

Le transistor et les progrès de la physique qui l'ont rendu possible

Travail de maturité réalisé au Lycée Denis-de-Rougemont de Neuchâtel
sous la direction de M. Michel Favez

Nathan Bernier

Introduction

Aujourd'hui, l'électronique nous accompagne en permanence. Nous ne faisons pas un pas sans prendre avec nous nos appareils indispensables, et nous ne pourrions plus travailler ou communiquer sans eux. Parce qu'elle est pratique et abordable, cette technologie est devenue omniprésente, et nous aide (ou prétend nous aider) au quotidien.

Il surprend de constater que la base sur laquelle s'est bâtie cette science reste très jeune dans l'Histoire de l'humanité. Tout repose sur une invention de 1947, le transistor.

Le transistor est une valve dans un circuit électrique. Il peut remplir deux fonctions: d'abord, il sert à amplifier un signal; ensuite, on l'utilise comme un interrupteur à deux positions, ouvert ou fermé. De ce fait, les applications du transistor sont premièrement les communications et l'amplification audio, et ensuite les circuits logiques. En effet, le transistor est la brique élémentaire de tout ordinateur ou autre machine à calculer. Les circuits intégrés et les microprocesseurs en contiennent des millions. Le langage binaire est basé sur les états ouvert et fermé qui correspondent à autant de uns et de zéros.

La lampe

Bien que le transistor constitue une révolution technique, il n'a pas réellement offert de nouvelles possibilités. Sa révolution tient plus de la forme que du fond. En effet, tout ce que peut faire un transistor, une triode en est capable.

La triode est un tube à vide (figure 1) dans lequel se trouvent trois électrodes: une cathode, un filament comme celui d'une ampoule; une anode, chargée positivement; et une grille, chargée négativement. Le filament est chauffé et un potentiel négatif y est appliqué. Il se met à émettre des électrons. Ceux-ci sont attirés par l'anode, mais repoussés par la grille. En faisant varier la charge négative de la grille, on fait varier le courant entre la cathode et l'anode.

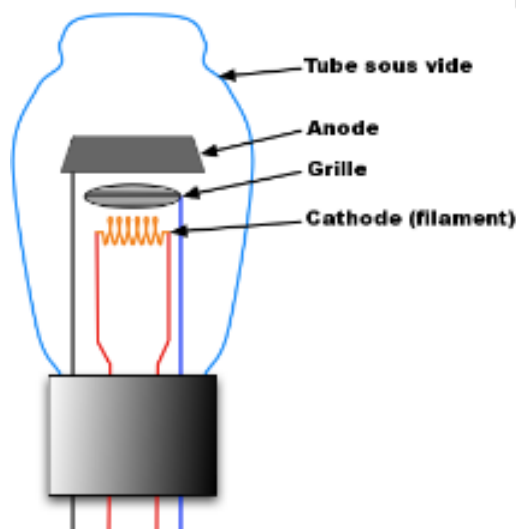


fig 1. Schéma d'une triode (tube électronique) (traduction d'un document provenant de Wikipedia)

Le transistor fut développé par Bell Labs, la branche de recherche de l'entreprise de télécommunications American Telephone & Telegraphs (AT&T). En effet, le premier but du développement du transistor consistait à remplacer le tube à vide pour l'amplification de signal électrique. AT&T s'en servait comme relais pour convoier des conversations téléphoniques à très longues distances.

Dans les années quarante, les lampes semblaient en bout de course. Pendant la deuxième Guerre mondiale, les Anglais et les Américains les remplaçaient par des dispositifs à base de semi-conducteurs dans les radars pour pouvoir utiliser des ondes radios plus courtes. Les tubes présentent de nombreux inconvénients. Premièrement, ils sont encombrants et difficiles à miniaturiser. De plus, ils demandent des précautions pour leur utilisation: on doit les laisser chauffer avant de pouvoir les employer; ils peuvent surchauffer; et ils génèrent de fortes tensions très dangereuses. Enfin, ils consomment énormément d'énergie.

But et Structure du travail

Le transistor est la base de l'électronique moderne. Il a façonné la deuxième moitié du XXe siècle. Et aujourd'hui, ma génération est en quelque sorte sa progéniture. Je trouve intéressant de remettre cette invention dans son contexte, et d'adopter le point de vue des scientifiques qui ont contribué à sa découverte.

L'invention du transistor constitue l'aboutissement final d'un demi siècle d'ébullition scientifique, pendant lequel les fondements de la physique classique furent renversés et une toute nouvelle théorie prit son envol. De 1897, lorsqu'on découvrit l'électron, jusqu'à la construction du premier transistor en 1947, 50 ans d'avancées scientifiques ont redéfini la manière d'appréhender le microscopique.

Dans cette optique, j'ai trouvé intéressant de séparer mon travail en deux parties. En premier lieu, il s'agit d'un historique de l'évolution de la physique de la première moitié du XXe siècle, comment la compréhension de l'électronique s'est modifiée. Je finis par un bref récit de l'invention du transistor. Dans la seconde partie, je présente un certain type de transistor et explique son fonctionnement.

Un demi-siècle de révolutions

L'électronique à la fin du dix-neuvième siècle

Dès l'invention de la pile par Volta en 1799, les physiciens ont étudié les circuits électriques. Plus particulièrement, à la fin du 19e siècle, ils s'intéressaient aux tubes à vide tels que ceux de Geissler ou de Crookes. Ces derniers consistaient en une anode et une cathode, montées dans l'atmosphère appauvrie d'un bulbe de verre. Ils produisaient de mystérieux rayons cathodiques entre les deux électrodes, qui, par exemple, provoquent des lueurs lorsque qu'ils percutent une surface du tube. Ces recherches aboutirent à l'invention de l'ampoule, à la découverte des rayons X et plus tard ces tubes engendrèrent le tube électronique et l'écran cathodique. Pendant tout le dix-neuvième siècle, les scientifiques employèrent l'électricité sans jamais en comprendre la nature, la prenant pour un flux d'énergie continu.

La nature de l'électricité

Le passage du siècle connut la fin de cette ignorance. En effet, Joseph John Thomson, qui publia ses résultats en 1897 et 1899, découvrit que les rayons cathodiques étaient en réalité des particules chargées et les briques élémentaires de toute matière: les électrons. Ce fut la première fois (mais pas la dernière) dans l'histoire de la physique que ce que l'on croyait fluide et même onde se révéla en réalité constitué de corpuscules solides de nombre fini. En effet, on se rendit compte que le courant électrique consistait en un flot d'électrons circulant à travers un matériau conducteur.

La conductivité que l'on observe dans les métaux provient ainsi d'un certain nombre d'électrons libres qui peuvent être "pompés" par une tension appliquée au métal et dont le

déplacement génère un courant. Les différences de conductivité des métaux proviennent d'une facilité variable à libérer des électrons. Cependant, comment seulement une fraction des électrons pouvait quitter l'atome et pourquoi le reste d'entre eux semblait figé et incapable d'emmagasiner de l'énergie thermique restaient incompréhensibles.

La quantification de la lumière

Bien que s'écartant légèrement du sujet, ce chapitre se révèle indispensable. Dans l'histoire de la physique, la quantification de la lumière a ouvert la porte à de nouveaux concepts dont les fructifications ont révolutionné la compréhension de la matière à l'échelle microscopique. Elle fut le premier pas vers la mécanique quantique.

En parallèle à la découverte de l'électron, à la fin du 19^e siècle, un groupe de physiciens (dont Wien, Boltzman et Planck) s'acharnaient à unifier deux domaines de la physique, l'électromagnétisme et la thermodynamique. Dans ce but, ils s'employaient à déterminer une loi générale décrivant l'émission lumineuse d'un corps noir en fonction de sa température. Un corps noir est un objet qui absorbe tout le rayonnement électromagnétique auquel il est soumis. Un exemple classique consiste en un orifice donnant sur une chambre noire. Un tel objet émet de l'énergie dans toutes les fréquences du spectre, indépendamment de la matière dont il est constitué, uniquement en fonction de sa température. De par sa nature universelle, le corps noir semblait parfait pour servir de base à une loi générale.

Il y eut beaucoup d'essais infructueux pour trouver cette miraculeuse fonction, dont certaines divergeaient avec l'expérience et d'autres prédisaient une énergie totale infinie. Enfin, en 1900, par déduction et intuition, Planck découvrit une formule fidèle aux résultats expérimentaux. Insatisfait, il lui fallait encore pouvoir la prouver théoriquement. Dans une étape de ses calculs, il décomposa l'énergie totale en un nombre entier de petites énergies. Il pensait pouvoir plus tard rendre infiniment petits ces "morceaux" d'énergie qu'il prenait seulement pour un moyen d'arriver à ses fins. Cependant, il finit par se rendre compte que pour que sa fonction tienne debout, il fallait que ces énergies ε soient finies et dépendent de la fréquence d'émission : $\varepsilon = hf$. Où h est la constante universelle de Planck et f la fréquence de l'onde électromagnétique.

Il fallut attendre jusqu'en 1905 pour qu'Einstein tire les conclusions de ce résultat. Dans le deuxième de ses trois articles de cette "année miraculeuse", il déduisit du travail de Plank que la lumière possède une double nature, à la fois onde et particule, et se compose de quanta (que l'on nomme plus volontiers aujourd'hui photons). Alors qu'après des siècles de discussion, on venait juste de se persuader que la lumière était constituée d'ondes électromagnétiques, on apprenait qu'elle est en réalité faite de corpuscules! Einstein continue son article en expliquant le phénomène photoélectrique à l'aide de ce nouveau point de vue (ce qui lui vaudra son prix Nobel).

La place de l'électron

Les modèles atomiques

Depuis un moment déjà, on pensait que des charges électriques bougeant dans l'atome étaient responsables de l'émission de lumière. En 1896, suivant une idée de Faraday qu'il n'avait pu prouver dû aux limitations matérielles de son temps, Zeeman réussit à modifier le spectre¹ de la vapeur de sodium grâce à un champ magnétique. Il fit part de ce résultat à Lorentz, qui comprit que le champ magnétique agissait sur des particules chargées dont le mouvement provoquait l'émission électromagnétique. Ensemble, ils calculèrent le rapport charge sur masse de cette particule. C'est le même rapport que Thomson calcula plus tard pour les rayons cathodiques, ce qui prouva qu'il s'agissait des mêmes électrons.

¹ Un spectre correspond à l'ensemble des fréquences des ondes émises par un objet. Ici, il s'agit d'ondes électromagnétiques.

Bien que les physiciens du début du vingtième siècle possédaient de bonnes connaissances sur certaines particules radioactives et sur l'électron, ils ignoraient presque tout de la structure de l'atome. La théorie généralement acceptée décrivait des électrons chargés négativement flottant dans une gelée positive. Cependant, Rutherford, en 1911 après avoir assisté à la célèbre expérience où l'on bombarde une feuille d'or de particules alpha, déduit qu'il existait un noyau positif très dense sur lequel rebondissaient certains rayons alpha.

Cette découverte encouragea une représentation planétaire de l'atome, avec les électrons orbitant autour du noyau à la manière des satellites autour d'un corps céleste. Ce modèle posait des problèmes de stabilité. En effet, sans rien pour les retenir en place, les électrons dissiperaient en un instant toute leur énergie cinétique en lumière et s'écraseraient lamentablement sur le noyau. De plus, rien ne déterminerait le rayon atomique, que l'on savait stable. Il fallait maintenir les électrons en place et Bohr eut la bonne idée.

Le physicien danois fit vers 1913 la supposition suivante: l'électron peut effectuer un certain nombre d'orbites stables autour du noyau, chacune possédant une énergie propre. Ensuite, il peut changer d'orbite uniquement en émettant ou en absorbant un quantum dont l'énergie correspond à la différence d'énergie entre les deux niveaux. Il absorbe de la lumière pour gagner une orbite plus grande et en émet pour redescendre sur une orbite plus basse. Bohr utilisa le fait que l'on connaissait assez précisément le spectre de l'hydrogène à cette époque. Grâce à cela, il réussit à trouver des formules prédisant précisément la fréquence d'émission en fonction des changements de niveaux d'énergie. Son modèle possédait même une énergie minimale en dessous de laquelle l'électron ne pouvait pas descendre, assurant la stabilité de l'atome. Un des résultats surprenant de cette théorie, fut que non seulement l'énergie de l'électron ne prend que des valeurs discrètes, mais son moment cinétique¹ également. La norme de ce dernier ne peut valoir que des multiples entiers de $\hbar = h/2\pi$.

Ce modèle, loin d'être parfait, permettait de décrire le comportement de l'hydrogène, mais ne fonctionnait déjà plus pour l'hélium, l'atome suivant le plus simple. Adoptant une démarche empirique, en observant plus précisément les spectres, de nombreux physiciens réussirent à améliorer les formules. A la suite de ces développements, on se rendit compte d'un troisième nombre quantique décrivant l'électron, fixant l'orientation dans l'espace de son vecteur moment angulaire.

La nouvelle avancée théorique qui permit à la physique quantique de repartir fut apportée tôt en 1925 par Wolfgang Pauli. Celui-ci, ayant étudié avec soin tous les résultats expérimentaux de l'étude des spectres, avança l'idée qu'une certaine orbite (définie par 4 nombres quantiques pour un atome) peut être occupée par un seul électron au maximum. C'est le principe d'exclusion de Pauli. La découverte du quatrième nombre quantique, le spin² vint bientôt étayer sa thèse. Enfin, on comprenait pourquoi tous les électrons n'occupaient pas la position avec l'énergie la plus basse (comme ils le feraient dans un système classique), mais restaient à différents niveaux.

La naissance de la mécanique quantique

Un grand pas restait encore à franchir pour se départir définitivement de la physique classique. Louis de Broglie ne se satisfaisait pas de la définition de la lumière. En effet, on n'avait réussi qu'à prouver qu'elle agissait parfois comme une onde et parfois comme une particule, sans pouvoir unifier sa description. De plus, il s'était aperçu que le comportement de l'électron dans l'atome ressemblait à des interférences et des modes de vibration, caractéristiques d'une

¹ Le moment cinétique est le produit vectoriel de la quantité de mouvement par la position $\vec{L} = \vec{r} \times \vec{p}$, où la quantité de mouvement est $\vec{p} = m\vec{v}$, avec m la masse, \vec{v} la vitesse de l'objet et \vec{r} vecteur position.

² Le spin est un paramètre quantique caractérisant l'électron représenté comme la rotation de l'électron sur lui-même. Il ne peut valoir que $\pm\hbar/2$ (voir Diu, 2005).

onde. Il développa ses théories entre 1923 et 1924, et arriva à la relation $\lambda = \frac{h}{p}$ qui lie la quantité de mouvement p d'une particule à la longueur d'onde λ associée. Voilà que toute matière possède également les propriétés d'une onde.

A partir de là, Schrödinger prit le relais en entreprenant de développer une théorie d'onde complète. Il finit par trouver sa célèbre équation, dans un travail publié en janvier 1926, qui décrit le comportement d'un fermion¹ :

$$\nabla^2\Psi(x,y,z) + \frac{2m}{\hbar^2} \{E - U(x,y,z)\}\Psi(x,y,z) = 0$$

Où m est la masse du fermion, \hbar est la constante de Planck divisée par 2π , E est l'énergie totale de la particule, U est l'énergie potentielle de la particule à un certain point. Enfin, ψ est la fonction d'onde de la particule à ce point et $\nabla^2\Psi$, sa deuxième dérivée².

Cette équation, bien que paraissant difficile, arbore la forme normale d'une équation de propagation d'onde, et était donc déjà bien maîtrisée à l'époque. Une seule question subsistait par rapport à la nature de la fonction ψ , qui donne un nombre complexe dépendant d'un point de l'espace. Dans un premier temps, Schrödinger crut que le carré de sa norme $|\psi^2|$ représentait la densité de l'électron en un point. La fonction décrirait ainsi des nuages électroniques où l'électron se diluerait (voir Annexe 1). On comprit plus tard qu'il s'agissait en fait de la probabilité de trouver l'électron en un point. Un électron confiné dans un atome occupe des volumes particuliers correspondant à des niveaux d'énergie. On peut les comparer aux différents modes de vibrations d'une corde tendue.

Les statistiques Fermi-Dirac

On commençait à maîtriser le comportement des électrons dans un atome, mais il restait à établir leur répartition dans un solide, où l'on fait face à de très grands nombres d'éléments adoptant des structures cristallines.

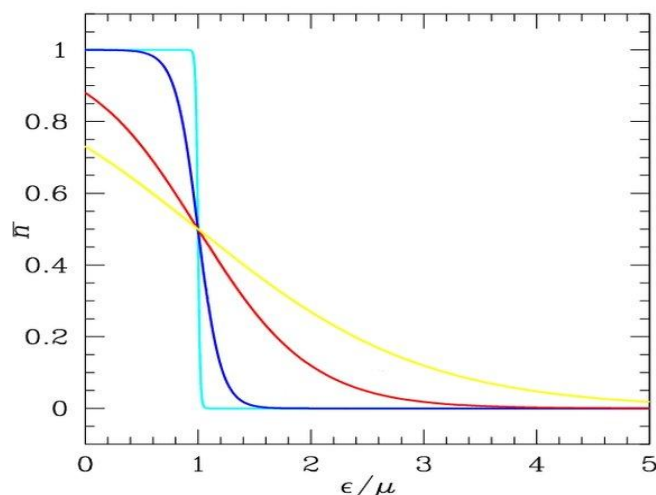


fig2. Représentation graphique des statistiques Fermi-Dirac pour différentes températures avec les niveaux d'énergie en abscisses et les probabilités en ordonnées. En bleu clair, la température absolue tend vers zéro. Les températures sont croissantes jusqu'à la courbe jaune. (Simplification d'un document provenant de Wikipedia (<http://en.wikipedia.org/wiki/Fermi-Dirac>))

¹ Un fermion est une particule qui possède un spin et suit la loi de Pauli (comme l'électron).

² ∇^2 est l'opérateur laplacien, c'est la deuxième dérivée d'une fonction par rapport à plusieurs variables de dimensions (ici x , y et z)

La première étape fut l'avènement de la notion de probabilité par Paul Dirac et Enrico Fermi. Ensemble, ils développèrent en 1926 les statistiques Fermi-Dirac (figure 2), qui donnent la probabilité de trouver des fermions pour chaque niveau d'énergie, à l'équilibre thermique.

La probabilité qu'un niveau d'énergie ε soit occupé à une température absolue T est donnée par : $\bar{n} = \frac{1}{e^{(\varepsilon-\mu)/kT} + 1}$ où μ représente le potentiel chimique, c'est à dire la dérivée de la modification de l'énergie interne du système par la quantité de matière ajoutée, et k est la constante de Boltzman.

Cette théorie se base sur le fait que les fermions suivent le principe d'exclusion de Pauli, et ainsi ne peuvent jamais tous descendre à une énergie minimale (même au zéro absolu). On peut observer qu'il existe une énergie maximale dans un système où la température absolue tend vers zéro, que l'on appelle énergie de Fermi (elle est égale au potentiel chimique correspondant à cette température).

Les bandes d'énergie

D'autre part, deux étudiants, Felix Bloch et Rudolf Peierls, établirent le concept de bandes d'énergie, dans des articles publiés en 1928 et 1929. Un solide est constitué d'un très grands nombres d'atomes. Il existe ainsi une quasi-infinité de niveaux d'énergie différents pour les électrons. Ils sont si nombreux que l'on n'a plus besoin de les séparer, et on peut considérer que les énergies possibles forment un continuum. En revanche, les interférences entre les ondes créent des plages d'énergie où il n'existe aucun niveau possible. Tout fonctionne comme si les électrons occupaient des bandes d'énergies permises alternées à des bandes interdites.

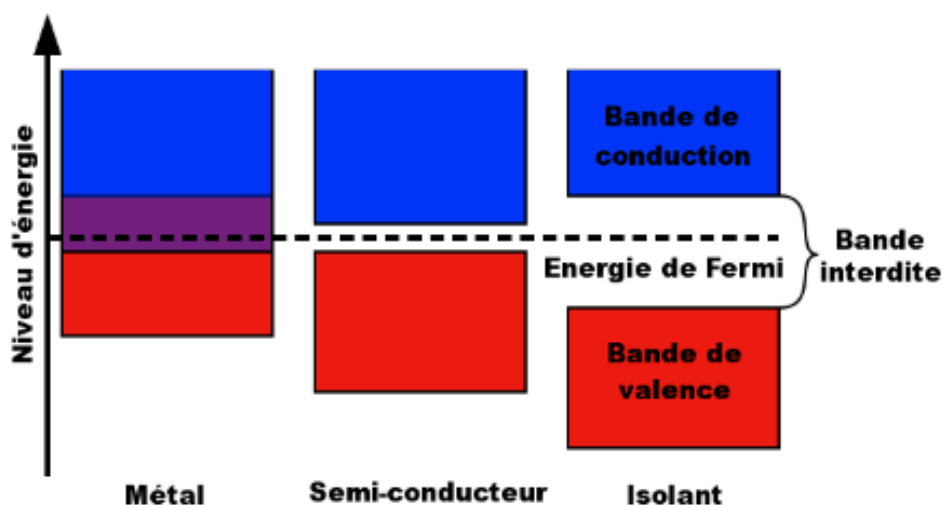


fig 3. Schéma des bandes d'énergie des trois types de matières (Traduction d'un document provenant de Wikipedia, http://en.wikipedia.org/wiki/Energy_bands)

A partir de là, Alan Wilson comprit ce qui différenciait les conducteurs des isolants et des semi-conducteurs (figure 3), et publia ses travaux en 1931. Alors qu'auparavant on pensait qu'il ne s'agissait que d'une question de quantité d'électrons libres (et tous les éléments en posséderaient plus ou moins), Wilson comprit qu'en réalité seule la structure métallique était en cause. Pour les métaux, l'énergie de Fermi se situe à l'intérieur d'une bande d'énergie, ce qui laisse de la place aux électrons de cette dernière bande (la bande de valence) pour bouger. Au contraire, la dernière bande des isolants et des semi-conducteurs est totalement remplie (au zéro absolu). Les électrons doivent donc gagner une énergie relativement grande pour passer sur la bande supérieure (de conduction) pour ensuite être libres de participer à un

courant. Enfin, on se rendit compte que la particularité des semi-conducteurs provient d'une bande interdite relativement mince entre la couche de valence et de conduction.

L'invention du transistor

Longtemps, des dispositifs à semi-conducteurs furent utilisés avant que l'on ne comprenne leur fonctionnement. Les premiers furent des redresseurs de courant¹, qui jouent un rôle essentiel dans le fonctionnement de la radio. Une invention dont le brevet date de 1906, le détecteur à galène fut très populaire jusqu'à qu'il tombât en désuétude dû à l'avènement de la diode à vide (dont on comprenait bien mieux le fonctionnement). Cette technologie ressortit de l'oubli dans les années trente. En effet, on se rendit compte que ces redresseurs à cristaux permettaient de travailler à des fréquences bien plus élevées que leurs homologues à lampes. A l'approche de la guerre, ceci devenait d'autant plus nécessaire que les radars ne fonctionnaient qu'avec des ondes relativement courtes.

Russel Ohl, qui travaillait pour les laboratoires Bell, comprit cela et tâcha de déterminer quels matériaux donnaient les meilleures performances. Il découvrit qu'il s'agissait du silicium et d'autres éléments de la colonne IV (A) du tableau périodique.

Le principal problème qui apparaissait lors de l'usage de ces dispositifs était leur grande versatilité. Ils ne fonctionnaient bien qu'en faisant contact à certains endroits précis. Pour les chercheurs de Bell Labs, ce devait être la conséquence de la présence d'impuretés. Ils entreprirent donc de purifier au maximum des morceaux de silicium. Rapidement, on découvrit qu'il se formait deux sortes de silicium avec des propriétés électriques différentes: certains dans lesquels des charges positives se déplaçaient, les p-types; et d'autres avec des porteurs de charges négatives, les n-types. Encore plus étrange, à leur frontière (la jonction P-N) existait une barrière qui ne laissait passer le courant que dans un sens et qui pouvait produire une tension lorsqu'on l'exposait à de la lumière. On comprit bientôt que des traces infimes d'impuretés restantes étaient responsables de cette différence, les impuretés provenant d'éléments de la colonne III donnaient du p-type alors que ceux provenant d'éléments de la colonne V, du n-type.

Dû à l'analogie avec les lampes, les chercheurs de Bell Labs espéraient réussir à créer un amplificateur avec des semi-conducteurs. En effet, pour transformer une diode, un redresseur à lampe, en une triode qui permet l'amplification du courant, il suffit de rajouter une troisième électrode (la grille) entre les deux premières. Charger la grille sert à modifier la résistance entre la cathode et l'anode, et limite le flux d'électrons. Il suffisait donc en théorie de trouver comment placer ce régulateur dans la jonction P-N. Cette idée se révéla cependant impossible, de par la nature même de cette jonction, mal comprise à l'époque. En revanche, elle ouvrit la voie à d'autres projets.

L'un des concepts était d'appliquer un champ électrique sur la surface afin de libérer et d'attirer des porteurs de charges. Malheureusement, il existe des états de surface pour les électrons, dans lesquels ces derniers se retrouvent bloqués, incapables de participer au courant et "cachant" le champ aux charges inférieures. Il fallait donc faire un contact plus direct avec le semi-conducteur pour pouvoir modifier sa conductivité.

L'équipe de Bell Labs continuait d'avancer dans le noir à coups d'essais et d'erreurs. Ils faisaient des découvertes empiriques qu'ils tâchaient ensuite d'expliquer théoriquement. A force de persévérance, John Bardeen et Walter Brattain, tous deux faisant partie de ce groupe, réussirent enfin, le 16 décembre 1947, à élaborer le premier vrai transistor fonctionnel: le transistor à point de contact (photographie en Annexe 2).

Le transistor

¹ Un redresseur est un élément qui ne laisse passer le courant que dans une direction.

Introduction

Pour illustrer et comprendre le fonctionnement du transistor, il paraît approprié d'observer le transistor de jonction. Bien qu'il ne soit pas le premier historiquement, il reste tout à fait représentatif et sa structure est suffisamment simple pour servir de premier pas dans l'étude de la physique des semi-conducteurs. Bien que de nos jours, il ait perdu beaucoup de sa popularité au profit de technologies plus économiques en énergie (principalement le CMOS, un transistor à oxyde métallique), il fut pendant des décennies le modèle de prédilection dans les circuits intégrés et autres dispositifs d'électronique.

Tout comme le transistor de point de contact, il fut développé dans les laboratoires de Bell. William Shockley inventa ce type de transistor en 1948.

Structure

Le transistor de jonction (NPN) est constitué de trois couches de semi-conducteur (du silicium en général) superposées. Les deux couches externes ont été dopées pour avoir un excès d'électrons libres négatifs (n-type) alors que la couche du milieu a un excès de trous positifs (p-type). Chaque couche est munie d'une électrode. La première électrode (correspondant à la première couche n-type) s'appelle l'émetteur, l'électrode moyenne, la base et la dernière, le collecteur.

Principe

Les semi-conducteurs

Le principe fondamental du transistor repose sur une caractéristique propre aux semi-conducteurs. Ce sont de mauvais conducteurs électrique dont on peut artificiellement faire varier la conduction. Pour comprendre pourquoi, nous devons d'abord comprendre ce qui fait des métaux de bons conducteurs.

Les métaux sont des éléments qui ont peu d'électrons de valence (en général moins de quatre), ils ont ainsi tendance à perdre ces électrons et devenir des cations pour être électroniquement plus stables. C'est ce qui se produit dans un métal solide, où les cations métalliques adoptent des structures cristallines, entourés par des électrons libres qui peuvent très facilement participer à un courant électrique (figure 4).

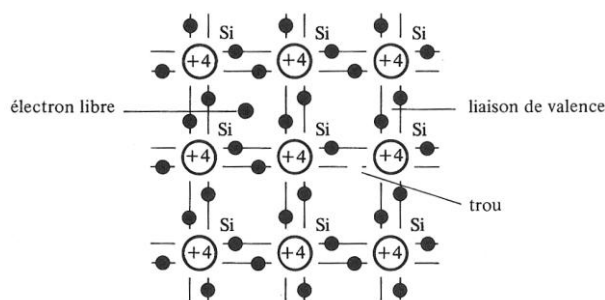


fig 4. Structure cristalline du Silicium. Reproduction d'un document provenant du Traité d'électricité, volume VII (p.6) (voir bibliographie)

En fait, la mécanique quantique nous apprend que les électrons ne peuvent prendre que certaines énergies, occuper uniquement certaines positions dans les couches électroniques des atomes et dans la structure cristalline du métal. Pour qu'un électron quitte l'atome et puisse ainsi conduire l'électricité il doit acquérir une énergie de libération. Nous nous souvenons qu'il peut exister une bande interdite juste au-dessus de la bande de valence qui rend l'énergie de libération importante. Or les métaux n'en possèdent pas, ce qui leur permet, avec l'énergie

thermique répartie aléatoirement dans la matière, de pouvoir libérer facilement de grandes quantités d'électrons.

De leur côté, les semi-conducteurs élémentaires (tels que le silicium et le germanium) se situent sur la colonne IV (A) du tableau périodique, ils ont donc quatre électrons de valence: la moitié de leur dernière couche remplie. La charge positive plus grande de l'atome sous sa couche de valence rend plus difficile aux électrons de se libérer. Cela se traduit par une mince bande interdite et une plus grande énergie nécessaire pour libérer des électrons. De plus, afin de devenir plus stables, ils utilisent leur quatre électrons libres pour former des liaisons covalentes et ainsi obtenir une structure cristalline. De rares électrons acquièrent l'énergie nécessaire pour "sauter" de l'autre côté de la bande interdite. Ils quittent leur liaison et sont libres de transporter le courant. Ils laissent derrière eux une absence d'électron dans la liaison, un "trou" qui peut se déplacer et qui se comporte comme un transporteur de charge positive. Tout repose sur le fait que la bande interdite des semi-conducteurs empêche les électrons de se libérer facilement en temps normal, mais reste suffisamment fine pour qu'un peu d'aide extérieure parvienne à les affranchir.

La chaleur, chez les métaux, nuit à la conductivité: elle augmente l'agitation des particules et ainsi le nombre de chocs entre électrons libres et ions métalliques, ce qui augmente la résistance du métal. Au contraire, l'énergie thermique aide les semi-conducteurs à libérer des électrons supplémentaires, qui avaient juste besoin de cette énergie supplémentaire pour se libérer, et laissant encore une fois des trous derrière eux. Nous observons ainsi que la résistance des semi-conducteurs diminue quand la température augmente.

Pour les mêmes raisons, la lumière peut aussi augmenter la conductivité, en fournissant l'énergie nécessaire aux électrons de valence pour se libérer. Ces propriétés permettent d'utiliser les semi-conducteurs comme appareils sensibles à la chaleur ou la lumière.

Dans le cas du transistor, la propriété qui nous intéresse provient de la présence d'impuretés. On se souvient que l'expérience a appris aux chercheurs de Bell Labs quelles impuretés choisir pour doper les semi-conducteurs et quels effets elles produisent.

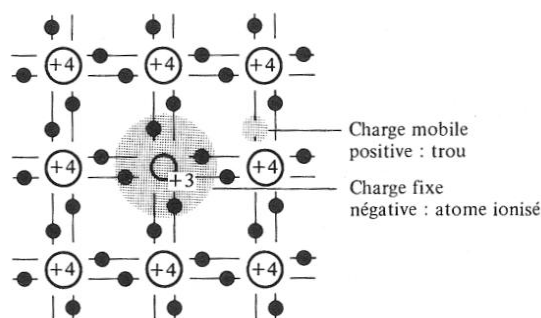


fig 5. Structure cristalline d'un semi-conducteur avec une impureté de la colonne III. Reproduction d'un document provenant du *Traité d'électricité*, volume VII (p.14)

Pour obtenir un p-type, ils faut utiliser des éléments de la colonne III (A) (tels que le bore ou l'aluminium) (figure 5). Ceux-ci ne possèdent que trois électrons de valence. Il leur manque un électron pour former les quatre liaisons nécessaire à la structure cristalline du semi-conducteur. Ainsi, les électrons alentours auront tendance à venir occuper la place vide dans la liaison, laissant un vide derrière eux. On observe donc une augmentation du nombre de trous, alors que celui des électrons libres reste constant. Un p-type possède plus de transporteurs de charges positives que négatives. La mécanique quantique nous explique qu'un élément de la troisième colonne ajoute un niveau d'énergie possible pour les électrons sur la bande interdite, ce qui donne la possibilité à plus d'électrons de partir en créant des trous, mais sans leur permettre de participer au courant eux-mêmes (figure 6).

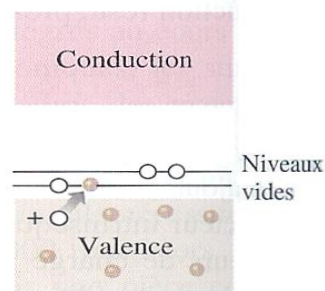


fig 6. Représentation des bandes d'énergie pour un semi-conducteur dopé p-type. Reproduction d'un document provenant de *Physique* par E. Hecht (p.890)

D'autre part, pour obtenir un n-type, il faut doper le semi-conducteur avec des éléments de la colonne V (A) (tels que le phosphore). Ces éléments possèdent cinq électrons de valence, un de trop pour former les liaisons nécessaires. Ils ont donc naturellement tendance à libérer l'électron de trop (sans pour autant créer de trou). Les n-types sont des semi-conducteurs possédant plus d'électrons libres que de trous. D'un point de vue quantique, cela se traduit par ajouter des électrons dont l'énergie se situe sur la bande interdite du semi-conducteur, ceux-ci se libèrent plus facilement que les autres, sans pour autant créer de trous (figures 7 et 8).

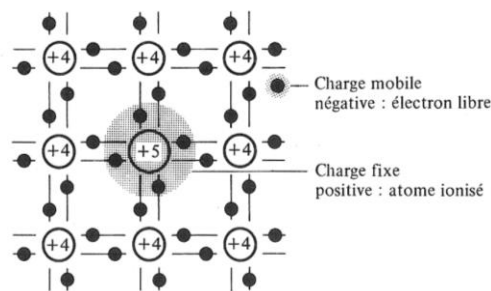


fig 7. Structure cristalline d'un semi-conducteur avec une impureté de la colonne V. Reproduction d'un document provenant du *Traité d'électricité*, volume VII (p.12)

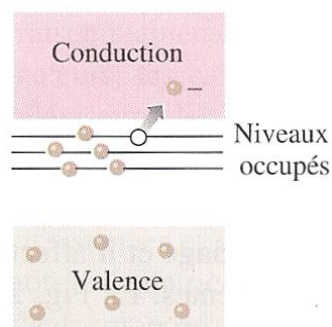


fig 8. Bandes d'énergie d'un semi-conducteur dopé n-type. Reproduction d'un document provenant de *Physique* par E. Hecht (p.890)

La jonction P-N

Considérons un cristal de semi-conducteur dopé de deux manières différentes et possédant une frontière entre ses parties n et p (figure 9). A proximité de la transition, les électrons libres du n-type et les trous du p-type vont s'attirer mutuellement et s'annuler. Il se crée alors

une couche de déplétion dans laquelle il n'existe plus de porteurs de charges. Leur absence provoque l'apparition de charges dans la zone, positives du côté du n-type et négatives de l'autre. Cette différence de potentiel crée une tension qui empêche la zone de se répandre, au point d'équilibre. La couche de déplétion, ne possédant pas de porteurs de charge, se comporte comme un isolant.

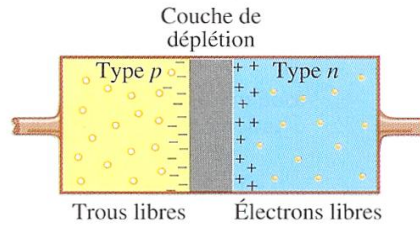


fig 9. Schéma de la jonction P-N. Reproduction d'un document provenant de *Physique* par E. Hecht (p.890)

On se souvient que la jonction P-N permet de redresser le courant. Pour comprendre cela, il suffit d'y appliquer une tension externe. Dans le premier cas, la borne positive est branchée à la partie n et celle négative à la partie p. Les électrons libres et les trous sont attirés vers l'extérieur par leur borne. Les deux parties se vident de leurs porteurs de charge et la couche de déplétion s'élargit. On observe un courant minime.

Dans le deuxième cas, les bornes sont inversées. Elles injectent des porteurs de charge de chaque côté, qui vont compenser les charges au milieu. La couche de déplétion s'amincit et disparaît à partir d'une certaine tension (environ 0.7 V pour le silicium). Le courant augmente à mesure que la zone diminue et devient très grande à sa disparition. On observe qu'ici, le courant n'implique pas des porteurs de charges se déplaçant sur un circuit complet. La plupart des électrons libres s'annulent avec des trous avant de parvenir à la borne opposée, et inversement.

Émetteur, Base et Collecteur

Notre transistor est formé de deux jonctions P-N. C'est la couche de déplétion entre l'émetteur et la base qui va servir de résistance variable. On y applique la tension d'entrée (celle du signal que l'on veut amplifier) avec le potentiel positif à la base et négatif à l'émetteur. La largeur de la zone isolante varie ainsi de manière inversement proportionnelle à cette tension.

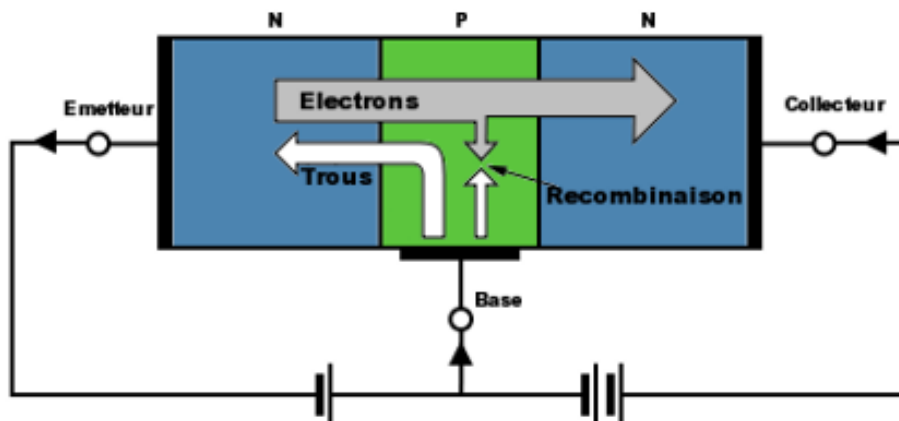


fig 10. Schéma du fonctionnement d'un transistor NPN. Traduction d'un document provenant de Wikipedia (<http://en.wikipedia.org/wiki/Bjt>)

On applique une forte tension fixe entre l'émetteur et le collecteur, avec la borne positive branchée au collecteur et celle négative à l'émetteur. La tension "aspire" les électrons libres du collecteur et élargit ainsi la couche de déplétion entre base et collecteur. De l'autre côté, elle

augmente le flux d'électrons passant de l'émetteur à la base. Il y a ainsi plus d'électrons libres dans la base que ceux que les trous parviennent à annuler. Ce surplus d'électrons se répand, et une partie atteint la jonction avec le collecteur. Dès lors, ces électrons "voient" le fort potentiel positif du collecteur et sont aspirés à travers la couche de déplétion (figure 10).

Le principe nécessite plusieurs éléments. D'abord, la couche de la base est suffisamment mince et le contact avec la base suffisamment petit pour que les électrons libres aient le temps d'arriver jusqu'à la jonction base-collecteur. De plus, la couche émetteur est très dopée pour offrir une bonne conduction, alors que la couche base l'est beaucoup moins afin que le minimum de trous se mettent au travers du courant principal.

En pratique, le transistor agit comme une résistance variant dans un rapport inverse à la tension d'entrée base-émetteur. Nous observons ainsi un courant à la sortie du collecteur proportionnel à cette tension.

Conclusion

Le transistor possède un aspect paradoxal. Il s'agit d'un élément d'électronique de base dont le principe d'utilisation reste relativement simple. Cependant, son invention repose sur le développement de la mécanique quantique, un domaine de la physique des plus complexes. Et malgré cette maîtrise théorique, elle a nécessité beaucoup d'expérimentation et d'essais à l'aveugle.

Une constatation intéressante que j'ai pu faire pendant l'étude de la partie historique est le contraste qui semble exister entre les scientifiques européens et américains. En Europe, les physiciens se complaisent dans la théorie et l'abstrait. Il ne leur est pas nécessaire de poursuivre un but concret. Il vont même jusqu'à discuter de l'éthique de leurs découvertes. Je prends comme exemple les longues discussions entre Einstein et Bohr sur les conséquences philosophiques et religieuses de l'incertitude quantique. En revanche, les chercheurs du Nouveau Monde restent en tout temps pragmatiques. Ce sont les experts de la démarche empirique. Je pense entre autres aux expériences de Clinton Davisson, dont les résultats furent cités comme preuve du comportement d'onde de l'électron, alors même que l'Américain n'était pas au courant de la théorie de de Broglie. Dès qu'ils eurent vent des nouvelles théories quantiques européennes, les physiciens américains s'empressèrent d'y chercher une application pratique. Le fait que le premier transistor fut construit dans le New Jersey n'est vraiment pas anodin.

L'exception Lilienfeld

Je souhaite dédier mon travail à Julius Edgar Lilienfeld, un physicien austro-hongrois qui émigra plus tard aux Etats-Unis pour fuir le nazisme. A partir de 1926, il déposa plusieurs brevets décrivant la structure d'un transistor à effet de champ. Il n'existe pas de preuve qu'il ait jamais construit un tel dispositif. En revanche, le transistor décrit par son brevet fonctionne. Il fut concrétisé plus tard, entre autre, par les chercheurs de Bell Labs. Lilienfeld n'a jamais réussi à intéresser l'industrie, peut-être parce qu'il n'arrivait pas très bien à expliquer le fonctionnement de son invention.

Lilienfeld est un exemple de ce que le monde scientifique peut avoir d'injuste. Bien qu'ayant en premier eu l'idée de l'amplification par un dispositif solide, il ne fut jamais reconnu (contrairement à l'équipe de Bell Labs, qui reçut plusieurs prix Nobel), et son nom n'a jamais atteint l'oreille du grand public.

Bibliographie

Aspects historiques

HODDESON Lillian et RIORDAN Michael, *Crystal Fire: the Invention of the Transistor and the Birth of the Information Age*, New York, Norton, 1998.

SEGRÈ Emilio, *From X-Rays to Quarks: Modern Physicists and Their Discoveries*, New York, Dover, 1980.

Aspects techniques

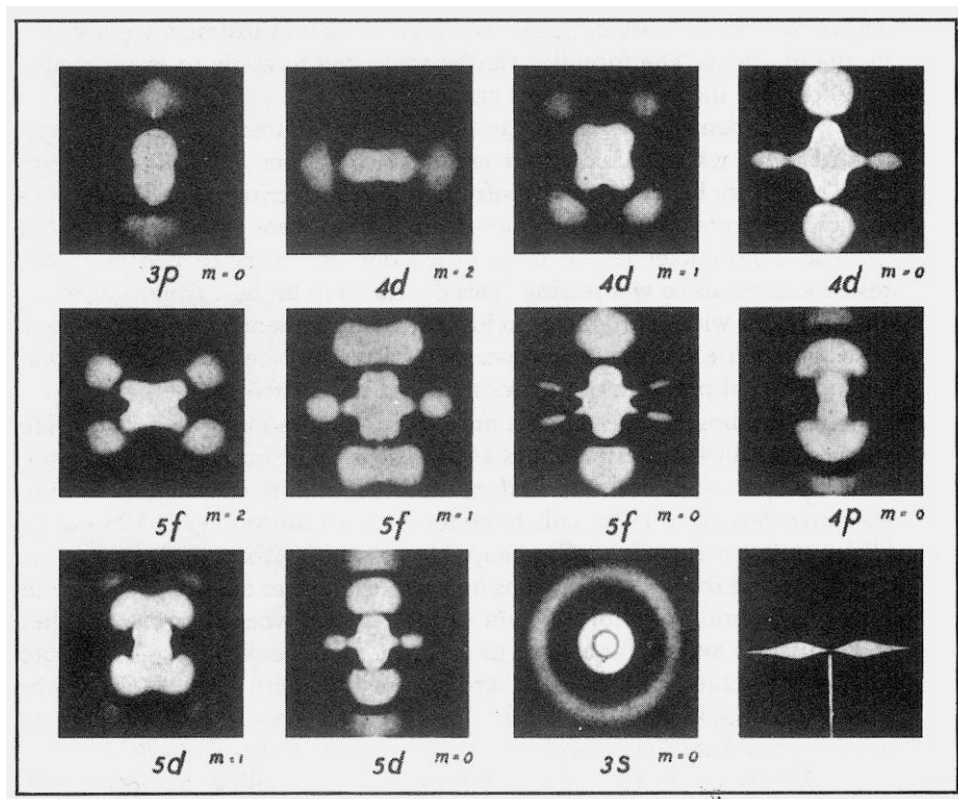
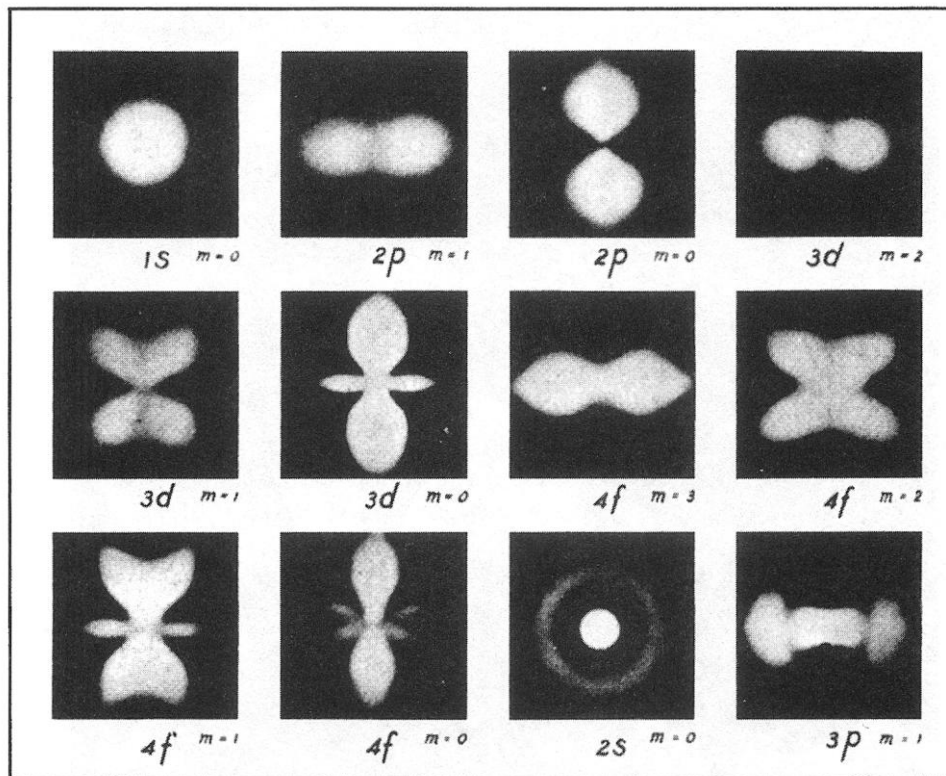
DIU Bernard, LECLERCQ Bénédicte, *La physique mot à mot*. Odile Jacob, 2005.

HECHT Eugene, *Physique*, [s. l.], De Boek, 1999.

NEIRYNCK Jacques, édit., *Traité d'électricité*, volumes II et VII, St-Saphorin, Editions Georgi, 1979.

Annexes

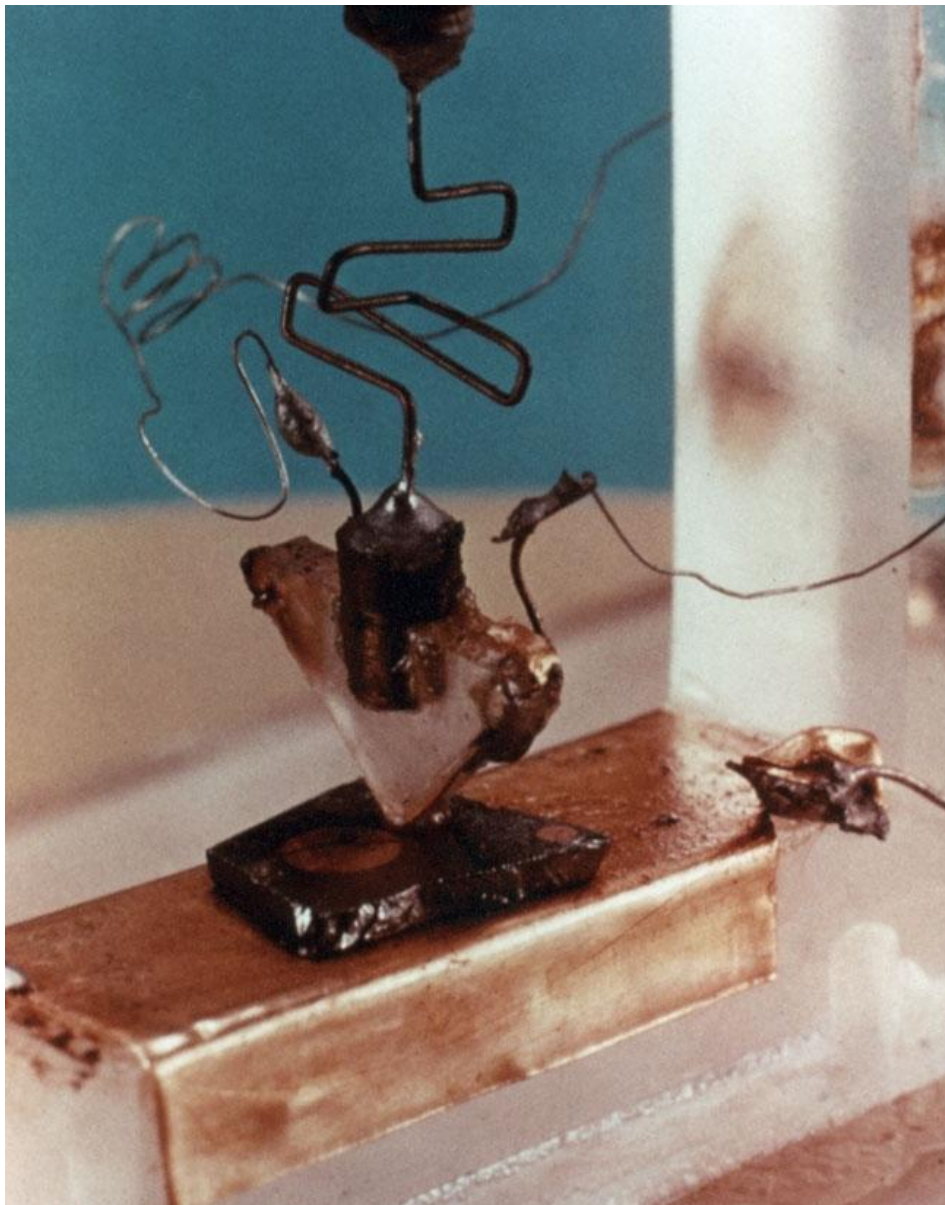
Les orbitales de Schrödinger



Volumes occupés par l'électron de l'hydrogène dans différents états quantiques. Ces derniers sont décrits par trois nombres quantiques: le premier chiffre correspond au nombre quantique principal (le niveau d'énergie); les lettres s , p , d et f décrivent le moment angulaire; et m

donne l'orientation de l'atome. Ces documents sont tirés de *From X-Rays to Quarks* par E. Segrè (pp.162-163).

Le transistor à point de contact



Photographie du premier transistor construit. Document provenant du site internet de l'IEEE (www.ieee.org).