

Di Davide Plumetti

TRANSISTOR

TRA PASSATO E FUTURO

*Come è fatto: il transistor.
Un'analisi dettagliata
del processo produttivo
attuale e uno sguardo
alle barriere tecnologiche
superate nel corso
degli anni e alle tante sfide
future ancora da affrontare.*

Transistor: dopo aver analizzato la logica e la fisica di base e la loro evoluzione all'interno dei processori, dagli albori a oggi, affrontiamo l'argomento dal punto di vista costruttivo, ovvero come sia possibile inserire in spazi tanto ristretti un numero così elevato di elementi delicatissimi, con tolleranze e specifiche quasi impossibili da immaginare. I numeri in campo sono impressionanti: alcuni particolari elementi dei transistor moderni sono infatti costituiti da uno strato spesso non più di un singolo atomo, mostrando come la miniaturizzazione esponenziale degli ultimi anni abbia raggiunto livelli estremi.

Nel seguito analizzeremo non solo la storia della progressiva riduzione dei transistor dettata dalla famosissima legge di Moore, ma ci soffermeremo sugli elementi che hanno ostacolato periodicamente tale costante evoluzione e come sono stati superati. Inoltre daremo una visione di stampo ingegneristico sulle metodologie produttive moderne e, infine, uno sguardo ai prossimi 10 anni, in cui il settore si propone di vivere una nuova primavera grazie ad alcune tecniche rivoluzionarie che le grandi aziende stanno studiando.

LA LOGICA DEI TRANSISTOR

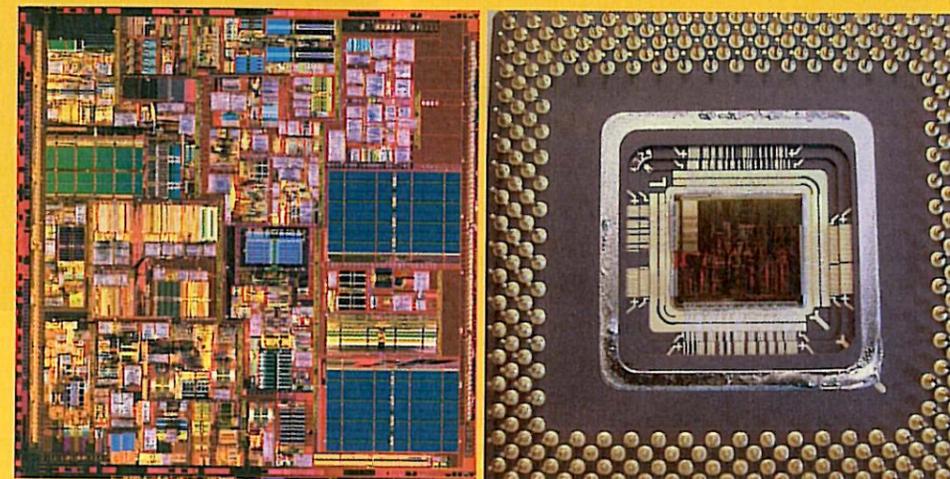
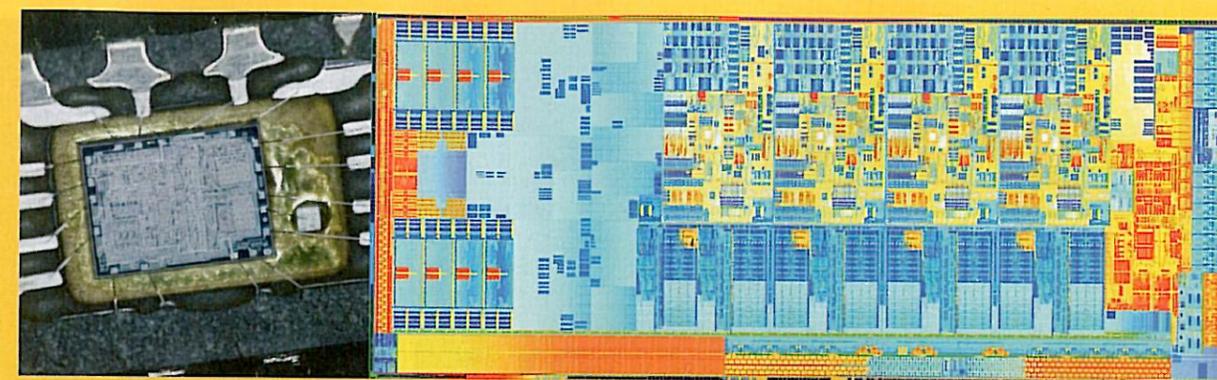
Come descritto nel dettaglio negli articoli dei mesi precedenti, il funzionamento di un singolo transistor è molto semplice: esso funziona infatti come un interruttore ai cui capi (*source* e *drain*) scorre o meno corrente elettrica in base allo stato del terzo connettore (il *gate* controllante). Inserendo in serie o parallelo alcuni

transistor si possono creare le porte logiche classiche (*And*, *Or*, *Not*) e le derivate, assemblando le quali si costruiscono i circuiti veri e propri. Progettando opportunamente un chip si possono disporre transistor (porte logiche) in modo che, inserendo a un estremo due valori binari, si abbia in uscita la loro somma, la sottrazione o molte altre combinazioni. Nei processori il funzionamento è dinamico, ovvero esiste un punto di ingresso al quale vengono fornite le "istruzioni", una serie di bit che identificano una particolare operazione da eseguire con i dati presenti in punti predefiniti. A seconda dell'istruzione iniziale (una serie di bit che configurano le porte logiche in una maniera piuttosto che in un'altra) il sistema eseguirà un calcolo diverso, ottenendo il risultato desiderato in uscita. Complicando ulteriormente il tutto possiamo immaginare un input in termini di istruzione corredato da altri due (o più) input che identificano l'indirizzo di memoria dove prelevare

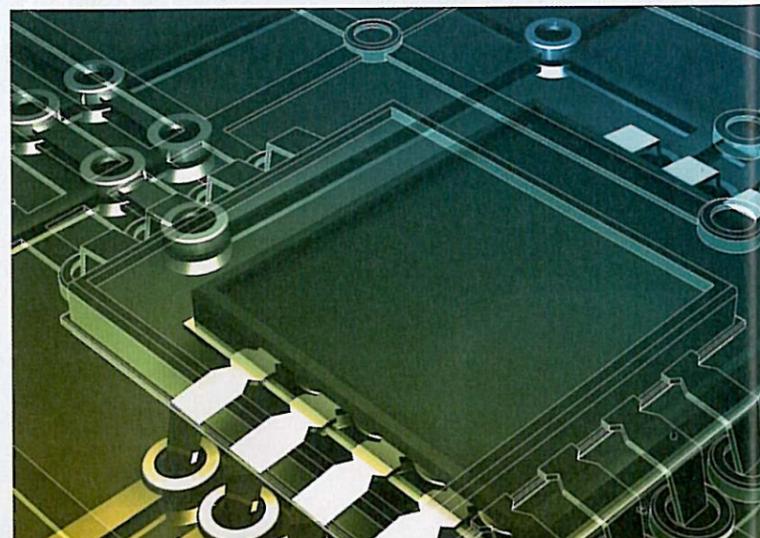
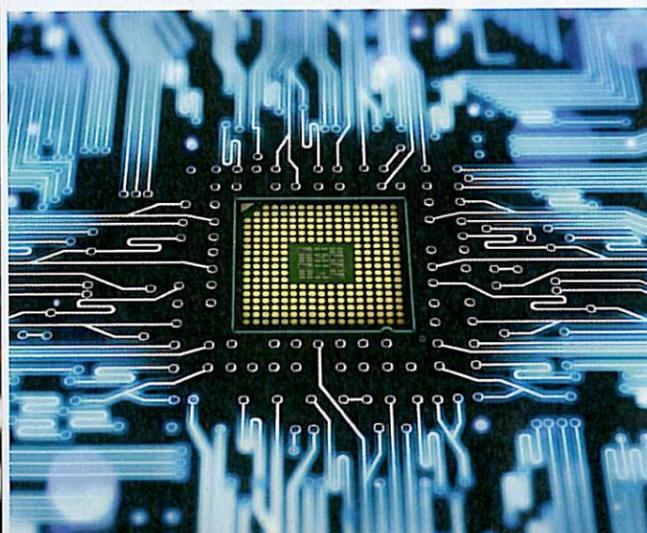
i dati in ingresso e dove scrivere quello o quelli in uscita. Riuscendo a immaginare questo abbiamo già chiaro come funziona una Cpu. Il principio è questo, sia il 4004 di Intel del 1971, il processore del vostro smartphone o quello del vostro Pc, così come quelli dei supercomputer sparsi per il mondo.

I NUMERI CONTANO

Ovviamente la "potenza" di un processore dipende dalle operazioni che è in grado di eseguire per unità di tempo, sia di tipo semplice sia di tipo complesso. I transistor presenti nel primo 4004 di Intel erano solo 2.300, mentre nelle moderne Cpu e Gpu siano ben oltre i 2 miliardi. Il funzionamento base è lo stesso ma (anche escludendo 1 miliardo abbondante di transistor solamente dedicati alla cache) avere più transistor permette di compiere più operazioni, sia come tipologia sia come numero, parallelizzando i calcoli.



L'evoluzione nella dimensione e nella forma di transistor e Cpu. Il 4004 (1971) aveva 2.300 transistor, il primo Pentium (1993) arrivava a 3,1 milioni mentre il Pentium 4 (2000) a 42 milioni e i nuovi Core i7 a oltre 1,4 miliardi.



Avere più transistor permette dunque di fare le cose più velocemente, per due motivi principali: il primo è che su una Cpu semplice i calcoli devono essere fatti in maniera elementare. Ad esempio per moltiplicare 9 x 5 su una Cpu non dotata di una logica moltiplicativa (che impiega molti transistor) è necessario procedere per passi, sommando 9 + 9 per cinque volte. Se una singola somma impiega 1 secondo per calcolare 9 x 5 sono necessari 4 secondi (9+9; 18+9; 27+9 e 36+9). Per calcolare 17 x 22 sono necessari 16 secondi (sommiamo 22 per 17 volte). Inserendo altri transistor capaci di effettuare una moltiplicazione diretta possiamo incrementare in maniera enorme la velocità di calcolo in questo specifico processo. Se tale logica dovesse impiegare ad esempio 1,5 secondi per qualunque moltiplicazione avremmo un miglioramento del 62% in termini di velocità effettuando 9 x 5 e addirittura del 90% per un'operazione come 17 x 22. Per questo motivo i processori integrano sempre più

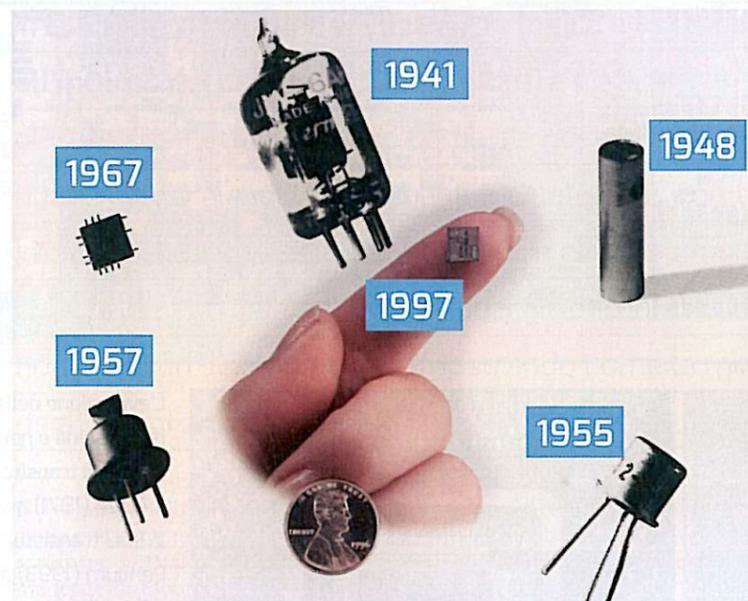
transistor in grado di costruire circuiti specializzati per operazioni complesse. Nelle Cpu di ultima generazione sono presenti istruzioni dedicate ai calcoli più disparati (con circuiteria dedicata), in grado di fare calcoli trigonometrici complessi in una singola iterazione o operazioni di codifica o decodifica in hardware rendendo tali applicazioni molto più veloci. Il secondo motivo per cui più transistor rendono più veloce un processore è relativo alla parallelizzazione. Se una Cpu deve fare una serie di calcoli indipendenti (9 x 5, 6 x 7 e 22 x 41) può agire in due modi in base alla propria struttura interna. Il processore di prima impiegherà 4,5 secondi totali per eseguire queste tre operazioni (a 1,5 secondi ciascuna), mentre una Cpu di pari caratteristiche intrinseche ma dotata di 4 core elaborativi può assegnare un'operazione a ciascun core (lasciandone anche uno libero), impiegandoli

in parallelo completando il calcolo in 1,5 secondi. Miglioramento del 66%. Ma servono più transistor.

LE DIMENSIONI CONTANO

Seguendo il principio appena accennato sembra possibile costruire processori sempre più potenti semplicemente integrando al loro interno porzioni sempre più nuove in grado di svolgere via hardware operazioni complesse. Il problema che si pone davanti agli ingegneri è però di tipo completamente opposto: dal punto di vista progettuale farebbero carte false per poter inserire miliardi e miliardi di transistor in una Cpu, ma in questi casi si pone il problema delle dimensioni. Realizzare una Cpu moderna con le tecniche produttive del Intel 4004 porterebbe ad avere un processore grande circa 16 metri quadrati, quanto un box auto di grandi dimensioni.

Miliardi di gate
I chip più complessi (le Gpu) sono composte da oltre 7 miliardi di transistor



Evoluzione del transistor dalla sua invenzione a oggi. I primi prototipi erano ben visibili a occhio nudo, mentre oggi sono molto più piccoli di un virus.



Su un moderno chip, illuminato lateralmente, si possono notare le strutture principali che lo costituiscono. I disegni derivano dalle diverse interconnessioni presenti.



Un laboratorio di ricerca sulla fotolitografia. La luce gialla tipica è utilizzata perché meno ricca di componenti Uv, estranee alla fotolitografia.

Per questo motivo l'ingegnerizzazione elettronica delle Cpu corre di pari passo con l'evoluzione delle tecniche produttive, che cercano di realizzare transistor sempre più piccoli che possano essere stipati in maniera ordinata all'interno di una Cpu. A oggi tutti i produttori utilizzano sistemi simili e, a parte qualche tecnologia appena affacciata sul mercato (di cui parleremo molto nel seguito), affrontano problemi identici ormai da decenni.

Dall'inizio dell'era del silicio tutti i processori sono stati costruiti inserendo i transistor direttamente su un wafer di silicio in maniera planare. Ovvero tutte le Cpu moderne sono costruite come un mosaico bidimensionale, con miliardi di transistor posti uno di fianco all'altro e interconnessi tra di loro (sia lateralmente sia con "impalcature" multipiano di estrema complessità). Per inserire un numero sempre più elevato di transistor in uno spazio sempre identico (quasi tutte le Cpu per questioni di spazio e rese produttive hanno una superficie da 50 a 250 mm²) è necessario rimpicciolire i transistor.

Visto il loro sviluppo quasi bidimensionale la misura dei transistor viene solitamente definita come la larghezza media del canale del gate. Le dimensioni totali sono maggiori ma, essendo proprio le dimensioni del gate la sfida

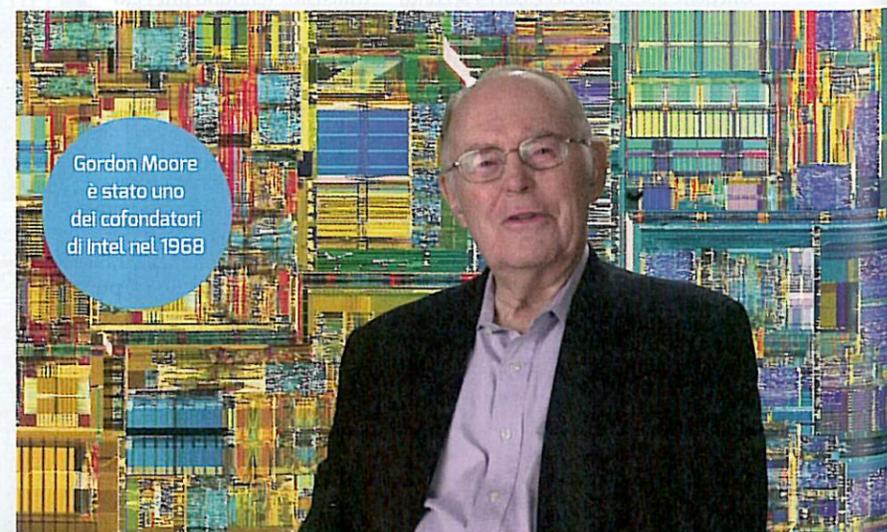
tecnologica più complicata da vincere, si utilizza tale misura per definire l'evoluzione delle tecniche produttive. I primi transistor ad essere inseriti in Cpu vere e proprie erano costruiti con un processo a 10 micrometri o micron. In pratica il gate aveva una dimensione media di poco superiore a quella di un globulo rosso (8 micron) e decisamente più sottile di un capello umano (75 micron).

LA STORIA E GORDON MOORE

Nei primi anni '70 l'evoluzione tecnologica relativa ai processi produttivi ebbe un'impennata tale da permettere di ridurre drasticamente più volte, nel

giro di pochi anni, le dimensioni medie di un transistor. Dopo i 10 micrometri del 1971 con i quali furono costruiti i 2.300 transistor dell'Intel 4004 si passò rapidamente a 6 micron e 4.500 transistor della Cpu 8080 del 1974 e a 3 micron per i 29.000 transistor del 8086 del 1978. Ancora prima di questa esplosione tecnologica uno dei padri fondatori di Intel, Gordon Moore, effettuò quando ancora lavorava per Fairchild Semiconductor, alcune osservazioni sulle prime tecnologie produttive degli anni '60, enunciando una propria teoria che ha in seguito guidato l'innovazione tecnologica per 50 anni esatti.

Nel 1965 Moore profetizzò che le prestazioni dei microprocessori sarebbero



Gordon Moore è stato uno dei cofondatori di Intel nel 1968

raddoppiate ogni 12 mesi. Dopo 10 anni di previsioni azzeccate, nel 1975 la legge venne modificata e resa più uniforme per un mercato più maturo, decretando prima che il raddoppio di prestazioni sarebbe avvenuto ogni 24 mesi e ritocata subito dopo nella definitiva "raddoppio ogni 18 mesi".

Il raddoppio delle prestazioni è stato considerato in maniera duplice, inizialmente riguardava la sola frequenza di funzionamento, mentre in seguito la legge è stata applicata sia alle prestazioni pure sia al numero di transistor che costituiscono una Cpu commerciale. In maniera quasi strabiliante, dati alla mano, possiamo vedere come la legge di Moore ricalchi perfettamente l'andamento dell'evoluzione nelle tecniche produttive da ormai 50 anni.

I processori moderni raddoppiano il numero di transistor che li compongono all'incirca ogni 18-20 mesi. In termini pratici siamo passati da 55 milioni nel 2003 a 291 milioni nel 2006, 410 milioni nel 2008 e 1,4 miliardi nel 2012. Oggi siamo a quota 2 miliardi. Trovate una descrizione dettagliata su come, praticamente, vengono inseriti tutti questi transistor su un supporto e prodotti in massa in un box dedicato nelle prossime pagine, utile anche per comprendere tutte le tecniche più avanzate messe in atto dai produttori.

La costante riduzione delle dimensioni dei transistor, unita alla loro crescita in frequenza operativa necessaria per aumentare le prestazioni, ha iniziato ad incontrare alcune difficoltà nel corso degli ultimi 10 anni. Con i processi produttivi sotto i 90 nanometri (nm) gli ingegneri hanno incontrato per la prima volta alcuni fenomeni fisici che non si erano mai manifestati con grandezze superiori.

Con un canale del gate largo solo 90 nm è necessario costruire un isolante tra il

canale e il gate vero e proprio molto sottile, dello spessore di pochi atomi di silicio. Con uno spessore tanto ridotto emergono però fenomeni di *current leakage*, durante i quali gli elettroni riescono a superare tali barriere isolanti come se non esistessero, creando correnti parassite che impegnano elettricamente i transistor. Processori costruiti a 90 nm hanno segnato il culmine in termini di potenza dissipata soprattutto per via di queste correnti parassite. In quegli anni le grandi aziende si sono rese conto, contemporaneamente, di due cose: con le tecnologie classiche era impossibile raggiungere la soglia dei 4 GHz per via del rumore elettronico prodotto troppo elevato per permettere di distinguere chiaramente tra "0" ed "1" e per via del calore generato.

Le Cpu dell'epoca sono arrivate a dissipare in calore anche 135 watt su meno di 200 mm². Un valore che da solo non rende abbastanza bene l'idea della quantità di potenza rilasciata da tali Cpu. In proporzione una Cpu di questo tipo ha un'emissione di calore di circa 0,8 MWatt per metro quadro (80 Watt al cm²). Tanto? Poco? Come paragone possiamo valutare il flusso di potenza media dissipata da un reattore nucleare BWR (quelli costruiti negli anni '90): circa 0,5 MWatt per metro quadro, con un picco massimo di 1 MWatt per metro



IL PROCESSO FOTOLITOGRAFICO

La produzione dei moderni microprocessori rappresenta probabilmente il processo tecnologico più complesso nell'intera storia dell'umanità. Qui possiamo vederne le fasi principali nella declinazione Intel a 22 nm.



1 Tutto ha inizio dalla sabbia, ricca di silicio (il secondo elemento più abbondante sulla terra).



2

La sabbia viene fusa e purificata fino a ottenere un lingotto di silicio monocristallino, 30 cm di diametro per oltre 100 Kg.



3

Il lingotto viene tagliato in orizzontale in quelli che prendono il nome di wafer, sui quali verranno costruiti i transistor.



4

I processi di costruzione si compongono di centinaia di passaggi. Qui l'applicazione di una maschera fotoresistente (reagisce alla luce e resiste agli acidi).



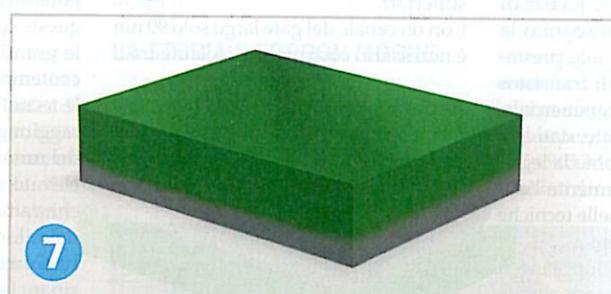
5

Esponendo il wafer alla luce UV tramite una maschera si indurisce una parte del wafer secondo i circuiti che si vogliono imprimere.



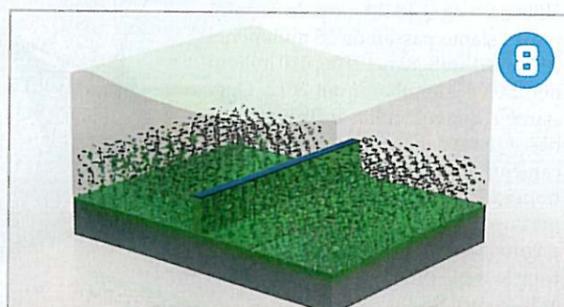
6

Grazie a un impianto ionico si droga il silicio a seconda della necessità.



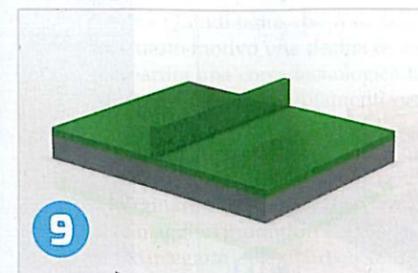
7

Un ingrandimento della zona dove verrà realizzato un transistor. In scala con una larghezza di circa 100 nm.



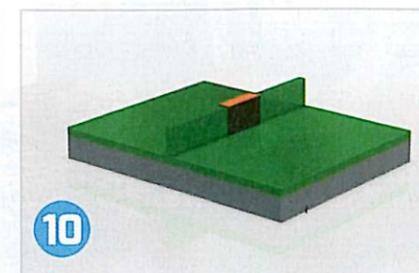
8

Si rimuove con un trattamento chimico parte del materiale. Resteranno intatte solo le zone a suo tempo coperte dal materiale fotoresistente.



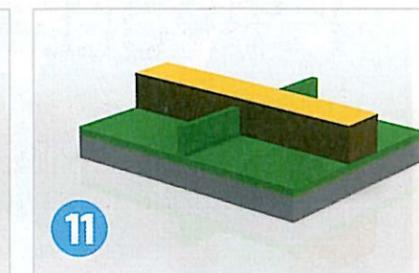
9

Si rimuove anche questo materiale (in blu nella foto precedente). Ecco l'aletta che diventerà il gate.



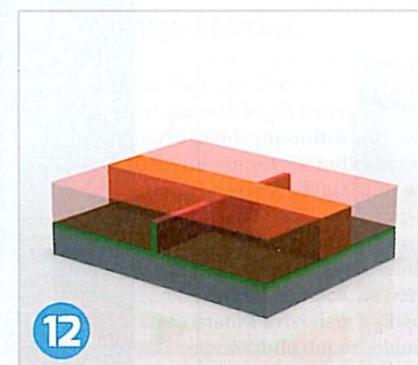
10

Sempre utilizzando la fotolitografia su una porzione del gate e si deposita solo su di esso un materiale isolante.



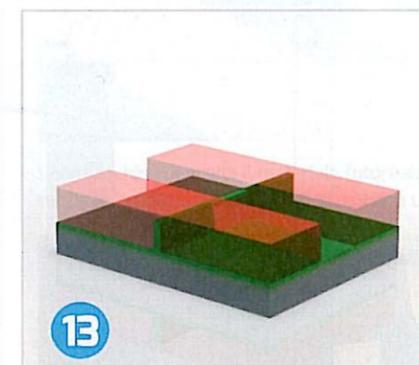
11

Di nuovo, con la fotolitografia, viene inserito un elemento di silicio policristallino che serve solo come forma per incisioni successive.



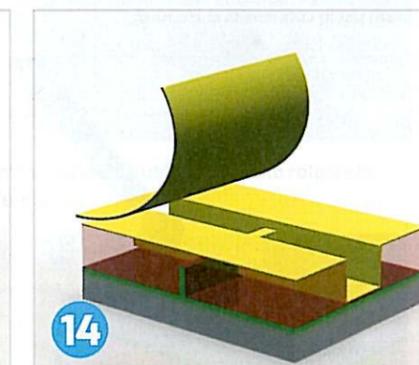
12

Si inserisce l'isolante, una copertura apposita che isola gli elementi interni del transistor dai vicini.



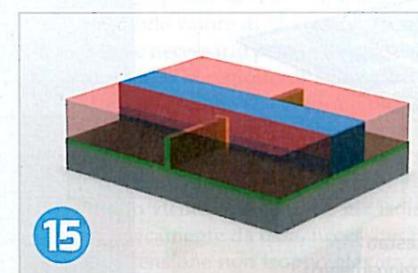
13

Tramite la tecnica "gate last" Intel rimuove il gate provvisorio scoperciando il cuore del transistor.



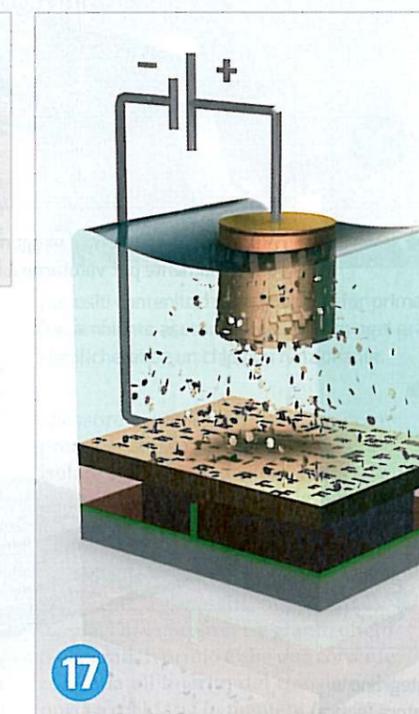
14

Viene applicato uno strato di dielettrico isolante ad alto coefficiente isolante su tutta la zona (high-k).



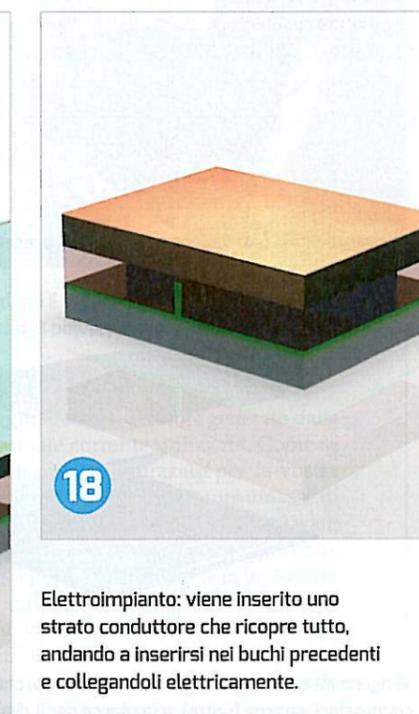
15

Viene riempito lo spazio precedente con un gate metallico, per dare migliori performance in combinazione con high-k.



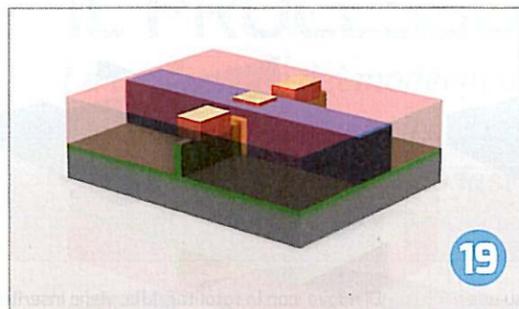
17

Sono scavati tre buchi nell'isolante che faranno da punti di contatto per source, drain e gate.

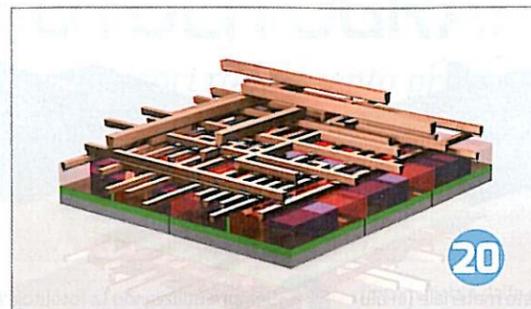


18

Elettroimpianto: viene inserito uno strato conduttore che ricopre tutto, andando a inserirsi nei buchi precedenti e collegandoli elettricamente.



Lo strato viene limato fino a mostrare i tre contatti metallici sulla superficie, non più in connessione tra loro.



Vari livelli di connessioni e isolanti sono impilati (fino a 10 piani) per interconnettere tra loro i transistor di una Cpu.



Separati uno a uno vengono tagliati con precisione e prelevati da macchinari appositi.



I processori sono pronti e vengono testati sommariamente per valutarne il funzionamento.



Vengono inseriti nel package che conosciamo, che integra le connessioni esterne (sotto) e la placca per il dissipatore (sopra).

quadro. Quindi tanto, davvero tanto. Per questo motivo una decina di anni fa è partita una corsa tecnologica tale da introdurre più cambiamenti negli ultimi 3 anni rispetto a quanto fatto nei precedenti 50 in termini di tecnologie produttive.

Per arginare inizialmente tali problemi i due maggiori produttori di processori hanno integrato alcune particolari strutture chimiche all'interno delle Cpu, in modo da ridurre le correnti parassite e avere processori più veloci e potenti (rispettando la legge di Moore), senza fondere su sé stessi appena accesi.

CURRENT LEAKAGE

Per superare i problemi relativi ai fenomeni quantistici gli ingegneri delle maggiori aziende produttrici di silicio hanno dovuto innovare radicalmente un settore che si era adagiato su uno standard ormai consolidato. L'impossibilità di salire in frequenza oltre i 4 GHz senza incorrere in gravi fenomeni di disturbo a livello atomico, ha portato in prima istanza a rivedere il singolo processo responsabile del problema. Punto focale della discussione è che, con processi produttivi sempre più raffinati (al di sotto dei 130 nm), lo spessore del materiale isolante (del semplice silicio drogato) posto tra il gate e il canale del gate, raggiunge l'incredibile valore di "1 atomo". Questo si rese necessario per avvicinare il più possibile la regione di svuotamento al gate (regione che si forma dall'applicazione di tensione al gate e la cui ampiezza varia in base alla tensione applicata sul gate). Quindi un gate molto vicino alla regione, ma isolato elettricamente da essa, necessiterà di una tensione non troppo elevata per generare un campo elettrico al suo interno sufficiente per modificarne lo stato da isolante a conduttore.

Questo significa anche che, vista la struttura atomica del silicio cristallino (immaginatelo disposto come un cubo di Rubik in cui ogni "cubetto" è un atomo di silicio), l'isolamento complessivo era costituito da un "foglio" spesso un singolo atomo. Paragonare tale valore alla tensione applicate al gate (tipicamente di poco superiore a 1,2 volt) risulta davvero complesso. Un voltaggio del genere è quello prodotto da una comune pila ricaricabile, e pensare di offrire tale tensione su un materiale molto semplice di uno



Una macchina deposita il materiale fotoresistente sul wafer. Si utilizza un piatto rotante in modo da "spalmare" tale strato in maniera uniforme.



Appositi contenitori accolgono i wafer prima e dopo il trattamento. Sono mantenuti separati e in ambiente asettico e controllato; ogni granello di polvere depositato in questa fase significherebbe un chip non funzionante.

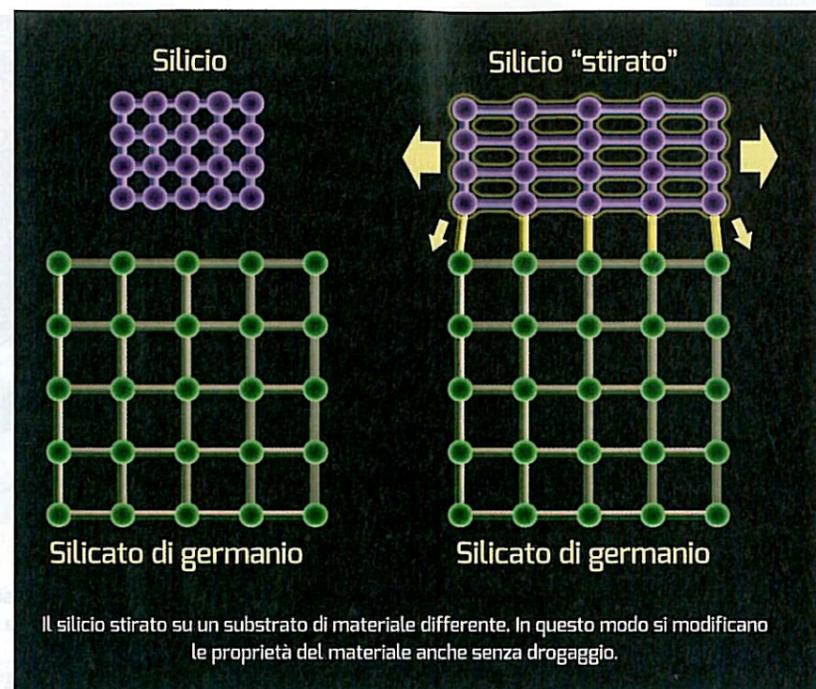
spessore tanto ridotto aiuta a comprendere come non fosse più possibile isolare elettricamente in maniera adeguata il gate dal suo canale. Questo comporta che parte dell'isolamento risulta permeabile e che una piccola corrente può fluire tra gate e drain. Una corrente parassita, ovvero indesiderata, che comporta tre grandi effetti collaterali. Il primo è che una corrente continua all'interno del transistor lo porta a scaldarsi in maniera costante,

aggiungendosi al calore generato dalla normale corrente utilizzata. Come se l'interruttore utilizzato per la vostra lampadina da 100 watt consumasse di suo, sia acceso sia spento, circa 10 watt, facendo passare corrente in ogni caso. Il secondo è che il segnale in uscita sarà disturbato da questa corrente costante che va a sommarsi a quella utile che attraversa il dispositivo. Sempre seguendo l'esempio precedente quando l'interruttore è spento la lampadina

non sarà mai spenta del tutto ma produrrà comunque una debole luce. Il terzo è che, vista la disposizione dei transistor nei circuiti logici, ovvero con l'uscita di molti connessa al gate di altri per controllarne lo stato, porti a una reazione a valanga tale da mantenere in tensione costante tutto il processore, non permettendo tecniche di risparmio energetico ottimali. In pratica, sempre per seguire l'esempio della lampadina, come se in casa vostra non fosse possibile staccare l'interruttore generale, ma ci fosse sempre un consumo di fondo.

STRAINED SILICON, HIGH-K E METAL GATE

Il primo passo evolutivo è stato raggiunto nel 2005 con la prima adozione del processo battezzato *Strained Silicon*, traducibile in maniera empirica come "Silicio stirato". La tecnica prevede proprio di stirare o comprimere il silicio di cui sono composti i transistor facendolo aderire a un materiale cristallino facilmente reperibile e integrabile nei circuiti elettronici (come il silicato di germanio *SiGe*), dotato di una struttura cristallina identica nella forma a quella del silicio, ma con dimensioni differenti. Stirando o comprimendo il normale silicio si ottiene una modifica delle proprietà di conduzione senza la necessità di dopare in maniera eccessiva il materiale, azione che ne ridurrebbe altre proprietà fondamentali. Introdotto da Intel con il processo produttivo a 65 nm risultò il primo passo verso una rapida transizione, abbandonando l'ultimo baluardo della

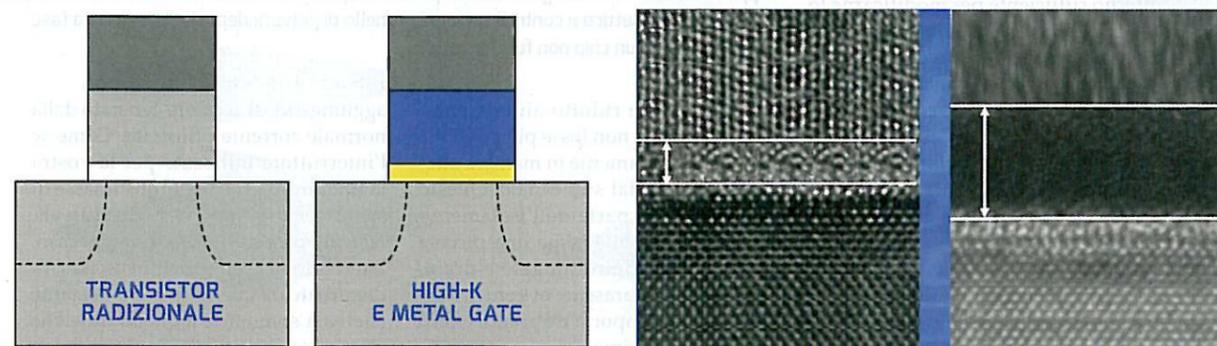


Il silicio stirato su un substrato di materiale differente. In questo modo si modificano le proprietà del materiale anche senza drogaggio.

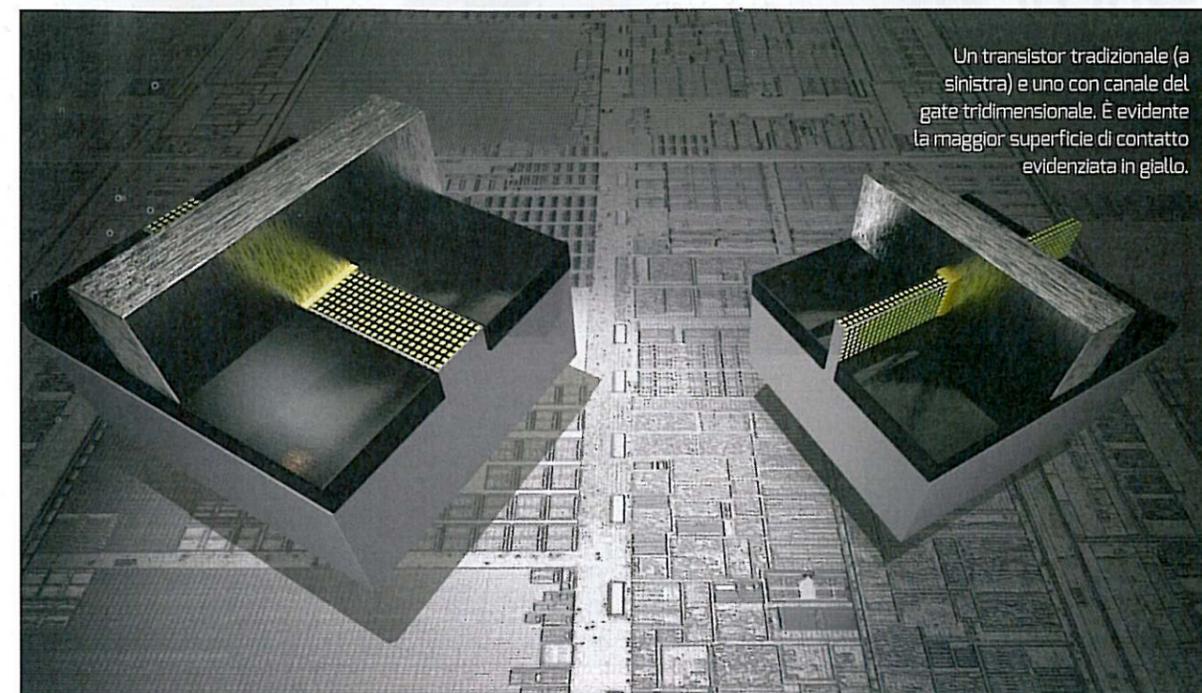
staticità produttiva (i 90 nm che tanti problemi hanno creato) per dirigersi verso il presente. Il secondo metodo per limitare gli effetti quantistici e le correnti parassite, introdotto da Intel nel 2007 con il proprio processo produttivo a 45 nm, è stato il cambiamento dei materiali utilizzati per la costruzione delle varie parti del transistor. Se il problema maggiore era infatti quello relativo al passaggio di corrente tra gate e canale fu introdotto il concetto di *high-k*. La sigla, che ingegneristicamente

parlando significa ad alta costante dielettrica, indica l'utilizzo di un materiale diverso dal semplice silicio per la costruzione dell'isolamento all'interno del transistor. Il processo di produzione resta nel complesso identico a quello in essere (vedi box dedicato), modificando solo il materiale da inserire. Passando semplicemente da uno strato di ossido di silicio a uno costituito da ossido di afnio (con una costante k di 25, ben 6 volte superiore a quella del silicio) è stato possibile ridurre in media le correnti parassite

METAL GATE E HIGH-K



Rimpiazzando il tradizionale dielettrico con un nuovo materiale, utilizzato insieme a un gate metallico, si riducono drasticamente le correnti parassite.



Un transistor tradizionale (a sinistra) e uno con canale del gate tridimensionale. È evidente la maggior superficie di contatto evidenziata in giallo.

di un fattore 100, riducendo di conseguenza il calore generato e tutte le controindicazioni espresse precedentemente. Questo strato è inoltre più spesso ("ben" 2 atomi) rispetto al precedente, rendendo necessaria una riconfigurazione anche del gate per ovviare alla distanza doppia che il campo elettrico deve coprire per generare lo svuotamento nella regione del canale.

Contestualmente fu infatti introdotto il concetto di "metal gate", sostituendo anche in questo caso l'unico elemento con il quale erano costruiti i transistor fino ad allora (il silicio) con un materiale metallico adatto. Questo materiale, diverso per transistor di tipo p o di tipo n visto l'utilizzo differente del gate, permette di utilizzare una tensione elettrica minore per ottenere lo stesso risultato in termini di campo elettrico, andando ad azzerare il gap introdotto dall'allontanamento del gate dal proprio canale controllato.

UN CAMBIO DI FORMA

Storicamente Intel e Amd hanno utilizzato metodi differenti per ottenere lo stesso scopo, ovvero quello di limitare le correnti parassite. I due contendenti hanno però gradualmente adottato tutte le tecniche possibili, andando con

il tempo a offrire tecnologie piuttosto simili, fino almeno al 2012 quando Intel portò a conclusione un progetto decennale che rivoluzionò in maniera estrema la costruzione stessa del transistor e che ha tracciato la strada per i successivi 10 anni. Con il processo produttivo a 22 nm Intel propose un cambio radicale della forma stessa del transistor. Se in oltre 50 anni di storia il transistor ha sempre vissuto in maniera bidimensionale, ovvero con source e drain posti sul medesimo piano e il canale del gate piatto tra di essi, Intel diede vita commercialmente parlando ai primi transistor di tipo tridimensionale. La grande differenza è che il canale del gate non è più complanare a source e drain, ma risulta sollevato rispetto

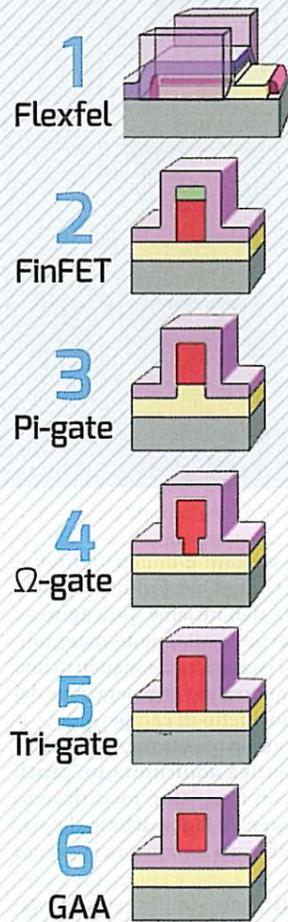


Dopo oltre 50 anni il transistor a struttura 3D rivoluziona il modo stesso di concepire l'elettronica

agli altri contatti e immerso direttamente nel gate. Quest'ultimo, vista la geometria dell'insieme, può agire su una superficie nettamente maggiore rispetto al passato, necessitando di una tensione minore per la generazione del medesimo effetto di campo. Intel integra in questa nuova tecnologia produttiva addirittura tre canali per ciascun transistor, modificando definitivamente l'idea di base che siamo stati abituati a vedere nei decenni passati. I risultati sono stati da subito evidenti, così come quelli ottenuti con l'evoluzione a 14 nm (l'ultima oggi disponibile), del 2014. Rispetto agli ultimi modelli completamente in silicio (ad esempio un Pentium 4 Prescott) i processori attuali integrano oltre 10 volte i transistor stipati nel processo a 90 nm (125 milioni contro 1.400 milioni), hanno prestazioni medie di 25 volte superiori (facendo una media sui vari ambiti) e un consumo ridotto di oltre il 50% (135 watt contro gli attuali 65 watt). Negli ultimi 2 anni abbiamo assistito a un affinamento di tale tecnologia, con il processo a 14 nm che ne migliora le caratteristiche sotto ogni punto di vista. La forma e l'idea originale resta la stessa, ma i canali penetrano molto più in profondità nel gate (vedere immagine in questa

MULTIGATE

Alcune delle geometrie proposte per il futuro del transistor. Gli ingegneri, liberi dai vincoli del passato, stanno studiando sistemi innovativi.



pagina), permettendo di ridurre da 3 a 2 i picchi 3D necessari al funzionamento, compattando gli elementi e permettendo un'integrazione eccezionale a livello di silicio.

LA NUOVA PRIMAVERA DEI TRANSISTOR

A oggi, con il processo Intel a 14 nm, l'evoluzione tecnologica ha raggiunto un punto estremamente interessante. Nuove frontiere si stagliano all'orizzonte grazie alla spinta del settore di ricerca e sviluppo nata negli ultimi 10 anni. La presa di coscienza che le tecnologie standard non avrebbero portato altri risultati ha spinto i maggiori produttori in ricerche avanzate su diversi fronti che, prevedibilmente, stanno dando i propri frutti dopo qualche anno di studio e sviluppo.

Il primo punto su cui l'industria si concentra rappresenta un déjà-vu, come una decina di anni fa i primi metodi evolutivi riguardarono il cambiamento degli ultimi elementi modificati anche in questo caso sono molti gli studi riguardanti la forma geometrica del transistor. I nuovi processi produttivi in grado di costruire forme tridimensionali hanno infatti aperto le porte

a una progettazione più creativa di questo elemento, con diverse forme già progettate e in fase di prototipazione per riuscire a scendere sotto la soglia dei 10 nm.

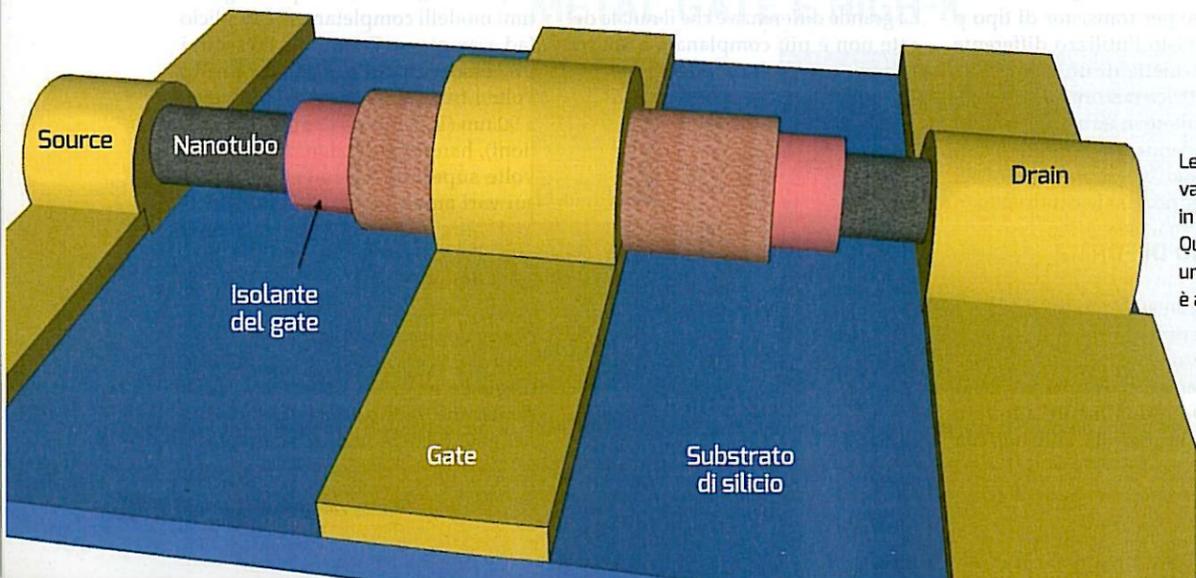
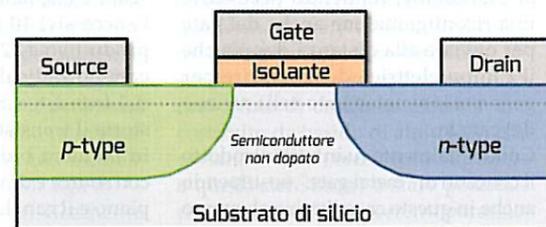
Le più interessanti e in fase di studio si concentrano sulla ridefinizione diretta della forma specifica del gate e del canale. Gli ingegneri hanno infatti rivisto il concetto stesso del Mosfet, partendo da un foglio bianco disegnando forme tali da rendere possibile una riduzione massima degli ingombri fermo restando il funzionamento da interruttore tripolare del dispositivo. In questa pagina trovate alcune delle forme già studiate e in parte realizzate sotto forma di prototipi, tramite i quali valutare le specifiche e realizzare su larga scala il più promettente.

Dal punto di vista tecnico le necessità sono le stesse di sempre: produrre transistor sempre più piccoli (per inserirne di più nel medesimo spazio) e operanti a frequenze più elevate con consumi più bassi possibili.

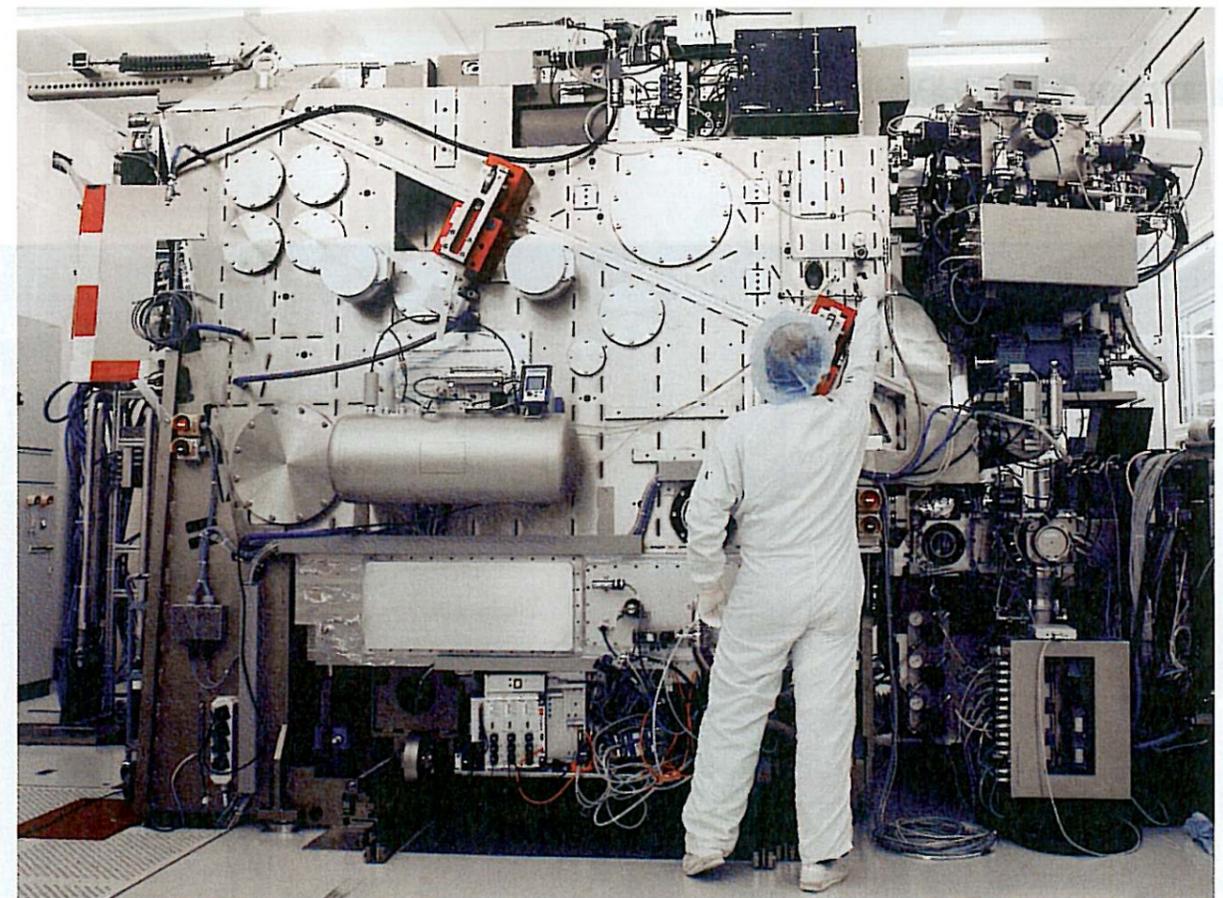
I più promettenti ad oggi sono il GAAFET (*Gate All Around FET*), il CNFET (*Carbon Nanotube FET*) e il TFET (*Tunnel FET*). Nel GAAFET il gate circonda completamente il canale su ogni lato,

TFET: TRANSISTOR A EFFETTO TUNNEL

Cambia il concetto del transistor: non più pMos e nMos, ma un unico elemento con source-p e drain-n.



Le geometrie possono variare in maniera continua in base ai materiali adottati. Qui a lato un CNFET, in cui un nanotubo di carbonio è avvolto nel gate.



Una macchina per la fotolitografia in un laboratorio sperimentale in Germania.

in modo da poter utilizzare una piccola tensione per creare al suo interno una zona con differente conducibilità. Il principio è quello di un "girarrosto", nel quale il "calore" (ovvero il campo elettrico) viene fornito da ogni direzione, permettendo di "riscaldare" (modificare la conducibilità) in maniera uniforme e non solo da un lato. Questa tipologia è la più vicina alla realizzazione pratica in quanto i materiali utilizzati sono pressappoco i medesimi degli attuali. Non ci stupiremmo se Intel utilizzasse una tecnica di questo tipo per le prossime generazioni con processi produttivi da 10 e da 7 nm.

Il CNFET rappresenta invece un'evoluzione importante sia dal punto di vista dei materiali sia da quello concettuale. Innanzitutto si abbandona il rassicurante silicio per passare a materiali più esotici, in questo caso i nanotubi di carbonio. Un nanotubo di carbonio non è nient'altro che un foglio di grafene (un materiale composto da un foglio di singoli atomi di carbonio) arrotolato su se stesso. Diversi esperimenti fatti utilizzando questo materiale in sostituzione diretta del canale del gate hanno portato a vari prototipi

con forme e caratteristiche differenti. Il più promettente è relativo a un'evoluzione del GAAFET con un cambio completo dei materiali, utilizzando un nanotubo di carbonio che collega Source e Drain avvolto in un gate metallico tradizionale che, con una tensione applicata, modifica le proprietà interne del nanotubo permettendo o meno la conduzione. In rapporto a un transistor tradizionale le prestazioni sono eccezionali: a parità di tensione sul gate la velocità di reazione è tripla, così come la conducibilità elettrica risultante. Nella pratica è possibile utilizzare tensioni di 0,5 volt per ottenere prestazioni doppie rispetto ai modelli attuali, con conseguenti consumi irrisori e possibilità di avere frequenze operative ben oltre i 6 GHz.

Altra possibile evoluzione sono i transistor a effetto tunnel, che utilizzano le specifiche conoscenze in meccanica quantistica elaborate negli ultimi decenni per creare un transistor che utilizzi un principio di funzionamento opposto a quello tradizionale. La struttura geometrica è molto simile a quella di un Mosfet, con source e drain planari separati da un canale del gate sovrastato dal

contatto di controllo. La differenza è che source e drain sono costituiti da silicio drogato in maniera opposta (p ed n, mentre nei Mosfet sono entrambe di tipo p o entrambe di tipo n, con il canale del tipo opposto), mentre il canale diretto non è drogato e costituito da silicio puro.

Il principio di funzionamento deriva dalle scoperte fatte durante i primi problemi emersi con le correnti parassite. Gli ingegneri, studiando tale fenomeno, hanno trovato un modo per utilizzarlo invece del processo tradizionale, trasformando una debolezza della tecnologia in una nuova promettente frontiera.

Dando una tensione ben definita al gate ci si trova nella condizione in cui parte della carica fluisce per effetto tunnel quantistico attraverso l'isolante, andando a ionizzare direttamente il canale e permettendo (per via del differente drogaggio di source e drain) il passaggio controllato di corrente.

LA LUCE NEI TRANSISTOR

Come abbiamo visto nel processo produttivo attuale, la luce gioca un ruolo di primaria importanza nell'incisione degli elementi che compongono i transistor. Uno degli elementi che può contribuire a frenare l'evoluzione

Nuove forme

Il cambio radicale nella forma stessa ha permesso un enorme salto tecnologico in avanti

è rappresentato dal fatto che la luce utilizzata per incidere tali elementi ha una lunghezza d'onda tipica che, se in passato non ha mai rappresentato un problema, inizia a risultare limitante. La luce adottata per la fotolitografia è prodotta dall'eccitazione di gas nobili in modo da ricreare una luce con uno spettro di emissione il più stretto possibile.

Nella pratica si scalda un gas ad alta temperatura che emette di conseguenza fotoni con una determinata lunghezza d'onda. Utilizzare un gas nobile permette di avere fotoni tutti (o quasi) con la stessa frequenza, in modo da incidere il silicio in maniera adeguata.

Se i primi modelli con lampade al mercurio producevano luce con lunghezza d'onda di 436, 405 e 365 nm, i successivi affinamenti hanno portato all'utilizzo degli attuali laser a fluoro di argon con spettro strettissimo a 193 nm. Attualmente dunque si utilizza luce con lunghezza d'onda di 193 nm per incidere transistor grandi mediamente 14 nm.

In pratica è come se si utilizzasse un'onda marina alta 1,93 metri per riempire il fossato di un castello di sabbia alto 14 centimetri. È evidente l'impossibilità della cosa, perlomeno senza gli accorgimenti (vedi lenti

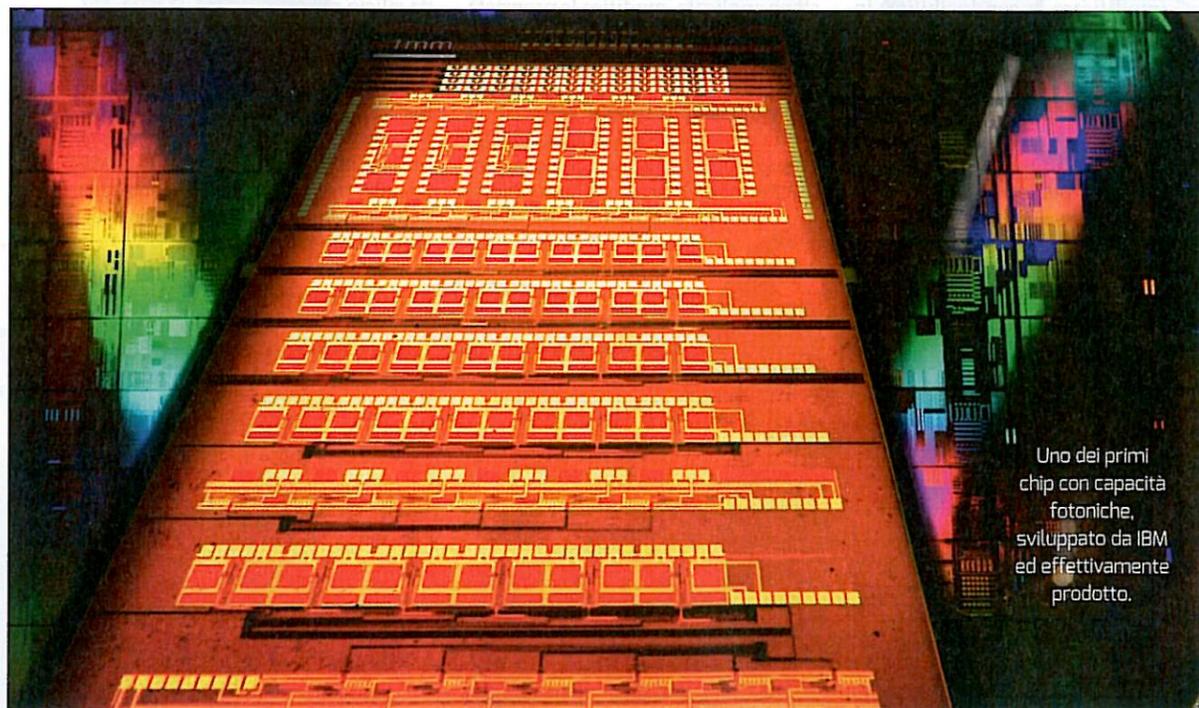


Il primo prototipo per la Extreme UltraViolet Lithography ottenuto al Fraunhofer Institute.

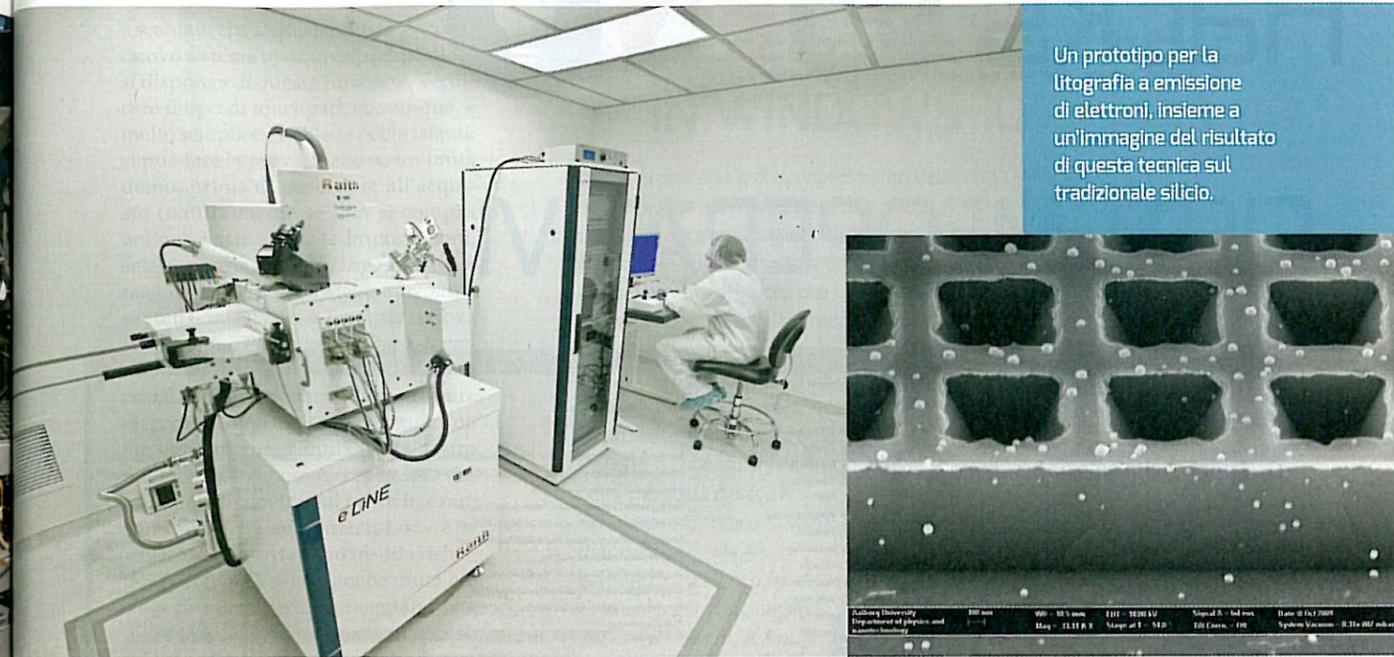
speciali che contraggono l'onda luminosa o immersione del wafer in acqua) messi in atto a livello produttivo. Questi accorgimenti sono però sempre più complessi e ci si rende conto di essere molto vicini al limite massimo della tecnologia.

In dirittura d'arrivo tra le varie

tecniche di nanolitografia, essenziale per il nodo produttivo a 7 nm previsto tra qualche anno, risulta essere la tecnica detta EUV (*Extreme Ultraviolet Lithography*) in cui è utilizzato un gas altamente ionizzato in grado di emettere fotoni estremamente energetici con lunghezza d'onda di 13,5 nm.



Uno dei primi chip con capacità fotoniche, sviluppato da IBM ed effettivamente prodotto.



Un prototipo per la litografia a emissione di elettroni, insieme a un'immagine del risultato di questa tecnica sul tradizionale silicio.

Questa tecnica rappresenta il punto di svolta dell'intero settore e permetterà di scendere sempre più nell'integrazione atomica dei transistor. I problemi ancora da affrontare sono molti e soprattutto relativi alla volontà degli ingegneri di creare strutture dei transistor completamente nuove. Fotoni tanto energetici sono infatti in grado di rovinare molti materiali che, per ovvi motivi, non possono essere utilizzati per la costruzione.

ELECTRON BEAM LITHOGRAPHY

In alternativa alla luce è allo studio l'integrazione di una litografia basata sull'emissione di elettroni, in grado di scavare nel silicio tracce di dimensioni infinitesimali (siamo ben al di sotto del singolo nanometro) e superare in questo modo di slancio qualunque ostacolo posto dal dimensionamento dei transistor.

Utilizzare elettroni e direzionarli in maniera adeguata è molto più semplice di quanto possa sembrare e adottare materiali reattivi a essi una sfida non troppo complessa.

Questa tecnologia, che sulla carta sembra immediatamente vincente, si scontra però con la pessima "abitudine" degli elettroni di interagire con gli atomi circostanti. Elettroni sparati ad alta velocità contro un materiale lo possono incidere senza problemi, ma

contestualmente lo caricano elettricamente e si legano a esso in maniera intrinseca. Quando non lo fanno interagiscono con gli atomi prendendo il posto di un elettrone esistente che viene espulso dagli orbitali in una direzione casuale, che comporta la creazione di tracce non perfettamente definite sulla superficie del materiale. Superati questi "piccoli" inconvenienti, soprattutto utilizzando elementi diversi per transistor e substrato (carbonio, grafene e simili), la litografia elettronica potrà spingere l'immaginazione dei progettisti a livelli finora impensati.

CMOS PHOTONICS

In un futuro ancora più lontano gli architetti delle grandi aziende stanno già ipotizzando un cambiamento ancora più epocale, ovvero l'abbandono dei segnali elettrici e del funzionamento

elettronico dei processori rimpiazzati dal fenomeno più rapido che si conosca: la luce. Transistor (intesi come interruttori a tre poli, indispensabili per la realizzazione della logica binaria) che si basano sugli effetti dei fotoni e non più sugli elettroni. Transistor che utilizzano impulsi luminosi anziché elettrici per trasmettere informazioni. Se la logica del processo è molto semplice la sua realizzazione pone all'attenzione degli ingegneri un solo grande problema: costruire dei transistor miniaturizzati con materiali semplici (ricordiamo che devono poi essere prodotti in maniera industriale) in grado di emettere luce e risultare sensibili ad essa. Elementi del genere esistono da decenni, basti pensare alle comunicazioni in fibra ottica, nelle quali un laser viene modulato (l'equivalente di acceso/spento) e l'informazione viene ricevuta da un fotomoltiplicatore che converte gli impulsi in segnali elettrici.

“*L'Extreme Ultraviolet Lithography (Euv) si basa sulla produzione di luce a frequenze di 13,5 nanometri, al limite della zona ultravioletta, a ridosso dei raggi X.*”