

Les nanotransistors : être ou ne pas être en CMOS sur silicium ?

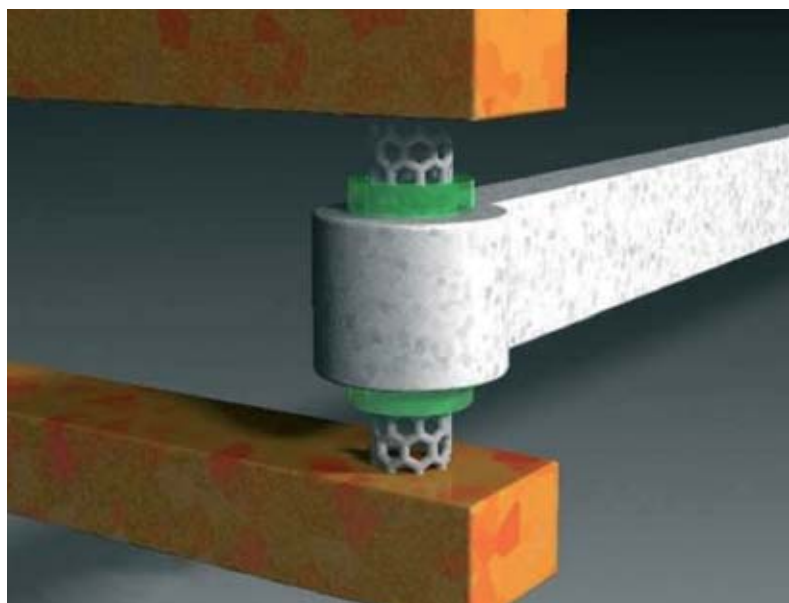
À l'échelle du nanotransistor, les électroniciens doivent aller encore plus loin dans l'optimisation des caractéristiques des transistors afin de minimiser les effets parasites. À moins d'en tirer astucieusement parti !

Les caractéristiques électriques des nanotransistors CMOS sur silicium dépendent principalement de leur contrôle dimensionnel horizontal et vertical, lié à la fabrication, à l'architecture du composant ou aux matériaux utilisés. L'optimisation dépendra étroitement des effets parasites à l'accès à ce dispositif ou de l'isolation entre composants.

De façon récurrente, l'optimisation des dispositifs CMOS repose sur le contrôle de la charge dans le canal du transistor via les électrodes de drain, de source, de grille et du substrat sur lequel est fabriqué le dispositif (figure 1 et figure de l'encadré D, *Le transistor, composant de base des circuits intégrés*, p. 28) afin d'éviter un courant de fuite excessif.

Pour les dispositifs de longueur de grille inférieure à 5 nm (qui correspond au nœud 65 nm dans la feuille de route de l'industrie de la microélectronique, voir figure de l'encadré *La loi de Moore*, p. 32), la diminution des dimensions verticales s'accompagne de l'augmentation des fuites à travers l'isolant de grille (SiO_2 de l'ordre de 1 nm) ainsi que l'augmentation de la résistance série en amont du canal (source et drain de profondeur 10 nm). De nombreuses recherches sont en cours en vue d'introduire des matériaux à constante diélectrique plus forte que celle de SiO_2 (exemple : HfO_2 , La_2O_3 , HfAlO_x , etc.) permettant d'épaissir l'isolant de grille tout en gardant une valeur de capacité de couplage compatible avec une commande rapide du canal.

L'augmentation de résistance des source et drain vient de la présence d'un faible nombre d'atomes de dopants dans ces zones et de leur variabilité statistique, des défauts d'activation électrique (faible volume de silicium et présence de défauts) et de la contrainte



Un concept de nanotransistor à canal tubulaire, le canal étant ici constitué par un nanotube de carbone.

mécanique exercée par les couches enrobant le dispositif. La technique de recuit laser (excimère, par exemple) permet d'augmenter l'activation des dopants au-delà de leur solubilité limite classique mais nécessite encore beaucoup d'efforts pour son intégration dans un procédé de fabrication.

Il est possible de profiter de l'état de contrainte intentionnelle de ces matériaux (figure 1) pour favoriser le transport des électrons et des trous dans des bandes à masse effective faible. La maîtrise de la contrainte dans les petites dimensions de façon sélective est un sujet de recherche majeur.

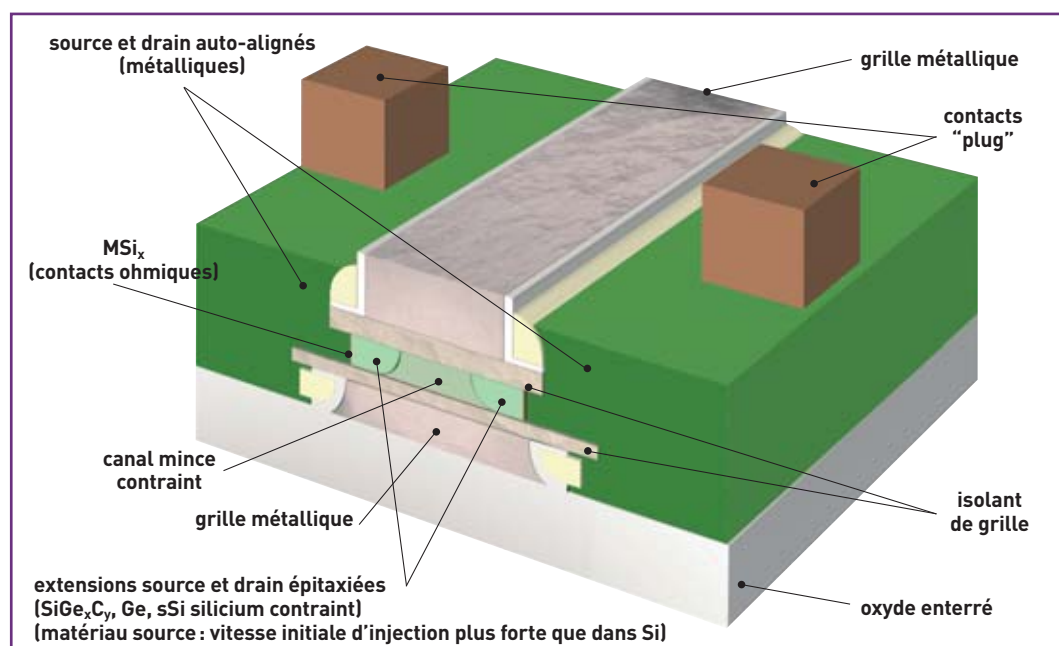


Figure 1. Architecture de transistor MOS multigrille et à source/drain métalliques, comportant des extensions permettant d'augmenter la vitesse d'injection des porteurs dans le canal (brevets CEA-Leti).

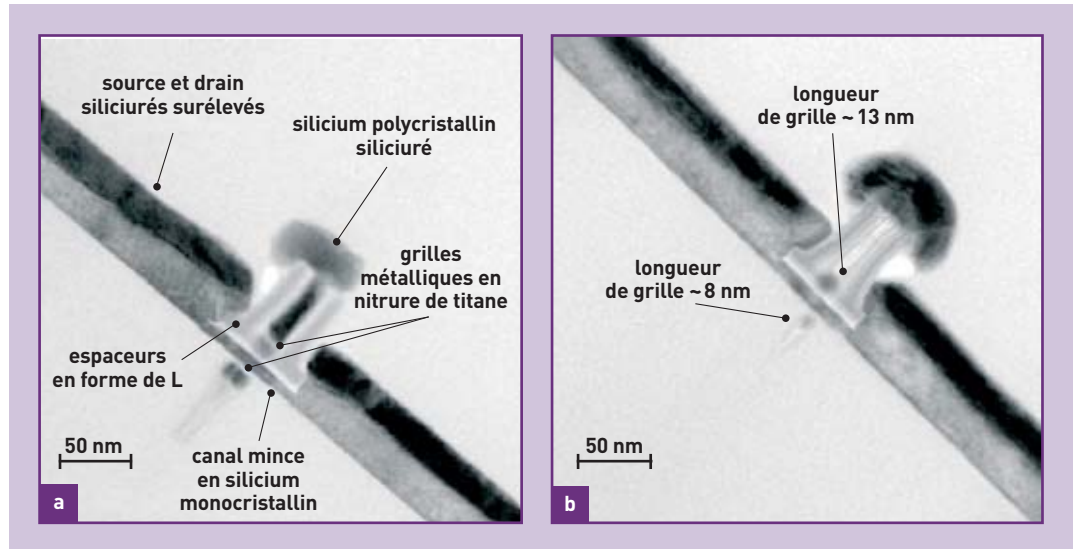


Figure 2. Coupes au microscope électronique à transmission de transistors à double grille de longueur 20 nm (a) et autour de 10 nm (b), réalisés au Leti afin de mieux contrôler les courants de fuite entre la source et le drain. Réf. : *Bonded planar Double metal Gate NMOS transistors down to 10 nm*, M. VINET, T. POIROUX, J. WIDIEZ, J. LOLIVIER, B. PREVITALI, C. VIZIOZ, B. GUILLAUMOT, Y. LE TIEC, P. BESSON, B. BIASSE, F. ALLAIN, M. CASSÉ, D. LAFOND, J.-M. HARTMANN, Y. MORAND, J. CHIARONI and S. DELEONIBUS, *IEEE Electron Devices Letters*, mai 2005.

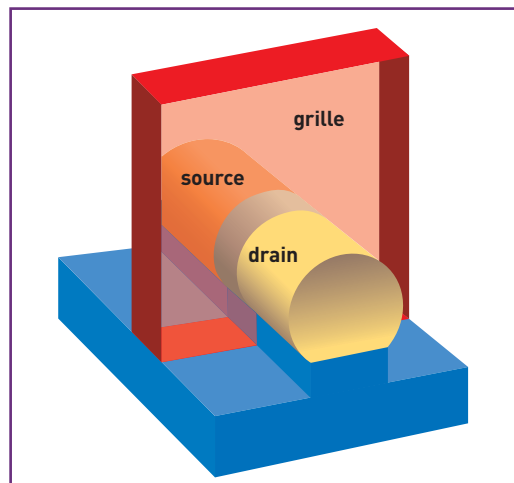


Figure 3. Transistor tubulaire. Le contrôle électrostatique du canal par la grille est total. Les exigences sur les caractéristiques géométriques sont moins contraignantes que dans le cas d'un transistor classique. Représenté ici dans sa configuration horizontale, il peut aussi être vertical.

Le nanotransistor est un dispositif dans lequel le transport des charges se ferait de façon **ballistique** : le libre parcours moyen des électrons (de l'ordre de 100 nm) est supérieur ou égal à la longueur du canal. La technologie impose toutefois ses limites : dans les dispositifs d'une longueur de 25 nm, moins de 50 % des électrons sont ballistiques, les autres étant diffusés. Les limitations viennent de la présence d'impuretés distribuées de façon aléatoire dans le canal ou le drain, du ralentissement des **porteurs de charge** à la source et de la rugosité à l'interface isolant de grille/canal. Dans des canaux de longueur inférieure à 10 nm, les porteurs peuvent passer de la source vers le drain par **effet tunnel** direct, augmentant ainsi la fuite à l'état bloqué du dispositif. La situation sera mieux gérée si le film constituant la zone du canal a une épaisseur inférieure à l'extension des source et drain (jusqu'à la moitié). C'est le cas des dispositifs réalisés sur silicium sur isolant à désertion forte (*Fully Depleted-Silicon On Insulator*, FD-SOI). Un parfait contrôle **électrostatique** du canal nécessite de l'enrober par une ou plusieurs électrodes de grille. Des efforts

de recherche importants sont menés par l'ensemble des acteurs de la nanoélectronique sur les dispositifs à "double grille" (figures 1, 2 et 3), FinFET (transistors "à ailerons"), *Gate All Around* (canal totalement enrobé par la grille).

Le nanotransistor idéal sera donc le transistor à canal tubulaire (figure 3) présentant quatre caractéristiques, la première étant que le canal non dopé sera contraint de façon sélective et différenciée pour les électrons et pour les trous. Ce canal sera constitué de ce que la nature ou les ingénieurs offriront : molécules ou **nanotubes** (carbone, **semi-conducteurs silicium**, germanium, arséniure de gallium, etc.). Par ailleurs, la grille sera métallique, minimisant ainsi le couplage capacitif au canal et la résistance de la grille pour les applications haute fréquence. Troisième caractéristique : des zones de contacts à encombrement minimal et peu résistantes (figure 1). Enfin, le matériau de source permettra de maximiser la vitesse des porteurs à l'entrée du canal contraint et favorisera le transport balistique.

La polarisation indépendante des "multigrilles" (figure 1) donne accès à une véritable plate-forme pour les applications de systèmes embarqués (*Systems On Chip*, SOC). C'est en ce sens que le Laboratoire d'électronique et de technologie de l'information (Leti) du CEA travaille aujourd'hui dans le cadre de ses différents projets et programmes.

Les nanoélectroniciens sont devant le célèbre "mur en briques rouges" réputé *terra incognita* : à l'instar des adeptes de l'escalade, ils doivent saisir toute bonne prise qui permet de l'escalader et le franchir sans dommage. À moins que l'un d'entre eux ne réinvente un nouveau dispositif permettant de franchir le mur par effet tunnel conjugué aux effets balistiques, ce qui serait une bonne utilisation des effets dits "parasites" !

➤ **Simon Deleonibus**

Direction de la recherche technologique
CEA-Leti, centre de Grenoble

E La lithographie, clé de la miniaturisation

La **lithographie** optique (ou photolithographie), application majeure de l'interaction particules/matière, est le procédé traditionnel de fabrication des **circuits intégrés**. Étape clé de la définition des motifs de ces circuits, elle reste le verrou de leur développement. La résolution étant en première approximation directement proportionnelle à la longueur d'onde, la finesse des motifs a d'abord progressé avec la diminution, qui s'est effectuée par sauts, de la longueur d'onde λ du rayonnement utilisé.

L'opération consiste en l'*exposition via* une optique réductrice d'une résine photosensible à des particules énergétiques, depuis les **photons ultraviolet (UV)** actuellement utilisés jusqu'aux **électrons** en passant par les photons **X** et les **ions**, au travers d'un masque représentant le dessin d'un circuit. Le but ? Transférer cette image sur l'empilement de couches isolantes ou conductrices qui le constitueront, déposées précédemment (phase de *couchage*) sur une plaquette (*wafers*) de matériau **semi-conducteur**, en général de **silicium**. Ce processus est suivi de la dissolution de la résine exposée à la lumière (*développement*). Les parties exposées de la couche initiale peuvent être alors gravées sélectivement, puis la résine est retirée chimiquement avant le dépôt de la couche suivante. Cette étape de lithographie peut intervenir plus d'une vingtaine de fois au cours de la fabrication d'un circuit intégré (figure).

Dans les années 1980, l'industrie de la microélectronique utilisait des lampes à mercure délivrant dans l'**UV proche** (raies g, h, i), à travers des optiques en quartz, un rayonnement d'une longueur d'onde de 436 nanomètres (nm). Elle gravait ainsi des structures d'une largeur de trait de 3 microns (μm). Employées jusqu'au milieu des années 1990, ces lampes ont été remplacées par des **lasers à excimères** émettant dans l'**UV lointain** (krypton-fluor KrF à 248 nm, puis argon-fluor ArF à 193 nm, les photons créés ayant une énergie de quelques **électronvolts**), permettant d'atteindre des résolutions de 110 nm, et même inférieures à 90 nm avec de nouveaux procédés.

Le Laboratoire d'électronique et de technologie de l'information (Leti) du CEA a été un des pionniers, dans les années 1980, dans l'utilisation des lasers en



Zone de photolithographie en salle blanche dans l'usine STMicroelectronics de Crolles (Isère).

lithographie et dans la réalisation des circuits intégrés par les lasers à excimères, qui constituent aujourd'hui les sources employées pour la production des circuits intégrés les plus avancés. Pour l'industrie, l'étape suivante devait être le laser F₂ ($\lambda = 157 \text{ nm}$), mais cette lithographie a été quasiment abandonnée face à la difficulté de réaliser des optiques en CaF₂, matériau transparent à cette longueur d'onde.

Si la diminution de la longueur d'onde des outils d'exposition a été le premier facteur à permettre le gain en résolution considérable déjà obtenu, deux autres ont été déterminants. Le premier a été la mise au point de résines photosensibles basées sur des matrices de **polymères** peu absorbantes aux longueurs d'onde utilisées et mettant en œuvre des mécanismes de

propagation de l'énergie reçue toujours plus innovants. Le second a consisté en l'amélioration des optiques avec une diminution des phénomènes parasites liés à la diffraction (meilleure qualité de surface, augmentation de l'**ouverture numérique**).

Au fil des années, la complexité accrue des systèmes optiques a ainsi permis d'obtenir des résolutions *inférieures* à la longueur d'onde de la source. Cette évolution ne pourra se poursuivre sans une rupture technologique majeure, un saut important en longueur d'onde. Pour les générations des circuits intégrés dont la résolution minimale est comprise entre 80 et 50 nm (le prochain "nœud" se situant à 65 nm), différentes voies basées sur la projection de particules à la longueur d'onde de plus en plus courte ont été mises en concurrence. Elles mettent respectivement en œuvre des rayons X "mous", en extrême ultraviolet (dans la gamme des 10 nm), des rayons X "durs" (à la longueur d'onde inférieure à 1 nm), des ions ou des électrons.

L'étape consistant à atteindre des résolutions inférieures à 50 nm conduira à s'orienter plutôt vers la nanolithographie à l'aide d'électrons de basse énergie (10 eV) et d'outils plus adaptés comme le microscope à **effet tunnel** ou l'**épitaxie par jets moléculaires** (encadré C) pour la réalisation de "super-réseaux".

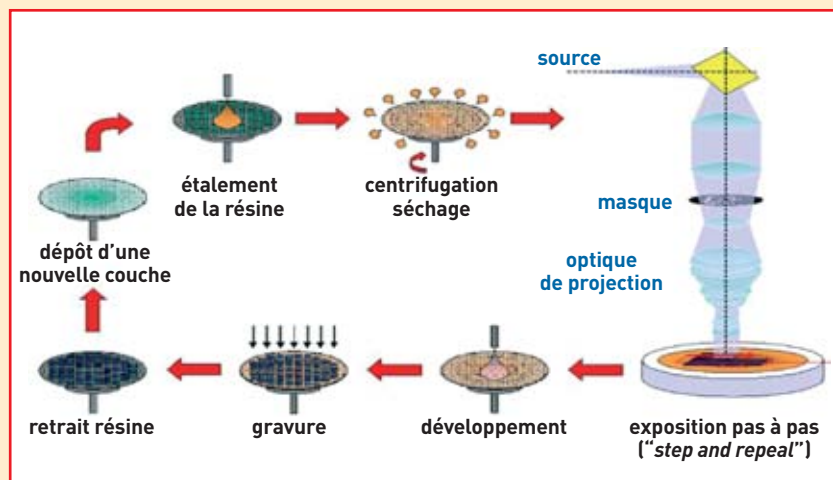


Figure. Les différentes phases du processus de lithographie dont le but est de délimiter les couches de matériaux conducteurs ou isolants qui constituent un circuit intégré. Cette opération est l'enchaînement d'un étalage de résine photosensible, de la projection du dessin d'un masque par une optique réductrice, suivis de la dissolution de la résine exposée à la lumière (développement). Les parties exposées de la couche initiale peuvent être alors gravées sélectivement, puis la résine est retirée avant le dépôt de la couche suivante.

A Du monde macroscopique au nanomonde, ou l'inverse...

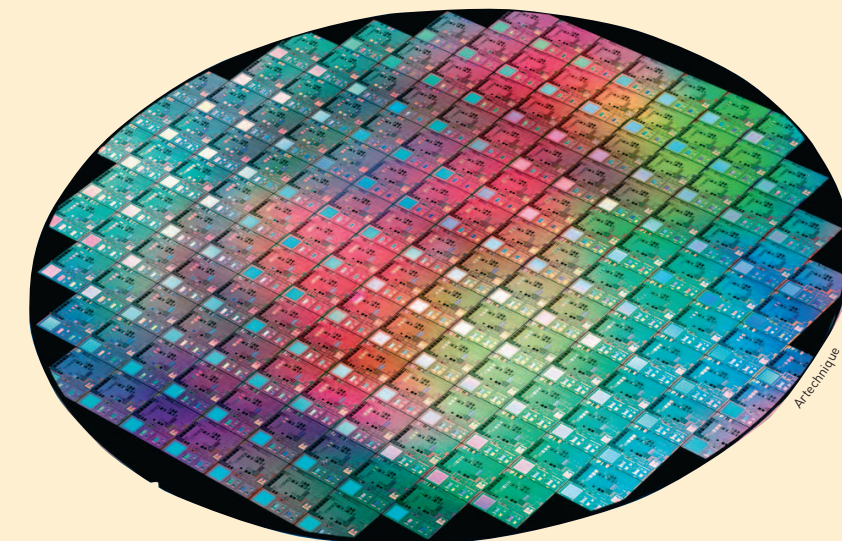
Afin de se représenter plus aisément les dimensions des objets micro et nanoscopiques*, il est pratique de procéder à des comparaisons et courant de faire correspondre différentes échelles, par exemple celle du monde du vivant, de la molécule à l'homme, et celle des objets manipulés ou fabriqués par lui (figure). Cette correspondance entre "artificiel" et "naturel" permet, par exemple, de voir que des **nanoparticules** fabriquées artificiellement sont plus petites que des globules rouges.

Un autre mérite de cette juxtaposition est d'illustrer les deux grandes façons

*Du grec *nano* qui signifie "tout petit" et est utilisé comme préfixe pour désigner le milliardième (10^{-9}) d'une unité. En l'occurrence, le **nanomètre** ($1 \text{ nm} = 10^{-9} \text{ m}$, soit un milliardième de mètre) est l'unité reine du monde des nanosciences et des nanotechnologies.

d'élaborer des objets ou des systèmes nanométriques : la voie descendante (**top-down**) et la voie ascendante (**bottom-up**). Deux chemins mènent en effet au nanomonde : la fabrication moléculaire, qui passe par la manipulation d'**atomes** individuels et la construction à partir de la base, et l'ultraminaturisation, qui produit des systèmes de plus en plus petits.

La voie descendante est celle du monde artificiel, qui part de matériaux macroscopiques, ciselés par la main de l'homme puis par ses instruments : c'est elle qu'a empruntée l'électronique depuis plusieurs dizaines d'années, principalement avec le **silicium** comme substrat, et ses "tranches" (**wafers**) comme entités manipulables. C'est d'ailleurs la microélectronique qui a largement contribué à donner à cette voie le nom anglais sous laquelle elle



Tranche de silicium de 300 mm réalisée par l'Alliance Crolles2, illustration de la démarche *top-down* actuelle de la microélectronique.

est connue. Mais il ne s'agit plus seulement d'adapter la miniaturisation de la filière silicium actuelle, mais aussi de prendre en compte, pour s'en prémunir ou les utiliser, les phénomènes physiques, **quantiques** en particulier, qui apparaissent aux faibles dimensions.

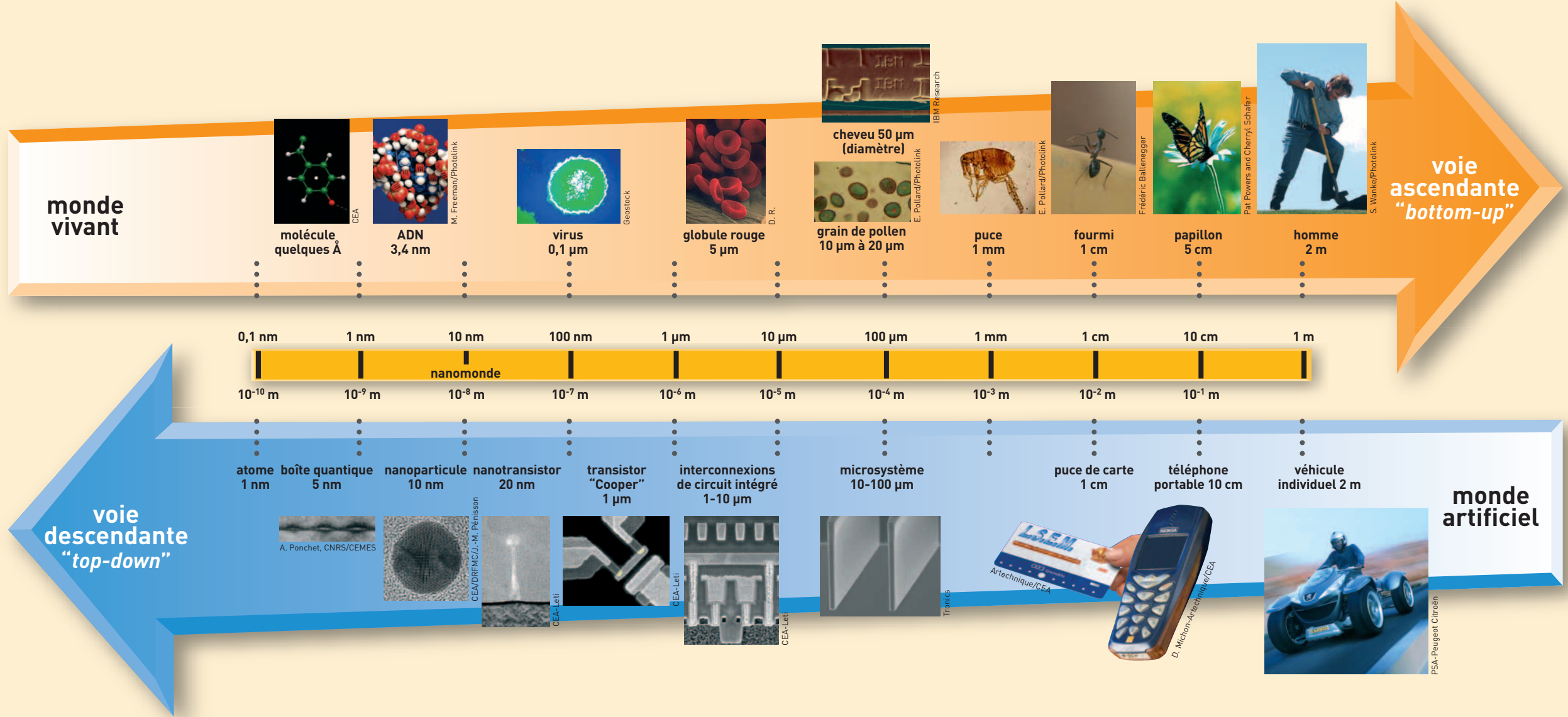
La voie ascendante peut permettre de passer outre ces limites physiques et aussi de réduire les coûts de fabrication, en utilisant notamment l'**auto-assemblage** des composants. C'est elle que suit la vie en pratiquant l'assemblage de molécules pour créer des **protéines**, enchaînement d'acides aminés que des super-molécules, les **acides nucléiques** (ADN, ARN), savent faire produire au sein de cellules pour former des organismes, les faire fonctionner et se reproduire tout en se complexifiant. Cette voie, dite "*bottom-up*", vise à orga-

niser la matière à partir de "briques de base", dont les atomes eux-mêmes sont les plus petits constituants, à l'instar du monde vivant. La nanoélectronique du futur cherche à emprunter cette voie d'assemblage pour aboutir à moindre coût à la fabrication d'éléments fonctionnels.

Les **nanosciences** peuvent ainsi être définies comme l'ensemble des recherches visant à la compréhension des propriétés (physiques, chimiques et biologiques) des nano-objets ainsi qu'à leur fabrication et à leur assemblage par auto-organisation.

Les **nanotechnologies** regroupent l'ensemble des savoir-faire qui permettent de travailler à l'échelle moléculaire pour organiser la matière afin de réaliser ces objets et matériaux, éventuellement jusqu'à l'échelle macroscopique.

A (Suite)



monde vivant

voie ascendante
"bottom-up"

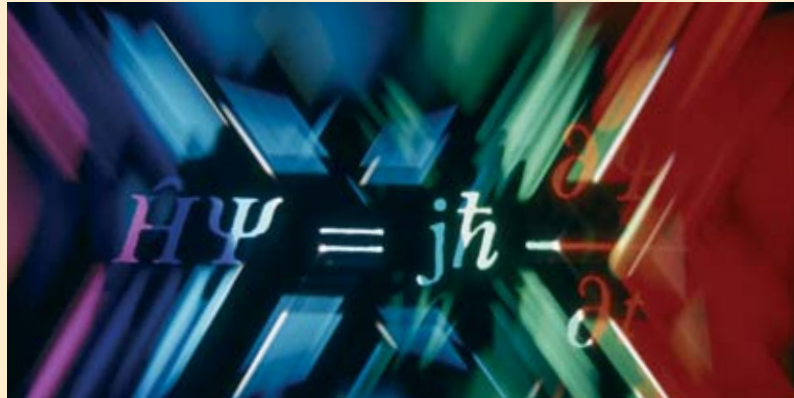
voie descendante
"top-down"

monde artificiel

B Quelques repères de physique quantique

La **physique quantique** (historiquement dénommée mécanique quantique) est l'ensemble des lois physiques qui s'appliquent à l'échelle microscopique. Fondamentalement différentes de la plupart de celles qui semblent s'appliquer à notre propre échelle, elles n'en constituent pas moins le socle global de la physique à toutes ses échelles. Mais à l'échelle macroscopique, ses manifestations ne nous apparaissent pas étranges, à l'exception d'un certain nombre de phénomènes *a priori* curieux, comme la **supraconductivité** ou la superfluidité, qui justement ne s'expliquent que par les lois de la physique quantique. Au demeurant, le passage du domaine de validité des lois paradoxales de cette physique à celui des lois, plus simples à imaginer, de la physique classique peut s'expliquer d'une façon très générale, comme cela sera évoqué plus loin.

La physique quantique tire son nom d'une caractéristique essentielle des objets quantiques : des caractéristiques comme le moment angulaire (**spin**) des particules sont des quantités **discrètes** ou discontinues appelées **quanta**, qui ne peuvent prendre que des valeurs multiples d'un *quantum* élémentaire. Il existe de même un **quantum d'action** (produit d'une énergie par une durée)



D. Sarrault/CEA

“Vue d’artiste” de l’équation de Schrödinger.

appelé **constante de Planck** (h), dont la valeur est de $6,626 \cdot 10^{-34}$ joule-seconde. Alors que la physique classique distingue *ondes* et *corpuscules*, la physique quantique englobe en quelque sorte ces deux concepts dans un troisième, qui dépasse la simple dualité onde-corpuscule entrevue par Louis de Broglie, et qui, quand nous tentons de l’appréhender, semble tantôt proche du premier et tantôt du deuxième. L’objet quantique constitue une entité inséparable de ses conditions d’observation, sans attribut propre. Et cela, qu’il s’agisse d’une particule – en aucun cas assimilable à une bille minuscule qui suivrait une quelconque trajectoire – de lumière

(**photon**) ou de matière (**électron, proton, neutron, atome...**).

Cette caractéristique donne toute sa force au **principe d’incertitude d’Heisenberg**, autre base de la physique quantique. Selon ce principe [d’*indétermination* plutôt que d’*incertitude*], il est impossible de définir avec précision à un instant donné à *la fois* la position d’une particule et sa vitesse. La mesure, qui reste possible, n’aura jamais une précision meilleure que h , la constante de Planck. Ces grandeurs n’ayant pas de réalité intrinsèque en dehors du processus d’observation, cette détermination simultanée de la position et de la vitesse est simplement impossible.

B (Suite)

C'est qu'à tout instant l'objet quantique présente la caractéristique de *superposer* plusieurs états, comme une onde peut être le résultat de l'*addition* de plusieurs autres. Dans le domaine quantique, la hauteur d'une onde (assimilable à celle d'une vague par exemple) a pour équivalent une **amplitude de probabilité** (ou onde de probabilité), nombre complexe associé à chacun des états possibles d'un système qualifié ainsi de quantique. Mathématiquement, un état physique d'un tel système est représenté par un **vecteur d'état**, fonction qui, en vertu du principe de superposition, peut s'ajouter à d'autres. Autrement dit, la somme de deux vecteurs d'état possibles d'un système est *aussi* un vecteur d'état possible du système. De plus, le produit de deux espaces vectoriels est aussi la somme de produits de vecteurs, ce qui traduit l'**intrication** : un vecteur d'état étant généralement étalé dans l'espace, l'idée de localité des objets ne va plus de soi. Dans une paire de particules intriquées, c'est-à-dire créées ensemble ou ayant déjà interagi l'une sur l'autre, décrite par le *produit* et non par la *somme* de deux vecteurs d'état individuels, le destin de chacune est lié à celui de l'autre, quelle que soit la distance qui pourra les séparer. Cette caractéristique, également appelée *l'enchevêtrement quantique d'états*, a

des implications vertigineuses, sans parler des applications imaginables, de la cryptographie quantique à – pourquoi ne pas rêver ? – la téléportation.

Dès lors, la possibilité de prévoir le comportement d'un système quantique n'est qu'une prédictibilité probabiliste et statistique. L'objet quantique est en quelque sorte une "juxtaposition de possibles". Tant que la mesure sur lui n'est pas faite, la grandeur censée quantifier la propriété physique recherchée n'est pas strictement définie. Mais dès que cette mesure est engagée, elle détruit la **superposition quantique**, par *réduction du paquet d'ondes*, comme Werner Heisenberg l'énonçait en 1927.

Toutes les propriétés d'un système quantique peuvent être déduites à partir de l'équation proposée l'année précédente par Erwin Schrödinger. La résolution de cette **équation de Schrödinger** permet de déterminer l'énergie du système ainsi que la **fonction d'onde**, notion qui a donc tendance à être remplacée par celle d'amplitude de probabilité.

Selon un autre grand principe de la physique quantique, le **principe (d'exclusion) de Pauli**, deux particules identiques de spin $\frac{1}{2}$ (c'est-à-dire des **fermions**, en particulier les électrons) ne peuvent avoir à la fois la même position, le même spin et la même vitesse (dans les limites

posées par le principe d'incertitude), c'est-à-dire se trouver dans le même *état quantique*. Les **bosons** (en particulier les photons), ne suivent pas ce principe et peuvent se trouver dans le même état quantique.

La coexistence des **états superposés** donne sa **cohérence** au système quantique. Dès lors, la théorie de la **décohérence quantique** peut expliquer pourquoi les objets macroscopiques ont un comportement "classique" tandis que les objets microscopiques, atomes et autres particules, ont un comportement quantique. Plus sûrement encore qu'un dispositif de mesure pointu, "l'environnement" (l'air, le rayonnement ambiant, etc.) exerce son influence, éliminant radicalement toutes les *superpositions d'état* à cette échelle. Plus le système considéré est gros, plus il est en effet couplé à un grand nombre de degrés de liberté de cet environnement. Et donc moins il a de "chances" – pour rester dans la logique probabiliste – de sauvegarder une quelconque cohérence quantique.

POUR EN SAVOIR PLUS

Étienne KLEIN, *Petit voyage dans le monde des quanta*, Champs, Flammarion, 2004.

C L'épitaxie par jets moléculaires

La fabrication des **puits quantiques** utilise la technique d'**épitaxie** (du grec *taxi* (ordre) et *epi* (dessus) par jets moléculaires (en anglais MBE, pour *Molecular Beam Epitaxy*). Le principe de cette technique de dépôt physique, développée initialement pour la croissance **crystalline** des **semi-conducteurs** de la famille III-V, est fondé sur l'évaporation des différents constituants purs du matériau à élaborer dans une enceinte où est maintenu un vide poussé (pression pouvant être de l'ordre de $5 \cdot 10^{-11}$ mbar) afin d'éviter toute pollution de la surface. Un ou des jets thermiques d'**atomes** ou de **molécules** réagissent sur la surface propre d'un substrat monocristallin, placé sur un support maintenu à haute température (quelques centaines de °C), qui sert de trame pour former un film dit épitaxique. Il est ainsi possible de fabriquer des empilements de couches aussi fines que le milliardième de millimètre, c'est-à-dire composées de seulement quelques plans d'atomes.

Les éléments sont évaporés ou sublimés à partir d'une source de haute pureté, placée dans une **cellule à effusion** (chambre dans laquelle un flux moléculaire passe d'une région où règne une pression donnée à une région de plus basse pression) chauffée par **effet Joule**.

La croissance du film peut être suivie *in situ* et en temps réel en utilisant diverses sondes structurales et analytiques, en particulier des techniques d'étude de la qualité des surfaces et de leurs transitions de phase par diffraction électronique en incidence rasante, LEED (pour *Low energy electron diffraction*) ou RHEED (pour *Reflection high-energy electron diffraction*) et diverses méthodes **spectroscopiques** (spectroscopie d'**électrons** Auger, SIMS (spectrométrie de masse d'**ions** secondaires), spectrométrie de photoélectrons XPS par **rayons X** et UPS (*Ultraviolet photoelectron spectroscopy*).

La technique d'épitaxie par jets moléculaires s'est étendue à d'autres semi-

conducteurs que les III-V, à des métaux et à des isolants, se développant avec les progrès des techniques d'*ultravide*. Le vide régnant dans la chambre de croissance, dont la conception varie en fonction de la nature du matériau à déposer, doit en effet être meilleure que 10^{-11} mbar pour permettre la croissance d'un film de haute pureté et d'excellente qualité cristalline à des températures de substrat relativement basses. Il s'agit de qualité de vide lorsque le bâti est au repos. Pour la croissance d'arséniures, par exemple, le vide résiduel est de l'ordre de 10^{-8} mbar dès que la cellule d'arsenic est portée à sa température de consigne pour la croissance. Le pompage pour atteindre ces performances fait appel à plusieurs techniques (pompage ionique, cryopompage, sublimation de titane, pompes à diffusion ou turbomoléculaires). Les principales impuretés (H_2 , H_2O , CO et CO_2) peuvent présenter des pressions partielles inférieures à 10^{-13} mbar.

D Le transistor, composant de base des circuits intégrés

En décembre 1947, John Bardeen et Walter H. Brattain réalisaient le premier **transistor** en germanium. Avec William B. Shockley, aux Bell Laboratories, ils développaient l'année suivante le transistor à jonction et la théorie associée. Au milieu des années 1950, les transistors seront réalisés en **silicium** (Si), qui reste aujourd'hui le **semi-conducteur** généralement utilisé, vu la qualité inégalée de l'interface créée par le silicium et l'oxyde de silicium (SiO_2), qui sert d'isolant.

En 1958, Jack Kilby invente le **circuit intégré** en fabriquant cinq composants sur le même substrat. Les années 1970 verront le premier microprocesseur d'Intel (2250 transistors) et les premières mémoires. La complexité des circuits intégrés ne cessera de croître exponentiellement depuis (doublement tous les deux-trois ans, selon la "loi de Moore") grâce à la miniaturisation des transistors. Le transistor (de l'anglais *transfer resistor*, résistance de transfert), composant de base des **circuits intégrés** micro-

électroniques, le restera *mutatis mutandis* à l'échelle de la nanoélectronique : adapté également à l'amplification, entre autres fonctions, il assume en effet une fonction basique essentielle : laisser passer un courant ou l'interrompre à la demande, à la manière d'un commutateur (figure). Son principe de base s'applique donc directement au traitement du langage binaire (0, le courant ne passe pas ; 1, il passe) dans des circuits logiques (inverseurs, portes, additionneurs, cellules mémoire).

Le transistor, fondé sur le transport des électrons dans un solide et non plus dans le vide comme dans les tubes électroniques des anciennes **triodes**, est composé de trois **électrodes** (*anode, cathode et grille*) dont deux servent de réservoirs à **électrons** : la **source**, équivalent du filament *émetteur* du tube électronique, le **drain**, équivalent de la plaque *collectrice*, et la **grille**, le "contrôleur". Ces éléments ne fonctionnent pas de la même manière dans les deux principaux types de transistors utilisés aujourd'hui, les *transistors bipolaires à jonction*, qui ont été les premiers à être utilisés, et les *transistors à effet de champ* (en anglais FET, *Field Effect Transistor*). Les transistors bipolaires mettent en œuvre les deux types de porteurs de charge, les électrons (charges négatives) et les **trous** (charges positives), et se composent de deux parties de substrat semi-conducteur identiquement

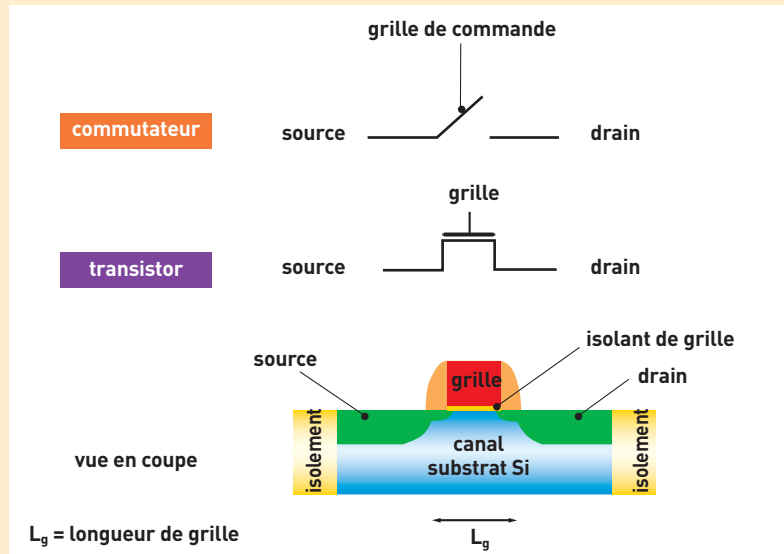


Figure. Un transistor MOS est un commutateur qui permet de commander le passage d'un courant électrique de la source (S) vers le drain (D) à l'aide d'une grille (G) isolée électriquement du canal de conduction. Le substrat en silicium est noté B (pour *Bulk*).

D (Suite)

dopées (p ou n), séparées par une mince couche de semi-conducteur inversement dopée. L'assemblage de deux semi-conducteurs de types opposés (jonction p-n) permet de ne faire passer le courant que dans un sens. Qu'ils soient de type n-p-n ou p-n-p, les transistors bipolaires sont fondamentalement des amplificateurs de courant, commandés par un courant de grille⁽¹⁾ : ainsi dans un transistor n-p-n, la tension appliquée à la partie p contrôle le passage du courant entre les deux régions n. Les circuits logiques utilisant des transistors bipolaires, appelés TTL (*Transistor Transistor Logic*), sont plus consommateurs de courant que les transistors à effet de champ, qui présentent un courant de grille nul en régime statique et sont commandés par l'application d'une tension.

Ce sont ces derniers, sous la forme **MOS** (Métal oxyde semi-conducteur), qui composent aujourd'hui la plupart des circuits logiques du type CMOS (C pour complémentaire)⁽²⁾. Sur un cristal de silicium de type p, deux régions de type n sont créées par dopage de la surface. Appelées là aussi **source** et **drain**, ces deux régions ne sont donc séparées que par un petit espace de type p, le **canal**. Sous l'effet d'une tension positive sur une électrode de commande placée

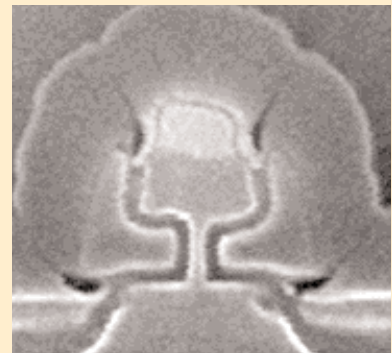
au-dessus du semi-conducteur et qui porte tout naturellement le nom de **grille**, les trous sont repoussés de sa surface où viennent s'accumuler les quelques électrons du semi-conducteur. Un petit canal de conduction peut ainsi se former entre la source et le drain (figure). Lorsqu'une tension négative est appliquée sur la grille, isolée électriquement par une couche d'oxyde, les électrons sont repoussés hors du canal. Plus la tension positive est élevée, plus la résistance du canal diminue et plus ce dernier laisse passer de courant. Dans un circuit intégré, les transistors et les autres composants (diodes, condensateurs, résistances) sont d'origine incorporés au sein d'une "puce" aux fonctions plus ou moins complexes. Le circuit est constitué d'un empilement de couches de matériaux conducteurs ou isolants délimitées par **lithographie** (encadré E, *La lithographie clé de la miniaturisation*, p. 37). L'exemple le plus emblématique est le microprocesseur placé au cœur des ordinateurs et qui regroupe plusieurs centaines de millions de transistors (dont la taille a été réduite par 10 000 depuis les années 1960) et bientôt un milliard, ce qui amène les industriels à fractionner le cœur des processeurs en plusieurs sous-unités travaillant en parallèle !

(1) Figurent dans cette catégorie les **transistors de type Schottky** ou à **barrière Schottky** qui sont des transistors à effet de champ comportant une grille de commande de type métal/semi-conducteur qui améliore la mobilité des porteurs de charge et le temps de réponse au prix d'une plus grande complexité.

(2) On parle alors de transistor **MOSFET** (*Metal-Oxide Semiconductor Field Effect Transistor*).



Le tout premier transistor.



Transistor 8 nanomètres développé par l'Alliance Crolles2 réunissant STMicroelectronics, Philips et Freescale Semiconductor.

E La lithographie, clé de la miniaturisation

La **lithographie** optique (ou photolithographie), application majeure de l'interaction particules/matière, est le procédé traditionnel de fabrication des **circuits intégrés**. Étape clé de la définition des motifs de ces circuits, elle reste le verrou de leur développement. La résolution étant en première approximation directement proportionnelle à la longueur d'onde, la finesse des motifs a d'abord progressé avec la diminution, qui s'est effectuée par sauts, de la longueur d'onde λ du rayonnement utilisé.

L'opération consiste en l'*exposition via* une optique réductrice d'une résine photosensible à des particules énergétiques, depuis les **photons ultraviolet (UV)** actuellement utilisés jusqu'aux **électrons** en passant par les photons **X** et les **ions**, au travers d'un masque représentant le dessin d'un circuit. Le but ? Transférer cette image sur l'empilement de couches isolantes ou conductrices qui le constitueront, déposées précédemment (phase de *couchage*) sur une plaquette (*wafers*) de matériau **semi-conducteur**, en général de **silicium**. Ce processus est suivi de la dissolution de la résine exposée à la lumière (*développement*). Les parties exposées de la couche initiale peuvent être alors gravées sélectivement, puis la résine est retirée chimiquement avant le dépôt de la couche suivante. Cette étape de lithographie peut intervenir plus d'une vingtaine de fois au cours de la fabrication d'un circuit intégré (figure).

Dans les années 1980, l'industrie de la microélectronique utilisait des lampes à mercure délivrant dans l'**UV proche** (raies g, h, i), à travers des optiques en quartz, un rayonnement d'une longueur d'onde de 436 nanomètres (nm). Elle gravait ainsi des structures d'une largeur de trait de 3 microns (μm). Employées jusqu'au milieu des années 1990, ces lampes ont été remplacées par des **lasers à excimères** émettant dans l'**UV lointain** (krypton-fluor KrF à 248 nm, puis argon-fluor ArF à 193 nm, les photons créés ayant une énergie de quelques **électronvolts**), permettant d'atteindre des résolutions de 110 nm, et même inférieures à 90 nm avec de nouveaux procédés.

Le Laboratoire d'électronique et de technologie de l'information (Leti) du CEA a été un des pionniers, dans les années 1980, dans l'utilisation des lasers en



Zone de photolithographie en salle blanche dans l'usine STMicroelectronics de Crolles (Isère).

lithographie et dans la réalisation des circuits intégrés par les lasers à excimères, qui constituent aujourd'hui les sources employées pour la production des circuits intégrés les plus avancés. Pour l'industrie, l'étape suivante devait être le laser F₂ ($\lambda = 157 \text{ nm}$), mais cette lithographie a été quasiment abandonnée face à la difficulté de réaliser des optiques en CaF₂, matériau transparent à cette longueur d'onde.

Si la diminution de la longueur d'onde des outils d'exposition a été le premier facteur à permettre le gain en résolution considérable déjà obtenu, deux autres ont été déterminants. Le premier a été la mise au point de résines photosensibles basées sur des matrices de **polymères** peu absorbantes aux longueurs d'onde utilisées et mettant en œuvre des mécanismes de

propagation de l'énergie reçue toujours plus innovants. Le second a consisté en l'amélioration des optiques avec une diminution des phénomènes parasites liés à la diffraction (meilleure qualité de surface, augmentation de l'**ouverture numérique**).

Au fil des années, la complexité accrue des systèmes optiques a ainsi permis d'obtenir des résolutions *inférieures* à la longueur d'onde de la source. Cette évolution ne pourra se poursuivre sans une rupture technologique majeure, un saut important en longueur d'onde. Pour les gé-

nerations des circuits intégrés dont la résolution minimale est comprise entre 80 et 50 nm (le prochain "nœud" se situant à 65 nm), différentes voies basées sur la projection de particules à la longueur d'onde de plus en plus courte ont été mises en concurrence. Elles mettent respectivement en œuvre des rayons X "mous", en extrême ultraviolet (dans la gamme des 10 nm), des rayons X "durs" (à la longueur d'onde inférieure à 1 nm), des ions ou des électrons.

L'étape consistant à atteindre des résolutions inférieures à 50 nm conduira à s'orienter plutôt vers la nanolithographie à l'aide d'électrons de basse énergie (10 eV) et d'outils plus adaptés comme le microscope à **effet tunnel** ou l'**épitaxie par jets moléculaires** (encadré C) pour la réalisation de "super-réseaux".

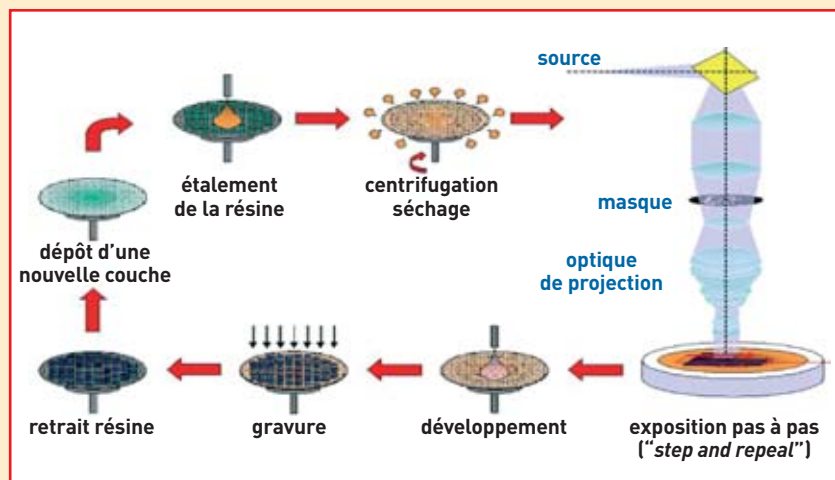


Figure. Les différentes phases du processus de lithographie dont le but est de délimiter les couches de matériaux conducteurs ou isolants qui constituent un circuit intégré. Cette opération est l'enchaînement d'un étalement de résine photosensible, de la projection du dessin d'un masque par une optique réductrice, suivis de la dissolution de la résine exposée à la lumière (développement). Les parties exposées de la couche initiale peuvent être alors gravées sélectivement, puis la résine est retirée avant le dépôt de la couche suivante.

G

L'effet tunnel, un phénomène quantique

La **physique quantique** prédit des comportements inhabituels et difficiles à accepter par notre intuition immédiate, comme l'**effet tunnel**. Prenons le cas d'une bille devant franchir une bosse. En physique classique, si l'énergie communiquée à la bille est insuffisante, elle ne peut pas franchir la bosse et retombe vers son point de départ. En physique quantique, une particule (**proton, électron**) peut franchir la bosse, même si son énergie initiale est insuffisante : elle peut passer de l'autre côté comme par un petit tunnel. L'effet tunnel peut ainsi permettre à deux protons de surmonter leur répulsion électrique à des vitesses relatives plus basses que celles qu'indique le calcul classique. La microscopie à effet tunnel est basée sur le fait qu'il y a une probabilité non nulle qu'une particule d'énergie inférieure à la hauteur d'une barrière de potentiel (la

bosse) puisse la franchir. Les particules sont des électrons traversant l'espace séparant deux **électrodes**, une fine pointe métallique terminée par un **atome** unique et la surface métallique ou **semi-conductrice** de l'échantillon. La physique classique donne d'une surface l'image d'une frontière étanche, les électrons étant strictement confinés à l'intérieur du solide. Par contre, la physique quantique enseigne que chaque électron a un comportement ondulatoire : sa position est "floue". En particulier, au voisinage de la surface existe un nuage d'électrons dont la densité décroît très rapidement, de façon exponentielle, lorsque l'on s'éloigne du solide. L'électron a une certaine probabilité de se trouver "en dehors" du solide. Quand la fine pointe métallique est approchée de la surface, à une distance inférieure au **nanomètre**, la **fonction d'onde** associée à l'électron n'est

pas nulle de l'autre côté de la barrière de potentiel, et les électrons passent de la surface à la pointe, et réciproquement, par effet tunnel. La barrière de potentiel franchie par les électrons est appelée **barrière tunnel**. Lorsqu'une faible tension est appliquée entre la pointe et la surface, un **courant tunnel** peut être détecté. La pointe et la surface étudiée forment localement une **jonction tunnel**. L'effet tunnel se manifeste également dans les **jonctions Josephson** où un courant continu peut passer à travers une étroite discontinuité entre deux éléments **supraconducteurs**. Dans un **transistor**, l'effet tunnel peut se révéler de manière parasite quand l'isolant de **grille** devient très mince (de l'ordre du nanomètre). Il est par contre mis à profit dans de nouvelles architectures, tels les **transistors à barrière tunnel Schottky** ou à base de **nanotubes de carbone**.

La loi de Moore

En 1965, Gordon Moore, cofondateur d'Intel Corporation, publia un article visionnaire prévoyant la croissance exponentielle du nombre de **transistors** des **circuits intégrés**, soulignant aussi le lien entre intégration, coût, performances et fiabilité. Selon lui, le nombre de transistors sur une même surface de circuit intégré (CI) devait doubler chaque année. Son observation a évolué par la suite dans le sens d'un doublement du nombre de transistors tous les dix-huit mois à coût constant, puis, depuis 1995, tous les deux ans, voire trois.

Cette célèbre "loi de Moore", parfaitement empirique, n'a pour l'essentiel (à savoir la croissance exponentielle du nombre de composants par unité de surface) jamais été mise en défaut, malgré l'infléchissement constaté ces dernières années. Elle devrait rester valable pendant plusieurs années (jusqu'en 2012?). Au-delà, l'évolution des procédés de gravure (encadré E, **La lithographie, clé de la miniaturisation**, p. 37) ou des architectures pourrait ralentir ce taux d'expansion, jusqu'au moment où sera atteinte une limite physique. Actuellement, dans un circuit intégré, la barrière isolante qui sépare une grille de transistor et le canal entre source et drain n'est que de quelques atomes. En poursuivant la miniaturisation, elle devient si fine que l'**effet tunnel quantique** commence à se manifester, créant une fuite de courant lorsque le commutateur est en position fermée, fuite qui constitue une fraction significative du courant du canal quand le dispositif est en position ouverte, ce qui pourrait réduire la fiabilité des transistors. La loi de Moore se traduit aujourd'hui par une feuille de route ITRS (*International Technology Roadmap for Semiconductors*), où les générations technologiques sont appelées "nœuds", dont chacun s'est vu associer des dimensions caractéristiques dont la référence est, jusqu'à maintenant, le demi-pas de gravure des mémoires DRAM (figure).

L'industrie de la microélectronique a ainsi pu progresser en suivant jusque-là un cercle vertueux : la recherche pour des dimensions réduites a permis la production à moindre coût de circuits intégrés aux fonctionnalités augmentées, créant de nouveaux marchés qui financent la R&D de nouvelles générations de dispositifs, le processus s'auto-entretenant.



Artechnique

Contrôle visuel des circuits intégrés dans les installations de l'Alliance Crolles2.

