

DISPOSITIVO DI MEMORIA NON-VOLATILE MULTILIVELLO E RELATIVO METODO DI LETTURA

Ricevete la proposta di brevetto in oggetto per l'invenzione sotto descritta e il documento allegato "MEMORIE FLASH MULTILIVELLO" che descrive i principi di funzionamento secondo l'arte nota.

Nei dispositivi ad alta capacità più recenti il metodo di lettura (fig. 4) è soggetto ad un margine di errore inaccettabile in quanto i livelli di tensione di lettura (V_{READ1} , V_{READ2} e V_{READ3}) sono fissi e predeterminati e così anche la tempistica (Δt) dei rispettivi intervalli. Inoltre, l'elevato consumo non rispetta le specifiche per applicazioni mobili e/o con alimentazione a batteria.

Secondo il nuovo metodo di lettura qui proposto, mediante un GENERATORE DI TENSIONE 780 si polarizzano simultaneamente ed identicamente le tre celle di riferimento 721, 722 e 723 (FIG. 7) (precedentemente programmate con estrema precisione ai valori di tensione di soglia V_{THREF1} , V_{THREF2} e V_{THREF3}) e la cella di memoria multilivello (MLC) 730. In particolare, si applicano tensioni di lettura costanti ai nodi di drain, source e body (ad es.: $V_D=1.0V$ e $V_S=V_B=0.0V$) mentre, secondo l'invenzione, si applica una tensione V_{READ} crescente nel tempo al nodo comune 740 accoppiato al nodo di gate di controllo CG delle celle 721, 722, 723 e 730. Preferibilmente la tensione V_{READ} è una rampa lineare (eventualmente a gradini, la cui ampiezza e durata sono controllati digitalmente); ad es. in fig. 6A, V_{READ} cresce da 0.0V a 4.5V. Ciascuna cella di memoria è inizialmente "spenta" ($I_{CELL}=0\mu A$) e si "accende" e quindi inizia ad assorbire corrente quando la tensione applicata al rispettivo nodo CG ne supera la tensione di soglia (V_{THREF1} , V_{THREF2} , V_{THREF3} e V_{THCELL}). In virtù della linearità della rampa di tensione V_{READ} , ciò avviene ai tempi $t(V_{THREF1})$, $t(V_{THREF2})$ e $t(V_{THREF3})$ per le tre celle di riferimento 721, 722 e 723, e al tempo $t(V_{THCELL})$ per la cella multilivello MLC 730, rispettivamente (come illustrato in fig.6A, ove è rappresentato il caso in cui $V_{THREF1} < V_{THCELL} < V_{THREF2} < V_{THREF3}$, corrispondente al dato 10).

Durante la lettura, la cella multilivello MLC 730 inizierà ad assorbire corrente prima o dopo ciascuna rispettiva cella di riferimento a seconda che al dato programmato in MLC corrisponda una tensione di soglia inferiore o maggiore di quella della cella di riferimento. I 2 bits vengono ricostruiti in base alla relazione temporale tra gli istanti in cui gli stati conduttivi sono modificati – figure 6 e 7.

Un RIVELATORE DI SOGLIA 750 identifica quando la corrente assorbita dalla corrispondente cella di riferimento 721, 722 e 723 supera un valore predefinito I_0 , ad esempio $10\mu A$ (fig.6A), e fa commutare il segnale L_{REF1} , L_{REF2} e L_{REF3} di rispettivi LATCH da basso ad alto in fig.6B. Nell'UNITA' DI RIFERIMENTO 760 i segnali L_{REF1} , L_{REF2} e L_{REF3} sono combinati per produrre il segnale sul bus BIT (fig.6 C) che, variando nel tempo, rispecchia in ciascun istante in che intervallo sia la lettura e quindi determina il corrispondente valore logico per entrambi i bits.

Contemporaneamente a quanto sopra descritto, viene anche polarizzata la cella MLC 730 e la corrente da essa assorbita è confrontata con la stessa corrente I_0 predefinita (fig.6A) in un altro RIVELATORE DI SOGLIA 755. Al superamento della corrente I_0 il segnale L_{CELL} commuta (fig.6D) e l'UNITA' LOGICA 770 determina, in base al valore attuale di BIT, il dato OUT che può essere fornito in uscita per entrambi i bits. Il metodo può essere esteso anche al caso di $N \neq 2$ bits. In sintesi, modificando le condizioni di polarizzazione del nodo 740 comune ai nodi CG della cella 730 e delle celle di riferimento 721, 722 e 723, se ne modificano i rispettivi stati conduttivi in istanti diversi e si determina il contenuto della cella 740 basandosi su una relazione temporale tra di essi.

1

Il nuovo metodo evita l'uso di tensioni di lettura predefinite e tempistica di campionamento fissa, assicurando la corretta lettura in ogni condizione. Inoltre, "spegnendo" sia le celle di riferimento 721, 722 e 723 che la cella MLC 730 (ad esempio disconnettendone il nodo di DRAIN dal rispettivo RIVELATORE DI SOGLIA 750 o 755 che lo alimenta) subito dopo che il rispettivo stato conduttivo è modificato, cioè quando la corrente supera I_0 (fig.6A), il consumo è notevolmente ridotto.

MEMORIE FLASH MULTILIVELLO (Arte nota)

Le memorie Flash possono immagazzinare uno o più bits di informazione in modo non volatile.

A differenza di un transistor convenzionale 110, la sezione della cui struttura è mostrata in figura 1A, una memoria Flash (nel seguito anche detta "cella di memoria" o semplicemente "cella") è costituita da un transistor a gate flottante 120, mostrato in figura 1B. Mentre nel transistor convenzionale 110 tutti i nodi sono accessibili elettricamente (possono cioè essere polarizzati alla tensione desiderata), nel caso della cella a gate flottante 120, la tensione può essere applicata solo ai nodi di DRAIN, SOURCE, BODY e GATE DI CONTROLLO (CG) ma non al nodo GATE FLOTTANTE (FG), che è completamente circondato da materiale isolante. Il nodo FG è accoppiato capacitivamente soprattutto al nodo CG (i due elettrodi sono separati da un sottile strato dielettrico), e quindi tende a seguirne la tensione; ne risulta che il transistor 120 può essere pilotato (acceso/spento) dal nodo CG.

Le caratteristiche elettriche e le prestazioni di un transistor convenzionale 110, quali ad esempio la tensione di soglia e la trans-caratteristica, sono determinate dai parametri costruttivi (drogaggio del canale, spessore dell'ossido di gate, ecc.).

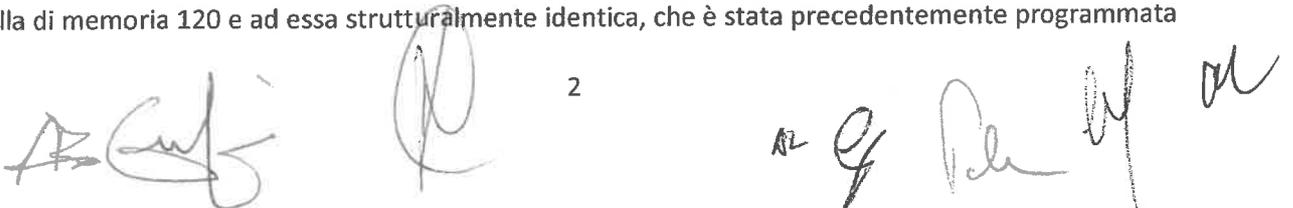
Le caratteristiche elettriche di un transistor a gate flottante 120, invece, oltre che dipendere dai parametri costruttivi, sono anche influenzate dalla eventuale presenza di carica Q_{FG} nel nodo FG. Infatti, la condizione di inversione nel canale (e quindi lo stato di conduzione) dipende dalla tensione effettiva del nodo FG, affacciato al canale stesso, che a sua volta è determinata dalle tensioni agli altri nodi e da Q_{FG} .

Come si diceva sopra, la tensione del nodo FG segue capacitivamente quella del nodo CG. Nella figura 2 è mostrata in linea continua la trans-caratteristica di un transistor a gate flottante 120 (ad esempio quello di fig.1B) in due possibili stati di carica Q_{FG} della gate flottante FG. La curva (a), con tensione di soglia V_{TH_1} , corrisponde ad uno stato di carica nullo ($Q_{FG} = 0$); la curva (b), con tensione di soglia V_{TH_0} , corrisponde ad uno stato di carica negativo ($Q_{FG} = -q < 0$, dovuto ad esempio alla presenza di elettroni nel nodo FG). In sostanza, come illustrato in fig.2, la trans-caratteristica risulta traslata di una quantità ($V_{TH_0} - V_{TH_1}$) che dipende dalla carica Q_{FG} presente nel nodo FG.

Questo effetto è utilizzato per immagazzinare informazione binaria in modo non volatile. Infatti, con riferimento alla figura 2, tenendo a massa i nodi di SOURCE e BODY, polarizzando il nodo di DRAIN, ad esempio a 1.0V, ed applicando al nodo CG-GATE DI CONTROLLO una tensione di lettura V_{READ} , intermedia tra V_{TH_1} e V_{TH_0} , si osserva che una cella 120 con carica $Q_{FG} = 0$ (curva (a) e soglia V_{TH_1}) conduce corrente, mentre la stessa cella 120 con carica $Q_{FG} = -q$ (curva (b) e soglia V_{TH_0}) non ne conduce. E' quindi possibile associare uno stato logico binario il cui valore può essere, ad esempio, "1" se fluisce corrente (curva (a) e $Q_{FG}=0$) e "0" nel caso contrario (curva (b) e $Q_{FG} = -q$).

Per distinguere tra i due stati è possibile confrontare in un COMPARATORE (non mostrato) la corrente di lettura I_{CELL} con una corrente di riferimento I_{REF} opportunamente predefinita (ad es. $10\mu A$) o, meglio, generata mediante una cella di riferimento CR (non mostrata), polarizzata nelle stesse condizioni della cella di memoria 120 e ad essa strutturalmente identica, che è stata precedentemente programmata

2



ad un valore di tensione di soglia V_{TH_REF} intermedio tra V_{TH_1} e V_{TH_0} , e la cui trans-caratteristica è rappresentata dalla linea tratteggiata (c) in figura 2.

E' possibile immagazzinare più di un singolo bit in una sola cella di memoria fisica 120, purché si sia in grado di controllare con grande precisione la quantità di carica Q_{FG} sul nodo FG durante la fase di programmazione. Ciò è fattibile secondo tecniche note, che qui non descriviamo. E' però necessario adottare un metodo di lettura leggermente modificato rispetto a quanto sopra descritto.

La figura 3 illustra la trans-caratteristica della stessa cella di memoria 120 nel caso di quattro diversi valori di carica Q_{FG} (curve a tratto continuo). Le stesse considerazioni fatte rispetto alla figura 2 si applicano anche alla fig.3: la trans-caratteristica identificata da "11" con tensione di soglia V_{TH_11} può risultare da una condizione di carica nulla o sostanzialmente nulla nel nodo FG: $Q_{FG} = 0$; le trans-caratteristiche identificate da "10", "01" e "00", con rispettive tensioni di soglia V_{TH_10} , V_{TH_01} e V_{TH_00} , corrispondono a quantità di carica Q_{FG} negativa via via crescente nel nodo FG e sono quindi corrispondentemente traslate verso destra. Si determinano quindi 4 possibili livelli a ciascuno dei quali è possibile associare un valore logico di 2 bits: 11, 10, 01 e 00, appunto.

Per poter distinguere i 4 livelli, in modo del tutto analogo a quanto descritto con riferimento alla figura 2 per distinguere 2 livelli, si utilizzano 3 celle di riferimento CR1, CR2 e CR3 rispettivamente programmate a valori di tensione di soglia intermedi (non mostrati per semplicità), le cui trans-caratteristiche sono illustrate in figura 3 dalle curve tratteggiate denominate REF1 per la cella CR1, REF2 per la cella CR2 e REF3 per la cella CR3, rispettivamente. Le curve di riferimento REF1, REF2 e REF3 sono sostanzialmente intermedie tra "11" e "10", tra "10" e "01" e tra "01" e "00".

La figura 5 mostra schematicamente un dispositivo di memoria Flash 500 che comprende la cella di memoria multilivello MLC (2 bits per cella) – per chiarezza, in figura è mostrata solo la cella MLC selezionata, ma il dispositivo comprende moltissime altre celle MLC, tipicamente milioni, che però sono selettivamente disconnettibili e quindi elettricamente inattive se non indirizzate.

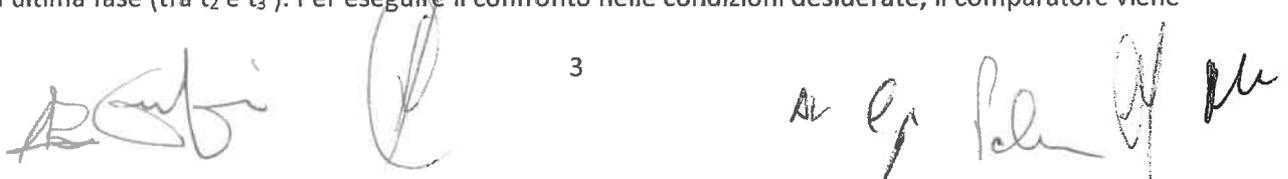
Un GENERATORE DI TENSIONE è connesso al nodo 540 comune, collegato ai rispettivi nodi di GATE DI CONTROLLO delle 3 celle di riferimento CR1, CR2 e CR3, e al quale è (selettivamente) collegabile il nodo CG della cella MLC. Un CIRCUITO DI LETTURA è accoppiato al nodo di DRAIN della cella MLC.

Il metodo di lettura della cella multilivello MLC prevede la polarizzazione, mediante il GENERATORE DI TENSIONE, del nodo comune 540 a ciascuno dei valori di tensione V_{READ1} , V_{READ2} e V_{READ3} , indicati nelle figure 3 e 4. Il nodo di GATE DI CONTROLLO della cella MLC e, simultaneamente, delle celle di riferimento CR1, CR2 e CR3 sono dunque polarizzati alla stessa tensione. I nodi di SOURCE e BODY sono tenuti a massa e il nodo di DRAIN polarizzato a 1.0V sia per MLC che per CR1, CR2 e CR3.

Le tre tensioni di lettura V_{READ1} , V_{READ2} e V_{READ3} sono applicate sequenzialmente secondo una sorta di rampa a gradoni. La figura 4 illustra la sequenza sopra descritta ed in particolare la tensione erogata dal GENERATORE DI TENSIONE (linea continua) e la tensione effettivamente presente ai nodi CG della cella MLC e delle celle CR1, CR2 e CR3 (linea tratteggiata), che evolve con un ritardo dovuto al carico resistivo-capacitivo (RC) del nodo e dell'interconnessione con il GENERATORE DI TENSIONE.

Si identificano 3 fasi: tra t_0 e t_1 (tensione target V_{READ1}), tra t_1 e t_2 (tensione target V_{READ2}) e tra t_2 e t_3 (tensione target V_{READ3}). Solitamente, ma non necessariamente, le tre fasi hanno durata uguale Δt .

Nel CIRCUITO DI LETTURA si confronta la corrente I_{CELL} della cella MLC con la corrente I_{REF} di una delle celle di riferimento, in particolare con I_{REF1} della cella CR1 durante la prima fase (ovvero, tra t_0 e t_1), con I_{REF2} della cella CR2 durante la seconda fase (tra t_1 e t_2) e con I_{REF3} della cella CR3 durante la terza ed ultima fase (tra t_2 e t_3). Per eseguire il confronto nelle condizioni desiderate, il comparatore viene



attivato da un segnale ENABLE appositamente generato al termine di ciascuna fase, quando il valore della tensione di lettura è ormai stabile. Il CIRCUITO DI LETTURA determina (mediante elementi non mostrati) il valore OUT per i due bits in uscita considerando il risultato al termine di ciascuna fase.

Si richiede al candidato di redigere una serie di rivendicazioni idonee a tutelare nel modo più efficace il trovato sopra descritto.

R. Conti

P

Al. G. Pel

Al
Al

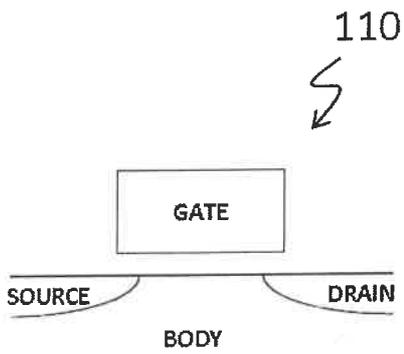


FIGURA 1A

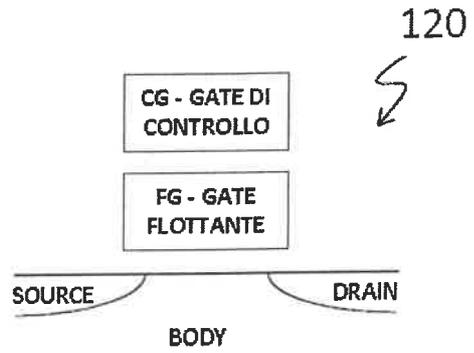


FIGURA 1B

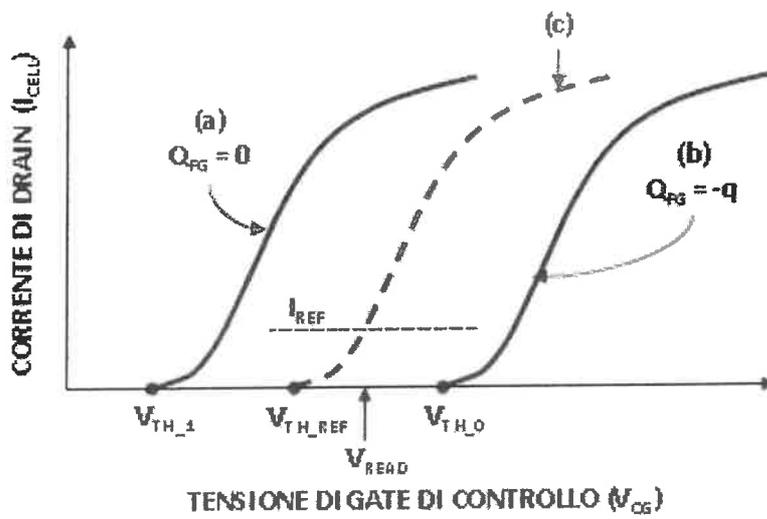


FIGURA 2

B. Conforti

AL G. Pal...

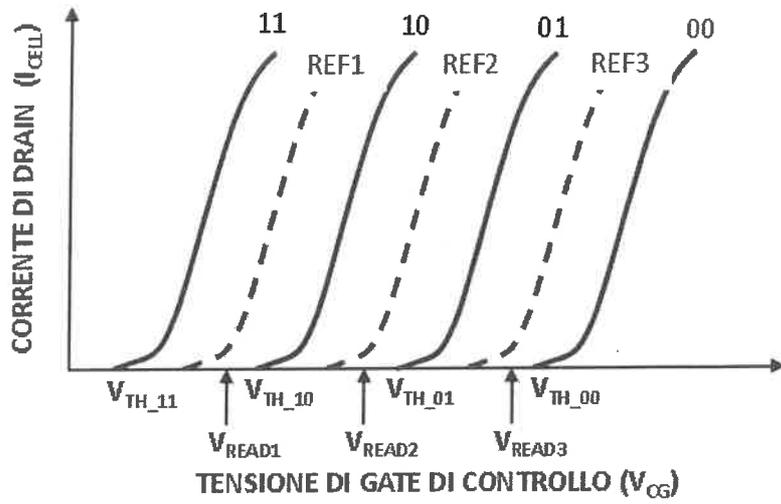


FIGURA 3

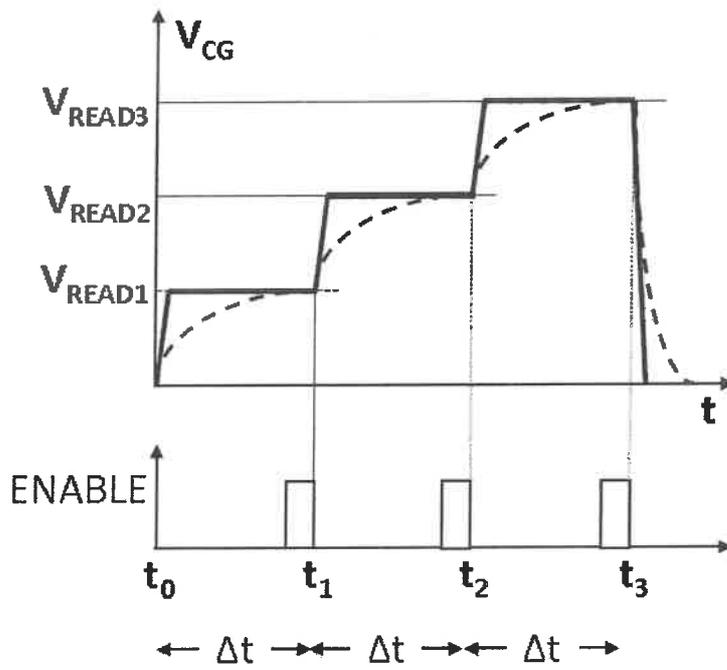


FIGURA 4

Handwritten signature

Handwritten mark

Handwritten signature

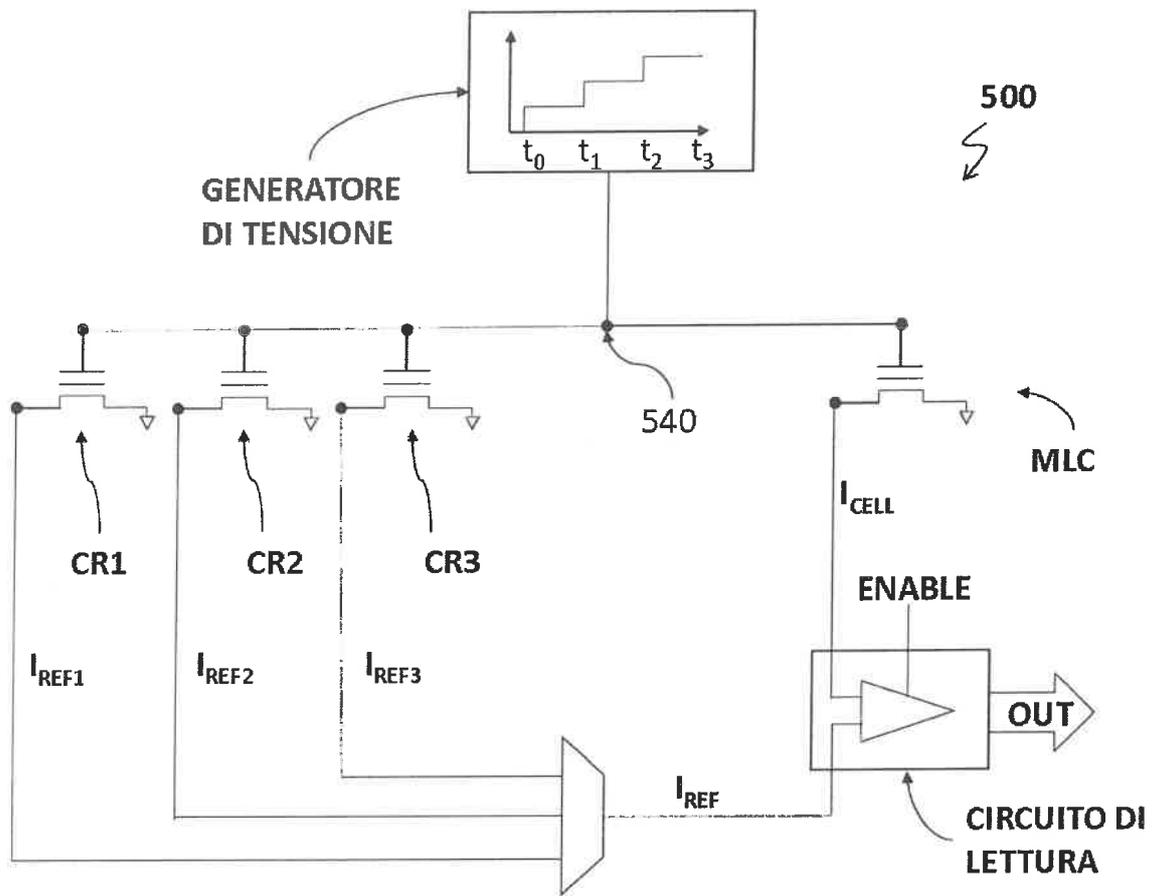


FIGURA 5

A. Saffari

g. m. m.

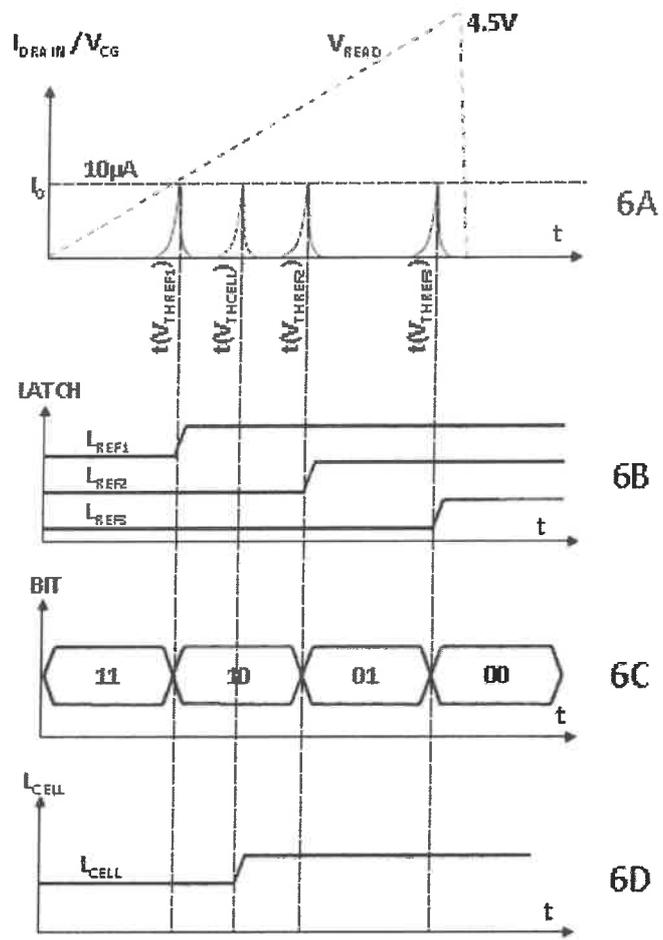


FIGURA 6

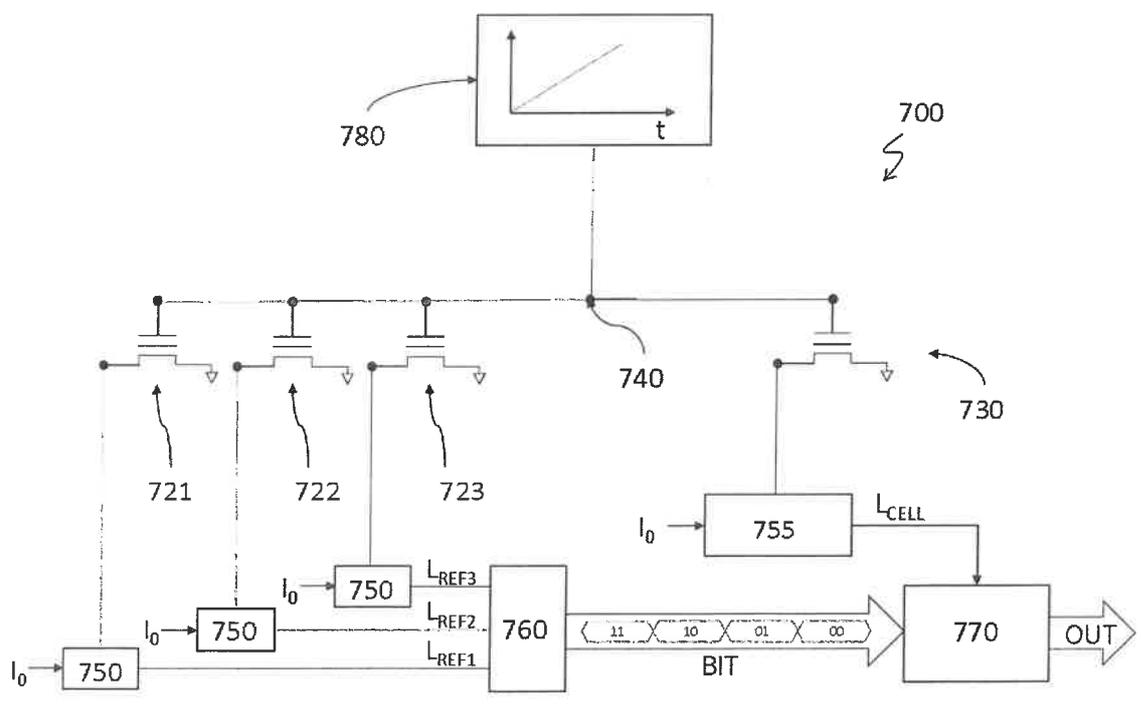


FIGURA 7

Handwritten signatures and notes:
 AL of [unclear] [unclear] [unclear]