

III.TEST CHIP

1.Obiettivi del Test Chip

intervenire sulla tecnologia.

Il design del *Test Chip* multilivello si basa su caratterizzazioni relative ad una memoria flash convenzionale a due livelli, per cui, il primo obiettivo consiste nella realizzazione pratica del chip e come secondo obiettivo, una volta prodotto, la sua caratterizzazione. Nella progettazione vengono utilizzati dei *tools* di simulazione software, i quali permettono di prevedere il comportamento reale del dispositivo. Purtroppo però le simulazioni non sempre rispecchiano il comportamento reale del dispositivo sul silicio, perché nella realtà intervengono molti fenomeni come i disturbi in lettura e scrittura, l'invecchiamento, le trappole nell'ossido, ecc.

Fenomeni questi difficilmente modellabili per la loro imprevedibilità, e di conseguenza il dispositivo potrebbe avere caratteristiche reali differenti rispetto a quelle di progetto.

La caratterizzazione consiste quindi nel testare il *Test Chip*, cioè effettuare tutte le possibili misure che permetteranno di valutare le reali potenzialità dell'approccio multilivello, e la possibilità di effettuare una analisi statistica sul funzionamento del dispositivo. Inoltre permetterà anche di valutare le varie tecniche di *programming* (algoritmica e analogica) e di *sensing*.

2.Schema a blocchi

Fino ad oggi il design è stato realizzato nella tecnologia 0.4u (u = micron) Texas Instruments. La Fig.III.1 mostra lo schema a blocchi del *Test Chip*.

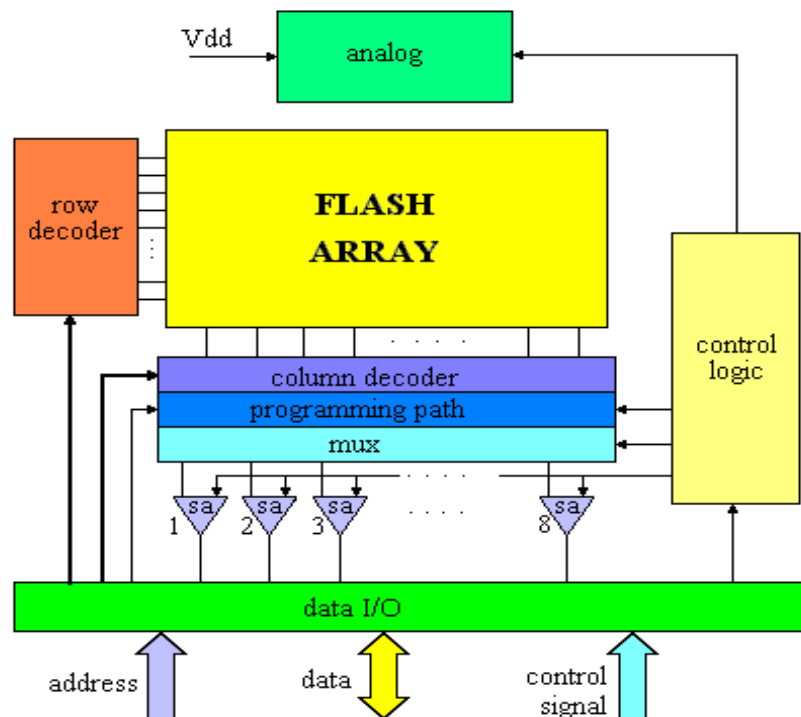


Fig.III.1.Schema a blocchi del Test Chip

Nel presente, e nei successivi paragrafi, viene fatta una descrizione dei singoli blocchi.

Flash array

Il blocco flash array costituisce la matrice di celle della memoria nella quale andranno immagazzinati i dati.

La caratteristica di una memoria flash è di avere una divisione in settori della matrice di celle, inoltre le celle dello stesso settore hanno in comune il terminale di source. Questo spiega il perché l'operazione di erase comporta la cancellazione di un intero settore, infatti le celle dello stesso settore nella cancellazione si trovano nelle stesse condizioni di polarizzazione, se non ci fosse tale suddivisione si avrebbe la cancellazione dell'intera memoria. Certamente la settorizzazione comporta un'ulteriore complessità dei circuiti di decodifica di colonna i quali dovranno dapprima permettere una selezione per settore, e successivamente una decodifica di colonna.

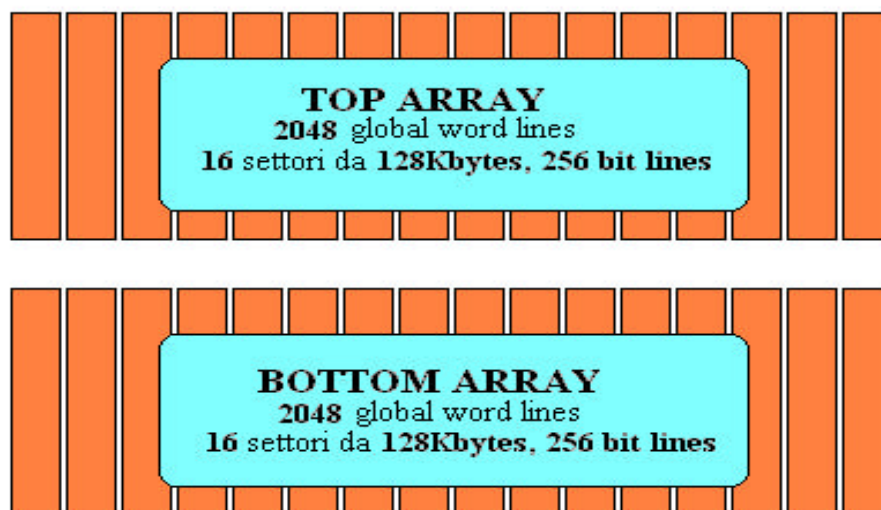


Fig.III.2.Settorizzazione dell'array di celle

Dalla fig.III.2 notiamo una divisione dell'array di celle in due blocchi, un blocco TOP e l'altro BOTTOM. A sua volta su ciascun blocco è operata una settorizzazione verticale in sedici settori, ogni settore è costituito da 2048 righe definite *word lines* e 256 colonne definite *bit lines*. Per quanto riguarda le *word lines* bisogna distinguere in *global word lines*, realizzate in metal2 e che percorrono orizzontalmente tutto l'array; e in *local word lines*, realizzate in poly2 e che percorrono localmente il singolo settore.

L'architettura della matrice di celle è di tipo NOR, in questa architettura le celle appartenenti alla stessa riga hanno le *control gate* collegate alla stessa *local word lines*, e le celle sulla stessa colonna hanno le *drain* collegate alla stessa *bit lines*, tutte le celle del settore hanno le source collegate allo stesso *array source*(fig.III.3).

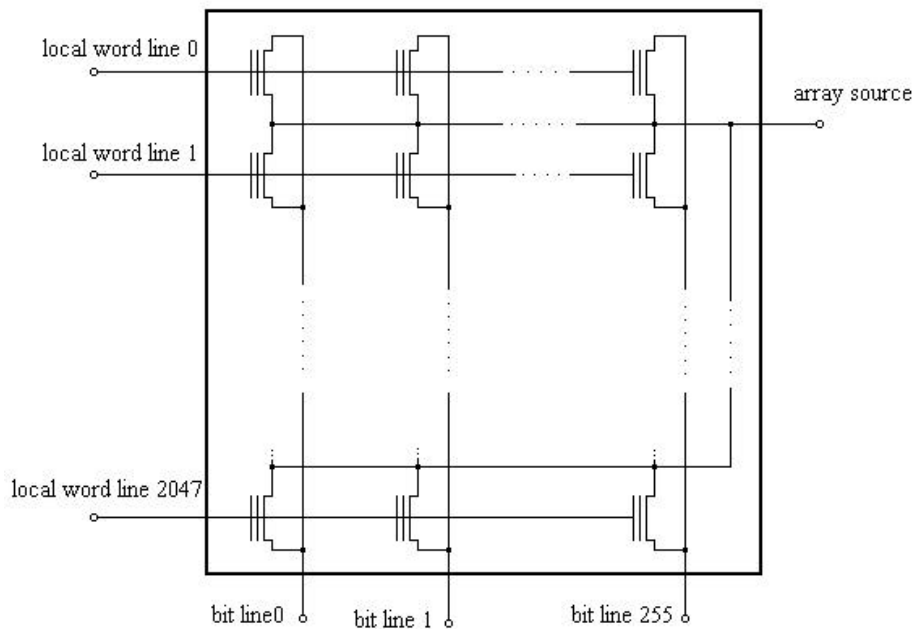


Fig.III.3.Settore da 128Kbytes

La divisione tra *global* e *local word lines* è necessaria perché quando viene selezionata una *global word lines* si vuole accedere al singolo settore, ciò significa che le *global word lines* dell'array non sono collegate direttamente alle corrispondenti *local word lines* dei settori, ma fra loro si frappongono dei transistor di selezione. Quindi il collegamento si realizza soltanto quando il corrispondente transistor di *local word lines* del settore sul quale si vuole accedere è selezionato.

Per motivi di occupazione di area è stato realizzato un raggruppamento a coppia dei settori, e tra ciascuno di questi raggruppamenti vi sono inseriti i transistor di selezione delle *local word lines* (Fig.III.4).

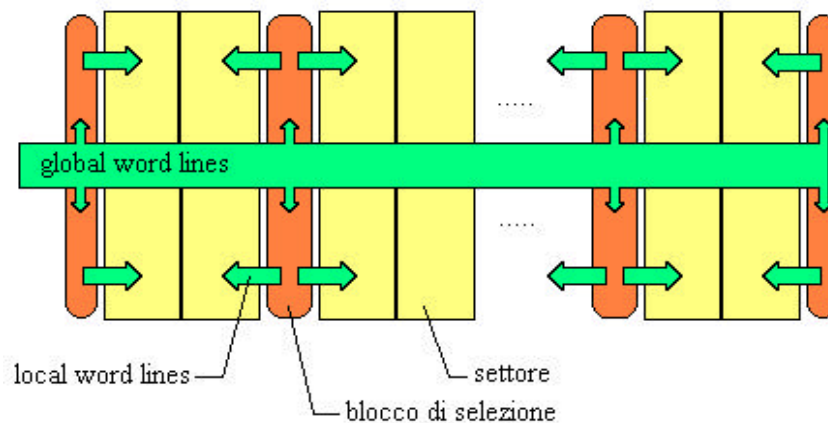


Fig.III.4.Raggruppamento a coppia dei settori con blocchi di selezione.

Row e column decoder

I due circuiti opportunamente indirizzati, permettono di accedere ad una parola di memoria per effettuare operazioni di scrittura e lettura.

In particolare il row decoder permette di selezionare una tra 4096 righe dell'array, mentre il column decoder apre un percorso dall'array verso l'esterno.

In una memoria costituita da 16 mega celle, le linee di indirizzo sono complessivamente 20, di cui le prime 10 vengono utilizzate per l'indirizzamento di riga, le restanti permettono la selezione di settore e di colonna.

Programming Path

Questo blocco attiva il percorso alla corrente di programmazione dal blocco analogico alle celle di memoria indirizzate per la programmazione.

Mux

Questo livello di *mux* è stato introdotto per evitare, durante l'operazione di programmazione, che le alte tensioni dal *programming path* raggiungano i *sense amplifiers*.

Data I/O

È il blocco che permette al chip di comunicare con l'esterno; in esso convengono i dati d'ingresso e d'uscita, le linee di indirizzo e i segnali di controllo relativi alle diverse modalità di funzionamento del Test Chip.

3.Sense amplifiers

Lettura di memorie multilivello

La lettura di una cella multilivello viene eseguita sentendo la corrente *I_{cell}* che la cella eroga in determinate condizioni di polarizzazione, cioè quando sul *control-gate* e sul *drain* vengono applicati determinati valori di tensione. La caratteristica *tensione-corrente* di una cella programmata a differenti valori di tensione di soglia mostra che il valore di *I_{cell}* risulta correlato in modo univoco alla tensione di soglia programmata, e quindi all'informazione immagazzinata (fig.III.5).

I valori delle tensioni di polarizzazione V_G (tensione di *control-gate*) e V_D (tensione di *drain*) devono essere scelti adeguatamente. Da un lato non possono essere troppo elevati per minimizzare gli effetti di disturbo in lettura, infatti per valori di V_G superiori ai 7V si ha, in lettura, una variazione di carica immagazzinata nella *floating-gate*. Per permettere una memorizzazione e una lettura affidabile è necessario che l'intervallo di corrente di cella sia sufficientemente largo, per allocare i diversi livelli di corrente programmabili. Questo induce a non scegliere delle tensioni di polarizzazione troppo basse.

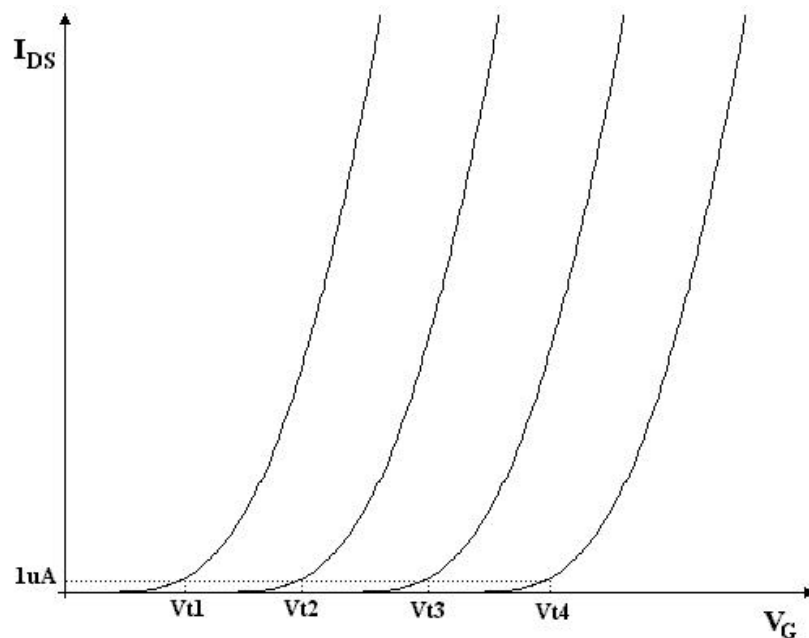


Fig.III.5. Caratteristiche tensione-corrente di una cella per diversi valori di tensione di soglia programmata

Attualmente un valore adeguato per la tensione di *drain* V_D risulta essere dell'ordine di 1V. Per quanto riguarda la tensione di *control-gate*, un buon compromesso si raggiunge con valori nell'ordine di 5-6V. Questa tensione è più elevata di quella usuale di alimentazione (5V, e nei circuiti più moderni 3V e anche 1.8V), e deve essere generata internamente al *chip* con opportuni elevatori di tensione integrati definiti pompe di carica (*charge pumps*).

La lettura di una cella viene in genere eseguita confrontando la corrente di cella *I_{cell}* con opportuni riferimenti. I riferimenti sono generati da celle di memoria fisicamente identiche a quelle dell'array, che vengono programmate in modo adeguato e in fase di lettura vengono polarizzate con le stesse tensioni. Si assicura così un ottimo *matching* costruttivo tra i riferimenti e le celle da leggere. Ciò a sua volta garantisce una buona indipendenza del risultato della lettura dai valori dei parametri fisici ed elettrici del circuito, che possono subire variazioni notevoli in funzione del processo di fabbricazione e delle condizioni ambientali (come tensione di alimentazione e temperatura).

Tecniche di lettura

La Fig.III.6 mostra lo schema a blocchi di un sense amplifier usato nella memoria multilivello. Il sense amplifier è caratterizzato dai seguenti segnali di ingresso:

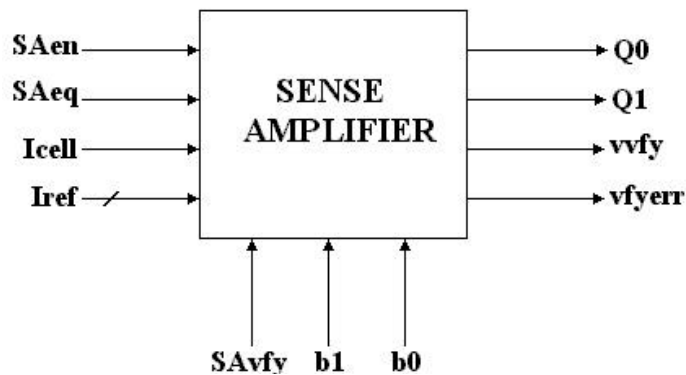


Fig.III.6. Blocco generale di sensing

- **Saen** : è un segnale di abilitazione il quale seleziona una tra le possibili tecniche di lettura;
- **Icell** : è la corrente proveniente dalla cella di memoria, la quale verrà confrontata con opportuni riferimenti di corrente;
- **Iref** : è la corrente di riferimento;

- **Saeq** : è il segnale di equalizzazione, abilitato prima di ogni operazione di lettura o di verifica, per evitare possibili instabilità dei segnali;
- **Savfy** : è un segnale di controllo, e viene abilitato nell'operazione di verify, operazione richiesta nella programmazione algoritmica;
- **b0, b1** : utilizzati nell'operazione di verify, e rappresentano il dato da memorizzare.

Quando la cella viene letta, i due bit **Q0** e **Q1** vengono mandati in uscita al sense amplifier. Quando la cella è verificata, in uscita al sense amplifier vengono generati due segnali:

- **vvfy** : il quale è '0' quando la cella non ha raggiunto lo stato di programmazione richiesto, '1' quando la cella è programmata con opportuni margini;
- **vyerr** : il quale vale '1' quando la cella si è programmata troppo.

Due tecniche di lettura sono implementate nel *Test Chip*, definite una **Three stages sense amplifier** e l'altra **Distance sense amplifier**. Ciascuna di queste tecniche verrà caratterizzata e la migliore sarà usata nel chip finale.

Three stages sense amplifier

Lo schema a blocchi è riportato in Fig.III.7 (TSA).

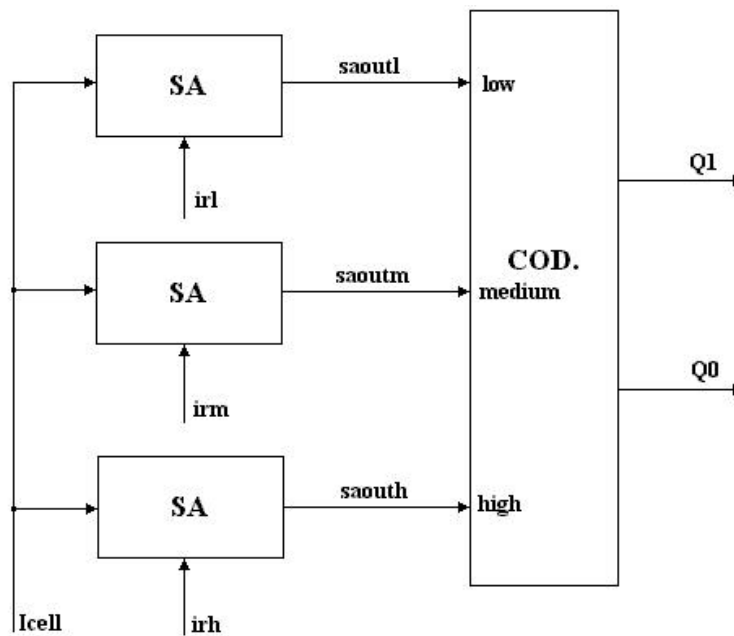


Fig.III.7.Schema a blocchi del TSA

La corrente di cella viene confrontata con tre riferimenti di corrente: **irl** (low), **irm** (medium) e **irh** (high), utilizzando tre sense amplifiers tradizionali, ciascuno dei quali fornisce una tensione di uscita, rispettivamente, **saoutl**, **saoutm** e **saouth**. I tre segnali di tensione vengono mandati in ingresso a un codificatore a priorità, il quale dà in uscita due bit **Q0** e **Q1** associati alla corrente di cella **Icell**.

Le tre correnti di riferimento variano di un certo margine per l'operazione di verify, rispetto alla lettura. La Fig.III.8 mostra i margini di lettura e di verify.

Se la corrente di cella è maggiore del margine superiore, relativo allo stato di programmazione da raggiungere, la condizione di verify non è soddisfatta e la programmazione dovrà continuare. Se invece la **Icell** si trova dentro l'intervallo di corrente compreso tra i limiti superiore ed inferiore, la condizione di verify risulta vera e la programmazione termina positivamente. Se la corrente di cella andrà al di sotto del limite di corrente inferiore, ciò significa che la cella si è programmata troppo, e il sense amplifier produrrà in uscita il segnale di errore **vferr**.

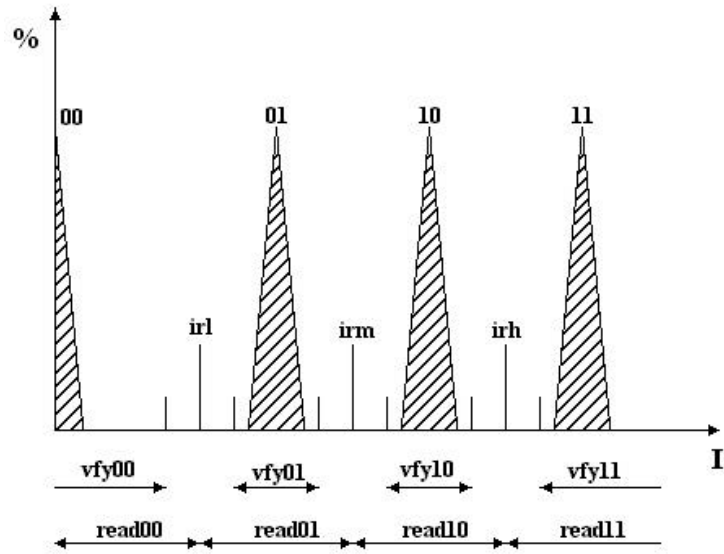


Fig.III.8.Margini di lettura e di verify TSA

Per ciascuna delle strutture di Fig.III.7, devono essere generati quattro valori di corrente: tre per i riferimenti e uno per il margine di verify. Queste correnti vengono generate a partire da celle FAMOS programmabile esternamente a differenti valori.

Distance sense amplifier

Nella Fig.III.9 è mostrato lo schema a blocchi del distnce sense amplifier (DSA).

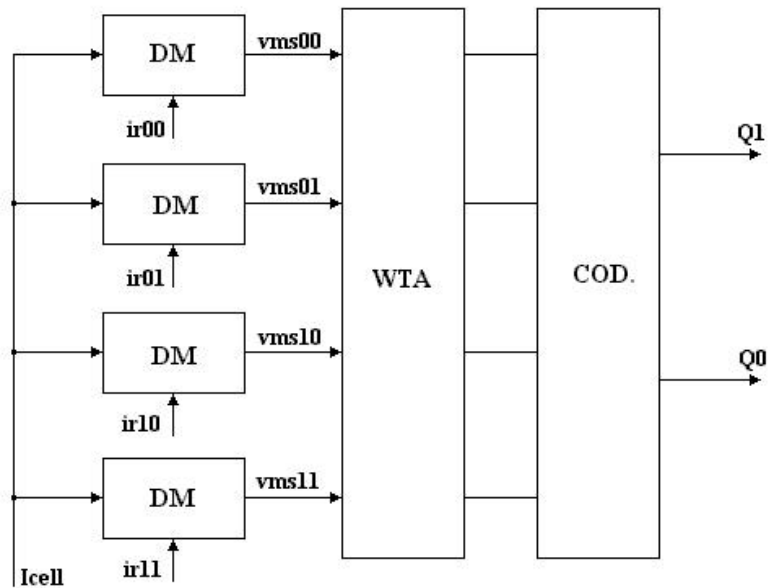


Fig.III.9.Schema a blocchi del DSA

Lo schema è costituito da quattro blocchi DM, ciascuna dei quali misura la distanza tra la corrente di cella I_{cell} e il corrispondente riferimento di corrente, fornendo in uscita un segnale di tensione. Le uscite $vms00$, $vm01$, $vms10$ e $vms11$ vanno in ingresso a un circuito, definito winner take all, che determina tra le quattro distanze quella minore fornendo in uscita quattro segnali digitali, i quali verranno codificati in bit **Q0** e **Q1** da un blocco codificatore.

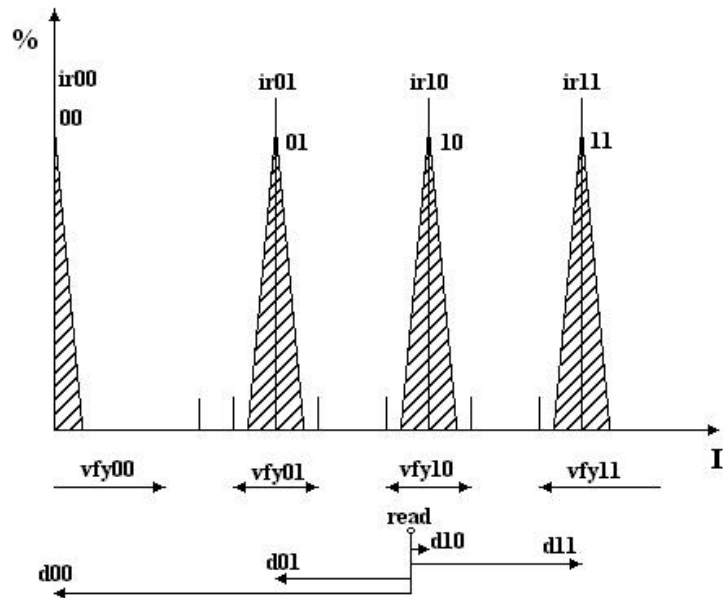


Fig.III.10.Margini di lettura e di verify per DSA

Il distance sense amplifier è provvisto anche di un circuito per l'operazione di verify con gli opportuni margini riportati in Fig.III.10. Inoltre è caratterizzato anche da un circuito che segnala la condizione di errore quando nella programmazione è stato superato il limite di verify.

Per ciascuna struttura di Fig.III.9 devono essere generati cinque correnti: quattro per i riferimenti e uno per il margine di verify.

4.Analog

L'analog costituisce la parte analogica, dove vengono generate le diverse tensioni necessarie per le modalità di funzionamento del *Test Chip*. Si utilizzano pompe di carica per la generazione delle diverse tensioni di lavoro, a partire dalla tensione di alimentazione Vdd. Sono previste nel *Test Chip* le seguenti pompe:

- pompa ad alta tensione (High Voltage Pump: **HVP**);
- pompa di corrente (High Current Pump: **HCP**).

High Voltage Pump

Lo schema a blocchi della pompa HV è mostrato in Fig.III.11.

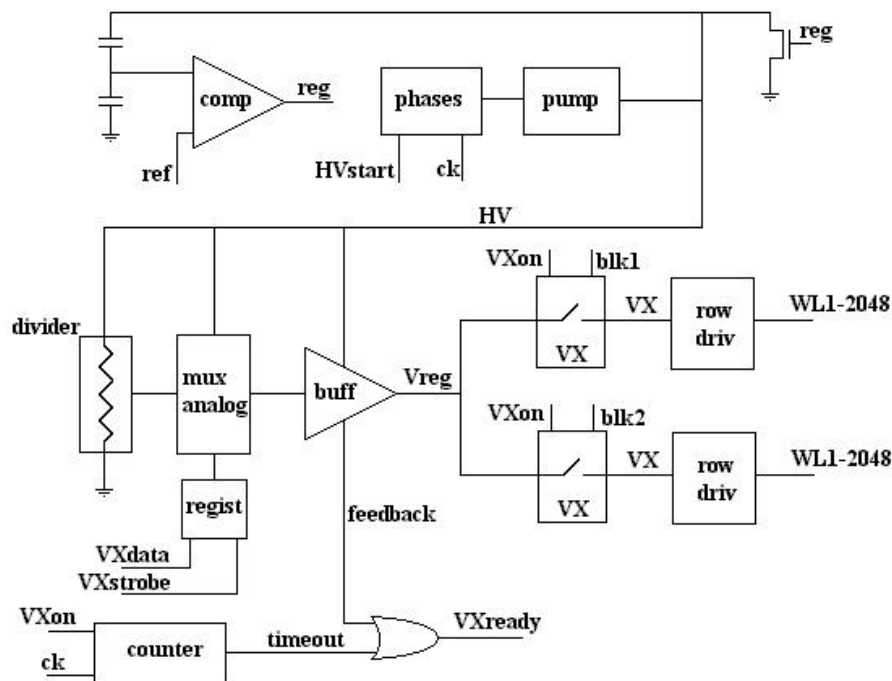


Fig.III.11.Schema a blocchi della pompa HVP

La pompa HV è costituita da un generatore di alta tensione HV, e un generatore di *Vx*. Il generatore HV è una pompa regolata capace di generare una tensione di 12

V, questa tensione viene applicata a un divisore di tensione, il quale può fornire delle tensioni intermedie. Il valore della tensione viene selezionato attraverso un mux analogico, pilotato da 8bits V_x data, provenienti dalla write state machine.

Infine la tensione intermedia V_x viene applicata alla word line selezionata dal blocco di decodifica. La V_x può variare da 0 a 12 V, con step di incremento di 45mV.

High Current Pump

La Fig.III.12 mostra lo schema a blocchi della pompa HC.

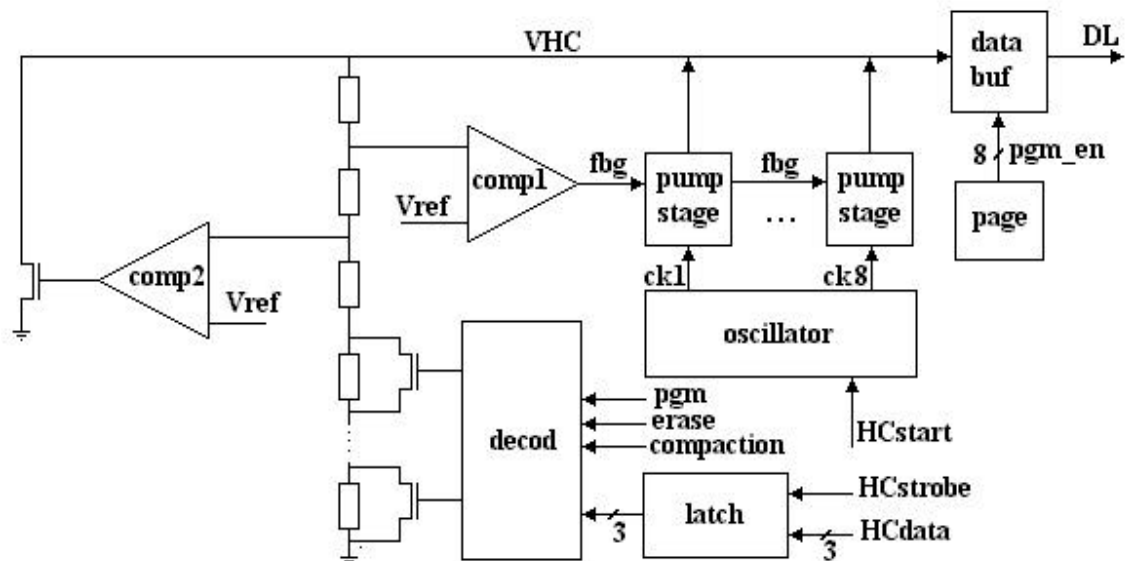


Fig.III.12.Schema a blocchi della pompa HCP

Lo schema è costituito da una pompa a otto stadi sincronizzati da un oscillatore, da un regolatore basato su un partitore resistivo, e da due comparatori (comp1 e comp2).

I valore VHC è regolato per ottenere una tensione diversa per la cancellazione, la compattazione e la programmazione. In particolare, per la programmazione, il valore di VHC è compreso tra 4.5V e 7V (con step di 0.5V).

I segnali HCdata, HCstrobe e HCstart arrivano alla pompa HC, dalla write state machine.

5. Control logic

È il blocco che gestisce il funzionamento della memoria in conformità all'operazione da svolgere. Esso è costituito da una **write state machine** avente un set di istruzioni molto ridotto, e da una memoria per immagazzinare le istruzioni da eseguire.

Write state machine

In Fig.III.13 è mostrato lo schema a blocchi della write state machine (WSM).

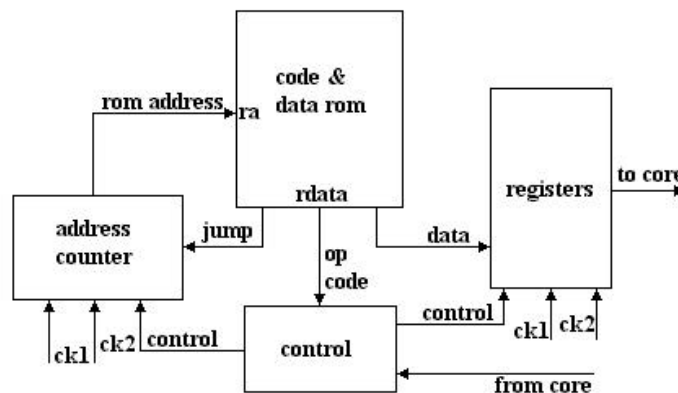


Fig.III.13.Schema a blocchi della WSM

La WSM è un controllore basato su un set ridotto di istruzioni caricate nel blocco code & data ROM. Nel *Test Chip* il codice delle istruzioni sarà caricato su una Flash EEPROM organizzata in 2048 righe, ciascuna di 32bits.

Il set di istruzioni è il seguente: IDLE, SET , PULSE, TEST, JUMP, JUMPIF1, JUMPIF0, CALL, RET e WAIT.

La write state machine e' sincronizzata da due segnali di clock CK1 e CK2 non overlapping, di periodo 100nsec.

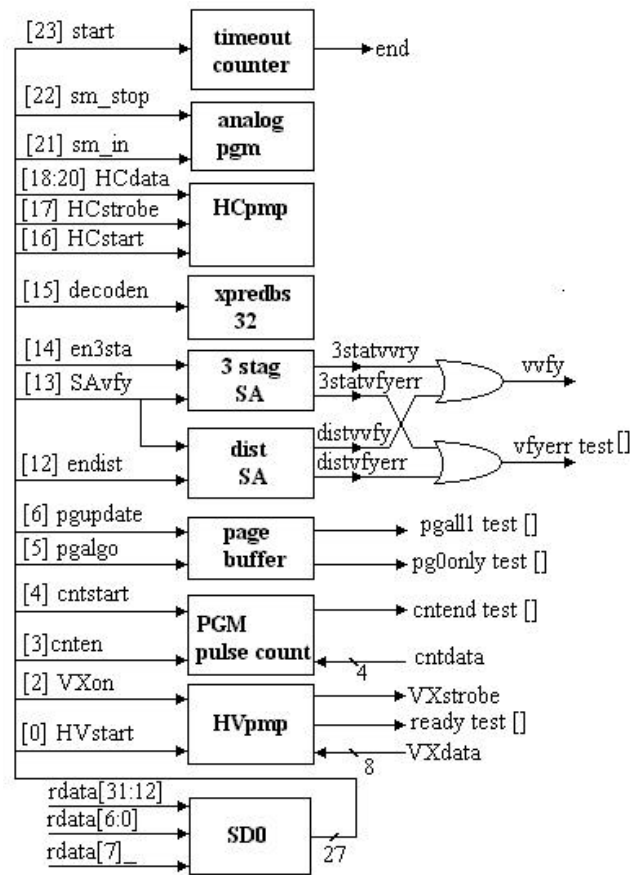


Fig.III.14.Registro sd0

I registri rappresentano l'interfaccia con l'hardware, in particolare per la MLC sono organizzati in due gruppi sd0 e sd1, come mostrato nelle figure Fig.III.14 e Fig.III.15. per la programmazione analogica la WSM, provvede a generare i segnali **sm_in** e **sm_stop** (il loro significato verrà descritto nel prossimo capitolo, dove si approfondirà la programmazione analogica).

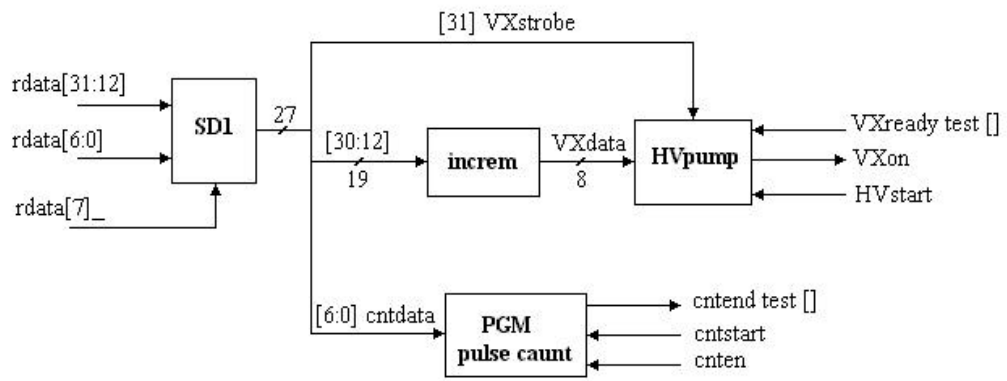


Fig.III.15.Registro **sd1**