

Le memorie NAND Flash, facciamo il punto ...



LINK (<https://www.nexthardware.com/focus/ram-memorie-flash/183/le-memorie-nand-flash-facciamo-il-punto-.htm>)

Cosa sono e come vengono utilizzate nei moderni SSD.

Introduzione

Per chi possiede un minimo di conoscenze di elettronica e di informatica, la prima grande differenza che intercorre tra le memorie Flash di tipo NAND più comunemente usate, SLC, MLC e TLC, risulterà intuibile direttamente dalla definizione dei relativi acronimi:

- **SLC** - *Single Level Cell* (1 bit per cella, 2 livelli di tensione)
- **MLC** - *Multi Level Cell* ↔ (2 bit per cella, 4 livelli di tensione)
- **TLC** - *Triple Level Cell* ↔ (3 bit per cella, 8 livelli di tensione)

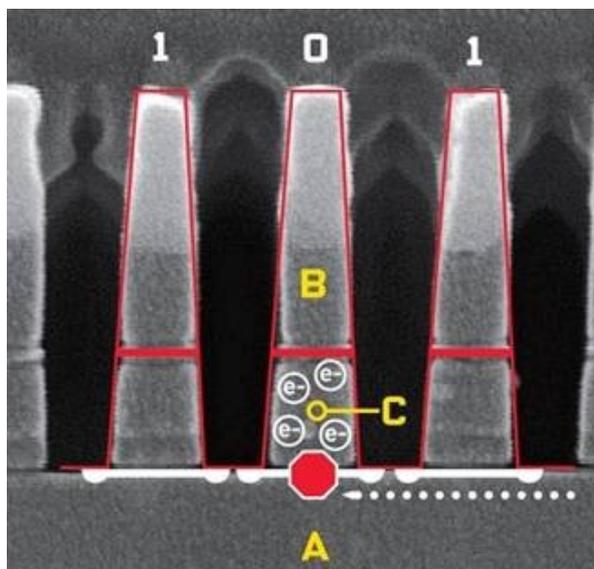


Figura 1: celle NAND 110.000x al microscopio elettronico (fonte Micron)

A (substrato) **B** (gate di controllo) **C** (gate flottante)

Una singola cella di memoria NAND Flash rappresenta di fatto l'unità elementare di memorizzazione nelle moderne unità di *storage* allo stato solido. E' costituita essenzialmente da un transistor relativamente particolare ad effetto di campo (**FET**) in tecnologia **MOS** (*Metal Oxide Semiconductor*), in quanto presenta la caratteristica di esser dotato di due **terminali di attivazione** (*gate*).

A partire dagli elementi costitutivi di un MOSFET è infatti previsto un secondo *gate*, di tipo flottante (*floating gate*), sottostante al gate di controllo (*control gate*): per questo motivo il transistor venne in origine definito **FGMOS** (*Floating Gate MOS*).

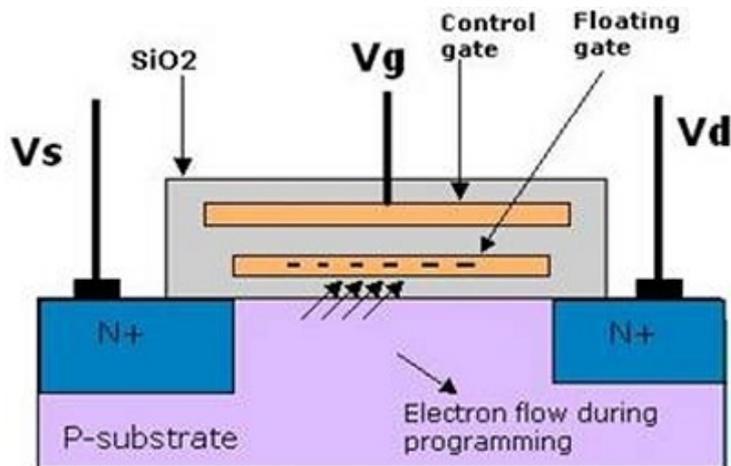


Figura 2: grafico di una tipica cella di memoria NAND Flash

Il *terminale di attivazione* aggiuntivo flottante viene usato per collegare le wordline al gate di controllo e costituisce uno spazio ben definito all'interno del transistor: ha capacità conduttive, è in genere di tipo *poli-siliconico* ed è in grado di trattenere della carica elettrica, riuscendo a *mantenerla inalterata nel tempo*.

Questa caratteristica riassume in sintesi la più essenziale delle proprietà delle memorie NAND Flash: la *non-volatilità*.

Il lasso di tempo per cui questa viene garantita, priva di apporto esterno di energia, va a definire il valore di *Data Retention Period*, in genere dell'ordine di uno o più anni: questo parametro è considerato, a detta dei più, fondamentale in quanto denoterebbe in qualche modo la qualità della tecnologia MOS utilizzata.

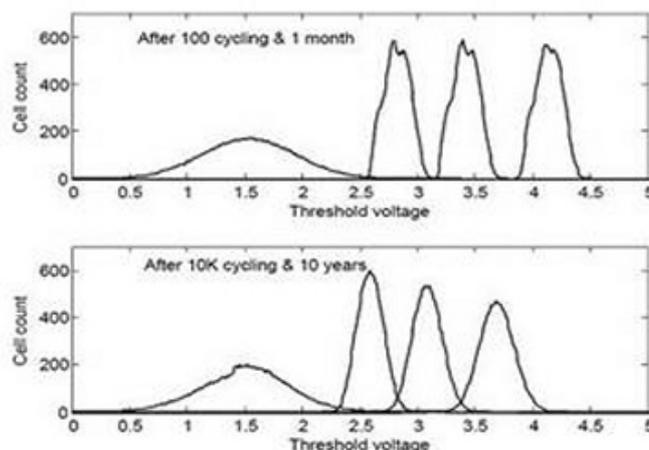


Figura 3: grafico della perdita media di efficienza con l'uso e il passare del tempo

Dal punto di vista dell'organizzazione a livello fisico, un *die* è tipicamente suddiviso in pagine da **4kB** agglomerate in 1280 blocchi, in modo da formare un singolo insieme uniforme di 512MB: a livello di pagina, nelle memorie NAND Flash possono avvenire un numero illimitato di letture senza che questo fatto

ne possa pregiudicare in alcun modo la durata.

Anche al momento della prima scrittura di una cella, sostanzialmente a drive giunto con impostazioni di fabbrica o a seguito di una procedura di *secure erase*, il procedimento di memorizzazione non comporta alcuna penalizzazione nella durata delle pagine interessate.

Per descrivere le implicazioni inerenti il procedimento finalizzato invece alla scrittura in esercizio, valevole per tutte le famiglie di memorie NAND, bisogna precisare che per una qualsiasi successiva *risrittura* una cella deve essere prima di tutto bloccata e cancellata, o meglio in qualche modo *svuotata* della carica precedente (che, logicamente, rappresenta delle ben precise *informazioni*).

Questa operazione, definita *tunnel erase*, è attivata al momento che una tensione intermedia viene applicata al *gate* di controllo: l'energia applicata dà modo alla corrente di poter liberamente fluire, permettendo agli elettroni già presenti nel *floating gate* di esser rilasciati e svuotare, così, il gate flottante.

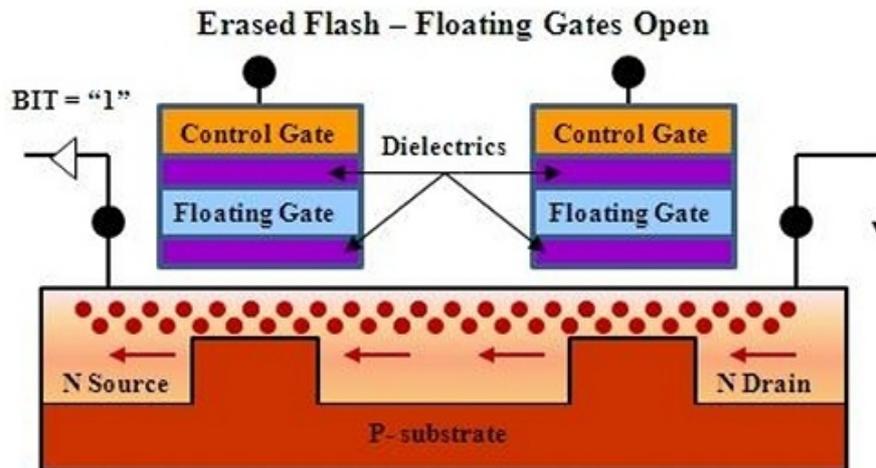


Figura 4: fase di cancellazione di una cella NAND

Successivamente la cella viene sbloccata ed è pronta per esser nuovamente utilizzata, ovvero *riscritta* tramite un procedimento di programmazione detto *tunnel injection*.

Esso consiste, qualora si volesse imporre il valore logico "0", binario, nell'applicazione di un potenziale di carica positiva tramite il quale una quantità di elettroni, di carica negativa, vengono trasferiti (*iniettati*) attraverso gli strati dielettrici del transistor, per esser infine trattenuti all'interno del *floating gate*.

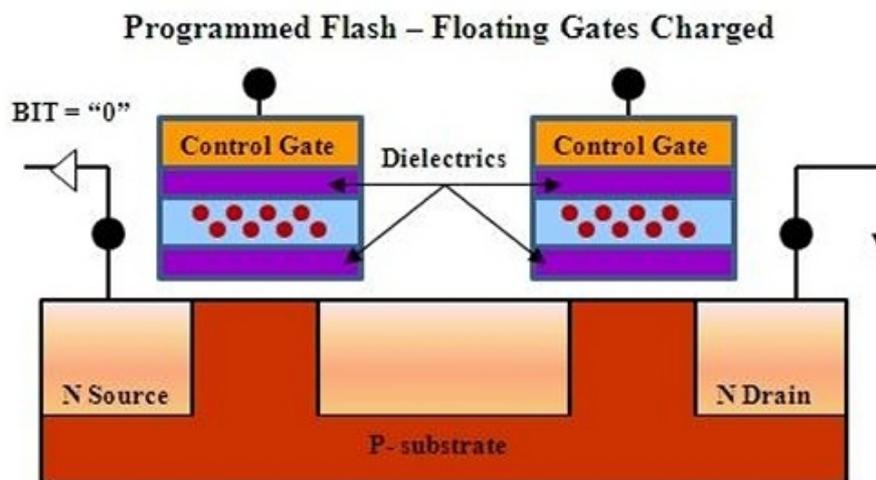


Figura 5: fase di programmazione di una cella NAND

L'altra alternativa possibile, quella di identificare cioè il valore "1", è circostanziata dalla *non emissione* di alcuna nuova carica, lasciando pertanto *vuota di elettroni* la cella in precedenza cancellata.

Per inciso, talvolta si può anche leggere in alcuni articoli o su pagine web che la corrispondenza tra gli stati e il valore di tensione viene descritta con modalità invertita, ovvero con lo "0" a significare uno stato privo di carica, ma è solo una descrizione a livello puramente teorico.

Bisogna, anche in questo caso, approfondire il fondamentale concetto di assenza di carica susseguente alle operazioni di cancellazione, puntualizzando quanto in realtà avviene: sebbene si parli comunemente del termine *assenza*, risulta in realtà molto complesso azzerare la tensione presente all'interno del *floating gate*.

La carica elettrica non può essere completamente eliminata, in quanto uno o più elettroni tendono in genere a rimanere naturalmente imprigionati all'interno dello strato di ossido, producendo così un seppur minimo livello di tensione.

Questo fatto non pregiudica le corrette funzionalità delle memorie NAND Flash in quanto tale carica permane ad un livello ben inferiore a quello del primo valore minimo di soglia, tale quindi da non divenire influente nelle successive rilevazioni durante le attività di lettura: almeno finché la cella continuerà a trovarsi, come vedremo, in stato di efficienza.

La sequenza di scritture, cancellazioni e successive riscritture, in continua alternanza sulle pagine costituite dalle celle, è detto nelle memorie NAND ciclo di *program/erase*.

Il valore accreditato dal produttore alle proprie celle di memoria è da considerare pertanto di vitale importanza ed utile quale criterio per stabilire e quantificare la durata del dispositivo di storage che le adotta.

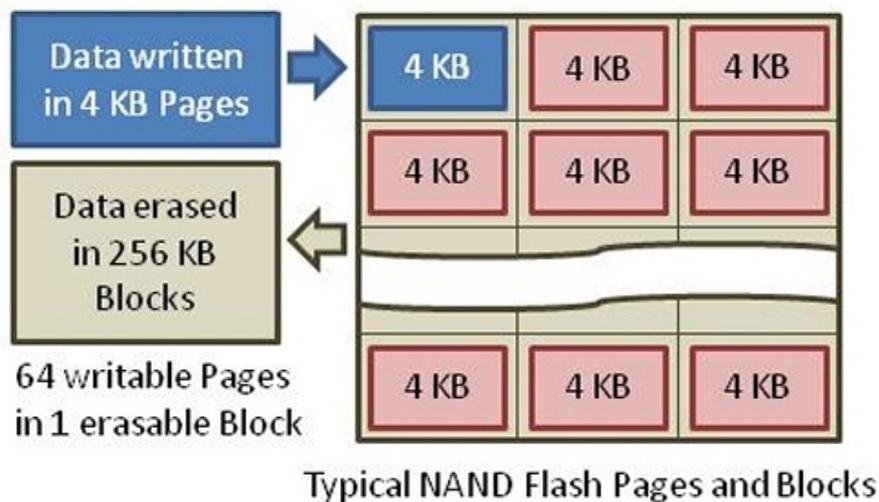


Figura 6: programmazione/cancellazione a pagine e blocchi di celle

Per denotare in qualche modo la complessità funzionale di queste memorie, bisogna chiarire che l'operazione di *programmazione/cancellazione*, seppur destinata in origine ad un'unica serie di *celle/pagina*, non può essere eseguita in modalità diretta solo su queste: è necessario infatti procedere alla preventiva cancellazione di un intero *blocco cumulativo* di pagine da *4kB* contigue, in genere in numero di **64**.

Stabilito che, spesso, all'interno del blocco di pagine interessato potrebbero essere già memorizzati altri dati non inerenti l'operazione in esecuzione, è facile intuire come questo fatto comporti spesso anche la necessità dello spostamento di una quantità di dati (*256kB*) non tutti sostanzialmente interessati alle operazioni originate dal sistema operativo.

Il controller dell'unità SSD, che in queste circostanze attiva contestualmente le cosiddette strategie di *wear leveling* per salvaguardare la durata dell'unità nel tempo, procederà quindi nelle attività tramite due differenti operazioni: la prima, rappresentata dalla scelta di un nuovo blocco di pagine vuote di destinazione (cancellate già in precedenza) dove memorizzare i dati, la seconda comprendente sia la *scrittura* dei dati modificati nonché la *riscrittura* vera e propria (*copia*) del contenuto *immutato* del blocco appena cancellato.

Oltre al concreto handicap a livello di spazio di memorizzazione che non può, in questo modo, essere utilizzato nel lungo periodo in maniera del tutto efficiente, tutta questa serie di operazioni supplementari richiedono ovviamente anche un tempo ampiamente superiore all'ideale necessità di modifica dei dati originari.

All'esordio di questa tecnologia le prestazioni globali in scrittura rappresentavano, pertanto, un consistente limite, soprattutto nelle operazioni *random*, il tempo necessario alla copia tanto inevitabile quanto imprevedibile della serie di blocchi costituiva e costituisce ancora il fattore più limitante in assoluto, in quanto al crescere di questo intervallo vanno a diminuire in misura diretta le performance.

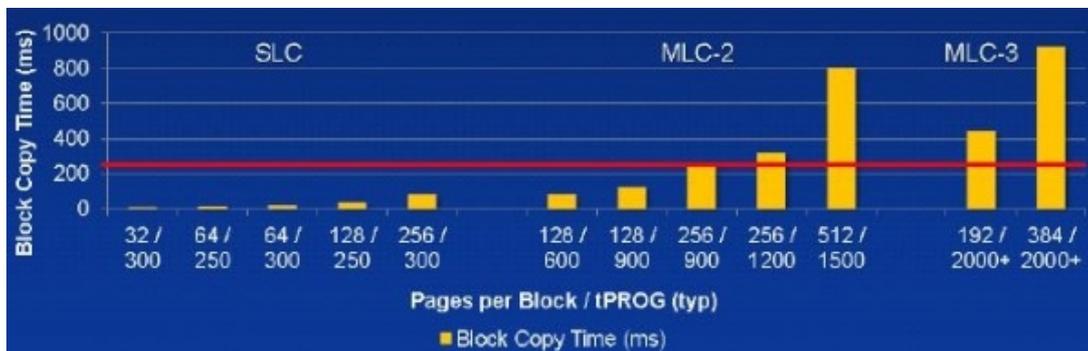


Figura 7: dilatazione delle temporizzazioni di copia dei blocchi in relazione all'innovazione tecnologica

Andando a riassumere sinteticamente le circostanze particolarmente negative di questa tecnologia, i fattori chiave che vanno ad impattare le performance in scrittura sono:

- l'elevato numero di pagine per blocco;
- l'aumentare del tempo necessario alla programmazione;
- la dilatazione delle temporizzazioni nei trasferimenti di I/O.

Questi limiti, uniti ad una iniziale non brillantezza nelle operazioni atomiche di scrittura delle NAND, furono in origine considerati fattori troppo penalizzanti nell'uso di questa tecnologia per pensare di poter soppiantare del tutto i dispositivi di storage, all'epoca più largamente utilizzati, a tecnologia meccanico-magnetica.

1. SLC, MCL, TLC: quali le differenze?

1. SLC, MCL, TLC: quali le differenze?

Passando a fare una rapida carrellata sulle diverse tecnologie sviluppate nel tempo, prenderemo a spunto, d'ora in avanti, le temporizzazioni tipiche delle NAND Flash ottenute tramite processo produttivo dell'ordine dei "2nm", dando modo di paragonare in maniera sintetica quanto accade nell'uso delle varie tipologia di memorie e, contemporaneamente, indicando anche i riferimenti sui tempi medi necessari alle complete attività di lettura, scrittura e cancellazione.

SLC

Le memorie NAND Flash SLC, le prime ad esser state utilizzate, adottano tipicamente una corrispondenza diretta tra la *presenza/assenza* di carica nel *floating gate* ed i valori binari "0" o "1", a livello logico assegnati.

In queste memorie, al momento in cui è necessario prelevare il contenuto della cella dovrà esser valutata l'esistenza, o meno, di una certa carica elettrica: l'operazione completa comporta in genere un tempo di lettura di $\sim 25\text{micro/s}$ e la circostanza potrà condurre a determinare uno solo dei due valori digitali.

SLC: 1 bit/cell
 Level 0 = Erased
 Level 1 = Programmed

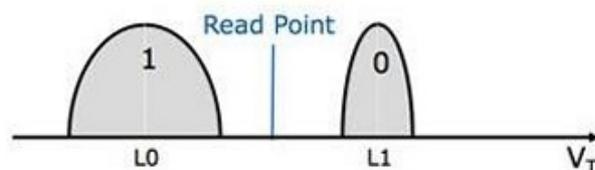


Figura 8: sistema funzionale a 1bit/2 livelli di tensione

Le operazioni di scrittura necessiteranno la generazione di un unico e ben definito livello di tensione e saranno eseguite in un tempo medio di $\sim 250\text{micro/s}$, mentre per quelle di cancellazione dei blocchi saranno necessari ben $\sim 1.5\text{milli/s}$.

Questo tipo di tecnologia a *singolo bit*, pur caratterizzandosi in eccellenti prestazioni ed in una invidiabile longevità operativa, non consente d'altro canto di produrre unità allo stato solido di elevata capacità a prezzi comprensibilmente accettabili a livello *SOHO*.

Per questo motivo, un SSD che utilizza memorie NAND SLC ha la condizione essenziale di dover disporre di quantitativi davvero elevati di celle di memoria: tale circostanza, di conseguenza, diviene economicamente sconveniente relegandone l'uso in ambito Enterprise su server *mission critical*, potenti workstation professionali o *appliance* di storage di alto profilo, tutti estremamente costosi.

MLC

La prima evoluzione verso uno scenario maggiormente produttivo è stata quella dell'adozione di una più avanzata tecnologia in grado di mettere a disposizione una nuova generazione di celle polivalenti, congegnate alla base sull'idea di poter accettare differenti valori di carica o *multilivello*.

Questo nuovo tipo di transistor è pur sempre costruito a partire da dispositivi a semiconduttore identici ai precedenti, ma l'idea funzionalmente alla base è quella di rendere possibile, a livello logico, la memorizzazione al proprio interno di una quantità raddoppiata di informazioni.

Si amplia, così, il concetto d'uso delle celle NAND tramite l'implementazione di una serie equidistante di valori di soglia da programmare nel *floating gate*: ad ognuno di questi valori (compreso lo stato di *assenza di carica*) è possibile associare a livello logico uno dei quattro valori digitali, "00" "01" "10" e "11", utilizzabili tramite due *bit*.

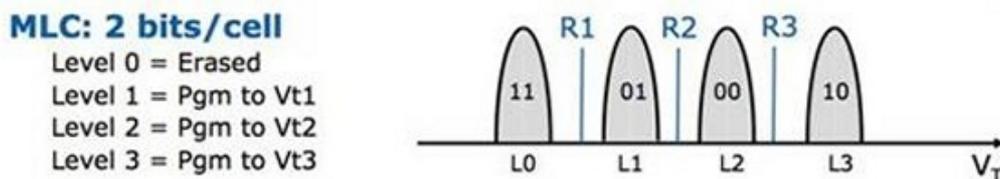


Figura 9: sistema funzionale a 2bit/4 livelli di tensione

Queste non sono solo inerenti il concreto aumento della velocità di esecuzione dei comandi in *scrittura* e *lettura* provenienti dal controller, ma vanno a produrre purtroppo anche una sensibile dilatazione dei tempi in fase di *cancellazione*, sebbene con tempi meno marcati in percentuale.

In fase di programmazione della cella queste latenze intervengono a causa dell'incremento del lasso di tempo necessario alla logica di controllo per determinare il corretto livello di tensione, tra i quattro possibili da poter applicare: livello che andrà poi a determinare, logicamente, la relativa coppia di *bit*.

Quest'attività porta il tempo di scrittura mediamente a quadruplicarsi rispetto all'equivalente scenario delle SLC, attestandosi sui $\sim 900 \text{micro/s}$, mentre le operazioni di cancellazione dei blocchi di pagine salgono a $\sim 3 \text{milli/s}$.

Sul fronte della lettura delle informazioni un'ulteriore sfida nell'uso delle celle *multilivello* è rappresentata, allo stato dell'attuale tecnologia di misurazione, dalla difficoltà di estrapolare in maniera diretta, con una singola verifica, il valore del livello di carica contenuto nella cella.

Tale circostanza operativa costituisce di fatto una limitazione aggiuntiva che rende, in assoluto, meno performante anche la lettura: il tempo necessario va infatti a raddoppiare rispetto alle migliori NAND SLC, giungendo a temporizzazioni dell'ordine di $\sim 50 \text{micro/s}$.

Queste strategie consistono, in sintesi, nell'assumere preventivamente un valore ben definito di tensione, prenderlo come riferimento e valutare, tramite una serie di veloci comparazioni sulla base di algoritmi specifici, la variazione del livello di carica dai singoli valori di soglia, permettendo di contenere le operazioni in genere a sole due verifiche.

↔

TLC

A seguito dell'affinamento delle tecnologie produttive e le superiori caratteristiche delle più recenti celle di memoria NAND Flash, i produttori hanno pensato di ampliare il concetto alla base dell'idea multilivello, sviluppando una tecnologia in grado di impiegare otto distinte soglie di tensione.

In questo modo, potendo gestire l'uso di una corrispondenza digitale di otto *bit*, è possibile usufruire di altrettante permutazioni su base binaria e poter singolarmente memorizzare uno dei valori "000", "001", "010", "011", "100", "101", "110" o "111", con un evidente aumento dell'efficienza del dispositivo elettronico originario.

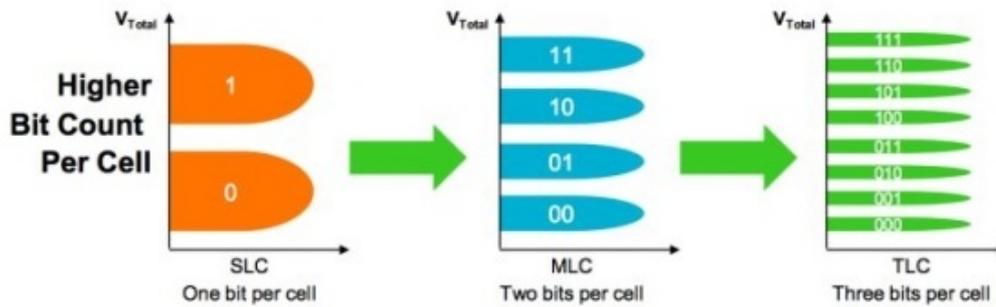


Figura 10: confronto tra i livelli funzionali delle varie tipologie di NAND

In misura del tutto analoga ai concetti espressi in precedenza per le MLC, il discorso diviene nelle TLC ancora più complesso, e penalizzante, al momento di attivare le operazioni di scrittura: alla luce del numero lievitato dei livelli di carica, deve esser gioco forza considerato un procedimento di maggior precisione ed affidabilità nella generazione del corretto valore di tensione (*programmazione*), il quale prevede ora un numero di soglie divenute anch'esse otto, utili a determinare a livello logico il contenuto, da tre bit, del dato da memorizzare.

Diventa pertanto inevitabile che le operazioni di scrittura sulle NAND TLC arrivino a risultare più lente in genere del 50% rispetto alle MLC, portando le temporizzazioni a salire a $\sim 1300\text{micro/s}$, mentre per la cancellazione dei blocchi sono necessari $\sim 4.5\text{milli/s}$.

Altro aspetto non secondario da tenere in considerazione è quello della diminuzione del *delta* di carica tra i distinti valori di tensione in gioco da attribuire alle nuovo numero complessivo di soglie.

La diminuita differenza di tensione tra i valori di soglia contigui, fa inoltre sì che le operazioni di programmazione possano anche divenire soggette ad una serie ulteriore di disturbi del segnale all'insorgere delle *correnti di leakage*, soprattutto nei modelli che utilizzano chip con più ridotte geometrie costruttive: la coesistenza di queste circostanze determina, tendenzialmente, un numero maggiore di probabilità della rilevazione di errori nelle successive fasi di lettura.

The Drawbacks of NAND Scaling: Decreasing Endurance, Increasing ECC

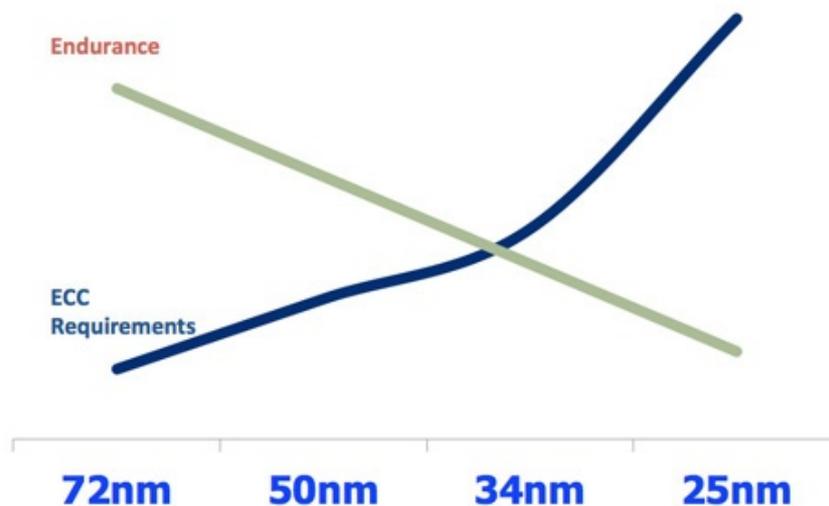


Figura 11: durata delle celle e quantità dei dati ECC sono inversamente proporzionali (fonte LSI)

Per la gestione di questo handicap si rende necessaria, di concerto, l'adozione di algoritmi di correzione di errore più complessi: questi, a loro volta, portano non solo all'esigenza di spazi di memorizzazione

aggiuntivi ma, sostanzialmente, inducono al momento della loro elaborazione latenze aggiuntive nonché la necessità, nel controller, di una capacità di calcolo numerico più rilevante.

Così come già riferito per le Flash NAND MLC, bisogna in ultimo sottolineare che in fase di lettura non è possibile neanche nelle TLC verificare il valore di carica presente nel *floating gate* tramite rilevazione diretta.

Per mezzo di diverse strategie e mediante l'uso di specifici algoritmi si riesce, nelle migliori delle circostanze, a rendere in qualche modo più efficienti le operazioni di rilevazione del valore di tensione nel gate flottante, tramite una serie di misurazioni, in genere quattro, che portano la latenza in lettura a divenire sostanzialmente non inferiore $\sim 75\text{micro/s}$.

2. 3D Vertical NAND

2. 3D Vertical NAND

Procedendo in questi ultimi anni nello sviluppo di soluzioni più efficienti, due sono state le grandi problematiche riscontrate dai produttori nel tentativo di procedere nella naturale evoluzione delle NAND Flash, una volta raggiunta una tecnologia costruttiva estremamente ridotta ($2xnm$).

La prima è rappresentata dalle grandi difficoltà nel raggiungimento di uno *scaling* ancora più ridotto ma contestualmente efficiente nelle due dimensioni, lunghezza e larghezza, della superficie del die.

Mentre per quanto riguarda la seconda difficoltà, il limite è invece costituito dal numero troppo esiguo di elettroni gestibili, a quelle dimensioni, per mantenere coerenti ed efficaci le funzionalità di una cella di memoria NAND.

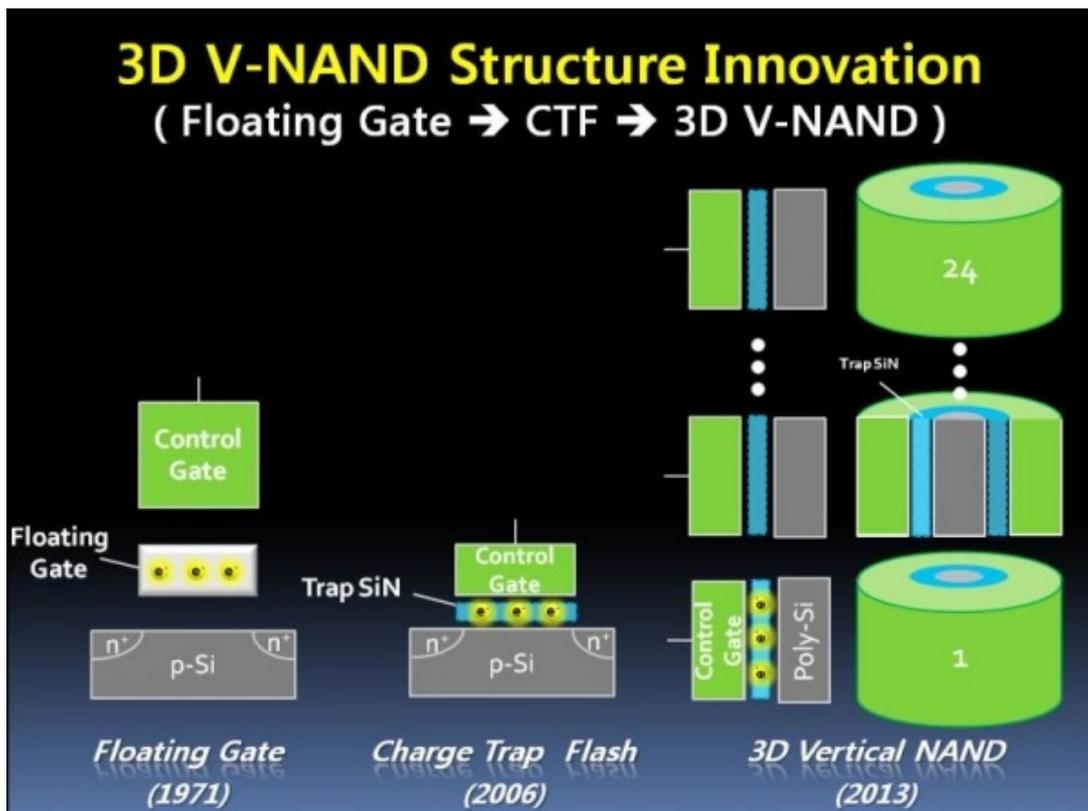


Figura 12: evoluzione tecnologica delle celle NAND (fonte Samsung)

Partendo da questi limiti e alla ricerca di una valida ed efficace soluzione, Toshiba giunse ad inventare e sviluppare qualche tempo fa, le prime NAND 3D: questa evoluzione è stata così innovativa e promettente nel settore, da essere già nell'immediatezza identificata come la strada giusta da percorrere nel futuro, anche da altri grandi protagonisti del settore come Samsung.

L'idea concettualmente alla base è davvero semplice ma allo stesso tempo geniale: a partire dall'ipotetica e lunga *stringa* longitudinale che ospita le celle di memoria Flash NAND standard, è utile creare uno spazio privo di celle al centro, in modo da dividere la stringa in due rami di ugual misura con una giusta base di appoggio.

A questo punto occorre piegare un ramo verso l'altro fino a renderli paralleli e dislocarli in maniera verticale.

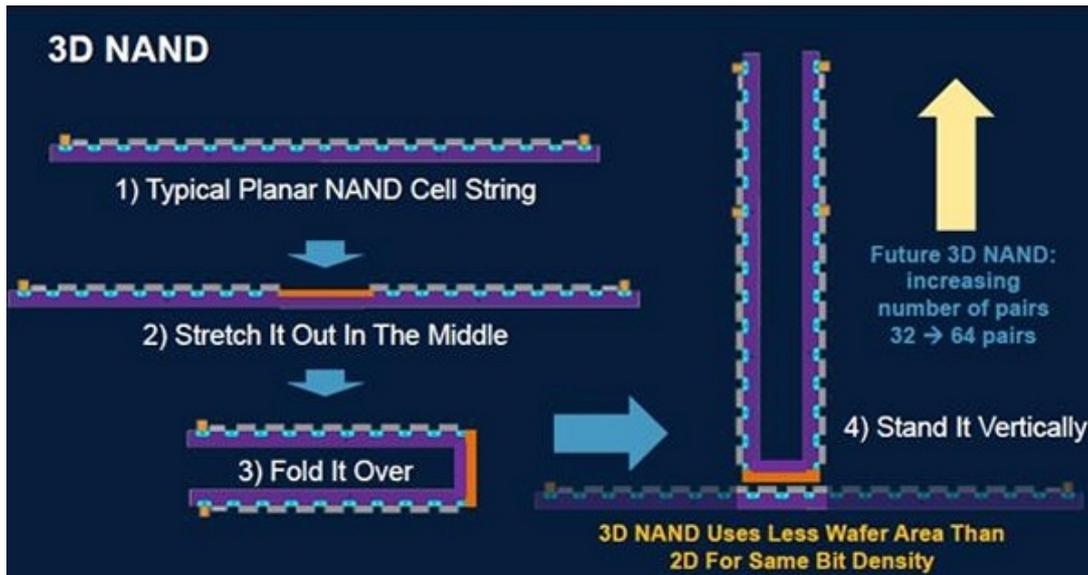


Figura 13: ideazione e sviluppo di una cella NAND 3D Vertical (fonte Samsung)

Con lo stesso procedimento, vengono così disposte tutta una serie di *doppie stringhe* ottenute in modo analogo: queste possono essere infine impilate alla base su una nuova stringa principale, planare, che ne estende le connessioni con le relative celle e soprattutto in grado di prenderne a livello elettrico il controllo.

Gli ingegneri Toshiba pensarono bene di ideare una duplice strategia: una prima fu adottata per l'implementazione della stringa planare, del tutto semplificata rispetto al passato, e una seconda, completamente differente, fu destinata a quella verticale.

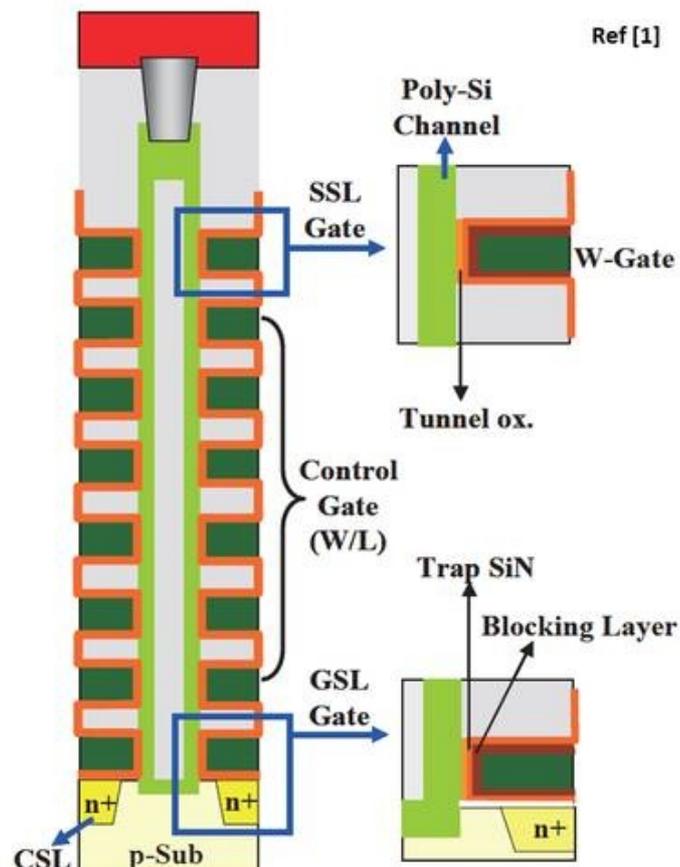


Figura 14: schema dello stack di una cella 3D Vertical NAND (fonte Samsung)

Bisogna innanzitutto chiarire che il processo costruttivo col quale vengono prodotte le *stringhe verticali*, con nodi dalle geometrie di $\sim 40\text{nm}$, per i motivi a cui si è accennato non può essere il medesimo usato comunemente per quelle planari.

Ma soprattutto è in realtà impossibile *ripiegare* una stringa di celle NAND, per come la conosciamo, senza pregiudicare a livello funzionale le caratteristiche tipiche dei transistor a doppio *gate* in essa contenuti: sotto l'aspetto della dislocazione verticale si erano concentrate purtroppo alcune difficoltà mai affrontate prima.

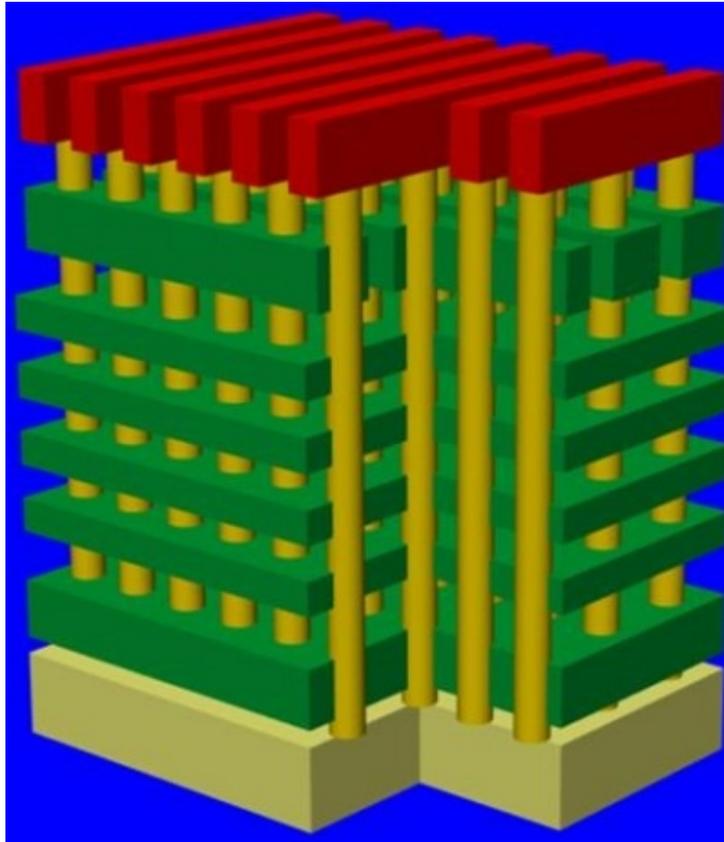


Figura 15: rappresentazione tridimensionale di un array V-NAND

In sostanza tali difficoltà derivano da un assoluto punto fermo, l'impossibilità cioè di usare per questa soluzione il procedimento litografico standard: la luce infatti non può, tramite la litografia, essere gestita per costituire pattern a sviluppo verticale, pertanto l'idea di mutuare una sorta di duplicazione degli ormai collaudati *gate flottanti*, non poteva in alcun modo essere presa in considerazione.

Così gli ingegneri stabilirono che sulla stringa planare 3D NAND venissero inseriti i soli *gate di controllo* mentre verticalmente, al posto dei vecchi *flottanti*, decisero invece di posizionare dei transistor composti con materiali differenti: i *charge trap* in Nitrato di Silicio.

La differenza tra i vecchi e i nuovi è sostanziale in quanto, contrariamente ai primi, questi ultimi sono costituiti da *layer non conduttivi*, ma ugualmente in grado di trattenere e mantenere nel tempo una carica elettrica.

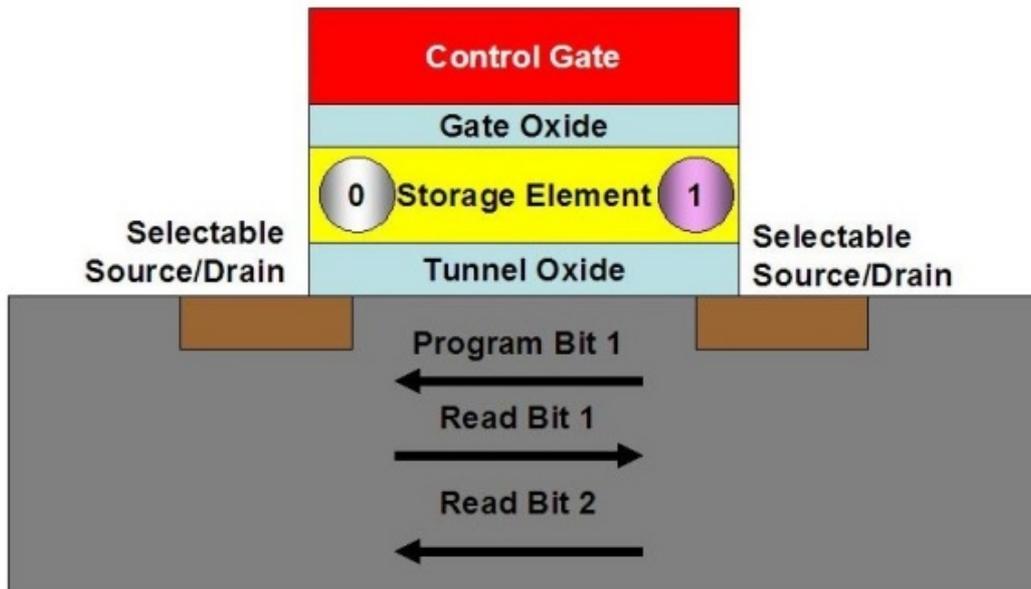


Figura 16: schema di una cella NAND 1-bit con charge trap

Analogamente a quanto già indicato in precedenza riguardo la differenza tra *gate flottanti* a singolo o *multilivello*, anche nel caso dei *charge trap* non cambia sostanzialmente il discorso: ciò sta a significare che all'origine di questa tecnologia era previsto l'uso di solo un livello di tensione mentre, allo stato attuale, i *charge trap* adottati nelle 3D Vertical sono tipicamente *multilivello* a quattro tensioni e permettono quindi di combinare l'uso di due bit.

Il necessario strato destinato al trasporto dei segnali viene pertanto costituito da un *foglio* non conduttivo posto planarmente e ricoperto in superficie da uno strato conduttivo *polisiliconico*.

Il principio costruttivo delle 3D V-NAND prevede la sovrapposizione di svariati strati costituiti fondamentalmente da Ossido di Silicio, isolati l'un l'altro da relativi fogli *non conduttivi*.

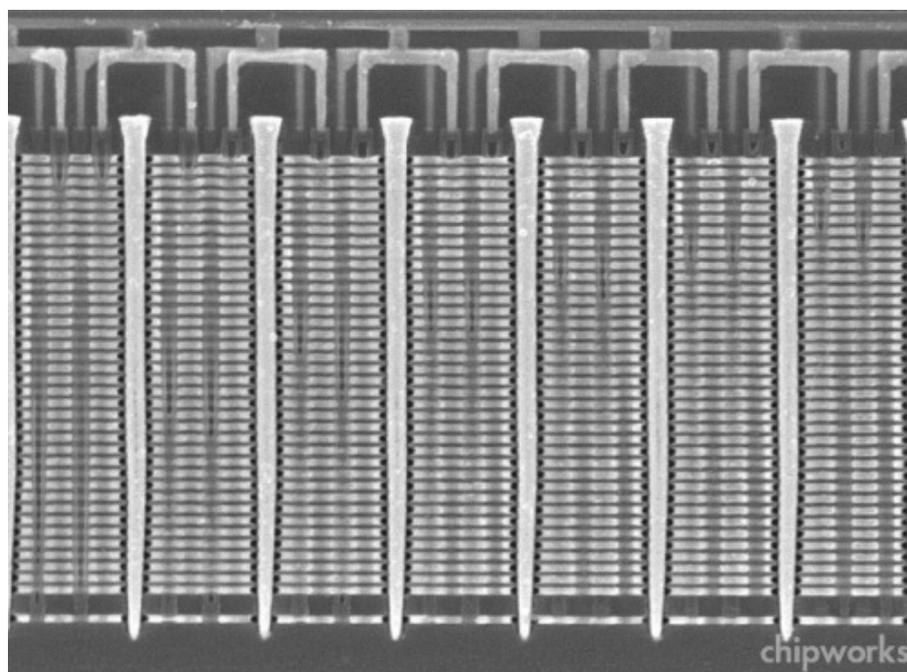


Figura 17: un array di Flash V-NAND al microscopio elettronico (fonte Chipworks)

Per implementare questa soluzione gli ingegneri rispolverarono una tecnica d'avanguardia, adottata oltre una decina di anni prima da Fujitsu e qualche anno dopo applicata in più larga scala su particolari memorie Flash NOR: le cosiddette **MirrorBit** del produttore americano di semiconduttori **Spansion**.

Questa tecnologia prevede che i *gate* di controllo sulla stringa planare siano interconnessi ai *charge trap* tramite un opportuno canale di comunicazione, il quale segue in questo modo uno sviluppo verticale: sia

Toshiba che Samsung per raggiungere 128Gbit nella prima generazione di 3D V-NAND utilizzarono un pattern verticale da 24 *strati*, divenuti 32 nell'attuale seconda generazione.

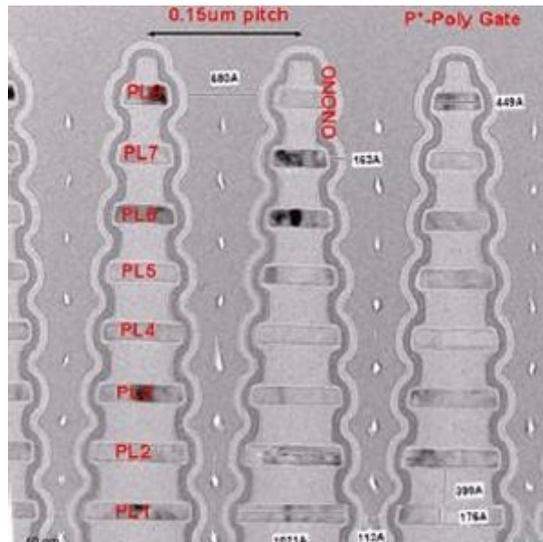


Figura 18: una serie di V-NAND al microscopio elettronico (fonte Spansion)

Per quanto riguarda la seconda generazione del produttore coreano, quella attualmente in commercio, a fronte di un corposo incremento del 33% utile ad incrementare la densità di memorizzazione, sembrerebbe invece che questi ulteriori *layer* possibilmente non siano stati usati come spazio di memorizzazione aggiuntivo, ma che Samsung abbia voluto mantenere la densità di 128Gbit scegliendo di adottare un'impronta planare più snella, in modo da liberare in qualche modo le *wordline* verticali dall'insorgere di indesiderate interferenze elettriche che con questa tecnologia possono instaurarsi tra celle contigue.

3. Quale efficienza e durata per le NAND Flash?

3. Quale efficienza e durata per le NAND Flash?

A rendere, sulla carta, in qualche modo più complesse le valutazioni sulla scelta di soluzioni MLC e TLC sviluppati con geometrie costruttive più ridotte, bisogna anche considerare la circostanza negativa che interessa la riduzione, nel tempo, della durata di esercizio dei cicli *program/erase*, in conseguenza diretta alle innumerevoli operazioni di cancellazione e scrittura su vasti numeri di celle.

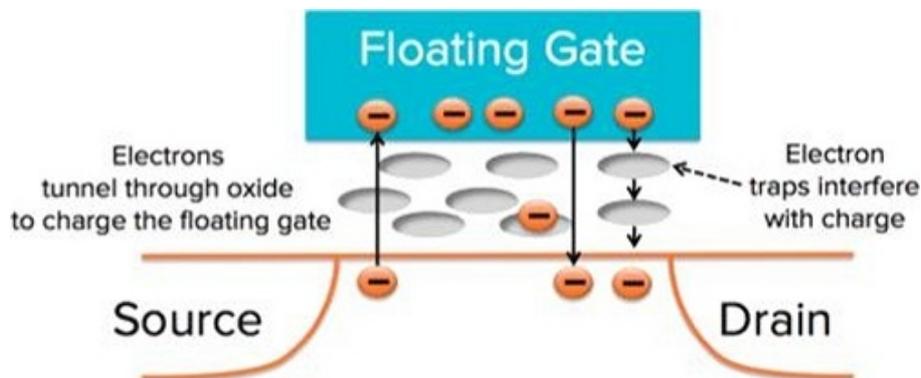


Figura 19: gli elettroni intrappolati nello strato di ossido interferiscono col livello di carica

Questo deleterio decadimento continuerà inesorabilmente a progredire fino a che, in buona sostanza, i valori di tensione *iniettati* al momento della programmazione (*scrittura*) non andranno più a corrispondere in fase di rilevazione (*lettura*) alle corrispondenti soglie prestabilite, falsando così a livello logico la coerenza dei dati sottostanti.

1 die	30	40	40	40	7	14	28	40
2 die	60	80	80	80	14	28	56	80
4 die	120	160	160	160	28	56	112	160

E' possibile notare come in lettura il bandwidth raggiungibile sia equivalente; in scrittura le SLC giungono invece a saturare la banda offerta dalle specifiche ONFI 1.x (160 MB/s) con 4 die, già a partire dall'uso di due soli canali, mentre per le MLC ne necessitano tutti e otto.

A titolo di paragone, le tabelle del *bandwidth* raggiungibile dalle SLC e MLC a partire dalle specifiche **ONFI 2.x** (max 200 MB/s per canale), con l'incremento dei die e dei canali.

Performance NAND SLC (MB/s): die per canale vs numero di canali

↔	lettura				scrittura			
	1	2	4	8	1	2	4	8
1 die	120	200	200	200	28	56	112	200
2 die	240	400	400	400	56	112	224	400
4 die	480	800	800	800	112	224	448	800

Performance NAND MLC (MB/s): die per canale vs numero di canali

↔	lettura				scrittura			
	1	2	4	8	1	2	4	8
1 die	88	176	200	200	8	16	32	64
2 die	176	352	400	400	16	32	64	128
4 die	352	704	800	800	32	64	128	256

Con le specifiche ONFI 2.x si può notare come la situazione in lettura rimane tutto sommato inalterata ed equivalente; in scrittura il divario si amplia invece in maniera abnorme, con le SLC che riescono a saturare il bandwidth di 800 MB/s in situazione di massimo parallelismo (4 die ed 8 canali), mentre nelle stesse condizioni di efficienza le MLC si fermano ad un parziale ~30% della banda di memoria offerta, attestandosi a soli 256 MB/s.

Parallelamente hanno anche provveduto ad innalzare la velocità del bus *memoria-controller*, migliorato in maniera contestuale l'efficienza delle tecnologie di *interleaving* e affinato lo sviluppo degli algoritmi **ECC** implementati in firmware, necessari alla correzione dell'accresciuta percentuale di errori rilevati in fase di lettura.

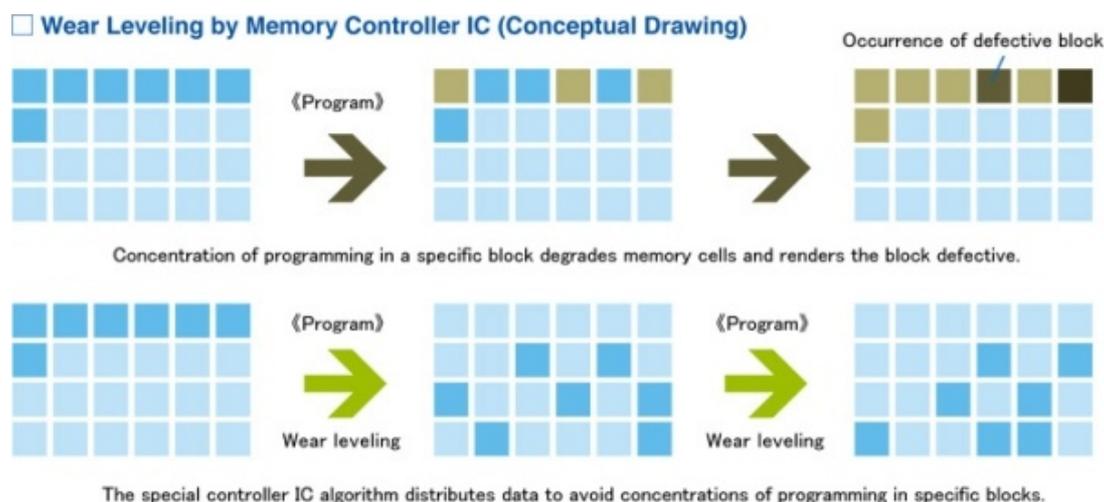


Figura 22: funzionalità di Wear Leveling

Tutte queste migliorie, unitamente allo sviluppo di ancor più avanzate strategie di gestione di *wear leveling*, *garbage collection* e *overprovisioning* sono divenute indubbiamente significative anche di quei fattori che hanno portato ad incidere in maniera più che rilevante al raggiungimento di una superiore longevità operativa delle più recenti NAND Flash multilivello.

Cosa cambia in concreto?

Ma allora, in pratica, cosa cambierebbe sostanzialmente per l'utente nell'acquisto di un SSD con le più recenti memorie NAND Flash da far propendere alla scelta delle nuove TLC piuttosto che le MLC o addirittura delle vecchie SLC, soprattutto alla luce dei più recenti processi produttivi?

La risposta è destinata ad essere inevitabilmente articolata: da una parte abbiamo il raggiungimento di un livello di rendimento prestazionale globalmente inferiore, sebbene parzialmente mitigato dal maggiore parallelismo funzionale ottenibile, un valore maggiore di **BER** (*Bit Error Rate*) e, in assoluto, soprattutto una minor longevità .

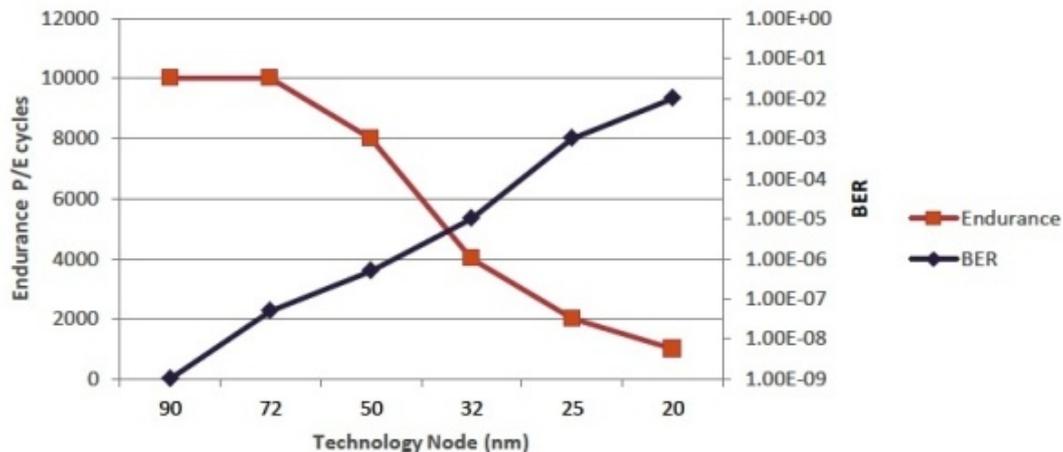


Figura 23: Bit Error Rate in relazione alla riduzione della geometria costruttiva

D'altro canto, con l'acquisto di una unità basata su NAND TLC viene garantita soprattutto la possibilità di ottenere capacità più elevate ad un prezzo tutto sommato di maggior convenienza, scaturito dal favorevole rapporto di *costo per GB* che lo sfruttamento delle ultime tecnologie riesce ad offrire

Facciamo alcuni calcoli

Tralasciando la valutazione sulle NAND Flash SLC a motivo della loro non-economicità , per giungere concretamente ad una quantificazione delle differenze a livello di rendimento tra le due ultime tecnologie, MLC e TLC, prendiamo a riferimento alcuni modelli di SSD prodotti da Samsung per avere un quadro sintetico con i valori dichiarati sulle prestazioni in lettura e scrittura sequenziali:

- Samsung 830 256GB - [MLC 27nm]↔ 520/400 MB/s (*lettura/scrittura*)
- Samsung 840 250GB - [TLC 21nm]↔ 530/240 MB/s (*lettura/scrittura*)
- Samsung 840 PRO 256GB - [MLC 21nm] 540/520 MB/s (*lettura/scrittura*)
- Samsung 840 EVO 240GB [TLC 19nm] 540/520 MB/s (*lettura/scrittura*)
- Samsung 850 PRO 256GB [3D-V 40nm] 550/520 MB/s (*lettura/scrittura*)

A parità di capacità (256GB) e processo costruttivo (21nm), il primo modello della serie 840, equipaggiato con NAND TLC, presenta una velocità in scrittura (240 MB/s) pressoché dimezzata rispetto al modello 840 Pro con memorie MLC (520 MB/s): tale deficit si mantiene sensibile anche nei confronti dell'unità 830 (400 MB/s) di precedente generazione (27nm).

Con l'uscita del modello 840 EVO di seconda generazione e dall'850 PRO 3D a sviluppo verticale, le prestazioni in scrittura sono decisamente accresciute allineandosi al top, conseguite genericamente dalle attività complementari alla riduzione del processo produttivo a 19nm e favorite dallo sviluppo di soluzioni più sofisticate a livello architetturale, come le 3D Vertical.

Volendo focalizzare la durata nel tempo delle soluzioni tecnologiche fin qui descritte e giungere così alla determinazione di un **valore** tangibile di longevità di un dispositivo SSD, è necessario introdurre il concetto di **TBW** (*Total Byte Written*, espresso in terabyte): TBW determina, in sostanza, la quantità di dati **scrivibili** durante l'arco di vita garantito da parte del produttore.

In primo luogo il suo valore è in diretta relazione alla capacità complessiva di dati memorizzabili e al numero di cicli **P/E** (*programmazione/cancellazione*) accreditati alle specifiche memorie NAND adottate; alcuni produttori indicano il procedimento in **W/E** (*write/erase, scrittura/cancellazione*), volendo però significare lo stesso medesimo concetto.

Per giungere ad un paragone per quanto possibile imparziale ed affidabile tra le unità SSD, in condizioni in un certo qual modo *standardizzate* di confronto, viene introdotto il termine di **Write Amplification (WA)** e ne è comunemente adottato un fattore pari a **1**.

Il significato di questo importante parametro è un concetto che vale la pena di essere approfondito: *WA* rappresenta il *rapporto* che intercorre tra la quantità di *dati scritti* sull'unità SSD da parte del controller (*Writes to NAND*) e la quantità, invece, di quelli originariamente *inviati* allo stesso controller dal sistema operativo (*Writes from Host*).

WA : N = (DATI↔ scritti↔ sulle↔ NAND / DATI↔ scritti↔ dall'Host)

Nel caso vengano usati controller che gestiscono in maniera del tutto trasparente le comunicazioni da e verso le memorie e qualora si trasferiscano dati in modalità sequenziale, il rapporto tra le due quantità si mantiene infatti equivalente, per via della diretta corrispondenza, e solo leggermente superiore a **1**.

Nella maggior parte degli SSD e così come accade negli hard disk, al momento di trasferire dati in maniera sequenziale, il valore di *Write Amplification* si mantiene infatti ~ 1 , e ci sentiamo quindi di adottare tale valore, seppur *ideale* ma in qualche modo *ben definito*, come riferimento per le nostre attuali e future determinazioni di durata di questi dispositivi.

Circostanze differenti, e anche molto difficilmente quantificabili, sono infatti tutte quelle in cui il trasferimento dei dati è eseguito in modalità *random*: in questi frangenti, dove molto più importanti sono le complete attività di intervento del controller (*wear leveling, garbage collection* etc.) il fattore di *WA* è costretto a subire dinamicamente una variazione anche superiore alle dieci volte, potendo giungere facilmente al valore ~ 12 .

Questo aspetto, oltre che estremamente penalizzante per la durata del nostro dispositivo, è anche molto difficile da valutare nell'insieme, rappresentando un delta di scostamento significativamente troppo ampio per la determinazione di un valore da adottare quale riferimento univoco.

Quando il controller è infatti congegnato in modo da *comprimere/decomprimere* in tempo reale i dati in transito, il valore di *WA* subirà una diminuzione più o meno drastica a seconda dei casi, apportando innegabili benefici alla durata delle NAND, a prescindere da quale modalità, *sequenziale* o *random*, venga utilizzata.

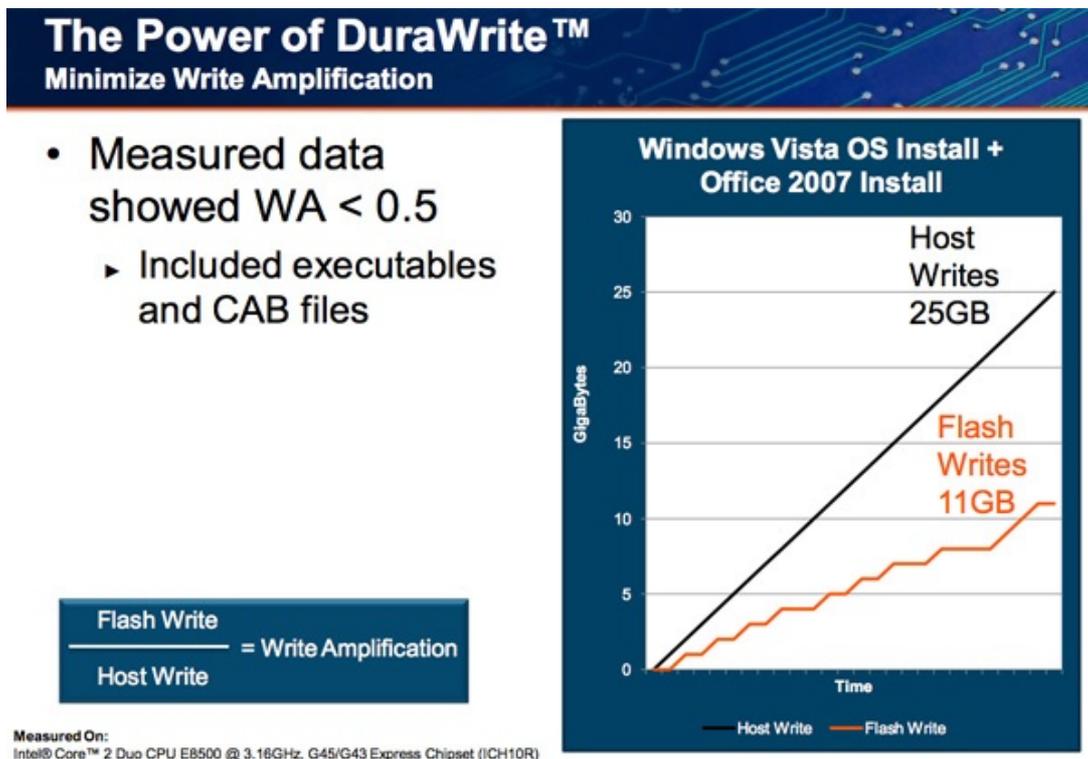


Figura 24: diminuzione del fattore di Write Amplification in relazione alla compressione dei dati

In tali circostanze la quantità di dati inviati dal s.o. (*Writes from Host*), una volta *compressi* in tempo reale dall'unità di calcolo del controller, diverrà sensibilmente inferiore al momento di essere memorizzata all'interno delle celle (*Writes to NAND*) rispetto a quella originata, potendo giungere nelle situazioni particolarmente favorevoli anche ad un valore di ~ 0.2 .

Un simile valore, sebbene direttamente connesso alla comprimibilità dei dati trattati, se mantenuto nel lungo periodo è in grado di poter anche quintuplicare la durata delle memorie NAND Flash.

L'esplicazione della formula risolverà qualsiasi dubbio: basterà dividere la capacità complessiva dell'unità SSD (in gigabyte) per il valore di **WA**, moltiplicare quanto ottenuto per il numero dei cicli **P/E** accreditati ed infine dividerne il risultato per mille.

$$\mathbf{TBW : TB = [(CAPACITA' \text{ in GB} / WA) * (CICLI \leftrightarrow P/E)] / 1000}$$

Per stimare invece la presumibile durata, in anni, basterà dividere il valore **TBW**, espresso in terabyte, per il cosiddetto **GB/day**, ovvero il carico di lavoro giornaliero che descrive la quantità di dati che si assume venga scritta dall'host (il sistema operativo) durante una tipica giornata d'uso, espresso in gigabyte.

Il valore risultante deve a questo punto essere moltiplicato per il valore derivante dalla divisione di 1000 per il numero di giorni dell'anno.

$$\mathbf{LONGEVITA' : ANNI = [(TBW TB) / (GB/day)] * (1000 / 365)}$$

Ipotizzando un carico di lavoro medio in scrittura da parte del s.o. pari a circa **20 GB** giornalieri, che rappresenta una quantità di dati abbastanza omogenea ipotizzando un uso generico, è possibile quindi giungere alla stima sia dei dati scrivibili (TBW) in terabyte, che della longevità in anni.

- Samsung 830 256GB (MLC 27nm, 3000 P/E) **768 TB** (105 anni)
- Samsung 840 250GB (TLC 21nm, 1000 P/E) **250 TB** (34 anni)
- Samsung 840 PRO 256GB (MLC 21nm, 3000 P/E) **768 TB** (105 anni)
- Samsung 840 EVO 240GB (TLC 19nm, 1200 P/E) **288 TB** (40 anni)
- Samsung 850 PRO 256GB (3D-V 40nm, 3000 P/E @ 50 MB/s, 35000 P/E @ 36 MB/s) **768 TB** (105 anni).

4. Conclusioni

4. Conclusioni

Naturalmente teniamo a specificare come le stime in gioco possano comunque costituire un fattore ragionevolmente approssimativo: i numeri usati in precedenza si basano infatti su modelli e previsioni statistiche fornite dai singoli produttori e, pertanto, nei frangenti di reale utilizzo potrebbero verosimilmente discostarsi anche in misura considerevole da quelli effettivi, dove possono confluire necessità spesso contingenti ed esigenze differenziate ed in parte non prevedibili, a monte, dall'utilizzatore.

Al di là di questo, ci sembra utile aggiungere che si potrebbe benissimo arrivare a scrivere una quantità di dati anche quattro o cinque volte superiore a quella dei 20GB preventivata negli esempi illustrati in precedenza, senza per questo doversi ancora preoccupare della salute dei propri dati.

La durata di un qualsiasi modello di SSD, a prescindere dalla tipologia di moderne NAND Flash utilizzate, risulterebbe infatti ugualmente idonea alle pressoché totali esigenze anche oltre l'arco di vita *efficace* del dispositivo, seppur con un carico di lavoro giornaliero in scrittura vicino ai **100GB**.

Bisogna a questo scopo anche valutare a titolo indicativo che le statistiche e la nostra esperienza ci portano a ritenere, infatti, come raramente un dispositivo di storage potrà superare i cinque/sette anni dal momento della sua commercializzazione senza, contestualmente, divenire *tecnologicamente obsoleto* a seguito del decadimento nel tempo della propria intrinseca efficacia prestazionale.

Tali soluzioni interessano in genere l'adozione di sempre nuovi e più ampi canali e bus di comunicazione al fine, fondamentalmente, di incrementare il bandwidth di I/O.

Di pari passo a questa evoluzione giunge spesso inevitabile anche l'introduzione di nuovi chipset e storage-controller nonché di più efficienti protocolli i quali insieme *rischiano*, nel giro di pochi anni, di rendere se non propriamente obsoleti, di certo non più all'altezza i nostri "vecchi" dispositivi di memorizzazione nell'uso sulle nuove piattaforme.

Ad ogni buon conto, nel momento in cui scriviamo il nostro ottimismo sulla possibile adozione di una unità allo stato solido quale dispositivo principale del proprio sistema, diviene sempre più concreto all'aumentare della capacità dell'unità SSD: significando che potendo scegliere un taglio a partire dai 240/256GB, o meglio se da 480/512GB, a maggior ragione ci sentiamo senz'altro di poter tranquillizzare tutti i futuri utilizzatori sull'estrema affidabilità nel tempo delle moderne unità allo stato solido.

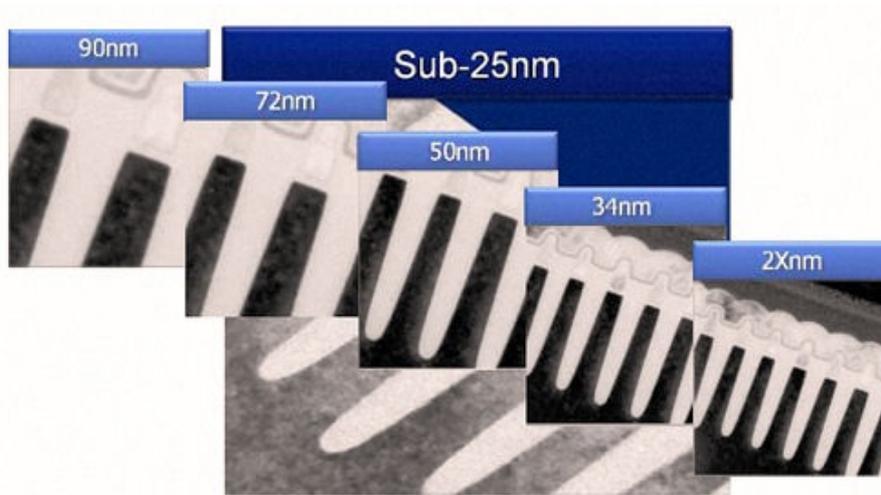


Figura 25: scaling delle geometrie costruttive (fonte Samsung)

Prendendo come riferimento gli ultimi SSD "entry level" rilasciati sul mercato e caldeggiando, come sempre, di attivare tutti gli accorgimenti e le ottimizzazioni consigliate nell'uso in Windows, la durata sarà con ogni probabilità anche superiore a quella raggiungibile da un moderno disco fisso di tipo tradizionale (*meccanico-magnetico*) al momento sul mercato, persino tra i migliori di classe Enterprise.

Soprattutto, non finiremo mai di sottolineare il vero fattore fondamentale per l'utente che ha in mente la reattività del proprio sistema come criterio basilare nella scelta del proprio dispositivo di storage: parliamo del **tempo di accesso**, ovvero quello strettamente necessario al dispositivo di memorizzazione per accedere al suo **strato fisico**, dove i dati sono effettivamente contenuti.

Ebbene, sotto questo profilo il peggiore degli SSD offre un valore di accesso di soli $\sim 0,1ms$, circa un cinquantesimo del tempo rispetto al più veloce degli Hard Disk ($\sim 5ms$); questa netta superiorità a livello temporale dà garanzia di avvii immediati del sistema operativo e caricamenti fulminei anche di vastissime quantità di informazioni, sia in modalità sequenziale che random, per tutti gli scenari di utilizzo, *gaming* compresi, dove ricordiamo che la gran parte delle operazioni avviene per lo più in lettura.