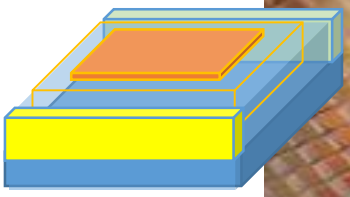
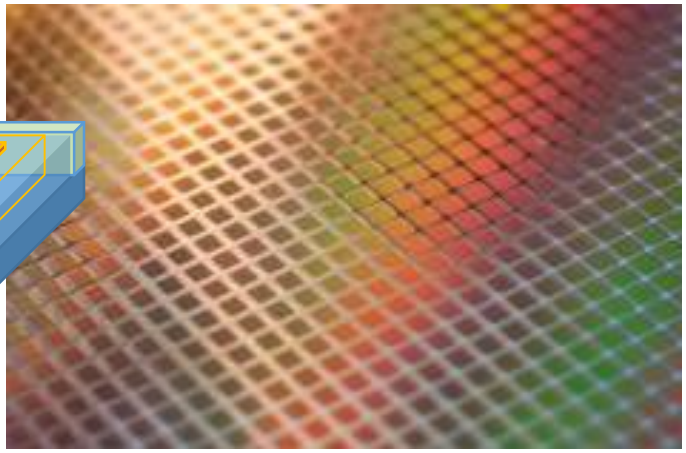
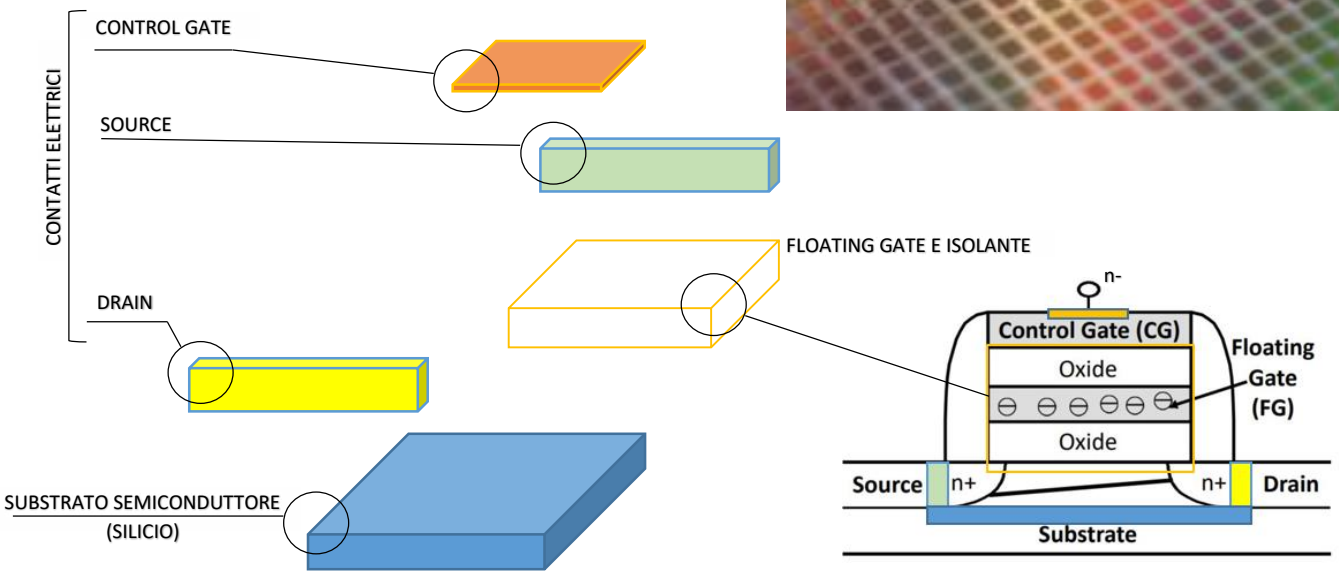
CELLA NAND FLASH… LA TEORIA DEL FUNZIONAMENTO

PRENDO COME ESEMPIO IL TRANSISTOR DI TIPO SLC PLANARE
CON TECNOLOGIA A GATE FLOTTANTE,
E' LA CELLA NAND PIU' SEMPLICE CUI SONO SEGUITE VARIE EVOLUZIONI, MA IL PRINCIPIO DI FUNZIONAMENTO E' SOSTANZIALMENTE RIMASTO LO STESSO.







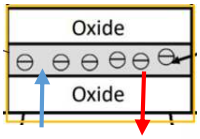
LA BASE DI QUALSIASI CELLA DI QUESTO TIPO E' IL **SEMICONDUTTORE**. IL SILICIO, E' ISOLANTE MA, DROGATO AD ESEMPIO CON FOSFORO, CHE HA UN ELETTRONE IN PIÙ, DIVENTA CONDUTTORE, CON UNA CONCENTRAZIONE DI ELETTRONI PARI ALLA CONCENTRAZIONE DI ATOMI DI FOSFORO. INFATTI L'ELETTRONE IN PIÙ* RISUTA POCO LEGATO E CON POCA ENERGIA DIVENTA LIBERO, OSSIA PASSA NELLA BANDA DI CONDUZIONE, MENTRE GLI ALTRI ELETTRONI PARTECIPANO AL LEGAME ATOMICO.
IL FENOMENO VIENE DETTO: **OSSIDAZIONE**

LA MEMORIZZAZIONE NELLA CELLA NAND FLASH AVVIENE GRAZIE A QUESTI ELETTRONI LIBERI, PER MEZZO DI UNA **CARICA ELETTRICA** BASSA CHE VIENE FATTA PASSARE NEL SEMICONDUTTORE E CHE GENERA OSSIDAZIONE, CON PASSAGGIO DI ELETTRONI AD UN APPOSITO MATERIALE LA CUI CARICA (NEGATIVA) PERSISTE AL CESSARE DELLA CORRENTE ELETTRICA CHE L'HA GENERATA.

SONO 3 I FENOMENI FISICI CHE AVVENGONO IN UNA CELLA NAND:
LA **CORRENTE ELETTRICA**, (DA SOURCE A DRAIN) OSSIA L'ECCITAZIONE DEGLI ELETTRONI CHE NON PARTECIPANO AL LEGAME ATOMICO*;
L'**OSSIDAZIONE** DI UN MATERIALE (SUBSTRATO) CAPACE DI RILASCIARE ELETTRONI A BASSISSIME CARICHE ELETTRICHE (SEMICONDUTTORE)
L'**EFFETTO TUNNEL** (TUNNELING) NELLO SPOSTAMENTO DEGLI ELETTRONI, CIOE' IL VERIFICARSI DEL FENOMENO QUANTISTICO DI FISICA ONDULATORIA DELLE PARTICELLE CHE HA COME EFFETTO, LA PENETRAZIONE DEGLI ELETTRONI ATTRAVERSO UNA BARRIERA ISOLANTE ANCHE A BASSA ENERGIA.

TALE BARRIERA ESSENDO ISOLANTE, TRATTIENE GLI ELETTRONI AL CESSARE DELLA CORRENTE ELETTRICA CHE HA PROVOCATO OSSIDAZIONE E NEL CONTEMPO NON SI LASCIA ATTRAVERSARE DAGLI ELETTRONI SENZA CHE VI SIA CORRENTE ELETTRICA CHE INTERVIENE.
ECCO IL PRINCIPIO CHE SPIEGA LA **RITENZIONE DEL BIT NELLA CELLA** ANCHE QUANDO IL DISPOSITIVO NON E' ALIMENTATO.

IL DATO MEMORIZZATO DALLA CELLA NON E' ALTRO QUINDI CHE LA PRESENZA O MENO DI **ELETTRONI INTRAPPOLATI NEL GATE**.



FLOATING GATE (POLISILICIO) E STRATI ISOLANTI

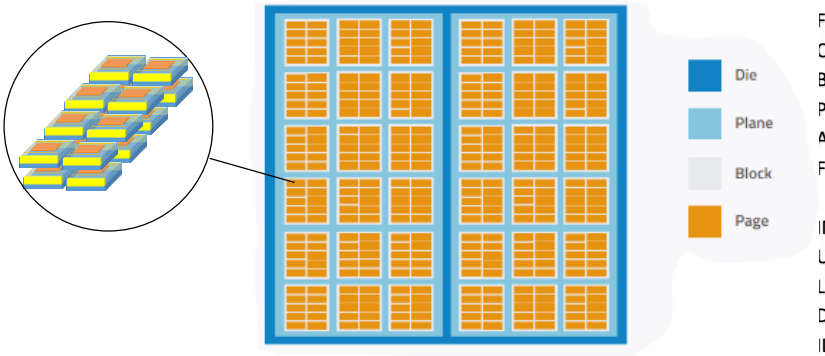
SCRIVERE UNA CELLA VUOL DIRE FAR PASSARE UNA LEGGERA CORRENTE TRA DRAIN E SOURCE, IN MODO CHE IL SEMICONDUTTORE RILASCI UNA SPECIFICA QUANTITA' DI ELETTRONI.
GLI ELETTRONI PER L'EFFETTO TUNNEL SUPERANO LA BARRIERA ISOLANTE E RESTANO INTRAPPOLATI NEL FLOATING GATE.

CANCELLARE UNA CELLA VUOL DIRE APPLICARE UNA CARICA DI TENSIONE PIÙ ELEVATA (SEMPRE TRA SOURCE E DRAIN), CHE INTERROMPE IL TUNNELING E RESTITUISCE AL SEMICONDUTTORE LA CARICA ACCUMULATA NEL FLOATING GATE.

QUANDO SI VUOLE INVECE **LEGGERE** SE UNA PARTICOLARE CELLA HA IL VALORE "0" PIUTTOSTO DI "1", OCCORRE FARE PASSARE UN DETERMINATO VOLTAGGIO (RAPPRESENTA LA SOGLIA DI CARICA DI QUEGLI ELETTRONI INTRAPPOLATI), ATTRAVERSO IL CONTROL GATE:
SE LA CORRENTE ELETTRICA TRA SOURCE E CONTROL GATE PASSA, VUOL DIRE CHE NON CI SONO ELETTRONI NEL FLOATING GATE, PER CUI IL VALORE E' "1", SE LA CORRENTE ELETTRICA E' OSTACOLATA (ELETTRONI PRESENTI NEL FLOATING GATE=SOGLIA DIVERSA), LA CELLA CORRISPONDE AL VALORE "0".

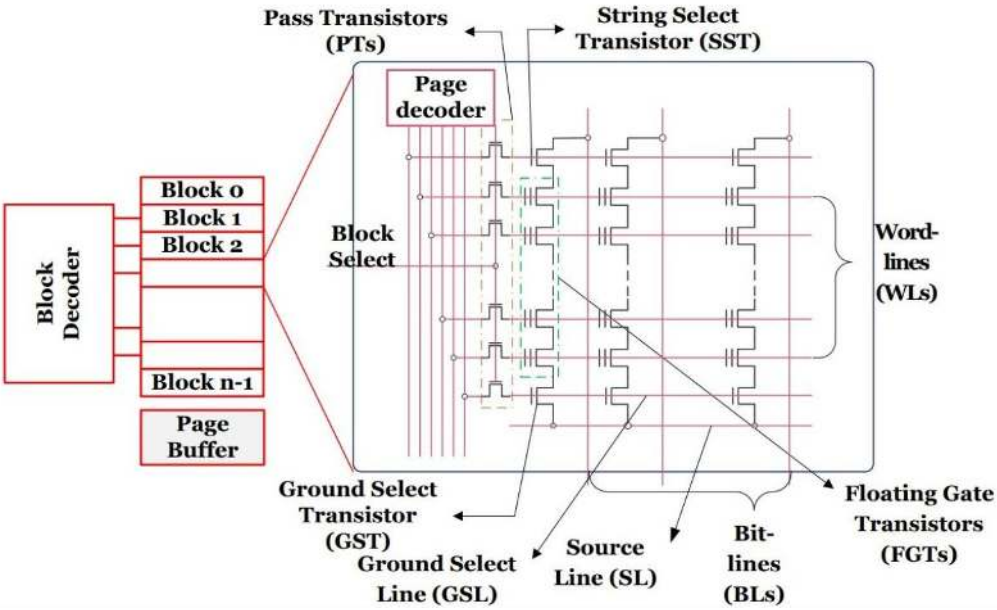
PER FARE IN MODO CHE DA OGNI CELLA SI RILEVI UN VALORE (BIT) E CHE NELL'INSIEME COSTITUISCANO DEI DATI MEMORIZZATI, LE CELLE VENGONO ORGANIZZATE **STRUTTURALMENTE** IN UN ARRAY, GRUPPI, CHE CONDIVIDONO LINEE DI CORRENTE ELETTRICA E ORGANIZZATE **LOGICAMENTE** NELLA RIPARTIZIONE GIA' UTILIZZATA DAI PRECEDENTI STRUMENTI DI MEMORIZZAZIONE.
VEDIAMO NELLO SPECIFICO L'ORGANIZZAZIONE DAL CHIP DI MEMORIA CHE VEDIAMO SULL'SSD SEMPRE PER L'IPOTETICA SINGOLA CELLA (UNA CELLA PLANARE SLC) IN UN VIRTUALE TEARDOWN:

SULL'SSD IDENTIFICHIAMO, OLTRE AL CONTROLLER E LA/LE EVENTUALI MEMORIE DRAM, DEI CHIP DI MEMORIA CHE SONO COSTITUITI DA UN RIVESTIMENTO ESTERNO SERIGRAFATO.
OGNI CHIP CONTIENE UNO O PIÙ STAMPI (AD ESEMPIO UNO, DUE O QUATTRO). IL **DIE** È L'UNITÀ PIÙ PICCOLA CHE PUÒ ESEGUIRE INDIPENDENTEMENTE COMANDI.
ALL'INTERNO UNA DIE E' FORMATA DA PIU' **PIANI** INTERCONNESSI (ALMENO 2 COME EVIDENZIATO IN FIGURA). I PIANI HANNO PERMESSO DI FAR STARE PIU' CELLE NELL'UNITA' DI AREA, E HANNO DATO POI L'IDEA ALLA BASE DELLE FUTURE CELLE 3D. I PIANI SONO IMPORTANTI PER LA STRUTTURA DEL CHIP MA IN SE NON PARTECIPANO AL MECCANISMO LOGICO DI FUNZIONAMENTO DELLA NAND FLASH.

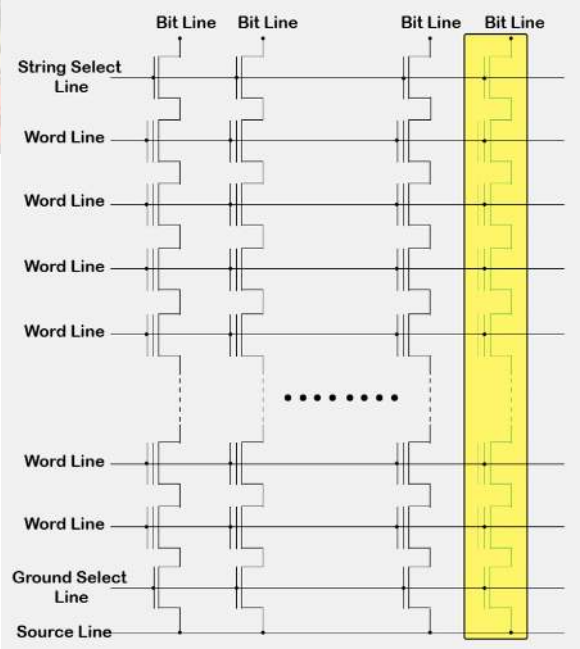


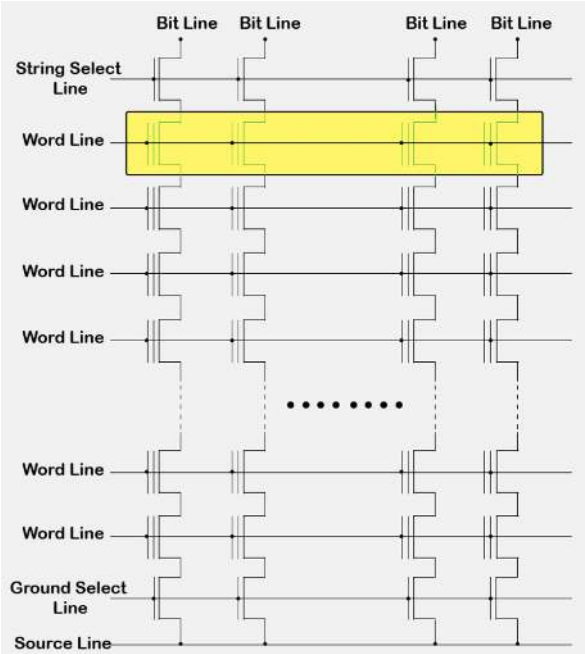
INVECE, OGNI DIE E' ORGANIZZATO A LIVELLO LOGICO DA UN DEFINITO NUMERO DI **BLOCCHI** CHE COSTITUISCONO L'UNITA' DI MEMORIZZAZIONE PIU' PICCOLA DI UN CERTO DATO MEMORIZZATO (4K O 512 BYTE EMUNATI).
IL BLOCCO E' UN INSIEME DI CELLE MA QUESTE SONO ORGANIZZATE DAL PUNTO DI VISTA STRUTTURALE E LOGICO IN FRAZIONI PIU' PICCOLE, OVVERO LE **PAGINE**.
UNA PAGINA COSTITUISCE INVECE L'ARRAY DI CELLE PIU' PICCOLO IN CUI ESSE SONO ORGANIZZATE, ALMENO DAL PUNTO DI VISTA LOGICO, I BLOCCHI.
VEDIAMO ORA COME E' ORGANIZZATO UN BLOCCO E COME FUNZIONA LA MEMORIZZAZIONE (PROGRAMMAZIONE), LA LETTURA E INFINE LA CANCELLAZIONE DI UN DATO, A LIVELLO DI UN VERO E PROPRIO CIRCUITO ELETTRICO.

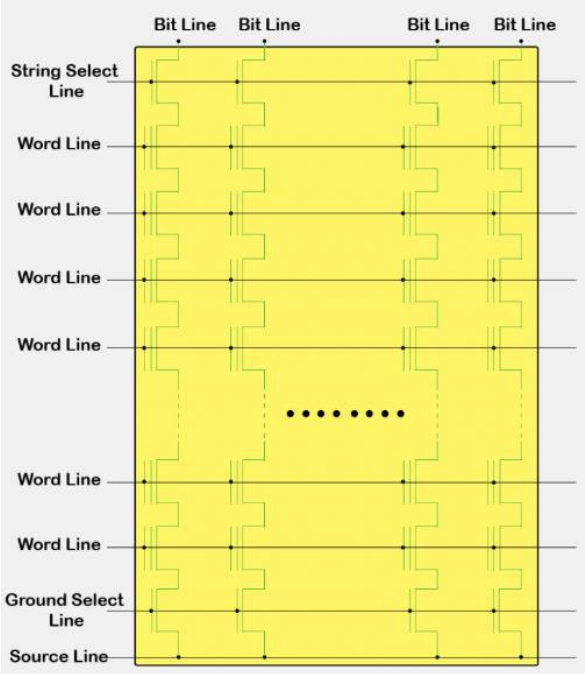
QUESTO E' LO SCHEMA SEMPLIFICATO DI UNA PORZIONE DI ARRAY DELLE PRECEDENTI CELLE SLC PLANARI, NELLA STRUTTURA ORIGINALE:



OGNI BLOCCO, COME SI PUO' VEDERE, E' UN INSIEME, UN ARRAY DI CELLE ORGANIZZATE.
DI PER SÉ, UNA SINGOLA CELLA NON AVREBBE MOLTO VALORE. MA LA COMBINAZIONE DI MOLTE CELLE È CIÒ CHE CONSENTE L'ARCHIVIAZIONE DI QUANTITÀ SIGNIFICATIVE DI DATI. IL PRIMO PASSO PER COMBINARE SINGOLE CELLE NAND È LA STRINGA NAND (STIAMO GUARDANDO DENTRO UN SINGOLO BLOCCO):







UNA STRINGA (A LATO IN GIALLO) E' L'INSIEME DELLE CELLE COLLEGATE ALLA MEDESIMA BIT LINE E COLLEGATE IN UN PUNTO ALLA SOURCE LINE (COLLEGA LE SORGENTI "S" DELLE SINGOLE STRINGHE).
A SUA VOLTA OGNI SINGOLA CELLA DELLA STRINGA E' COLLEGATA ALLA WORD LINE (ORTOGONALMENTE QUINDI).
OGNI STRINGA CONTIENE QUINDI **DUE MECCANISMI DI CONTROLLO** IN SERIE CON LE ALTRE CELLE NAND.
UN MECCANISMO SERVE PER PROGRAMMARE E CANCELLARE, MENTRE UN ALTRO CONSENTE DI LEGGERE LA CELLA.
UNA STRINGA DI CELLE NAND PUÒ MEMORIZZARE 32 BIT DI DATI, QUESTO SI TRADUCE COMUNQUE SOLO IN 4 BYTE DI DATI O SUFFICIENTI PER 4 CARATTERI. LE STRINGHE (MOSTRATE COME COLONNE) SONO L'UNITÀ MINIMA DA LEGGERE E SONO IN GENERE COMPOSTE DA 32 O 64 CELLE NAND. PERTANTO, LE STRINGHE VENGONO COMBINATE IN ARRAY PIÙ GRANDI PER OTTENERE UNA QUANTITÀ UTILE DI SPAZIO DI ARCHIVIAZIONE.

MA COME SI RIESCE AD INVIARE GLI INPUT DI PROGRAMMAZIONE, ERASE, O LEGGERE IL DATO SULLA STRINGA?

I TRANSISTOR STRING SELECT E GROUND SELECT SONO COLLEGATI ALLA STRING SELECT LINE (SSL) E ALLA GROUND SELECT LINE (GSL).
QUESTI ELEMENTI DI CONTROLLO SONO A LORO VOLTA COLLEGATI ATTRAVERSO DELLE LINEE AL CANALE DI COMUNICAZIONE DEL **CONTROLLER**.

LA STRINGA RAPPRESENTA IL VERSO DI PROGRAMMAZIONE O CANCELLAZIONE DELLE CELLE NAND FLASH.

DATO CHE L'ELEMENTO FONDAMENTALE DELLA MEMORIZZAZIONE E' LA LETTURA, TUTTE LE CELLE CHE CONDIVIDONO LA STESSA WORD LINE SONO L'UNITÀ MINIMA DA PROGRAMMARE ED ANCHE IL SENSO IN CUI VENGONO LETTE LE CELLE DELLA STRINGA.

QUESTE SONO LE **PAGINE** E VENGONO RAPPRESENTATE ORTOGONALMENTE ALLE STRINGHE (SI VEDA LA SECONDA FIGURA).

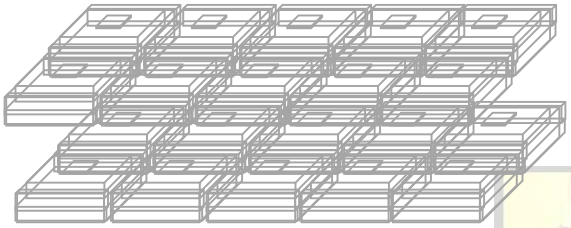
QUINDI COSI' COME LA STRINGA RAPPRESENTA IL VERSO DI PROGRAMMAZIONE E CANCELLAZIONE DELLE CELLE, LA PAGINA NE RAPPRESENTA IL VERSO DI LETTURA.

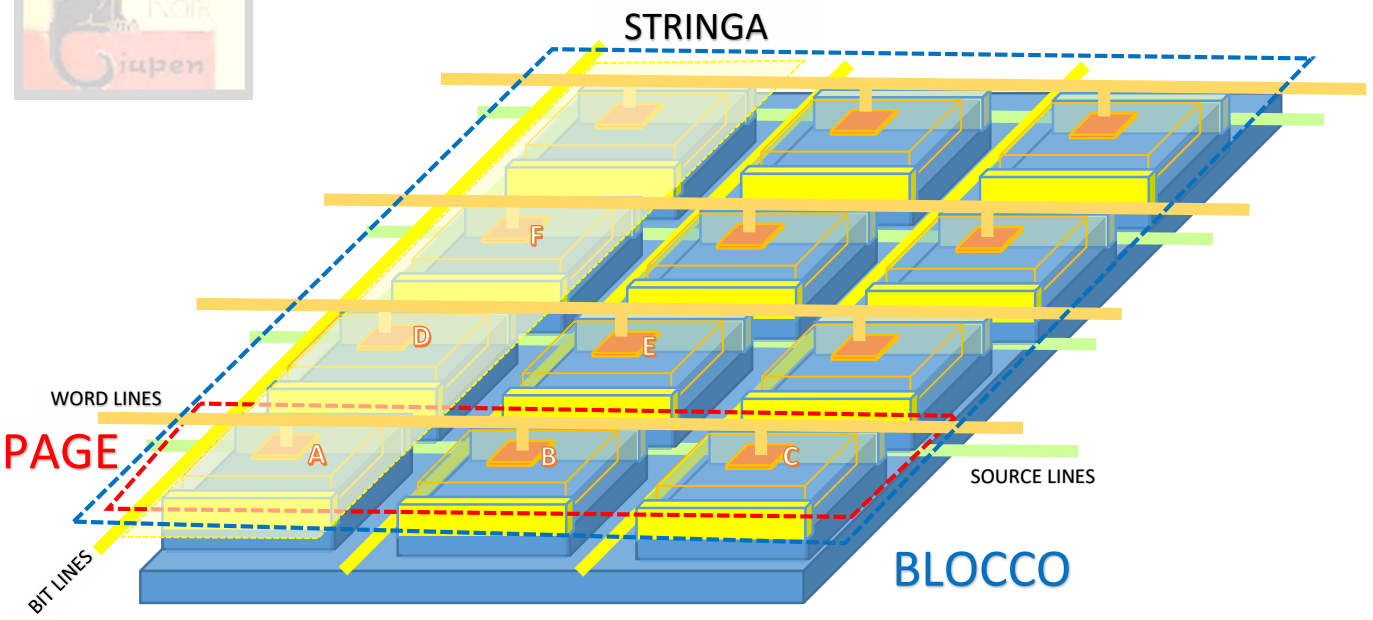
IN GENERE UNA PAGINA E' COSTITUITA DA ALMENO QUALCHE DECINA DI MIGLIAIA DI CELLE NAND, CON MOLTI DEI PIÙ RECENTI DISPOSITIVI NAND HANNO DIMENSIONI PAGINA DI 128K SE NON 256K O 512K.
LA MAGGIOR PARTE DELLE DIMENSIONI DELLA PAGINA SONO INDICATE COME 2K, 4K, 8K, ECC. CIÒ INDICA LA DIMENSIONE DELLA PAGINA IN BYTE.

L'INSIEME DELLE PAGINE COSTITUISCE UN **BLOCCO**.

UN BLOCCO È UNA MATRICE BIDIMENSIONALE COMPOSTA DA PAGINE (RIGHE) E STRINGHE (COLONNE). IL NUMERO TOTALE DI BIT IN UN BLOCCO PUÒ ESSERE CALCOLATO MOLTIPLICANDO IL NUMERO DI STRINGHE PER IL NUMERO DI PAGINE.

PER CAPIRE IL MECCANISMO DELLE TRE OPERAZIONI DELLE CELLE (PROGRAMMAZIONE/CANCELLAZIONE/LETTURA), HO PENSATO DI FARVI VEDERE UNA IPOTETICA VISTA TRIDIMENSIONALE DELL'ARRAY:





SI IMMAGINA UN BLOCCO IDEALE REALIZZATO CON POCHE CELLE NAND FLASH; ALCUNE DI QUESTE SONO A,B,C,D, ECC.

AL COMANDO DI NECESSITÀ DI **SCRITTURA** DA PARTE DELL'HOST, IL CONTROLLER ESEGUE UNA RICERCA E IDENTIFICA I BLOCCHI VUOTI (CANCELLATI).
QUINDI INIZIA LA SCRITTURA RIEMPIENDO SEQUENZIALMENTE LE PAGINE DEL BLOCCO E ANDANDO AL BLOCCO LIBERO SUCCESSIVO.
PER SCRIVERE AD ESEMPIO NELLA 1A PAGINA LA 1° CELLA "A", LA BASE DELLA CELLA VIENE FATTA OGGETTO DI UNA TENSIONE ELETTRICA SPECIFICA (NE TROPPO DEBOLE, NE TROPPO FORTE, MA "QUELLA" GIUSTA CHE PERMETTE DI STACCARE LA QUANTITÀ DI ELETTRONI LIBERI E ATTRAVERSO IL TUNNELING DEPOSITARLI SUL FLOATING GATE DI "A".

SOURCE LINE A (CARICA MAGGIORE) + BIT LINE A = SCRITTURA A [LA CELLA DA 1 PASSA AL VALORE 0]

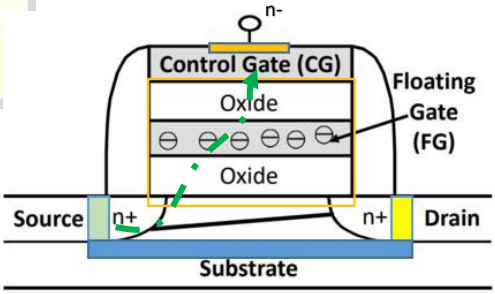
NATURALMENTE SE IL PRIMO VALORE DELLA PAGINA FOSSE UN 1, LA CELLA NON VIENE SCRITTA E RIMANE IL VALORE 0.

SIMILMENTE LA SCRITTURA PER LE ALTRE CELLE DELLA PAGINA O PER LE ALTRE CELLE DELLA PAGINA SUCCESSIVA: VIENE ATTIVATA LA RISPETTIVA SOURCE LINE E FATTO PASSARE UNA LEGGERA CORRENTE ELETTRICA.

DURANTE UN'OPERAZIONE DI **CANCELLAZIONE** , È NECESSARIA UNA GRANDE TENSIONE NEGATIVA PER RESPINGERE GLI ELETTRONI DAL GATE FLOTTANTE, CIÒ SI OTTIENE COLLEGANDO A TERRA IL GATE DI CONTROLLO E APPLICANDO 20 V AL SUBSTRATO, ATTRAVERSO LE BIT LINE E SORCE LINES (MATRICE IP-WELL). DI CONSEGUENZA, GLI ELETTRONI VENGONO RIMOSSI DAL GATE FLOTTANTE A CAUSA DELL'EFFETTO TUNNELING GIA' VISTO.

SI DICE CHE "UN SSD PUO' PROGRAMMARE OGNI SIGOLA CELLA MA QUANDO CANCELLA, PUÒ SOLO CANCELARE L'INTERO BLOCCO", VERO MA… VEDIAMO PERCHÉ.
QUANDO L'HOST HA UN FILE DA SCRIVERE L'SSD INIZIA A SCRIVERE UNA PAGINA ALLA VOLTA. PER FARLO UTILIZZA LA SOUCE LINE E LA BIT LINE A BASSA TENSIONE SUL SUBSTRATO, ATTIVATE NEL MOMENTO PER OTTENERE SU OGNI CELLA UN VALORE 1/0 (COME VISTO PRECEDENTEMENTE).
DOVENDO CANCELLARE LA SINGOLA CELLA PER RISCRIVERE DATI NUOVI (VEDERE FIGURA SOPRA), OCCORRE SVUOTARE/CANCELLARE LA RELATIVA PAGINA CUI APPARTIENE LA CELLA. AD ESEMPIO SE DEVO SCRIVERE INIZIO CON 1 BIT SU "A", POI SU "B", ECC
L'OPERAZIONE DI CANCELLAZIONE APPLICA SULLA SOURCE LINE UNA ALTA TENSIONE CHE RENDE CANCELLATA (VALORE 1) QUINDI L'INTERA PAGINA.
qUINDI ORA ABBIAMO UN BLOCCO COSTITUITO DA PAGINE CANCELLATE E PAGINE ANCORA VALIDE… TEORICAMENTE LA COSA (programmazione parziale delle pagine) DOVREBBE FUNZIONARE; Sebbene abbiamo ritardato l'esecuzione delle operazioni di cancellazione spostando in modo intelligente i dati su pagine diverse, a un certo punto chiaramente non rimarranno pagine vuote e le cancellazioni diventeranno essenziali. È qui che arrivano le cattive notizie: per eseguire una cancellazione ci vuole molte volte più tempo che per eseguire una lettura o un programma . E questo ha chiaramente delle conseguenze sulle prestazioni se non gestito correttamente.
OLTRE AL RALLENTAMENTO DELL'SSD C'è UN SECONDO FATTORE PIù IMPORTANTE CHE HA RESO LA PROGRAMMAZIONE PARZIALE DELLE PAGINE NON ADOTTATA CONCRETAMENTE: LA TENSIONE PIù ALTA DI CANCELLAZIONE INVESTENDO tutto il substrato del blocco (che è unico), CREA UN DISTURBO CHE PUò CANBIARE LO STATO DELLE CELLE ATTIGUE.
lA SOLUZIONE AD ENTRAMBI I PROBLEMI è STATA QUELLA DI CONSIDERARE I BLOCCHI ISOLE AUTONOME (LO SONO ELETTRICAMENTE).
Un'operazione di cancellazione su un chip flash cancella i dati da tutte le pagine nel blocco , quindi se alcune delle altre pagine contengono dati attivi (DATI DA conservare) OCCORRE prima copiarli altrovE.
ECCO perché una pagina può essere modificata (cioè scritta) solo dopo che l'intero blocco a cui appartiene è stato cancellato
lA CONSEGUENZA PIù IMMEDIATA (COME GIà ACCENNATO) è CHE è necessario preparare nuovi blocchi vuoti (cancellati) prima della programmazione (cioè, cancellare prima della scrittura).
Inoltre, l'operazione di cancellazione ha una latenza molto più lunga rispetto alle operazioni di lettura e programmazione.

PER LA **LETTURA** DELLE CELLE VENGONO UTILIZZATE LE WORD LINES.



UNA CARICA DI SOGLIA VIENE FATTA PASSARE TRA SOURCE E CONTROL GATE. SI TRATTA DI UNA TENSIONE MOLTO BASSA SENSIBILE ALLA PRESENZA DI ALTRE CARICHE ELETTRICHE EVENTUALMENTE PRESENTI SUL SUO PERCORSO.
SE AL CONTROL GATE LA TENSIONE È LA STESSA DI SOURCE, VUOL DIRE CHE NON SONO PRESENTI ELETTRONI NEL FLOATING GATE E QUINDI IL VALORE ASSEGNATO È "1".
SE LA TENSIONE NEL CONTROL GATE HA SUBITO UN ABBASSAMENTO (ANCHE IN QUESTO CASO SARANNO DELLE SOGLIE), A CAUSA DELLA PRESENZA DI ELETTRONI NEL FLOATING GATE, ALLORA IL VALORE ASSEGNATO SARÀ "0"

https://www.youtube.com/watch?v=M-wNC3Z3ZX4