

Dalla Microelettronica Alla Nanoelettronica

Federico Faggin

Introduzione

L'elettronica ha compiuto cent'anni. Nata nel 1907 con l'invenzione del triodo, la prima valvola termoionica in grado di amplificare un segnale elettrico, l'elettronica ha avuto uno sviluppo prodigioso, fornendo il principale materiale da costruzione per il sistema nervoso della società umana.

Per i primi decenni di vita, la funzione essenziale delle valvole fu l'amplificazione di segnali, permettendo così lo sviluppo della radiotelegrafia, della radio, degli amplificatori e della televisione. Durante la seconda guerra mondiale la gamma di applicazioni delle valvole si estese al radar, ai sistemi di controllo e al calcolo elettronico, culminando nel Colossus (1944), il calcolatore elettronico sviluppato dagli scienziati inglesi e specializzato alla decodifica dei messaggi segreti tra Hitler ed il suo alto comando; e nell'ENIAC (1946), considerato il primo calcolatore elettronico. In queste ultime macchine la valvola fu usata come interruttore, invece di amplificatore, operando su segnali binari, e fornendo un'alternativa molto più veloce al relais elettromeccanico.

La ricerca per sostituire la valvola con un dispositivo più piccolo, meno costoso ed energeticamente più efficiente, iniziata negli anni trenta, portò finalmente all'invenzione del transistor nel 1947, ad opera di tre scienziati della Bell Labs.

È con il transistor che nasce la microelettronica.

Per i primi dieci anni i transistori furono costruiti uno alla volta usando il germanio come elemento semiconduttore. Nel 1959, l'invenzione del processo planare alla Fairchild Semiconductor, ad opera dell'ingegnere svizzero Jean Hoerni, fa cambiare tutto. Si iniziò così a costruire i transistori, un centinaio alla volta, su di una piastrina di silicio -- elemento semiconduttore con caratteristiche più favorevoli del germanio -- facendo scomparire il germanio nel giro di pochi anni. Con il processo planare, visto che i transistori erano costruiti uno accanto all'altro, venne naturale pensare di collegarli insieme, realizzando così il circuito integrato -- una svolta decisiva per la microelettronica. Con l'avvento del circuito integrato, la microelettronica, nota anche come elettronica allo stato solido, cominciò a sostituire non soltanto i componenti attivi, ma anche quelli passivi: resistenze, condensatori, diodi, ecc., nonché la porzione del circuito stampato necessaria a collegarli insieme.

Ulteriore tappa fondamentale della microelettronica fu l'invenzione di un nuovo tipo di transistor, il transistor MOS (Metallo Ossido Semiconduttore), che una volta perfezionato cominciò a sostituire i transistori bipolari usati in tutti i primi circuiti integrati. Con l'invenzione

del processo MOS autoallineante con porta di silicio policristallino, fu possibile costruire le prime memorie a semiconduttore ed il primo microprocessore. Il campo applicativo della microelettronica si allargò ancora una volta, rimpiazzando le memorie a ferrite magnetica fino ad allora usate nei computer e creando per la prima volta un intero computer allo stato solido. Nel giro di altri dieci anni i circuiti MOS finirono per soppiantare quasi totalmente i circuiti integrati bipolari, contemporaneamente estendendo la loro capacità applicativa ai circuiti integrati analogici, alle memorie non volatili e ai sensori d'immagine; le ultime due applicazioni impossibili da fare sia con transistori bipolari sia con transistori MOS con gate di alluminio.

Nel 1965, Gordon Moore, uno dei due fondatori dell'Intel (1968), allora capo dei laboratori di ricerca della Fairchild Semiconductor, osservò che ogni anno il numero di transistori in un circuito integrato raddoppiava, prognosticando che questo comportamento sarebbe continuato nel futuro. Questa osservazione venne più tardi chiamata la legge di Moore, anche se ovviamente non è una legge fisica come la legge di Newton e la sua validità è limitata nel tempo. La longevità della legge di Moore è dovuta al principio dello *scaling*, scoperto qualche anno dopo, secondo il quale, riducendo le dimensioni critiche del transistor MOS nella stessa proporzione, si diminuisce non solo l'area del transistor e la sua potenza dissipata, ma si aumenta anche la sua velocità. A partire dalla metà degli anni Settanta, lo *scaling* ha rappresentato la strategia fondamentale per migliorare le prestazioni e ridurre il costo dei circuiti integrati.

Nel 1970 la litografia più avanzata in produzione era in grado di stampare linee di 6 μm . Oggi siamo arrivati a 45 nm, riducendo l'area di un transistor di circa 20.000 volte in 38 anni. Nel 1970, la memoria RAM più avanzata era statica e aveva 256 bit, con un tempo di accesso di 1,5 μsec . Oggi la RAM statica più avanzata ha 64Mbit con un tempo di accesso di 2,6 nsec. Ancora più rapido è stato il progresso nel campo delle memorie RAM dinamiche e nelle memorie flash, che non esistevano ancora nel 1970. La memoria flash più avanzata oggi in produzione ha 512 milioni di celle, con un tempo di accesso di circa 50 nsec, integrando circa 1,6 miliardi di transistori. Con la memoria flash è addirittura possibile immagazzinare più di un bit per cella, dosando il numero di elettroni che sono iniettati nel *floating gate* della cella di memoria e raggiungendo fino a 8Gbit di memoria con un chip da 512 milioni di celle. Faccio notare che la differenza tra uno stato e l'altro è dell'ordine di cento elettroni in più o in meno caricati nel *floating gate*.

Nel campo dei microprocessori il progresso è stato altrettanto sorprendente. Il primo microprocessore, l'Intel 4004, introdotto nel mercato nel 1971, integrava circa 2300 transistori in un'area di 12 mm quadrati, con litografia da 6 μm , e usando 6 maschere per la sua fabbricazione. Il 4004 era in grado di eseguire circa 100,000 istruzioni al secondo, operanti su 4 bit. Ciascun ciclo di istruzione richiedeva 8 periodi di clock, alla frequenza di circa 750 kHz. Il 4004 aveva 16 pin, la tensione di lavoro era di 15 volt, la potenza dissipata era 750 mW e costava \$35 in quantità di 1000 pezzi. Se confrontiamo il 4004 con uno degli ultimi microprocessori Intel, lo Xeon L5420, introdotto all'inizio di quest'anno, possiamo verificare lo straordinario progresso fatto in 37 anni. Lo Xeon integra circa 820 milioni di transistori con litografia da 45 nm, occupa un'area di 210 mm quadrati ed è fabbricato usando circa 30 maschere -- la maggioranza dei transistori sono usati per due livelli di memoria *cache*. Lo Xeon ha quattro unità centrali funzionanti in parallelo, con cui esegue fino a 20 miliardi di istruzioni per secondo, ciascuna operante su 64 bit. La frequenza di clock è di 2.5 GHz e durante ciascun periodo di clock, il chip

esegue fino ad 8 istruzioni in parallelo. Lo Xeon ha 771 pin, dissipa 50 W, usa una tensione di lavoro di 1,2 volt e costa \$380 in quantità di 1000 pezzi; più o meno lo stesso costo del 4004 se si considera l'effetto dell'inflazione. Un paio di mesi fa l'Intel ha annunciato una nuova serie di microprocessori Xeon, la serie 7400, con 6 *core* e 16 MB di memoria *cache*, integrando ben 1,9 miliardi di transistori nello stesso chip.

Nella storia del progresso tecnologico, la microelettronica ha superato di gran lunga qualsiasi altra disciplina nell'aumentare le prestazioni nel tempo, mantenendo lo stesso costo.

Ma fino a quando sarà possibile continuare la corsa esponenziale degli ultimi 40 anni? Oggi lo *scaling* sta incontrando difficoltà sempre più grandi, a mano a mano che ci avviciniamo alle dimensioni molecolari. Già con la litografia da 90 nm, che risale a 5 anni fa, lo spessore dell'isolante del gate – fatto con biossido di silicio -- fu ridotto a 1,2 nm, 5 strati molecolari in tutto. Se si riduce lo spessore a 4 strati molecolari, il *leakage* del gate dovuto all'effetto *tunneling*, aumenta al punto da rendere il transistor quasi inservibile. Pertanto le due generazioni successive hanno evitato di scalare lo spessore dell'ossido riducendo così l'aumento delle prestazioni dei transistori che altrimenti sarebbero state possibili.

Questa limitazione è stata anticipata da tempo dall'industria dei semiconduttori che ha fatto ricerca per molti anni per scoprire materiali ad alta costante dielettrica per sostituire il biossido di silicio. Solo così è possibile aumentare le prestazioni del transistor senza dover diminuire lo spessore dell'isolante che ha l'effetto negativo di aumentare la corrente di *leakage*. Questa ricerca si è rivelata molto più ardua del previsto in quanto non è soltanto necessario trovare un isolante adeguato, ma l'intero sistema metallo-ossido-silicio -- chiamato *stack* -- deve essere compatibile a tutti i livelli richiesti per la costruzione, il funzionamento e l'affidabilità del transistor. Il sistema precedente, costituito da silicio policristallino-biossido di silicio-silicio è rimasto inalterato dal 1968, anno in cui il primo circuito integrato commerciale fatto con la tecnologia *silicon gate* – tecnologia e circuito da me sviluppati -- fu introdotto nel mercato.

Finora, soltanto l'Intel è riuscita a sviluppare uno *stack* adeguato, dopo uno sforzo decennale, portando in produzione alla fine dell'anno scorso i primi circuiti integrati a 45 nm con un isolante basato sull'elemento hafnio, ed un metallo che non è stato specificato. Si stima che la permittività dielettrica relativa del nuovo isolante abbia un valore tra 10 e 15 – l'Intel non ha ancora pubblicato il valore esatto – che è un fattore da 2,5 a 4 volte superiore alla permittività dielettrica relativa del biossido di silicio, che è 3,9. Con questo nuovo *stack*, l'Intel ha potuto usare isolante più spesso, facendo marcia indietro e riducendo drasticamente la corrente di *leakage* e quindi la potenza dissipata, e contemporaneamente aumentando le prestazioni dei transistori come sarebbe stato possibile fare usando 3-4 strati molecolari di ossido di silicio.

L'Intel ha dichiarato che lo stesso *stack* sarà utilizzato nel nodo litografico successivo, 32 nm, che entrerà in produzione l'anno venturo. Tuttavia, per la generazione a 22 nm, c'è ancora qualche incertezza se sarà possibile usare gli stessi materiali senza conseguenze negative. Il problema sta diventando sempre più difficile e già al livello dei 22 nm il metodo usato per le due generazioni precedenti è in dubbio. Il *silicon gate*, durato per circa venti generazioni, deve essere d'ora in poi sostituito da materiali nuovi che richiedono un grandissimo sforzo economico per il loro sviluppo e che durano soltanto qualche generazione.

Malgrado queste difficoltà, penso che sarà possibile avere chip a 22 nm in produzione nel 2011-2012, allungando solo di poco la tabella di marcia che ha recentemente registrato una media di circa due anni per generazione. Prevedo comunque che le generazioni successive cominceranno a richiedere tempi di sviluppo sempre più lunghi, con i 16 nm in produzione nel periodo 2014-2015, e gli 11 nm, in produzione non prima del 2019.

Gli esperti non sono ancora sicuri che sarà possibile andare sotto i 10 nm, tuttavia se il vantaggio economico è sufficiente a giustificare l'enorme costo dello sviluppo e soprattutto se non ci sono alternative migliori, penso che l'industria riuscirà a ridurre ulteriormente le dimensioni, magari fino a 4-5 nm, usando però tempi di sviluppo sempre più lunghi. Non c'è alcun dubbio comunque che la strada maestra dello *scaling* del transistor MOS ha una vita limitata e si devono cercare altre strutture per continuare la corsa al sempre più piccolo, più veloce e meno costoso.

Una strategia evolutiva che sta già maturando sotto i nostri occhi consiste nel fare chip con più di uno strato attivo. Il metodo più elegante, ma anche più complesso, è di costruire chip con più di un livello attivo sulla stessa fetta di silicio. Per esempio, la Foveon ha costruito sensori di immagine dove i tre fotodiodi per il rilevamento dei colori primari sono uno sopra l'altro invece che uno accanto all'altro. Il metodo più pratico, e forse alla fine anche meno costoso, è tuttavia quello di montare più di un chip, uno sopra l'altro, sullo stesso *package*. Già gran parte delle memorie flash usano queste tecniche per aumentare il numero di bit di memoria contenuti nello stesso dispositivo. Però la strada più avanzata si baserà sulle nuove tecniche che oggi vanno sotto il nome generico di *wafer-scale-packaging*. Metodi in grado di estendere lo stesso principio usato nel processo planare anche all'incapsulamento dei chip, assemblando quindi tutti in una volta i chip che si trovano in una fetta di silicio, e separandoli alla fine del processo, pronti per essere montati sui circuiti stampati.

Tutta questa attività sta creando la conoscenza necessaria per iniziare una nuova strada che va sotto il nome generico di 3D ICs, circuiti integrati a tre dimensioni. Le tecniche di base per fare 3D ICs sono l'assottigliamento delle fette di silicio ad uno spessore di qualche decimo di mm, lo sviluppo di tecniche per fare *vias*, da usare per il collegamento verticale tra i circuiti integrati, e tecniche di *stacking* per sovrapporre e allineare sia chip sia wafer. Una delle ragioni pratiche che favoriscono il movimento verso le tre dimensioni è il fatto che la potenza dissipata dovuta alle interconnessioni diventa sempre più alta a mano a mano che l'area dei chip aumenta. Facendo dei chip più piccoli e montandoli uno sopra l'altro si può ridurre la lunghezza media delle interconnessioni e quindi la potenza dissipata.

Non c'è dubbio che in futuro i 3D ICs, che oggi sono soltanto all'inizio del loro sviluppo, riceveranno sempre più attenzione e diventeranno di uso comune a mano a mano che lo *scaling* diventerà sempre più difficile e costoso, prolungando così il periodo di continui miglioramenti per la microelettronica anche dopo che lo *scaling* tradizionale avrà finito il suo corso.

Faccio notare che uso il termine "microelettronica" anche quando le dimensioni dei transistori sono di qualche decina di nm, quando potrei usare benissimo il termine "nanoelettronica." Questo perchè vorrei riservare il termine "nanoelettronica" ad una nuova classe di dispositivi elettronici più piccoli e più veloci, basati su nuovi principi di funzionamento, che promettono di sostituire i transistori MOS. La nanoelettronica offre quindi la nuova strada per continuare ad

aumentare le prestazioni e ridurre le dimensioni e il costo dei circuiti integrati una volta che i transistori MOS hanno raggiunto il limite fisico dello *scaling*.

La nanoelettronica è oggi più scienza che tecnologia. Di fatto non ha ancora raggiunto il livello di commercializzazione benchè il progresso negli ultimi dieci anni sia stato fenomenale. Alla scala del nanometro esistono proprietà elettriche e magnetiche sorprendenti in materiali sia vecchi sia nuovi, dando vita a molti nuovi filoni di ricerca, ciascuno basato su una proprietà o su un materiale promettente. Potrei dire che oggi c'è anche troppa dovizia di possibilità, rendendo difficile discernere la strada maestra da strade minori o dai vicoli ciechi. Molti ricercatori stanno cercando di trovare un dispositivo molto più piccolo di un transistor MOS, senza preoccuparsi troppo dei problemi di fabbricazione. Questo va bene, purché non si dimentichi che il problema da risolvere è ben maggiore di quello di fare semplicemente un transistor più piccolo. Un transistor più piccolo è solo una condizione necessaria ma non sufficiente per sostituire il transistor MOS. Bisogna anche che il nuovo dispositivo sia molto più veloce, dissipino meno energia, sia collegabile ad altri dispositivi analoghi per fabbricare circuiti integrati con decine di miliardi di componenti a costo inferiore e con affidabilità equivalente a quella dei circuiti integrati convenzionali. Un problema monumentale.

Penso che la scala del problema da risolvere sia talmente vasta che l'unica strategia possibile è quella di trovare nanomateriali e nanostrutture in grado di sostituire materiali e dispositivi tradizionali gradualmente, a mano a mano che i vecchi materiali si riveleranno limitanti. Solo così si potrà far leva sull'investimento collettivo che ha portato la microelettronica all'incredibile livello di sviluppo contemporaneo. Faccio ora un esempio di come l'evoluzione dalla microelettronica alla nanoelettronica potrebbe avvenire, considerando che i problemi più critici da risolvere per fare circuiti integrati sempre più avanzati sono il problema delle interconnessioni ed il problema della potenza dissipata, dovuto soprattutto alla capacità e alla resistenza elettrica delle interconnessioni. Oggi impieghiamo fino a nove livelli di interconnessioni, usando sottilissimi strati di rame. Però, come riduciamo lo spessore e la larghezza del nastrino di rame che collega i vari transistori, la resistenza per unità di lunghezza aumenta rapidamente, contribuendo negativamente sia alla potenza dissipata sia alle prestazioni del circuito. Se riuscissimo a scalare soltanto le dimensioni delle interconnessioni, senza però degradarne le caratteristiche elettriche, potremmo continuare a ridurre le dimensioni dei chip ancora per anni senza che sia necessario trovare un dispositivo attivo più piccolo. Per far ciò è necessario trovare un nanomateriale con proprietà di conduzione elettrica molto migliori del rame, quando le dimensioni sono ridotte a qualche nanometro. Recentemente pare che sia stato trovato un materiale promettente: il grafene.

Il grafene è una delle forme allotropiche del carbonio che sono state scoperte recentemente, cominciando con la fullerite, scoperta nel 1985 e nota anche con il nome di "buckyball." La fullerite è una sfera vuota composta da 60 atomi di carbonio regolarmente disposti sulla sua superficie. Nel 1991 il nanotubo di carbonio fu riscoperto e portato all'attenzione scientifica. Si tratta di un cilindro vuoto dal diametro minimo di poco più di un nanometro, la cui superficie è formata da atomi di carbonio con disposizione esagonale. Nel 2004 il grafene fu isolato per la prima volta: si tratta di un solo strato molecolare di carbonio dove gli atomi sono disposti in un reticolato esagonale, esattamente come in un nanotubo, però su di una superficie piatta. Questo è uno sviluppo insospettato visto che la grafite, che è nota da sempre, è composta da tanti strati di grafene sovrapposti caoticamente.

Bene, nel giro di pochi anni gli scienziati hanno scoperto che il grafene ha proprietà incredibili, simili a quelle dei nanotubi, ma più promettenti in quanto la sua forma a due dimensioni è compatibile con la tecnologia dei circuiti integrati che sono costruiti sulla superficie di una fetta di silicio. Per esempio, un nastro di grafene della larghezza di pochi atomi può essere sia un semiconduttore sia un metallo. Non solo, ma le proprietà di conduzione elettrica della forma metallica sono molto superiori a quelle del rame, aprendo la strada a potenzialmente sostituire il rame per le interconnessioni dei circuiti integrati. Recentemente è stato anche dimostrato un transistor di grafene che ha proprietà elettroniche superiori a quelle dei dispositivi fatti sul silicio. Ecco quindi prospettarsi per la prima volta una strada che non solo è compatibile con la tecnologia esistente, ma che si presta anche ad una adozione graduale.

Ciò che rende il grafene un materiale straordinario è che esso può esistere in nanoscala senza decomporsi, ossidarsi o diventare instabile; proprietà che nemmeno il silicio possiede. Si intravede quindi la possibilità reale, anche se ancora a divenire, di creare circuiti integrati in uno o due strati molecolari di carbonio -- proposta impensabile soltanto dieci anni fa.

Un altro filone importante è l'uso di materiali magnetici su nano scala in circuiti integrati basati sul silicio per fare memorie permanenti di caratteristiche superiori alle memorie flash e DRAM tradizionali. Questa nuova strada è molto promettente poiché sposa il circuito integrato contemporaneo con la nuova disciplina chiamata "spintronics." Nata con la scoperta nel 1988 della magnetoresistenza gigante (GMR), la spintronics usa sia le proprietà di carica elettrica sia le proprietà magnetiche dell'elettrone che si manifestano nel suo spin quantico. I dispositivi spintronici si basano sull'abilità di creare e controllare correnti di elettroni con lo spin polarizzato in una certa direzione. Sensori magnetici basati sui principi spintronici -- la *spin valve* -- sono già in uso da alcuni anni nelle memorie magnetiche rotanti. L'industria dei semiconduttori da anni sta sviluppando memorie che, usando principi spintronici, promettono di sostituire le memorie RAM statiche, le DRAM e le memorie flash con un'unica memoria universale, una MRAM, che assomma le caratteristiche principali di velocità, costo per bit e non-volatilità che hanno giustificato finora l'esistenza di tre tipi diversi di memoria.

Recentemente l'IBM ha annunciato lo sviluppo di una nuova tecnologia per memorie non-volatili che ha chiamato "RaceTrack," con cui spera di soppiantare le memorie flash e le memorie magnetiche rotanti. Si tratta di un dispositivo allo stato solido dove le pareti dei domini magnetici si muovono all'interno di un nanoconduttore, con un processo simile ad un registro a scorrimento, permettendo di ottenere densità di memoria cento volte superiori alle attuali memorie flash, con velocità, affidabilità, costo e potenza dissipata favorevoli rispetto ad esse. L'IBM prevede che questa nuova tecnologia permetterà di fare memorie magnetiche con densità fino a 2 Gbit per mm quadrato. Benché questa tecnologia sia ancora allo stadio di ricerca e sviluppo, e potrebbe anche rivelarsi non adeguata alla commercializzazione, come minimo essa attesta all'enorme vitalità di questa nuova direzione tecnologica.

Benché io abbia fatto soltanto alcuni esempi delle nuove frontiere della nanoelettronica, non è ancora possibile discernere la strada evolutiva principale che sarà imboccata a livello industriale-commerciale tra circa dieci anni. Posso però dire che si intravedono già possibilità concrete di integrazione del vecchio e del nuovo che non solo permetteranno di prolungare la validità della legge di Moore ancora per molti decenni, ma che anche arricchiranno sostanzialmente le possibilità espressive del nuovo hardware del secolo XXI.

Non posso fare a meno di far notare la possibilità concreta e inaspettata che l'hardware del futuro possa anch'esso essere basato sul carbonio, lo stesso elemento magico che ha permesso alla natura di evolvere la vita e creare la nanomacchina più avanzata del sistema solare: l'uomo.

*Distribuito in occasione del Convegno organizzato da
AICA, CNIPA e Confindustria Servizi Innovativi e Tecnologici*

“Professioni ICT: una “lezione” per il futuro
Roma, 14 Novembre 2008