

Le memorie

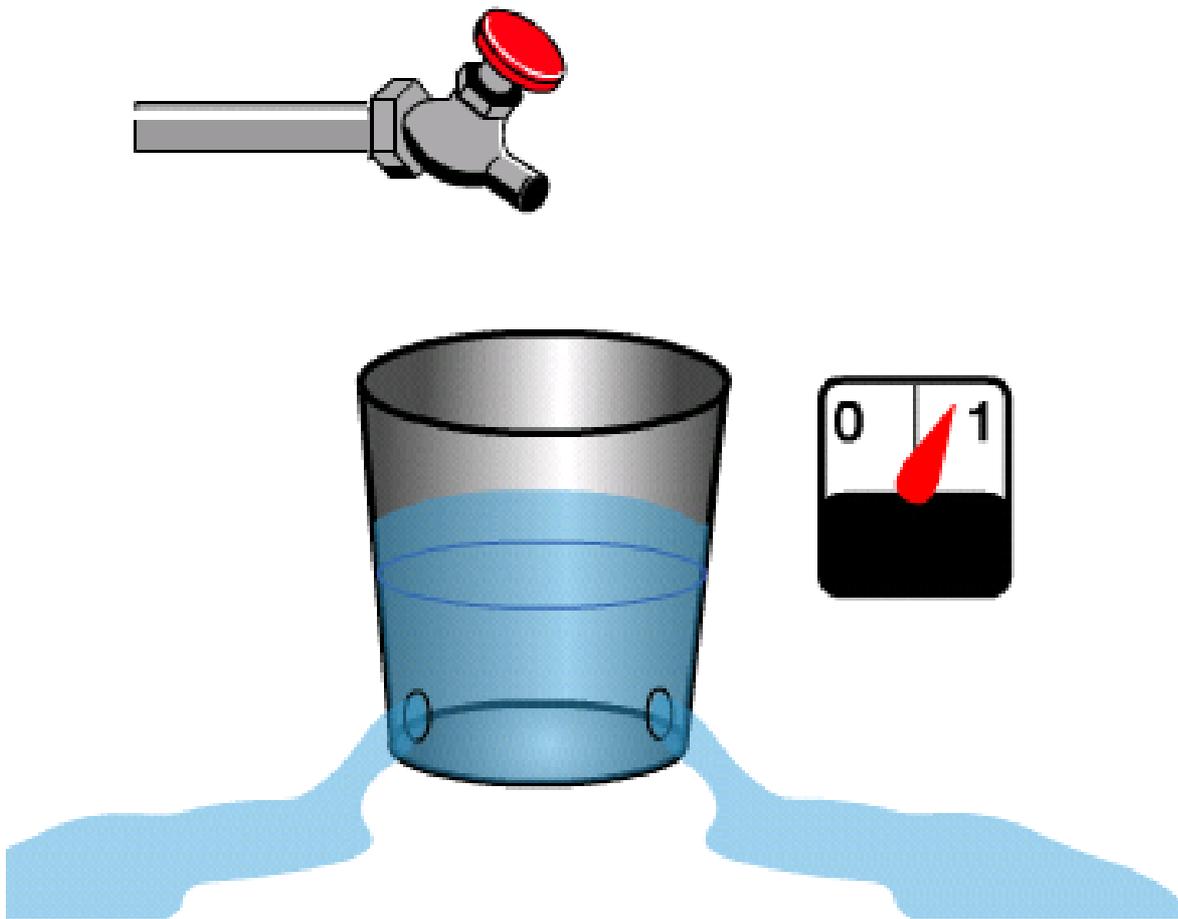


Le RAM (Random Access Memory) costituiscono la forma meglio conosciuta di memorie per computer. La Ram è considerata ad accesso casuale poiché si può accedere ad ogni cella di memoria direttamente. L'opposto delle Ram sono le SAM (Serial Access Memory). Le SAM immagazzinano i dati in una serie di celle di memoria cui si può accedere solo in sequenza come in un nastro di audiocassetta in cui si può accedere ad una canzone del nastro solo dopo aver fatto scorrere tutta la porzione precedente del nastro. Se il dato richiesto non è nella locazione corrente, viene controllata ogni cella di memoria seguente finché non viene trovato il dato richiesto. Le memorie SAM funzionano molto bene per i buffer di memoria, in cui i dati vengono immagazzinati normalmente nell'ordine in cui dovranno essere utilizzati (un buon esempio è il buffer di memoria di una scheda video). Dall'altro lato ai dati della RAM si può accedere in qualsiasi ordine.



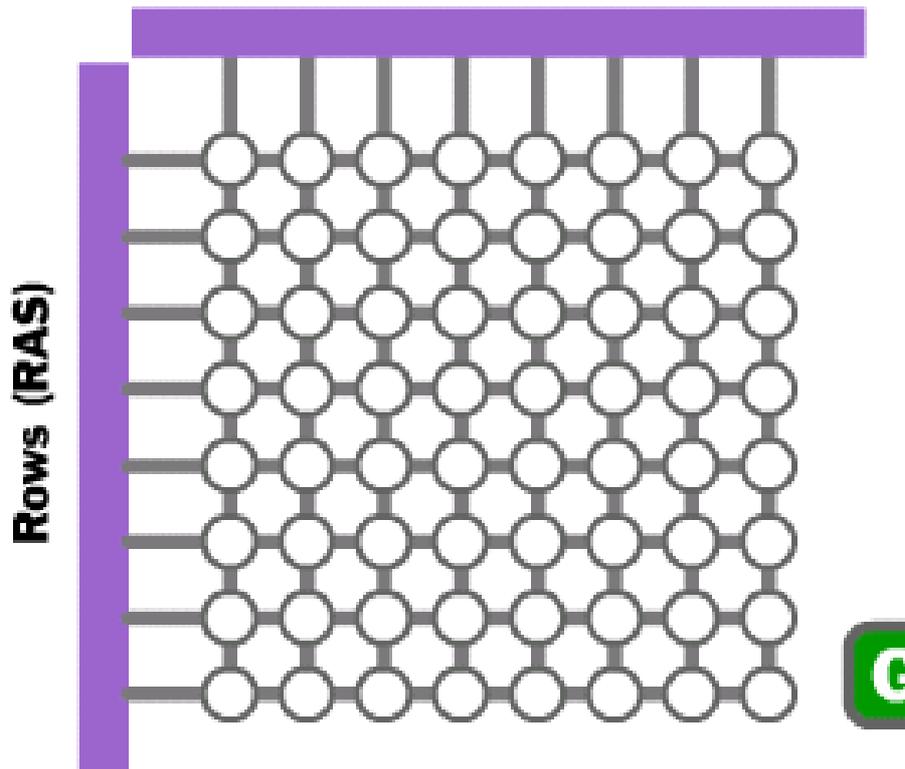
Un chip di memoria è un circuito integrato fatto di milioni di transistor e condensatori . Nella forma più comune di memorie per computer , le memorie ram dinamiche (DRAM Dynamic Random Access Memory) sono accoppiati un transistor e un condensatore per creare una cella di memoria, che rappresenta un singolo bit. Il transistor agisce come un interruttore che permette alla circuiteria di controllo esterna di leggere lo stato di tensione del condensatore cambiare il suo stato.

Un condensatore può essere visto come un piccolissimo secchio capace di immagazzinare elettroni. Per immagazzinare il bit 0, il secchio va svuotato. Il problema con questo tipo di memorie è che il "secchio perde". In pochi millisecondi la cella di memoria si scaricherebbe . perciò per poter lavorare con memorie dinamiche , la CPU o il controller di memoria deve ricaricare tutti i condensatori che immagazzinano un 1 prima che essi si scarichino. Per fare questo il controller di memoria legge la memoria e la riscrive . questa operazione di refresh avviene migliaia di volte al secondo.



IL nome di memorie dinamiche deriva da questa operazione. Le RAM dinamiche devono essere rinfrescate continuamente o perdono il loro contenuto. Il lato negativo sta nella lentezza che si introduce in queste memorie.

Le celle di memoria sono ricavate in un wafer di silicio in un array di colonne (linee dei bit) e righe (linee delle parole). L'intersezione di una colonna e di una riga costituisce l'indirizzo della singola cella di memoria



Le RAM dinamiche lavorano inviando una carica attraverso la colonna appropriata per attivare il transistor di ogni bit della colonna. In scrittura le linee di riga contengono lo stato che dovrà possedere ogni bit. In lettura un rilevatore determina il livello di carica della cella. Se esso è più del 50% del valore massimo possibile, si ha la lettura di un 1, altrimenti si ha la lettura di uno 0. Il tempo necessario per effettuare queste operazioni è dell'ordine dei nanosecondi.

Le celle di memoria da sole sarebbero inutili senza qualche mezzo per ottenere informazioni o immagazzinare informazioni. Così le celle hanno un'intera struttura di supporto di altri circuiti specializzati. Questi circuiti effettuano operazioni come

- identificazione di ogni riga e colonna (row address select e column address select).
- Tenere traccia della sequenza di rinfresco (contatore)

- Lettura e ripristino del segnale in ogni singola cella (sense amplifier)
- Ordine ad una cella di immagazzinare una carica o meno (write enable)

Altre funzioni del controller di memoria includono una serie di compiti che includono l'identificazione del tipo, velocità ed ammontare di memoria e controllo degli errori.

Le RAM statiche utilizzano una tecnologia completamente differente. In esse un flip flop contiene ogni bit di memoria. Un flip flop per ogni cella necessita di 5-6 transistor e qualche filo ma non ha bisogno dell'operazione di rinfresco. Questo rende le RAM statiche significativamente più veloci delle RAM dinamiche. Dall'altro lato esso occupa più spazio a causa della circuiteria più complessa. In sostanza abbiamo meno memoria per chip e molto più costosa. Per tale motivo le memorie statiche sono riservate alla realizzazione delle memorie cache e le RAM dinamiche per la realizzazione dei banchi di memoria RAM del computer.

I chip di memoria originariamente utilizzati in un desktop computer utilizzavano una configurazione dei pin detta dual in-line package (DIP). Questa configurazione consentiva la saldatura dei pin in fori nella motherboard o l'incastro in un socket saldato sulla scheda. Questo metodo lavorava bene quando i computer avevano tipicamente una coppia di megabyte di RAM, ma con la crescita della memoria il numero di chip da collocare sulla motherboard crebbe enormemente.

La soluzione fu quella di piazzare i chip di memoria insieme a tutti i componenti di supporto su di un circuito stampato separato (printed circuit board PCB) che poteva essere serrato in un connettore speciale (banco di memoria) sulla scheda madre. La maggior parte di questi chip usa una configurazione detta small outline J-lead (SOJ), ma pochi produttori utilizzano anche la configurazione del tipo thin small outline package (TSOP). La differenza chiave fra questi nuovi tipi di pin e il DIP è che SOJ e TSOP sono a montaggio superficiale. In altre parole i pin sono saldati direttamente sulla superficie della scheda e non inseriti in fori o socket.

I chip di memoria sono ora normalmente disponibili soltanto montati come parte di una card detta modulo. Avrete probabilmente sentito parlare di chip 8 X 32 o 4 X 16 . Questi numeri rappresentano il numero di chip moltiplicato la capacità di ogni singolo bit , misurata in megabit. Per esempio 4 X 32 significa che abbiamo 4 chip da 32 megabit che fa un numero totale di 128 megabit. Dividendo per otto abbiamo che si tratta di un chip da 16 megabyte.

Il tipo di schede e connettori per desktop computer si è evoluto negli ultimi anni. I primi tipi erano proprietari il che significa che ogni produttore di computer realizzava chip di memoria utilizzabili soltanto sui loro specifici sistemi. Poi vennero le SIMM (single in.line memory module). Questa scheda di memoria usava un connettore a 30 pin. Nella maggior parte dei computer si dovevano installare questi moduli a

coppie di uguale capacità e velocità, questo perché l'ampiezza del bus è più grande di quella di una singola SIMM.

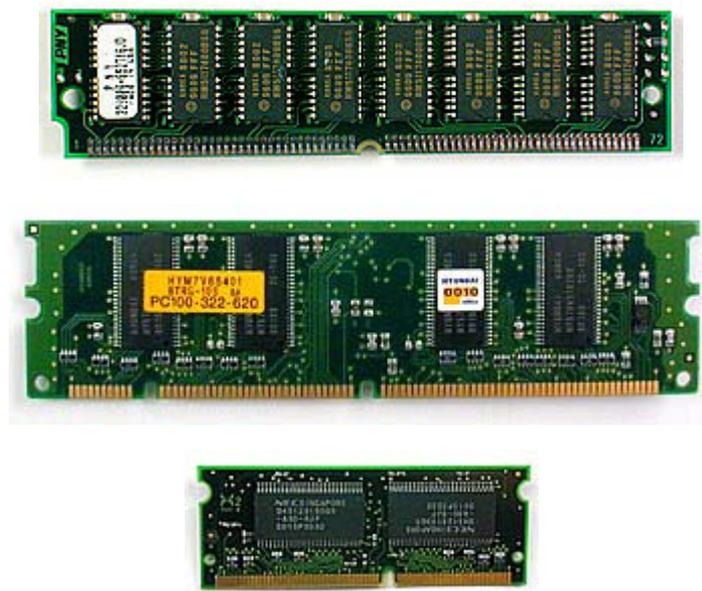


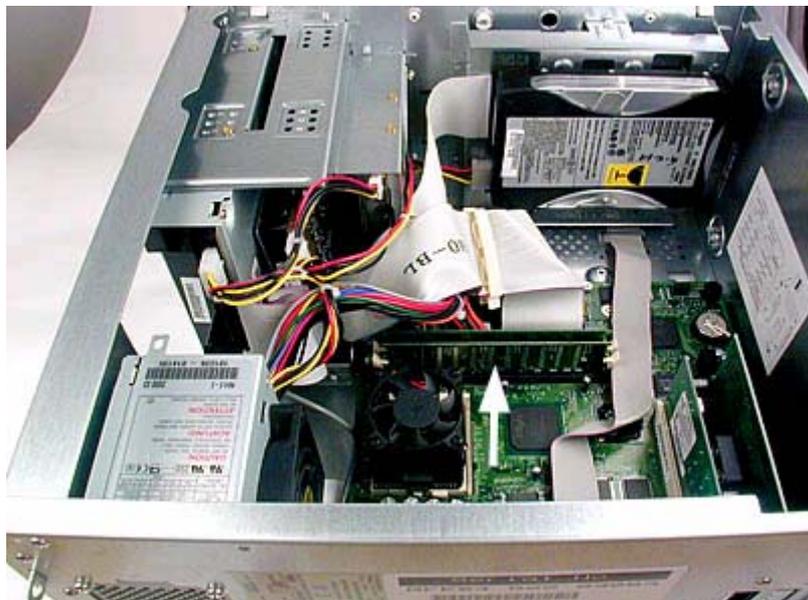
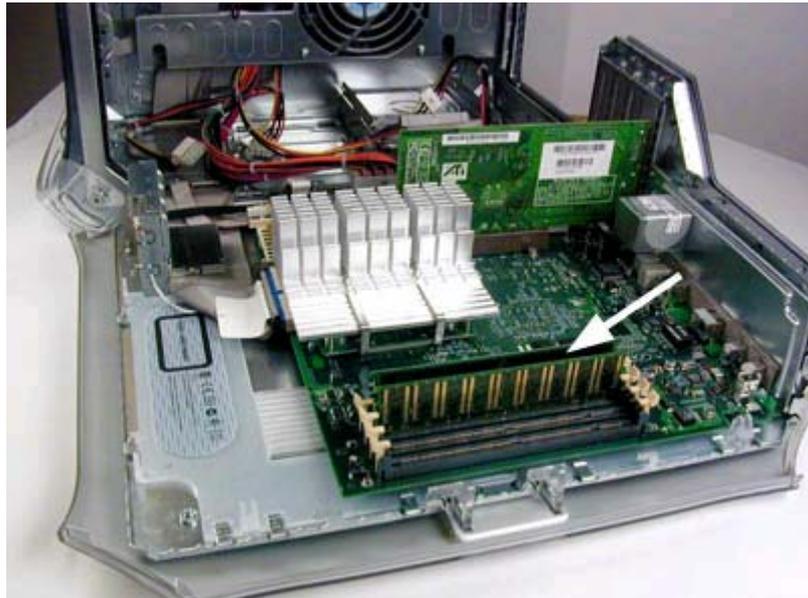
Tabella 1 dall'alto in basso: SIMM; DIMM e SODIMM

Con la crescita di velocità e ampiezza di banda dei processori, l'industria adottò il nuovo standard dual in-line memory module (DIMM) con un connettore di 168 o 184 pin. Le DIMM possono giungere alla capacità di 1 Gigabyte per modulo. Un altro standard Rambus inline module (RIMM) è comparabile in ampiezza e configurazione dei pin alle DIMM ma usa uno speciale bus di memoria per incrementare la velocità.

Molte marche di notebook usano moduli di memoria proprietari, ma diversi produttori usano Ram basate sulla configurazione small outline dual in-line memory module (SODIMM).

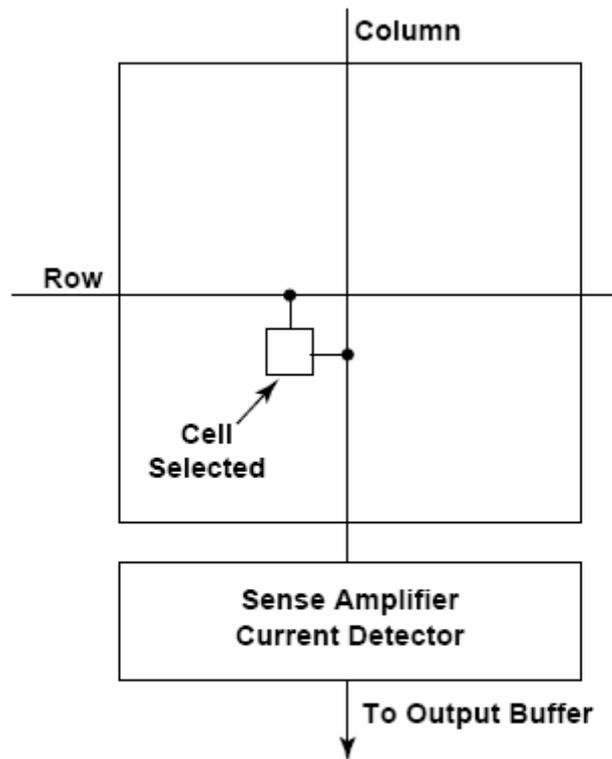
La maggior parte delle memorie attuali sono molto affidabili. La maggior parte dei sistemi hanno semplicemente un controllo sulla presenza di errori alla partenza e fanno completo affidamento su di esso. I chip di

memoria con controllo di errore integrato utilizzano tipicamente il controllo di parità. Il problema con il controllo di parità è che esso rileva gli errori ma non fa nulla per correggerli. Computer in ruoli critici abbisognano di livelli di sicurezza superiore per cui si utilizzano sistemi di correzione dell'errore (ECC Error Correction Code)



Rom

In figura vediamo lo schema di principio di funzionamento di una memoria a sola lettura

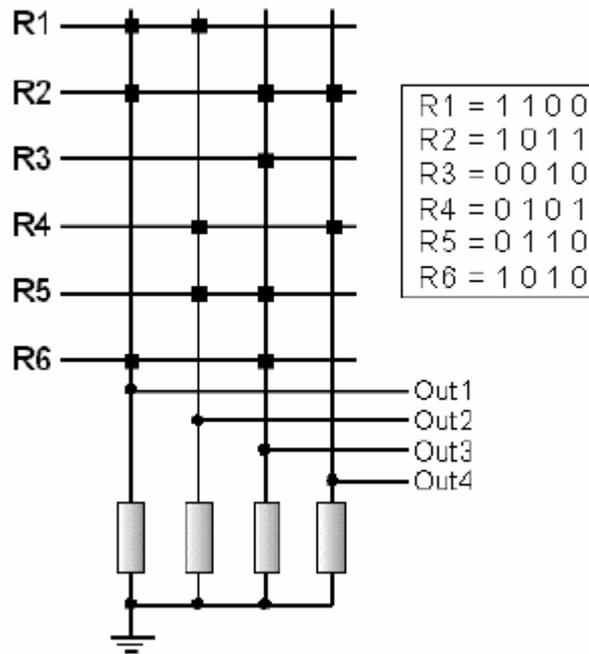


Source: ICE, "Memory 1997"

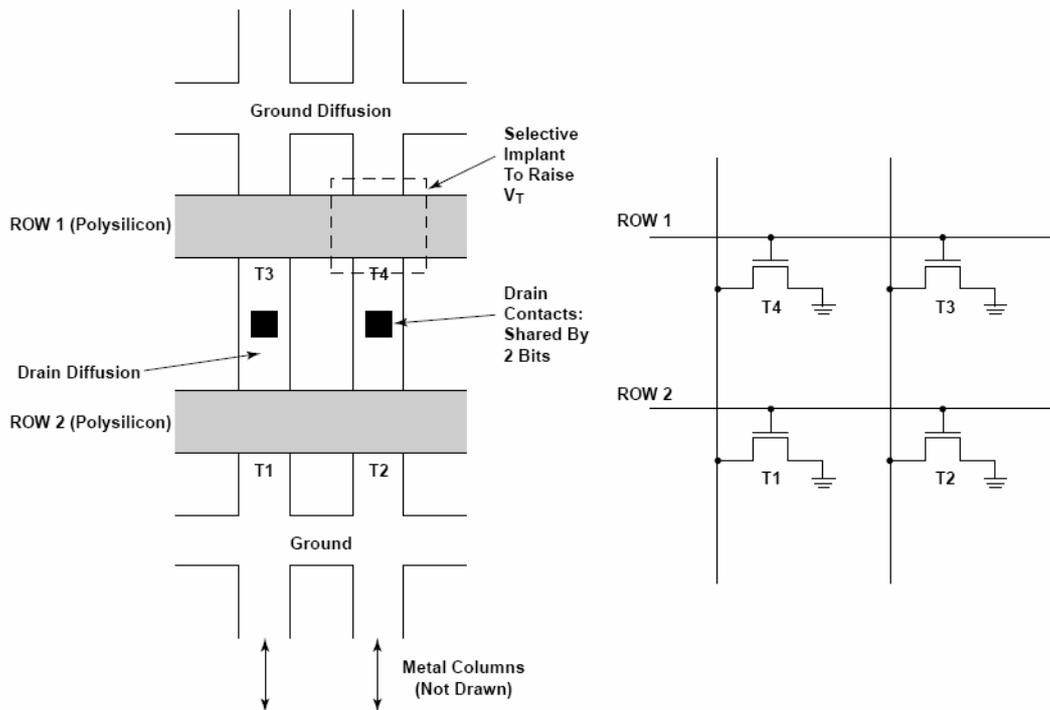
19956

Figure 9-1. Read Only Memory Schematic

La cella che unisce riga e colonna consiste di un solo transistor nelle memorie rom ed eprom e di due transistor nelle memorie eeprom. In fase di lettura viene imposta una tensione sul gate del transistor che forma la cella. In funzione della tensione di soglia programmata il transistor coinvolto consentirà o meno il passaggio di corrente. Un amplificatore "sente" la corrente o la mancanza di essa interpretandola come uno 0 o un bit ad 1.



In sostanza, come si può vedere meglio dallo schema precedente, se riga e colonna sono collegate si ha la memorizzazione di un 1 altrimenti di uno 0.

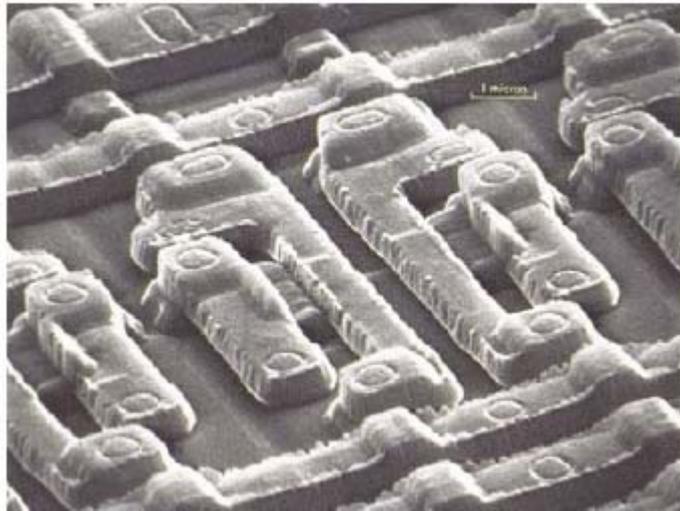


Source: ICE, "Memory 1997"

20845

Figure 9-2. ROM Programmed by Channel Implant

La figura precedente mostra una ROM programmata mediante impiantazione. Un transistor senza canale impiantato (transistor enhancement mode) ha una normale tensione di soglia per cui una tensione sul gate lo manda in conduzione. Un transistor con impiantazione di canale ha una tensione di soglia così elevata da superare la V_{cc} per cui risulta sempre interdetto.

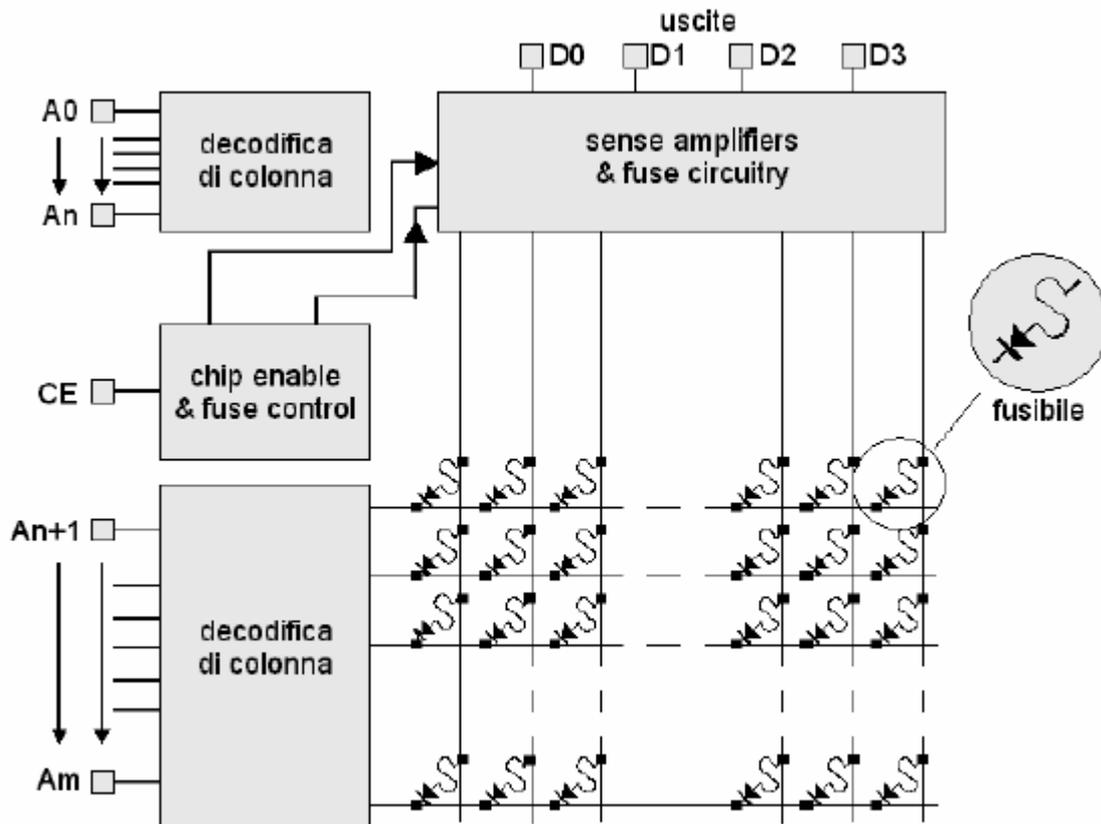


In fabbrica i nodi vengono ottenuti tramite "metallizzazione", ovvero interconnettendo opportunamente tramite sottilissime piste di alluminio (vedi figura a lato) una matrice di connessioni preesistenti. Le memorie di questo tipo vengono anche dette "mask-ROM", poiché la personalizzazione delle interconnessioni viene realizzata tramite "mascheratura", con un processo di tipo fotolitografico.

Prom

La PROM (acronimo di Programmable Read-Only Memory) è una tipologia di memoria informatica, in particolare una tipologia memoria scrivibile una sola volta a stato solido.

Nasce come evoluzione della ROM a maschera volta a ridurre i notevoli costi di produzione dovuti alla progettazione delle maschere per l'impiantazione ionica. La ROM a maschera infatti richiedeva di cambiare l'intera linea di produzione ogni qual volta fosse necessario modificare anche una minima parte dei circuiti logici.



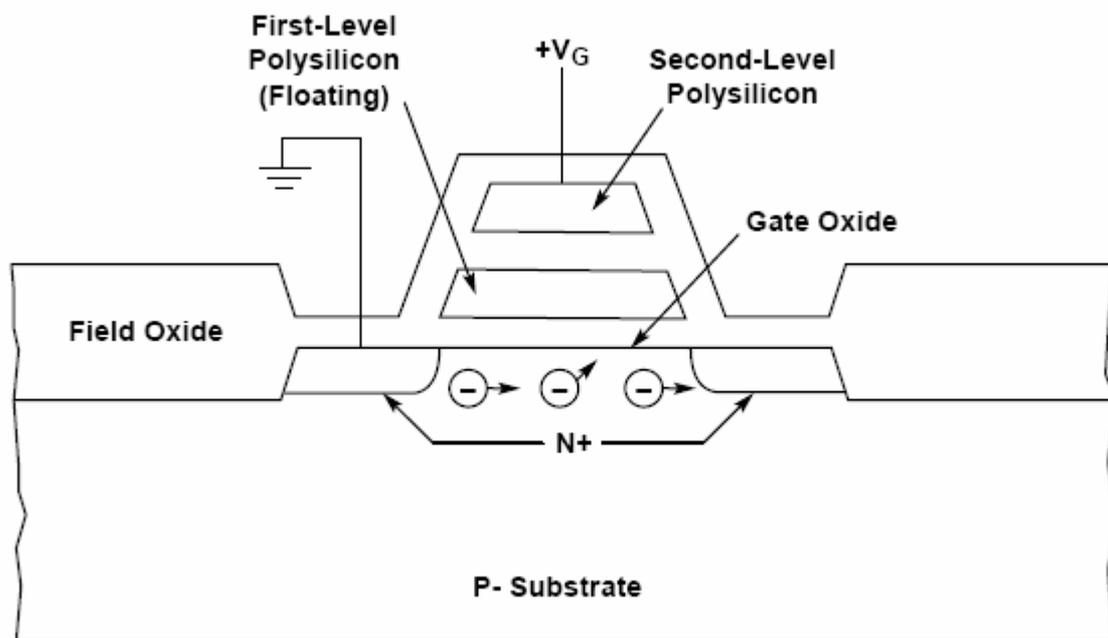
La PROM invece contiene dei fusibili che possono essere bruciati secondo le esigenze per creare i circuiti logici richiesti e richiede un'apparecchiatura speciale per le operazioni di scrittura. Sono state fatte però anche delle PROM dove invece dei fusibili venivano usati degli antifusibili nei quali il collegamento viene creato invece di venire bruciato.



EPROM

La EPROM, acronimo di Erasable Programmable Read Only Memory, è una memoria di sola lettura cancellabile tramite raggi ultravioletti. Si tratta di un'evoluzione della PROM che, una volta programmata, non poteva essere più modificata. La EPROM invece può essere totalmente cancellata (per un numero limitato ma consistente di volte) e riprogrammata a piacimento.

La scrittura avviene forzando una carica elettrica in un piccolo pezzo di materiale policristallino detto gate flottante localizzato nella cella di memoria. Usualmente la presenza di carica corrisponde ad uno zero logico e la sua assenza corrisponde ad un uno.

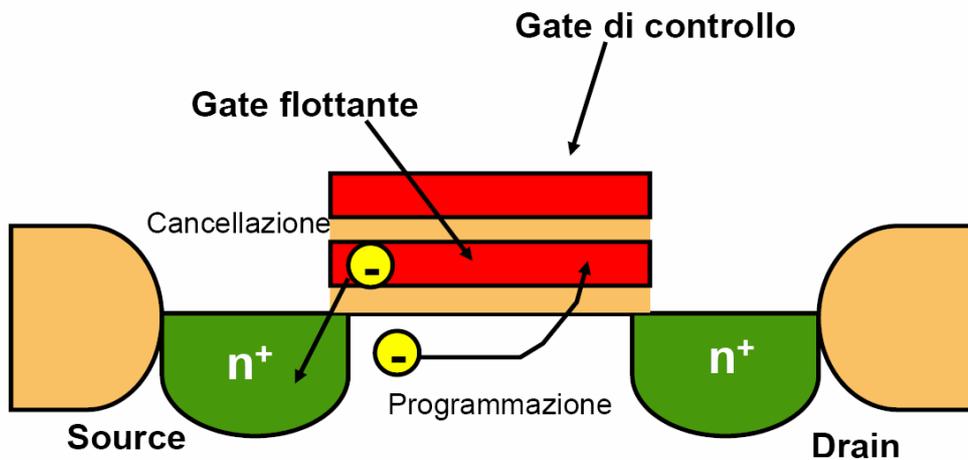


Source: Intel/ICE, "Memory 1997"

18474

La scrittura nelle celle di una memoria EPROM si realizza tramite un fenomeno di breakdown a valanga. Infatti viene fatta scorrere una forte corrente tra Drain e Source. Con una certa probabilità, alcuni elettroni "caldi" riescono a penetrare nel Gate flottante. La cella è programmata quando nel gate flottante è presente una carica tale da inibire l'accensione del dispositivo perché innalza la tensione di soglia oltre la V_{cc} . Infatti le cariche negative degli elettroni tendono ad attrarre lacune sotto il gate quindi diventa più difficile creare il canale.

Programmazione e cancellazione



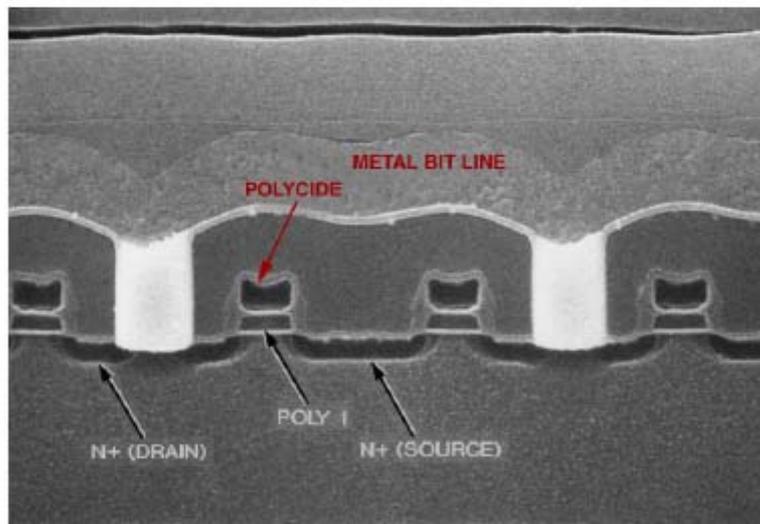
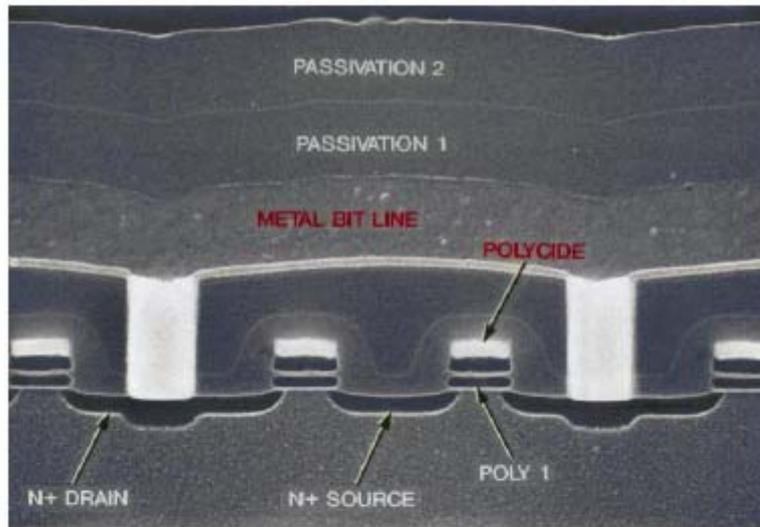
Per cancellare il dato si usano raggi ultravioletti, che danno energia agli elettroni e scaricano il gate (EPROM) oppure tensioni di segno opposto, che scaricano il gate (EEPROM). Nel caso di molte operazioni di scrittura, il passaggio di elettroni deteriora l'isolante tra il gate flottante e il gate o il canale. Per questo motivo dispositivi di memorizzazione EPROM o EEPROM possono essere scritti solo qualche migliaio di volte.

Per permettere l'irradiazione ultravioletta, nel package vi è una finestrella trasparente. Nelle prime versioni, la finestra veniva chiusa con una sottile lastrina di quarzo, in seguito fu utilizzato il semplice vetro, utilizzando la quantità necessaria a coprire come una lente il foro tondo ricavato nella metà superiore del package ceramico destinato a questo dispositivo. La cancellazione avveniva per mezzo di un attrezzo chiamato Eprom Eraser, una semplice scatola contenente una lampada a raggi UV. La cancellazione consisteva nel portare tutti i bit (celle o locazioni) a valore alto (1), la EPROM a questo punto tornava a essere programmabile. Le EPROM sono ormai da considerare obsolete e non vengono

praticamente più usate. Sono state sostituite dalle memorie flash o EEPROM che possono essere cancellate elettricamente senza dover ricorrere ai raggi UV.



Sono stati prodotti diversi tipi di EPROM che si differenziano - oltre che per la capacità di memoria disponibile, anche per la velocità (tempo di accesso) e per le tensioni di lavoro e di programmazione.

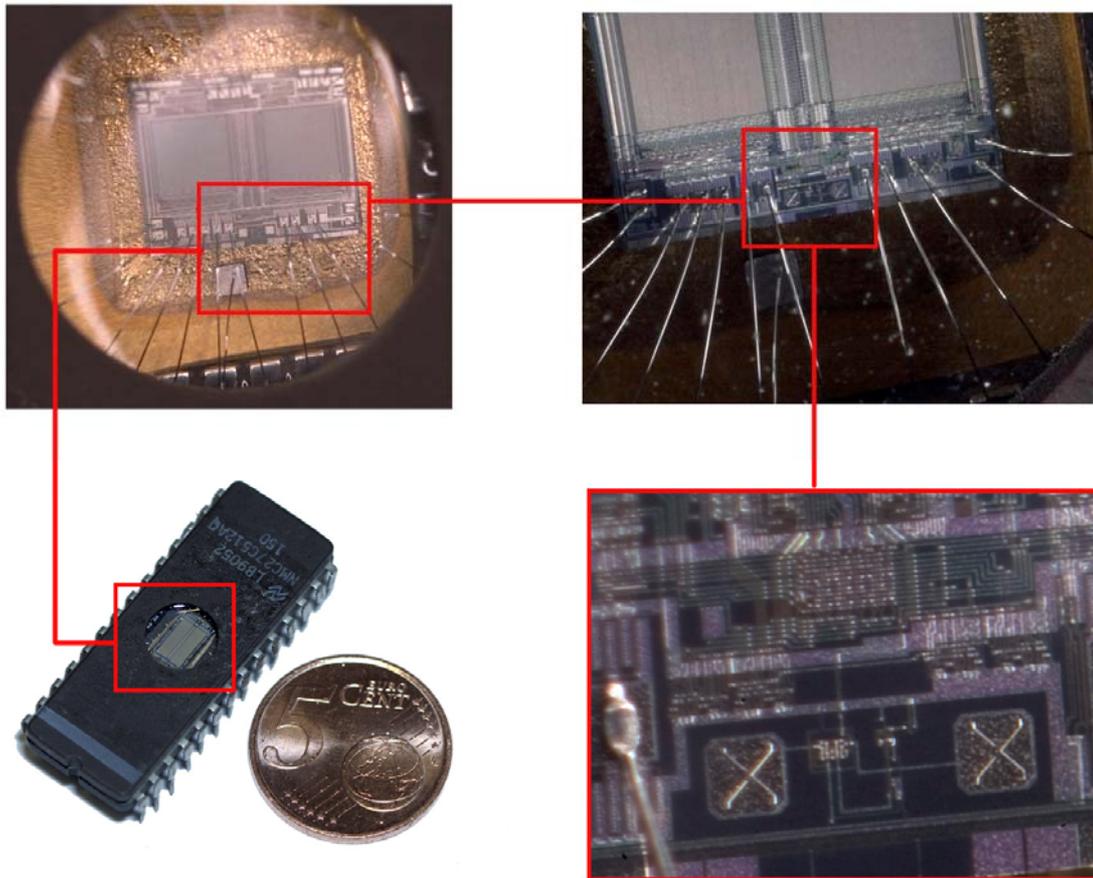


Source: ICE, "Memory 1997"

22462

Figure 9-8. Typical 1Mbit EPROM Cells

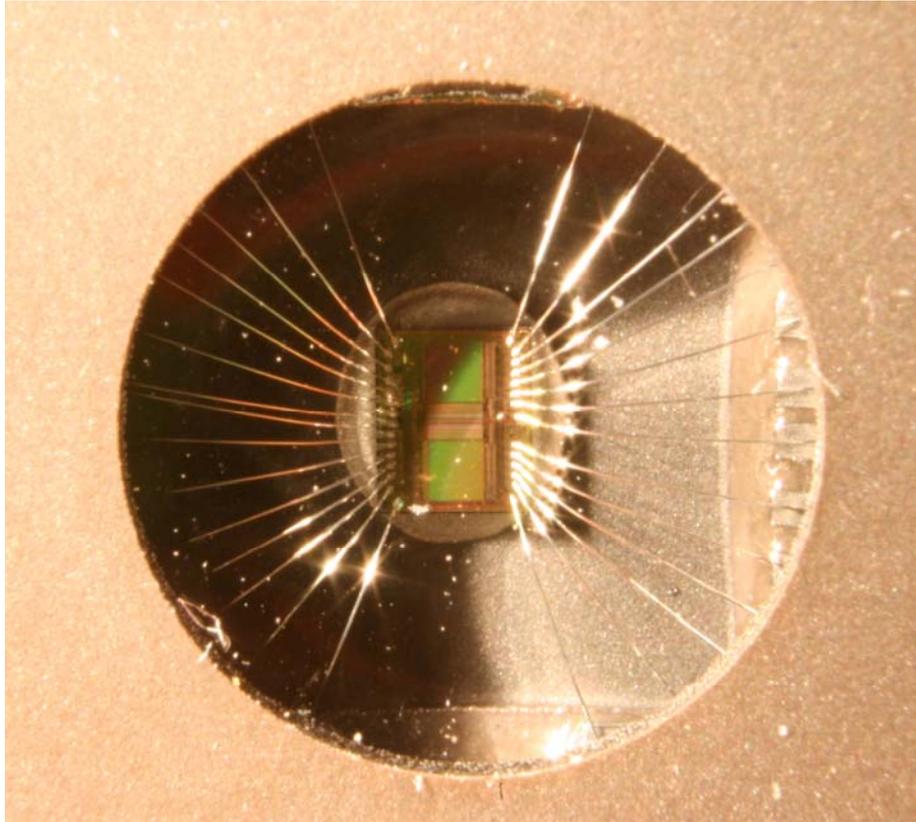
A partire dal chip con capacità di 8 Kbyte in poi, alcuni di questi tipi di memoria furono prodotti anche usando la tecnologia CMOS, consentendo una forte riduzione nel consumo di corrente a beneficio anche di una minore produzione di calore.



EEPROM

La EEPROM, acronimo di Electrically Erasable and Programmable Read Only Memory, è una memoria ROM programmabile da parte dell'utente. Le operazioni di scrittura, cancellazione e riscrittura hanno luogo elettricamente.

Ciascuna cella di memoria capace di memorizzare un singolo bit è costituita da due transistori MOS, uno "di memoria" e uno "di accesso".



La cella di memoria più comune è costituita da due transistor. Il transistor di memorizzazione ha un gate fluttuante simile a quello delle eprom che intrappolerà elettroni. Occorre notare che una cella di memoria eprom è cancellata quando gli elettroni vengono eliminati dal gate fluttuante mentre, inversamente, una cella eeprom viene cancellata quando gli elettroni vengono intrappolati nel gate fluttuante.

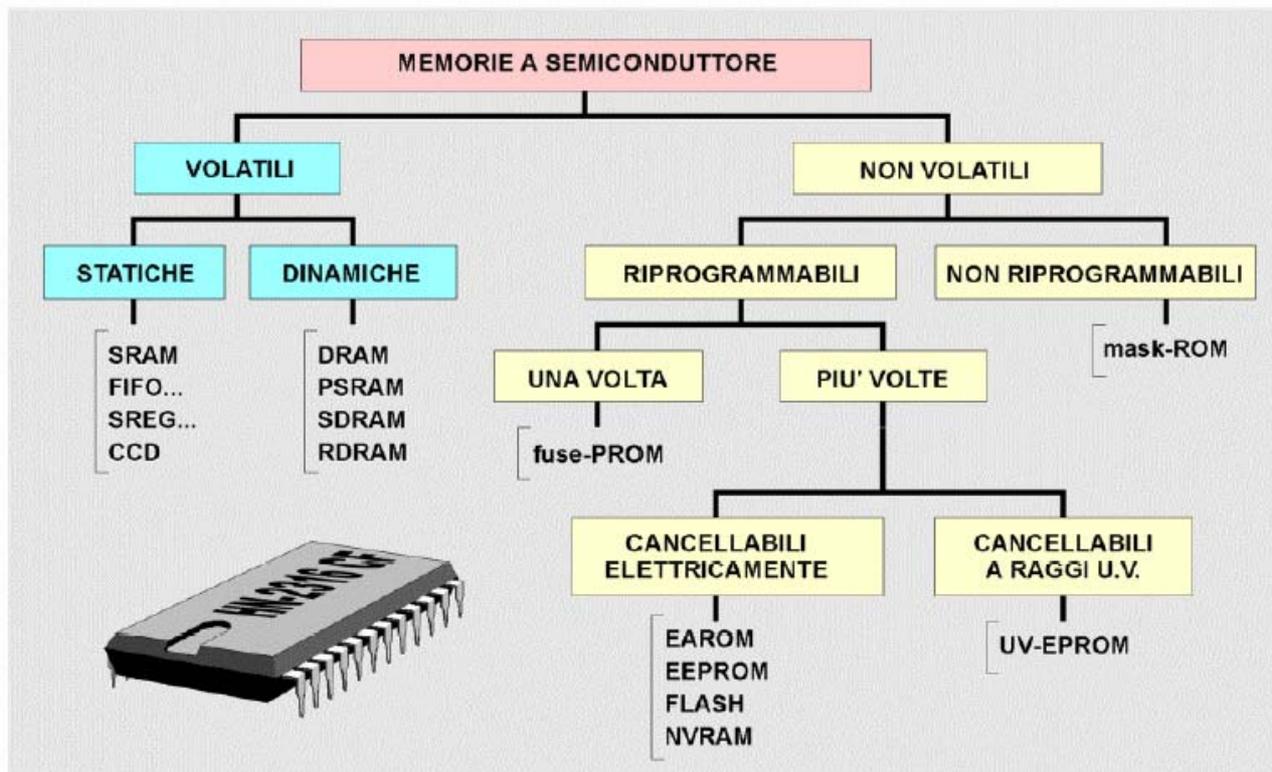
A differenza delle EPROM, nelle EEPROM vi è una regione, in prossimità del drain, in cui lo spessore dello strato di ossido che separa il floating gate dal canale è ridotto al punto tale da permettere il passaggio di elettroni per effetto tunnel (Fowler-Nordheim).

L'effetto tunnel è un effetto quanto-meccanico che permette una transizione ad uno stato impedita dalla meccanica classica.

Nella meccanica classica la legge di conservazione dell'energia impone che una particella non possa superare un ostacolo (barriera) se non ha l'energia necessaria per farlo. Questo corrisponde al fatto intuitivo che, per far risalire un dislivello ad un corpo, è necessario imprimergli una certa velocità ovvero cedergli dell'energia.

La meccanica quantistica invece prevede che una particella abbia una probabilità, piccola ma finita, di attraversare spontaneamente una barriera arbitrariamente alta. Infatti, applicando i postulati della meccanica quantistica al caso di una barriera di potenziale in una dimensione, si ottiene che la soluzione dell'equazione di Schrödinger all'interno della barriera è rappresentata da una funzione esponenziale decrescente. Dato che le funzioni esponenziali non raggiungono mai il valore di zero si ottiene che esiste una piccola probabilità che la particella si trovi dall'altra parte della barriera dopo un certo tempo t .

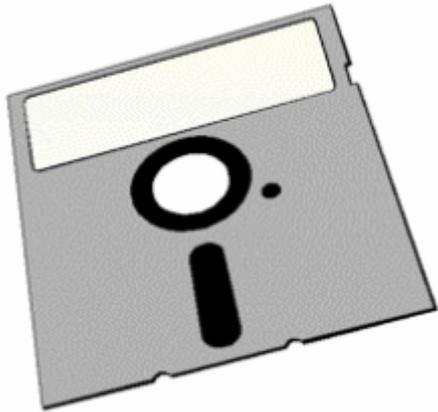
Il transistor di memoria viene programmato attraverso il pilotaggio in tensione del control gate. La variazione del potenziale a cui questo si trova esposto determina, nella zona in cui il ridotto spessore dell'ossido isolante lo rende possibile, il manifestarsi dell'effetto tunnel e la conseguente attrazione di elettroni dal drain al gate sepolto.



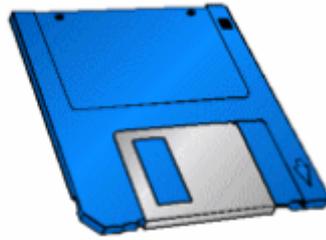
Memorie di massa magnetiche

Le memorie magnetiche - quali ad esempio le musicassette e le videocassette - memorizzano i dati sotto forma di differenti intensità di magnetizzazione dei micro-granuli di polvere di materiale ferromagnetico (generalmente ossido di ferro o di cromo) annegati all'interno del nastro di materiale plastico che fa da supporto. La magnetizzazione (scrittura) e la lettura del nastro avvengono mediante un trasduttore magnetico costituito da un avvolgimento di sottili fili di rame attorno ad un materiale magnetico dolce (testine). La scrittura delle tracce deve avvenire dopo la completa cancellazione della precedente magnetizzazione mediante un'apposita testina

FLOPPY-DISC



5 pollici - 360 Kbyte



3½ pollici - 1.4 Mbyte

Nei dischi floppy il supporto è costituito da un disco di materiale plastico morbido (da cui la denominazione di "floppy") in cui è annegata una sottile polvere di materiale ferromagnetico, mentre negli hard-disk il materiale ferromagnetico è depositato sotto forma di sottilissimo strato sulla superficie di un disco metallico.

In un hard-disk - proprio per consentire elevate capacità di memorizzazione - si utilizzano non solo entrambe le facce del disco, ma anche più dischi sullo stesso asse, ciascuno con due testine.

I dati sono generalmente memorizzati su disco seguendo uno schema di allocazione fisica ben definito in base al quale si può raggiungere la zona dove leggere/scrivere i dati sul disco. Uno dei più diffusi è il cosiddetto CHS acronimo per il termine inglese Cylinder/Head/Sector (Cilindro/Testina/Settore); in questa struttura i dati sono memorizzati avendo come indirizzo fisico un numero per ciascuna delle seguenti entità fisiche:

Piatto

un disco rigido si compone di uno o più dischi paralleli, di cui ogni superficie, detta "piatto" e identificata da un numero univoco, è destinata alla memorizzazione dei dati.

Traccia

ogni piatto si compone di numerosi anelli concentrici numerati, detti tracce, ciascuna identificata da un numero univoco.

Cilindro

l'insieme di tracce alla stessa distanza dal centro presenti su tutti i dischi è detto cilindro. Corrisponde a tutte le tracce aventi il medesimo numero, ma diverso piatto.

Settore

ogni piatto è suddiviso in settori circolari, ovvero in "spicchi" radiali uguali ciascuno identificato da un numero univoco.

Blocco

L'insieme di settori posti nella stessa posizione in tutti i piatti.

Testina

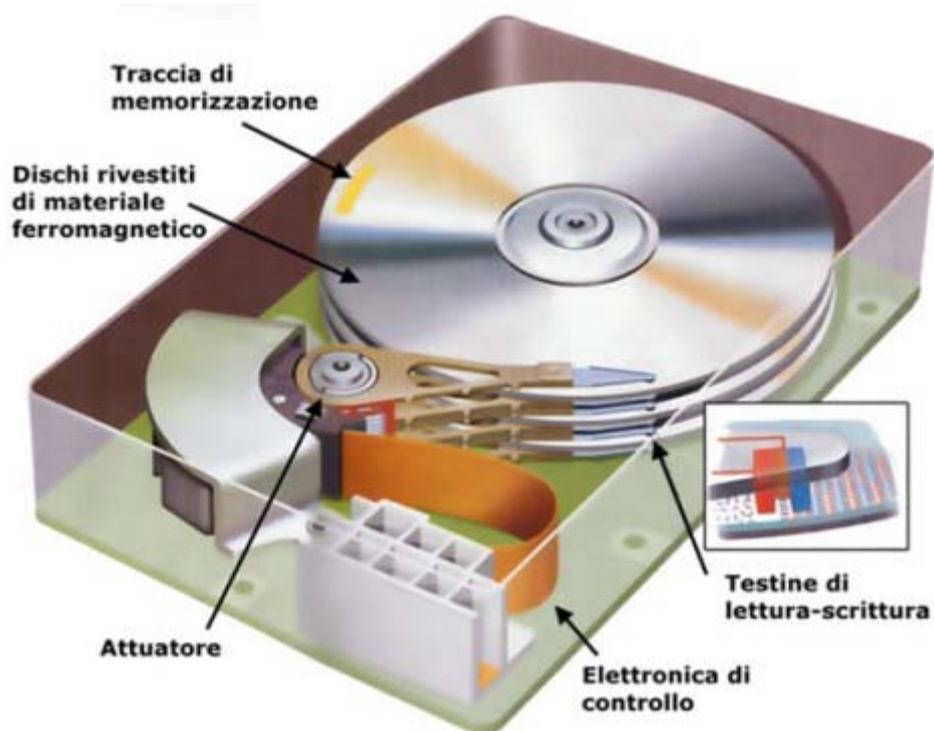
Su ogni piatto è presente una testina per accedere in scrittura o in lettura ai dati memorizzati sul piatto; la posizione di tale testina è solidale con tutte le altre sugli altri piatti. In altre parole, se una testina è posizionata sopra una traccia, tutte le testine saranno posizionate nel cilindro a cui la traccia appartiene.

Questa struttura introduce una geometria fisica del disco che consta in una serie di "coordinate" CHS, esprimibili indicando cilindro, testina,

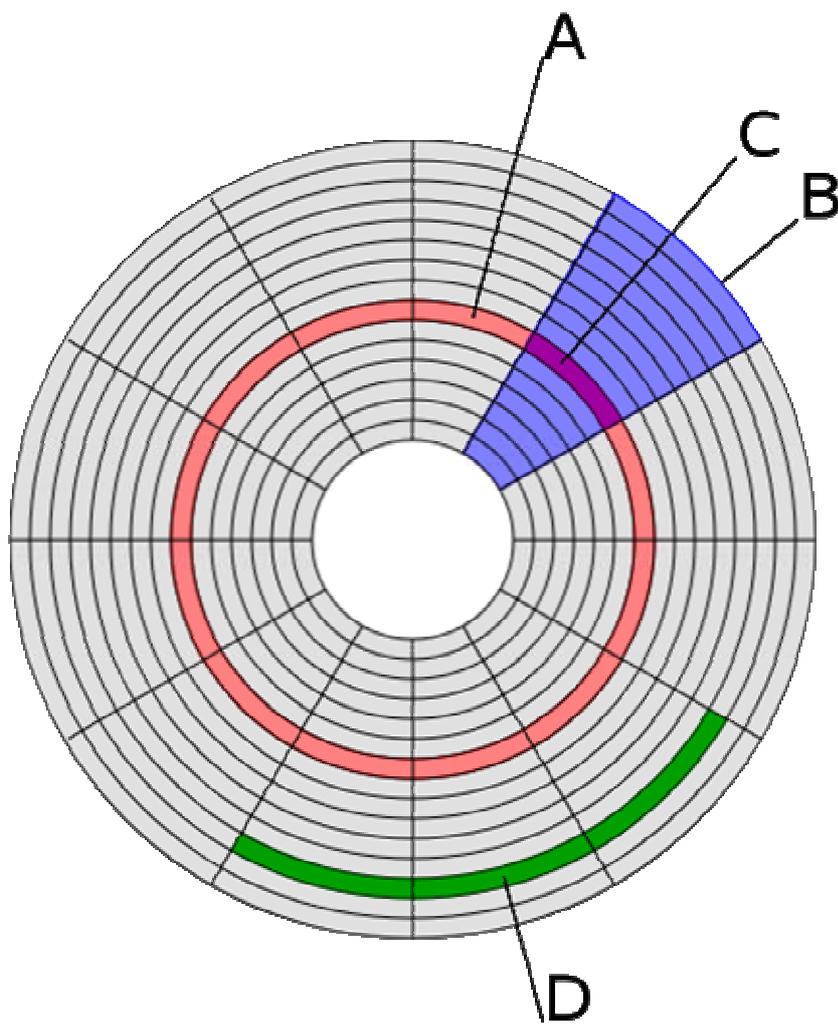
settore. In questo modo è possibile indirizzare univocamente ciascun blocco di dati presente sul disco. Ad esempio, se un disco rigido si compone di 2 dischi (o equivalentemente 4 piatti), 16384 cilindri (o equivalentemente 16.384 tracce per piatto) e 16 settori di 4096 byte per traccia, allora la capacità del disco sarà di $4 \times 16384 \times 16 \times 4096$ byte, ovvero 4 GiB.

Il fattore di interleaving è il numero dei settori del disco rigido che si deve saltare per leggere consecutivamente tutti quelli della traccia. Ciò dipende strettamente dalle caratteristiche prestazionali del disco rigido stesso, cioè dalla velocità di rotazione del disco, dal movimento dei seekers con le relative testine e dalla velocità di lettura-scrittura della stessa testina.

Tale processo è stato introdotto poiché inizialmente le cpu, che ricevevano e rielaboravano i dati letti, compivano queste azioni ad una velocità inferiore della velocità di lettura/scrittura sul disco rigido, quindi, una volta rielaborati i dati provenienti da un settore, la testina si troverebbe già oltre l'inizio del settore successivo. Alternando i settori in modo regolare e leggendoli secondo lo specifico interleaving factor, si velocizzava il disco rigido e il calcolatore. I moderni dischi rigidi non necessitano di interleaving.



Le memorie ottiche (Compact-Disk o CD e Digital Versatile Disk o DVD) sono quelle introdotte più recentemente sul mercato (anni '90) e si sono potute imporre solo grazie all'impiego del laser (la "luce" che viene impiegata per la lettura e la scrittura del supporto) e delle tecniche di "digitalizzazione" del suono. In altri termini, mentre le memorie di tipo magnetico possono essere di tipo sia analogico che digitale, le memorie ottiche sono esclusivamente di tipo digitale.



Struttura della superficie di un piatto:

A) Traccia

B) Settore

C) Settore di una traccia

D) Cluster, insieme di settori contigui



Nelle memorie a disco ottico la memorizzazione avviene effettuando dei piccolissimi "fori" sulla superficie di un sottile strato di alluminio annesso all'interno di un materiale plastico che fa da supporto del disco stesso. I minuscoli fori vengono realizzati vaporizzando l'alluminio tramite impulsi di luce laser all'infrarosso, generata dalla testina di scrittura del "masterizzatore", che è il nome del drive in grado di scrivere i CD.

