette di accedere alle proprietà di cristallizzazione (in termini di nucleazione e crescita) del materiale sotto una luce del tutto nuova, rendendone di fatto più accessibile un'analisi statistica dall'elevato valore aggiunto anche in termini di previsioni affidabilistiche per la ritenzione del dato. Questo lavoro costituisce un punto di partenza per ulteriori caratterizzazioni sperimentali su diverse tecnologie, dimensioni e strutture di cella, al fine di validare e di migliorare il semplice modello dei *Joule heating* finora sviluppato.

Nel complesso, la tesi affronta alcuni tra i problemi più urgenti che limitano lo sviluppo di PCM ad alte prestazioni e piccole dimensioni. Comprendere i dettagli degli effetti di programmazione e gli effetti di *drift* della soglia, infatti, può condurre allo sviluppo di nuovi algoritmi di programmazione, nuove metriche di lettura della cella, e nuovi materiali per una più rapida programmazione ed una migliore affidabilità. Ma soprattutto, conoscere i dettagli di questi fenomeni può aiutare a prevedere l'entità di questi effetti sui prossimi nodi tecnologici con dispositivi più scalati, pertanto gettando le basi per un corretto *scaling* della PCM.

# Capitolo 1

## Le Memorie non Volatili (NVM)

*Nel presente capitolo viene presentata brevemente la tecnologia delle memorie non volatili. Dopo un'introduzione sulla tecnologia Flash, dominante il settore NVM (Sezione 1.1), viene mostrata una panoramica delle memorie emergenti (Sezione 1.2). Tra queste, la tecnologia PCM è discussa in maggiore dettaglio nella parte conclusiva (Sezione 1.3), dove vengono introdotti i concetti di switching elettronico e di switching di fase che verranno approfonditi nel Capitolo 3 e nel Capitolo 4.*

## 1.1 Le memorie non volatili e la tecnologia Flash

Con il termine memoria non volatile (*Non Volatile Memory*, NVM) si intende una tipologia di memoria elettronica in grado di mantenere l'informazione anche quando non viene alimentata. Nell'ultimo decennio i requisiti richiesti alla tecnologia delle memorie in termini di consumo di potenza, densità di integrazione e velocità di accesso sono divenuti sempre più stringenti. Per comprendere queste esigenze si pensi a quanto il trend del mercato spinga sempre più all'integrazione su unico dispositivo portatile di diverse applicazioni come la fotografia, il video, la musica, oppure a come la pubblica amministrazione si stia indirizzando verso la digitalizzazione di tutti i dati anagrafici, o ancora al recente sviluppo delle tecnologie di *Cloud Computing*, che richiedono una quantità di memoria crescente ai *server*. La tecnologia che fino ad oggi ha dominato questo ramo dell'elettronica è sicuramente la tecnologia *Flash* (Figura 1.1). La quale ha un dispositivo di memoria del tutto simile ad un transistor MOS, tranne che per l'aggiunta di un *gate* supplementare elettricamente isolato dagli altri terminali del transistor, e per questo denominato *Floating Gate*. La capacità di memorizzare un bit di informazione risiede proprio nella presenza di questo terminale flottante, che, a seconda del valore della *control gate*, immagazzina o rilascia della carica elettrica attraverso un sottile strato di ossido (*tunnel oxide*) interposto tra *floating gate* e canale del MOSFET. La variazione di tensione di soglia indotta dalla presenza o meno di un pacchetto di carica all'interno della *floating gate* permette la lettura del dato memorizzato dalla cella.

Attualmente le memorie *Flash* hanno dimostrato di poter essere scalate fino al nodo 1X nm [4], ma si prevede che lo *scaling* sarà sempre più difficile, a causa dei limiti fisici che limitano l'affidabilità delle celle in tecnologie ultrascalate (Figura 1.2).

La componente più critica delle celle di memoria *Flash* è l'ossido di tunnel.

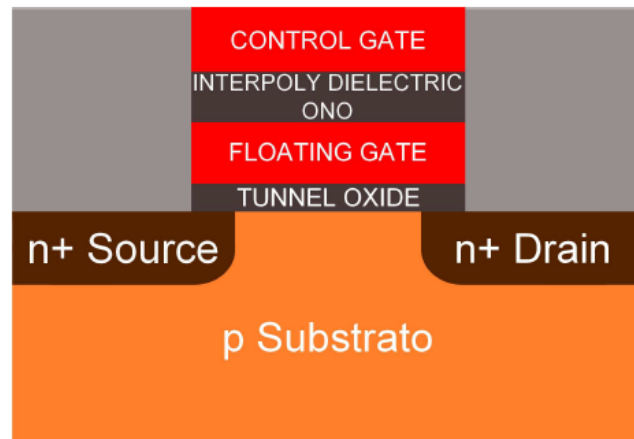


Figura 1.1: Schema in sezione della struttura di una cella di memoria *Flash floating gate*. Figura tratta da [3].

Difatti, lo *scaling* delle tensioni impone che la capacità specifica dell'ossido cresca per mantenere sufficientemente elevata la densità di corrente di sopra-soglia. Ricordando la formula  $C_{ox} = \epsilon_{ox} / t_{ox}$  si comprende come lo *scaling* dei dispositivi imponga la diminuzione dello spessore del *tunnel oxide*, ma così facendo ci si scontra con l'aumento delle correnti di perdita attraverso l'ossido per effetto tunnel, con conseguente diminuzione della capacità ritenzione del dato della cella. La soluzione a questo problema che è stata adottata è l'utilizzo di materiali alternativi all'ossido di silicio che presentino una costante dielettrica maggiore (materiali *High-k*), in modo tale che la capacità specifica dell'ossido sia superiore senza dover diminuire eccessivamente lo spessore dell'ossido stesso. Negli ultimi anni, il trend di scalabilità di queste celle di memoria è stato lineare: la densità di celle per una memoria Flash (sia NOR sia NAND) è raddoppiata ogni uno-due anni, in accordo con la legge di Moore. Tutto ciò è stato possibile grazie al costante miglioramento delle tecniche litografiche e della pulizia del processo, che hanno consentito di fornire le giuste garanzie di affidabilità, ritenzione del dato e resa di array per una commercializzazione a

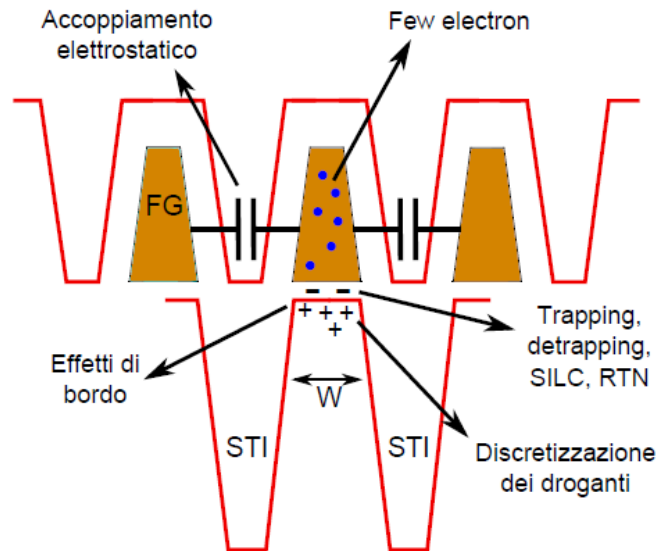


Figura 1.2: Schema riassuntivo delle problematiche associate allo *scaling* delle memorie *Flash*. Figura tratta da [3].

basso costo delle memorie basate su questa tecnologia. Tuttavia, nei prossimi anni, si prevede un notevole rallentamento di questo trend [1, 2]. Ciò è dovuto principalmente a due motivi: l'ossido di tunnel non potrà più essere ridotto ulteriormente (si ritiene che il limite ultimo sia il nodo tecnologico dei 7 nm) senza causare errori nella ritenzione del dato e le interazioni elettrostatiche fra celle adiacenti saranno troppo elevate per permettere un corretto salvataggio del dato. Inoltre, dato che man mano che le celle diventano sempre più piccole il numero di elettroni posti nella *Floating Gate* e presenti nel canale del dispositivo diminuisce, nascono effetti come il cosiddetto *Random Telegraph Noise (RTN)* che può causare instabilità di soglia (fra stato di *set* e stato di *reset*) ed errori di lettura. Infine, poiché le celle vengono programmate applicando elevate tensioni su ossidi ultra sottili, entrano in gioco fenomeni di correnti di perdita indotte da stress (*Stress-Induced Leakage Current, SILC*) che possono alterare il dato immagazzinato.

## 1.2 Le tecnologie emergenti

Viste queste problematiche da tempo sono allo studio diverse possibili alternative NVM alle memorie *Flash*, che è possibile suddividere in tre macro-categorie basate sul principio fisico di immagazzinamento del dato: le FeRAM (*Ferroelectric RAM*), le MRAM (*Magnetoresistive RAM*) e le RRAM (*Resistive RAM*).

Le FeRAM si basano su particolari materiali ferroelettrici, che sono in grado di cambiare il loro stato di polarizzazione dielettrica in risposta ad un campo elettrico, e di mantenere lo stato di polarizzazione anche quando il campo viene spento. I materiali più utilizzati per questo tipo di tecnologia sono il PZT ( $\text{PbTiZrO}$ , *titanato di piombo e zirconio*) e il SBT ( $\text{SrBi}_2\text{Ta}_2\text{O}_9$ , *tantalato di stronzio e bismuto*).

Le MRAM si basano invece su una particolare giunzione detta MTJ (*Magnetic Tunnel Junction*), formata da uno strato di materiale ferromagnetico a

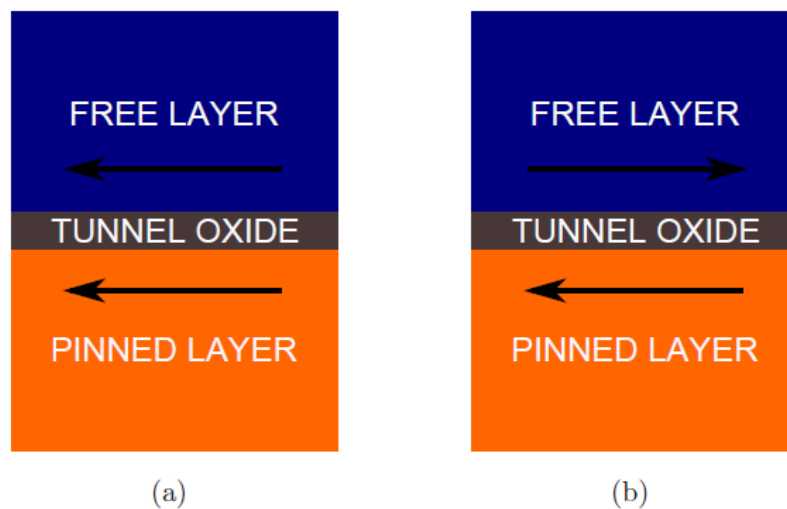


Figura 1.3: Stati logici di una memoria MRAM: (a) polarizzazione parallela, bassa resistenza (b) polarizzazione antiparallela, alta resistenza

polarizzazione permanente (*Pinned Layer*), un sottile strato di ossido di tunnel, e uno strato di materiale ferromagnetico a polarizzazione variabile e programmabile elettricamente (*Free Layer*) (Figura 1.3). La memorizzazione del dato avviene agendo sulla polarizzazione del free layer, che per effetto di induzione magnetica è modificabile facendo passare della corrente elettrica in una pista metallica adiacente. Per il fenomeno fisico dell'*effetto tunnel magnetico* alla condizione di polarizzazione magnetica parallela corrisponde una resistenza elettrica attraverso il *tunnel oxide* bassa, mentre ad una polarizzazione anti-parallela corrisponde una resistenza elevata. È così possibile memorizzare in maniera non volatile il dato sottoforma di polarizzazione magnetica.

Le memorie RRAM si basano sulla capacità di particolari materiali di presentarsi in stati di aggregazione caratterizzati da diversa resistenza elettrica, uno stato basso resistivo (detto *stato di set*, valore logico 1) e uno alto resistivo (detto *stato di reset*, valore logico 0). Questa famiglia di dispositivi è molto ampia, e si può suddividere in due grosse famiglie: RRAM basate su materiali organici o polimerici e RRAM basate su materiali inorganici. Quelle polimeriche sfruttano o complessi metallo-organici, in grado di variare la propria resistenza a seguito dell'applicazione di un campo elettrico, o le proprietà di singole molecole o *monolayer*. Le prospettive di scalabilità di quest'ultime sono in linea teorica enormi, vista la possibilità di arrivare a una cella composta da una singola molecola, ma i limiti della tecnologia attuale e i problemi di affidabilità legati alle dimensioni del dispositivo impediscono ad ora prospettive di commercializzazione. Le RRAM inorganiche comprendono a loro volta le memorie elettrolitiche (PMC, *Programmable Metallization Cell*) e le memorie basate su ossidi metallici e le PCM (*Phase Change Memory*).

Le memorie PMC sfruttano reazioni di ossido-riduzione, e sono realizzate iniettando nanoparticelle di metallo in un elettrolita solido. Tramite impulso elettrico è possibile formare o rompere il filamento, passando da uno stato

basso resistivo (1 logico) ad uno alto resistivo (0 logico). Le RRAM basate su ossido metallico sono costituite da un sottile film di ossido, la cui resistenza è modificabile tramite l'applicazione di un campo elettrico, che forma (ON) o dissolve (OFF) un filamento metallico all'interno dell'ossido stesso (Figura 1.4). I materiali utilizzati sono molto vari, da ossidi quaternari come  $\text{SrTi}(\text{Zr})\text{O}_3$ ,

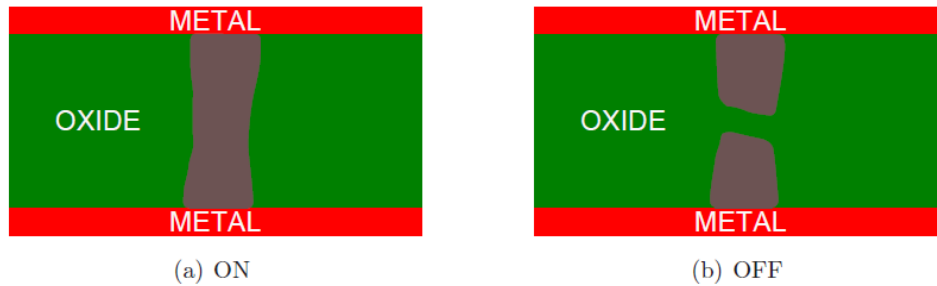


Figura 1.4: Stati logici di una memoria PMC: (a) filamento metallico formato, bassa resistenza (b) filamento metallico dissolto, alta resistenza. Figura tratta da [3].

o semplici ossidi binari come  $\text{HfO}_2$ ,  $\text{TiO}_2$ ,  $\text{ZrO}_2$ ,  $\text{NiO}_2$ . Mentre i primi sono difficilmente compatibili con la tecnologia CMOS, si preferiscono ad oggi quelli binari, tra i quali il più promettente sembra essere  $\text{NiO}_2$ .

### 1.3 La tecnologia PCM

Tra le nuove proposte di soluzioni di *storage* è possibile includere diversi tipi di memoria: magneto-resistive (MRAM) [5], ferro-elettriche (FeRAM) [6], a cambiamento di fase (PCM) [7] a switch resistivo (RRAM) [8]. Ognuna di queste soluzioni basa il suo principio di funzionamento su un cambiamento della resistenza del materiale attivo, l'unica eccezione sono le FeRAM, nelle quali il dato è immagazzinato nella polarizzazione di un materiale ferro-elettrico e l'elemento di memoria è sostanzialmente un condensatore. In particolare, le memorie a cambiamento di fase hanno una posizione privilegiata, a causa del



livello di maturità della loro tecnologia e del loro potenziale nella scalabilità [9, 10]. In questa sezione, dopo una breve introduzione delle tipiche architetture delle celle e delle soluzioni nel loro accorpamento, ci concentreremo nei principi fisici come la conduzione e le proprietà di *switching* elettronico, le quali saranno poi oggetto di approfondimenti teorici e sperimentali presentati nei capitoli seguenti.

### 1.3.1 Principio di funzionamento delle PCM

Le memorie a cambiamento di fase contano sulla proprietà di alcune leghe calcogenure di cambiare fase da amorfa a policristallina (alla quale ci si riferisce abitualmente come cristallina) e viceversa. L'informazione binaria è associata con la struttura fisica del materiale, la quale si distingue nelle proprietà ottiche ed elettriche, è poi facilmente deducibile da tecniche di misura dell'uno o dell'altro tipo.

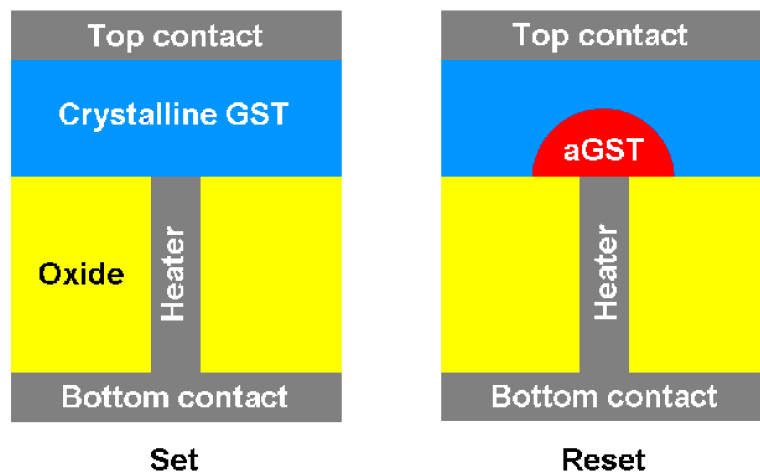


Figura 1.5: Schema di una semplice PCM cilindrica. Il materiale attivo è contattato da un sottile pilastro metallico, il quale permette al GST di riscaldarsi tramite impulsi elettrici. Figura tratta da [3].

Un fattore chiave nelle prestazioni delle PCM è il materiale attivo. Materiali a cambiamento di fase sono in generale calcogenuri, i quali sono stati ampiamente studiati dagli anni sessanta sia per le loro proprietà di cambiamento di fase (oggi sfruttate nei dispositivi ottici come BD, DVD e CD riscrivibili), sia per l'interessante fenomeno di *switching* elettronico [11]. Il materiale più usato è  $Ge_2Sb_2Te_5$  (*GST*) [12, 13], eventualmente drogato con azoto o ossigeno, ma svariati differenti materiali sono stati presi in considerazione di recente, come SbTe, GeSb, InGeTe, GeTe e nanofili in GST o GE .

La Figura 1.5 mostra il principio di funzionamento di una cella PCM. Il materiale attivo è interposto tra due contatti elettrici (detti *Top* e *Bottom*). A sinistra c'è la cella nel suo stato di bassa resistenza (fase cristallina, *set*), mentre a destra è nello stato alto resistivo (presenza di una calotta di materiale amorfizzato, *reset*). Il cambiamento di fase è ottenuto col riscaldamento per effetto Joule della regione attiva: impulsi di corrente ampi (centinaia di microampere) e relativamente rapidi (da decine a centinaia di nanosecondi, v. Capitolo4).

L'operazione di *reset*, o amorfizzazione, è ottenuta riscaldando il materiale oltre la sua temperatura di fusione  $T_m$  (di solito nell'intorno di 600°C nel GST) e facendolo raffreddare repentinamente, tanto da congelarsi in uno stato di disordine strutturale (la fase amorfa). La fase cristallina è ristabilita dall'operazione di *set*, si riscalda il materiale amorfo ad una temperatura prudentemente sotto quella di fusione e comunque sufficiente a garantire un veloce processo di cristallizzazione. Solitamente quest'ultima operazione richiede durate di impulso maggiori rispetto a quelle di *reset*. Siccome si vuole che queste procedure siano efficienti è necessario un forte confinamento del calore nella cella. Ciò è reso possibile da un sottile elettrodo metallico (spesso realizzato in TiN o TiSiN) al quale ci si riferisce con la parola *heater*, vedi Figura 1.5. Questo elemento è di fondamentale importanza: deve dissipare molta potenza

e sostenere un forte gradiente di temperatura, perciò sono necessarie una bassa resistenza elettrica e una bassa conducibilità termica. A causa della forma dell'*heater*, la cella in Figura 1.5 è solitamente chiamata di tipo *lance* [12] o *mushroom*[14].

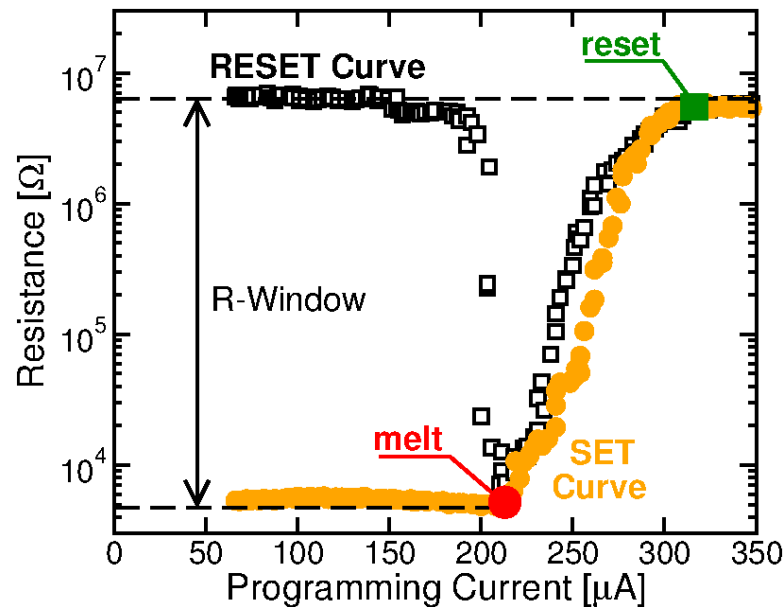


Figura 1.6: Tipica *curva R-I* con stato iniziale *reset* (alta resistenza) o *set* (bassa resistenza). L'ampia finestra tra i due stati rappresenta uno degli aspetti chiave di questa tecnologia.

Per quanto riguarda l'operazione di programmazione, le tipiche curve che la caratterizzano sono le *R-I* mostrate in Figura 1.6. Esse sono acquisite misurando la resistenza subito dopo un breve impulso (circa  $100ns$ ) di corrente sulla cella in un determinato stato. Quindi prima di ogni misura viene applicato un impulso iniziale di *set* o *reset*, rispettivamente si ricava la *curva di set* o la *curva di reset* mostrate in Figura 1.6. Nella curva di *reset* per correnti relativamente alte la resistenza programmata comincia a mostrare una cospicua discesa, a causa della progressiva cristallizzazione di parte del volume amorfo nella cella, causata dalla dissipazione di potenza elettrica e dal conseguente riscaldamento per effetto

Joule. Con il raggiungimento della corrente di fusione, la resistenza mostra il suo valore minimo e poi ricomincia a crescere (punto di fusione), infatti il riscaldamento è sufficiente a fondere il materiale calcogenuro ed eventualmente ad avere come risultato l'amorfizzazione, dato che il volume fuso si raffredda durante il veloce fronte di discesa dell'impulso di programmazione. La parte crescente della curva di *reset* è chiamata *regione di reset*, mentre alla prima parte ci si può riferire come *regione di set*. Per la curva di *set*, correnti di programmazione sotto il punto di fusione lasciano inalterata la fase cristallina, mentre correnti più alte inducono un incremento del volume amorfo fino ad ottenere lo stato di *full reset*. Notare che anche se l'operazione di *reset* è tipicamente eseguita su una cella in stato di *set*, lo stato di *reset* finale è indipendente dallo stato iniziale. Va fatto notare che la R-I e in particolare la possibilità di prevedere la resistenza della cella con appropriati impulsi di programmazione [15, 16], ha una fondamentale importanza per l'uso in memorie multilivello.

### 1.3.2 *Switching a soglia*

Come accennato in precedenza, i materiali calcogenuri nella loro fase amorfa sono soggette al fenomeno di *switching* a soglia, che consiste in una transizione improvvisa e reversibile da uno stato alto-resistivo ad uno basso resistivo, visibile come uno *snap-back* nella caratteristica IV del dispositivo PCM in stato iniziale amorfo. Poiché lo *switching* a soglia determina le tensioni di funzionamento e la velocità di programmazione dei dispositivi PCM, la capacità di prevedere la tensione di soglia  $V_T$ , la corrente di soglia  $I_T$  e la velocità del fenomeno è di importanza fondamentale per la scelta della corretta lega calcogenura per applicazioni NVM.

Benché lo *switching* a soglia sia stato attribuito da tempo ad un processo elettronico legato alla distribuzione dei portatori in trappole nel *bandgap*, i

meccanismi fisici precisi che lo generano non sono ancora stati ben chiariti. In questo lavoro lo *switching* a soglia verrà spiegato in termini di acquisto di energia degli elettroni in moto a spese del campo elettrico, rallentato dalla perdita di energia a causa delle interazioni elettrone-fonone. L'acquisto di energia porta ad un'aumento di conduttività e ad un collasso del campo elettrico all'interno del materiale calcogenuro, fenomeno visibile dalla zona a resistenza differenziale negativa (NDR) visibile nella curva IV.

Il modello sviluppato servirà a spiegare i meccanismi alla base del *drift* della tensione di soglia, ossia dell'aumento di  $V_T$  nel tempo dopo la programmazione della cella, fenomeno questo che va tenuto sotto controllo per un corretto funzionamento dei dispositivi basati su questa tecnologia.

### 1.3.3 Cristallizzazione indotta da campo elettrico

La fase amorfa del materiale calcogenuro non è termodinamicamente stabile, e tende spontaneamente alla fase cristallina tramite dei processi di nucleazione e crescita. E' evidente che il tempo di cristallizzazione è un parametro fondamentale per uno studio di ritenzione del dato su memorie basate su questa tecnologia. Il processo di cristallizzazione è termicamente attivato e presenta un'andamento all'Arrhenius, almeno alle temperature tipiche degli esperimenti in *bake*. Quello che avviene è che i tempi di fallimento a basse temperature ( $T < 160C$ ) sono nell'ordine dei mesi e anni, e per questa ragione sono necessarie misure accelerate ad alte temperature per una rapida valutazione delle proprietà di cristallizzazione. L'energia di attivazione che controlla la cinetica di cristallizzazione è un parametro fondamentale che permette una stima al prim'ordine del tempo di fallimento alla massima temperatura operativa dei dispositivi PCM, e quindi gioca un ruolo determinante nello studio di ritenzione della tecnologia. Il fenomeno di cristallizzazione è altresì decisivo nel processo di scrittura della cella di memoria a cambiamento di fase, in cui un impulso

elettrico applicato alla cella può provocare il passaggio del calcogenuro dalla fase amorfa a quella cristallina. La cristallizzazione indotta da impulso elettrico (*set*) è un aspetto critico nelle PCM perché limita i tempi di programmazione e la resa di array. Infatti, lo *spread* nelle caratteristiche di cristallizzazione costringe ad aumentare il tempo di set per garantire la corretta programmazione di tutte le celle, con conseguente limitazione della velocità di scrittura. Inoltre, un eccessivo *spread* nella resistenza di set può comportare l'impossibilità di distinguere i due stati della memoria, diminuendo la resa di array. È dunque decisivo capire la cinetica, la statistica e i meccanismi di cristallizzazione, oltre all'impatto dei materiali attivi su di essa, per sviluppare dispositivi PCM ad alta resa, basso costo e alta densità. In tale contesto, la disponibilità di modelli di *Joule heating* che forniscano una stima locale della temperatura durante l'impulso di programmazione è di fondamentale importanza. Verranno pertanto presentati i risultati dell'applicazione di un modello di questo tipo per la costruzione di un *Arrhenius plot*, che verrà poi confrontato coi risultati sperimentali ottenuti in *bake* presenti in letteratura.

### 1.3.4 Struttura della cella

Il primo lavoro con array di celle PCM risale a 40 anni fa [17]. Fu una matrice 16x16 che su un *die* di  $9\text{mm}^2$  forniva 256 bit di memoria. Ogni elemento aveva in serie un diodo a giunzione p-n che agiva da selettore. In quegli anni a causa delle dimensioni, le celle richiedevano un grande dispendio di energia per la programmazione: 5s, 200mA a 25V per il processo di *reset* e 10ms, 5mA a 25V per quello di *set*. Per questo motivo, nonostante l'interesse nell'idea, il loro sviluppo rimase pressoché inattivo, finché i noti avanzamenti nella microelettronica hanno recentemente reso competitiva questa tecnologia [18]. Nell'ultima decade grazie all'intensa attività di ricerca in questo campo le prestazioni sono drasticamente aumentate. Un aspetto critico da questo punto di vista è la

geometria della cella, che determina l'efficienza nel riscaldamento del materiale e quindi nella corrente di programmazione. Quest'ultima è probabilmente uno dei problemi maggiori nella produzione di massa perché la corrente di *reset* è ancora nell'ordine delle decine di microampere, rappresentando così un fattore limitante nella programmazione contemporanea di più elementi. L'ambizione di ridurre la corrente di *reset* ha motivato lo sviluppo di una grande varietà di strutture differenti. La Figura 1.7 mostra una breve indagine delle possibili architetture elementari, che possono essere approssimativamente divise in due categorie maggiori:

- i) *contact-minimized cells*, nelle quali la corrente è forzata a passare attraverso una piccola sezione di contatto tra l'elettrodo e il materiale a cambiamento di fase;
- ii) *volume-minimized cells*, che confinano il volume del materiale attivo in modo da creare una piccola sezione trasversale nel dispositivo PCM.

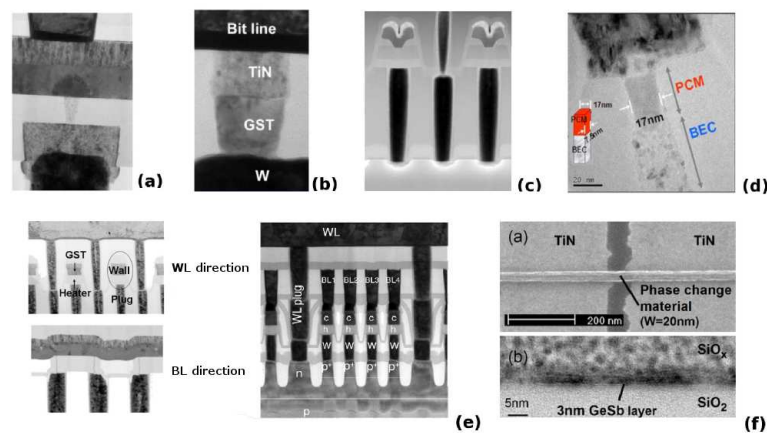


Figura 1.7: TEM o SEM di alcuni dispositivi di memoria PCM con differenti architetture.

Nel contesto delle *contact-minimized cells* l'esempio più comune è quello rappresentato dalla cella a fungo (più comunemente cella *mushroom*), nella

quale uno stretto elettrodo metallico di forma cilindrica contatta un sottile strato di materiale a cambiamento di fase, come mostrato in Figura 1.7a [19]. In questo caso la parte più piccola e quindi più critica della cella è l'elettrodo inferiore, che ha dimensioni sub-litografiche. Altri esempi di *contact-minimized cells* sono la struttura *edge-contact* [20] o la struttura ad anello che combina i vantaggi delle dimensioni sub-litografiche a quelli del controllo dello spessore del materiale [21]. Per quel che riguarda le *volume-minimized cells*, la struttura più ovvia è rappresentata dalla cella di tipo *pillar* (Fig 1.7b), nella quale una stretta porzione del materiale giace tra due elettrodi. Le Figura 1.7c e 1.7 rappresentano combinazioni dei due tipi di struttura, rispettivamente la *μ-trench cell*[16, 22] e la *dash-confined cell*[23].

### 1.3.5 Dispositivo di selezione

In parallelo all'ottimizzazione dell'elemento di memoria PCM in sé, un punto su cui è necessario soffermarsi per un buon funzionamento della cella è il selettore utilizzato per fornire l'impulso di *set/reset* al materiale calcogenuro. Vista la distribuzione statistica dei parametri che caratterizzano la cella PCM su array, la quantità di corrente che il dispositivo di selezione è in grado di fornire dev'essere agevolmente superiore a quella che serve nel caso peggiore per il *reset* dell'elemento di memoria. I MOSFET convenzionali, a causa della relativamente bassa densità di corrente che possono raggiungere, non sono i dispositivi di accesso più adatti per array PCM scalati. In effetti questa scelta limiterebbe il layout della cella a più di  $8F^2$ . Quindi BJT, diodi o altri nuovi dispositivi di accesso sono stati presi in considerazione per ridurre la dimensione della cella senza perdere la capacità di fornire corrente (densità di corrente dell'ordine di  $10MA/cm^2$  o più devono essere prese in considerazione per i prossimi nodi tecnologici). Usando un bipolare o un diodo in silicio come selettore si riesce in effetti a raggiungere un'area vicino a  $5F^2$ , ma tuttavia



circuiti di decodifica e di selezione di riga in tecnologia CMOS condividono il substrato con selettori in silicio, con la conseguenza di una diminuzione della resa di array. Quindi, nonostante la tecnologia PCM sembra potersi candidare per prestazioni e affidabilità come NVM del futuro, visto la semplicità di integrazione e la scalabilità che mostra, bisogna rendere confrontabile il costo di questa tecnologia con quello delle *Flash* NAND, leader attuale del mercato memorie (area cella  $< 4F^{2/3}$ ). Per raggiungere questo obiettivo, oltre che indagare la possibilità di sviluppo di memorie PCM multilivello, è necessario cercare di sviluppare un layout più efficiente, attraverso l'evoluzione del concetto di architettura di memoria 3D *stacked cross-point* (celle disposte all'intersezione di piste ortogonali di selezione, con array impilati nella terza direzione). In un'architettura di questo tipo, è stato recentemente proposto come dispositivo di selezione un OTS (*Ovonic Threshold Switch*), costituito da un materiale che sfrutta la presenza di *switching* a soglia per indirizzare un elemento all'interno dell'array. Questa alternativa prevede che l'elemento di memoria (PCM) e il selettore (OTS) siano verticalmente integrati a formare la cosiddetta PCMS. Inoltre, unitamente alla compatibilità di processo, l'OTS è in grado di fornire alte densità di corrente rispetto ai transistori convenzionalmente integrati, fornendo una delle più interessanti soluzioni per il selettore dell'elemento di memoria PCM.

### 1.3.6 Affidabilità della PCM

Le memorie PCM sono una classe di memorie programmabili e non volatili, pertanto i principali requisiti di affidabilità sono la ciclabilità e la ritenzione del dato memorizzato. La ciclabilità rappresenta la capacità di una cella di memoria di mantenere inalterato il suo funzionamento dopo un certo numero di cicli di programmazione, che deve essere chiaramente il più alto possibile. La tecnologia PCM sotto questo punto di vista è molto affidabile, in quanto

garantisce  $10^9$  cicli di programmazione. È stato osservato che le problematiche legate alla ciclatura provengono dallo stato di *reset*. Infatti, mentre l'applicazione di una lunga sequenza di impulsi di *set* non altera il funzionamento della cella conservando le proprietà di cambiamento di fase che la caratterizzano, applicando una sequenza di impulsi di *reset* o una sequenza di impulsi di *set* e di *reset* alternati si osserva in entrambi i casi che la resistenza della cella si abbassa fino ad arrivare al valore minimo corrispondente allo stato di *set*. Una volta raggiunta la resistenza di *set* lo stato della cella rimane "bloccato" e non è più possibile programmarla, nemmeno aumentando ampiezza e durata degli impulsi di programmazione. Le cause di questo comportamento, detto di *stuck set*, sono da imputare all'operazione di *reset*, in quanto la fusione e la successiva rapida solidificazione richieste per portare la cella PCM nello stato amorfo lentamente degradano il materiale GST. Inoltre si può assistere anche ad un fenomeno duale di *stuck reset*, provocato dalla separazione fisica del GST dall'elettrodo inferiore, dove la cella permane in uno stato con resistenza elevata (centinaia di  $M\Omega$  o  $G\Omega$ ).

La ritenzione del dato è la capacità della cella di memoria di mantenere l'informazione memorizzata per il maggior tempo possibile. Nella tecnologia PCM vi sono diverse cause che portano ad una alterazione o perdita del dato. La quale può succedere a causa di un disturbo provocato durante la fase di lettura, oppure in fase di scrittura, a seguito del forte calore sviluppato durante la programmazione. Ciò comporta un'interferenza tra celle adiacenti. Infatti l'eccessivo calore può aumentare la temperatura in una cella vicina alterandone di conseguenza lo stato programmato. La variazione dello stato programmato non è sistematica, dipende sia dallo stato in cui era programmata la cella prima di ricevere il disturbo sia dall'impulso applicato alla cella che si vuole effettivamente scrivere. Ciò può tradursi in un aumento o una diminuzione della resistenza programmata. In una memoria multilivello la questione è più gravosa

che in una memoria a due livelli, in quanto a seguito della riduzione della finestra di lettura anche una piccola alterazione dello stato programmato può risultare dannosa. Particolare cura deve essere dedicata quindi all'isolamento termico tra le celle nella progettazione di un array di memoria. Infine c'è un aspetto della ritenzione del dato che non è legato a disturbi di programmazione o lettura, ma alla fisica del GST. La fase amorfa è meta-stabile e tende lentamente a diventare cristallina. Questo comporta la perdita del dato memorizzato perché la cella dopo un determinato tempo transisce nello stato di *set*. Questa transizione, che non è distruttiva per la cella, si compone di due fasi ben distinte. Inizialmente si assiste ad un aumento della resistenza programmata, in una prima fase detta *drift*, infine nella seconda fase denominata cristallizzazione la resistenza decresce molto lentamente fino al valore di *set*. Il tempo impiegato dal GST in fase amorfa a portarsi nello stato cristallino rappresenta il tempo di ritenzione della cella. Questo tempo è funzione dello stato programmato, in quanto a diversi valori di resistenza corrispondono differenti volumi di materiale amorfo. Volumi maggiori cristallizzano più lentamente, e questo è un aspetto da tenere in considerazione nella realizzazione di una memoria multilivello.

# Capitolo 2

## Apparato Sperimentale

*Nel seguente capitolo è descritta l'attività sperimentale svolta. Dopo un'introduzione sui diversi dispositivi PCM (Sezione 2.2) segue la caratterizzazione svolta sul MOSFET di selezione utilizzato nelle misure di cristallizzazione (Sezione 2.3). Grazie a questi aspetti preliminari, è possibile procedere all'esposizione del lavoro sperimentale svolto, il quale è suddiviso in due parti relative all'ottimizzazione dell'apparato sperimentale, quello per le misure a temperatura ambiente (Sezione 2.4) e quello per le misure a temperature criogeniche (Sezione 2.5).*

## 2.1 Strumenti di misura

Verranno ora mostrati gli strumenti che sono stati utilizzati durante l'attività sperimentale di caratterizzazione e studio della tecnologia PCM.

Per la programmazione di una cella PCM è fondamentale disporre di uno strumento che possa fornire impulsi di tensione di durata dell'ordine della decina di nanosecondi e di ampiezza variabile fino a qualche volts. Questo ruolo è svolto da un generatore di segnali arbitrari *tti TGA1244 40MHz* oppure *tti TGA12100 100MHz*, che per semplicità verrà indicato nel seguito come *Pulser*. Come sarà chiarito meglio in seguito è importante misurare le tensioni presenti sui nodi intermedi del circuito durante l'invio dell'impulso, e questa funzione è ricoperta da un oscilloscopio *LeCroy LT372*. Infine per caratterizzare lo stato finale della cella dopo l'impulso di programmazione si effettua una misura di resistenza con *HP 4155B Semiconductor parameter analyzer*, indicato nel seguito come *Parameter*.

## 2.2 Struttura dei dispositivi sperimentali

Tra le attività associate alla ricerca sulle memorie non volatili, ce ne sono alcune che hanno bisogno di un prototipo della tecnologia in esame. Questi prototipi, detti celle analitiche, sono utilizzate per l'analisi delle prestazioni, la verifica di teorie e dei relativi modelli, ma anche per la ricerca di comportamenti inaspettati che rendono due tecnologie differenti.

In questo lavoro, sono stati usati prototipi di cella di memoria PCM dalle stesse caratteristiche nominali. Questi dispositivi sono a disposizione sottoforma di *die*, è quindi necessario accedervi elettronicamente tramite dei contatti creati in laboratorio (vedi Figura 2.1). Vediamo ora brevemente le tipologie di struttura di cella con cui sono state effettuate le misure sperimentali.

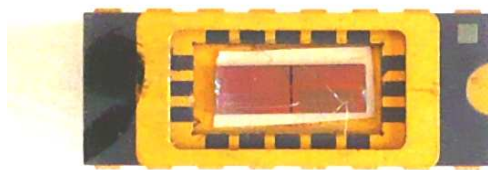


Figura 2.1: Campione contattato con i tipici fili d'oro.

### 2.2.1 La struttura 1R

Il tipo più semplice di struttura che si può immaginare è quello di una cella posta in serie ad un resistore. Nelle nostre misure il resistore è posto esternamente al circuito integrato, montata su una *Printed Circuit Board* (PCB), e il suo valore è pertanto modificabile dall'utente. Il valore della resistenza adottata è un parametro molto importante nello studio dei fenomeni fisici che caratterizzano la tecnologia PCM (Figura 2.2).

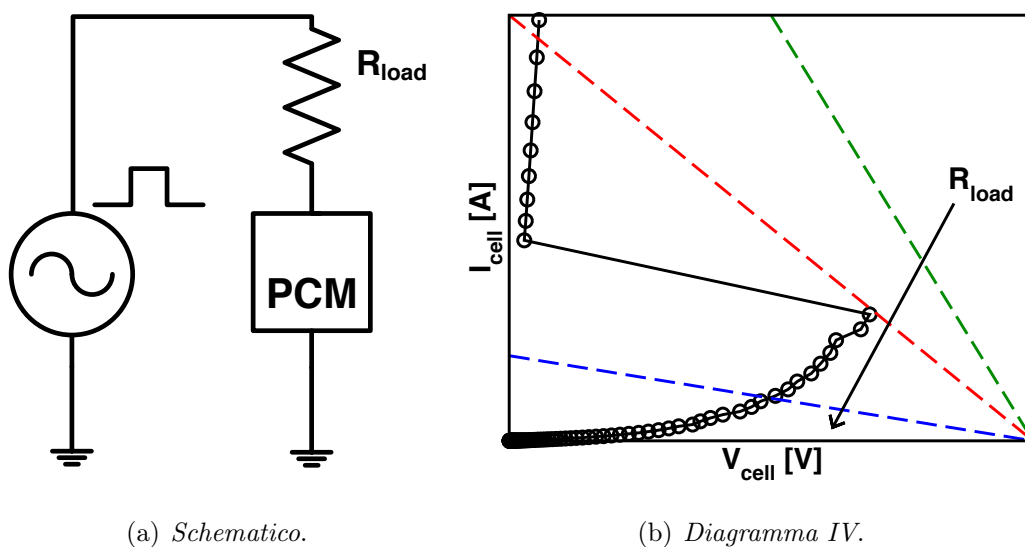


Figura 2.2: Caratteristica IV di cella PCM in stato di *full reset* (pallini), insieme alle rette di carico di resistori di valore crescente (linee tratteggiate).

Difatti la tensione sulla serie resistenza-cella è imposta, e da una lettura ad oscilloscopio della tensione sul resistore è possibile dedurre la tensione effettiva sulla cella e la corrente che fluisce in essa, tramite la semplice equazione (2.1):

$$I_{\text{cell}} = \frac{V_{\text{resistor}}}{R} \quad (2.1)$$

Evidentemente carichi elevati consentono di esplorare con un passo di discretizzazione di corrente più piccolo la regione di interesse sul piano I-V, anche se limitano la massima corrente massima erogabile al materiale calcogenuro, e quindi la possibilità di porre in stato di *reset* la cella.

### 2.2.2 La struttura 1T1R

Un altro tipo di struttura adottata è quella di una cella con geometria di tipo *wall* a  $90\text{nm}$ , avente in serie un transistor n-MOS con lunghezza di canale  $L = 1\mu\text{m}$ , da cui la denominazione 1T1R,  $T$  per transistor e  $R$  per resistore. Il circuito equivalente è mostrato in Figura 2.3).

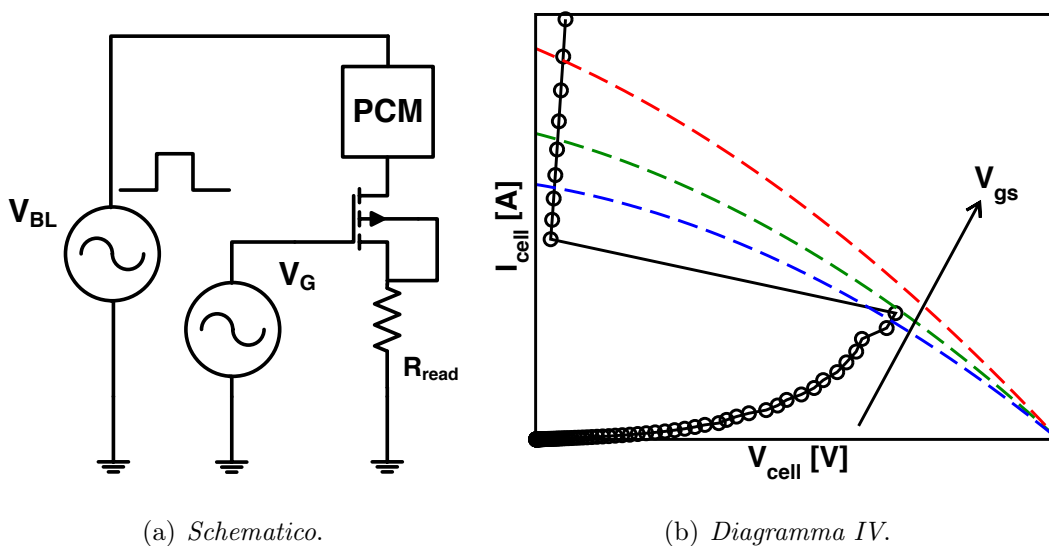


Figura 2.3: Caratteristica IV di cella PCM in stato di *full reset* (pallini), insieme alle caratteristiche di transistori polarizzati a diverse  $V_{\text{gs}}$  (linee tratteggiate).

Per eseguire uno studio approfondito del comportamento del dispositivo, è necessario conoscere la tensione ai capi dell'elemento di memoria. Ciò è possibile misurando la corrente che attraversa il campione. Infatti, conoscendo la tensione ai lati e la caratteristica di uscita del MOS, si può stimare la caduta sul selettore e sottrarla alla differenza di potenziale totale.

Chiaramente l'utilizzo di un transistor al posto di una resistenza da un lato consente con un solo dispositivo di raggiungere i punti di lavoro di interesse con diverse "rette di carico" (Figura 2.3), dall'altro aumenta la complessità della struttura e richiede uno studio a parte del comportamento del selettore MOSFET. A questo scopo disponibili sullo stesso *wafers* dei campioni con il solo selettore, integrato esattamente con la stessa tecnologia e con gli stessi parametri di quello in serie alla cella. Ciò ha permesso di collocarne uno nel criostato per determinarne sperimentalmente la caratteristica al variare della temperatura.

## 2.3 Struttura 1T1R: caratterizzazione del selettore in temperatura

La misura per caratterizzare il MOS si può effettuare col *Parameter*, si collegano 3 dei suoi canali sui contatti di *Gate*, *Drain* e *Source* (in cortocircuito col *Body*), poi si richiama un programma apposito che genera delle rampe di tensione sul *Gate* e delle rampe di tensione sul *Drain* mentre traccia la corrente di *Source*. Questo tipo di misura è in "quasi" continua e si approssima che la caratteristica sia la stessa anche alle frequenze degli impulsi di programmazione.

La caratterizzazione è stata eseguita a differenti temperature, le stesse poi usate per le altre misure. In Figura 2.4 si mostra il grafico della corrente di *Drain* in funzione della tensione di *Gate*.



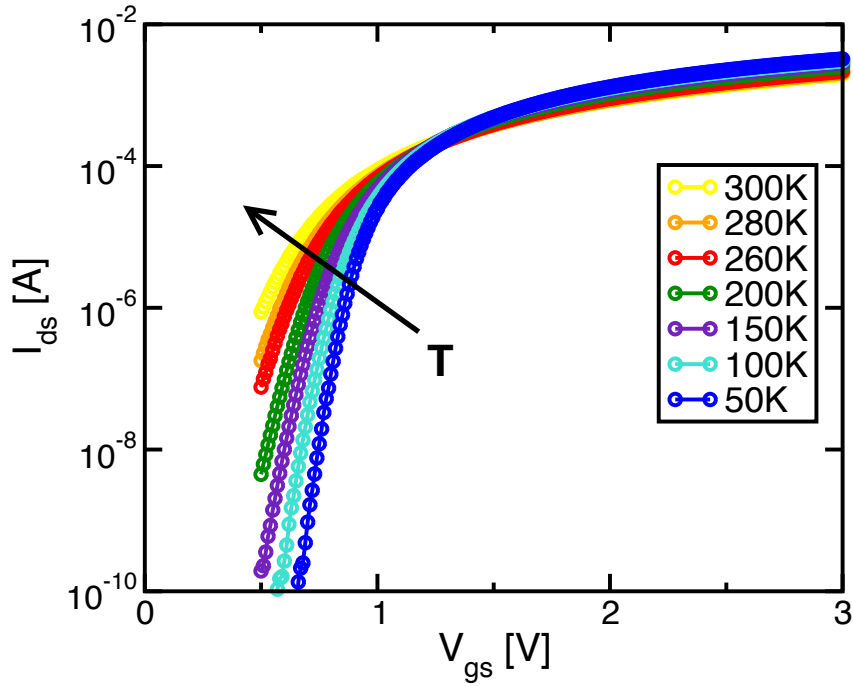


Figura 2.4: Trans-caratteristica di ingresso del selettore a  $V_{ds}$  pari a 3V.

Nel frattempo si è stimata la temperatura dalla *Sub-Threshold-Slope* del transistor, usando la formula nota dalla teoria del transistor MOSFET sottosoglia:

$$STS = \frac{V_{th}(1 + m)}{\log_{10}e} \quad (2.2)$$

approssimando  $m$  costante e pari a quello ricavato a temperature più alte. Il risultato è in Figura 2.5.

Per le misure successive, si è ritenuto più affidabile il sensore del criostato, evitando per ragioni cautelative di scendere a temperature inferiori ai 100K.

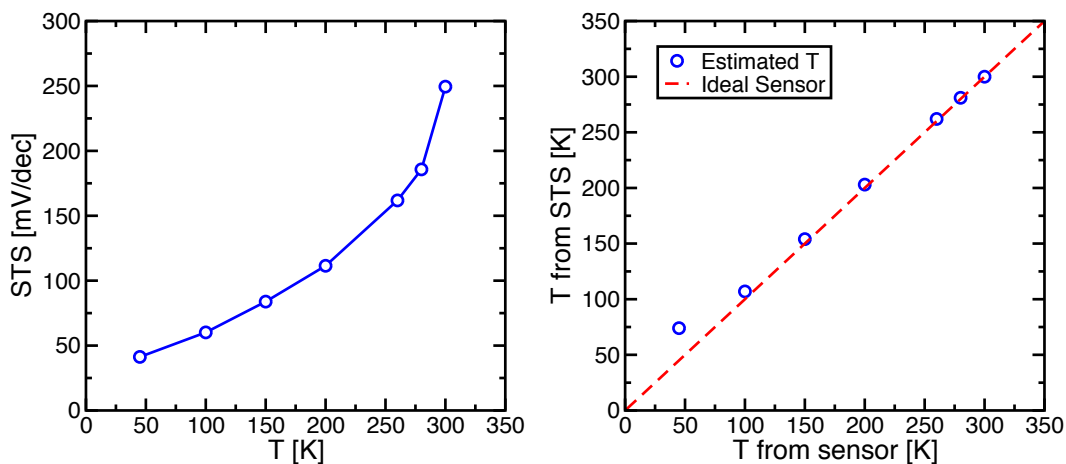


Figura 2.5: *Sub-Threshold-Slope* a sinistra e temperatura stimata a destra, entrambe sono in funzione della temperatura misurata dal sensore del criostato.

## 2.4 Apparato di misura a temperatura ambiente

### 2.4.1 Metodologia di misura

L'apparato di misura che verrà mostrato da qui in poi è quello relativo alla struttura 1T1R, che è quello che ha richiesto maggior lavoro per il suo miglioramento in termini di rumore, tempi di misura e ripetibilità.

Per poter raccogliere una sufficiente quantità di dati è necessario automatizzare il ciclo di misura. Infatti gli strumenti sono dotati di una funzionalità che ne permette il controllo totale tramite una connessione ad un computer (GPIB). Da questa necessità sorge un problema: gli strumenti devono poter funzionare senza variazione dei collegamenti. Quindi serve un circuito che permetta di disaccoppiare i vari strumenti, permettendo a ciascuno di funzionare nella fase di misura di loro competenza. Il primo circuito pensato a questo scopo è quello mostrato in Figura 2.6.

La scatola metallica ha cinque connettori: due sono riservati al *Parameter*, uno all'oscilloscopio e due al generatore di segnali, il *Pulser*. Le connessioni

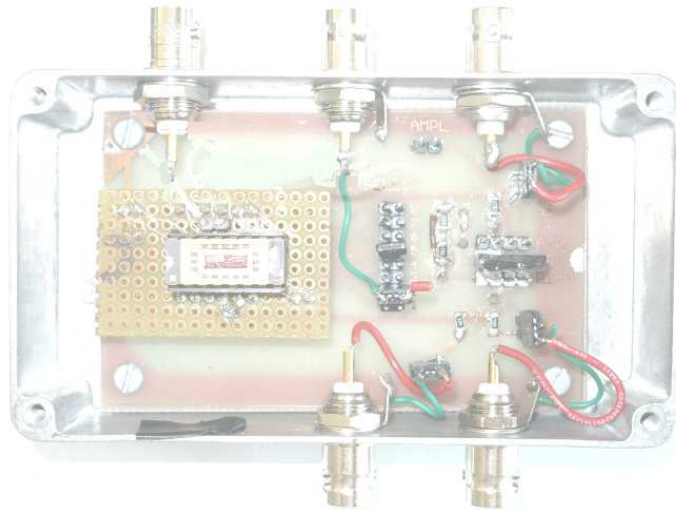


Figura 2.6: Contenitore schermato per il circuito per le misure a temperatura ambiente.

risultanti danno vita allo schema elettrico di Figura 2.7.

Per descrivere brevemente il funzionamento, è possibile considerare il circuito in figura come se fosse suddiviso in due parti, ognuna si attiva a turno per la programmazione o per la lettura dello stato della cella. La parte funzionante a frequenze più alte (nel dominio di frequenza degli impulsi di programmazione) è composta dal *Pulser*, dal circuito e dall'oscilloscopio. Infatti, il *Parameter* ha i canali in serie a resistori di valore elevato, così durante la fase di programmazione la corrente che scorre in essi è di valore trascurabile. La seconda parte funziona a bassa frequenza, sufficientemente bassa da concedere il tempo di caricarsi al condensatore. Dunque il *Pulser*, durante questa fase, non genera impulsi, permettendo al *Parameter* di misurare la resistenza. Il *Parameter* funziona, in questo caso, con una tipica configurazione di misura: quella *Kelvin*. Essa consiste nell'imporre una corrente da un canale, e rilevare la tensione dell'altro. Siccome su quest'ultimo passa una corrente idealmente nulla, la caduta di tensione sulla resistenza in serie è pressoché uguale a zero. Così mediante una

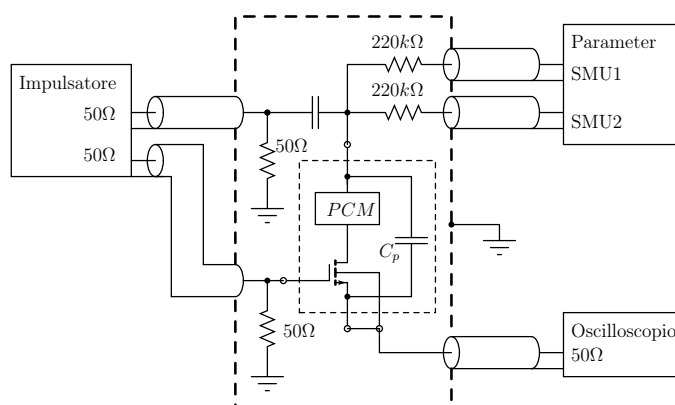


Figura 2.7: Schema dei collegamenti per l'apparato convenzionale.

semplice divisione tra i valori di tensione rilevata e corrente imposta abbiamo il risultato desiderato.

La funzione degli ultimi due collegamenti è più semplice ed intuitiva: Il secondo canale del *Pulser* controlla il gate del transistor in entrambe le fasi. L'oscilloscopio è impostato per operare con un'impedenza di  $50\Omega$ , quindi permette di ricavare la corrente in serie al prototipo. Infatti è sufficiente campionare la tensione sul canale e dividere i campioni per il valore dell'impedenza mostrata dall'oscilloscopio, ovvero  $50\Omega$ .

## 2.4.2 Ottimizzazione dell'apparato a temperatura ambiente

In questa sezione sono messe in discussione le prestazioni dell'apparato descritto precedentemente e usato in lavori antecedenti a questo. In particolare, si è ritenuto necessario provare a diminuire l'ampiezza del rumore sulla misura tempo-risolta della corrente.

Il risultato della modifica al circuito è mostrato in Figura 2.8. Ovviamente la sollecitazione corrispondente ai dati illustrati è un impulso di programmazione,

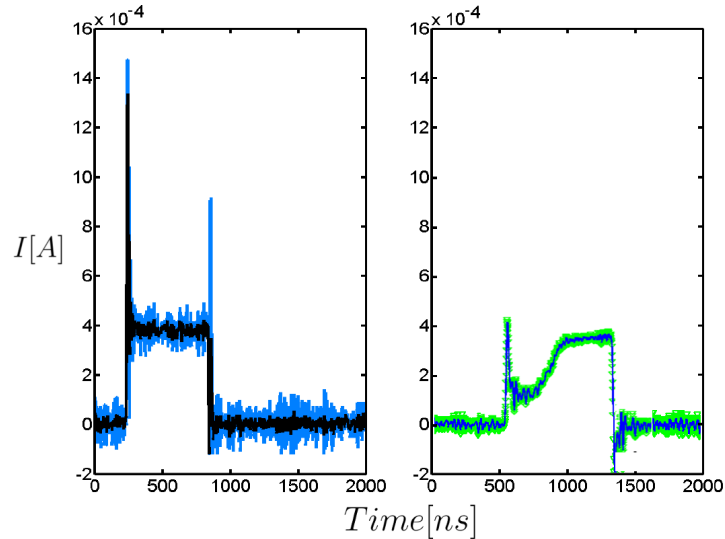


Figura 2.8: A sinistra, la corrente misurata col vecchio circuito; a destra dopo la modifica. I due impulsi sono stati ottenuti su diverse tecnologie, tant'è che nella figura a destra è visibile il fenomeno di *switching* a soglia, mentre a sinistra a no. È importante osservare come il nuovo *setup* consenta una riduzione del rumore, permettendo di misurare con successo la corrente di off (corrente prima dello *switching* a soglia), che nel vecchio *setup* non è rivelabile perché confrontabile col rumore elettronico.

quindi di forma quadra. Questo miglioramento sostanziale del rumore sulla corrente è stato ottenuto con il circuito modificato di Figura 2.9.

I contatti di *Source* e *Body* del prototipo erano collegati direttamente ad un canale dell'oscilloscopio, il quale aveva impedenza pari a  $50\Omega$ . Nella nuova configurazione invece, la corrente attraversa un resistore di valore 10 volte maggiore, la conseguenza è che la tensione campionata è scalata dello stesso fattore. È stato poi verificato empiricamente che il rapporto segnale-rumore migliora. Il potenziamento è dovuto al fatto che la fonte di rumore prevalente è situata a valle del circuito, altrimenti il rumore sarebbe aumentato allo stesso modo del segnale.

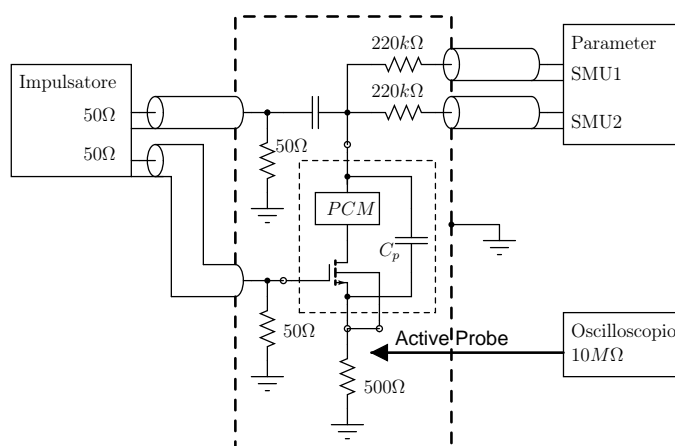


Figura 2.9: Schema dei collegamenti per l'apparato ottimizzato.

L'idea insita nella modifica ha permesso di misurare con ottima precisione piccole correnti. Il beneficio è soprattutto nella parte di caratterizzazione, che vuole una stima della cosiddetta *corrente di off*. Per comprendere il problema, si faccia riferimento al grafico di destra della Figura 2.8, la corrente di off è la corrente media all'inizio dell'impulso, tra il primo e il secondo fronte di salita. Si osservi come con la nuova configurazione la corrente di rumore sia più piccola della corrente di off, consentendo una sua misurazione.

## 2.5 Apparato di misura a temperature criogeniche

Al fine di indagare a fondo i meccanismi fisici che regolano il funzionamento delle celle è possibile sfruttare un grado di libertà importante nella caratterizzazione, cioè la possibilità di variare la temperatura del campione. Per questo elaborato, le misure di interesse sono state ripetute da  $300K$  ad un minimo di  $50K$ .

Quando si deve scendere a temperature criogeniche per effettuare misure

sulle celle, si usa un altro apparato sperimentale che verrà esposto cominciando con l'introdurre la Figura 2.10, che riassume schematicamente tutte le parti che lo compongono e le loro interconnessioni.

### 2.5.1 Principi di funzionamento del criostato

La parte principale è quella centrale, il criostato, all'interno del quale si raggiungono temperature prossime allo zero assoluto, idealmente circa  $4K$ . Il criostato si compone di una camera a vuoto (parte superiore), all'interno della quale si trova la *testa fredda* sulla cui sommità vengono posizionati i campioni (rettangolino rosso in figura), e di un *cryocooler* (parte inferiore).

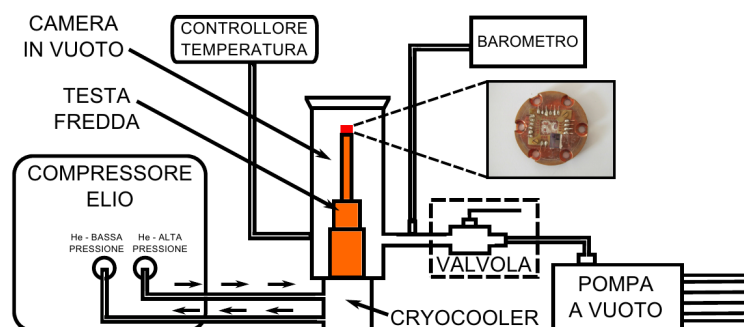


Figura 2.10: Schema rappresentativo dell'apparato sperimentale atto alle misure a basse temperature.

L'apparato in questione è del tipo a ciclo chiuso con raffreddamento ad elio liquido. La testa fredda viene necessariamente a trovarsi in una camera a vuoto in quanto deve essere eliminato il trasporto di calore che può avvenire per conduzione o per convezione dall'esterno. Per un approfondimento sul criostato e sul vecchio *set-up* è possibile fare riferimento alla documentazione di laboratorio o ad altri elaborati [3]. Di seguito si descrive brevemente la vecchia configurazione per poi introdurre il nuovo *set-up*.

## 2.5.2 Metodologia di misura

Come per il circuito fuori dal criostato, la misura va automatizzata. Questo scopo è stato raggiunto usando un circuito più evoluto e complicato di quello a temperatura ambiente, infatti esso fa uso di interruttori analogici per connettere o disconnettere gli strumenti alla cella. Lo schema elettrico è mostrato in Figura 2.11.

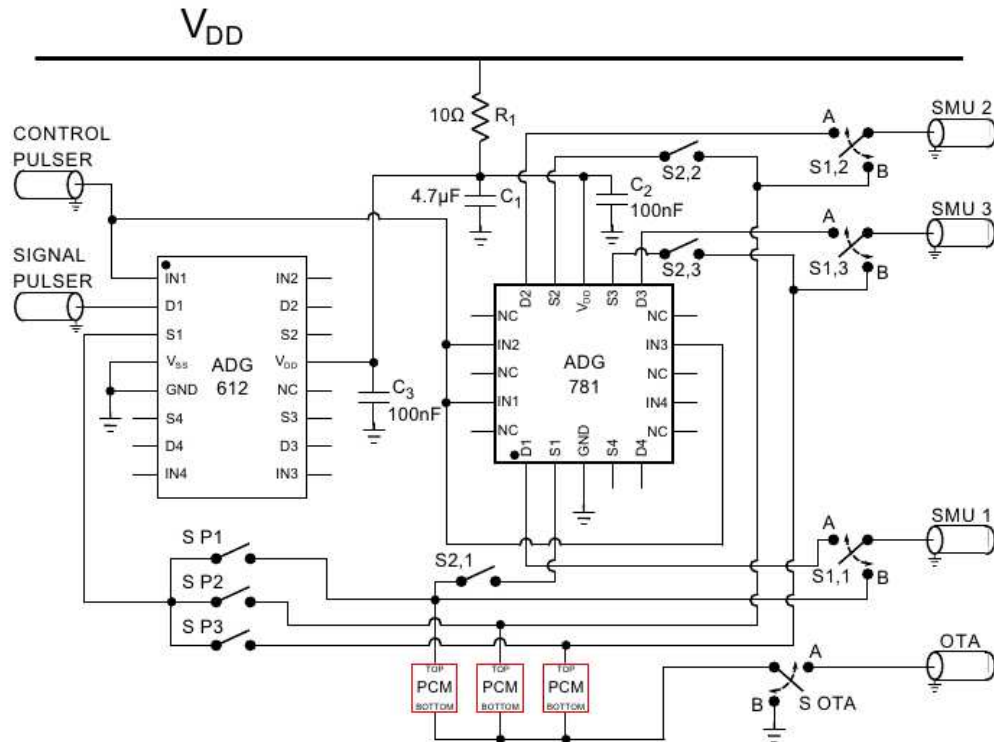


Figura 2.11: Schema circuitale del dispositivo di interfacciamento col criostato.

Il dispositivo associato allo schema di Figura 2.11 funziona da interfaccia tra i segnali che entrano nel criostato e quelli all'esterno. Infatti, il grosso connettore industriale della scatola metallica in figura va inserito nella corrispondente presa del criostato. Così facendo, si collegano elettricamente il circuito e i fili di circa 80cm, che partono dal connettore per arrivare a contattare il prototipo sulla testa fredda. Le prese coassiali sono usate per la strumentazione.



### 2.5.3 Un nuovo *set-up* per la struttura 1T1R [24]

I componenti descritti nella prima parte di questa sezione sono stati usati in lavori antecedenti a questo per fare misure non risolte in tempo. In questo tipo di misure è di fondamentale importanza il campionamento della corrente che attraversa la cella. Per questo motivo, l'ottimizzazione dell'apparato è assolutamente necessaria, come già evidenziato in questo capitolo. Un'altra differenza importante in questa attività di caratterizzazione è che c'è un prototipo con selettore, dunque serve un segnale, non previsto nel vecchio *set-up*, per la polarizzazione del transistor.

Il progetto delle modifiche ha impedito per diversi giorni l'attività di misura. A causa dei problemi di ripetibilità riscontrati durante le prove, è stata addirittura presa in esame la possibilità di rinunciare ai dati sulla corrente. In seguito è descritta la piccola modifica apportata per controllare il selettore, per poi trattare dettagliatamente l'aspetto del miglioramento nella risoluzione temporale del segnale rivelato.

#### 2.5.3.1 Abilitazione dell'apparato al controllo del selettore

Il transistor di selezione deve poter essere controllato e attivato a piacere in ogni momento, sia in fase di programmazione sia durante la lettura.

Per accedere elettricamente al *Gate* serve un canale isolato nel dispositivo di interfacciamento. Quindi è stato usato uno dei canali ausiliari, i quali possono collegare il campione con una linea coassiale esterna. Inoltre è opportuno adattare l'impedenza al termine della linea coassiale, e va dunque inserito un resistore da  $50\Omega$ .

Lo strumento adatto al comando del *Gate* può essere il *Pulser*, il quale ha quattro canali. Perciò è possibile controllare con esso sia i segnali di *Gate*, sia gli impulsi sul nodo superiore della cella, sia il segnale per gli interruttori analogici.

### 2.5.3.2 Progetto della modifica per la misura tempo-risolta

L'obiettivo di questa parte del lavoro è stato aggiungere al vecchio *set-up* la possibilità di misurare la corrente. A causa dell'aspetto sperimentale di quest'attività non è possibile progettare il circuito senza fare prove sul campo. Infatti, si possono verificare errori difficilmente prevedibili a priori, come problemi di rumore, di accoppiamento capacitivo e di *crossstalk* tra i cavi.

Il primo tentativo di misurare la corrente è stato quello di collegare il contatto di *Source* del prototipo ad un canale dell'oscilloscopio, impostando l'impedenza di ingresso a  $50\Omega$ . Questa soluzione è simile a quella adottata a temperatura ambiente (prima dell'ottimizzazione), tranne per il fatto che, per poter effettuare il collegamento, è necessario un filo elettrico di circa  $80\text{cm}$  tra scatola metallica e prototipo. Il rapporto segnale-rumore (SNR) di questa configurazione è risultato inaccettabile, peggiore di quello previsto. Successivamente è stata provata la soluzione adottata a temperatura ambiente (vedere sezione 2.4.2), cioè con sonda attiva e resistore da  $500\Omega$ , tuttavia le prestazioni non sono migliorate. Dunque, si è ritenuto necessario indagare sull'origine del rumore elettronico riscontrato.

### 2.5.3.3 Rumore sulla misura

Le fonti di rumore prevalenti sono state identificate variando i diversi collegamenti e accendendo gli strumenti uno alla volta. Da questa analisi sperimentale si è visto che il controllore di temperatura dell'apparato dà un contributo significativo, mentre quello della pompa per il vuoto risulta trascurabile. Un altro disturbo è stato riscontrato con l'accensione del *Parameter*.

Per risolvere questi problemi è stato possibile e sufficiente effettuare dei collegamenti per migliorare i riferimenti di massa. Sono stati aggiunti dei cavi in rame con un diametro di circa  $3\text{mm}$  tra le masse delle fonti di rumore e la struttura del criostato, collegata elettricamente a terra. È inoltre

stata cortocircuitata la massa comune del *Parameter* con le masse degli altri strumenti.

### 2.5.3.4 Prove preliminari

Le misure preliminari hanno dato risultati non coerenti sotto il profilo della corrente misurata. La curva RI ricavata era traslata e dispersa lungo l'asse della corrente, il confronto con i risultati a temperatura ambiente è mostrato in Figura 2.12.

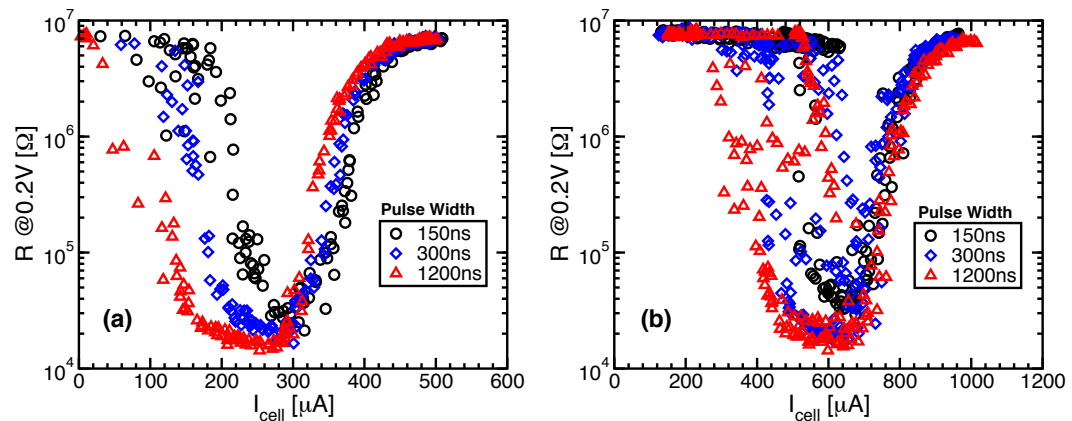


Figura 2.12: In (a) sono riportate le corrette curve di programmazione, rilevate con l'apparato per le misure a temperatura ambiente. In (b) i risultati (errati) delle prove preliminari rilevati nel criostato a temperatura ambiente. Si noti come in (b) vi sia una sovrastima della corrente di circa un fattore due, oltre che una indeterminazione nel suo valore (*spread* dei valori di corrente). Entrambi i grafici rappresentano dati a tre diverse durate dell'impulso di programmazione.

La presenza di un'accoppiamento capacitivo tra il cavo che porta l'impulso di BitLine all'interno del criostato e la camera a vuoto del criostato stesso è in grado di spiegare il comportamento anomalo osservato:

- i) la presenza di una corrente di spostamento significativa per segnali ad alta frequenza (come gli impulsi di programmazione sulla *BitLine*, dell'ordine

dei  $100ns$ ) porta a leggere una corrente sulla resistenza di degenerazione al *Source* superiore rispetto a quella che fluisce in realtà nella serie PCM/MOSFET. Ne discende una *traslazione* verso destra della curva di programmazione lungo l'asse delle correnti;

- ii) la corrente di spostamento è una frazione della corrente che scorre nella serie PCM/MOSFET, e avrà quindi in prima approssimazione un valore proporzionale alla tensione di *BitLine*. Conseguenza di ciò è il fatto che a diverse coppie *WordLine/BitLine* cui competono in teoria la stessa corrente, mostrano sulla resistenza di degenerazione una corrente diversa, perché diversa è la corrente di spostamento ad esse associata. Questo è il motivo della *dispersione* delle curve di programmazione lungo l'asse delle correnti.

### 2.5.3.5 Collegamenti del nuovo *set-up*

I problemi comparsi durante le prove preliminari riguardano solo il funzionamento ad alta frequenza. Si è concepito un modo per rendere il circuito precedente immune ad accoppiamenti o disturbi, è stata valutata la possibilità di un inserimento nel criostato di un cavo coassiale.

Un cavo coassiale rispetto ai fili già presenti ha un diametro totale superiore, dunque è capace di trasportare molto più calore alla testa del criostato. Questo effetto è stato mitigato con due accorgimenti. Il primo è l'avvolgimento del cavo coassiale attorno alla parete interna del criostato, così il cavo arriverà in prossimità del prototipo con una temperatura considerevolmente minore. Il secondo è interporre tra cavo e prototipo qualche centimetro di filo più sottile. Il sistema progettato ha lo schema equivalente di Figura 2.13.

I contatti *Source* e *Body* del prototipo sono posti a massa, in questo modo la tensione di *Gate* corrisponde sia alla  $V_{gs}$ , sia alla  $V_{g-body}$  del selettore. Così facendo si elimina si ha reiezione dell'*effetto body* e delle possibili fluttuazioni

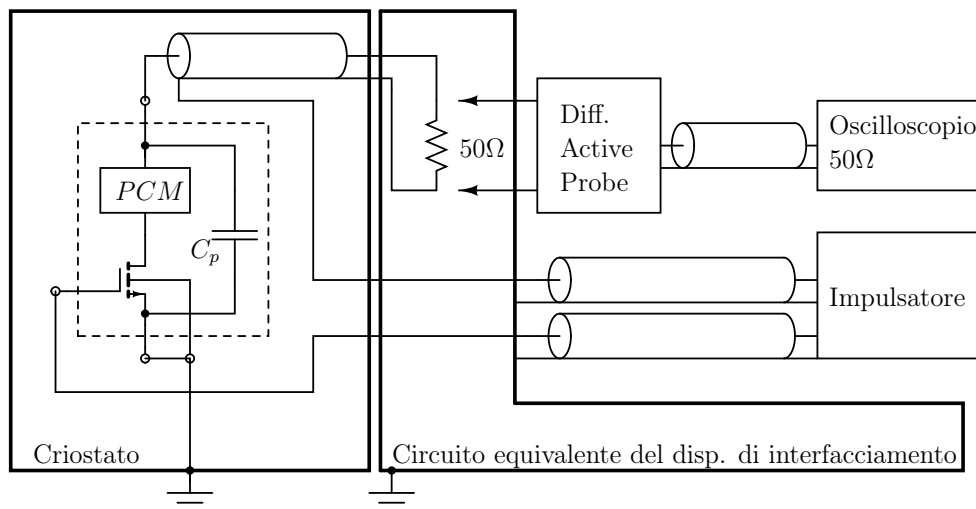


Figura 2.13: Schema dei collegamenti del nuovo *set-up* con il circuito equivalente per la fase di programmazione.

della tensione sul nodo di *Source*. Per la lettura di corrente, è stata ideata una configurazione particolare, che è descritta nelle prossime righe.

Un filo elettrico porta il segnale del *Pulser* sulla sommità della testa fredda. Questo segnale vede un'impedenza di  $50\Omega$  verso il lato superiore della cella, l'impedenza in questione è quella vista all'estremità del cavo coassiale ben terminato. La misura di corrente è effettuata campionando la tensione all'altro lato del cavo coassiale. Infatti siccome il collegamento è adattato, la tensione generata vicino al prototipo viene integralmente trasportata in fondo al cavo, dove una sonda attiva differenziale la può campionare. Basta così dividere i valori campionati dall'oscilloscopio per 50 per ottenere una precisa misura della corrente che effettivamente scorre in prossimità del prototipo.

La soluzione adottata è immune al rumore (grazie alla schermatura del cavo) e anche al *crosstalk* tra i segnali (la sonda è differenziale). Il risultato

finale è mostrato nella Figura 2.14



Figura 2.14: A sinistra, il dispositivo di interfacciamento con il connettore costruito in laboratorio per la sonda attiva differenziale. A destra, il criostato è aperto e pronto per le misure. Si vede il cavo coassiale bianco partire dal fondo e arrivare alla sommità girando intorno alla parete interna.

#### 2.5.3.6 Contromisure adottate per preservare il prototipo

Durante l'attività sperimentale, è stato necessario accendere e spegnere la strumentazione mentre il campione era connesso, ciò può provocare picchi di corrente indesiderati provenienti dalle linee di alimentazione. Questi disturbi possono, nei casi peggiori, danneggiare il campione.

Dall'esperienza fatta in laboratorio con questo apparato, si è visto che è sufficiente cortocircuitare i connettori della scatola metallica e lasciare alimentati gli interruttori analogici. Così si crea un circuito chiuso tra contatti del prototipo e massa.

#### 2.5.3.7 Considerazioni sulla durata delle misure

L'Impulsatore è stato il soggetto di alcune riflessioni sull'ottimizzazione del tempo di misura, le quali sono riassunte nelle prossime righe. Le prove di automatizzazione del funzionamento del *Pulser*, hanno evidenziato tempi di

programmazione decisamente superiori, sia al periodo effettivo degli impulsi sia ai tempi di programmazione degli altri strumenti.

La durata totale del ciclo di misura (reset-lettura-set-lettura) ha delle conseguenze non banali. La prima è che può ritardare la disponibilità dei dati per l'analisi dei risultati. Un altro aspetto importante è la comodità di poter predire la fine di una sessione di misura, per poi essere presenti per iniziarne un'altra. Ma soprattutto, il vantaggio fondamentale nell'avere un ciclo breve è avere una sessione di misura a durata minima, ciò permette di ridurre al minimo i rischi di effettuare misure affette da errori accidentali.

Per esempio durante l'attività svolta è stato necessario spegnere l'apparato per riaccenderlo il giorno successivo. Le operazioni di spegnimento e accensione rendono alta la probabilità di danneggiamento del dispositivo, rendendo inutili tutti i dati raccolti in precedenza. Per diminuire il tempo di misura, è stata fatta un'ottimizzazione del *software* di controllo degli strumenti, ad esempio, il *Pulser* può salvare in memoria le impostazioni e poi lo si può sollecitare a richiamarle. Insieme all'uso di questa funzionalità, è stato aggiunto un secondo impulsatore, il quale può essere utilizzato a turno con l'altro. Il ciclo di misura è stato così ridotto da circa 50s a 27s, un miglioramento importante.

## 2.6 Caratterizzazione preliminare del nuovo apparato

Per verificare il buon funzionamento del nuovo apparato, sono state fatte delle misure preliminari a temperatura ambiente, in Figura 2.15 è mostrata una curva R-I ricavata da questi dati. Si osservi come ora, a differenza di quanto avveniva nel *setup* precedente (Figura 2.12), l'accoppiamento capacitivo dell'impulso di BitLine con la linea di massa sia trascurabile. La riduzione dell'accoppiamento è osservabile da due evidenze sperimentali:

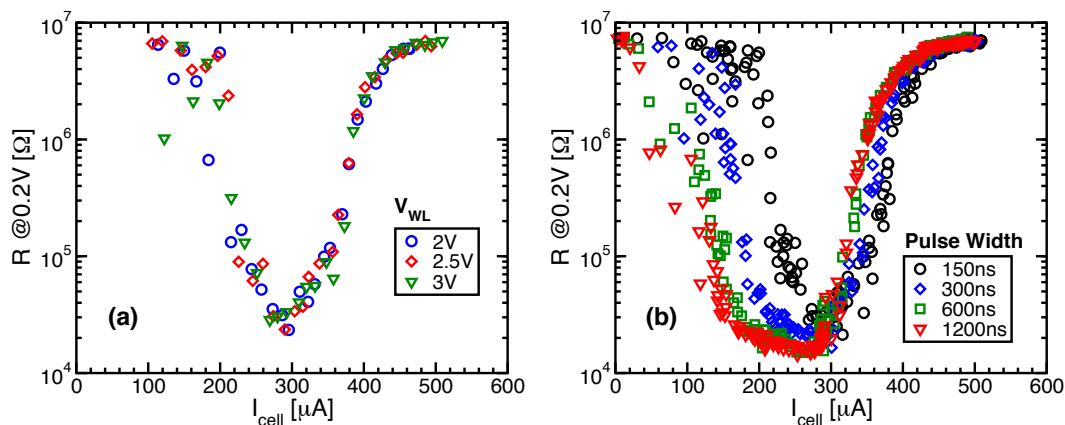


Figura 2.15: In (a) è riportata una curva di programmazione RI per tempo di programmazione  $t_{prog} = 150ns.$ , in cui ciascuna serie corrisponde ad una diversa tensione di WordLine. Si osservi come con il nuovo *setup* la resistenza raggiunta non dipenda dalla particolare combinazione WordLine/BitLine (cosa che avveniva in Figura 2.12(b)), ma solo dalla corrente fornita dall'impulso. In (b) sono riportate curve di programmazione della stessa cella ottenute per diverse durate dell'impulso di programmazione  $t_{prog}$ : i valori di corrente sono ora coerenti con quelli ottenuti con l'apparato convenzionale (v. Figura 2.12(a)).

- i) ad una certa resistenza programmata corrisponde ora un ben preciso valore di corrente di programmazione. La ragione di ciò è chiara osservando al Figura 2.15(a): a diverse coppie WL/BL non corrispondono più diversi accoppiamenti capacitivi, e quindi le curve di programmazione ottenute con diversa tensione di *WordLine* coincidono;
- ii) non c'è più una sovrastima della corrente, e le curve ottenute sono ora coerenti con quelle ottenute col *setup* convenzionale (v. Figura 2.12(a)). Il motivo è ancora lo stesso: la riduzione dell'accoppiamento capacitivo porta a non leggere la corrente di spostamento in eccesso, che precedentemente era confrontabile con la corrente che realmente fluiva nella serie cella-MOSFET.



## Capitolo 3

# Modello di *drift* della tensione di soglia

*Nel seguente capitolo viene introdotto il fenomeno dello switching a soglia (Par.3.1.1), analizzando sia la dipendenza dal rilassamento strutturale (Par.3.1.2), sia dallo stato programmato (Par.3.1.3). Nella Sezione 3.2 vengono riassunti i risultati della teoria classica della conduzione sottosoglia ad alto campo nei semiconduttori amorfi, distinguendo tra due diversi meccanismi di abbassamento di barriera, il Poole e il Poole-Frenkel. Nella Sezione 3.3 si riprende la teoria esposta in [25] dello switching elettronico visto in termini di acquisto di energia dei portatori alle spese del campo elettrico. Questa teoria viene applicata al fenomeno del drift della tensione di soglia, con l'introduzione innovativa di una regola di Meyer-Neldel per il calcolo della corrente di sottosoglia. Segue nella Sezione 3.4 il confronto tra le previsioni del modello con i dati sperimentali ottenuto su una cella in tecnologia 90nm, comparando i risultati ottenuti con l'abbassamento di barriera Poole e Poole-Frenkel.*

### 3.1 Lo *switching* a soglia e i fenomeni di *drift*

#### 3.1.1 Lo *switching* a soglia (*threshold switching*)

Lo *switching* a soglia (*Threshold Switching*) consiste in un marcato e repentino aumento di conduttività elettrica osservato nella fase amorfa di diversi materiali calcogenuri, cui corrisponde una regione sul piano IV a resistenza differenziale negativa (*Negative Differential Resistance*, NDR, Fig 3.1).

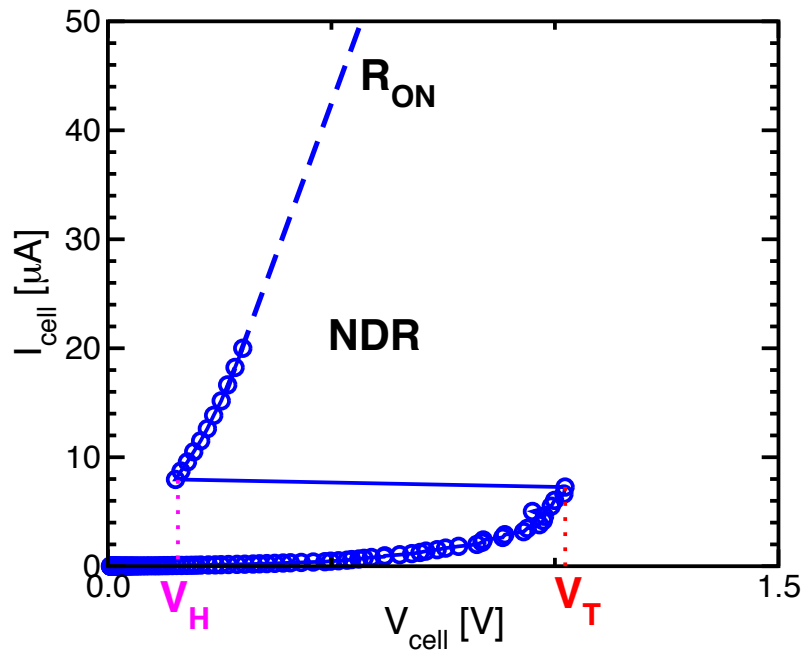


Figura 3.1: Diagramma IV di una cella PCM posta in stato iniziale di *full reset* che presenta *switching* a soglia. Sono visibili la tensione di soglia  $V_T$ , la tensione di *hold*  $V_H$  e la zona a resistenza differenziale negativa (NDR). Dopo lo *switching* a soglia la resistenza associata al calcogenuro è molto bassa, e domina la resistenza dell'*heater* posto in serie alla cella (comportamento ohmico,  $R_{ON}$ ).

È importante sottolineare che questo fenomeno è distinto dalla transizione di fase amorfa/cristallina, anche se in certe condizioni e per certi materiali, come il GST, allo *switching* a soglia corrisponde anche una transizione di fase

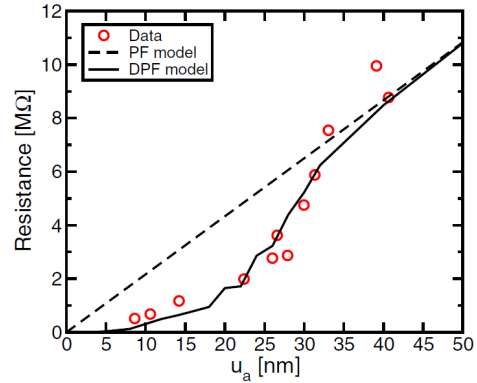
[26]. La concomitanza dei due eventi può essere spiegata osservando che il forte incremento di corrente causato dalla diminuzione di resistenza elettrica genera un aumento consistente di temperatura nel materiale calcogenuro per *Joule heating*. A seconda della durata dell'impulso elettrico, lo *switching* può contribuire alla transizione di fase per i materiali aventi bassa temperatura di fusione [27].

La tensione elettrica per cui si innesca il comportamento NDR di un cella PCM è detta *Tensione di Soglia* ( $V_T$ ) ed è un parametro molto importante dal punto delle applicazioni, perché definisce il confine tra l'intervallo di tensioni utili per la lettura e quelle utili per la scrittura della cella. Difatti si desidera che la tensione di lettura abbia un valore inferiore a  $V_T$ , in modo tale che l'operazione di *Read* sia non distruttiva, e che la tensione di scrittura abbia invece un valore superiore a  $V_T$ , cosicchè la temperatura raggiunta per *Joule heating* nella cella sia sufficiente a far cristallizzare la fase amorfa in tempi sufficientemente rapidi (centinaia di nanosecondi). Inoltre, il tempo di *switching*, stimato essere in un range di tempi inferiore al nanosecondo [28], rappresenta il limite ultimo per la velocità di programmazione nelle PCM.

D'altra parte comprendere il fenomeno di *switching* a soglia è un punto importante anche per lo studio di base della struttura a bande e dei meccanismi di trasporto nei semiconduttori amorfi. Difatti lo *switching* a soglia è stato scoperto essere di natura elettronica, ma non è ancora certo se la conduzione nella fase amorfa sia da attribuire a portatori intrappolati (polaroni) oppure ad emissione termica per meccanismo Poole-Frenkel di elettroni intrappolati. Una descrizione fisica dettagliata dello *switching* a soglia può dunque contribuire alla comprensione concettuale dei meccanismi di trasporto nei semiconduttori amorfi, in particolare in condizione di non equilibrio [25].

### 3.1.2 Dipendenza dallo stato programmato

Un modo semplice di modificare il valore di  $V_T$  è quello di porre la cella in uno stato intermedio tra quello di *set* e quello di *full reset*, imponendo una tensione di programmazione posta tra quella di questi due stati. Così facendo si ottiene un raggio della calotta amorfa  $u_a$  di diversa grandezza (Figura 3.2(a)).



(a) Rappresentazione in sezione della variazione di  $u_a$  in seguito all'applicazione di tensioni di programmazione crescenti.

(b) Resistenza di basso campo in funzione di  $u_a$ , dedotta dalla STS. Sono mostrati i risultati della teoria DPF e della PF.

Figura 3.2: Variazione di  $u_a$  a seguito di diversa tensione di programmazione (a), e suo effetto sulla resistenza di basso campo (b). Figure tratte da Fugazza et al., IEDM 2009 [29]

Come ovvia conseguenza, l'aumento di  $u_a$  provoca un aumento di resistenza. Tuttavia, per effetti legati alla percolazione degli elettroni nel materiale amorfo, non vi è una proporzionalità diretta tra il raggio della calotta amorfa  $u_a$  e la resistenza di basso campo, come si può vedere dalla Figura 3.2(b)). Difatti, se ci si limita a considerare l'effetto della geometria la resistenza di basso campo dovrebbe scalare linearmente con  $u_a$ , ma ad una calotta più piccola corrisponde una maggiore probabilità di percorsi percolativi attivi, il che si traduce in un'energia di attivazione per la conduzione in fase amorfa che diminuisce con  $u_a$  [29], da cui l'andamento sub-lineare di Figura 3.2(b).

Oltre che influire sulla resistenza di basso campo, ad un diverso valore di  $u_a$  corrisponde una diversa tensione di soglia [30]. Un primo modo per comprendere intuitivamente questo andamento è quello di osservare che il volume della zona amorfa scala con  $u_a$ , e così la tensione necessaria a portare il gas di elettroni sufficientemente fuori equilibrio termodinamico. Queste argomentazioni verranno approfondite più dettagliatamente nella Sezione 3.3. Per adesso basti osservare la Figura 3.3, dove è illustrata la correlazione tra resistenza di basso campo e  $V_T$ , dovuta alla variazione della tensione di programmazione, e quindi di  $u_a$ .

### 3.1.3 Dipendenza dal tempo dopo programmazione (*drift*)

Come sottolineato in precedenza, la tensione di soglia è un parametro fondamentale per il funzionamento della PCM come elemento di memoria, e per questo motivo qualsiasi meccanismo responsabile di una sua variazione deve essere attentamente considerato. In tale contesto, un fenomeno sicuramente da tenere in considerazione è l'aumento della tensione di soglia col tempo trascorso dopo che l'impulso di programmazione ha posto la cella in uno stato di *partial* o *full reset* (*drift* di soglia).

Come nel caso di diverso stato programmato, ad un aumento di tensione di soglia corrisponde un aumento di resistenza di basso campo della fase amorfa (*drift* di resistenza, Figura 3.3), secondo l'andamento:

$$R(t) = R_0 \left( \frac{t}{t_0} \right)^\nu \quad (3.1)$$

dove  $R_0$  e  $t_0$  sono delle costanti, e  $\nu \approx 0.1$  fornisce la pendenza del diagramma bilogaritmico di  $R$  in funzione di  $t$ . La presenza di una causa comune al *drift* di soglia e al *drift* di resistenza viene considerata e sviluppata nelle prossime sezioni 3.3.2.

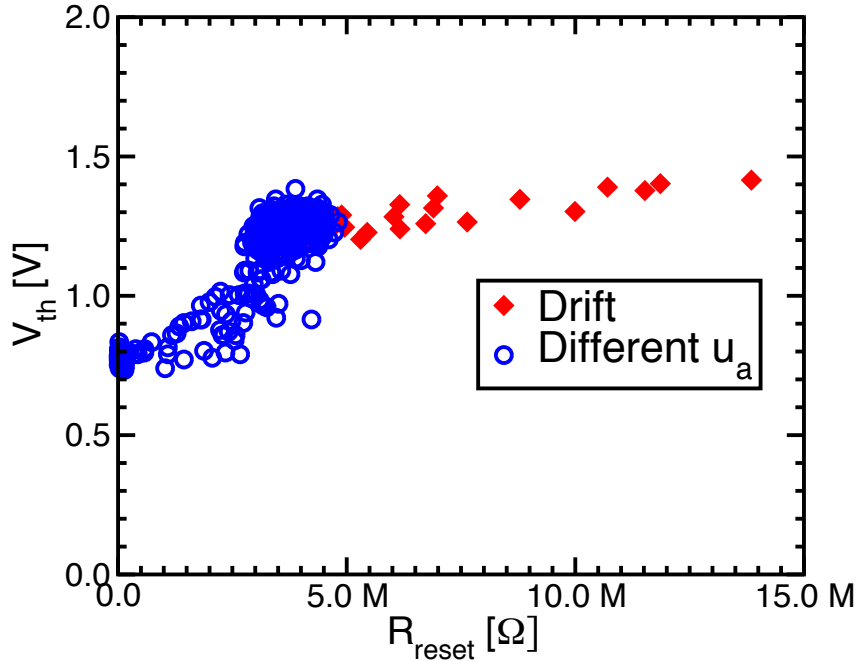


Figura 3.3: Scatter plot di  $V_T$  Vs.  $R$  di basso campo. Cerchi: punti corrispondenti a variazione dello stato programmato (i.e. diversa  $u_a$ ). Rombi: punti a diverso tempo dopo programmazione (*drift*).

## 3.2 Modelli di conduzione nella fase amorfa di GST

### 3.2.1 Diagramma a bande nel GST amorfo

La struttura a bande del GST amorfo differisce sostanzialmente da quella del GST cristallino. Facendo riferimento alla Figura 3.4, le tre principali differenze che possono essere evidenziate sono [31]:

- i) generalmente l'*Energy Gap* della fase amorfa è grande comparato con quello della fase cristallina. Nel caso del GST, i dati di assorbimento ottico e di fotoconduttività indicano che il *gap* per la fase amorfa è di circa  $0.7eV$ , mentre per la fase cristallina è di circa  $0.5eV$ ;

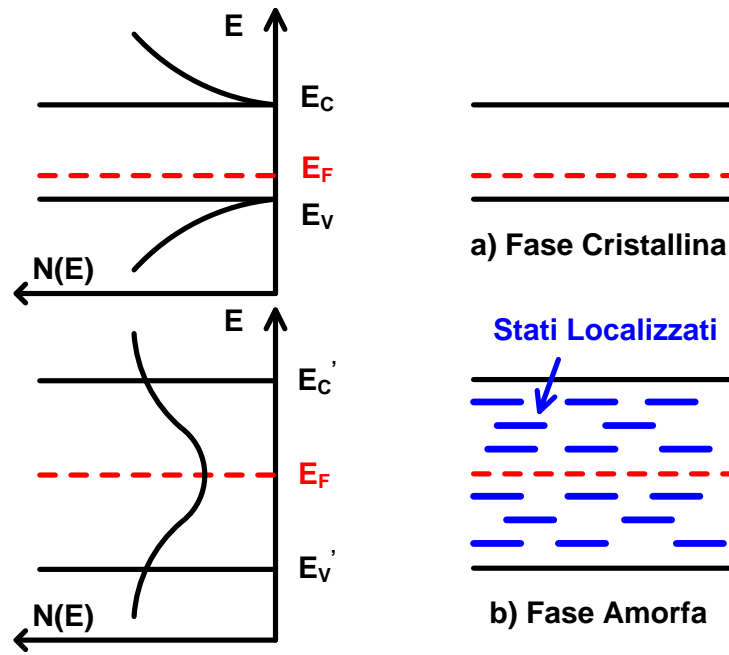


Figura 3.4: Rappresentazione schematica del diagramma a bande di un semiconduttore calcogenuro nella fase cristallina (a) e nella fase amorfa(b).

- ii) mentre il *gap* proibito e le bande di conduzione/valenza sono chiaramente distinte nella fase cristallina, la fase amorfa è caratterizzata da una larga concentrazione di stati localizzati nel *gap*. Questi stati localizzati contribuiscono significativamente alla densità di stati, di conseguenza la grossa differenza tra *gap* e bande di conduzione/valenza non risiede nella disponibilità o meno di stati occupabili, come avviene nella fase cristallina, ma quanto nella loro posizione all'interno dell'*Energy Gap*. Stati profondi sono più legati da interazioni di tipo coulombiane con gli atomi vicini mentre la mobilità aumenta per energie che si avvicinano alla banda di conduzione o di valenza. Per questo motivo, l'*Energy Gap* nel semiconduttore amorfo è generalmente denominato *Mobility Gap*.
- iii) nel GST amorfo il livello di Fermi, a causa della grande concentrazione di stati localizzati all'interno del *Band Gap*, è bloccato a circa metà *gap*

(*pinning* del livello di Fermi). Per questa ragione, la resistenza elettrica e la rispettiva energia di attivazione  $E_A$  sono relativamente alte. Nel GST amorfo si osservano energie di attivazione intorno a  $0.25 - 0.35 eV$ , mentre nella sua fase cristallina si osservano energie di attivazione più basse, che sono compatibili con un comportamento di *self-doping*, probabilmente causato da difetti di punto (e.g. vacanze).

Vista la grande concentrazione di stati donori e accettori provocata dall'elevata densità di difetti della struttura disordinata, il livello di Fermi tende a posizionarsi al centro dell'*Energy Gap*, portando ad un'alta resistività e ad un'alta energia di attivazione per la conduzione [32]. Questo aspetto è stato evidenziato anche in esperimenti sul GST amorfo, per cui l'energia di attivazione per la conduzione è stata dedotta essere all'incirca pari a  $E_G/2$ .

### 3.2.2 Conduzione ad alto campo

Per sviluppare un modello realistico di *switching* a soglia, è necessario comprendere la conduzione elettrica nella fase amorfa dei materiali calcogenuri in condizioni di quasi-equilibrio. Per studiare questi aspetti, dobbiamo considerare la regione sotto-soglia nel piano IV. Uno studio in temperatura delle caratteristiche IV permette di ottenere il valore dell'energia di attivazione per la conduzione  $E_A$  al variare della tensione di sotto-soglia applicata. Le curve IV indica un andamento lineare per tensioni molto basse (regione lineare), e un incremento esponenziale della corrente a tensioni più alte (regione esponenziale). Quello che si osserva dalla misura della resistenza in regione esponenziale al variare della temperatura è che l'energia di attivazione per la conduzione decresce all'aumentare della tensione applicata. Questo andamento è un segno evidente del trasporto dei portatori per effetti di emissione termica da barriera, che si abbassa in risposta all'applicazione di un campo elettrico [33].



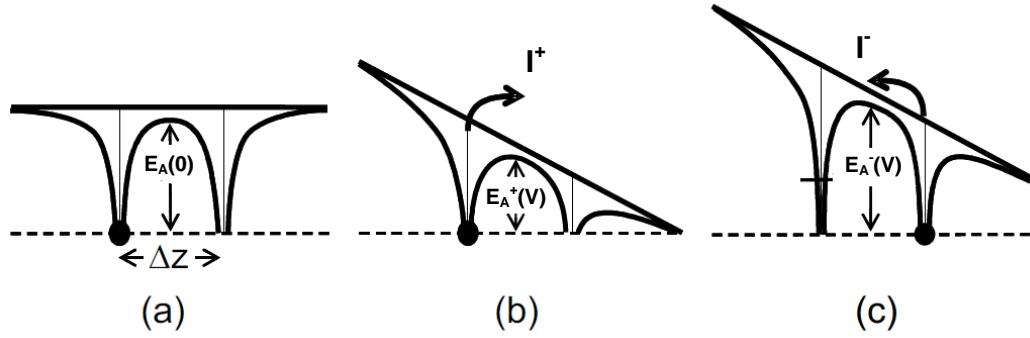


Figura 3.5: Rappresentazione schematica della barriera di potenziale nell'intorno di una trappola. La corrente è dovuta all'emissione degli elettroni da una trappola ad una adiacente sopra la barriera di potenziale. La barriera di potenziale è pari a  $E_A(0)$  quando nessuna tensione è applicata (a), e scende a  $E_A^+(V)$  quando è applicata una tensione  $V$ , col risultato di un aumento esponenziale di  $I^+$  (b). Il contributo di corrente opposto  $I^-$  al campo è esponenzialmente depresso per l'aumento di barriera  $E_A^-$ . Immagine tratta da D. Ielmini, Phys.Rev.Lett 2008 [25].

Ora, facendo riferimento alla Figura 3.5, dove è rappresentata la situazione di un portatore posto in uno stato trappola, detta  $\Delta z$  la distanza media tra una trappola e la prima vicina ed  $\varepsilon$  la costante dielettrica del materiale amorfo, è possibile ottenere l'andamento del potenziale elettrico in funzione del campo elettrico  $F$ , che risulta essere:

$$\phi(r) = \frac{1}{4\pi\varepsilon|r|} - \frac{1}{4\pi\varepsilon|\Delta z - r|} - qFr \quad (3.2)$$

Il valore dell'abbassamento di barriera indotto dal campo elettrico  $F$  si ottiene cercando il punto di massimo del potenziale di Eq.(3.2) per  $0 < r < \Delta z$ :

$$\frac{d\phi}{dr} = 0 \Rightarrow \frac{q^2}{4\pi\varepsilon} \left( \frac{1}{(\Delta z - r)^2} - \frac{1}{r^2} \right) - qF = 0 \quad (3.3)$$

e sottraendolo al punto di massimo ottenuto quando  $F = 0$ . Vengono ora presi in considerazione due diversi modelli che esprimono l'abbassamento di barriera indotto da campo elettrico.

### 3.2.2.1 Abbassamento di barriera secondo Poole

Se la densità volumetrica di trappole è molto elevata, una semplice espressione analitica dell'abbassamento di barriera è ottenuto risolvendo l'Eq.(3.3) nel limite  $\Delta z \rightarrow 0$ . In questa condizione il punto di massimo si trova circa nel punto medio tra le due trappole (*i.e.* per  $r = \Delta z/2$ ). Considerando una cella PCM avente raggio della calotta amorfa  $u_a$ , si ottiene un abbassamento di barriera (*Poole Barrier Lowering*) pari a:

$$\frac{d\phi}{dr} = -qF \Rightarrow \Delta\varphi(V) = -qV \frac{\Delta z}{2u_a} \quad (3.4)$$

La barriera per l'emissione termica al netto dell'abbassamento di barriera è quindi data da:

$$E_A^\pm(V) = E'_C - E_{F0} \mp qV \frac{\Delta z}{2u_a} \quad (3.5)$$

dove col  $E_A^+$  si indica l'altezza di barriera per emissione termica in direzione del campo e con  $E_A^-$  in direzione opposta ad esso. Facendo riferimento alla Figura 3.5b, gli elettroni con energia tra  $E$  e  $E + dE$  contribuiscono alla corrente in direzione del campo con un contributo  $dI^+$ , dato da [31]:

$$dI^+ = qA_a N_T(E) e^{-\frac{E_A^+(V)}{k_B T}} \frac{\Delta z}{\tau_0} dE \quad (3.6)$$

dove  $q$  è la carica dell'elettrone,  $A_a$  è l'area trasversale della calotta amorfa,  $N_T$  è la concentrazione totale di trappole che contribuiscono alla conduzione (*i.e.* da  $E_{F0}$  a  $E'_C$ ). La velocità media elettronica è data dal rapporto tra  $\tau_0$ , che è il tempo caratteristico di fuga di un elettrone posto nel *Mobility Gap* (tipicamente  $\tau \sim 10^{-13} - 10^{-14} s$ ), e la probabilità di fuga,  $P_f = \exp(-E_A(V)/k_B T)$ .

In aggiunta alla corrente in diretta di Eq.(3.6), deve essere preso in considerazione anche il contributo di corrente in direzione opposta al campo (Figura 3.5c). In questo caso, la barriera per l'emissione termica è incrementata di un valore pari all'opposto di  $\Delta\varphi$  di Eq.(3.4). Quindi il contributo di corrente opposta al campo degli elettroni con energia tra  $E$  ed  $E + dE$  nel *Mobility Gap*

è:

$$dI^- = qA_a N_T(E) e^{-\frac{E_A^-(V)}{k_B T}} \frac{\Delta z}{\tau_0} dE \quad (3.7)$$

La corrente totale sarà data dalla somma integrale dei contributi netti di corrente energia per energia:

$$I_P = \int_{E_{F0}}^{E'_C} (dI^+ - dI^-) dE = 2qA_a N_T \frac{\Delta z}{\tau_0} e^{-\frac{E'_C - E_{F0}}{k_B T}} \sinh\left(\frac{qV}{k_B T} \frac{\Delta z}{2u_a}\right) \quad (3.8)$$

Il modello prevede una dipendenza lineare di  $\log I$  dalla tensione.

### 3.2.2.2 Abbassamento di barriera secondo Poole-Frenkel

Se la densità volumetrica di trappole è bassa, un'espressione analitica semplice dell'abbassamento di barriera è ottenibile risolvendo l'Eq.(3.3) nel limite di  $\Delta z \rightarrow \infty$ . Indicando con  $r_{max}$  la distanza dalla trappola del massimo di barriera, si ottiene (*Poole-Frenkel Barrier Lowering*):

$$r_{max} = \sqrt{\frac{q}{4\pi\epsilon F}} \quad (3.9)$$

e quindi, inserendo l'espressione nell'Eq.(3.2):

$$\Delta\varphi(F) = -q\sqrt{\frac{qF}{\pi\epsilon}} \Rightarrow \Delta\varphi(V) = -q\sqrt{\frac{qV}{\pi\epsilon u_a}} \quad (3.10)$$

L'altezza di barriera sarà perciò data da:

$$E_A^\pm = E'_C - E_{F0} \mp q\sqrt{\frac{qV}{\pi\epsilon u_a}} \quad (3.11)$$

Inserendo l'altezza di barriera ottenuta nell'espressione della corrente ricavata in precedenza per il caso Poole (Eq.(3.8)) otteniamo l'espressione:

$$I_{PF} = 2qA_a N_T \frac{\Delta z}{\tau_0} e^{-\frac{E'_C - E_{F0}}{k_B T}} \sinh\left(\frac{q}{k_B T} \sqrt{\frac{qV}{\pi\epsilon u_a}}\right) \quad (3.12)$$

Si osservi che in questo caso il modello prevede una dipendenza di  $\log I$  dalla radice quadrata della tensione.

### 3.2.3 Resistenza di basso campo

Un'espressione significativa della resistenza di basso campo può essere ottenuta studiando il comportamento dell'Eq.(3.8) per piccole polarizzazioni (*i.e.*  $\Delta\varphi(V) \ll k_B T$ ). Ricordando che, per  $x \ll 1$ ,  $\sinh x \sim x$ , otteniamo una corrente di basso campo pari a:

$$I_{LF} = qA_a N_T \frac{\Delta z}{\tau_0} e^{-\frac{E'_C - E_{F0}}{k_B T}} \frac{qV}{k_B T} \frac{\Delta z}{2u_a} \quad (3.13)$$

e quindi una resistenza pari a:

$$R_{LF} = \left( \frac{I_{LF}}{V} \right)^{-1} = R_0 e^{\frac{E'_C - E_{F0}}{k_B T}} \quad (3.14)$$

Come è semplice osservare, la resistenza di basso campo fornisce un metodo diretto per il calcolo dell'energia di attivazione per la conduzione. Basta difatti eseguire delle misure in temperatura e valutare la pendenza del grafico  $\log R$  Vs  $1/k_B T$  (*Arrhenius plot*).

## 3.3 Modello di *switching* a soglia

### 3.3.1 Il criterio di potenza di soglia costante

La Figura 3.6a mostra schematicamente il *Mobility Gap* di un calcogenuro amorfo, delimitato dalle energie  $E'_C$  ed  $E'_V$ , con il livello di Fermi  $E_{F0}$  bloccato a mezzo *gap* (3.2.1). L'applicazione di una tensione elevata può portare all'acquisto di energia dei portatori a spese del campo elettrico, stabilendo una distribuzione elettronica di fuori-equilibrio, come in Figura 3.6b. Considerando per semplicità solo il trasporto elettronico, l'eccesso medio di energia per un elettrone in posizione  $z$  lungo la direzione del trasporto è dato dalla differenza tra il quasi-livello di Fermi  $E_F$  e il livello di Fermi all'equilibrio termodinamico,  $E_{F0}$ .

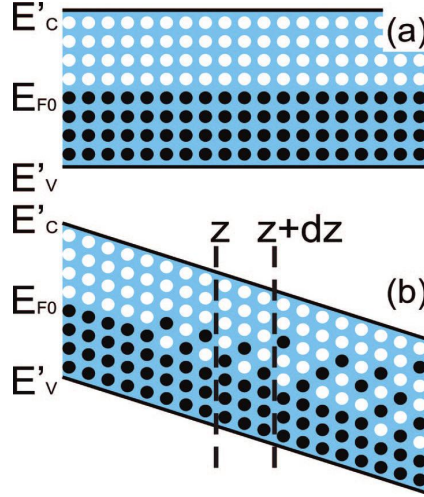


Figura 3.6: (a) Rappresentazione schematica della distribuzione in energia degli elettroni nel calcogenuro amorfo all'equilibrio.  $E'_c$  ed  $E'_v$  rappresentano le energie di confine per il *Mobility Gap* in banda di conduzione e in banda di valenza, rispettivamente.  $E_{F0}$  è il livello di Fermi. (b) Stessa rappresentazione di (a) ma in condizioni di fuori-equilibrio per alti campi.

Mentre il campo elettrico tende ad incrementare  $E_F$ , i processi di *scattering* anelastico riducono l'energia in eccesso. L'effetto dello *scattering* anelastico può essere espresso al prim'ordine come:

$$\left. \frac{dE_F}{dt} \right|_{\text{scatt}} = - \frac{E_F - E_{F0}}{\tau_{rel}} \quad (3.15)$$

dove il tempo di rilassamento dell'energia  $\tau_{rel}$  può essere associato ad interazioni elettrone-fonone negli stati localizzati. L'equazione di bilancio tra il guadagno di energia dal campo e lo *scattering* anelastico nel tratto di materiale amorfo tra  $z$  e  $z + dz$  è data da:

$$\frac{J(z)E_F(z)}{q} - \frac{J(z+dz)E_F(z+dz)}{q} + J(z)F(z)dz - \frac{E_F - E_{F0}}{\tau_{rel}} n_T(z)dz = 0 \quad (3.16)$$

Nell'Eq.(3.16) il primo e il secondo termine tengono conto del flusso energetico verso il materiale in  $z$  e fuori da esso in  $z + dz$ . Il terzo ed il quarto termine rappresentano rispettivamente il guadagno e la perdita di energia, avendo

indicato con  $n_T$  la concentrazione di elettroni intrappolati che contribuiscono alla corrente. Il valore di  $n_T$  può essere ottenuto dalla densità degli stati come:

$$n_T = N_T \frac{k_B T}{E'_C - E_{F0}} \quad (3.17)$$

dove è assunto che la distribuzione di trappole  $N_T$  sia uniformemente distribuita in energia e che solo le trappole entro un  $k_B T$  dal livello di Fermi siano popolate, il che è equivalente ad integrare la distribuzione esponenziale decrescente di Maxwell-Boltzmann. Per l'equazione di continuità della corrente  $J(z) = J(z + dz)$ , l'Eq. (3.16) può essere semplificata in:

$$\frac{dE_F}{dz} = qF - \frac{qn_T}{J} \frac{E_F - E_{F0}}{\tau_{rel}} \quad (3.18)$$

dove il primo ed il secondo termine sulla destra descrivono rispettivamente i meccanismi di guadagno e di perdita di energia dei portatori.

Ora, se si indica con  $\gamma$  l'eccesso di energia in unità di  $k_B T$  [ $\gamma = (E_F - E_{F0})/k_B T$ ], il fenomeno di *switching* a soglia avviene solo quando  $\gamma$  raggiunge il valore critico  $\gamma_T$  nella regione di ON (che si verifica essere dell'ordine dell'unità, v. [25]), che causa il collasso del campo elettrico e l'inizio del comportamento NDR. Se si considera l'Eq.(3.17) e la situazione stazionaria associata all'Eq.(3.18) (ottenuta imponendo  $dE_F/dz = 0$ ) si può traslare la presenza di un  $\gamma_T$  in un *criterio di potenza di soglia costante*:

$$P_T''' = JF = \frac{\gamma_T N_T (k_B T)^2}{\tau_{rel} (E'_C - E_{F0})} \quad (3.19)$$

Una conferma sperimentale di questo criterio è mostrato in Figura 3.7(d), che mostra il valore di  $P_T$  al variare della tensione di programmazione (e quindi di  $u_a$ , v. Figura 3.7(a) e (b)). La potenza di soglia risulta praticamente indipendente dal raggio della calotta amorfa  $u_a$ . Per giungere a questa figura è stato necessario sottrarre alla potenza elettrica misurata la potenza dissipata nell'*heater*, ottenibile dal prodotto della corrente di soglia  $I_T$  e della tensione di hold  $V_H$ .

---

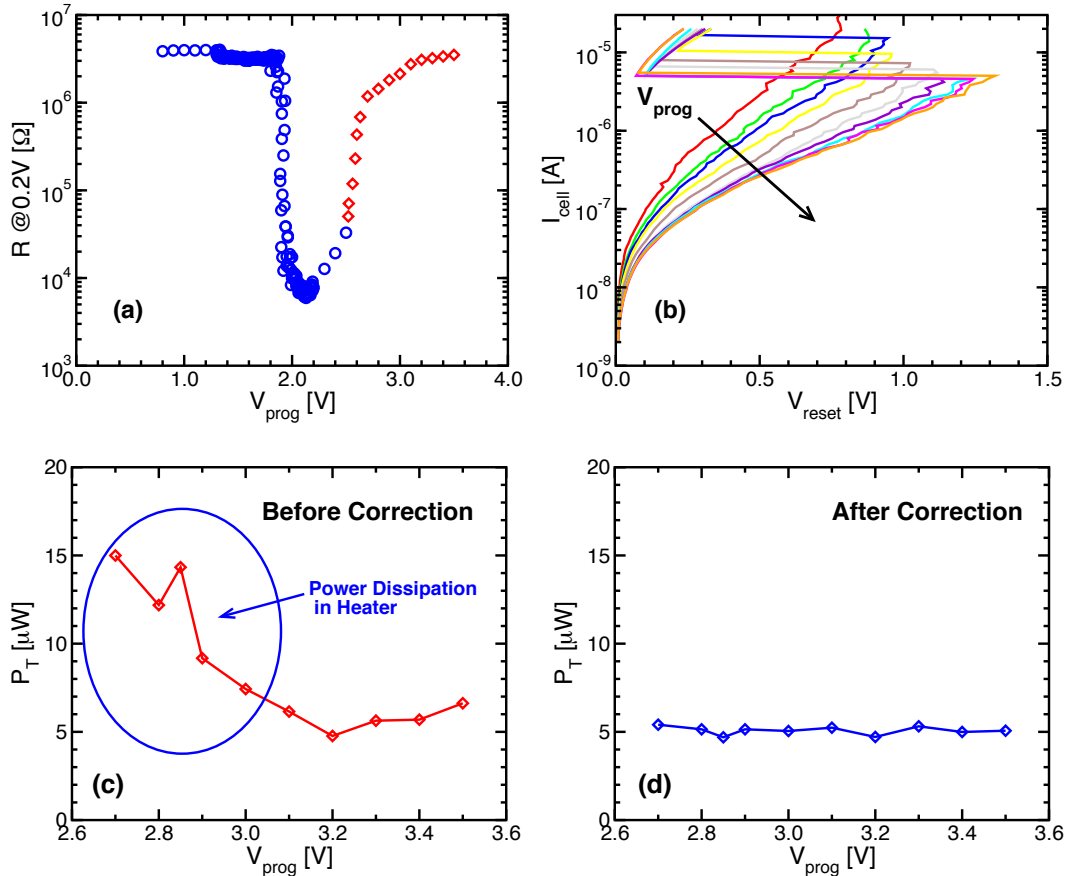


Figura 3.7: In (a) è mostrata la curva di programmazione della cella in esame (90nm 1R,  $R_{load} = 2k\Omega$ ), dove col simbolo ' $\diamond$ ' sono indicati gli stati di *reset* (i.e. gli stati con diverso  $u_a$ ). In (b) sono riportate le caratteristiche IV della cella posta nei diversi stati di *reset* evidenziati in (a). In (c) viene mostrato il valore della potenza di soglia  $P_T = I_T V_T$  in funzione della tensione di programmazione (e quindi per diverso  $u_a$ ). In (d) è riportata la potenza di soglia effettiva, ottenuta sottraendo ai punti in (c) la potenza dissipata nell'*heater*. Evidentemente il criterio  $P_T = costante$  è verificato.

In Figura 3.8(b) è riportato l'andamento di  $P_T$  in funzione del tempo dopo programmazione (*drift time*), ottenuto dalle caratteristiche IV in Figura 3.8(b). Come è evidente, la potenza di soglia è un parametro che a meno di fluttuazioni statistiche risulta indipendente dai fenomeni associati al *drift* che verranno introdotti nel Paragrafo 3.3.2. Si osservi come il valore della potenza di soglia

in Figura 3.8(b) e in Figura 3.7(d) siano lo stesso, pari circa a  $5\mu A$ .

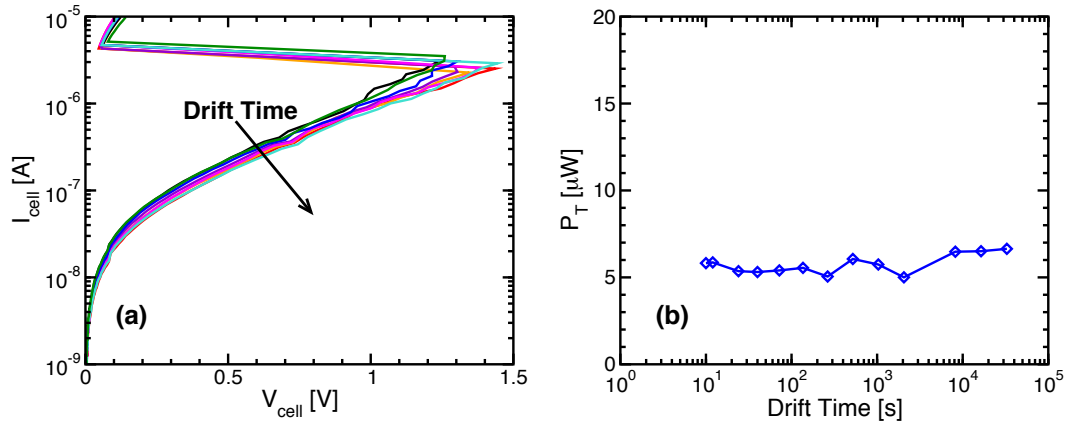


Figura 3.8: In (a) è sono riportate le caratteristiche IV della cella in esame al variare del tempo dopo programmazione (*drift time*). Dal prodotto  $P_T = I_T V_T$  di ciascuna curva è possibile ricavare il grafico in (b), che rappresenta il valore della potenza di soglia  $P_T$  in funzione del tempo dopo programmazione. Il criterio  $P_T = costante$  è verificato.

### 3.3.2 Il rilassamento strutturale e il *drift* dell'energia di attivazione

Il rilassamento strutturale (*Structural Relaxation*, SR) consiste in un riarrangiamento a livello atomico della struttura disordinata di un semiconduttore amorfo, che evolve spontaneamente verso una configurazione termodinamicamente più favorevole, ossia con minore energia libera di Gibbs. Quindi il rilassamento strutturale non deve essere visto come un vero e proprio cambiamento di fase, ma come un riordinamento a corto raggio della struttura amorfa, che mantiene comunque un sostanziale disordine a lungo raggio. Il fenomeno di rilassamento strutturale è stato osservato da analisi fisiche in molti materiali amorfi, tra cui i metalli vetrosi come Si, Ge, SiC, Se, o materiali a cambiamento di fase come le leghe calcogenure. Esperimenti di calorimetria suggeriscono che



SR è un processo esoenergetico, il che è concorde con l'interpretazione del fenomeno di rilassamento come spostamento del sistema verso una configurazione termodinamicamente più favorevole. Esperimenti di diffrazione indicano che la struttura a lungo raggio rimane inalterata, con un trascurabile cambiamento della distanza tra primi vicini, e un aumento dei legami del reticolo amorfo.

Per queste ragioni in questo lavoro si attribuisce il rilassamento strutturale ad una annichilimento di difetti nel calcogenuro amorfo. I difetti includono legami mancanti, distorsioni dell'angolo e/o della lunghezza dei legami, oppure vacanze, e sono modellizzati come stati metastabili nella struttura amorfa disordinata. Lo stato metastabile può rilassare ad uno stato di energeticamente più favorevole per emissione termica sopra una barriera di potenziale. Il tempo medio  $\tau$  per l'annichilimento del difetto a temperatura  $T$  sarà perciò dato da:

$$\tau(W_{ij}) = \tau_0 e^{\frac{W_{ij}}{k_B T}} \quad (3.20)$$

dove  $\tau_0$  è una costante di tempo tipica per la vibrazione atomica e  $W_{ij}$  è l'altezza di barriera per il rilassamento del difetto, che agisce come un'energia di attivazione per il processo SR. Da sottolineare il fatto che la regola di Meyer-Neldel è applicabile al fattore pre-esponenziale  $\tau_0$ , portando ad un'approccio unificato al trattamento dei fenomeni di rilassamento strutturale e di cristallizzazione nelle celle PCM [34].

L'energia di attivazione  $E_A$  nell'Eq.(3.20) si suppone avere valori molto distribuiti, vista la grande varietà di tempi di reazione osservata negli esperimenti. Mentre la cristallizzazione si osserva occorrere per un tempo ben preciso ad una determinata temperatura, l'evoluzione temporale di SR può essere osservata lungo tutto il tempo a disposizione nei tipici esperimenti di caratterizzazione, da qualche decina di nanosecondi, tempo in cui l'amorfo raggiunge una configurazione stabile dopo un riassetto iniziale [30], fino a qualche settimana. È evidente dalla Eq.(3.20) che l'estesa distribuzione dei tempi di rilassamento può essere spiegata solo con un'estesa distribuzione delle

energie di attivazione, comprese normalmente tra  $0eV$  e l'energia di attivazione tipica per la cristallizzazione posta circa a  $2.6eV$ .

Poiché i difetti corrispondono a stati localizzati nel *Mobility Gap*, la variazione nel tempo del numero di difetti per fenomeni di SR ha un notevole impatto sulle proprietà elettriche di conduzione e di *switching* a soglia del materiale calcogenuro nella sua fase amorfa. Di conseguenza, una descrizione quantitativa basata su modelli fisici del rilassamento strutturale nei dispositivi PCM richiede una modellizzazione dettagliata sia della cinetica dell'annichilazione dei difetti sia della conduzione termicamente attivata nella fase amorfa dei materiali calcogenuri. Per questi motivi le tipiche caratterizzazioni sperimentali includono misure accelerate in temperatura, al fine di diminuire i tempi caratteristici del rilassamento dei difetti.

Particolare attenzione è stata prestata allo studio del parametro  $\nu$  che governa il *drift* di resistenza (Eq. (3.1)), visto che questo parametro riveste un ruolo importante nella descrizione degli effetti del rilassamento strutturale sul comportamento elettrico dei dispositivi e sulla loro affidabilità. Da misure in temperatura su celle poste in diverso stato programmato, ossia con diverso valore di  $u_a$ , è possibile ottenere l'energia di attivazione per la conduzione  $E_A$  che compete a ciascuno stato.

Se per ciascuno degli stati programmati vengono effettuate misure di *drift* di resistenza, si è in grado di costruire un diagramma di correlazione tra il parametro  $\nu$  e l'energia di attivazione per la conduzione  $E_A$  (Figura 3.9). I dati suggeriscono che  $\nu \propto E_A$ . Combinando allora questo risultato sperimentale con l'Eq.(3.1) e l'Eq.(3.14) possiamo scrivere:

$$E_A(t) = E_{A0} \left[ 1 + \gamma \log \left( \frac{t}{t_0} \right) \right] \quad (3.21)$$

L'interpretazione fisica è immediata: il *drift* di resistenza può essere spiegato con un incremento di energia di attivazione per la conduzione causato dal rilassamento di difetti posti in stati metastabili. Gli stati metastabili rilassano

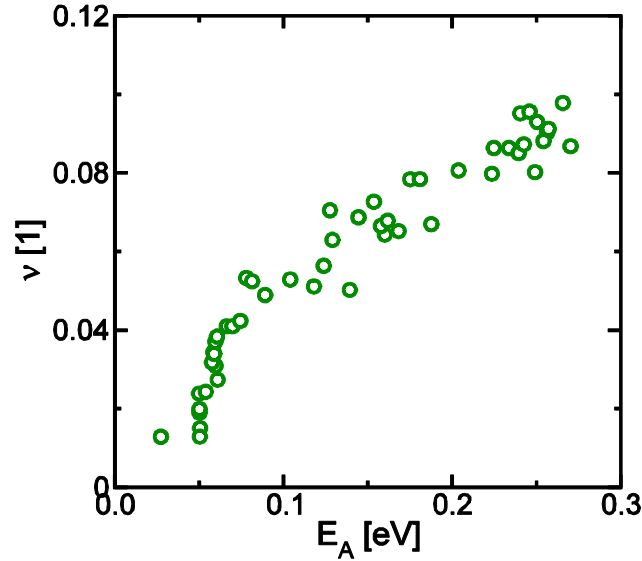


Figura 3.9: Diagramma di correlazione tra il parametro di *drift* di resistenza  $\nu$  e l'energia di attivazione per la conduzione  $E_A$ . Ogni punto rappresenta un diverso stato programmato, i.e. un diverso  $u_a$ . Grafico tratto da [35].

progressivamente mediante eccitazione termica ad uno stato di equilibrio energeticamente più favorevole con una cinetica governata dall'Eq.(3.20). Ciascun difetto, che come visto precedentemente(3.2.1) contribuisce alla conduzione in fase amorfa, dopo il rilassamento si trova ad avere una barriera energetica (distanza da  $E'_C$ ) accresciuta. Tenendo conto della statistica di rilassamento dei difetti, l'effetto complessivo è quello di un aumento di  $E_A$  nel tempo come da Eq.(3.21). Dall'Eq.(3.14) e l'Eq.(3.21) è infine possibile ottenere un'espressione di  $\nu$  in funzione dell'energia di attivazione per la conduzione e della temperatura:

$$\nu = \frac{d \log R}{d \log t} = \frac{\gamma E_{A0}}{k(1 - T/T_{MN})} \quad (3.22)$$

dove con  $T_{MN}$  si è indicata la temperatura di Meyer-Neldel. Applicando questa equazione ai dati di *drift* si ottiene un  $\gamma$  indipendente dall'energia di attivazione, a conferma della validità dell'Eq. (3.21).

### 3.4 Risultati sperimentali e previsioni teoriche

Vengono ora riportati i risultati sperimentali che riportano gli effetti del rilassamento strutturale su alcuni parametri elettrici significativi che influenzano il funzionamento delle celle PCM, come la tensione di soglia, la *Sub-Threshold Slope* (STS) e la resistenza di basso campo. Le misure sono state effettuate su una cella in tecnologia  $90nm$ , struttura 1R.

#### 3.4.1 Il *drift* della resistenza

La Figura 3.10 mostra l'andamento della resistenza di basso campo al variare del tempo dopo programmazione (*drift time*). L'andamento lineare in scala bilogarithmica è a conferma della validità dell'Eq.(3.1). Dal fitting dei punti sperimentali si ottiene  $\nu \simeq 0.11$ , e, facendo uso dell'Eq.(3.22),  $\gamma \simeq 2.53\%/dec$ .

#### 3.4.2 Applicazione della regola di Meyer-Neldel al calcolo della corrente

Si è già osservato come il tempo caratteristico di annichilimento di una trappola per fenomeni di rilassamento strutturale sia soggetta alla regola di Meyer-Neldel (Par.3.3.2)[34]. Per questi motivi è ragionevole supporre che il fattore pre-esponenziale  $\tau_0$  di Eq.(3.8) e di Eq.(3.12) sia soggetto alla regola di Meyer-Neldel:

$$\tau_0 = \tau_{00} e^{-\frac{E_A(V)}{k_B T_{MN}}} \quad (3.23)$$

Partendo da questa equazione possiamo trovare una forma conveniente della corrente di sotto-soglia in condizioni di alto campo. Sia nel caso di *Poole barrier lowering* (Eq.(3.8)), sia nel caso di *Poole-Frenkel barrier lowering* (Eq.(3.12)) si ottiene:

$$I(V) = I_{00} e^{\frac{E_A(V)}{k_B T_{MN}}} e^{-\frac{E_A(V)}{k_B T}} = I_{00} e^{-\frac{E_A(V)}{k' T}} \quad (3.24)$$

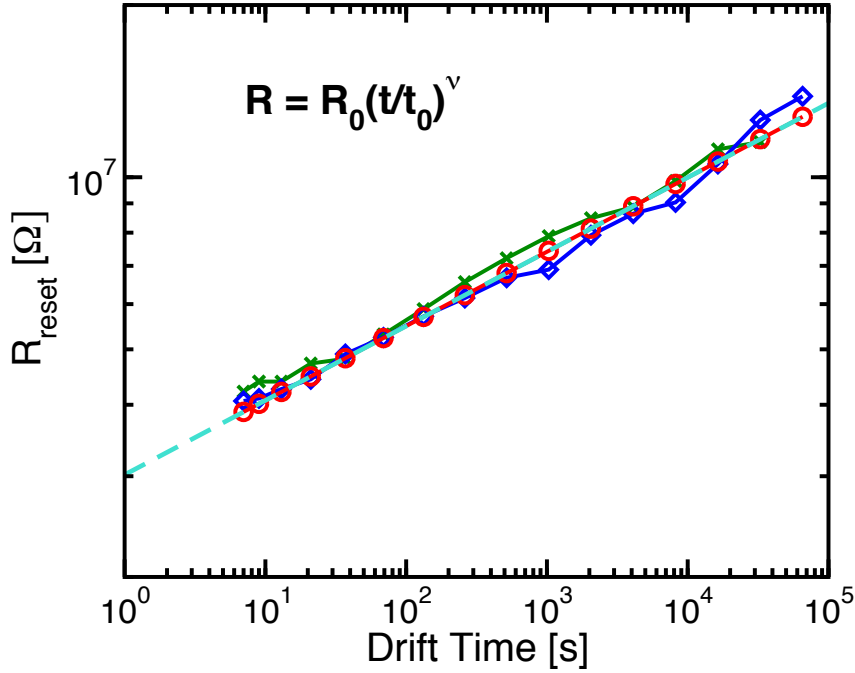


Figura 3.10: Resistenza di basso campo in funzione del tempo dopo programmazione (*Drift Time*) su un intervallo di circa 4 decenni, compreso tra circa 10s e un giorno. Le 3 serie corrispondono a tre stati di full reset, che differiscono leggermente nei valori di resistenza per motivi legati alla statistica dell'operazione di programmazione. Il  $\nu$  estratto (linea tratteggiata) è pari a 0.11.

avendo definito:

$$k' = \frac{k_B}{1 - T/T_{MN}} \quad (3.25)$$

In Figura 3.11 è riportato l'andamento delle curve IV al variare del tempo dopo programmazione.

L'intercetta della regione esponenziale con l'asse delle correnti è utilizzata per ottenere il valore di  $I_{00}(t)$ , ricavata dall'Eq.(3.24) come:

$$I_{00}(t) = I|_{V=0}(t) e^{\frac{E_A|_{V=0}(t)}{k'T}} \quad (3.26)$$

I valori dell'energia di attivazione al variare del *drift time* sono ottenuti dall'Eq.(3.21), facendo uso del valore di  $\gamma$  ricavato in precedenza (Par.3.4.1. I risultati ottenuti sono riportati in Figura 3.12.

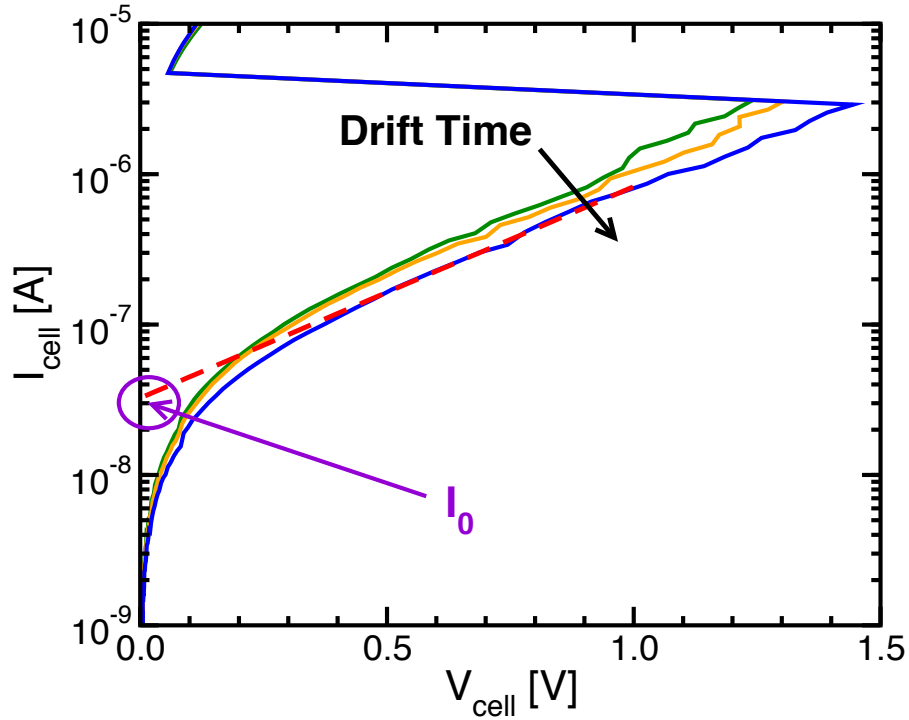


Figura 3.11: Curve IV nella regione di sotto-soglia su una cella PCM in tecnologia  $90nm$ , posta in stato di *full reset*. Ogni curva è stata misurata ad un diverso tempo dopo programmazione, in un range di tempi compreso tra  $10s$  e  $10^4s$  circa. Sono visibili l'aumento di resistenza e di tensione di soglia all'aumentare del *drift time*. Come mostrato in figura, da ogni curva è possibile estrarre il valore del fattore pre-esponenziale  $I_0$ .

Si noti come nel caso Poole il valore di  $I_{00}$  resti più costante rispetto al caso Poole-Frenkel, il che può suggerire di essere più vicini all'ipotesi di densità di trappole elevate. Questo argomento verrà ripreso in seguito. Nei due casi si può estrarre un valore medio  $\overline{I_{00}^P} = 21\mu A$  e  $\overline{I_{00}^{PF}} = 3\mu A$ .

### 3.4.3 Simulazione del *drift* della tensione di soglia

Partendo dall'equazione della corrente di sotto-soglia in condizione di alto campo (Eq.(3.24)) e dal criterio di potenza di soglia costante (Eq.(3.19)) è possibile ottenere un'espressione, che permette di calcolare la tensione di soglia

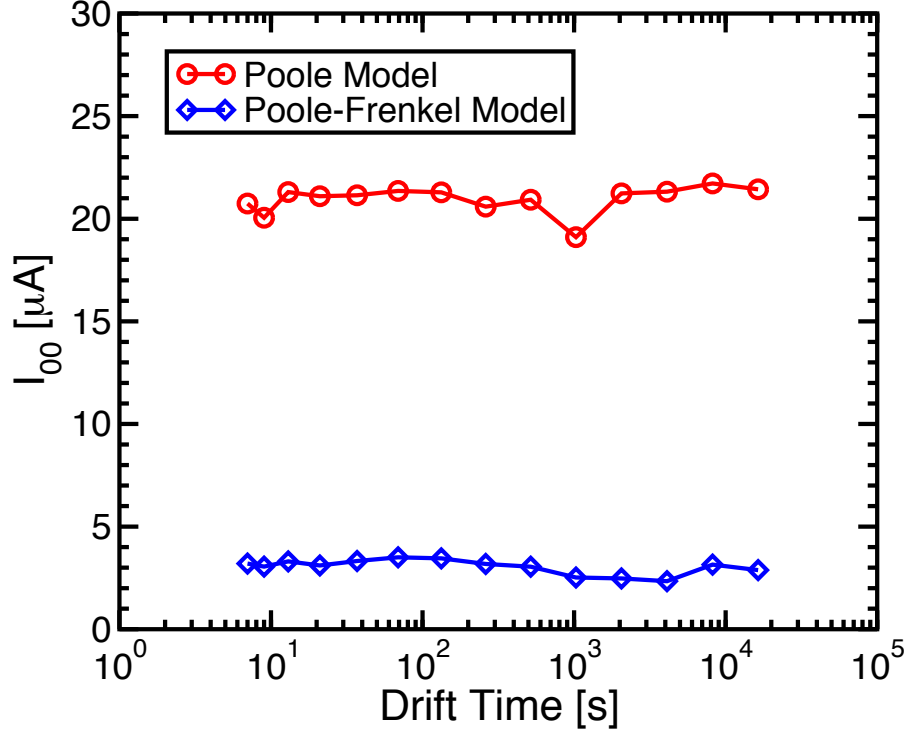


Figura 3.12: Valori di  $I_{00}$  estratti dalle curve IV (Figura 3.11) al variare del tempo dopo programmazione, nell'ipotesi di Poole e di Poole-Frenkel *barrier lowering*.

in funzione del tempo dopo programmazione.

$$P_T = V_T I_T = V_T I_{00} e^{-\frac{E_A | V_T, t}{k' T}} \quad (3.27)$$

Nel caso di meccanismo di abbassamento di barriera Poole, usando l'Eq.(3.5), ed esplicitando la dipendenza esponenziale di  $P_T$  da  $V_T$  si ottiene:

$$\begin{aligned} V_T &= \frac{2u_a}{\Delta z} \left[ \frac{E_A |_{V=0, t}}{q} + \frac{k' T}{q} \log \left( \frac{P_T}{V_T I_{00}} \right) \right] \\ &= \frac{2u_a}{\Delta z} \left[ \frac{E_A |_{V=0, t=0}}{q} + \frac{k' T}{q} \log \left( \frac{P_T}{V_T I_{00}} \right) + \frac{\gamma E_A |_{V=0, t=0}}{q} \log \left( \frac{t}{t_0} \right) \right] \end{aligned} \quad (3.28)$$

Nel caso di meccanismo di abbassamento di barriera Poole-Frenkel combinando l'Eq.(3.27) e l'Eq.(3.11), ed esplicitando la dipendenza esponenziale di  $P_T$  da  $V_T$  si ha:

$$\sqrt{V_T} = \sqrt{\frac{\pi \epsilon u_a}{q}} \left[ \frac{E_A |_{V=0, t}}{q} + \frac{k' T}{q} \log \left( \frac{P_T}{V_T I_{00}} \right) \right]$$

$$= \sqrt{\frac{\pi \varepsilon u_a}{q}} \left[ \frac{E_A|_{V=0, t=0}}{q} + \frac{k'T}{q} \log \left( \frac{P_T}{V_T I_{00}} \right) + \frac{\gamma E_A|_{V=0, t=0}}{q} \log \left( \frac{t}{t_0} \right) \right] \quad (3.29)$$

Non essendo possibile una soluzione analitica di Eq.(3.28) e di Eq.(3.29), si prospettano due possibilità per la risoluzione di Eq.(3.28) ed i Eq.(3.29):

- i) si ignora la dipendenza logaritmica da  $V_T$  e si sostituisce il termine incognito con il valore medio sperimentale della tensione di soglia  $\overline{V_{T,meas}} = 1.09V$  (*metodo analitico*);
- ii) partendo da un valore di  $V_T$  (parametro di *guess*  $V_T^0$ , posto pari a  $\overline{V_{T,meas}}$ ), le Eq.(3.28) e (3.29) possono essere riscritte rispettivamente in forma simbolica come  $V_T^{n+1} = f(V_T^n)$  e  $\sqrt{V_T^{n+1}} = g(V_T^n)$ . Il metodo procede per approssimazioni successive fino a convergenza (*metodo iterativo*).

Il confronto tra il risultato dei modelli a potenza di soglia costante e i risultati sperimentali è riportata in Figura 3.13(a) e Figura 3.13(b). I grafici mostrano i dati sperimentali e le previsioni del modello di *switching* a soglia a potenza di soglia costante, con  $V_T$  calcolata sia con il metodo analitico sia con quello iterativo. Si osservi come il modello Poole riesca a tenere conto meglio del modello Poole-Frenkel dei dati sperimentali. La capacità di prevedere il *drift* di soglia del modello Poole si estende su un range di tempi molto ampio di 10 decenni, compreso tra 100ns e 10000s.

Per meglio comprendere le ragioni fisiche della preferibilità di un modello rispetto all'altro è utile l'osservazione della Figura 3.14, che mostra la medesima curva IV con il logaritmo della corrente in funzione della tensione (a) e della radice quadrata della tensione (b).

Il logaritmo della corrente resta proporzionale a  $V$  per un intervallo di tensioni più esteso di non quanto succeda per  $\sqrt{V}$ , e la cosa più importante è che la deviazione del modello Poole-Frenkel dall'andamento esponenziale è marcato proprio vicino alla tensione di soglia, laddove si desidera che il modello possa prevedere correttamente la corrente.



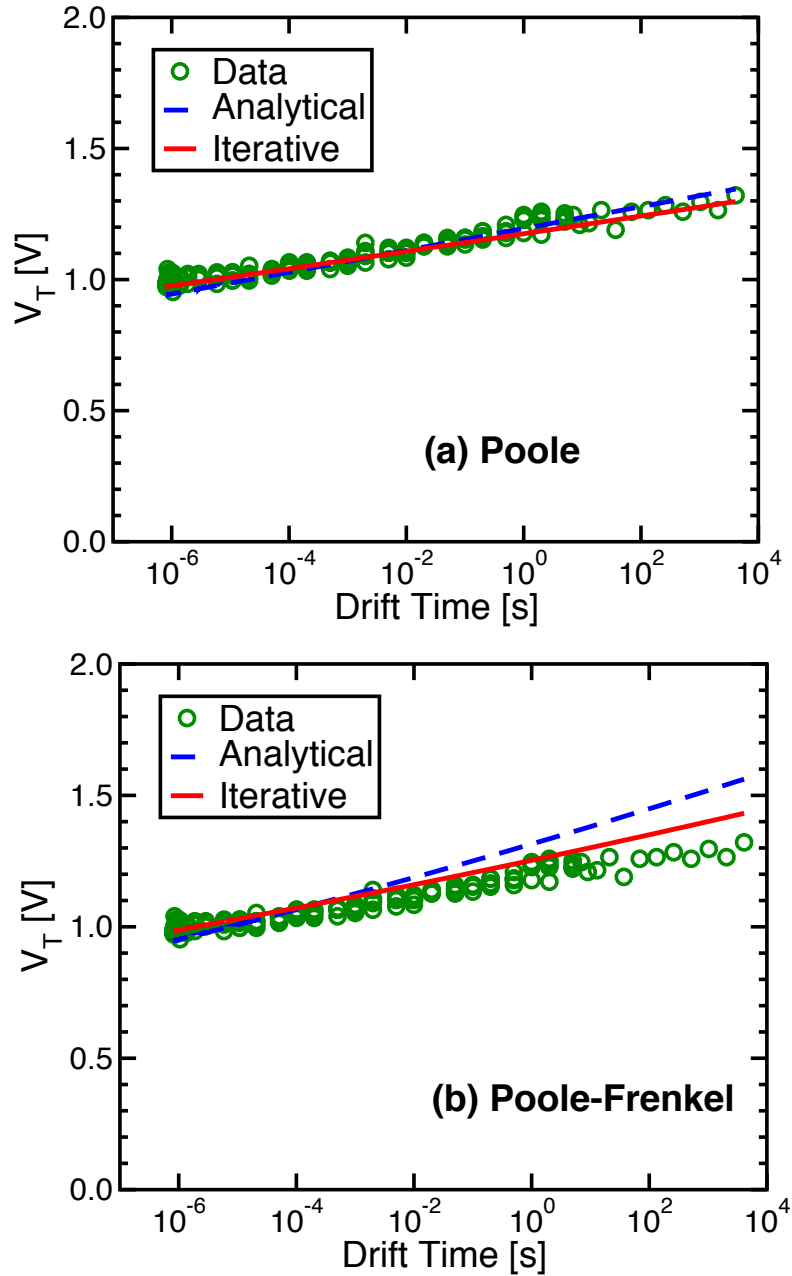


Figura 3.13: Tensione di soglia in funzione del tempo dopo programmazione (*drift time*). Sono riportati i dati sperimentali (cerchi), i risultati del modello Poole (a) e Poole-Frenkel (b) con il metodo analitico (linea tratteggiata) e con quello iterativo (linea continua). Si osservi come il modello Poole riesca a prevedere l'andamento di  $V_T$  sull'ampio range di tempi osservati, a differenza del modello Poole-Frenkel.

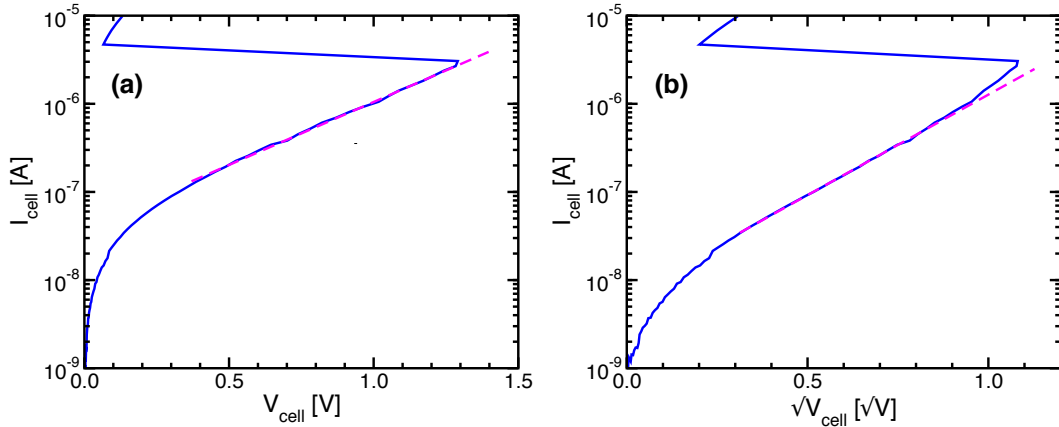


Figura 3.14: I due grafici rappresentano la stessa curva IV. Mentre in (a) la linea di fitting della regione esponenziale si allinea bene ai punti sperimentali anche vicino al punto di *switching* a soglia, in (b) il fitting sottostima la corrente proprio nella regione di interesse nell'intorno di  $V_T$ . Ciò fa comprendere il motivo della maggiore accuratezza del modello Poole rispetto al modello Poole-Frenkel.

Il lavoro esposto mostra come lo *switching* a soglia richieda una certa potenza di soglia  $P_T$  affinché possa avvenire. Il valore di  $P_T$  è risultato indipendente dallo stato di rilassamento strutturale. L'analisi del *drift* della tensione di soglia ha portato a concludere che nella tecnologia analizzata la distanza media tra le trappole nella direzione della conduzione  $\Delta z$  giustifichi un approccio di tipo Poole alla conduzione sotto-soglia. Chiaramente a seconda della cella in esame (tecnologia, struttura, dimensione, composizione del materiale attivo) bisogna riconsiderare l'Eq.(3.2) e verificare che sia più sensato un approccio di tipo Poole o di tipo Poole-Frenkel alla conduzione, o addirittura un caso intermedio tra i due. Una validazione ulteriore di questo approccio al fenomeno dello *switching* a soglia coinvolgerà perciò la caratterizzazione elettrica di diverse tecnologie al variare della dimensione del dispositivo, della temperatura, della struttura di cella, dello stato programmato.

# Capitolo 4

## Modello di cristallizzazione indotta da campo elettrico

*In questo capitolo viene discusso uno dei fenomeni fisici su cui si basa il funzionamento dei materiali calcogenuri come dispositivi di memoria, quello della cristallizzazione da fase amorfa. Nella Sezione 4.1 viene introdotta la teoria classica della transizione amorfo-cristallino, distinguendo tra i fenomeni di nucleazione (Paragrafo 4.1.1) e crescita (Paragrafo 4.1.2). Vengono poi riportati nella Sezione 4.2 i risultati fondamentali della cristallizzazione “lenta” del GST ottenuta tramite annealing termico nei tipici esperimenti di ritenzione. Nella Sezione 4.3 viene presentato il lavoro originale svolto sulla cristallizzazione “veloce” in GST indotta dall’applicazione di un impulso elettrico. Lo studio e la creazione di un modello analitico, che possa spiegare in maniera coerente come l’impulso elettrico generi per Joule heating un aumento di temperatura nel materiale attivo, è l’obiettivo principale di questa lavoro. Si conclude con il confronto tra l’Arrhenius plot dei tempi di fallimento ottenuto con misure in bake, e quelli ottenuti con impulsi elettrici mediante i modelli elettro-termici di Joule heating sviluppati nel corso del capitolo.*

## 4.1 Meccanismi di cristallizzazione

I processi fondamentali che sono coinvolti nella cristallizzazione di un liquido o di un solido amorfo sono due:

i) inizialmente la cristallizzazione è avviata dal meccanismo di nucleazione.

Nel caso più semplice ciò avviene all'interno della fase amorfa. Quando questo avviene si parla di *nucleazione omogenea*(Par.4.1.1.1). Se il cristallo si forma a partire da una sostanza estranea che agisce come centro di nucleazione preferenziale si parla invece di *nucleazione eterogenea*(Par.4.1.1.2);

ii) successivamente, un cristallo stabile cresce fino a raggiungere dimensioni macroscopiche, meccanismo chiamato *crescita*(Sez.4.1.2).

### 4.1.1 Fenomeno di nucleazione

I concetti fondamentali della teoria della nucleazione furono dati da Gibbs nel diciannovesimo secolo. Questa prima trattazione è puramente termodinamica e descrive la formazione di un *cluster* cristallino all'interno di una matrice amorfa. Usando questo approccio, il primo modello sulla cinetica della nucleazione fu proposto da Volmer e Weber nel 1926, ed è servito da punto di partenza per gli approfondimenti successivi di Becker e Döring nel 1935. Infine, nel 1949 Turnbull e Fisher ottennero un'espressione per il fattore pre-esponenziale che compare nell'espressione del tasso di nucleazione. Al giorno d'oggi, la teoria è nota come la *teoria classica della nucleazione* di Volmer, Weber, Becker, Döring, Turnbull e Fisher.

#### 4.1.1.1 Nucleazione omogenea

Nella matrice amorfa, gli atomi si avvicinano gli uni agli altri statisticamente a formare un *cluster* cristallino, secondo le leggi della termodinamica.

Supponiamo per semplicità che queste formazioni cristalline siano delle sfere di raggio  $r$ . La distribuzione dei cluster cristallini all'equilibrio termodinamico è data dalla statistica di Boltzmann:

$$N^{equ}(r) = N_0 e^{-\frac{\Delta G_{cluster}(r)}{k_B T}} \quad (4.1)$$

dove  $\Delta G_{cluster}(r)$  è il lavoro reversibile per la formazione di un cluster cristallino di raggio  $r$  e  $N_0$  è il numero totale di atomi nell'amorfo (normalizzato per unità di volume).  $\Delta G_{cluster}(r)$  può essere espresso come somma di due contributi:

$$\Delta G_{cluster}(r) = -\Delta G_{lc,V} \cdot \frac{4}{3}\pi r^3 + \sigma \cdot 4\pi r^2 \quad (4.2)$$

dove  $\Delta G_{lc,V}(T)$  è la differenza tra l'energia libera di Gibbs per unità di volume del liquido amorfo e del cristallo, mentre  $\sigma > 0$  è l'energia libera di Gibbs di interfaccia. Un'espressione significativa di  $\Delta G_{lc,V}$  in funzione della differenza di temperatura  $\Delta T$  rispetto alla temperatura di fusione  $T_m$  ( $\Delta T = T_m - T$ ) è [36]:

$$\Delta G_{lc,V}(\Delta T) = \frac{\Delta H_f \Delta T}{T_m} \quad (4.3)$$

dove  $\Delta H_f$  è il calore latente di fusione. Si osservi allora come  $\Delta G_{lc,V}$  nulla alla temperatura di fusione  $T_m$ , e cresce più si scende sotto  $T_m$ . La Figura 4.1 raffigura qualitativamente l'evoluzione di  $\Delta G_{cluster}(r)$ . Le curve presentano un massimo, che può essere ottenuta semplicemente risolvendo  $\frac{\partial \Delta G_{cluster}}{\partial r}(r) = 0$ . Il massimo si ha per un raggio  $r_c$  pari a:

$$r_c = \frac{2\sigma}{\Delta G_{lc,V}} \quad (4.4)$$

e l'altezza della barriera energetica vale:

$$\Delta G_c \triangleq \Delta G_{cluster}(r_c) = \frac{16\pi}{3} \frac{\sigma^3}{(\Delta G_{lc,V})^2} \quad (4.5)$$

La quantità  $r_c$  è detta *raggio critico*, ed è dell'ordine dei nanometri, mentre  $\Delta G_c(T)$  è il lavoro critico per la formazione di un *cluster* cristallino.  $\Delta G_c(T)$

---

cresce per  $r < r_c$ , quindi sfere di raggio minore di  $r_c$  sono energeticamente non favorevoli e si dissolvono spontaneamente. Invece per  $r > r_c$  le formazioni cristalline crescono per aumentare la loro energia libera. In conclusione  $\Delta G_c$  può essere considerata come una energia di attivazione che si oppone alla cristallizzazione. Combinando l'Eq.(4.3), l'Eq.(4.4) e l'Eq.(4.5) è possibile ottenere

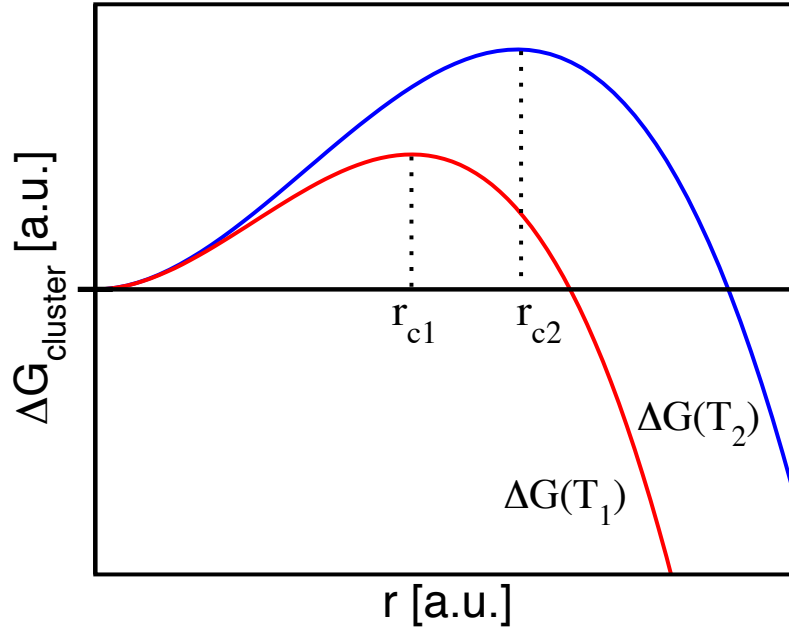


Figura 4.1: Andamento dell'energia libera di Gibbs per la formazione di un cristallo sferico di raggio  $r$  in matrice amorfa. Le due curve rappresentano l'effetto della temperatura ( $T_1 < T_2$ ) sulla dimensione (raggio critico  $r_c$ ) e sull'energia di nucleazione omogenea  $\Delta G_c$  di un cristallo, in accordo con le Eq.(4.6) e (4.7).

una espressioni del raggio raggio critico e dell'altezza di barriera in funzione del  $\Delta T$  di sovra-raffreddamento:

$$r_c = \frac{2\Delta G_{lc,v}T_m}{\Delta H_f \Delta T} \quad (4.6)$$

$$\Delta G_c^{om} = \frac{16\pi}{3} \frac{\sigma^3 T_m^2}{\Delta H_f^2 \Delta T^2} \quad (4.7)$$

L'esistenza di questa barriera fa sì che il sovra-raffreddamento di un liquido sotto  $T_m$  non provochi un'immediata e completa cristallizzazione. La nucleazione può quindi essere vista come la formazione di *cluster* cristallini di raggio post-critico, i.e.  $r > r_c$ .

Facendo uso della distribuzione delle sfere cristalline (Eq.(4.1)) si possono ottenere i parametri relativi alla cinetica di cristallizzazione nel caso di nucleazione omogenea. Poiché  $N^{equ}(r)$  diviene non fisica per  $r > r_c$ , nella prima teoria di Volmer e Weber venne ignorato, imponendo  $N^{equ}(r > r_c) = 0$ . In questa teoria si assume ragionevolmente che la nucleazione avvenga quando un *cluster* critico acquisti un atomo aggiuntivo. Con queste ipotesi il tasso di nucleazione è dato da:

$$I^{equ} = s_c k N_0 e^{-\frac{\Delta G_c}{k_B T}} \quad (4.8)$$

dove  $k$  è il tasso di arrivo di atomi dalla matrice amorfa verso il *cluster* critico (numero di atomi che arrivano alla superficie del cluster per unità di tempo, in  $1/s$ ), mentre  $s_c$  è la densità superficiale di atomi all'interfaccia cristallo-amorfo. Il valore del fattore pre-esponenziale  $k$  ha una diversa dipendenza dalla temperatura a seconda di quale sia il meccanismo limitante il movimento degli atomi nella fase amorfa. Si possono delineare due diversi meccanismi:

**i)** per cristallizzazione limitata da processi diffusivi, il cambio del numero di primi vicini e/o del numero di coordinazione sono necessari per la cristallizzazione. Questo comportamento si riscontra nelle leghe metalliche, nei materiali ionici, nei materiali covalenti ed è anche osservato nei materiali a cambiamento di fase. Dalla teoria di Turnbull e Fischer si ottiene:

$$k = \frac{6D}{\lambda^2} \quad (4.9)$$

dove  $D$  è associata alla diffusività nella fase amorfa. Il parametro  $\lambda$  è la distanza media interatomica.

ii) per cristallizzazione limitata da collisioni, i primi vicini in genere non devono cambiare posizione attraverso processi diffusivi per giungere a cristallizzazione. Quello che succede è che il movimento degli atomi dall'amorfo al cristallo è assistito da vibrazioni termiche. Questo meccanismo è generalmente dominante nei metalli puri e nei solidi di Van der Waals ma non nei materiali a cambiamento di fase, come detto in precedenza. Il valore di  $k$  è perciò pari a al tasso di collisione con il quale gli atomi tentano di unirsi al *cluster* cristallino:

$$k = \frac{u_s}{\lambda} \quad (4.10)$$

dove  $u_s$  è la velocità del suono nella fase amorfa, che è la velocità caratteristica delle vibrazioni degli atomi nell'amorfo.

Come osservato, nel caso di materiale calcogenuro siamo nel caso i). Sostituendo l'Eq.(4.9) in (4.8) otteniamo:

$$I^{equ} = s_c \frac{6D}{\lambda^2} N_0 e^{-\frac{\Delta G_c}{k_B T}} \quad (4.11)$$

Approssimando la diffusività  $D$  localmente con un andamento all'Arrhenius, i.e.  $D \propto \exp(-E_D/(k_B T))$ , dove  $E_D = -\partial \log D / \partial [1/k_B T]$  è l'energia di attivazione locale per la diffusione, si ha un'energia di attivazione complessiva per nucleazione omogenea pari a:

$$E_{I^{equ}} = E_D + \Delta G_c \quad (4.12)$$

Si osservi che si può esprimere l'Eq.(4.11) in funzione della viscosità del fluido amorfo  $\eta$ , che è più semplice da ottenere sperimentalmente rispetto alla diffusività  $D$ . Usando l'equazione di Einstein-Stokes:

$$\eta D = \frac{k_B T}{3\pi\lambda} \quad (4.13)$$

che mette in relazione  $\eta$  e  $D$  si ottiene:

$$I^{equ} = s_c \frac{2k_B T}{\eta\pi\lambda^3} N_0 e^{-\frac{\Delta G_c}{k_B T}} \quad (4.14)$$



È evidente dall'Eq.(4.13) che l'energia di attivazione locale per la diffusività  $E_D$  e per la viscosità  $E_\eta$  sono identiche ( $E_D = E_\eta$ ).

Il tasso di nucleazione  $I^{equ}$  diventa trascurabile quando prossimi alla temperatura di fusione  $T_m$ , dove  $\Delta G_c = \infty$  (Eq.(4.7)), oppure vicini alla temperatura di transizione vetrosa  $T_g$ , dove la viscosità cresce drammaticamente con la diminuzione della temperatura. Di conseguenza,  $I^{equ}$  mostra un massimo pronunciato ad una temperatura intermedia tra  $T_g$  e  $T_m$ .

#### 4.1.1.2 Nucleazione eterogenea

Nella pratica, è difficile identificare una cristallizzazione dominata puramente dalla nucleazione omogenea.

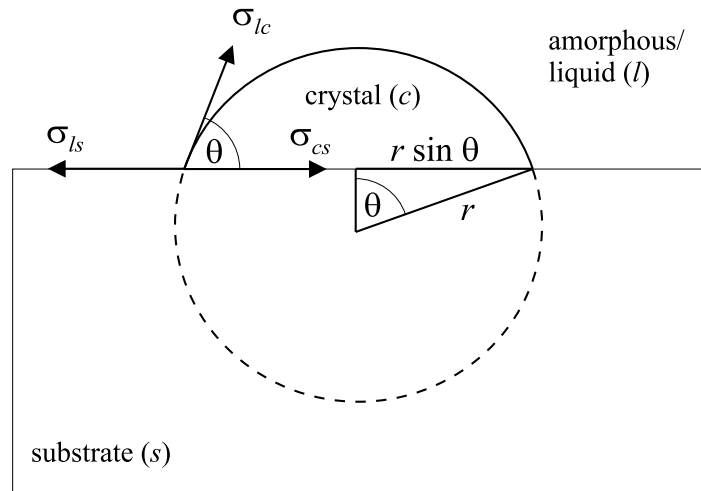


Figura 4.2: Rappresentazione schematica della nucleazione di una calotta sferica di raggio  $r$  a partire da un centro di nucleazione estraneo (etero-nucleazione). Il volume della sfera è pari a  $f(\theta) \cdot 4/3\pi r^3$ . Figura tratta da [18].

Di solito, delle fasi estranee come la pareti del contenitore in cui è posto il materiale (o gli ossidi di isolamento che circondano il materiale calcogenuro di una cella PCM) agiscono come centri di nucleazione preferenziali. Questi

processi, detti di nucleazione eterogenea, sono di natura estrinseca e possono essere influenzati dalle condizioni sperimentali.

Un semplice modello per la nucleazione eterogenea è ottenuto partendo dal modello di Gibbs (Eq.(4.2)), ma ragionando su una geometria che tenga conto dell'effetto di bordo generato dalla presenza di una fase estranea (Figura 4.2). In funzione dell'angolo di contatto  $\theta$  (*wetting angle*) il volume esposto è una frazione  $0 < f(\theta) < 1$  della sfera intera pari a:

$$f(\theta) = \frac{(2 + \cos \theta)(1 - \cos \theta)^2}{4} \quad (4.15)$$

Ripetendo i calcoli di Gibbs si può dimostrare che la nucleazione eterogenea è favorita rispetto alla nucleazione omogenea se è rispettata la condizione:

$$\sigma_{cs} - \sigma_{ls} = \sigma_{lc} \quad (4.16)$$

dove  $\sigma_{cs}$ ,  $\sigma_{ls}$ ,  $\sigma_{lc}$  sono rispettivamente l'energia libera per unità di superficie all'interfaccia cristallo-solido, all'interfaccia liquido-solido e all'interfaccia liquido-cristallo. In questa ipotesi si ha un valore del lavoro di formazione di un *cluster* eterogeneo pari a:

$$\Delta G_c^{et} = \frac{16\pi}{3} \frac{\sigma^3}{\Delta G_{lc,V}} f(\theta) \quad (4.17)$$

Si noti che se l'angolo di contatto è pari a  $180^\circ$  la Eq.(4.17) si riduce al caso omogeneo(Eq.(4.5)), come atteso.

La teoria classica della nucleazione omogenea (Eq.(4.11)) può essere applicata al caso di nucleazione eterogenea. Le uniche differenze sono il lavoro critico per la formazione di un *cluster* (Eq.(4.17)) e la riduzione degli atomi immersi nella fase amorfa che possono agire come centri di nucleazione. Difatti, mentre qualsiasi atomo può agire come centro di nucleazione omogenea, solo quelli in contatto con la fase estranea (impurità, substrato, etc.) possono agire come centro di nucleazione eterogenea. Detta  $\varepsilon$  la frazione di atomi immersi

nella fase amorfa in contatto con la fase estranea (normalmente  $\varepsilon \ll 1$ ), allora il tasso di nucleazione eterogenea,  $I^{equ,et}$  è dato da:

$$I^{equ,et} = \varepsilon s_c \frac{6D}{\lambda^2} N_0 e^{-\frac{\Delta G_c^{et}}{k_B T}} \quad (4.18)$$

con un'energia di attivazione locale:

$$E_{J^{et}} = E_D + \Delta G_c^{et} \quad (4.19)$$

Visto che  $\Delta G_c^{et} < \Delta G_c^{om}$  quando  $f(\theta) < 1$  (Eq.(4.17)). Il rapporto tra il tasso di nucleazione eterogeneo e quello omogeneo ad una certa temperatura è ottenuto come:

$$\frac{I^{equ,et}}{I^{equ,om}} = \varepsilon e^{\frac{\Delta G_c^{om}}{k_B T} [1-f(\theta)]} \quad (4.20)$$

La Figura 4.3 mostra un grafico dell'Eq.4.20, con un valore tipico del parametro  $\varepsilon = 10^{-6}$ , per tre valori diversi del lavoro critico per nucleazione omogenea  $\Delta G_c^{om}$ .

Dal grafico si evince come le impurezze possano provocare un aumento di ordini di grandezza del tasso di nucleazione. Generalmente, si osserva che il tasso di nucleazione eterogenea è molto più grande di quello omogeneo. In ogni caso, il tasso di nucleazione eterogeneo si somma a quello omogeneo,  $I^{equ,tot} = I^{equ,om} + I^{equ,et}$ .

### 4.1.2 Fenomeno di crescita

La teoria della nucleazione descrive la formazione di cluster cristallini termodinamicamente stabili. Queste formazioni possono crescere fino a raggiungere dimensioni macroscopiche. Sono due i possibili meccanismi che governano il fenomeno di crescita cristallina: la crescita controllata dall'interfaccia cristallo-amorfo, e quella da diffusione a lungo raggio.

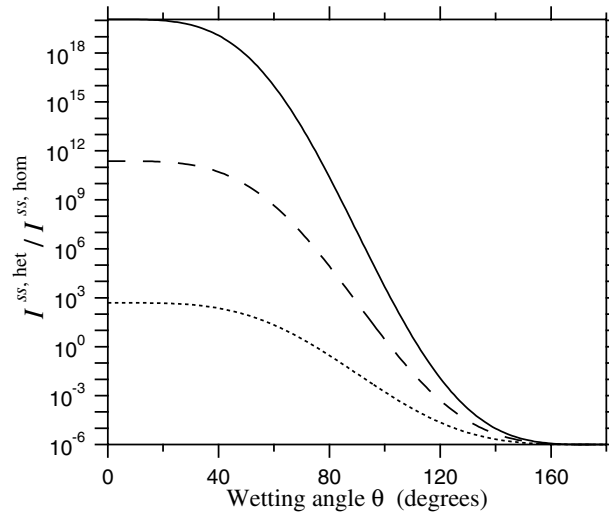


Figura 4.3: Rapporto tra il tasso eterogeneo di nucleazione e quello omogeneo a temperatura costante, calcolato dall'Eq.(4.20) e dall'Eq.(4.15) per  $\varepsilon = 10^{-6}$ . Linea continua:  $\Delta G_c^{om} = 60k_B T$ . Linea tratteggiata:  $\Delta G_c^{om} = 40k_B T$ . Linea a puntini:  $\Delta G_c^{om} = 20k_B T$ . Figura tratta da [18].

#### 4.1.2.1 Crescita controllata dall'interfaccia cristallo-amorfo

Se la cristallizzazione non provoca cambiamento della composizione chimica, la crescita è controllata dai processi di riarrangiamento strutturale all'interfaccia cristallo-amorfo. Questo riarrangiamento può coinvolgere sia un salto diffusivo dall'amorfo verso il cristallo nel caso di cinetica limitata da diffusione, sia una collisione tra atomi nel caso di cinetica limitata da collisioni. La frequenza degli urti tra atomi in fase amorfa e atomi all'interfaccia non dipende da dove si trova l'interfaccia, e la velocità di crescita  $u$  è perciò tempo-indipendente, e pari a:

$$u = \gamma_s \frac{2k_B T}{\eta \pi \lambda^2} \left[ 1 - e^{-\frac{\Delta G_{lc,atom}(T)}{k_B T}} \right] \quad (4.21)$$

dove  $0 \leq \gamma \leq 1$  è la frazione di siti dove un nuovo atomo può essere incorporato,  $\lambda$  è la distanza media interatomica (e quindi di quanto procede l'interfaccia ad ogni atomo incorporato),  $\eta$  la viscosità dell'amorfo, e  $\Delta G_{lc,atom} > 0$  è la

differenza di energia libera di Gibbs tra la fase amorfa e quella cristallina per ogni atomo incorporato. Il termine tra parentesi quadre di Eq.(4.21) risulta dal trasferimento atomico termicamente attivato all'interfaccia cristallo-amorfo. Alla temperatura di fusione  $T_m$  la velocità di crescita è nulla visto che  $\Delta G_{lc,atom}(T_m) = 0$ . In prossimità della temperatura di transizione vetrosa  $T_g$  il termine tra parentesi quadre varia lentamente rispetto alla viscosità  $\eta$  che aumenta di diversi ordini di grandezza, portando  $u$  a zero.

#### 4.1.2.2 Crescita controllata da diffusione a lungo raggio

Se c'è un cambio di composizione chimica in seguito alla cristallizzazione, la velocità di crescita è controllata da diffusione a lungo raggio, visto che la fase amorfa è priva degli elementi utili alla crescita cristallina in prossimità dell'interfaccia cristallo-amorfo. Questa assenza di elementi diviene sempre più pronunciata al trascorrere del tempo, e per questo la velocità di crescita  $u$  diminuisce col tempo. Da una semplice analisi dimensionale si ottiene:

$$u \propto \left(\frac{D}{t}\right)^{\frac{1}{2}} \quad (4.22)$$

Visto che una crescita cristallina dipendente dal tempo non è mai stata osservata per i materiali calcogenuri usati nei dispositivi PCM, non si discuterà ulteriormente questo meccanismo di crescita.

#### 4.1.3 Dipendenza dalla temperatura

Come visto nelle sezioni precedenti, lo sviluppo della fase cristallina di un amorfo super-raffreddato (i.e.  $T < T_m$ ) procede su una scala temporale lunga appena sotto la temperatura di fusione  $T_m$  e appena sopra la temperatura di transizione vetrosa  $T_g$ . La velocità di cristallizzazione più elevata è perciò osservata ad una temperatura intermedia  $\tilde{T}$  tra  $T_g$  e  $T_m$ . Quello che si osserva che per temperature prossime a  $\tilde{T}$  la cristallizzazione avviene in tempi dell'ordine

di 100ns. Un tempo così breve rende impossibile uno studio sperimentale del tasso di nucleazione e di crescita cristallina nell'intorno di  $\tilde{T}$ . Per questo motivo, solitamente la cristallizzazione è studiata appena sotto  $T_m$  o appena sopra  $T_g$ , dove la cristallizzazione è abbastanza lenta per poter essere osservata in laboratorio. La Figura 4.4 mostra il tasso di nucleazione omogeneo ottenuto sperimentalmente su diverse leghe calcogenure, in funzione della differenza di temperatura di temperatura  $\Delta T_r$  dalla temperatura del *liquidus*.

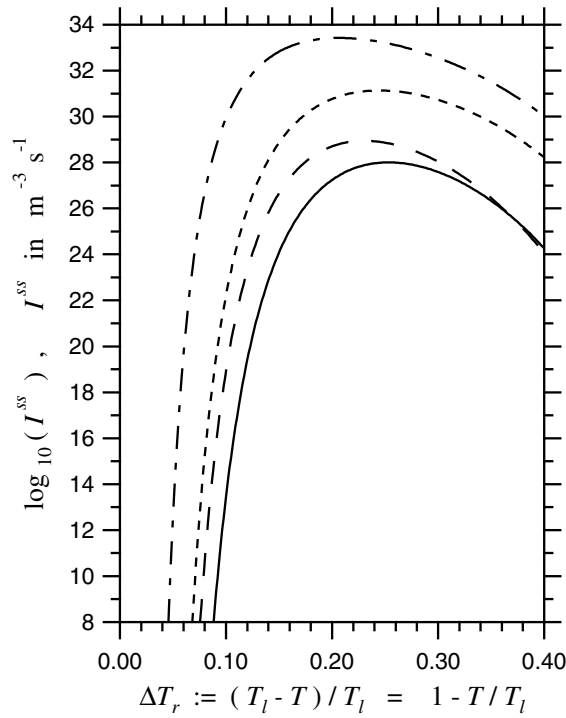


Figura 4.4: Limite superiore alla nucleazione omogenea  $I^{equ,om}$  ottenuta sperimentalmente minimizzando la nucleazione eterogenea, tramite isolamento del materiale a cambiamento di fase con un flusso liquido di  $B_2O_3$ .  $\Delta T_r \triangleq (T_l - T)/T_l$ . Linea tratto-punto:  $Ge_2Sb_2Te_5$ . Linea a tratti brevi:  $Ge_4Sb_1Te_5$ . Linea a tratti lunghi:  $AgInSb_2Te$ . Linea continua:  $Ge_{12}Sb_{88}$ .  $T_l$  è la temperatura del *liquidus* amorfo-cristallino. Figura tratta da [18].

Il tasso di nucleazione è più rapido per la lega  $GeSbTe$ , che mostra una cristallizzazione dominata dalla nucleazione, rispetto alle lege ricche di  $Sb$  come

$AgInSb_2Te$  e  $Ge_{12}Sb_{88}$ , che mostrano una cristallizzazione dominata dalla crescita. I tassi di crescita mostrati in Figura 4.4, frutto di esperimenti quasi-stazionari (tempi di cristallizzazione lunghi), sono troppo elevati per consentire l'amorfezzazione di dispositivi PCM sia elettrici che ottici. La nucleazione interferisce difatti anche con la velocità di raffreddamento più alti ottenibile, che sono dell'ordine di  $10^{10}Ks^{-1}$ . Ad esempio, se consideriamo un dispositivo per la memorizzazione ottica delle informazioni, il volume di un bit è dell'ordine di  $V_b = 1\mu m \times 1\mu m \times 10nm = 10^{-20}m^3$ . Con un tasso di raffreddamento di  $10^{10}Ks^{-1}$ , la temperatura decresce di  $100K$  in  $t_c = 100ns$ . Di conseguenza, la nucleazione interferisce con l'amorfezzazione se il tasso di nucleazione  $I^{equ,tot} > V_b^{-1}t_c^{-1} = 10^{28}m^{-3}s^{-1}$  su un intervallo di  $100K$ . Osservando la Figura 4.4 è evidente che ciò avviene sicuramente per le leghe  $GeSbTe$ , ma anche per quelle ricche di  $Sb$ , visto che hanno un tasso di nucleazione eterogenea elevata. Nel caso di dispositivi PCM elettronici, il volume da programmare ha un volume dell'ordine di  $V_b = 10^{-23}m^3$ . Quindi, la nucleazione interferisce se  $I^{equ,tot} > 10^{31}m^{-3}s^{-1}$ , che è appunto il caso del GST ( $Ge_2Sb_2Te_5$ ), la lega calcogenura utilizzata nel presente lavoro. In conclusione, si può affermare che:

- i) l'amorfezzazione (*quenching*) nelle condizioni operative dei dispositivi PCM avviene in un lasso di tempo in cui la nucleazione non è ancora cominciata (situazione non quasi-stazionaria, decine di nanosecondi), quando la popolazione di equilibrio  $N^{equ}$  di cluster cristallini non è ancora formata. Durante questo tempo, detto *tempo di incubazione*, il tasso di nucleazione resta a livelli più bassi della nucleazione di equilibrio per diversi ordini di grandezza. Visto che il tempo di incubazione è indipendente dal volume  $V_b$  del dispositivo utilizzato, il discorso fatto vale sia per i dispositivi elettronici sia per quelli ottici.
- ii) l'amorfezzazione non sarebbe possibile se il tempo di incubazione per la nucleazione fosse assente. Quindi, è proprio la presenza di questo tempo

morto per la nucleazione che rende possibile il funzionamento dei dispositivi di memorizzazione basati su materiali a cambiamento di fase.

## 4.2 Cristallizzazione indotta da *annealing*

Nella Sezione 4.1 è stato mostrato come la fase amorfa sia metastabile per  $T < T_m$ , e come il tasso di cristallizzazione sia una funzione fortemente dipendente dalla temperatura. Il processo naturale di cristallizzazione si scontra con l'utilizzo del materiale a cambiamento di fase come dispositivo di memoria non volatile. Per poter competere con le memorie *Flash*, il tempo tipico di cristallizzazione deve essere abbastanza lungo per garantire la ritenzione del dato per dieci anni alla massima temperatura di esercizio posta tra 85°C e 150°C, a seconda dell'applicazione.

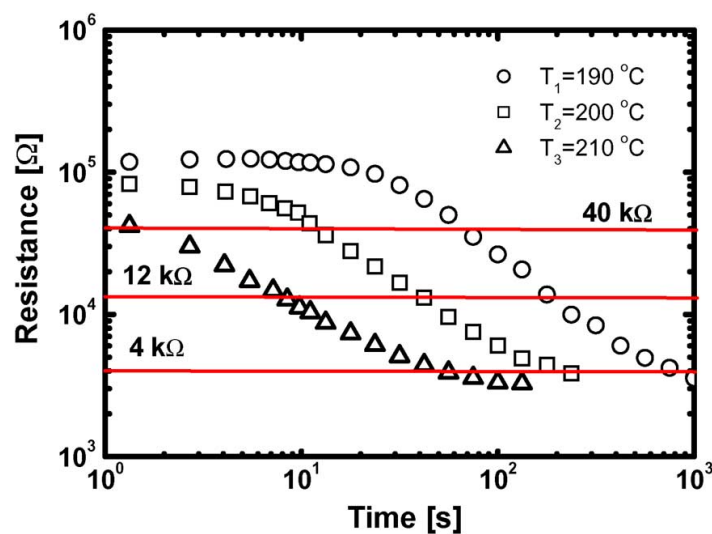


Figura 4.5: Resistenza media in funzione del tempo ottenuta su un dispositivo a cambiamento di fase in GST, a diverse temperature di annealing. Sono stati riportati diversi criteri per il calcolo di  $t_{fail}$ , con valori di  $R_{crit}$  pari a 40kΩ, 12kΩ, 4kΩ. Grafico preso da [37].



Visto che i tempi di ritenzione a temperatura ambiente sono dell'ordine dei mesi o degli anni, la tipica procedura per estrarre il tempo di ritenzione a basse temperature è quella di svolgere degli esperimenti di cristallizzazione accelerata in temperatura. La cella PCM, posta in stato amorfo (i.e. *full reset*), viene portata ad una elevata temperatura (tra 150°C e 230°C), ed infine viene misurata l'evoluzione temporale della sua resistenza elettrica (Figura 4.6). Il tempo di cristallizzazione  $t_{fail}$  può essere definito come il tempo necessario affinché la resistenza della cella scenda sotto un certo valore critico  $R_{crit}$  scelto a priori. La scelta di  $R_{crit}$  è fondamentale: a diversi valori di  $R_{crit}$  corrispondono diversi valori della frazione cristallina  $f_X$ , e quindi diverse energie di attivazione per la cristallizzazione (Figura 4.6).

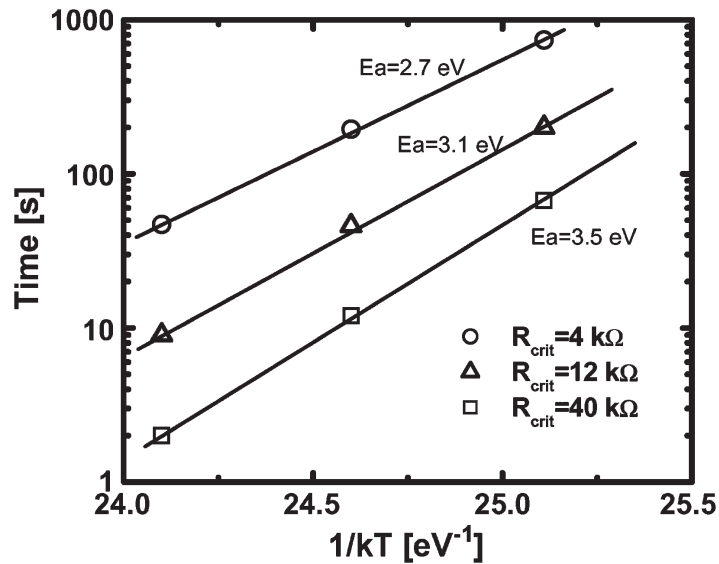


Figura 4.6: Arrhenius plot dei tempi di cristallizzazione per un dispositivo PCM basato su GST. Le tre serie sono ottenute per diversi valori della resistenza di soglia  $R_{crit}$ . È evidente come l'energia di attivazione che si estrae dal grafico dipenda dal valore scelto di  $R_{crit}$ , e quindi della frazione cristallina  $f_X$ . Grafico preso da [37].

Allo stesso modo, lo stesso valore di  $R_{crit}$  corrisponde ad una frazione  $f_X$  che cambia al variare della temperatura di *annealing*. Come dimostrato in [37], un

modo per aggirare questi problemi è quello di porre  $R_{crit}$  il più vicino possibile al punto di cristallizzazione, cui corrisponde  $f_X = 1$ , in modo tale che l'energia di attivazione per la conduzione sia praticamente indipendente dalla temperatura di *annealing*. In alternativa è possibile riportare le resistenze misurate a diversa temperatura alla medesima temperatura, sfruttando l'Eq.(3.14) e la conoscenza del legame tra l'energia di attivazione per la conduzione e  $R_0$ .

### 4.3 Cristallizzazione indotta da impulso elettrico nel GST

Come illustrato in 4.1.3, il processo di programmazione nei dispositivi PCM (sia ottici che elettronici) avviene portando il materiale calcogenuro ad una temperatura intermedia tra la temperatura di fusione  $T_m$  e quella di transizione vetrosa  $T_g$ , dove il tasso di nucleazione elevato consente la programmazione in tempi sufficientemente brevi per l'elettronica (decine di nanosecondi). Per raggiungere la giusta temperatura i dispositivi PCM studiati sono sottoposti ad un impulso elettrico che consenta la conversione della potenza elettrica in potenza termica per effetto Joule. Obiettivo di questo lavoro è comprendere quale sia il legame tra i parametri elettrici di tensione e corrente e la temperatura raggiunta nel materiale calcogenuro, e quali siano le differenze tra la cristallizzazione in *bake* e quella ottenuta mediante programmazione elettrica.

#### 4.3.1 Risultati di caratterizzazione

La raccolta dei dati sperimentali è proceduta fornendo degli impulsi di programmazione (tensione/corrente) di ampiezza e durata variabili, e misurando la resistenza elettrica raggiunta dalla cella dopo ogni impulso di programmazione.

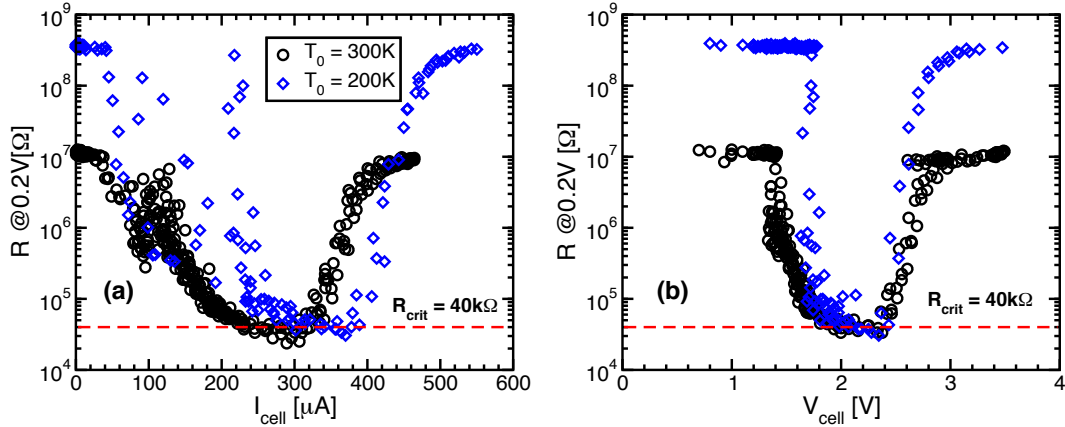


Figura 4.7: Curve di programmazione su cella PCM 90nm, struttura 1T1R, effettuate in criostato a due diverse temperature, in funzione dell corrente (a) e della tensione (b) sulla cella. Tempo di programmazione:  $300ns$ . Cerchi:  $T_0 = 300K$ . Rombi:  $T_0 = 200K$ . Si osservi come diminuendo la temperatura  $T_0$  aumenti la corrente/tensione necessaria per la completa cristallizzazione e per la fusione del GST. La linea tratteggiata rappresenta il valore di  $R_{crit}$  scelto per il calcolo di  $t_{fail}$ .

Per poter meglio comprendere come l'impulso elettrico induca un aumento di temperatura nel GST, alle misure a temperatura ambiente sono state affiancate misure in criostato. La Figura 4.7 riporta i risultati ottenuti in criostato a partire da due diverse temperature  $T_0$ ,  $300K$  e  $200K$ . Al fine di ovviare al problema della scelta di  $R_{crit}$  si è deciso di porre  $R_{crit}$  prossimo al valore di cristallizzazione, come suggerito nella Sezione 4.2.

### 4.3.2 Modello di *Joule heating* a resistenza termica costante

Per poter descrivere l'effetto di Joule heating indotto dal passaggio di elettroni attraverso il materiale calcogenuro, si consideri l'equazione di trasporto del calore di Fourier in una dimensione:

$$C_T \rho \frac{dT}{dt} = P''' - k_T \frac{d^2T}{dx^2} \quad (4.23)$$

dove con  $C_T$  si è indicato il calore specifico del GST, con  $k_T$  la sua conduttività termica, con  $\rho$  la sua densità e con  $P'''$  la generazione di potenza per unità di volume all'interno del materiale, in questo caso provocata dal passaggio di corrente elettrica. La quantità  $D_T = \frac{k_T}{C_T \rho}$  è detta diffusività termica (unità di misura  $m^2/s$ ). Da una semplice analisi dimensionale possiamo stimare una durata del transitorio termico  $\tau_T$  pari a:

$$\tau_T \sim \frac{L^2}{D_T} \quad (4.24)$$

dove con  $L$  si è indicata la lunghezza tipica del dispositivo, che è dell'ordine di  $10nm$ . Valori tipici delle grandezze termo-fisiche per il GST [38] sono:  $C_T = 250Jkg^{-1}K^{-1}$ ,  $k_T = 1Wm^{-1}K^{-1}$ ,  $\rho = 6400kgm^{-3}$ . Inserendo questi valori nell'Eq.(4.24) si ottiene  $\tau_T \sim 160ps$ , un valore più di due ordini di grandezza superiore ai tempi tipici di programmazione. Per questo motivo, possiamo ridurre l'Eq.(4.23) al caso stazionario ( $dT/dt = 0$ ). Schematizzando la cella come un tratto di materiale uniforme di lunghezza  $L$ , supponendo che la densità di potenza generata per unità di volume dall'impulso elettrico  $P'''$  sia uniforme, e imponendo le condizioni al contorno  $T(0) = T(L) = T_0$ , si ottiene un'andamento parabolico di  $T(x)$ , il cui massimo è posto per  $x = L/2$ , e pari a:

$$T_{max} = T_0 + \frac{L^2}{8k_T} P''' = T_0 + R_T P''' \quad (4.25)$$

dove con  $R_T$  si è indicata la resistenza termica del materiale calcogenuro. Poiché  $P'''$  è proporzionale al prodotto tra tensione e corrente sulla cella, da ora in poi verrà utilizzata la potenza elettrica  $P$  in luogo della potenza per unità di volume  $P'''$ ,  $R_T$  si adeguerà affinché l'Eq.(4.25) rimanga dimensionalmente corretta.

Per poter dedurre tramite questo semplice modello la temperatura raggiunta nel materiale a cambiamento di fase durante programmazione bisogna trovare un modo per estrarre il valore di resistenza termica  $R_T$ .

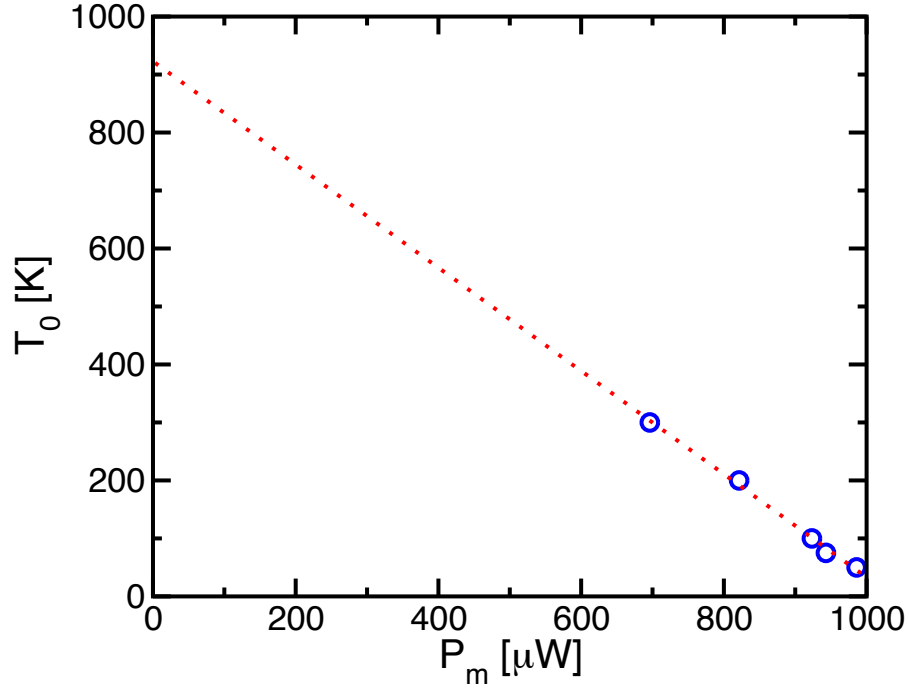


Figura 4.8: Il grafico mette in relazione la potenza elettrica  $P_m$  necessaria per ottenere la fusione del GST, in funzione della temperatura nel criostato  $T_0$ . Si osservi come la temperatura di fusione  $T_m$  del GST, rappresentata dall'intercetta con l'asse delle ordinate (Eq.(4.25)), sia prossima al valore noto in letteratura, intorno a  $900K$ .

A questo scopo basta applicare il modello di Eq.(4.25) al punto di fusione del GST, rappresentato sulle curve di programmazione di Figura 4.7 come il punto di risalita ( $I_m$  e  $V_m$ ) della resistenza di programmazione. In Figura 4.8 è riportato l'andamento della potenza di fusione,  $P_m = V_m I_m$ , in funzione della temperatura cui è stato posto il campione nel criostato. La temperatura di fusione prevista dall'Eq.(4.25) è l'intercetta del grafico  $P_m(T_0)$  con l'asse delle ascisse. Il valore estrapolato è vicino alla temperatura di fusione del GST misurabile sperimentalmente, posta intorno a  $900K$ , cosa che sembra confermare in prima istanza la bontà del modello, seppure nella sua estrema semplicità. Il valore di  $R_T$  estratto dalla regressione lineare dei punti sperimentali è  $R_T = 8.9 \cdot 10^5 K/W$ .

Dai dati sperimentali come quelli in Figura 4.7 ripetuti per diversi tempi di programmazione  $t_{prog}$  (i.e. durata dell'impulso elettrico) è possibile ottenere la Figura 4.9, dove il tempo di fallimento  $t_{fail}$  è evidentemente uguale a  $t_{prog}$ , visto che proprio in quel lasso di tempo la cella è passata da amorfa (*reset*) a cristallina (*set*).

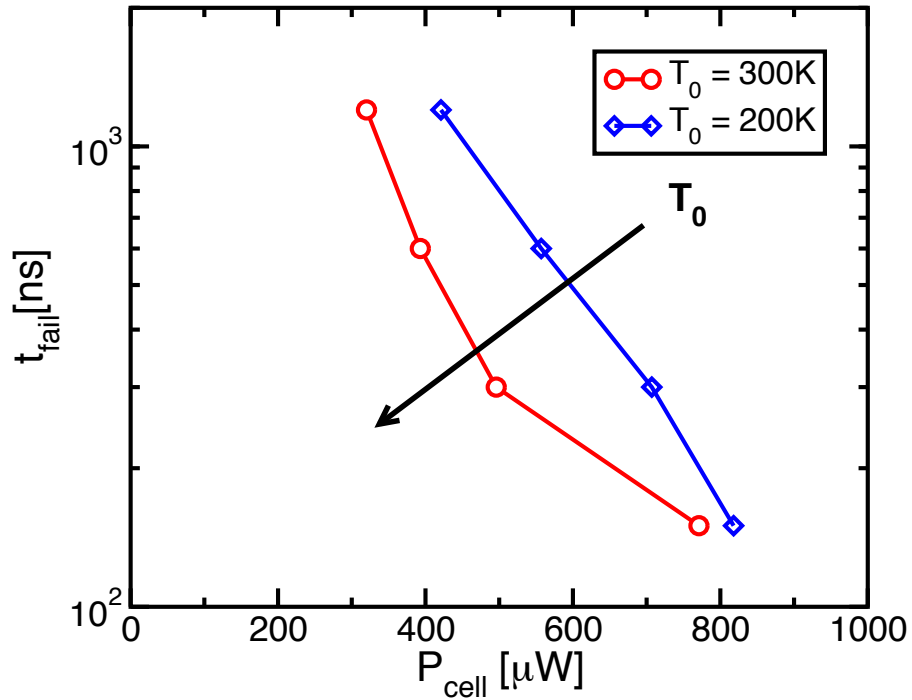


Figura 4.9: Le due serie rappresentano il tempo di fallimento  $t_{fail}$  in funzione della potenza elettrica sulla cella durante programmazione, con  $T_0 = 300K$  (cerchi) e  $200K$  (rombi). Il grafico è ottenuto a partire da curve di programmazione come quelle in Figura 4.7 per diversi tempi di programmazione  $t_{prog}$ . Si osservi che al diminuire della temperatura  $T_0$  debba crescere la potenza necessaria per la cristallizzazione, come predetto da Eq.(4.25).

Infine è possibile costruire un grafico Arrhenius dei tempi di cristallizzazione indotta da impulso elettrico (Figura 4.10), dove la temperatura non è quella realmente misurata nel campione (cosa impossibile da realizzarsi), ma quella dedotta attraverso il modello di *Joule heating* di Eq.(4.25).

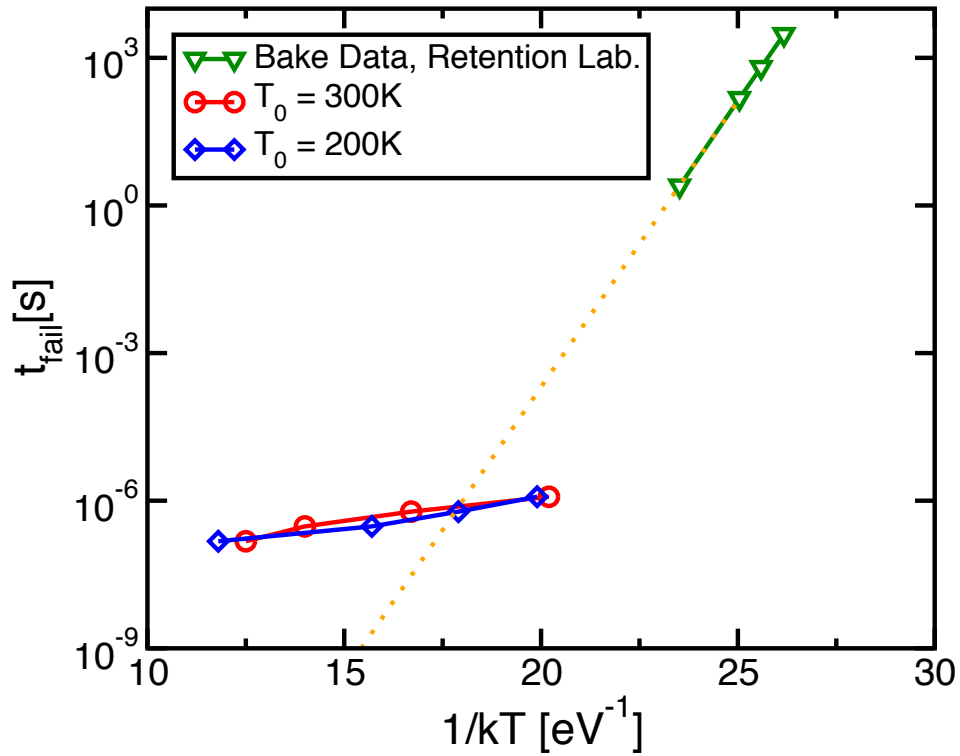


Figura 4.10: Arrhenius plot del tempo di fallimento  $t_{fail}$ . Cerchi:  $T_0 = 300K$ . Rombi:  $T_0 = 200K$ . Triangoli: dati ottenuti in Bake, Retention Lab. Si osservi come le temperature dedotte col modello di *Joule heating* a partire da  $300K$  e  $200K$  si sovrappongono bene, ma manca il raccordo con i dati ottenuti in *bake*, cosa che mina la validità del modello.

Pur riuscendo a stimare la stessa temperatura equivalente per  $T_0 = 300K$  e  $T_0 = 200K$  il modello presenta un'incongruenza di fondo: le serie non possono raccordarsi in alcun modo con i dati di ritenzione presi sulla stessa tecnologia in *bake*. Per questo motivo nel seguito viene discusso un raffinamento del modello di *Joule heating* fino ad ora proposto.

### 4.3.3 Modello di *Joule heating* a resistenza termica variabile

Per migliorare il modello è stata riconsiderato il valore di resistenza termica. Per motivi legati alla teoria della percolazione, un aumento della corrente è dovuto all'instaurarsi di più percorsi percolativi ad alta conducibilità elettrica (i.e. bassa energia di attivazione per la conduzione). Come è noto dalla scienza dei materiali, ad un aumento della conducibilità elettrica corrisponde un aumento della conducibilità termica, e quindi una diminuzione della resistenza termica  $R_T$ . Queste osservazioni hanno portato a considerare un nuovo modello di *Joule heating*, in cui la resistenza termica non è più costante, ma dipendente dalla corrente elettrica che attraversa il materiale calcogenuro durante la programmazione. La Figura 4.11 mostra schematicamente la dipendenza ipotizzata di  $R_T$  da  $I$ .

Per rendere la trattazione matematica più semplice e l'interpretazione fisica più immediata si è suddiviso il grafico in due casi limite, che si raccordano in  $I = I_{th}$ :

$$R_T(I) = \begin{cases} \frac{C}{I} & \text{per } I < I_{th}, \\ R_{T_0} & \text{per } I \geq I_{th}. \end{cases} \quad (4.26a)$$

**i)** per  $I < I_{th}$  ad un aumento della corrente corrisponde un aumento dei percorsi percolativi a bassa resistenza termica, o, in maniera equivalente, dell'area efficace  $A_{eff} \propto I$  ad essi associabile. Essendo  $R_T \propto A^{-1}$ , possiamo scrivere  $R_T = C/I$ ;

**ii)** per  $I \geq I_{th}$  tutto il materiale calcogenuro è passato in regime di alta conduzione elettro-termica, e  $R_T$  satura al valore di resistenza termica  $R_{T_0}$  associato all'intera area del dispositivo in stato di alta conduzione.

Affinchè questa descrizione abbia un senso fisico, il valore di  $I_{th}$  è atteso minore della corrente  $I_m$  necessaria alla fusione del GST (fondo della curva



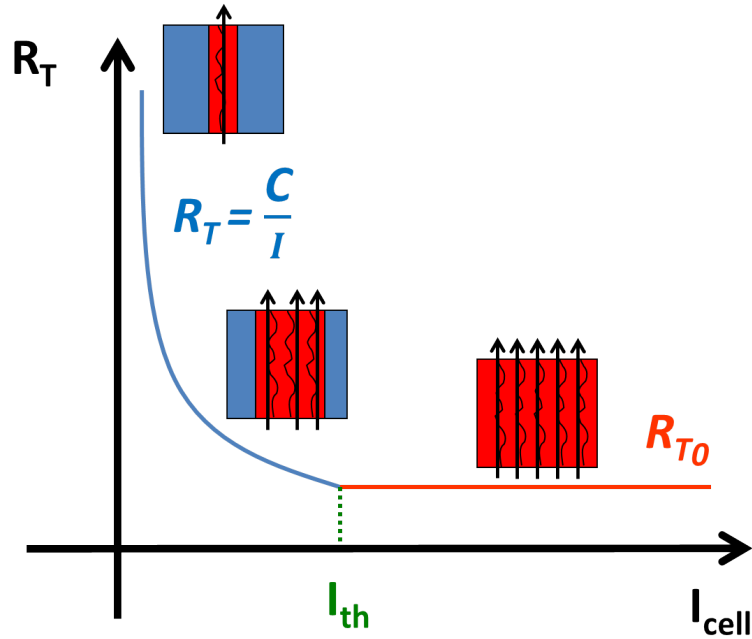


Figura 4.11: Il grafico mostra qualitativamente la dipendenza della resistenza termica  $R_T$  dalla corrente elettrica che attraversa la cella PCM durante l'impulso di programmazione. Per effetti percolativi nel GST amorfo, un aumento di corrente corrisponde ad un aumento dei percorsi a bassa resistenza elettrica e termica. In questo modello molto semplice si è supposta la presenza di una prima regione ( $I < I_{th}$ ) in cui  $R_T$  è inversamente proporzionale alla corrente  $I$ , e una seconda regione ( $I > I_{th}$ ) in cui  $R_T$  satura a un valore  $R_{T0}$ .

di programmazione), altrimenti si arriverebbe all'assurdo per cui il materiale cominci a fondere in prossimità dell'*heater* (che è il punto a maggiore densità di corrente), senza che tutto il materiale sia passato in regime di alta conduzione.

Il nodo successivo è quello di passare ad una forma dell'Eq.(4.25) in funzione della sola corrente elettrica  $I$ . A questo scopo si è espressa la tensione  $V$  in funzione della corrente come:

$$V = V_H + R_S I \quad (4.27)$$

dove  $V_H$  è la *tensione di hold* della cella in stato di *full reset* (Figura 3.1),

mentre  $R_S$  la resistenza serie della cella associata alla presenza di un *heater* metallico in  $TiN$ . Unendo l'Eq.(4.25) con l'Eq.(4.26) e l'Eq.(4.27) si ottiene:

$$T(I) = \begin{cases} T_0 + C(V_H + R_S I) & \text{per } I < I_{th}, \\ T_0 + R_{T_0}(V_H I + R_S I^2) & \text{per } I \geq I_{th}. \end{cases} \quad (4.28a)$$

$$(4.28b)$$

La resistenza associata all'*heater*  $R_S$  è estraibile da un'analisi delle curve IV della cella PCM al variare della temperatura nel criostato  $T_0$  (Figura 4.12(a)). Difatti, una volta avvenuto lo *switching* a soglia, il calcogenuro passa in regime di alta conduzione e la resistenza della serie GST+*heater* è dominata da quest'ultimo.

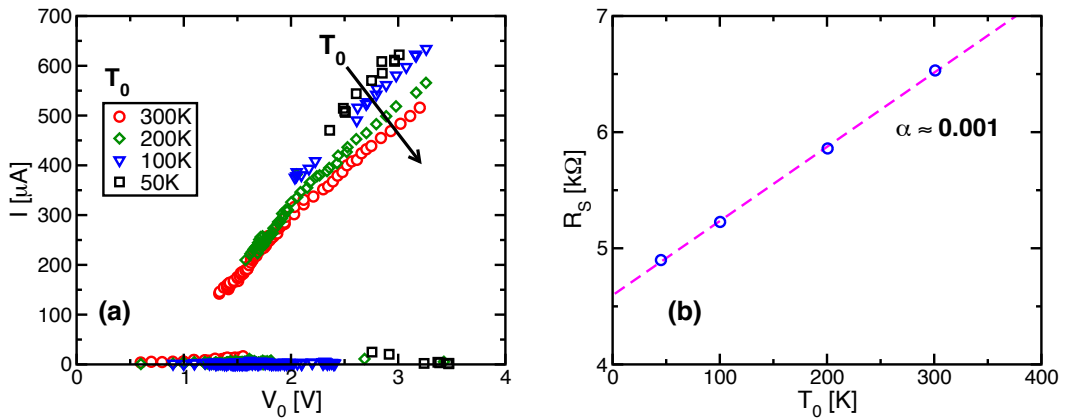


Figura 4.12: In (a) è riportato il diagramma IV della cella PCM 90nm 1T1R in esame, per diverse temperature nel criostato  $T_0$ . La resistenza di on  $R_{ON}$  corrisponde alla resistenza dell'*heater*  $R_S$  (v. Figura 3.1), e viene riportata nella figura (b) in funzione della temperatura nel criostato. Si osservi il comportamento metallico: la resistenza aumenta al crescere della temperatura. Il coefficiente termico  $\alpha$  definito dall'equazione  $R(T_0) = R_0(1 + \alpha(T_0 - T_{ref}))$ , con  $T_{ref} = 273K$ , vale  $1 \cdot 10^{-3}K^{-1}$ .

La resistenza estratta  $R_S$  in funzione dalla temperatura nel criostato  $T_0$  è riportata in Figura 4.12(b). Una prova importante per il modello è quella di riuscire a predire il valore della temperatura di fusione  $T_m$  del GST. Per fare ciò, è utile rappresentare in analogia a quanto fatto precedentemente (Figura

4.8) la dipendenza della corrente di fusione  $I_m$  in funzione della temperatura nel criostato  $T_0$ . I risultati sono riportati in Figura 4.13.

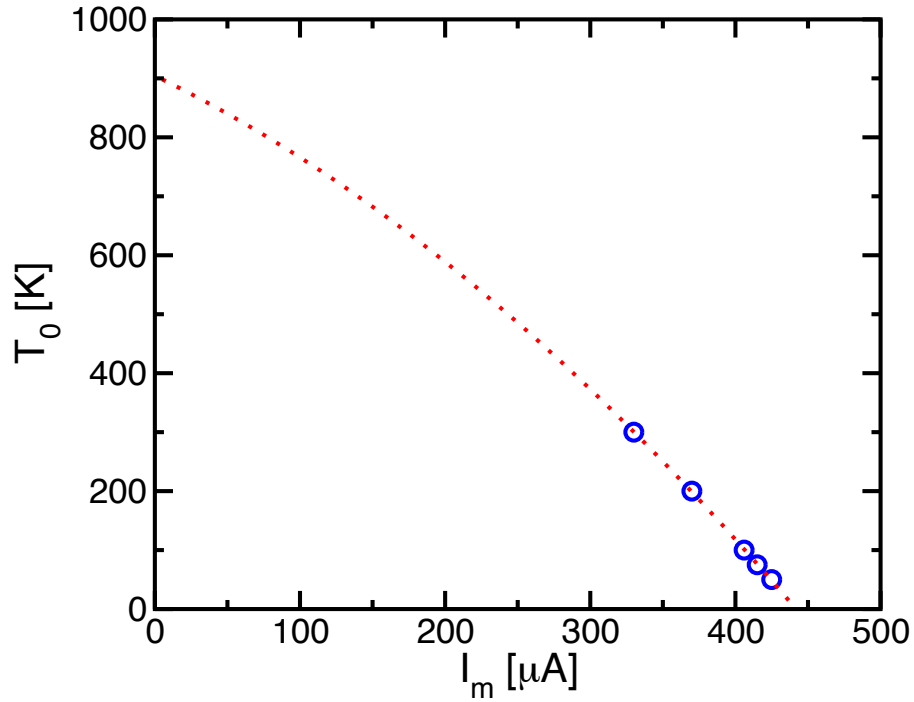


Figura 4.13: Il grafico mostra la relazione tra la corrente di fusione  $I_m$  presa sul fondo della curva di programmazione (Figura 4.7) e la temperatura nel criostato  $T_0$ . Applicando il modello elettro-termico di Eq.(4.28) con  $T = T_m$  si deve ottenere una dipendenza quadratica di  $T_0$  da  $I_m$ . L'intercetta con l'asse delle ordinate è la temperatura di fusione del GST, in questo caso prossima a  $900\text{K}$ , in ottimo accordo con il valore noto in letteratura.

Utilizzando l'Eq.(4.28) nella regione di *melting* ( $I_m > I_{th}$  e  $T = T_m$ ) si prevede una dipendenza quadratica della temperatura nel criostato  $T_0$  dalla corrente di *melting*  $I_m$ . Utilizzando come valore della resistenza dell'*heater* un valore medio  $\overline{R_S} = 5\text{k}\Omega$ , dal *fitting* parabolico dei dati in Figura 4.13 si può estrarre:  $T_m = 900\text{K}$ ,  $V_H = 0.55\text{V}$ ,  $R_{T_0} = 6 \cdot 10^5\text{K/W}$ , tutti valori in linea con i dati sperimentali.

A questo punto è possibile applicare il modello di Eq.(4.28) ai dati speri-

mentali, per produrre un grafico Arrhenius dove la temperatura è dedotta dal nuovo modello, con la stessa procedura adottata nel Paragrafo 4.3.2. I risultati sono riportati in Figura 4.14.

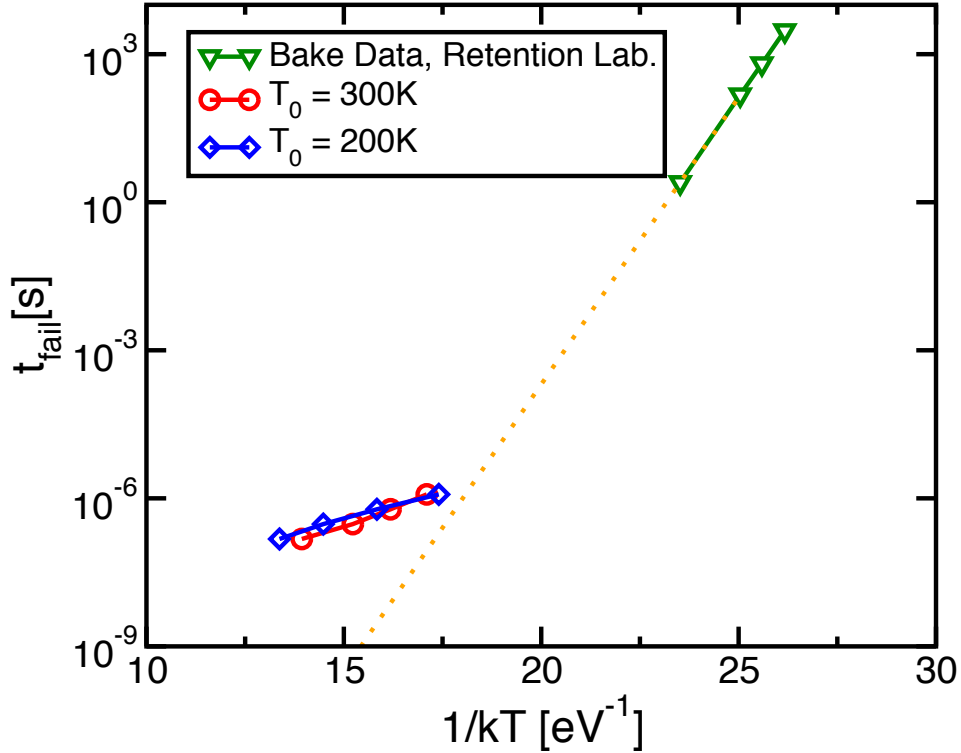


Figura 4.14: Arrhenius plot del tempo di fallimento  $t_{fail}$ . Si osservi come le temperature dedotte col modello di *Joule heating* a partire da  $300K$  e  $200K$  si sovrappongono bene, e come con l'introduzione di una descrizione più accurata della resistenza termica sia possibile un raccordo tra i dati di cristallizzazione in *bake* e quelli ottenuti mediante applicazione di impulso elettrico.

Si osservi come le serie ottenute con  $T_0 = 300K$  e con  $T_0 = 200K$  (Figura 4.9) tornino a sovrapporsi sul grafico Arrhenius. Il nuovo modello riesce inoltre a risolvere il problema evidenziato col modello a resistenza termica costante in Figura 4.10: risulta possibile un raccordo tra i dati ottenuti in *bake* e quelli ottenuti mediante applicazione di impulso elettrico.

In entrambi i modelli si nota una diversa energia di attivazione per la cristallizzazione tra le misure fatte in *bake* e quelle ottenute mediante applicazione di impulso elettrico. In particolare, l'energia di attivazione risulta inferiore nel caso di cristallizzazione indotta da impulso elettrico. Una spiegazione di quello che si osserva è possibile dall'analisi dall'Eq.(4.12) e (4.19). L'energia di attivazione per la cristallizzazione si compone di due contributi. Il contributo dovuto alla viscosità è trascurabile nel caso di cristallizzazione indotta da impulso elettrico, dove le temperature raggiunte sono ben sopra la temperatura di transizione vetrosa  $T_g$ , mentre è presente nel caso di *annealing* termico. Il contributo di energia di attivazione dovuto alla presenza di una dimensione critica per i nuclei cristallini cresce solo quando si è molto prossimi alla temperatura di fusione  $T_m$  (Eq.(4.7)), e perciò in entrambi i casi poco influisce.

Sebbene questa interpretazione dei risultati sia ragionevole, i meccanismi di cristallizzazione indotti da impulso elettrico non sono ancora ben chiari. Il presente lavoro apre la strada a successive indagini, che prevedono un'estesa caratterizzazione elettrica della transizione di *set*, al variare della tecnologia, della struttura, delle dimensioni della cella e della temperatura. Alla raccolta dei dati sperimentali dovrà essere affiancato un lavoro di modellistica per la simulazione fisica degli effetti elettro-termici che coinvolgono l'operazione di programmazione di una cella PCM.

# Conclusioni

Il presente lavoro di tesi si è soffermato su due fenomeni fisici fondamentali nel funzionamento dei dispositivi di memoria a cambiamento di fase (PCM): il *drift* della tensione di soglia e la cristallizzazione indotta da impulso elettrico (*switching* di fase, o transizione di *set*).

Il primo fenomeno riguarda la deriva spontanea nel tempo (*drift*) della tensione di soglia  $V_T$ , che è definita come il valore di tensione da applicare alla cella per innescare la transizione di conducibilità nel materiale attivo, e che quindi demarca i regimi operativi di lettura e di programmazione del dispositivo PCM. Le misure effettuate hanno indagato l'aumento della tensione di soglia nel tempo dopo la programmazione su un intervallo temporale di più di dieci decenni. Una corretta previsione del *drift* della tensione di soglia è essenziale al fine di permettere che le operazioni di lettura e scrittura del dispositivo PCM abbiano una bassa probabilità di fallimento in ogni momento del suo ciclo di funzionamento. È stato sviluppato un modello per la previsione del *drift* della tensione di soglia, che ha tenuto conto del rilassamento strutturale della fase amorfa e del fenomeno di *switching* dovuto all'aumento di energia dei portatori a causa dell'alto campo elettrico. Il modello ha permesso di ottenere previsioni coerenti con i dati sperimentali. Gli sviluppi futuri del modello avranno come obiettivo la possibilità di prevedere gli effetti della geometria di cella e dello *scaling* sul valore della tensione di soglia e del *drift* ad esso associato, in modo

tale che questo parametro fondamentale per il funzionamento dei dispositivi PCM possa essere controllato da ogni punto di vista.

La comprensione dello *switching* di fase è decisiva in quanto la cristallizzazione spontanea è il fattore che limita la ritenzione del dato nella tecnologia PCM. Da un altro punto di vista, la cristallizzazione è anche indispensabile per lo *switching* dallo stato alto resistivo a quello basso resistivo, pertanto alte prestazioni del dispositivo richiedono una veloce cristallizzazione indotta da impulso elettrico. Lo studio confronta i regimi di cristallizzazione lenta (ore e giorni) ottenibile da misure accelerate in temperatura (*annealing*) e di cristallizzazione rapida (10 – 100ns) indotta dall'applicazione di un impulso elettrico al dispositivo PCM. In particolare, ci si è concentrati sulla possibilità di confrontare i due regimi in un unico grafico di Arrhenius, mediante la deduzione della temperatura locale sviluppata per effetto Joule e responsabile dell'accelerazione della cristallizzazione veloce. La caratterizzazione elettrica della cristallizzazione ha richiesto diversi mesi di lavoro per poter effettuare misure affidabili a temperature criogeniche, che sono state pensate per lo studio di un modello analitico di *Joule heating* in grado di spiegare come l'impulso elettrico induca un aumento di temperatura nel materiale attivo. Il grafico Arrhenius ottenuto ha mostrato diverse energie di attivazione per la cristallizzazione indotta da *annealing* isoterma e per quella indotta da impulso elettrico. Tale risultato è stato interpretato alla luce della teoria classica della nucleazione e crescita cristallina, che prevede una dipendenza dalla temperatura del valore dell'energia di attivazione per la cristallizzazione. Sviluppi futuri di questo lavoro riguardano l'estensione della caratterizzazione a diverse temperature e diverse dimensioni di cella. L'obiettivo è quello di comprendere più approfonditamente i fenomeni elettro-termici coinvolti nell'operazione di programmazione, oltre che i fattori determinanti nella cristallizzazione dei materiali calcogenuri. Tali modelli potranno permettere di definire le proprietà dei materiali che meglio

garantiscono un'alta ritenzione alle basse temperature e/o una veloce transizione di fase indotta da impulso elettrico, con l'obiettivo di fare della tecnologia PCM una credibile candidata per una memoria universale in grado di assolvere alle diverse funzioni di memoria, dalla *embedded* SRAM in microprocessori alla memoria non volatile in dischi a stato solido.



# Bibliografia

- [1] Y. Koh, “Nand flash scaling beyond 20nm,” *IMW. IEEE*, 2009.
- [2] D. Ielmini, “Reliability issues and modeling of Flash and post-Flash memory,” *IMW. IEEE*, vol. 86, pp. 1870–1875, 2009.
- [3] G. Montemurro, “Studio dell’affidabilità di celle di memoria PCM multilivello,” pp. 29–39, 2010.
- [4] J. Hwang, J. Seo, Y. Lee, J. L. S. Park, J. Kim, T. Hong, S. Jeong, K. Lee, H. Heo, H. Lee, P. Jang, K. Park, M. Lee, S. Baik, J. Kim, H. Kkang, M. Jang, J. Lee, G. Cho, J. L. B. L. H. Jang, S. Park, J. Kim, S. Lee, S. Aritome, S. Hong, and S. Park, “A middle-1x nm NAND Flash Memory Cell (M1X-NAND) with highly manufacturable integration technologies,” *IEDM Conference*, 2011.
- [5] S. Tehrani, J. M. Slaughter, M. Deherrera, B. N. Engel, N. D. Rizzo, J. Salter, M. Durlam, R. W. Dave, J. Janesky, B. Butcher, K. Smith, and G. Grynkewich, “Magnetoresistive random access memory using magnetic tunnel junctions,” *Proc. IEEE*, vol. 91, pp. 703 – 714, 2003.
- [6] S. Y. Lee and K. Kim, “Future 1T1C FRAM technologies for highly reliable, high density FRAM,” in *IEDM Tech. Dig.*, 2002, pp. 547–550.
- [7] S. Lai, “Current status of the phase change memory and its future,” 2003, pp. 255 – 258.
- [8] I. Baek, M. Lee, S. Seo, M. Lee, D. Seo, D.-S. Suh, J. Park, S. Park, H. Kim, I. Yoo, U.-I. Chung, and J. Moon, “Highly scalable nonvolatile resistive memory using simple binary oxide driven by asymmetric unipolar voltage pulses,” 2004, pp. 587 – 590.
- [9] 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. and Dev.*, vol. 52, pp. 449 – 464, 2008.

- 
- [10] S. Lai, "Non-volatile memory technologies: The quest for ever lower cost," in *IEDM Tech. Dig.*, 2008, pp. 11–16.
- [11] S. Ovshinsky, "Reversible electrical switching phenomena in disordered structures," *Phys. Rev. Lett.*, vol. 21, pp. 1450–1453, 1968.
- [12] F. Pellizzer, A. Benvenuti, B. Gleixner, Y. Kim, B. Johnson, M. Magistretti, T. Marangon, A. Pirovano, R. Bez, and G. Atwood, "A 90nm phase change memory technology for stand-alone non-volatile memory applications," in *Symp. VLSI Technology, Dig. Tech.*, 2006, pp. 122–123.
- [13] M. Boniardi, D. Ielmini, A. Lavizzari, A. Lacaíta, A. Redaelli, A. Pirovano, I. Tortorelli, M. Allegra, M. Magistretti, C. Bresolin, D. Erbetta, A. Modelli, E. Varesi, F. Pellizzer, and R. Bez, "Impact of material composition on the write performance of phase-change memory devices," *Proc. IMW*, pp. 25–28, 2010.
- [14] T. Happ, M. Breitwisch, A. Schrott, J. Philipp, M. Lee, R. Cheek, T. Nirschl, M. Lamorey, C. Ho, S. Chen, C. Chen, E. Joseph, S. Zaidi, G. Burr, B. Yee, Y. Chen, S. Raoux, H. Lung, R. Bergmann, and C. Lam, "Novel one-mask self-heating pillar phase change memory," in *Symp. VLSI Tec. Dig.*, 2006, pp. 120–121.
- [15] D.-H. Kang, J.-H. Lee, J. Kong, D. Ha, J. Yu, C. Um, J. Park, F. Yeung, J. Kim, W. Park, Y. Jeon, M. Lee, Y. Song, J. Oh, G. Jeong, and H. Jeong, "Two-bit cell operation in diode-switch phase change memory cells with 90nm technology," 2008, pp. 98 – 99.
- [16] F. Bedeschi, R. Fackenthal, C. Resta, E. M. Donze, M. Jagasivamani, E. Buda, F. Pellizzer, D. Chow, A. Cabrini, G. Calvi, R. Faravelli, A. Fantini, G. Torelli, D. Mills, R. Gastaldi, and G. Casagrande, "A multi-level-cell bipolar-selected phase-change memory," in *Proc. ISSCC*, 2008, pp. 428–429.
- [17] R. G. Neale, D. L. Nelson, and G. E. Moore, "Nonvolatile and reprogrammable, the read-mostly memory is here," *Appl. Phys. Lett.*, vol. 96, p. 163501, 2010.
- [18] S. Raoux and M. Wuttig, *Phase Change Materials*. Springer, New York, 2009.
- [19] M. Breitwisch, T. Nirschl, C. Chen, Y. Zhu, M. Lee, M. Lamorey, G. Burr, E. Joseph, A. Schrott, J. Philipp, R. Cheek, T. Happ, S. Chen, S. Zaidi, P. Flaitz, J. Bruley, R. Dasaka, B. Rajendran, S. Rossnagel, M. Yang, Y. Chen, R. Bergmann, H. Lung, and C. Lam, "Novel lithography-independent pore phase change memory," in *Simp. VLSI Tech. Dig.*, 2007, pp. 100–101.

- [20] Y. Ha, J. Yi, J. Horii, J. Park, S. Joo, S. Park, U.-I. Chung, and J. Moon, "An edge contact type cell for phase change RAM featuring very low power consumption," in *Symp. VLSI Tech. Dig.*, 2003, pp. 175–176.
- [21] S. Ahn, Y. Hwang, Y. Song, S. Lee, S. Lee, J. Park, C. Jeong, K. Ryoo, J. Shin, Y. Fai, J. Oh, G. Koh, G. Jeong, S. Joo, S. Choi, Y. Son, J. Shin, Y. Kim, H. Jeong, and K. Kim, "Highly reliable 50nm contact cell technology for 256mb pram," in *Symp. VLSI Tech. Dig.*, 2005, pp. 98–99.
- [22] F. Pellizzer, A. Pirovano, F. Ottogalli, M. Magistretti, M. Scaravaggi, P. Zuliani, M. Tosi, A. Benvenuti, P. Besana, S. Cadeo, T. Marangon, R. Morandi, R. Piva, A. Spandre, R. Zonca, A. Modelli, E. Varesi, T. Lowrey, A. Lacaita, G. Casagrande, P. Cappelletti, and R. Bez, "Novel  $\mu$ trench phase-change memory cell for embedded and stand-alone non-volatile memory applications," in *Symp. on VLSI Tech. Dig.*, 2004, pp. 18–19.
- [23] S. Kim, B. Lee, M. Asheghi, G. Hurkx, J. Reifenberg, K. Goodson, and H.-S. Wong, "Thermal disturbance and its impact on reliability of phase-change memory studied by the micro-thermal stage," in *Proc. IRPS*, 2010, pp. 99–103.
- [24] S. Disegna, "Caratterizzazione tempo-risolta della cristallizzazione indotta da campo elettrico in memorie a cambiamneto di fase." pp. 22–31, 2011.
- [25] D. Ielmini, "Threshold switching mechanism by high-field energy gain in the hopping transport of chalcogenide glasses," *Phys. Rev. B*, vol. 78, p. 0353008, 2008.
- [26] A. P. A. Redaelli, F. Pellizzer, A. Lacaita, D. Ielmini, and R. Bez, "Electronic switching effect and phase-change transition in chalcogenide materials," *IEEE Electron Device Lett.*, vol. 25, pp. 684–686, 2004.
- [27] D. Ielmini, D. Mantegazza, A. Lacaita, A. Pirovano, and F. Pellizzer, "Parasitic reset in the programming transient of phase change memories," *IEEE Electron Device Lett.*, vol. 26, pp. 799–801, 2005.
- [28] D. Adler, M. S. Shur, M. Silver, and S. R. Ovshinsky, "Threshold switching in chalcogenide-glass thin films," *J. Appl. Phys.*, vol. 51, no. 6, pp. 3289–3309, 1980.
- [29] D. I. D. Fugazza, S. Lavizzari, and A. L. Lacaita, "Distributed-Poole-Frenkel modeling of anomalous resistance scaling and fluctuations in phase-change memory (pcm) devices," *IEDM*, vol. 78, pp. 723–726, 2009.

- [30] D. Ielmini, A. L. Lacaita, and D. Mantegazza, "Recovery and drift dynamics of resistance and threshold voltages in phase-change memories," *IEEE Trans. Electron Devices*, vol. 54, pp. 308–315, 2007.
- [31] S. Raoux, W. Welnic, and D. Ielmini, "Phase change materials and their application to nonvolatile memories," *Chem. Rev.*, vol. 110, pp. 240–267, 2010.
- [32] N. F. Mott and E. A. Davis, *Electronic processes in non-crystalline materials*. Clarendon Press, Oxford, 1979.
- [33] C. Thomas, D. Ielmini, S. Lavizzari, and A. L. Lacaita, "The temperature dependence of the non-ohmic current and switching characteristics of a chalcogenide glass," *J. Phys. D.*, vol. 9, pp. 2587–2596, 1976.
- [34] D. Ielmini and M. Boniardi, "A common signature of many-body thermal excitation in structural relaxation and crystallization of chalcogenide glasses," *Appl. Phys. Lett.*, vol. 94, p. 1906, 2009.
- [35] D. Ielmini, D. Fugazza, M. Boniardi, G. Montemurro, and A. Lacaita, "Kinetic of resistance drift in PCM by structural relaxation of the amorphous chalcogenide phase," *EP\*COS*, 2010.
- [36] J. Mullin, *Crystallization*. Oxford: Butterworth Heinemann, 1993.
- [37] A. Redaelli, A. Pirovano, I. Tortorelli, and D. Ielmini, "A reliable technique for experimental evaluation of crystallization activation energy in PCMs," *IEEE Electron Device Lett.*, vol. 29, pp. 41–43, Jan. 2008.
- [38] M. Kuwahara, O. Susuki, Y. Yamakawa, P. Foms, T. Yagi, N. Taketoshi, K. Tsutsumi, M. Suzuki, J. Tominaga, and T. Baba, "Measurement of thermal and optical properties at high temperature for optical disk materials," *ISTDTPE*, vol. 29, pp. 18–20, Jun. 2008.