Els premis Nobel de Física del 2000

Lluís Prat Viñas*

En l'anunci de concessió dels premis Nobel de Física de l'any 2000,[1] comunicat el passat 10 d'octubre, es pot llegir:

«The Royal Swedish Academy of Sciences has awarded the Nobel Prize in Physics "for basic work on information and communication technology".

The prize is being awarded with one half jointly to Zhores I. Alferov, A. F. Ioffe Physico-Technical Institute, St. Peterburg, Russia, and Herbert Kroemer, University of California at Santa Barbara, California, EUA, "for developing semiconductor heterostructures used in high-speed and opto-electronics" and one half to Jack S. Kilby, Texas Instruments, Dallas, Texas, EUA, "for his part in the invention of the integrated circuit".»

Per presentar els premis Nobel d'enguany, ens hem de referir, per tant, a les tecnologies de la informació i les comunicacions, als circuits integrats, a les heteroestructures i la seva aplicació a dispositius d'alta velocitat i a l'optoelectrònica.

Les tecnologies de la informació i les comunicacions. Una visió panoràmica

Les tecnologies de la informació i les comunicacions tenen una gran influència en la vida quotidiana de la humanitat dels nostres dies. La televisió, la ràdio, el telèfon, els ordinadors, Internet, etc. configuren un estil de vida radicalment diferent del que tenien els nostres avis, i ens seria molt difícil acostumar-nos a viure sense ell. Hi ha qui considera que la irrupció d'aquestes tecnologies és tan o més important que la Revolució Industrial o la invenció de la impremta. I tot això ha succeït en menys d'un segle, i continua creixent.

Per situar l'aportació dels premis Nobel del 2000 en el desenvolupament de les tecnologies de la informació i les comunicacions, res millor que presentar una breu cronologia dels principals fets que han marcat la seva història. Es pot considerar que hi ha tres grans etapes: una gestació que s'allarga fins a mitjan segle XIX; el naixement que coincideix amb el del segle XX; i una infantesa que dura uns cinquanta anys i que és seguida d'una metamorfosi i d'un creixement espectacular que continua fins avui.

A la fase de gestació s'han d'assenyalar alguns noms i dates importants:
1833: Babbage idea el precursor mecànic dels computadors actuals.
1837: Morse inventa la telegrafia elèctrica.
1847: Boole publica la seva àlgebra de proposicions lògiques.
1865: Maxwell presenta la teoria electromagnètica.
1876: Bell inventa el telèfon.
1877: Edison crea el fonògraf.
1896: Marconi transmet i detecta ones electromagnètiques a distància.
1897: Braun crea el tub de raigs catòdics.

S'acostuma a dir que l'electrònica neix el 1904 quan Fleming inventa el díode de buit. La primera infantesa de l'electrònica arriba el 1947, quan es descobreix el transistor bipolar. Durant aquesta etapa l'electrònica fa els primers passos íntimament lligada al desenvolupament de les comunicacions i amb el suport de la tecnologia de les vàlvules de buit. Es creen les grans teories i circuits que fan possible el desenvolupament de l'electrònica analògica. La teoria de la informació desenvolupada per Shanon i Kotelnikov i la de la realimentació per Black, Bode i Nyquist són d'aquesta època. També ho són la síntesi de circuits amplificadors i oscil·ladors, de l'amplificador operacional, i de circuits de comunicacions com per exemple el PLL. Neixen les grans aplicacions de l'electrònica, com ara la ràdio, la televisió i la computació. Comença, encara que molt tímidament, el desenvolupament de l'electrònica digital, i s'utilitza l'àlgebra de Boole per a l'anàlisi i el disseny de circuits de commutació. Unes dates ens ajudaran a situar aquests esdeveniments:
1904: Fleming crea el díode de buit.
1906: L. de Forest inventa el tríode (l'*audión*) i l'aplica a la telefonia.
1914-18: Primera Guerra Mundial: gran desenvolupament dels sistemes de ràdio portàtil.
1920: primera emissora comercial de ràdio.
1923: Baird transmet la primera imatge de televisió.
1928: primera demostració experimental de televisió.
1935: apareix el radar.
1939: primera televisió comercial als EUA.

***Lluís Prat** (Manlleu, 1946) és enginyer de telecomunicació per la Universitat Politècnica de Madrid (1971) i doctor per la Universitat Politècnica de Catalunya (1981). Actualment és catedràtic del Departament d'Enginyeria Electrònica de l'Escola Tècnica Superior d'Enginyeria de Telecomunicació de la UPC.

[1]Internet: http://www.Nobelprizes.com

1939-45: Segona Guerra Mundial: gran desenvolupament del radar.
1945: primer computador electrònic: l'ENIAC.

Les vàlvules de buit tenien dos inconvenients molt importants: un consum molt gran, ja que calia escalfar fins a posar vermell un filament que emetia electrons, i una vida molt curta a causa del desgast que sofria aquest filament. Altres característiques desfavorables eren el pes, el volum i la fragilitat. És fàcil d'entendre, per tant, que hi havia molt interès a trobar uns dispositius alternatius d'estat sòlid que permetessin superar aquestes limitacions. M. J. Kelly, director d'investigació als laboratoris Bell, va formar el 1945 un grup de recerca integrat per Shockley, Brattain i Bardeen, que el 23 de desembre de 1947 descobreix el transistor bipolar mentre intentava realitzar un transistor d'efecte de camp en un cristall de germani. Aquest descobriment va significar un ressorgiment de l'electrònica, un nou naixement, una reconstrucció de tot el que s'havia fet fins aleshores. En uns anys van desaparèixer els antics equips construïts amb vàlvules de buit, que van ser substituïts per nous circuits amb transistors. Algunes dates significatives d'aquesta època són:
1947: es descobreix el transistor bipolar de germani.
1954: primer transistor bipolar de silici.
1957: *Kroemer publica la teoria del transistor bipolar d'heterojunció.*
1958: *Kilby inventa el primer circuit integrat* (CI).
1960: primer làser (de robí).
1962: primer làser semiconductor (que funciona a temperatura del nitrogen líquid).
1963: *Alferov i Kroemer proposen el làser d'heteroestructura.*
1969: el Departament de Defensa dels EUA crea ARPANET, precursora d'Internet.
1970: *primer díode làser de sortida contínua, a temperatura ambient (Alferov i Panish).*
1971: primer computador en un CI (Intel 4004).
1975: tecnologia MBE per a heteroestructures. Neix el terme PC: (personal computer).
1983: primer sistema comercial de telefonia mòbil als EUA (AMPS).
1985: el sistema operatiu Windows és creat per Microsoft.
1992: es crea Internet Society.

Els primers transistors es van comercialitzar l'any 1951. Eren fets de germani, un semiconductor que té l'inconvenient de ser molt sensible a l'augment de temperatura. G. Teal, de Texas Instruments, va aconseguir fer transistors amb un altre semiconductor, el silici, que segueix sent el material electrònic bàsic dels nostres dies. Un nou salt endavant, d'una importància crucial, fou la invenció del circuit integrat (CI) per J. Kilby, un dels guardonats amb el Premi Nobel d'aquest any. A partir del seu descobriment l'electrònica va passar de ser discreta, basada en la connexió de components individuals, a esdevenir integrada, basada en xips, cristalls de silici que contenen tot el circuit.

Un altre fet important d'aquest ressorgiment de l'electrònica fou el naixement de la fotònica,[2] els circuits que tracten amb senyals òptics. Els díodes electroluminiscents, els làsers i les fibres òptiques són components essencials d'aquesta nova tecnologia. La fotònica va néixer l'any 1960 amb el làser de robí desenvolupat per Maiman. El desenvolupament, però, de l'electrònica dels fotons requeria nous materials semiconductors diferents del silici, els anomenats *semiconductors de gap directe*, i noves estructures combinant diferents semiconductors. Alferov i Kroemer van proposar el 1963 els làsers d'heterojunció que van permetre realitzar dispositius operatius per a comunicacions òptiques a començament dels anys setanta. Al mateix temps, les heteroestructures van permetre dissenyar uns dispositius més ràpids, els transistors d'heterojunció, que van empenyent cada cop més lluny la freqüència d'operació dels circuits electrònics, actualment per sobre dels 500 GHz. Encara que la idea bàsica va ser enunciada per Shockley el 1950, Kroemer va presentar la teoria d'aquests transistors l'any 1957. Calgué esperar, però, fins al final dels anys setanta perquè el desenvolupament de noves tecnologies fes possible la seva realització pràctica.

Durant aquesta segona joventut de l'electrònica l'accent ha passat dels circuits analògics als circuits digitals, i la computació ha experimentat un desenvolupament imparable. Amb el suport dels circuits integrats, els ordinadors han emprès una cursa de creixement en potència i velocitat que converteix en obsolets els equips que tenen una edat superior als cinc anys. La connexió d'aquests ordinadors en una xarxa planetària, Internet, obre les portes de la humanitat a una nova civilització global que està esborrant fronteres i muralles.

Jack S. Kilby i la invenció del circuit integrat

En un article (Kilby, 1976) Jack Kilby (1923, Jefferson City (Missouri, EUA) explica en detall la seva invenció del circuit integrat. Graduat en enginyeria elèctrica a la Universitat d'Illinois el 1947, va començar a treballar a l'empresa Centralab en el disseny d'amplificadors per a sords, equip que per les seves característiques específiques requeria l'aplicació de tècniques de miniaturització. El 1952 Centralab va començar a fabricar transistors bipolars de germani per a aquests equips, però aquella era una època de canvi tecnològic molt accelerat, que exigia inversions importants en noves tecnologies (el germani ja començava a ser obsolet enfront del silici) que superaven les capacitats de Centralab, i Kilby va decidir canviar d'empresa.

El maig de 1958 va ser contractat per Texas Ins-

[2] Internet: http://www.vislab.usyd.edu.au/photonics/revolution/index.html

truments, una gran empresa del sector electrònic, amb l'encàrrec de treballar a l'àrea de microminiaturització. A l'estiu es va quedar sol a la fàbrica, ja que no tenia dret a les vacances perquè acabava de ser contractat. Sense possibilitat d'incidir en els processos de fabricació, va dedicar-se a dissenyar un circuit digital simple, un biestable, construït només amb semiconductors: transistors, resistències i condensadors fets amb silici que interconnectava amb fils metàllics. Pensava que una fàbrica de semiconductors podria reduir d'aquesta manera el cost dels equips que produïa pel fet d'estalviar-se d'adquirir els altres components a l'exterior. Va muntar el circuit i el 28 d'agost d'aquell any va demostrar que funcionava.

El següent pas fou realitzar el circuit sobre un únic tros de semiconductor. Va aprofitar unes oblies de germani per realitzar un oscil·lador de desplaçament de fase, connectant amb fils d'or els components realitzats sobre un únic tros de germani. Va demostrar que funcionava el 12 de setembre de 1958. Acabava de néixer el primer circuit integrat. Com ell mateix deia, un circuit integrat molt bast, molt rudimentari, que s'assemblava molt poc als circuits integrats actuals. El 19 de setembre ampliava l'experiència construint de la mateixa manera tres biestables per demostrar que també es podien fer circuits digitals. Va sol·licitar la patent el febrer de 1959 (figura 1) i va presentar el nounat al gran públic en el congrés de l'IRE del març següent amb el nom de circuit sòlid.

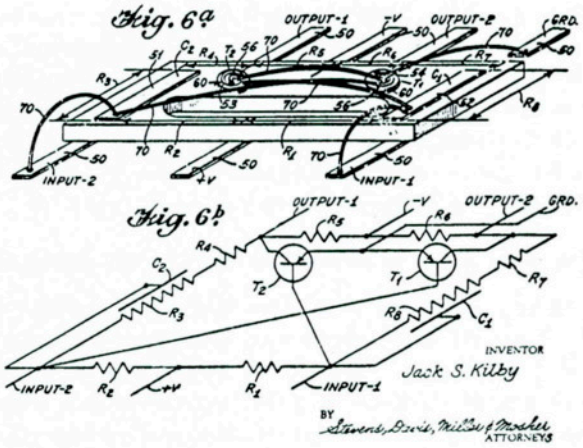

Figura 1: *Patent de circuit integrat presentada per J. Kilby el 6 de febrer de 1959*

Com molt sovint passa en el camp de la ciència, gairebé alhora i de manera totalment independent, R. S. Noyce va realitzar també un circuit integrat (Wolf, 1976). Sospito que si en Noyce no hagués mort el 1990 hauria compartit el Nobel amb Kilby, de la mateixa manera que el comparteixen Kroemer i Alferov per ser coinventors de les heteroestructures. Aquesta coincidència d'invents o descobriments, que són independents i quasi simultanis, té una freqüència massa elevada per ser producte de l'atzar. Jo ho comparo amb un gran roser a punt de florir. Una de les seves poncelles serà la primera a obrir-se i tindrà un color i un aroma ben definit, però si aquesta no hagués florit ho hauria fet la del costat. Perquè qui floreix en realitat és el roser.

Doncs bé, la societat d'aquella època reclamava la invenció del circuit integrat. En un congrés de components electrònics celebrat el 1952, l'anglès G. W. A. Dummer havia dit: «Amb l'adveniment del transistor i el treball en general amb semiconductors, sembla possible imaginar un equip electrònic realitzat en un bloc sòlid sense fils de connexió. Aquest bloc podria estar fet de capes de materials aïllants, conductors, rectificadors i amplificadors, i així es definirien les seves funcions elèctriques pel fet de tallar i connectar àrees de les diverses capes.» Era una intuïció visionària, ja que no existien capes amplificadores ni rectificadores, però sembla talment la descripció dels actuals circuits integrats per algú que hagués pogut dirigir una mirada fugaç al futur.

La necessitat d'aquests circuits ja venia de lluny. Els avions B-29, utilitzats a la Segona Guerra Mundial, requerien un miler de vàlvules de buit i unes quantes desenes de milers de dispositius passius. Era evident que el futur estava limitat pel cost, el volum i la fiabilitat d'aquests equips. No era gens estrany que l'Exèrcit, la Marina i la Força Aèria nord-americana financessin molts projectes d'investigació que cercaven la miniaturització dels equips electrònics, alguns dels quals apuntaven, fins i tot, cap al desenvolupament d'una electrònica molecular.

Robert Noyce va graduar-se a la Universitat de Grinnell el 1949, i va fer el doctorat en electrònica física a l'Institut Tecnològic de Massachusetts (MIT). Després de treballar divuit mesos a l'empresa que havia creat W. Shockley, un dels descobridors del transistor, va crear el 1957, juntament amb altres set col·legues, l'empresa Fairchild Semiconductors. Més tard, el 1968, fundava amb Gordon Moore l'empresa Intel, que, des que va crear el primer microprocessador en un xip el 1971, no ha deixat d'encapçalar l'electrònica de computadors (els Pentium dels nostres ordinadors personals actuals porten la seva signatura).

Fairchild es dedicava a la fabricació de transistors de silici d'alta velocitat. El 1958, J. Hoerni, un dels socis fundadors, va desenvolupar el procés planar, que permetia eliminar les inestabilitats superficials dels dispositius semiconductors d'aquella època cobrint la superfície del silici amb una capa d'òxid. Aprofitant aquesta nova tecnologia, Noyce dibuixava el 23 de gener de 1959 en el seu quadern de laboratori un esquema en què es feien múltiples dispositius sobre una única peça de silici i es realitzaven les interconnexions mitjançant pistes metàlliques dipositades sobre de la capa aïllant superficial,

com una part més del procés de fabricació. El 30 de juliol de 1959 presentava la patent del seu circuit integrat, que tenia totes les característiques dels circuits integrats actuals (figura 2). En aquest sentit anava més enllà que Kilby.

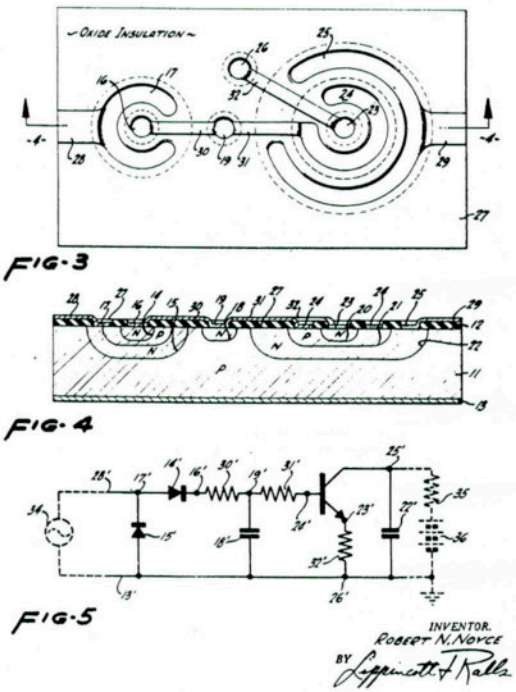

Figura 2: *Patent de circuit integrat presentada per R. Noyce el 30 de juliol de 1959*

L'impacte que significaria la creixent complexitat que permetien els circuits integrats queda perfectament reflectida en una frase de Noyce del 1977, pocs anys després d'haver creat el primer microprocessador comercial integrat en un xip: «Un microcomputador d'avui té més capacitat de computació que el primer gran computador electrònic (ENIAC). És 20 vegades més ràpid, té una memòria més gran, és milers de vegades més fiable, consumeix la potència d'una bombeta en lloc de la d'una màquina de tren, ocupa un volum 30.000 vegades inferior, costa com a molt 10.000 vegades menys, i pot adquirir-se per correu o a la botiga de la cantonada.» Vint-i-cinc anys després podríem dir una cosa similar dels computadors d'avui respecte d'aquells a què es referia Noyce.

Circuits integrats i ordinadors

Els circuits integrats han revolucionat tota l'electrònica, tant l'analògica com la digital. Malgrat això, el camp on la seva incidència ha tingut efectes més espectaculars ha estat el de la computació. Els circuits integrats han permès augmentar en ordres de magnitud la capacitat i velocitat de tractament de la informació dels ordinadors i això ha fet possible, en definitiva, que ja no sigui una quimera pensar que una xarxa planetària comunicarà tots els éssers humans mitjançant ordinadors personals.

Els circuits digitals s'han desenvolupat a través dels transistors MOS, que són els dispositius que permeten una major densitat d'integració. A la figura 3 es mostra l'estructura d'aquest transistor, que consisteix essencialment en un interruptor, en el qual es tanca o s'obre la connexió entre els terminals de drenador i de font, aplicant una tensió a un tercer terminal anomenat *porta*.

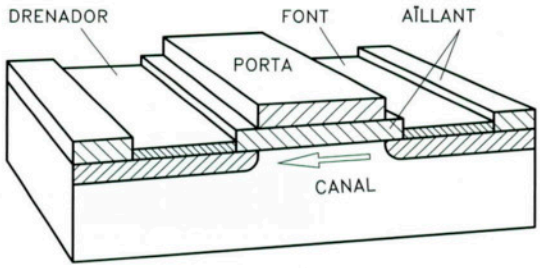

Figura 3: *Estructura d'un transistor MOS. El condensador de porta format per metall-òxid (aïllant)-semiconductor dóna nom a aquest transistor*

Per tenir una certa idea de la importància que ha tingut el CI en la revolució electrònica, cal fer una breu incursió en la tecnologia dels circuits integrats de silici. A la figura 4 hi ha un esquema de les principals etapes per fabricar un xip. L'inici és la sílice (sorra), un dels materials més abundants de la naturalesa. Mitjançant processos físics i químics s'aconsegueix silici de gran puresa: té un àtom d'impuresa per cada mil milions d'àtoms de silici, la mateixa proporció que un individu respecte a tota la població humana actual. Aquest material li manca, però, una condició per ser silici electrònic: la perfecta ordenació espacial dels seus àtoms, és a dir, una estructura perfectament cristal·lina. Això s'aconsegueix, per exemple, amb el mètode Czokralski, que permet crear un cilindre monocristal·lí a partir del silici fos i una llavor cristal·lina. Aquest cilindre es talla posteriorment en discs, anomenats *oblies*, que són la matèria primera per fer dispositius electrònics. El diàmetre de les oblies on es van fer els primers CI era d'uns cinc centímetres, però ha anat augmentant regularment des d'aleshores i avui és d'uns trenta centímetres.

Un circuit integrat consisteix en la interconnexió de regions N i P creades dins del cristall de silici mitjançant pistes metàl·liques dipositades sobre una capa d'òxid superficial. Les regions N són molt riques en electrons lliures i s'aconsegueixen inserint àtoms donadors al cristall de silici. Les regions P són deficitàries en electrons lliures i es fan afegint impureses acceptores al cristall. La inserció d'àtoms d'impuresa en el cristall de silici s'anomena *dopatge* i s'aconsegueix per tècniques com la difusió d'estat sòlid o la implantació iònica. Per dopar una regió de l'oblia de partida s'aprofita una propietat

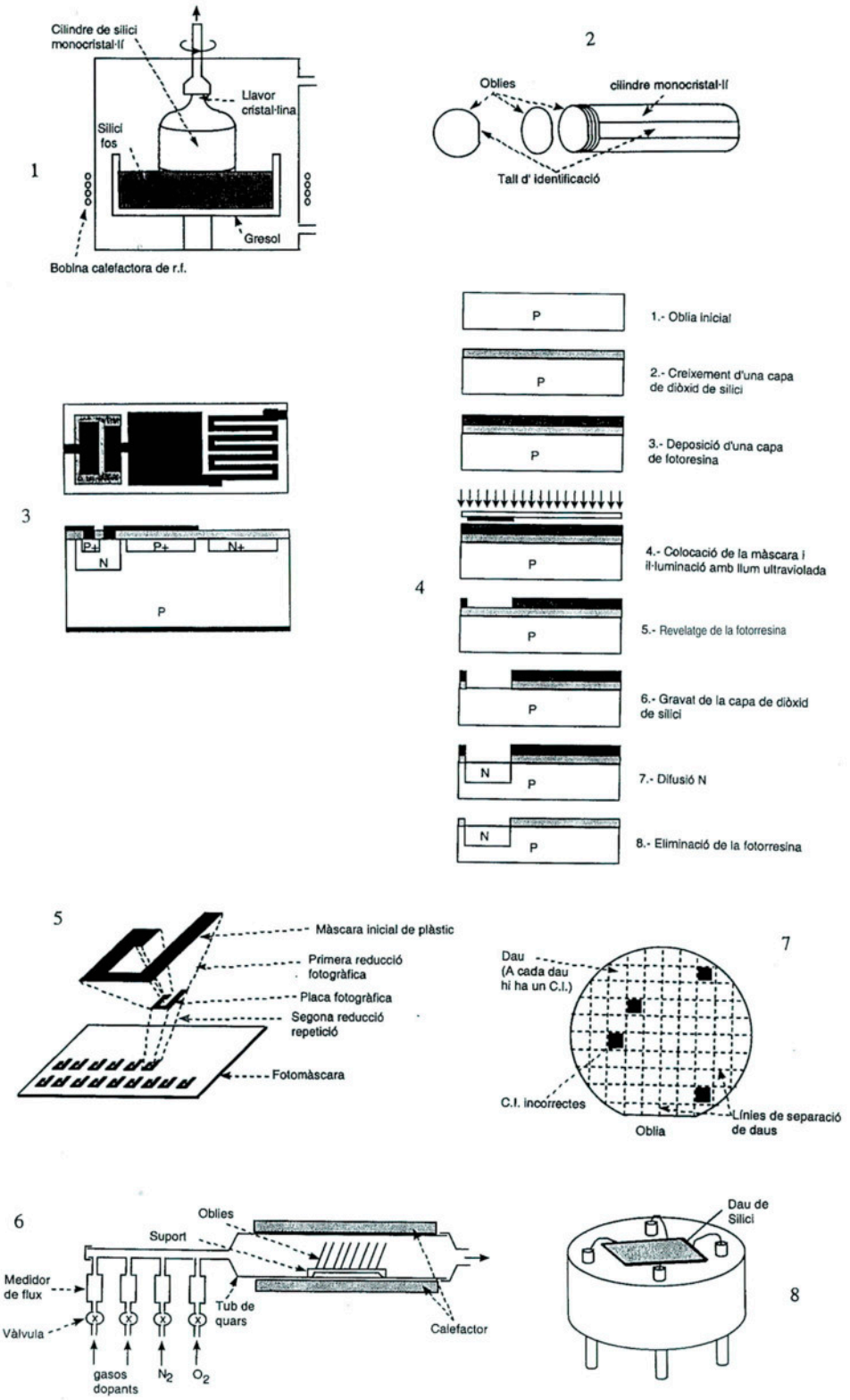

Figura 4: *Tecnologia dels circuits integrats. 1) Obtenció del silici monocristal·lí pel mètode Czokralski. 2) Obtenció de les oblies. 3) Estructura d'un circuit integrat. 4) Processos tecnològics bàsics: oxidació, fotolitografia, gravat, inserció d'impureses. 5) Màscara per fotolitografia. 6) Forn d'oxidació i difusió. 7) Oblia i daus que contenen cadascun un CI 8) Muntatge del dau en un encapsulat*

molt important del diòxid de silici que consisteix a no deixar-se travessar pels àtoms d'impuresa. Per això es comença per cobrir tota l'oblia d'una capa superficial de diòxid de silici i s'obren *finestres* en aquesta capa que deixen el silici de l'oblia al descobert. Les impureses penetren fins a una certa profunditat dins l'oblia de silici per aquestes finestres, i no s'altera el silici protegit per la capa d'òxid.

La creació de finestres en la capa d'òxid es realitza mitjançant la fotolitografia. Es diposita una capa de resina fotosensible sobre la capa d'òxid superficial, i s'aplica una màscara sobre aquesta resina. La màscara té àrees transparents i àrees opaques. En il·luminar l'oblia amb la màscara al damunt amb llum ultraviolada, els fotons d'aquesta radiació arriben a la resina a través de les àrees transparents de la màscara, la qual cosa provoca un canvi en les seves propietats, de manera que es dissol en un líquid de *revelat* i desapareix, deixant al descobert la capa d'òxid. En canvi, la resina no il·luminada es manté intacta. A continuació se submergeix l'oblia en un àcid que ataca el diòxid de silici però no la resina fotosensible. D'aquesta manera s'obren finestres en la capa d'òxid.

Un circuit integrat ocupa una àrea molt petita sobre l'oblia, anomenada *dau*. De fet, la superfície de l'oblia es divideix de forma regular en daus, de manera que es fabriquen simultàniament molts circuits integrats. El nombre de daus que es poden realitzar en una oblia augmenta en relació quadràtica al diàmetre d'aquesta. Si es té en compte que els processos tecnològics de fabricació es fan simultàniament sobre un nombre elevat d'oblies, s'entén que el cost de producció de circuits integrats disminueixi en augmentar el diàmetre de les oblies. Després es talla l'oblia, se separen els daus i es procedeix a l'encapsulació dels circuits que funcionen correctament. El xip que veiem en els equips electrònics conté en el seu interior el petit dau de silici que és el que fa les funcions electròniques.

L'altre element que incideix en la disminució del cost és la capacitat de fer els dispositius cada cop més petits. L'amplada d'una línia, que es pot dibuixar d'una manera segura i repetitiva sobre el silici, ha anat disminuint uniformement i sostingudament des del primer circuit integrat. Això ha permès encabir dins una mateixa àrea un nombre creixent de transistors. El primer microprocessador en un xip, el 4004 d'Intel, tenia 2.300 transistors. Un Pentium actual en té més de 10 milions. Això explica per què des de fa molts anys obtenim pel mateix preu ordinadors cada cop més potents que deixen obsolets els anteriors.

Gordon Moore, cofundador d'Intel amb Noyce, va adonar-se el 1965 que cada 18 a 24 mesos es doblava el nombre de components en el xip a causa de la reducció de l'amplada de línia. Per això va formular la seva famosa llei que preveia el manteniment d'aquesta característica en el futur. Gairebé quaranta anys després d'haver estat formulada se segueix complint. A la figura 5 es mostra com ha anat augmentant el diàmetre de les oblies i el nombre de components en el xip (per a microprocessadors el nombre de components es dobla cada 2 anys i per a memòries dinàmiques RAM cada 1,5 anys).

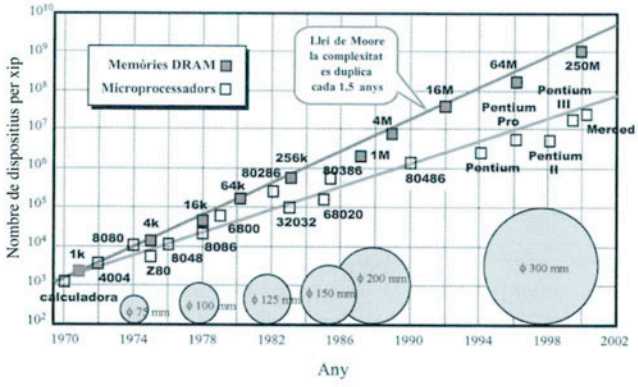

Figura 5: *Evolució del nombre de transistors integrats en un xip (llei de Moore)*

A mesura que els dispositius es fan més petits també es fan més ràpids. La velocitat d'un circuit digital ve limitada fonamentalment pel temps que es triga a carregar i descarregar les capacitats paràsites. Aquestes capacitats són proporcionals a l'àrea, de manera que en reduir-se les dimensions dels dispositius es pot augmentar la freqüència de treball del circuit. A la figura 6 es mostra l'evolució de l'amplada mínima de línia que permet la tecnologia i de la freqüència màxima dels circuits digitals.

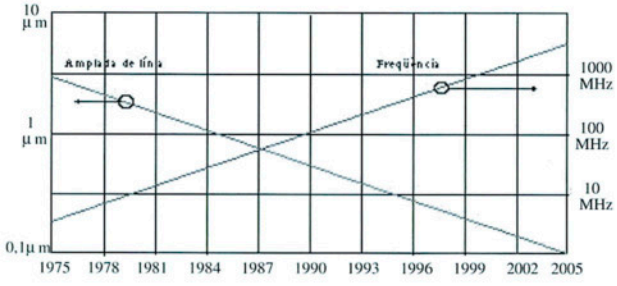

Figura 6: *Evolució de l'amplada mínima de línia que permet la tecnologia i de la freqüència màxima de treball d'un microprocessador*

Les xifres representatives de l'estat actual de la tecnologia de circuits integrats digitals i les previsions de futur[3] es mostren en la taula 1. Malgrat l'espectacularitat d'aquestes xifres, la llei de Moore segueix marcant la

[3]International Technology Roadmap for Semiconductors, 1998 Update. http://public.itrs.net/Files/1998Update/98Update.htm

ANY	1997	1999	2002	2005	2008
Amplada de línia (μm)	0,25	0,18	0,13	0,10	0,07
DRAM: bits/xip	267 M	1,07 G	4,29 G	17,2 G	68,7 G
mP: transistors/xip	11 M	21 M	76 M	200 M	520 M
Freqüència xip (MHz)	750	1.250	2.100	3.500	6.000
Grandària μP: (mm^2)	300	400	560	790	1.120
Núm. nivells de metal·lització	6	6-7	7	7-8	8-9
Diàmetre oblia (cm)	20	30	30	30	45
Tensió alimentació (V)	1,8-2,5	1,5-1,8	1,2-1,5	0,9-1,2	0,6-0,9
Cost transistor μP (10^{-8} \$)	3.000	1.735	580	255	110

Taula 1: *Algunes característiques específiques de l'estat actual de la tecnologia de circuits integrats i previsions de futur. (International Technology Roadmap for Semiconductors, 1998 Update)*

pauta. Intel acaba d'anunciar el passat 7 de novembre que disposa, pel que fa al laboratori, de la tecnologia de 0,13 mm, amb la qual preveu fabricar els xips previstos a la taula per a l'any 2002. D'aquí a cinc anys les CPU dels nostres ordinadors tindran 200 milions de transistors i treballaran a una freqüència de 3,5 GHz. El cost d'un transistor s'haurà dividit per quatre. Els transistors MOS d'aquests xips tindran un òxid de porta d'una amplada d'uns 70 namòmetres i un gruix d'una dotzena d'àngstroms.

La llei de Moore, però, té un final anunciat. No és possible anar doblant la densitat de components indefinidament cada any i mig (Birnbaum *et al*, 2000). Hi ha dues raons que ho impedeixen. Una de caràcter físic i una altra de caràcter econòmic. La primera és que el transistor MOS opera com un interruptor. Actualment la càrrega que s'ha de dipositar a la porta per fer-lo commutar és d'uns mil electrons. En cas de seguir el ritme de disminució marcat per la llei de Moore, aquesta càrrega seria, el 2020, d'una fracció d'electró, cosa que és físicament impossible.

L'altra limitació, de caràcter econòmic, és el cost del muntatge d'una fàbrica de circuits integrats. Aquest cost segueix fidelment la llei de Moore: es dobla cada tres anys, a causa de la modificació d'equips que exigeix la reducció de dimensions de l'amplada de línia i l'augment del diàmetre de les oblies. Per això se la coneix com la segona llei de Moore (figura 7). Aquest augment de cost creix més ràpidament que el mercat de circuits integrats. Així, d'aquí a pocs anys construir una fàbrica de xips costarà més que el volum de tot el mercat electrònic mundial. Per tant, cal esperar que l'actual ritme de creixement exponencial es moderi i es tendeixi cap a una saturació.

Herbert Kroemer i Zhores Alferov: les heteroestructures

La segona meitat del Premi Nobel és compartida entre aquests dos científics per la seva aportació al desenvolupament dels dispositius basats en heteroestructures

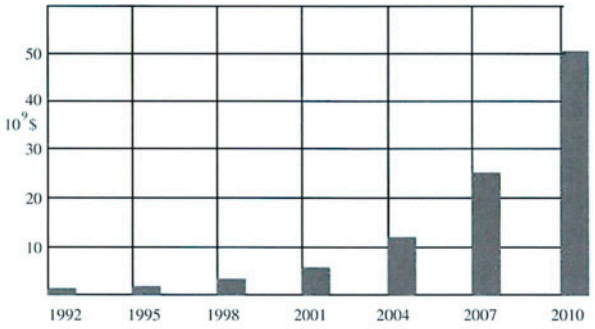

Figura 7: *Evolució del cost d'instal·lació d'una planta de fabricació de circuits integrats*

que ha permès crear l'electrònica d'alta velocitat i la fotònica. Com en el cas de Kilby i Noyce, les seves propostes es presenten simultàniament i de manera independent.

El silici no és l'únic semiconductor emprat en l'electrònica. Abans va ser-ho el germani, i més tard l'arseniür de gal·li (GaAs) i molts altres. S'anomena *heteroestructura* el cristall realitzat amb diferents semiconductors. Per fer dispositius basats en heteroestructures és important que no es trenqui l'ordenació cristal·lina global en passar d'un material a un altre. Aquest requisit va exigir disposar d'unes tecnologies adequades per fer possible unes estructures que van ser proposades molts anys abans que es poguessin realitzar. Shockley les havia suggerit en una patent de 1948, H. Kroemer va presentar la teoria dels transistors bipolars d'heterojunció l'any 1957, i va suggerir aquestes estructures per fer làsers el 1963. Al mateix temps i de manera independent, Alferov patentava aquesta estructura a l'URSS. Malgrat això, els primers làsers d'heteroestructura no van arribar fins el 1970 fent servir una tecnologia LPE (epitàxia de fase líquida), i la seva irrupció massiva no arribà fins els anys vuitanta després que es van posar a punt les tecnologies MBE (*molecular beam epitaxy*) i MOCVD

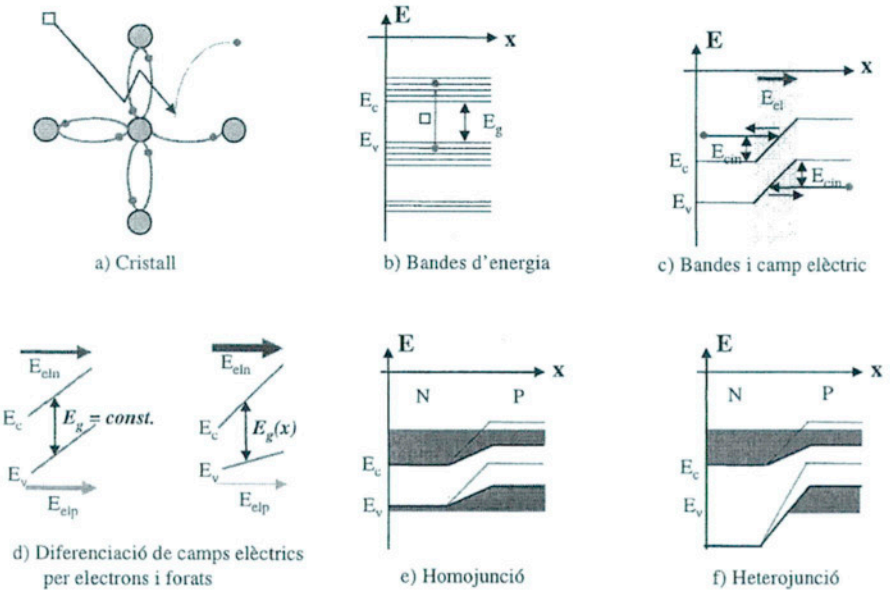

Figura 8: *Bandes d'energia d'un semiconductor. a) Estructura cristal·lina que mostra els enllaços covalents. b) Bandes d'energia que mostren la transició d'un electró quan trenca l'enllaç. c) Doblegament de les bandes per efecte d'un camp elèctric i confinament dels portadors. d) Diferenciació de camps elèctrics per electrons i forats en les heteroestructures. e) Diagrama de bandes en un díode d'homojunció i (f) d'heterojunció*

(*metal-organic chemical vapor deposition*), que van permetre fer créixer capes cristal·lines sobre un substrat, controlant el creixement de cada capa d'àtoms.

Per entendre les propietats de les heteroestructures cal tenir una certa noció del diagrama de bandes d'energia d'un semiconductor. Aquest diagrama (figura 8) descriu l'energia total que tenen o poden tenir els electrons en cada punt del semiconductor. De manera similar al que succeeix en l'àtom, els electrons en el semiconductor només poden tenir unes energies, uns *nivells d'energia*, determinats. Hi ha uns nivells d'energia especialment importants en el comportament elèctric del semiconductor: els que tenen les càrregues mòbils que *porten* el corrent. És a dir, els electrons lliures, que són aquells que han trencat la seva pertinença a un enllaç covalent, i els electrons de València, quan hi ha enllaços covalents trencats. Doncs bé, les energies que poden tenir els electrons lliures estan agrupades en un interval anomenat *banda de conducció* en el qual hi ha una gran densitat de nivells permesos, cosa que dóna la impressió de continuïtat de nivells. El mínim d'energia d'aquesta banda s'anomena E_c. Les energies que poden tenir els electrons que formen part dels enllaços covalents s'agrupen en la *banda de València*, i E_v és el seu valor màxim. L'interval entre E_c i E_v s'anomena *banda prohibida*, o *gap* en anglès, i és un dels paràmetres més significatius del semiconductor. Representa el mínim d'energia que cal donar a un electró d'un enllaç covalent per deslligar-lo. En el silici val 1,12 eV i en el GaAs 1,43 eV.

El diagrama de bandes d'energia permet descriure el moviment dels portadors a l'interior del cristall i la seva interacció amb els camps elèctrics presents. Un electró lliure d'energia E_e té una energia cinètica de valor $E_e - E_c$. Quan aquest electró es desplaça d'esquerra a dreta, tal com mostra la figura 8c, i entra en una regió en la qual hi ha present un camp elèctric de sentit esquerra a dreta, l'electró és frenat pel camp i perd velocitat, i, per tant, energia cinètica. Aquesta situació es representa en el diagrama de bandes fent que el pendent de la corba E_c sigui proporcional al camp elèctric present en aquest punt. Quan l'electró *toca* el nivell E_c, la seva energia cinètica és zero i, per tant, la seva velocitat s'ha anul·lat per l'acció de frenada del camp elèctric. A partir d'aquest moment el camp elèctric empeny l'electró cap a l'esquerra. Es diu que la barrera d'energia confina l'electró en la regió on E_e és major que E_c.

Quan hi ha un enllaç covalent trencat, un electró de València d'un enllaç proper salta cap a l'enllaç trencat per refer-lo, i deixa darrere seu, com a resultat d'aquest salt, un enllaç covalent trencat. Tot succeeix com si qui es mogués fos l'enllaç covalent trencat, que equival a una càrrega positiva que s'anomena *forat*. L'energia d'un forat és la del nivell d'energia buit que queda a la banda de València quan un electró trenca l'enllaç i salta a la banda de conducció. Un forat d'energia E_h té una energia cinètica $E_v - E_h$. Quan el forat es mou dintre el

semiconductor i *toca* el nivell E_v, s'anul·la la seva energia cinètica. Es diu que el forat està confinat a la regió on E_v és més gran que E_h.

El pendent de la corba E_c representa, per tant, el camp elèctric efectiu que actua sobre els electrons, mentre que el pendent de E_v representa el camp elèctric efectiu que actua sobre els forats. Una heterojunció permet que el camp elèctric efectiu que actua sobre els electrons sigui diferent del que actua sobre els forats. Aquest nou grau de llibertat que permet l'heterojunció s'acostuma a anomenar *enginyeria del band gap*.

A la figura 8 també es representen les bandes d'energia d'un díode d'homojunció i un altre, d'heterojunció. En el segon cas, el semiconductor de l'esquerra té un gap més gran que el de la dreta. A conseqüència d'aquesta diferència, el doblegament de E_c és menor que el de E_v, i per tant el camp elèctric que actua sobre els electrons és menor que el que actua sobre els forats. Quan la tensió de polarització fa baixar la barrera d'energia, circulen electrons de N a P i forats de P a N en l'homojunció, mentre que en l'heterojunció els forats queden confinats a la regió P, perquè que la barrera per a aquests portadors és molt més alta que la que han de superar els electrons per anar de N a P. Noteu que quan la polarització canvia de signe no circula corrent, ja que ni els electrons ni els forats poden superar les barreres d'energia.

Herbert Kroemer (1928, Weimar (Alemanya)) va defensar la seva tesi doctoral el 1952 a la Universitat alemanya de Göttingen sobre efectes d'electrons calents (*hotelectrons*) en el transistor bipolar que s'acabava de descobrir. Després de treballar en diversos laboratoris de recerca alemanys i americans, es dedica des de 1976 a la tecnologia de semiconductors compostos a la Universitat de Califòrnia a Santa Clara. El 1957 va publicar un article sobre la teoria de funcionament dels transistors d'emissor de gap ample, i el 1963 va proposar el concepte de làser de doble heteroestructura sense el qual no hagués estat possible el desenvolupament de la fotònica. La seva investigació actual gira entorn de la creació d'estructures híbrides, superconductors-semiconductors basats en pous quàntics de InAs-AlAs, i en la recerca d'oscil·ladors de terahertz basats en superxarxes.

Zhores I. Alferov (1930, Vitebsk (Bielorússia) va graduar-se en electrònica a l'Institut Electrotècnic de Leningrad el 1952. L'any següent va entrar a l'Institut Físico-Tècnic "A. F. Ioffe" de Sant Petersburg on ha desenvolupat la seva tasca científica fins ara. Ha anat accedint de manera regular a nivells superiors de coneixement i responsabilitat, fins a doctorar-se en Física i Matemàtiques l'any 1970. Des de 1962 treballa a l'àrea d'heteroestructures basades en semiconductors compostos, ha destacat en aportacions tan importants com la patent a l'URSS de l'estructura làser de doble heterojunció el 1963, al mateix temps que Kroemer, i el desenvolupament el 1970 del primer làser de sortida contínua a temperatura ambient basat en aquesta estructura i realitzat amb tecnologia LPE, al mateix temps que Panish en els laboratoris Bell (EUA). Actualment treballa en el desenvolupament de làsers basats en punts quàntics (*quantum dots* en anglès).

Heteroestructures i transistors d'alta velocitat

A la figura 9 es presenta l'espectre de freqüències per a comunicacions d'alta velocitat previstes a Europa l'any 2008. Aquestes aplicacions ja comencen a tenir un gran impacte en la nostra vida quotidiana: la televisió d'alta definició, la distribució de vídeo a la carta, les xarxes d'ordinadors connectades via ràdio, la telemàtica del transport per carretera, radars anticol·lisió en els automòbils, etc. Com es veu en la figura, les freqüències a utilitzar superen les desenes de GHz i exigeixen dispositius i circuits d'alta velocitat.

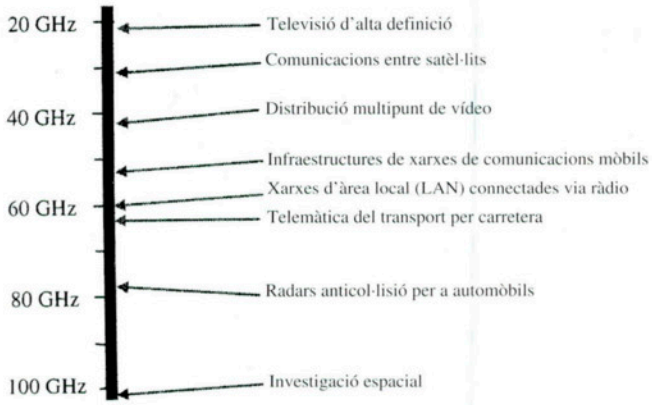

Figura 9: *Espectre de freqüències per a aplicacions de comunicacions d'alta velocitat previstes a Europa l'any 2008*

Els dos dispositius electrònics que encapçalen l'electrònica d'alta velocitat es basen en heteroestructures. Un d'ells és l'HBT (*heterojunction bipolar transistor*). L'altre és l'HEMT (*high electron mobility transistor*), també conegut com MODFET (*modulation doped field effect transistor*). Per fer aquests dispositius s'utilitzen normalment semiconductors formats per la combinació d'elements dels grups III i V de la taula periòdica. Els que han tingut un desenvolupament més important són els basats en el sistema AlGaAs/GaAs, però comencen a tenir un protagonisme creixent i una potencialitat de futur més gran els basats en InGaAs/InP. Cal assenyalar, però, que també es fan transistors d'heteroestructura partint del sistema Si/SiGe, que intenta aprofitar la maduresa de la tecnologia del silici i les potencialitats de les heterojuncions. A la figura 10 es representen els valors del gap de diversos semiconductors i la seva constant de xarxa cristal·lina.

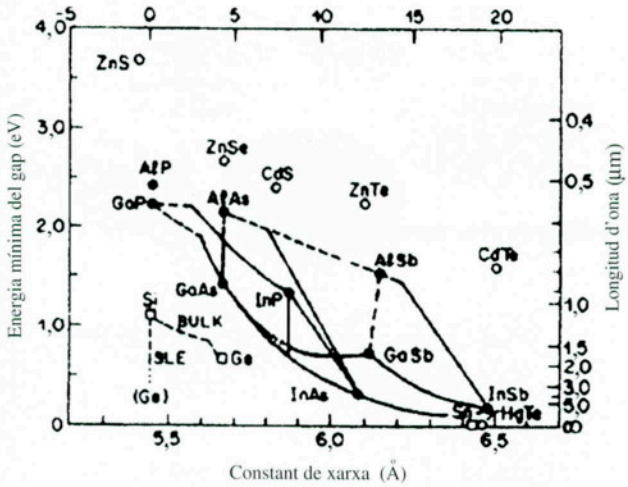

Figura 10: *Semiconductors per a heteroestructures: han de tenir la mateixa constant de xarxa cristal·lina*

Encara que Shockley ja havia suggerit l'HBT en una patent de 1948, el pare efectiu d'aquest dispositiu va ser Kroemer (Kroemer, 1957), que va presentar la seva teoria d'operació el 1957 i un estudi de les seves potencialitats el 1982 (Kroemer, 1982), just quan s'havien desenvolupat les tecnologies que el feien possible. La figura 11 mostra l'estructura i el diagrama de bandes d'aquest dispositiu. Com es pot veure, el gap de l'emissor és més gran que el de la base i el del col·lector. L'avantatge fonamental d'aquest dispositiu és que la barrera que han de superar els forats per anar de la base a l'emissor és molt més gran que la dels electrons per anar de l'emissor a la base. Això permet no dependre de la relació de dopatges entre emissor i base, per aconseguir que el corrent d'electrons d'emissor a base sigui molt més gran que el corrent de forats de base a emissor, que és la relació que determina el guany de corrent del transistor. A diferència dels transistors fets d'un sol semiconductor, els dopatges d'emissor i base d'un HBT es dissenyen per minimitzar les resistències i capacitats que determinen la màxima freqüència del senyal que pot amplificar el transistor.

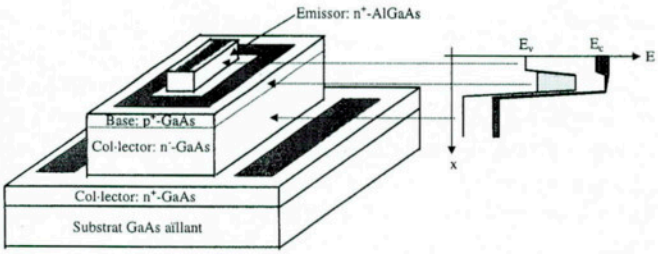

Figura 11: *Estructura i diagrama de bandes d'energia d'un HBT. Noteu que el gap de l'emissor és més gran que el de la base i el del col·lector*

Per la seva banda l'HEMT és un transistor d'efecte de camp, similar al MOS, però que aprofita una propietat que presenta la seva heteroestructura: el transvassament d'electrons d'una regió molt dopada i de gap ample, a una regió poc dopada i de gap petit. A la figura 12 es presenta l'estructura i el diagrama de bandes d'aquest transistor. El corrent que flueix entre el drenador i el surtidor circula a través d'un canal anomenat 2DEG, i així es pot controlar la quantitat d'electrons d'aquest canal a través de la tensió aplicada a la porta G. En aquesta heterojunció es forma una escletxa molt estreta a la part poc dopada de GaAs. Aquesta escletxa s'omple d'electrons procedents de la capa molt dopada de AlGaAs, i, com que és molt estreta, presenta fenòmens de quantificació d'energia típics d'un pou quàntic (el nom 2DEG fa referència a aquest fet: *2-dimensional electron gas*). Els electrons d'aquest pou tenen una alta mobilitat.

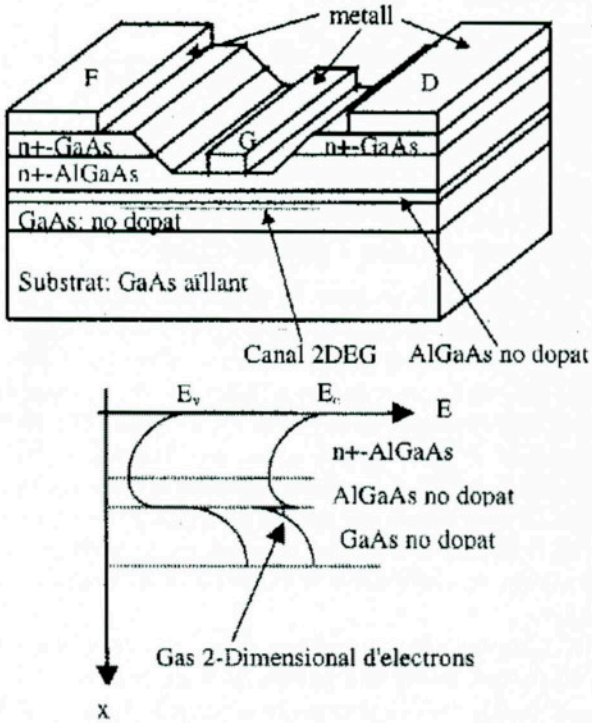

Figura 12: *Estructura i diagrama de bandes d'energia d'un HEMT. El canal conductor està constituït pels electrons atrapats a l'escletxa del nivell E_c en el material de gap menor*

Hi ha una pugna entre els HBT i els HEMT per encapçalar l'electrònica d'alta velocitat. A la figura 13 es presenta l'evolució de la freqüència màxima que poden amplificar aquests transistors. En una primera fase la cursa era dominada pels HEMT, però el resultat final és cada cop més incert. Actualment, tant uns com altres poden amplificar freqüències superiors als 500 GHz. Sembla, però, que els HBT de InGaAs tenen més potencialitat de futur que els altres, tant pel que fa referència

a la freqüència màxima que poden amplificar, com per la potència que poden controlar, i són, per tant, els dispositius del futur per a les comunicacions sense fils i les telecomunicacions espacials.

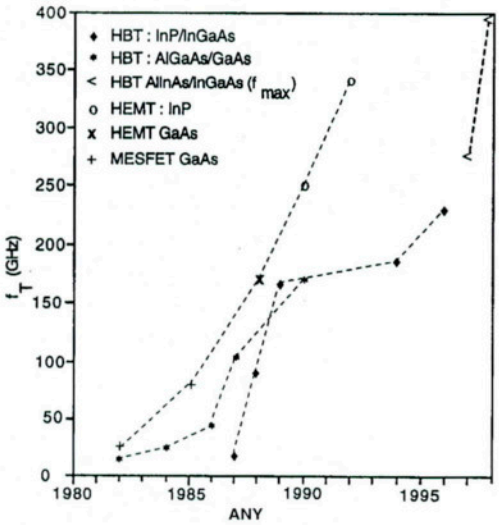

Figura 13: *Evolució de la màxima freqüència d'amplificació en els últims anys*

Heteroestructures i díodes làser

El làser (*light amplification by stimulated emission of radiation*) es basa en l'emissió estimulada de radiació electromagnètica descoberta en l'àmbit teòric per Einstein el 1917. Calgué esperar fins el 1960 per veure la confirmació experimental de la teoria de l'emissió estimulada mitjançant el làser de robí creat per Maiman, i fins el 1970 per tenir làsers de semiconductor que donessin un feix de llum continu i a temperatura ambient. La llum del làser és monocromàtica, direccional i coherent, propietats que la diferencien d'altres fonts de llum i la fan única.

A la figura 14 es presenten alguns camps d'activitat on els làsers tenen una presència significativa. La cirurgia ocular en el camp de la medicina, la lectura del codi de barres en el comerç, els discos compactes en la música i en la informàtica, la fotònica en les comunicacions òptiques, les impressores i l'holografia en les arts gràfiques, els sensors de fibres òptiques en el camp de l'electrònica, així com la seva utilització en els camps de la meteorologia i de la tecnologia dels materials, són algunes de les aplicacions dels làsers més recents i conegudes per tothom.

La llum que emeten alguns semiconductors neix de la recombinació de portadors. Un electró ha d'absorbir energia, un *quantum*, per trencar el seu lligam d'un enllaç covalent. Això li permet saltar de la banda de València a la de conducció. Doncs bé, quan un electró lliure passa a prop d'un enllaç covalent trencat pot ser

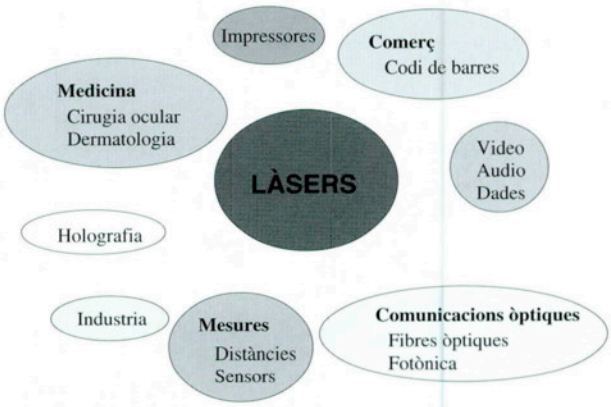

Figura 14: *Alguns camps d'aplicació dels làsers*

capturat i reconstruir aquest enllaç. Per fer-ho cal, però, que es desprengui de l'excés d'energia que va adquirir en alliberar-se. Si el semiconductor té una estructura de bandes anomenada de *gap directe*, l'energia d'excés la desprèn en forma de fotó; en cas contrari en forma de calor. El GaAs és un material de gap directe, mentre que el silici és de gap indirecte. Per això no es fan làsers de silici.

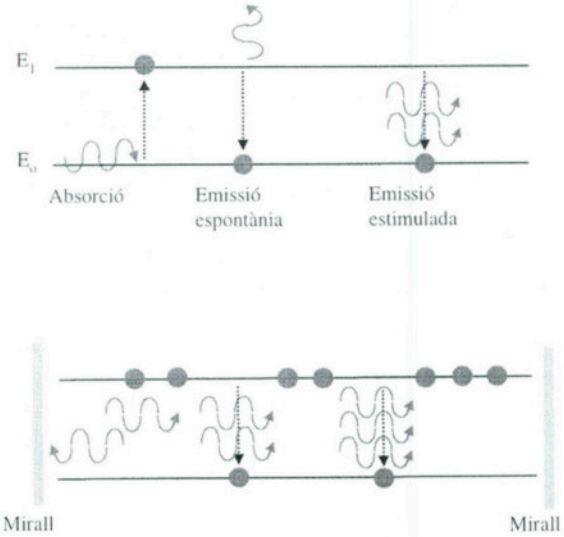

Figura 15: *Emissió estimulada i cavitat òptica ressonant en un làser*

L'emissió estimulada consisteix en el fet que un fotó d'energia adequada pot *estimular* la recombinació d'un electró. El fotó estimulador ha de tenir la mateixa energia que la del fotó emès a la recombinació. Aleshores el fotó estimulador i el fotó lliurat en la recombinació viatgen en fase, en coherència. A la figura 15 es representa aquest concepte. Suposem que E_0 i E_1 són dos nivells d'energia que pot tenir un electró. Evidentment, els fotons emesos en passar l'electró de E_1 a E_0 poden ser reabsorbits per electrons que estiguin en el ni-

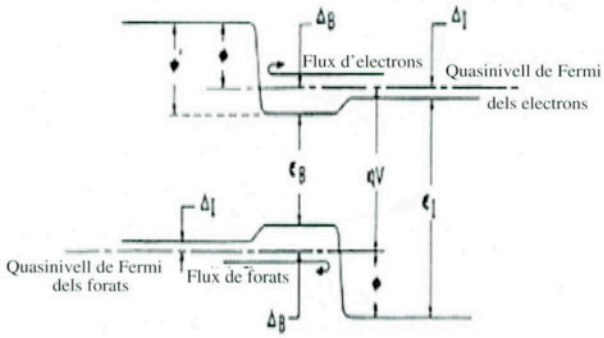

Figura 16: *Diagrama de bandes d'un làser de doble heterojunció proposat per Kroemer el 1963*

vell E_0. També pot donar-se la recombinació *espontània* d'un electró des de E_1 fins a E_0. Per tal que l'emissió estimulada domini sobre l'absorció i l'emissió espontània, s'han de complir dues condicions: que hi hagi *inversió* de població, la qual cosa vol dir que hi hagi més electrons en el nivell E_1 que en el E_0, i que hi hagi una cavitat òptica ressonant que faci créixer el feix d'electrons coherents, a base d'anar-los reflectint entre dos miralls perpendiculars a la regió activa del làser. A cada passada, a través de la capa activa es van afegint al feix coherent nous fotons generats per emissió estimulada.

s'aconseguia produir polsos de llum de durada molt curta. El problema fou superat gràcies a la utilització d'heteroestructures. A la figura 16 es mostra l'esquema proposat per Kroemer el 1963 (Kroemer, 1963). Com es pot veure, consisteix en una regió central de gap estret situada enmig de dues regions de gap ample. La regió de la dreta té un dopatge de tipus P i injecta forats a la regió central. La barrera de potencial que presenta el semiconductor de la dreta els confina a l'esquerra de la barrera. Quelcom similar succeeix als electrons que injecta el semiconductor de tipus N de la dreta cap a la regió central. A conseqüència d'aquests fenòmens la regió del mig de gap estret, té una gran riquesa simultània d'electrons i de forats, i d'aquesta manera s'aconsegueix la inversió de població sense necessitat de fer circular uns corrents enormes. La cavitat òptica ressonant acostuma a construir-se polint les parets del semiconductor perpendiculars a la capa activa. El feix de llum surt pel fet que una d'aquestes parets transmet parcialment la llum (figura 17).

A la figura 18 es mostra com ha anat evolucionant al llarg del temps el corrent mínim que cal fer circular en el díode per aconseguir un feix làser (Alferov, 1998). Com es pot veure, aquest corrent s'ha reduït en més de tres ordres de magnitud. També s'indiquen en aquesta figura les línies de progrés que han seguit els làsers en els anys posteriors per aconseguir aquesta disminució: la utilització de pous quàntics (*quantum wells*) i la de

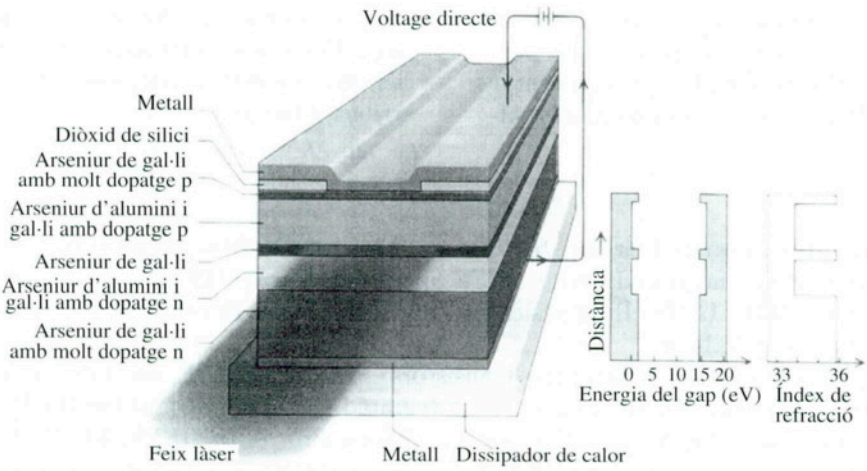

Figura 17: *Estructura d'un díode làser de doble heterojunció*

Aconseguir la inversió de població en un díode làser fou un repte difícil de superar. La inversió de població en un semiconductor consisteix a tenir en un mateix punt molts electrons lliures i molts forats. En els primers làsers això s'aconseguia molt difícilment fent circular uns corrents molt intensos en un díode construït amb regions molt dopades. Calia refrigerar el díode a la temperatura del nitrogen líquid i només

superxarxes de curt període SPS (*short-period superlattices*). A la taula 2 es mostren les característiques d'un làser modern basat en aquestes estructures. Crida l'atenció el seu corrent llindar de només 1 mA, la seva eficiència de potència del 50 %, així com el seu baix cost.

Cal subratllar com a fet important, la recent obtenció de làsers blaus (1996) per S. Nakamura fent servir GaN, la qual cosa ha permès completar tot l'espectre

visible. Aquests làsers també contribueixen a accelerar el progrés en el món de la computació, ja que tenen una longitud d'ona menor que permet multiplicar la capacitat d'emmagatzemament dels discos compactes.

Estructura	Pous quàntics (QW)
Nombre de capes	200
Gruix de cada capa	60 - 900 Å
Cavitat òptica ressonant	Perpendicular a la capa activa
Grandària	10 μm × 10 μm
Corrent llindar	1 mA
Potència làser de sortida	10 mW
Eficiència de potència	50 %
Amplada de banda de modulació	20 GHz

Taula 2: *Característiques d'un làser modern*

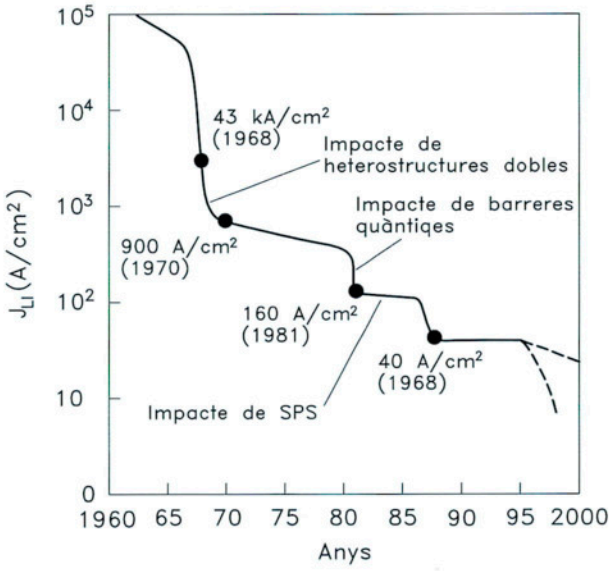

Figura 18: *Evolució del corrent llindar dels díodes làser i causes de la seva reducció*

A mode de cloenda

Com s'ha dit al llarg d'aquestes pàgines, les tecnologies de la informació i les comunicacions estan influint poderosament en la forma de viure de la humanitat actual. Afecten tant els nostres comportaments més personals com les estructures més col·lectives. Tant és així que acceleren el pas de la humanitat cap a una convergència planetària, esborrant fronteres i barreres de comunicació. Sembla per tant raonable que s'honori amb un Premi Nobel aquesta part de la ciència.

Les tecnologies de la informació es fonamenten en els circuits digitals i en els circuits de comunicacions. El desenvolupament d'aquests circuits no hauria estat possible sense les aportacions de Jack Kilby, que va inventar el circuit integrat, i de Herbert Kroemer i de Zhores Alferov, que van idear les heteroestructures que han fet possible l'aparició dels làsers i dels transistors d'alta velocitat. Ocupen, per tant, un lloc rellevant en aquesta ciència, la qual cosa els fa mereixedors de la distinció rebuda.

Potser les aportacions d'aquests científics no tenen la brillantor teòrica d'altres premis Nobel de física, però les conseqüències socials de les seves contribucions superen, sense cap mena de dubte, les de molts altres guardonats. Penso que el reconeixement de la tasca d'aquests científics ha arribat una mica tard. R. Noyce no hi ha arribat a temps.

Bibliografia

KILBY, J. S., Invention of the integrated circuit, *IEEE trans. Electron Devices*, **23**, 648 (1976).

WOLF, M. F., The genesis of the integrated circuit, *IEEE Spectrum*, August (1976).

BIRNBAUM, J., STANLEY WILLIAMS, R., Physics and the Information Revolution, *Physics Today*, **53**, (2000). http://physicstoday.org/vol-53/p38.html.

KROEMER, H., Theory of a Wide-Gap Emitter for Transistors, *Proc. of the IRE*, **45**, 1535 (1957).

KROEMER, H., Heterostructure Bipolar Transistors and Integrated Circuits, *Proc. of the IEEE*, **70**, January (1982).

KROEMER, H., A proposed class of heterojunction injection lasers, *Proc of the IEEE*, **51**, 1782 (1963).

ALFEROV, ZH. I., The history and future of semiconductor and heterostructures, *Semiconductors*, **32**, January (1998).