

PIERO OLIVO

EVOLUZIONE DELLE MEMORIE A SEMICONDUTTORE

Nel preparare questa chiacchierata ho considerato che la maggior parte dell'uditorio non fosse esperto di elettronica e di memorie in particolare. Pertanto, piuttosto che fare un'analisi dettagliata dei problemi delle memorie, ho pensato di iniziare con una carrellata su quello che è stato lo sviluppo dell'elettronica negli ultimi sessanta anni.

Nella prima parte ci soffermeremo sull'evoluzione dei sistemi di calcolo per l'elaborazione dei dati e poi entreremo più nel dettaglio dell'argomento "memorie", per arrivare a trattare, negli ultimi minuti, alcuni problemi di ricerca sui quali lavoriamo presso l'Università di Ferrara.

Lo sviluppo dell'elettronica per l'elaborazione delle informazioni ha una data fondamentale: il 1947, praticamente al termine della seconda guerra mondiale. Ed è proprio il periodo bellico che vede non solo il grande sviluppo dell'elettronica, ma anche quello delle sue figlie ovvero l'informatica e le telecomunicazioni. La guerra rappresenta, sfortunatamente, il vero traino all'evoluzione del campo dell'elettronica. Tutte le grandi rivoluzioni in questo ambito nascono per scopi militari e solamente dopo molti anni trovano applicazione in ambito civile. Per esempio, durante la seconda guerra mondiale si sono sviluppate in modo straordinario tutte le telecomunicazioni. Un'altra esigenza fondamentale, il calcolo delle tabelle di tiro dell'artiglieria, porta negli anni '40 alla partenza di importanti programmi per ricerca per la realizzazione di calcolatori elettronici in grado di sostituire i sistemi elettromeccanici.

Nel 1947 viene presentato il primo calcolatore elettronico, denominato ENIAC (FIG. 6.1.). Il suo sviluppo, iniziato nel corso della guerra, viene completato nel 1947. Questo calcolatore presentava 500.000 connessioni, saldate insieme, 6.000 interruttori, e circa 18.000 valvole: l'utilizzo di questo elemento elettronico rappresenta la grande novità!

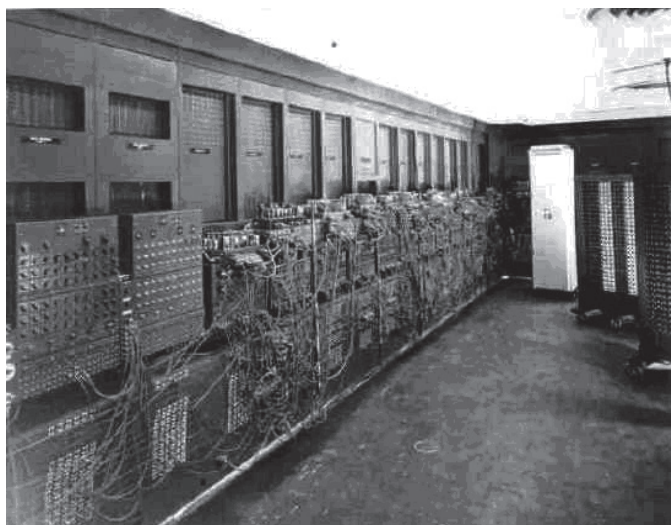
L'ENIAC è il primo calcolatore che adopera degli elementi elettronici al suo interno. Il tubo a vuoto (o valvola), utilizzato per le telecomunicazioni, viene quindi adoperato anche per l'elaborazione dei dati. Le sue dimensioni erano molto grandi: 24 metri di lunghezza per 1,5 metri di larghezza e 2,5 metri di altezza, con un peso di circa 30 tonnellate. Riusciva ad eseguire 300 moltiplicazioni al secondo, prestazione dieci volte superiore a quelle delle precedenti macchine puramente meccaniche.

I problemi dei primi calcolatori erano fondamentalmente quattro:

- il costo;
- le dimensioni;
- il consumo di potenza;
- l'affidabilità.

Quest'ultima è legata al numero di componenti, soprattutto valvole, che si bruciavano con un ritmo altissimo. Nel caso dell'ENIAC quando, dopo qualche anno di utilizzo, si riuscì ad ottenere un intervallo di tempo di due giorni tra un guasto e l'altro si trattò di una grande conquista: si trattava di un tempo abbastanza lungo se si pensa che all'inizio passavano solo poche ore prima che qualcosa si inceppasse.

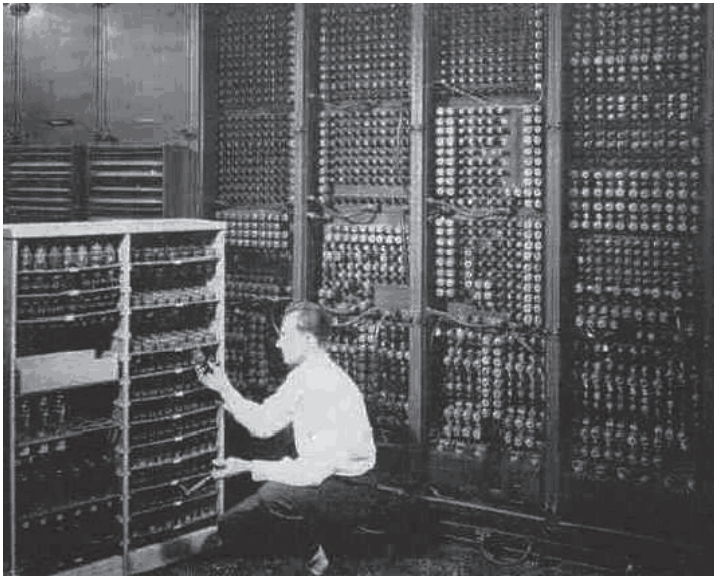
FIG. 6.1. ENIAC, il primo calcolatore elettronico



Il calcolatore ENIAC veniva programmato modificando delle connessioni al suo interno: il programmatore doveva cambiare le connessioni all'interno della macchina per poter eseguire le elaborazioni (Fig. 6.2). Al contrario dei calcolatori attuali, l'ENIAC non aveva ancora una memoria nella quale "salvare" il programma da eseguire. Alcuni tubi a vuoto erano utilizzati per il salvataggio di alcuni dati parziali generati durante i calcoli. I risultati finali delle elaborazioni venivano trasferiti all'utente attraverso schede perforate ma non vi era possibi-

tà, in alcun modo, di memorizzare l'informazione, e ciò rappresentava un grande problema e un grande limite!

FIG. 6.2. Programmazione dell'ENIAC



Prima di proseguire con l'evoluzione dei calcolatori, vediamo un confronto tra il calcolatore ENIAC del 1947 e il processore Intel "Santa Rosa", quello che troviamo nei PC portatili odierni. Si è passati da 5.000 addizioni al secondo a circa 18 miliardi di operazioni, con numeri decimali, al secondo. Per quanto riguarda la memoria, l'ENIAC aveva la possibilità di avere circa 200 byte per calcoli di risultati parziali. I calcolatori attuali hanno circa 2 miliardi di byte di memoria da utilizzare durante l'esecuzione di un programma e circa 200 miliardi di byte di memoria permanente, una memoria cioè in cui il dato può essere lasciato in maniera stabile e duratura. Per quanto attiene agli elementi elettronici si passa da 18.000 tubi a vuoto a più di 1 miliardo di transistori, presenti in un quadrato di silicio di poco più di 1 centimetro di lato. Le dimensioni sono quelle sopra riportate per l'ENIAC, contro quelle di un normale portatile. Si passa inoltre da 30 tonnellate a 2 kg di peso.

Facendo un confronto con il mondo dell'aviazione civile, una quarantina di anni fa, quando sono cominciati i primi voli diretti tra l'Italia e gli Stati Uniti, per raggiungere New York si impiegavano circa

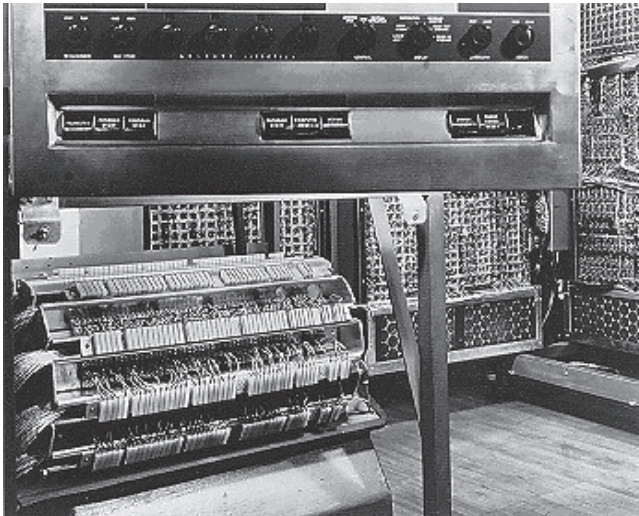
8-9 ore e il costo del volo era di circa 700-800 dollari. Nel 2007 per volare a New York dall'Italia si impiegano ugualmente 8-9 ore e il costo del volo è sempre di 700-800 dollari. Se lo sviluppo nel mondo dell'aeronautica fosse stato pari a quello dell'elettronica, al giorno d'oggi si potrebbe raggiungere New York in pochi secondi e al costo di qualche centesimo di euro. Questo per far comprendere come nel giro di 50-60 anni l'evoluzione dell'elettronica sia stata straordinaria.

Nello sviluppo dei sistemi di elaborazione diventava sempre più necessario memorizzare l'informazione e inizialmente si è pensato a una memorizzazione di tipo magnetico. Quello riportato in FIG. 6.3. è un insieme di tamburi magnetici, messi a punto dall'IBM, per memorizzare l'informazione in maniera permanente.

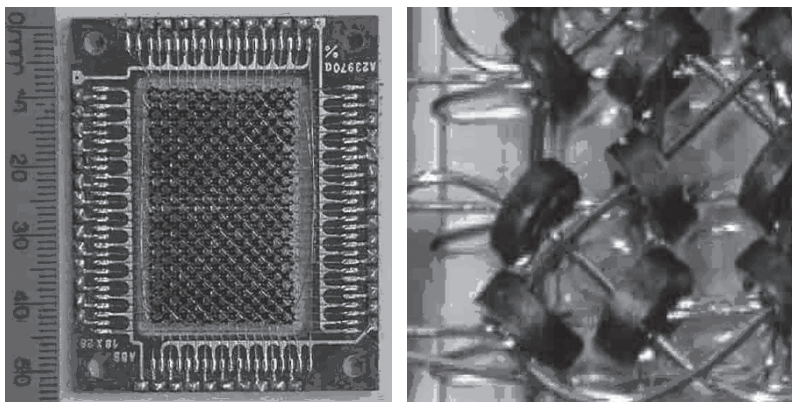
Successivamente si è cercato di creare anche delle memorie più piccole da poter utilizzare all'interno di calcolatori. Nelle FIGG. 6.4. e 6.5. è mostrata una memoria a nuclei magnetici. Siamo sempre negli anni '50 e in questa memoria troviamo degli elementi di materiale magnetico attraversati da fili di corrente, utilizzati per sfruttare il fenomeno dell'isteresi.

Ci troviamo di fronte a qualche centinaio di elementi di memoria in un'area che è di circa $5 \times 3 \text{ cm}^2$. Sempre per mettere in evidenza qual è stata l'evoluzione nel campo delle tecnologie elettroniche, nei sistemi elettronici attuali, come ricordavo in precedenza, in un centimetro quadrato trovano posto oltre 1 miliardo di transistori.

FIG. 6.3. Esempio di memorie a tamburi



FIGG. 6.4. e 6.5. Memorie a nuclei magnetici. In FIG. 6.4. è possibile osservare anche la scala di riferimento. In FIG. 6.5. è riportato un particolare



In questo ambito la prima grande rivoluzione ha avuto luogo sempre nel 1947. Nel corso di quell'anno, infatti, presso i laboratori Bell, negli Stati Uniti, viene inventato il transistor. Il transistor è un elemento elettronico che permette di amplificare il valore di corrente sfruttando le proprietà di alcuni materiali particolari: i semiconduttori.

Il semiconduttore più importante è il silicio, che è un materiale molto presente in natura. Il silicio ha delle caratteristiche particolari che fanno sì che, in condizioni neutre, sia un cattivo conduttore (lascia cioè passare una debole corrente elettrica). Se però si riesce a sostituire un atomo di silicio ogni circa dieci, cento milioni di atomi, con atomi di altre specie, come ad esempio boro, fosforo o arsenico, il materiale muta straordinariamente le proprie proprietà elettriche e da un pessimo conduttore diventa un buon conduttore.

Il grande vantaggio del transistor rispetto alla valvola è che ha delle dimensioni molto più piccole, è molto più affidabile, ha dei costi molto più bassi e consuma molto meno. Quindi i quattro problemi fondamentali dei primi calcolatori elettronici vengono nettamente migliorati grazie allo straordinario sviluppo dell'elettronica al silicio.

Il transistor è stato utilizzato in principio per amplificare la corrente, in particolare per applicazioni legate alle telecomunicazioni. Il transistor, però, può essere utilizzato per un altro scopo: il transistor può essere anche essere utilizzato come interruttore, che fa passare o non fa passare la corrente.

A questo punto è necessario fare un accenno a cosa significhi "elettronica digitale" in quanto tutta l'elettronica, tranne alcuni cam-

più specifici, è diventata digitale, cioè che elabora dei numeri. Apriamo quindi una parentesi sull'elettronica digitale: quando si scrive o si legge un numero, per esempio 624, la nostra mente esegue un calcolo, un algoritmo. Questo algoritmo, che ci viene insegnato nei primi anni della scuola elementare, noi lo eseguiamo in modo automatico, senza rendercene conto: il numero 624 significa che ci sono 6 centinaia, 2 decine e 4 unità. Abbiamo a disposizione dieci cifre, dallo 0 fino al 9, e se si vuole rappresentare un numero più elevato si riutilizzano le stesse 10 cifre e le si riadattano dando un loro significato che viene definito "posizionale", nel senso che, se si scrive 624, il 6 indica le centinaia perché è la terza cifra partendo da destra, il 2 indica le decine perché è la seconda cifra, e il 4 indica le unità perché è la cifra più a destra.

La numerazione decimale, da noi utilizzata, non è l'unica utilizzabile in ambito matematico. Esiste, per esempio, anche la possibilità di utilizzare la numerazione binaria. Secondo tale numerazione si hanno a disposizione solo due cifre, lo zero e l'uno, e si può rappresentare un qualunque numero esattamente come lo rappresentiamo secondo la numerazione decimale. Per esempio, se scrivessimo il numero binario 1101, utilizzeremmo solo due cifre, sempre attribuendo loro un diverso peso a seconda della posizione. La cifra più a destra rappresenta l'unità, poi abbiamo un'altra cifra con peso 2, poi una cifra con peso 4 (2×2), e poi ancora con peso 8 ($2 \times 2 \times 2$) e così via con potenze di 2. La cifra binaria 1101 equivale pertanto al numero decimale 13 ($1 \times 8 + 1 \times 4 + 0 \times 2 + 1 \times 1$).

È quindi possibile rappresentare qualunque numero decimale con la numerazione binaria.

Il termine bit, molto utilizzato in ambito informatico, altro non è che è l'acronimo inglese che sta per *binary digit*, ovvero cifra binaria.

Perché la numerazione di tipo binario è così importante nel campo dell'elettronica?

Quando si ha a che fare solo con le due cifre 0 e 1, noi possiamo associare a queste cifre la presenza o meno di un fenomeno fisico facilmente individuabile, ovvero: c'è o non c'è, è bianco o nero, è aperto o chiuso, passa corrente o non passa corrente.

Quindi il transistor può essere utilizzato anche come interruttore: qualora "faccia passare corrente" possiamo considerare la cifra "1", qualora "non faccia passare corrente", la cifra "0". Quindi diventa molto semplice utilizzare transistori per fare elaborazioni di tipo numerico, valutando o meno la presenza di corrente.

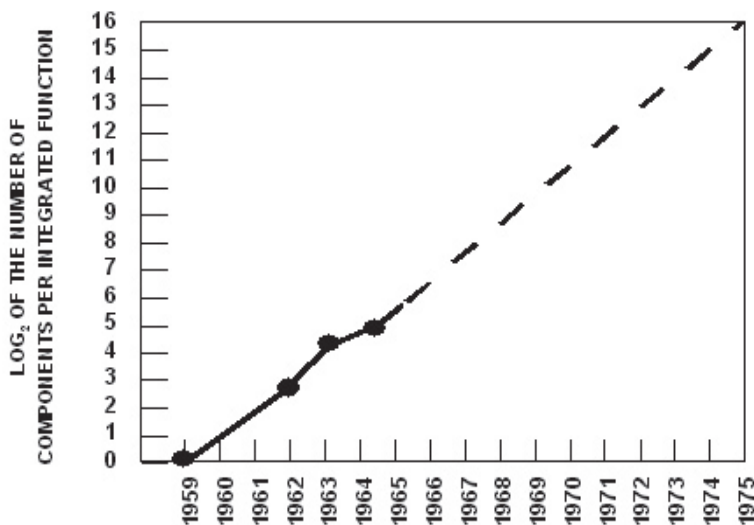
Proseguiamo nell'evoluzione dei sistemi di calcolo: nel 1958 è stato creato il primo circuito integrato. Esso è costituito da un pezzettino di silicio dove non troviamo un solo transistor, ma tanti transistori

che vengono connessi tra loro in modo da svolgere una data funzione logica o operazione matematica.

Nel '60 è stata presentata la prima "famiglia logica", cioè il primo insieme di circuiti logici che eseguono diverse funzioni logiche. Diventano quindi disponibili componenti commerciali da assemblare per realizzare dei sistemi digitali più complessi.

In FIG. 6.6. è riportata una curva elaborata nel 1965 da uno dei fondatori di una ditta californiana, diventata poi importantissima: l'Intel. Questo scienziato, di nome Moore, aveva riportato in un grafico quattro punti sperimentali che rappresentavano il numero di transistori che, di anno in anno, era possibile integrare nello stesso pezzo di silicio. I dati mostrano che, dal '59 al '65, ogni due anni si aveva un raddoppio del numero di transistori che poteva essere messo sullo stesso pezzo di silicio.

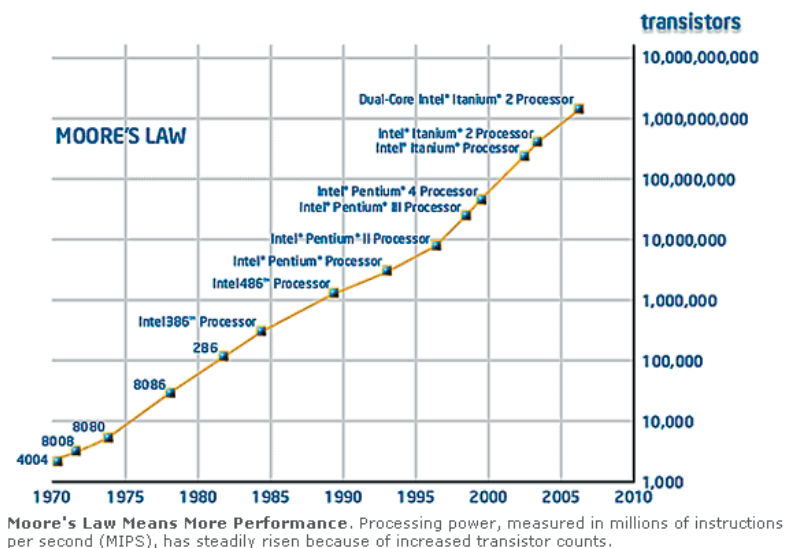
FIG. 6.6. Figura originale del lavoro di Moore



Moore osservò che la linea che univa i 4 punti era praticamente una retta. Facendo una proiezione, che si fermava al '75, ipotizzò che lo stesso andamento sarebbe proseguito almeno per una decina di anni. Aveva effettivamente ragione perché, se si va a guardare quella che è tuttora l'evoluzione delle dimensioni e del numero di transistori presenti in un circuito, si trova una retta che mostra un raddoppio circa ogni due anni.

La FIG. 6.7. mette in evidenza l'incremento del numero dei transistori che possono essere inseriti in un singolo circuito integrato.

FIG. 6.7. Incremento nel tempo del numero dei transistori integrati in un singolo circuito integrato. Tale evoluzione segue le previsioni di Moore del 1965



Seconda rivoluzione: siamo nel 1971, sempre all'Intel. Viene realizzato il primo microprocessore, un circuito integrato al cui interno ci sono tutti gli elementi per poter eseguire dei calcoli (tra i suoi inventori c'è anche l'italiano Faggin). Questo microprocessore aveva 2000 transistori al suo interno, faceva le sue elaborazioni a soli 4 bit, quindi un numero piccolissimo di cifre, ma aveva già una memoria al suo interno: una "memoria a semiconduttore".

Questo nuovo circuito aveva, dunque, la possibilità di memorizzare dati direttamente su una memoria a semiconduttore.

Ma perché sono così importanti le memorie nei sistemi elettronici?

Le tre motivazioni principali per cui noi abbiamo bisogno di memorizzare dati sono:

1) memorizzare dei risultati: perché quello che si fa non vada perso, senza avere la necessità di doverlo scrivere su un pezzo di carta o su schede. Dunque per la necessità di avere una permanenza dei ri-

sultati. Fino a qualche anno fa questa operazione veniva eseguita solamente su dischi magnetici: gli hard disk dei computer sono un classico esempio di supporto magnetico, all'interno del quale noi memorizziamo dei dati in modo permanente.

2) memorizzare dei dati parziali durante l'elaborazione: mentre si eseguono dei calcoli si ha bisogno che, all'interno del sistema si calcolano, alcuni dati siano memorizzati, così come quando facciamo un conto a mano scriviamo su un foglio di carta un risultato parziale per non dimenticarlo. Allo stesso modo un calcolatore ha al suo interno dei dati parziali;

3) memorizzare, all'interno del sistema, codici o istruzioni che devono essere eseguite, ovvero il programma.

Nei primi PC venduti all'inizio degli anni '80 queste memorie avevano densità (numero di bit memorizzabili) piccolissime e costavano moltissimo. La grande differenza tra un calcolatore e l'altro era proprio la quantità di memoria che rappresentava, dunque, l'elemento fondamentale che ne determinava il costo.

Fino a qualche anno fa la memorizzazione permanente veniva eseguita solamente su dischi magnetici: gli hard disk dei computer sono un classico esempio di supporto magnetico in cui è possibile "salvare" dati in modo permanente. Da una decina di anni, anche la memorizzazione permanente può essere eseguita su memorie a semiconduttore. Esempi di questo tipo di memorie sono le penne USB, le memorie delle fotocamere digitali o delle telecamere, alcune memorie degli MP3 utilizzati per ascoltare musica. Queste sono memorie che devono essere necessariamente a semiconduttore perché devono essere piccole, costare poco, essere veloci, e poter essere inserite all'interno dei sistemi di calcolo.

Vediamo nel dettaglio quali sono le caratteristiche fondamentali che distinguono una memoria da un'altra.

Il primo aspetto è la cella di memoria che è l'elemento che memorizza il bit.

La memorizzazione del dato può avvenire in modo circuitale o fisico. Nel primo caso avviene realizzando dei circuiti in maniera tale da permettere il mantenimento dell'informazione. Questa modalità però è una classica operazione da ingegnere progettista di circuiti che consiste nel connettere dei transistori in modo opportuno. La memorizzazione di un dato può altresì avvenire in modo fisico, cioè sfruttando dei fenomeni fisici presenti nei semiconduttori.

In questo ambito vi è una ricerca continua nel campo della fisica, della chimica, dell'elettronica, per scoprire delle nuove alternative, per trovare delle nuove proprietà per creare la cella, che deve essere

sempre più piccola, avere delle prestazioni migliori, essere più veloce, costare meno ed essere più affidabile.

In realtà esistono diversi problemi che riguardano la singola cella. Un aspetto è legato al consumo di potenza, cioè quanta energia richiede la cella per funzionare, valore che è in genere trascurabile. Detto così, non si comprende perché il problema del consumo di potenza sia così importante. In realtà è necessario tenere conto che in una memoria ci sono miliardi di celle (basti pensare alle memorie per una macchina fotografica digitale, nelle quali il numero di celle è ormai superiore a diversi miliardi). L'energia richiesta per il funzionamento della memoria dipende quindi dall'energia richiesta dalla singola cella moltiplicata per il numero delle celle! Ci si rende facilmente conto che diventa importantissimo creare celle che consumino sempre meno energia, dato che il numero delle celle che vengono realizzate nella stessa memoria è sempre crescente.

Leggendo le pubblicazioni scientifiche si osserva come vengano presentate continuamente nuove celle, nuove soluzioni. Il problema è che una soluzione, per essere adattata, non dipende solamente dalle sue prestazioni ma dipende anche dalla facilità con cui viene realizzata dal punto di vista tecnologico e dalla possibilità di fare operare miliardi di celle nello stesso pezzetto di silicio. Anche in questo caso, il problema non è, dunque, la singola cella, ma è come organizzare e far lavorare insieme miliardi di celle. Diventa un problema di organizzazione, ovvero di "architettura delle memorie".

L'architettura delle memorie è un parametro fondamentale, che tiene conto del modo in cui sono organizzate le memorie.

Gestire una memoria è come gestire la posta in una città: il dato da memorizzare equivale alla lettera da consegnare. È fondamentale, quindi, che sia disponibile un indirizzo della cella in cui memorizzare un dato (e dove andarlo a riprendere, quando questo deve essere "letto" e quindi utilizzato).

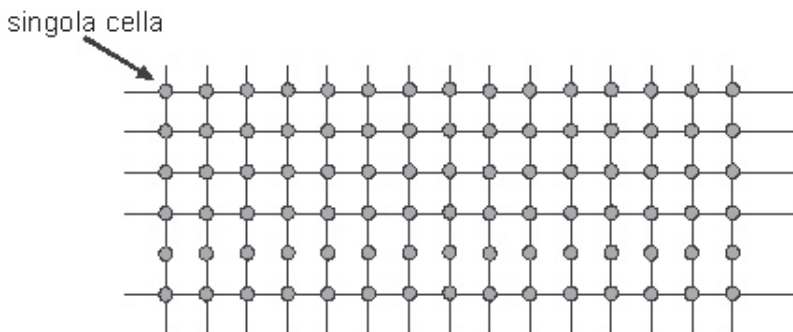
La cella di memoria può essere pensata come se fosse casa e il postino deve sapere dove andare portare la lettera. Nello stesso modo, il sistema elettronico che gestisce la memoria deve sapere l'"indirizzo" della cella (cioè dove il dato va memorizzato e prelevato).

Tralasciando le caratteristiche dei sistemi che gestiscono gli indirizzi, focalizziamo l'attenzione sull'organizzazione delle celle. Le celle sono organizzate come in FIG. 6.8., in un modo che potrebbe essere paragonato, in termini non tecnici, alla città di Manhattan, con le sue street e le sue avenue che corrispondono ad un'organizzazione a righe e colonne tipiche di una memoria a semiconduttore. Chi si trova a Manhattan e deve utilizzare un taxi non dà all'autista un indirizzo

come farebbe qua in Italia, ma gli fornisce un incrocio tra una street ed una avenue.

Nel campo delle memorie accade esattamente la stessa cosa: ogni singola cella è posizionata in modo univoco dal suo indirizzo di riga o di colonna.

FIG. 6.8. *Rappresentazione schematica della disposizione delle celle di memoria in una struttura a righe e colonne*



Vi sono però degli aspetti legati a questo tipo di struttura che evidenziano i problemi di questa organizzazione.

Come già osservato, si cerca di avere delle celle sempre più piccole. D'altro canto, per diminuire le dimensioni fisiche della memoria ed aumentare la densità della memoria (cioè il numero di celle), è necessario mettere le celle sempre più vicine fra loro ma, più vicine sono tra loro le righe e le colonne, più probabili sono i disturbi.

Sempre mantenendo una analogia con una città, se le strade sono molto vicine tra di loro e in una strada passano delle macchine, le macchine danno fastidio alle case presenti nella strada ma ne risentono anche le abitazioni nelle strade vicine.

Se si considera che le memorie sono dei circuiti in cui viaggiano dei segnali elettrici, è possibile immaginare che se le celle sono troppo vicine si disturbano a vicenda (perdita di dati, memorizzazione di dati errati ecc.).

Quindi celle più piccole e memorie più dense vogliono dire righe e colonne più vicine fra di loro e questo vuol dire disturbi ed il problema dei disturbi è uno dei principali temi di ricerca nel campo delle architetture delle memorie.

Un altro aspetto molto importante è la velocità di lettura.

La velocità di lettura rappresenta il tempo necessario per avere

la disponibilità di un dato memorizzato. Questo tempo naturalmente deve essere il più piccolo possibile perché le prestazioni del sistema dipendono proprio da quanto è veloce la memoria a restituire il dato. Questa proprietà dipende dalle caratteristiche della cella e dell'architettura. Inoltre è importante che questo tempo non dipenda da dove è posizionata la cella all'interno della memoria. Ritornando all'esempio di New York, se uno si trova all'80^a strada e deve scendere alla 5^a strada, impiega di certo molto tempo. Se invece dall'80^a deve scendere alla 75^a, naturalmente ne impiegherà molto meno! Nel campo della memoria questo non deve succedere: il tempo dopo il quale il dato è disponibile deve essere sempre lo stesso. È per questo che si parla di memorie RAM, acronimo di *random access memories* (memorie ad accesso *random* ovvero casuale), che vuol dire che, dovunque sia posizionata l'informazione, il tempo in cui si rende disponibile è sempre lo stesso. Anche questo è un problema legato all'architettura delle memorie.

Esistono poi altri tre parametri importanti: velocità di scrittura, *endurance* e gestione della ridondanza. La velocità di scrittura è data dal tempo impiegato a scrivere un dato. L'*endurance* è il numero di operazioni di scrittura effettuabili sulla stessa cella. In ultimo vi è la gestione della ridondanza. La ridondanza rappresenta la ruota di scorta della memoria. In effetti la probabilità che una memoria funzioni correttamente fin dall'inizio è quasi pari a zero. Si parla pur sempre di miliardi di celle e la probabilità che tutte funzionino contemporaneamente è quasi nulla. Viene quindi realizzato un certo numero di righe e di colonne in soprannumero che vengono utilizzate per sostituire quelle che funzionano e per evitare di buttare via tutti i pezzi che non funzionano correttamente al termine della produzione, il che comporterebbe dei costi che diventerebbero troppo elevati.

Un altro parametro importante è la "volatilità" della memoria.

La volatilità è la capacità di mantenere il dato anche quando il sistema è spento, ovvero non è alimentato. Le memorie tradizionali sono memorie "volatili". L'informazione permane solamente se la memoria è accesa. Quando si spegne il computer i dati che sono sulla memoria RAM vengono persi. È fondamentale, dunque, avere memorie "non volatili", cioè memorie all'interno delle quali il dato rimane anche in assenza di alimentazione. Un tipico esempio di memorie non volatili è dato dalle memorie magnetiche (vedi hard disk dei PC). È però fondamentale avere anche memorie a semiconduttore che siano non volatili, in modo da poter essere utilizzate in sistemi molto compatti (memorie per fotocamere digitali, lettori MP3, penne USB ecc.). Inoltre avere memorie non volatili a semiconduttore permette di non avere parti in movimento: uno dei grossi problemi dei dischi magnetici sta nel fatto

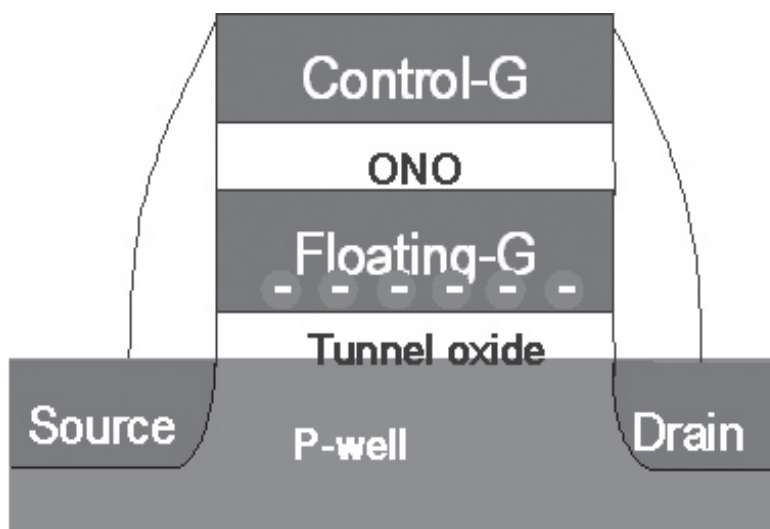
che avendo parti in movimento è possibile creare dei danni che possono rovinare il disco. Inoltre essi richiedono un maggiore consumo di energia. Di fatto, se fosse possibile sostituire tutti i dischi magnetici con memorie non “volatili” si risolverebbero molti problemi di affidabilità dei sistemi.

Infine, l’ultimo parametro ed il più importante di tutti, è il costo.

Il successo o l’insuccesso di un prodotto dipende solamente dal suo costo.

L’ultima parte di questa presentazione è dedicata alle “memorie mlash”. Le “memorie mlash” hanno rappresentato, a partire da metà degli anni ‘90, il componente elettronico con il più alto tasso di crescita. Hanno cominciato ad essere presentati negli anni ‘90 e adesso rappresentano circa il 50% del mercato delle memorie.

FIG. 6.9. Sezione schematica di una cella flash



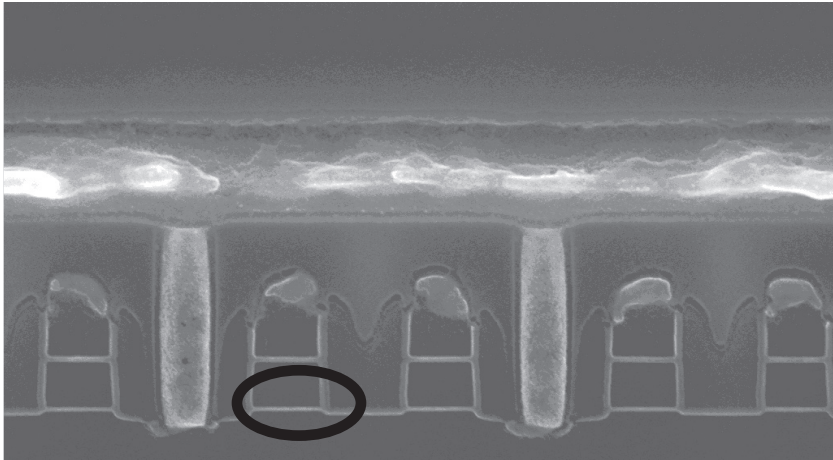
Le “memorie flash” sono memorie non volatili: permettono una memorizzazione di tipo permanente, garantita per dieci anni. In FIG. 6.9. è riportata in modo schematico la sezione di una cella flash. È importante tenere conto delle dimensioni della cella: le tecnologie più avanzate permettono di avere distanze tra le regioni indicate come *source* e come *drain* di alcune decine di nanometri, cioè di alcune decine di milionesimi di metro. Si può avere pertanto un’idea delle dimensioni

dell'intera cella e delle difficoltà tecnologiche per la sua realizzazione. Il funzionamento di queste celle dipende solamente dalla capacità di riuscire a trasferire alcuni elettroni dalla regione indicata come *p-well* a quella indicata come *floating-G*, passando attraverso un sottilissimo strato di isolante. La presenza di elettroni in questa regione denominata *floating-G* modifica drasticamente le caratteristiche elettriche della cella. Senza entrare nel dettaglio del funzionamento fisico della cella, vale la pena ricordare che per memorizzare un bit in una cella di memoria è sufficiente riuscire a trasferire qualche centinaio di elettroni, dunque un numero piccolissimo. La difficoltà non è solo riuscire a trasferirveli, ma è anche quella di farveli rimanere per tempi molto lunghi (fino a 10 anni!), altrimenti si perde l'informazione, si perdono i dati.

Esistono pertanto grossissimi problemi legati alla fisica e all'affidabilità di queste memorie che sono dovuti anche alle difficoltà di garantire la permanenza dei dati, definita tecnicamente come *data retention*.

In FIG. 6.10. è riportata una sezione di alcune celle ottenute mediante microscopio elettronico. La regione cerchiata in nero rappresenta il sottile strato di isolante che gli elettroni devono attraversare per poter "memorizzare" un "1". L'area occupata dalla cella è di circa 10 femtometri quadri, cioè un milionesimo di miliardesimo di metro quadro!

FIG. 6.10. Sezione di celle flash osservate al microscopio elettronico



In precedenza si accenna agli aspetti economici del mercato delle memorie. In FIG. 6.11. sono riportate le previsioni del mercato delle

“memorie flash”. In particolare va evidenziato il mercato previsto per le memorie cosiddette “NAND”, utilizzate per memorizzare i dati in modo permanente in lettori MP3, macchine fotografiche, penne USB. Per tali memorie le previsioni di mercato per il 2010 si aggirano attorno ai 40 miliardi di dollari.

Quali sono le sfide aperte?

Il mercato vuole memorie sempre più capienti: per esempio, non basterà poter memorizzare mille fotografie quando si fa un viaggio, se ne vorranno memorizzare sempre di più; si vorranno memorizzare filmati o addirittura film ad alta qualità. Questo richiede di dover fare delle celle sempre più piccole e, come si accennava in precedenza, il problema dei disturbi elettrici tra le celle diventerà via via più difficile da risolvere.

Ma c'è ancora un altro problema: più piccole sono le celle, minore è il numero degli elettroni che devono attraversare lo strato di ossido per memorizzare un bit e diventerà ancora più difficile mantenere l'informazione per tempi lunghi: è facile intuire che perdere 100 elettroni su 600 può essere ancora accettabile, ma perderne 100 su 200 significa aver di fatto perso l'informazione.

Un altro elemento riguarda la riduzione delle tensioni delle alimentazioni dei circuiti, che deve essere ridotta per diminuire il consumo di energia. Questi circuiti sono tutti alimentati o a batteria o con un alimentatore collegato alla rete elettrica. Per ridurre il consumo di potenza è stato necessario ridurre via via la tensione di alimentazione. Questi circuiti, attualmente, operano con tensioni di circa 1.2 volt ed i prossimi componenti che verranno creati scenderanno a tensioni sotto il volt. Si riuscirà così a ridurre il consumo di potenza ma questo porterà ad altri problemi dal punto di vista delle modalità di scrittura e lettura delle celle.

Presso l'Università di Ferrara ci occupiamo da tempo di tutti i problemi legati all'affidabilità delle memorie non volatili di tipo *flash*. Il problema più specifico su cui lavoriamo a Ferrara è “l'erraticità”. Questo è un fenomeno per il quale, in un circuito di qualche miliardo di celle, ce ne sono una decina che, in modo completamente incontrollato, hanno un comportamento diverso dalle altre. Questo accade senza sapere a priori quali sono le celle che sfuggiranno al controllo.

Lo studio dell'erraticità è un po' come lo studio di una malattia molto rara. Si pensi ad una malattia che su un miliardo di persone ne colpisca dieci, e quindi che sull'intera popolazione mondiale ne colpisca sessanta. Rispetto a problemi molto più importanti, difficilmente si troverebbero investitori interessati a finanziare la ricerca in questo campo.

Nel campo delle memorie il problema non sarebbe grave se non fosse per il fatto che, se non funziona quel dato bit, in un dato momento, possono accadere cose molto gravi: potrebbe cadere un aereo, o l'airbag di un'auto aprirsi quando non si deve aprire e viceversa. Esistono tecniche circuitali o software per cercare di ovviare al problema. È ovvio che la comprensione del fenomeno a livello fisico permetterebbe uno straordinario aumento delle prestazioni delle memorie ed una altrettanto significativa riduzione dei costi.

Infine, un'altra grande sfida della ricerca in campo di memorie è rappresentata dalle cosiddette "memorie multilivello".

Le "memorie multilivello" sono memorie in grado di memorizzare più bit all'interno della stessa cella. Queste memorie sono in grado di memorizzare due, tre o quattro bit per cella. In questo senso la ricerca si sta muovendo per anticipare le performance nella capacità di memorizzazione dei dati che lo sviluppo della tecnologia permetterebbe di raggiungere solo tra 6-8 anni. Cade il concetto fondamentale su cui si basa l'elettronica digitale: sì o no, acceso o spento. Non è più un problema di passaggio o meno di corrente. I circuiti elettronici devono essere in grado di valutare "quante" corrente passa.

Cambia completamente il paradigma con cui si è lavorato fino ad ora. Questa è una delle grandi sfide su cui si lavora in questi ultimi anni.

Le prime memorie multilivello sono già presenti sul mercato e sono memorie con due bit memorizzati nella singola cella.

Un breve accenno infine a possibili alternative: "memorie ferroelettriche" o "magneto-resistive a cambiamento di fase".

Farò un brevissimo accenno a quest'ultime, le "memorie a cambiamento di fase", perché proprio a Ferrara abbiamo diversi progetti finanziati dal Ministero dell'Università e della Ricerca, che si occupano proprio delle memorie a cambiamento di fase. Queste sono memorie in cui si utilizzano dei particolari materiali che per effetto di un rapido riscaldamento o di un rapido raffreddamento, cambiano la propria fase (lo stato chimico-fisico in cui si trovano), passando dalla fase cristallina alla fase amorfa o viceversa. Il riscaldamento o il raffreddamento rapido avvengono accendendo/spegnendo molto rapidamente la corrente che attraversa tali materiali. In questo modo cambia la resistenza del materiale, dato che un materiale nello stato amorfo ha una conducibilità diversa rispetto a quella che avrebbe nello stato cristallino. Queste sono memorie che vengono studiate come possibili alternative alle "memorie flash" e di queste memorie sono disponibili i primi prodotti commerciali.

Conclusione: le memorie a semiconduttore rappresentano uno

dei “mattoni” principali dei sistemi elettronici che hanno rivoluzionato il nostro modo di vivere negli ultimi trenta anni. Se sparissero le memorie a semiconduttore ci troveremmo completamente spiazzati perché ci mancherebbe tutto: non si accenderebbe l’automobile, non riusciremmo a comunicare se non a voce o via posta (scrivendo ovviamente le lettere “a mano”, non funzionando più i PC), ad ascoltare la musica. Negli ospedali la quasi totalità delle attrezzature utilizzate per tenere in vita i pazienti, per fare diagnosi e per fare ricerca, non funzionerebbe. Le memorie a semiconduttore rappresentano, insieme al processore (ovvero l’elemento di calcolo), il cuore di tutti i sistemi elettronici, in ambito civile (quindi *consumer*), industriale, militare, medico, dei trasporti, della ricerca. Se si vogliono sviluppare sempre nuovi sistemi è necessario avere memorie che siano sempre più performanti, con prestazioni migliori, con maggiori densità di dati memorizzabili, e con un’alta affidabilità.

Concludendo possiamo dire che ogni miglioramento richiederà di affrontare sempre nuove problematiche tecnologiche e di superare nuove sfide dal punto di vista fisico, chimico, circuitale, architeturale.

Ma questo è ciò che, da sempre, affascina e stimola i ricercatori!