

8. Semicondutores

8.1 – História dos semicondutores

“One shouldn’t work on semiconductors, that is a filthy mess; who knows whether any semiconductors exist!”

A frase é de Wolfgang Pauli, no ano de 1931, em uma carta para Rudolf Peierls. Ela ilustra bem a situação da física de semicondutores naquela época, quando não havia ainda amostras com qualidade suficiente para que resultados reprodutíveis pudessem ser sistematizados dentro de um arcabouço teórico convincente.

É curioso analisar o contraste entre a situação daquela época e a situação atual, em que a física de semicondutores é responsável por alguns dos resultados mais espetaculares da Física, tanto básica quanto aplicada. Podemos citar dois exemplos emblemáticos. Na área de Física Aplicada, a descoberta em 1947 do transistor e o desenvolvimento subsequente dos circuitos integrados - e sua miniaturização crescente - pavimentaram a sociedade da informação em que vivemos hoje em dia. Já na área de Física Básica podemos mencionar os experimentos fundamentais em semicondutores a baixas temperaturas e campos magnéticos que permitiram a descoberta do efeito Hall quântico, por exemplo.

Os estudos das propriedades de condução elétrica dos diversos materiais, desde o século XVIII, já apontavam a existências de “materiais de natureza semicondutora”, como observou Alessandro Volta. Tais materiais apresentavam condutividade intermediária entre os metais (bons condutores) e os isolantes (maus condutores). A propósito, a resistividade (inverso da condutividade) é uma das grandezas físicas cuja variação pode ser medida nas maiores faixas de magnitude, desde $10^{-6} \Omega \cdot \text{cm}$ para os metais típicos até $10^{22} \Omega \cdot \text{cm}$ para os isolantes. Os semicondutores ocupam uma faixa intermediária importante neste espectro, principalmente porque sua condutividade pode ser sintonizada pela adição criteriosa de impurezas (dopantes) e pela temperatura, como veremos neste Capítulo.

Já no século XIX, Michael Faraday percebeu que os semicondutores apresentavam comportamento distinto dos metais no que se refere à dependência da condutividade com a temperatura. Enquanto a condutividade diminui com o aumento da temperatura para os metais, ela aumenta para os semicondutores. Esta talvez seja uma das assinaturas experimentais mais características dos semicondutores, como explicaremos neste Capítulo. A Fig. 8.1 mostra os dados de J. W. Hittorf (1851), que ilustram este resultado para o Ag_2S e o Cu_2S . Note que a escala logarítmica no eixo vertical e o ajuste dos dados por uma reta indica uma lei do tipo $\sigma(T) \propto e^{-\Delta E/k_B T}$, ou seja, um processo termicamente ativado, com barreira de energia ΔE . Entenderemos este resultado na Seção 8.2.

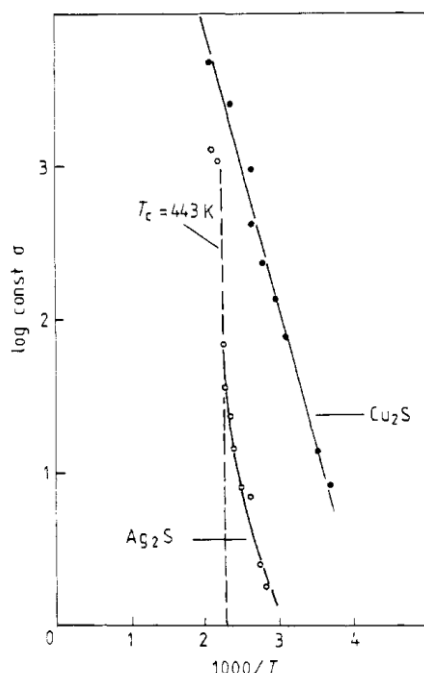


Figura 8.1 – Dados de J. W. Hittorf (1851) indicando condutividade ativada termicamente em semicondutores. **Fonte:** “Early history of the physics and chemistry of semiconductors - from doubts to fact in a hundred years”, G. Busch, Eur. J. of Phys. 10, 254 (1989).

O que provavelmente motivou o comentário de Pauli, com o qual ilustramos o início deste capítulo, foi a enorme quantidade de resultados experimentais de medidas de condutividade, aparentemente contraditórios, que já eram disponíveis na década de 1920. A Fig. 8.2 mostra uma coletânea de dados em 1924, indicando valores de condutividade para o Cu_2O que variam por diversas ordens de magnitude para a mesma temperatura! Como sabemos hoje, isso se deve à presença não controlada de impurezas que, como veremos a seguir, podem modificar de forma drástica a condutividade dos semicondutores.

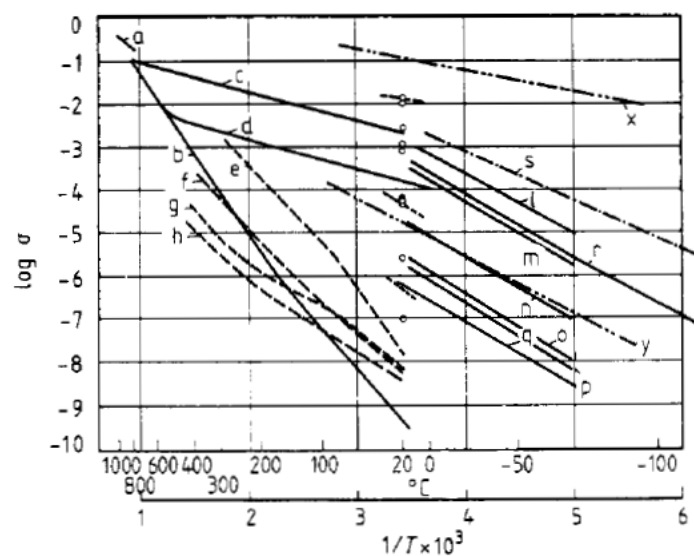


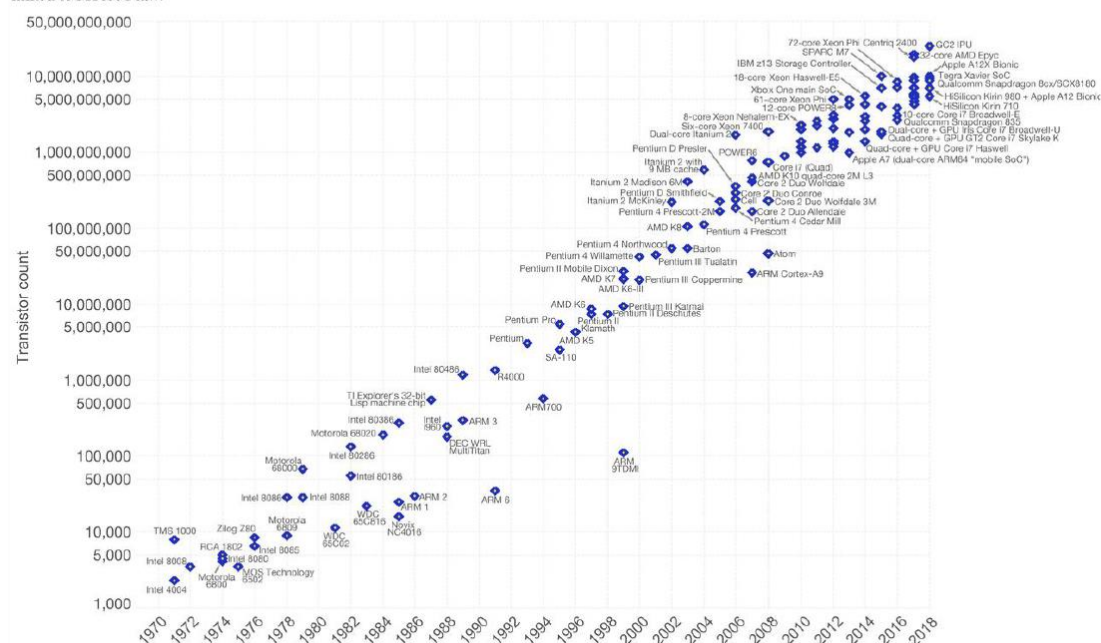
Figura 8.2 – Coletânea de medidas de condutividade em Cu_2O , realizada em 1924. **Fonte:** “Early history of the physics and chemistry of semiconductors - from doubts to fact in a hundred years”, G. Busch, Eur. J. of Phys. 10, 254 (1989).

Então, a partir da década de 1920, o desenvolvimento da Mecânica Quântica começou a exercer um papel fundamental no entendimento dos semicondutores. Notem que, entre o desenvolvimento da equação de Schrödinger (1926), o teorema de Bloch (1928) e o desenvolvimento da teoria de bandas eletrônicas nos cristais por Alan Wilson (1931), pouco tempo se passou. Wilson foi o primeiro a explicar a distinção entre metais e isolantes e também a distinção entre semicondutores intrínsecos (puros) e extrínsecos (com impurezas). Em paralelo a estes desenvolvimentos teóricos, a fabricação de amostras com pureza cada vez mais controlada na década de 1940 pavimentou a chegada do importante ano de 1947, que marcou a descoberta do transistor por Bardeen, Shockley e Brattain, um marco espetacular na física de semicondutores e Prêmio Nobel de Física de 1956.

Em 1958, Joyce e Kilby tiveram a ideia de integrar os diversos elementos de um circuito eletrônico em uma única placa de semicondutor e produziram o primeiro “circuito integrado” ou “*chip*”. Por esta descoberta, ganharam o Prêmio Nobel em 2000. A partir da década de 1960, iniciou-se uma progressiva e impressionante miniaturização destes dispositivos, com o intuito de colocar um número cada vez maior de transistores nos chips, aumentando desta forma o poder de processamento dos computadores.

Moore's Law – The number of transistors on integrated circuit chips (1971-2018)

Moore's law describes the empirical regularity that the number of transistors on integrated circuits doubles approximately every two years. This advancement is important as other aspects of technological progress – such as processing speed or the price of electronic products – are linked to Moore's law.



Data source: Wikipedia (https://en.wikipedia.org/wiki/Transistor_count)
The data visualization is available at OurWorldinData.org. There you find more visualizations and research on this topic.

Licensed under CC-BY-SA by the author Max Roser.

Figura 8.3 – Lei de Moore: o número de transistores em um circuito integrado ao longo dos anos.

A Fig. 8.3 ilustra bem este desenvolvimento, caracterizado pela chamada Lei de Moore. A Lei de Moore não é uma lei da Física, mas, digamos, da tecnologia e da economia. Ela indica, para a indústria de semicondutores, como deve ser a evolução futura na miniaturização dos transistores. Tipicamente, vem sendo seguida há várias décadas a previsão de Moore de que o número de transistores em um *chip* é multiplicado por 2 a cada dois anos. Já houve diversas previsões sobre o fim da Lei de Moore, já que

as dimensões dos padrões geométricos dos transistores estão se aproximando das distâncias interatômicas, quando então não se espera que a arquitetura atual dos transistores funcione mais. No entanto, a indústria tem conseguido, a partir de diversas ideias engenhosas, estender o desenvolvimento do poder computacional previsto pela Lei de Moore até os dias de hoje. Mais recentemente, um dos mecanismos para isso tem sido estender de 2D para 3D a arquitetura dos *chips*.

8.2 – Propriedades eletrônicas

Em relação à sua **estrutura cristalina**, os semicondutores mais tradicionais e de uso mais comum na indústria são aqueles que se cristalizam na *estrutura do diamante* (como silício e germânio) ou *zincblende* (como o GaAs e outros compostos III-V), como já discutimos no Capítulo 3. Desta forma, a Rede de Bravais é *fcc*. A forma predominante de **coesão cristalina** nestes materiais é a ligação covalente, em que cada átomo está no centro de um tetraedro formado por seus quatro primeiros vizinhos, fazendo portanto quatro ligações covalentes com eles. No caso dos compostos III-V, a ligação é também parcialmente iônica, como vimos no Capítulo 2.

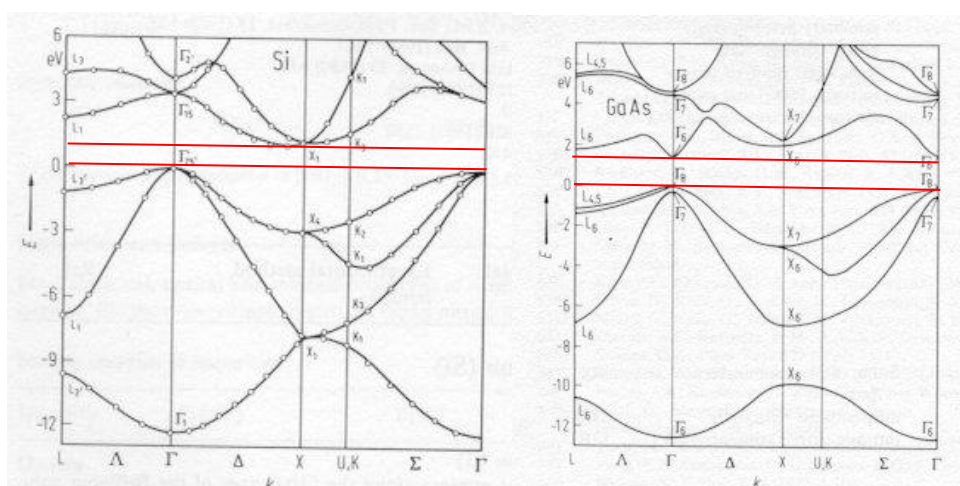


Figura 8.4 – Estrutura de bandas do Si (à esquerda) e do GaAs (à direita). Em vermelho estão indicados os gaps de energia em cada caso.

A Fig. 8.4 mostra a estrutura de bandas do Si e do GaAs. As linhas vermelhas delimitam os gaps de energia em cada caso. Semicondutores são isolantes com o gap pequeno¹, ou seja, as bandas de valência (abaixo do gap) estão totalmente ocupadas por elétrons, enquanto que as bandas de condução (acima do gap) estão totalmente vazias. Desta forma, semicondutores devem ter condutividade nula à temperatura zero. O que os torna interessante para as aplicações em eletrônica é o fato de que a condutividade pode ser sintonizada em uma ampla faixa pelo controle de impurezas e da temperatura, como veremos mais adiante. A Fig. 8.4 também introduz uma distinção importante entre dois tipos de semicondutores, aqueles de **gap direto**, ou seja, onde o topo da banda de valência e o fundo da banda de condução ocorrem no mesmo vetor de onda \mathbf{k} – no exemplo mostrado na Fig. 8.4, o GaAs, isto ocorre no ponto Γ – e aqueles de **gap indireto**, em que o topo da banda de valência e o fundo da banda de condução ocorrem

¹ Não há uma definição precisa quanto a isso, mas materiais com gaps da ordem de 2 eV ou menores são considerados semicondutores.

em pontos **k** distintos (como é o caso do Si). A distinção entre semicondutores de gap direto e indireto é muito importante para o estudo das propriedades óticas, como veremos mais adiante. A Tabela 8.1 mostra o gap de alguns semicondutores, indicando também se o gap é direto ou indireto.

	Material	Direct / Indirect Bandgap	Band Gap Energy at 300 K (eV)
Elements	C (diamond)	Indirect	5.47
	Ge	Indirect	0.66
	Si	Indirect	1.12
	Sn	Direct	0.08
Groups III-V compounds	GaAs	Direct	1.42
	InAs	Direct	0.36
	InSb	Direct	0.17
	GaP	Indirect	2.26
	GaN	Direct	3.36
	InN	Direct	0.70
Groups IV-IV compounds	<i>a</i> -SiC	Indirect	2.99
Groups II-VI compounds	ZnO	Direct	3.35
	CdSe	Direct	1.70
	ZnS	Direct	3.68

Tabela 8.1 – Propriedades do gap de energia de alguns materiais semicondutores.

Vamos recordar a fórmula da condutividade no modelo de Drude, que estudamos no Capítulo 6, $\sigma_D = ne^2\tau/m^*$ (onde m^* é a massa efetiva dos portadores de carga, que podem ser elétrons ou buracos). No caso de metais, a dependência da condutividade com a temperatura era determinada principalmente pelo tempo de relaxação τ , que tipicamente diminui com o aumento da temperatura, causando a diminuição da condutividade com a temperatura que se observa nos metais. No caso dos semicondutores, o principal fator que causa dependência de condutividade com a temperatura não é o tempo de relaxação, mas sim a densidade de portadores n , que poderá ser alterada não apenas pela temperatura, mas também por impurezas.

Vejamos como isto acontece, de forma qualitativa, na Fig. 8.5. A figura ilustra a estrutura de bandas do Si (à direita) e a distribuição de Fermi-Dirac (à esquerda). À temperatura finita, existe probabilidade não-nula de que estados no fundo da banda de condução sejam ocupados por elétrons, que são promovidos termicamente a partir da banda de valência, ali deixando uma certa população de buracos. Tanto os elétrons na banda de condução como os buracos na banda de valência podem conduzir eletricidade, pois temos agora uma situação com duas bandas com preenchimento parcial.

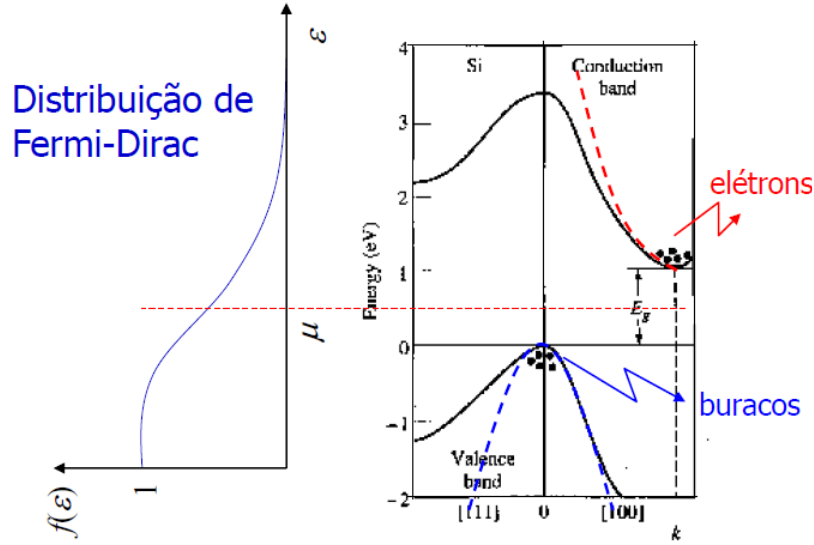


Figura 8.5 – Ilustração esquemática da excitação térmica de portadores de carga (elétrons e buracos) em um semiconductor (silício). O painel à esquerda mostra a distribuição de Fermi-Dirac. O potencial químico μ está posicionado próximo ao centro do gap em um semiconductor intrínseco, como veremos a seguir. As linhas tracejadas vermelha e azul indicam o ajuste do fundo da banda de condução e do topo da banda de valência por funções parabólicas, a partir das quais as massas efetivas podem ser obtidas.

Vamos agora fazer uma análise quantitativa deste fenômeno. Nosso objetivo é obter, como função da temperatura T a densidade de elétrons $n(T)$ na banda de condução e de buracos $p(T)$ na banda de valência. Supondo bandas parabólicas e isotrópicas tanto para valência como para condução, respectivamente com massas efetivas m_b e m_e , podemos estender o resultado que obtivemos no Capítulo 5 para a densidade de estados do elétron livre (Eq. 5.18) para o caso em questão. Desta forma, a densidade de estados para elétrons (por unidade volume) é:

$$D_e(\varepsilon) = \frac{1}{2\pi^2} \left(\frac{2m_e}{\hbar^2} \right)^{3/2} (\varepsilon - E_c)^{1/2} , \quad (8.1)$$

em que E_c é a energia do fundo da banda de condução. De modo semelhante, a densidade de estados para buracos é:

$$D_b(\varepsilon) = \frac{1}{2\pi^2} \left(\frac{2m_b}{\hbar^2} \right)^{3/2} (E_v - \varepsilon)^{1/2} , \quad (8.2)$$

em que E_v é a energia do topo da banda de valência. Assim, a densidade de elétrons na banda de condução é:

$$n(T) = \int_{E_c}^{\infty} D_e(\varepsilon) f_e(\varepsilon) d\varepsilon = 2 \left(\frac{m_e k_B T}{2\pi \hbar^2} \right)^{3/2} e^{(\mu - E_c)/k_B T} . \quad (8.3)$$

Notem a forte dependência (exponencial) da densidade de portadores com a temperatura. Podemos fazer um cálculo análogo a este para o caso dos buracos, lembrando que a distribuição de Fermi-Dirac para buracos é $f_b = 1 - f_e$, chegando no seguinte resultado:

$$p(T) = 2 \left(\frac{m_b k_B T}{2\pi \hbar^2} \right)^{3/2} e^{(E_v - \mu)/k_B T} . \quad (8.4)$$

Combinando (8.3) e (8.4), podemos calcular o produto das densidades de elétrons e buracos em um semiconductor intrínseco:

$$n(T)p(T) = 4 \left(\frac{k_B T}{2\pi\hbar^2} \right)^3 (m_e m_b)^{3/2} e^{-E_g/k_B T}, \quad (8.5)$$

onde $E_g = E_c - E_v$ é a energia do gap. A Eq. (8.5) é conhecida como “lei de ação das massas” e também é válida para semicondutores extrínsecos. Ela mostra que o produto np é independente do potencial químico (que pode ser modificado pela dopagem química ou eletrostática).

Em um semiconductor intrínseco, as densidades de elétrons e buracos são iguais ($n=p$), pois necessariamente a excitação térmica de um elétron para banda de condução deixa um buraco na banda de valência. Desta forma,

$$n(T) = p(T) = \sqrt{n(T)p(T)} = 2 \left(\frac{k_B T}{2\pi\hbar^2} \right)^{3/2} (m_e m_b)^{3/4} e^{-E_g/2k_B T}. \quad (8.6)$$

A Eq. (8.5) explica os resultados mostrados nas Figuras 8.1 e 8.2, em que densidade de portadores de carga (e, portanto, a condutividade) apresenta comportamento termicamente ativado. A Figura 8.6 também ilustra este comportamento. Note que, quanto maior a energia do gap, mais rápida é a queda da densidade de portadores com a diminuição da temperatura.

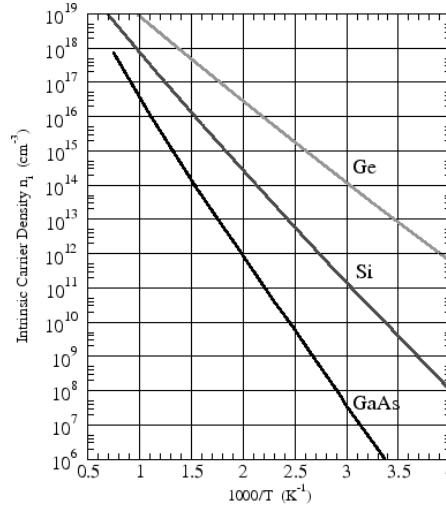


Figura 8.6 – Densidade de portadores em semicondutores intrínsecos, como função da temperatura. Os gaps do Ge, Si e GaAs são, respectivamente, 0,74 eV, 1,17 eV e 1,52 eV.

Impondo a igualdade entre n e p para semicondutores intrínsecos e usando as Eqs. (8.3) e (8.4), podemos obter o potencial químico como função da temperatura. Depois de alguma álgebra, obtemos:

$$\mu(T) = E_v + \frac{1}{2} E_g + \frac{3}{4} k_B T \ln \left(\frac{m_b}{m_e} \right). \quad (8.7)$$

Este resultado mostra que, para baixas temperaturas, o potencial químico permanece sempre em torno do centro do gap, como ilustrado na Fig. 8.5. De fato, ele apenas se

desvia substancialmente do centro do gap para temperaturas mais altas se as massas efetivas de elétron e buraco forem substancialmente distintas.

8.3 – Propriedades óticas e éxcitons

Em um semiconductor de gap direto, as propriedades óticas são claramente determinadas pela magnitude do gap. Em uma visão simplificada, como mostra o painel à esquerda da Fig. 8.7, um semiconductor será transparente para fótons com energia menor que a do gap, e apenas para fótons com energia maior do que esta haverá absorção de luz, a partir da promoção de elétrons da banda de valência para a banda de condução. O painel à direita da mesma figura mostra o coeficiente de absorção de luz do GaAs para várias temperaturas. De fato, como esperado, a absorção aumenta abruptamente a partir de um certo valor de energia, que identificamos como sendo aproximadamente a energia do gap (note que o gap aumenta com a redução da temperatura, um efeito que não pretendemos descrever aqui).

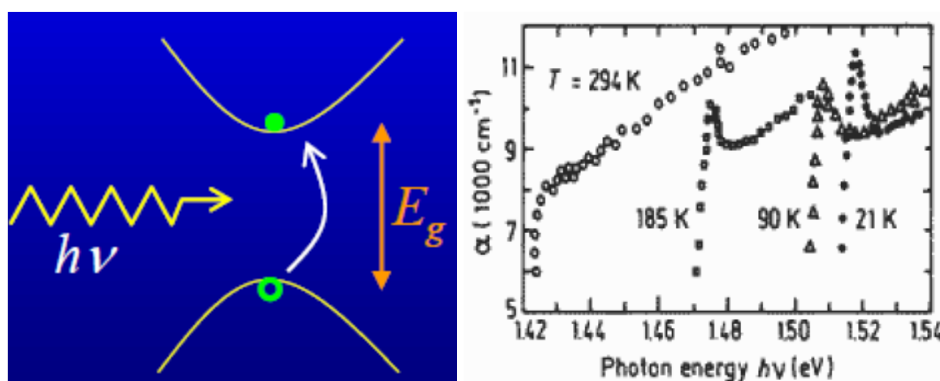


Figura 8.7 – À esquerda, indicamos a criação de um par elétron-buraco a partir da absorção de um fóton. À direita, mostramos o coeficiente de absorção do GaAs para diferentes valores de temperatura. O pico estreito que surge a baixas temperaturas é devido aos éxcitons.

A transição ótica ilustrada na Fig. 8.7 é conhecida como “transição banda-a-banda” (da banda de valência para a banda de condução) e é, para todos os efeitos práticos, “vertical”, ou seja o momento cristalino ou vetor de onda do elétron não é alterado de forma substancial. De fato, ocorre conservação do momento cristalino neste processo, de modo que o momento linear do fóton absorvido é transmitido para o momento cristalino do elétron, o que implica em que $\mathbf{k}_{\text{fóton}} = \Delta\mathbf{k}_e$. No entanto, para valores típicos da energia do fóton (~ 1 eV), seu vetor de onda é da ordem de $k_{\text{fóton}} \approx 10^6$ m⁻¹. Por outro lado, as dimensões típicas da 1ª Zona de Brillouin são da ordem do inverso do parâmetro de rede, ou seja, $\Delta k_{\text{ZB}} \approx 1/a \approx 10^{10}$ m⁻¹. Desta forma, conclui-se que a variação do vetor de onda do elétron é desprezível se comparada às dimensões da 1ª Zona de Brillouin, ou seja, podemos considerar que a transição ótica é praticamente “vertical”. Isto ocorre tanto para absorção como para emissão de luz.

O que deve ocorrer em um material de gap indireto? Isto está ilustrado esquematicamente na Fig. 8.8. Considere a situação ilustrada na figura, ou seja, o material tem um “gap mínimo” E_g , enquanto que a diferença de energias entre as bandas de valência e condução para o vetor de onda correspondente ao topo da banda de valência é E_0 , com $E_0 > E_g$. Pelo argumento desenvolvido acima, só haverá absorção

substancial de luz para energias do fóton maiores que E_0 . Isto, de fato, é o que ocorre, como ilustrado na figura. No entanto, é possível ocorrer uma absorção de luz menos intensa para energias menores que E_0 , devido a processos que envolvem a absorção ou emissão de um fônon (conjuntamente com a absorção do fóton). Os fônons, ao contrário dos fótons, podem carregar bastante momento linear (eles “existem” na mesma Zona de Brillouin dos elétrons), mas, comparativamente com o fóton, têm pouca energia (essencialmente porque a velocidade do som é muito menor que a velocidade da luz). Processos de absorção indireta, como indicados na figura, que envolvem um fóton e um fônon, ocorrem com menor probabilidade, e portanto a intensidade de absorção associada a eles é menor. Com relação à emissão de luz, algo semelhante ocorre. Depois de excitados, os elétrons tipicamente relaxam não-radiativamente para o mínimo da banda de condução, de onde poderão emitir luz ao recombinar radiativamente com um buraco da banda de valência. No entanto, para satisfazer a conservação do momento, elétron e buraco devem ter o mesmo vetor de onda, o que não ocorrerá com facilidade em um material de gap indireto. Assim, os materiais semicondutores com boa eficiência de emissão de luz, e portanto com maior uso em optoeletrônica, são apenas aqueles com gap direto.



Figura 8.8 – À esquerda, um diagrama de bandas esquemático para um material de gap indireto, ilustrando os processos de absorção indireta, que ocorrem por intermédio da absorção ou emissão de um fônon. À direita, o espectro de absorção correspondente. A absorção é mais intensa pra fótons com energia maior que E_0 , mas também ocorre de forma residual devido aos processos indiretos, com energias aproximadamente entre E_0 e E_g (desprezando a energia do fônon, que também deve ser levada em conta na análise conservação de energia).

Um olhar mais atento ao espectro de absorção mostrado na Fig. 8.7, revela que, à medida em que a temperatura diminui, começa a surgir um pico muito intenso para energias um pouco abaixo da energia do gap. Esta é assinatura experimental dos **éxcitons**. Como argumentamos, a E_g é a energia mínima para uma transição banda-a-banda, ou seja, para promover um elétron da banda de valência para a banda de condução. Nesta situação, tanto os estados de valência como o de condução são estados de Bloch, deslocalizados espacialmente, de modo que não há nenhuma correlação espacial entre as coordenadas do elétron e do buraco. No entanto, reduz-se a energia necessária para a transição ótica se considerarmos a possibilidade de que elétron e buraco, que são “quase-partículas” de carga oposta, possam se atrair de forma

eletrostática e formar um par ligado, de forma análoga a um átomo de hidrogênio². A este par elétron-buraco correlacionado, chama-se éxciton.

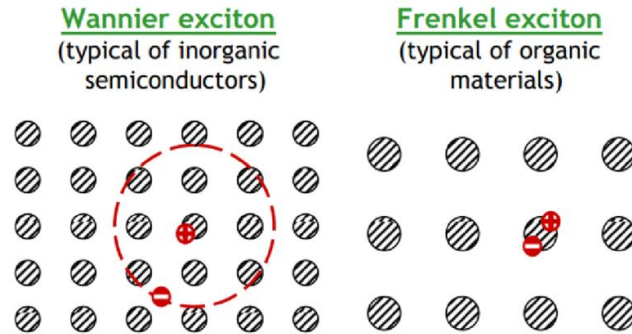


Figura 8.9 – Dois modelos para éxcitons, em duas situações extremas: éxcitons de Mott-Wannier (mais estendidos) e éxcitons de Frenkel (mais localizados).

Os dois modelos teóricos mais comuns para os éxcitons estão ilustrados na Fig. 8.9, e se referem a duas situações extremas da localização espacial relativa entre elétron e buraco. Se o par elétron-buraco é pouco localizado, de modo que a distância média entre eles é muito maior que o parâmetro de rede, considera-se o modelo de **Mott-Wannier**. No extremo oposto, em que o par é bastante localizado, como se fosse uma excitação atômica que pode viajar pelo cristal através de “saltos”, temos os éxcitons de **Frenkel**.

Vamos analisar em mais detalhe o modelo de Mott-Wannier, que é mais adequado para os semicondutores inorgânicos (mais tradicionais) e no qual a analogia com o átomo de hidrogênio pode ser utilizada. Lembrando dos níveis de energia do átomo de H (no S.I.):

$$E_n = -\frac{me^4}{2(4\pi\epsilon_0)^2\hbar^2n^2} = -\frac{13,6 \text{ eV}}{n^2} , \quad (8.8)$$

onde n é o número quântico principal e m é a massa do elétron livre. Para nossa discussão subsequente, será também útil analisar a expressão para o **raio de Bohr**:

$$a_0 = \frac{4\pi\epsilon_0\hbar^2}{me^2} = 0,529 \text{ Å} , \quad (8.9)$$

que mede o raio típico do átomo de H no estado fundamental. Para adaptarmos as expressões (8.8) e (8.9) para o problema excitônico, temos que fazer duas modificações. A primeira envolve substituir a massa do elétron pela massa reduzida μ do par elétron-buraco ($\mu^{-1} = m_e^{-1} + m_b^{-1}$), onde m_e e m_b são as massas efetivas da banda de condução e valência, respectivamente. A segunda modificação envolve blindar a interação eletrostática pela constante dielétrica macroscópica ϵ (em unidades de ϵ_0) do material. Com estas duas modificações, a energia do estado fundamental do éxciton (em relação ao elétron e buraco livres) é:

$$E_{exc} = -\frac{\mu}{m} \frac{1}{\epsilon^2} \times (13,6 \text{ eV}) . \quad (8.10)$$

² A analogia mais precisa seria em relação ao positrônio, o par ligado entre um elétron e sua anti-partícula, o pósitron.

O negativo deste valor é a energia de ligação do éxciton, ou seja, a energia necessária para dissociar o par elétron-buraco. Já o raio do éxciton torna-se:

$$a_{exc} = \frac{m}{\mu} \epsilon \times (0,529 \text{ \AA}) \quad (8.11)$$

A Tabela 8.2 mostra valores de massa efetiva e constante dielétrica para alguns semicondutores. Tipicamente, $\epsilon \sim 10$ e as massas efetivas são menores do que a massa do elétron por um fator entre 0,01 e 0,1. Isto faz com que as energias de ligação de éxcitons sejam na faixa de alguns meV, ou seja, muito menores do que a do átomo de H e comparáveis com a energia térmica à temperatura ambiente. É por este motivo que os picos excitônicos no espectro de absorção dos semicondutores são mais perceptíveis apenas a temperaturas baixas. Com relação ao raio do éxciton, a partir da equação (8.11) obtemos valores na ordem de 10^2 \AA , de fato muito maiores do que as distâncias interatômicas típicas, o que justifica (*a posteriori*) a abordagem de Mott-Wannier (incluindo o uso da constante dielétrica macroscópica do material para blindar a interação eletrostática).

8.4 – Impurezas doadoras e aceitadoras

Possivelmente, a característica mais importante dos semicondutores, no que diz respeito ao seu uso nas mais diversas aplicações, seja a possibilidade de dopá-los, ou seja, alterar sua condutividade de maneira drástica pela adição controlada de uma pequena quantidade de **impurezas**. Impurezas substitucionais são um tipo de “defeito” na estrutura cristalina em que um átomo do cristal é substituído por outro. No caso dos semicondutores, como veremos a seguir, o caso mais interessante ocorre quando o átomo substituinte tem valência distinta, ou seja, pertence a outra coluna da Tabela Periódica (contendo portanto um número distinto de elétrons de valência).

A adição de dopantes altera de maneira drástica a condutividade dos semicondutores. Por exemplo, adicionando-se apenas 1 átomo de B (grupo III) para cada 10^5 átomos de Si (grupo IV), ou seja, concentração atômica de 0,001%, é possível aumentar a condutividade do Si a temperatura ambiente por um fator 1000! A Fig. 8.10 ilustra este efeito de forma quantitativa: podemos observar as curvas de resistividade como função da temperatura em Ge (grupo IV) dopado com Sb (grupo V), para várias concentrações da impureza. Note a enorme faixa de variação da resistividade que pode ser obtida a partir da dopagem.

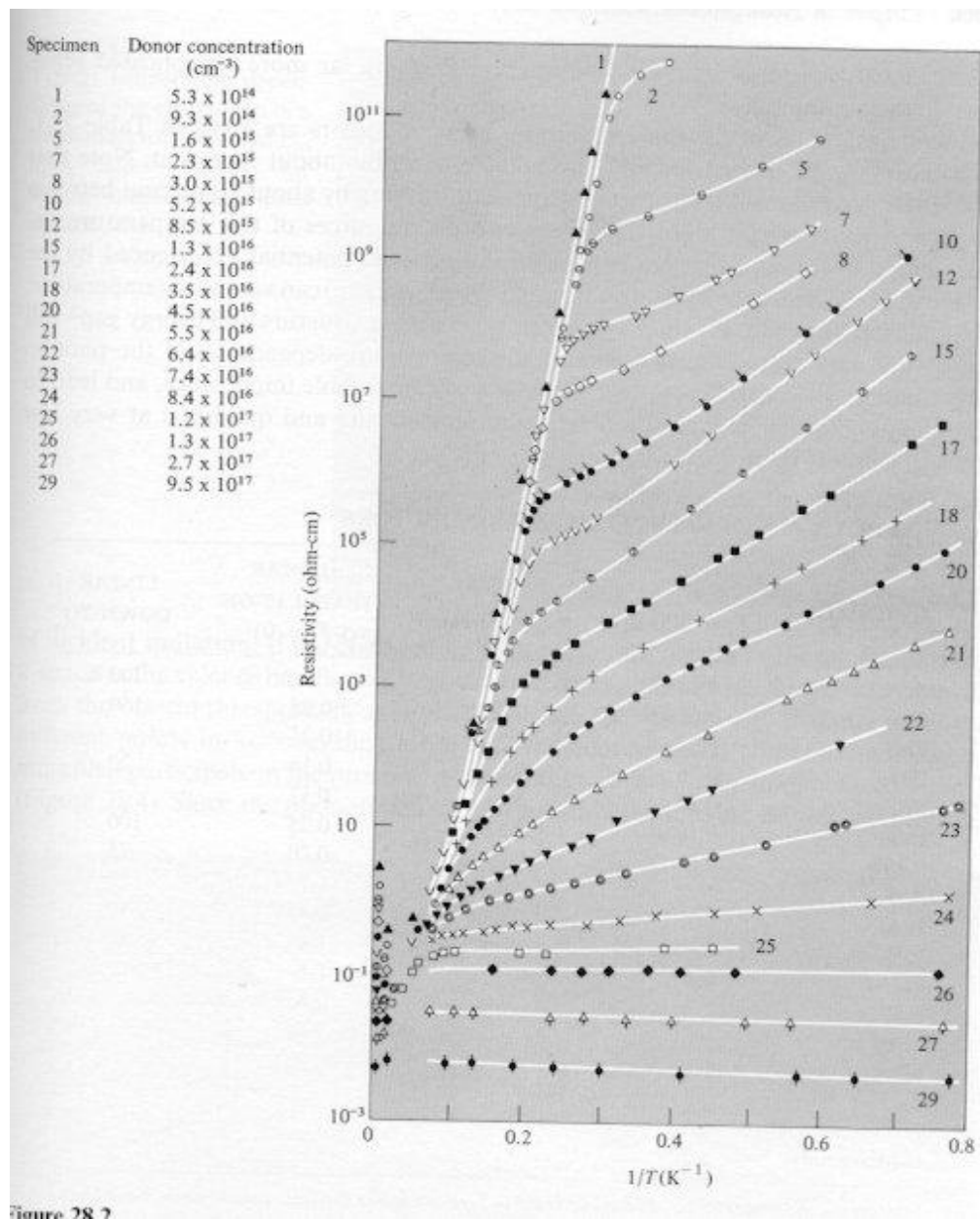


Figura 8.10 – Resistividade em função da temperatura em Ge dopado com Sb, para várias concentrações de impureza. **Fonte:** Ashcroft e Mermin.

Vamos entender como isto funciona de maneira mais detalhada. A Fig. 8.11 ilustra de forma esquemática o que acontece quando substituímos um átomo de Si por um de As (grupo V, à esquerda) ou de B (grupo III, à direita). No caso da impureza de As, sabemos que o átomo substituinte tem 5 elétrons de valência, ao contrário do Si, que tem 4. Os 4 elétrons de valência do Si são usados para formar as 4 ligações covalentes com seus vizinhos. Ao substituir o átomo central pelo As, as quatro ligações poderão ser feitas (por 4 elétrons que ficarão em estados ligantes na banda de valência) e ainda sobrar um elétron, que em princípio poderia ser promovido para a banda de condução e contribuir para a condutividade. No entanto, ao substituirmos um átomo do grupo IV por um do grupo V, não estamos apenas adicionando um elétron a mais, mas também um próton a mais (lembre-se que o cristal permanece neutro). Este próton extra funciona como um centro atrativo para o elétron extra, exercendo um potencial Coulombiano sobre ele. O sistema é (mais uma vez!) análogo ao átomo de H, e por isso

este tipo de impureza é também chamada de “hidrogenóide”. Desta forma, a baixas temperaturas, o elétron permanece preso à impureza mas, como veremos a seguir, a energia de ligação é pequena, de modo que mesmo à temperatura ambiente a energia térmica será suficiente para “ionizar” a impureza e “doar” o elétron para a banda de condução, onde poderá participar dos processos de condução elétrica. Por este motivo, este tipo de impureza também é chamado de “impureza doadora”. A energia de ligação da impureza doadora está indicada esquematicamente como E_D no painel central da Fig. 8.11.

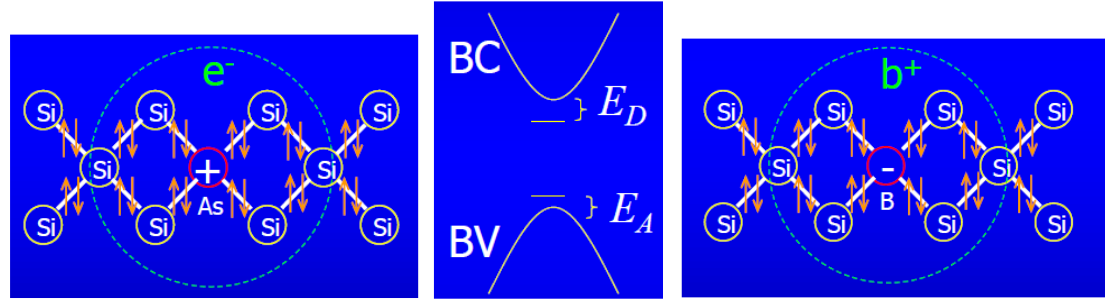


Figura 8.11 – Visão esquemática das impurezas doadoras (esquerda) e aceitadores (direita). Note que desenhamos a estrutura do Si no plano apenas para facilitar a visualização. Como sabemos, a estrutura correta é tridimensional.

O caso do dopante do grupo III (à direita na Fig. 8.11) é análogo. Há um elétron a menos e um próton a menos. A ausência do elétron pode ser vista como um buraco (temos que retirar ou “aceitar” um elétron da banda de valência) que está preso pela interação Coulombiana ao centro de carga negativa no átomo de B. Estas são as “impurezas aceitadoras” e seus níveis de energia ficam um pouco acima da banda de valência, como indicado na Fig. 8.11 (E_A).

Energia de ligação (meV) (doadores)				
	P	As	Sb	Teoria
Si	45	49	39	30
Ge	12,0	12,7	9,6	9,1

Tabela 8.3 – Energia de ligação do elétron em impurezas doadoras. As três primeiras colunas mostram valores experimentais, enquanto que a última coluna mostra o resultado da teoria de massa efetiva.

Para calcularmos a energia de ligação do elétron, podemos proceder de forma muito semelhante ao que fizemos no caso do éxciton, ou seja, por analogia com o átomo de H e fazendo as modificações necessárias nas equações (8.8) e (8.9). As fórmulas ficam quase idênticas às (8.10) e (8.11) e não vamos repeti-las aqui. A única diferença é que temos que usar as massas efetivas de elétron ou buraco (dependendo se estamos analisando doadores ou aceitadores) em lugar da massa reduzida do éxciton. Este tipo de abordagem é também conhecido como **teoria de massa efetiva** para impurezas. A Tabela 8.3 mostra as energias de ligação de alguns doadores (P, As e Sb) em cristais de Si e Ge. Note que este modelo teórico descrito aqui é bastante simples e não leva em

conta a natureza química específica da impureza, enquanto que as medidas experimentais mostram variações neste sentido. Obviamente, existem modelos mais sofisticados, mas vão além das necessidades deste curso. Reparem ainda os valores pequenos das energias de ligação, o que faz com que estes estados eletrônicos sejam também denominados “rasos” (pois ficam próximos das bandas de valência ou condução).

Assim como nos casos dos éxcitons, o raio de Bohr destes “átomos de hidrogênio” efetivos estão na faixa de dezenas ou centenas de angstroms. A Fig. 8.12 mostra o resultado de um cálculo “tight-binding” da densidade eletrônica ($|\psi|^2$) associada à função de onda do estado ligado à impureza. Note que o elétron se estende por distâncias muito maiores que o parâmetro de rede, mas, superimposta à função envelope aproximadamente exponencial que descreve o elétron, notam-se as oscilações de curto comprimento de onda associadas à parte periódica das funções de Bloch.

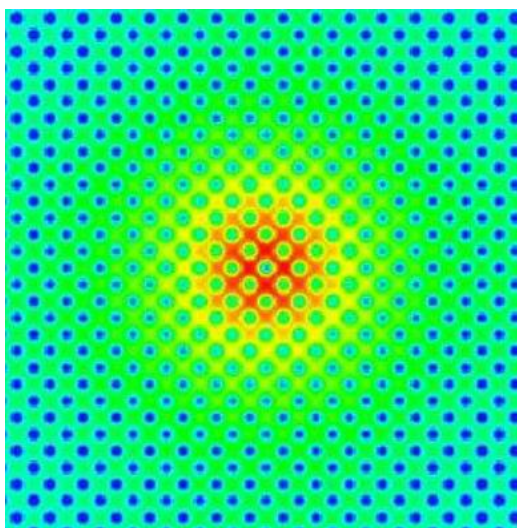


Figura 8.12 – Função de onda ao quadrado do elétron ligado a uma impureza doadora em um cristal da GaAs. Trata-se de um cálculo “tight-binding” com 10^6 átomos.

8.5 – Junção pn e dispositivos semicondutores

Como dissemos, a proximidade em energia entre o fundo da banda de condução e os níveis eletrônicos doadores faz com que um semicondutor com uma certa concentração de impurezas doadoras tenha uma densidade n de elétrons na banda de condução que pode ser substancial. Isto pode ser entendido de forma qualitativa simplesmente considerando a Eq. (8.3) na condição em que o potencial químico se posiciona aproximadamente no nível eletrônico da impureza, ou seja, a poucos meV do fundo da banda de condução. A um semicondutor dopado desta maneira chama-se **semicondutor do tipo-n**. É claro que, a uma certa temperatura não nula, haverá também, por excitação térmica, buracos na banda de valência (como havia no caso dos semicondutores intrínsecos que analisamos anteriormente), mas eles serão em muito menor número do que os elétrons, e por isso a eles denominamos **portadores minoritários**. Tudo ocorre de forma análoga para as impurezas aceitadoras, que dão origem aos **semicondutores do tipo-p**, em que há uma densidade p substancial de buracos e uma pequena densidade minoritária de elétrons.

Nesta Seção, vamos tentar explicar de forma muito breve o funcionamento de alguns dispositivos semicondutores, analisando com um pouco mais de detalhamento o mais simples deles: o **diodo** ou **junção pn**. Vamos analisar o que acontece quando colocamos em contato um semiconductor tipo-*p* com um outro semiconductor tipo *n*, como ilustrado na Fig. 8.13. Na região *n* existe uma densidade significativa de elétrons (bolinhas vermelhas) e na região *p* existe uma densidade significativa de buracos (círculos amarelos). Quando colocamos em contato os dois semicondutores, haverá uma tendência de difusão de elétrons para a região *p* e de buracos para a região *n*, gerando acúmulo de carga positiva na região *n* e carga negativa na região *p*. Esta transferência de cargas é restrita a uma região próxima à interface conhecida como **camada de depleção**. Surge então um campo elétrico na camada de depleção que impede que a difusão ocorra indefinidamente campo elétrico. Associado a este campo elétrico, há uma rampa de potencial na camada de depleção, também mostrada na figura. A espessura da camada de depleção pode variar bastante (entre 10 e 1000 nm).

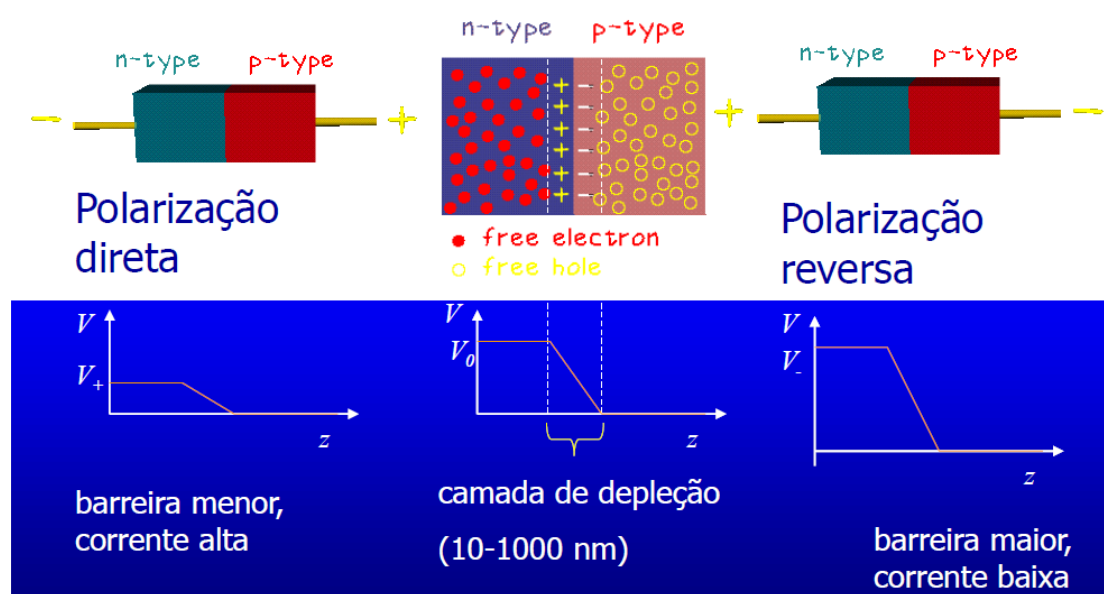


Figura 8.13 – A junção pn sem d.d.p. (centro), com polarização direta (esquerda) e com polarização reversa (direita).

Quando não há d.d.p. aplicada à junção, obviamente não circula corrente elétrica. Esta situação de corrente nula é causada pelo cancelamento ou equilíbrio entre duas contribuições para a corrente. Vamos considerar apenas a componente eletrônica destas duas contribuições, tudo se dá de forma análoga para buracos. Repare que a energia potencial para elétrons (carga negativa) tem o perfil inverso ao potencial eletrostático mostrado na Fig. 8.13:

- (1) **Corrente de recombinação:** alguns elétrons na região *n* chegam à camada de depleção com energia suficiente para vencer a barreira de potencial e chegar à região *p*. A probabilidade que isto aconteça é dada pelo fator de Boltzmann, de modo que a corrente de recombinação para eletrons J_{nr} também será proporcional ao mesmo fator: $J_{nr}(0) \propto e^{-eV_0/k_B T}$. Esta corrente representa um fluxo de elétrons da esquerda para a direita na Fig. 8.13 e, portanto, uma corrente elétrica da direita para a esquerda.
- (2) **Corrente de geração:** na região *p* não há apenas buracos, mas também elétrons (portadores minoritários). Se alguns desses elétrons chegam à

camada de depleção, são imediatamente acelerados pela rampa de potencial para a região n . Note que, ao contrário da corrente de recombinação, a corrente de geração J_{ng} não depende da altura da rampa de potencial V . A corrente de geração representa um fluxo de elétrons da direita para a esquerda na Fig. 8.13 e, portanto, uma corrente elétrica da esquerda para a direita.

Como argumentamos, se não houver d.d.p. aplicada, a corrente de geração deve ser igual à corrente de recombinação (em módulo), ou seja:

$$J_{nr}(0) = J_{ng}(0) , \quad (8.12)$$

onde estamos considerando a situação em que a d.d.p. é nula $\Delta V = 0$.

Vamos considerar agora o caso de uma d.d.p. aplicada. Por convenção, vamos considerar $\Delta V > 0$ a situação de **polarização direta**, em que a rampa de potencial na camada de depleção é reduzida de V_0 para $V_+ = V_0 - \Delta V$ (veja a Fig. 8.13) e $\Delta V < 0$ a situação de **polarização reversa**, em que a rampa de potencial é aumentada para $V_- = V_0 - \Delta V$. Como vimos, a corrente de geração é independente da voltagem, de modo que $J_{ng}(\Delta V) = J_{ng}(0)$. Por outro lado, a corrente de recombinação é modificada:

$$J_{nr}(\Delta V) \propto e^{-e(V_0 - \Delta V)/k_B T} \Rightarrow J_{nr}(\Delta V) = J_{nr}(0)e^{e\Delta V/k_B T} = J_{ng}(0)e^{e\Delta V/k_B T} , \quad (8.13)$$

onde usamos a Eq. (8.12). A corrente total será:

$$J_{tot}(\Delta V) = J_{nr}(\Delta V) - J_{ng}(\Delta V) = J_{nr}(0)[e^{e\Delta V/k_B T} - 1] . \quad (8.14)$$

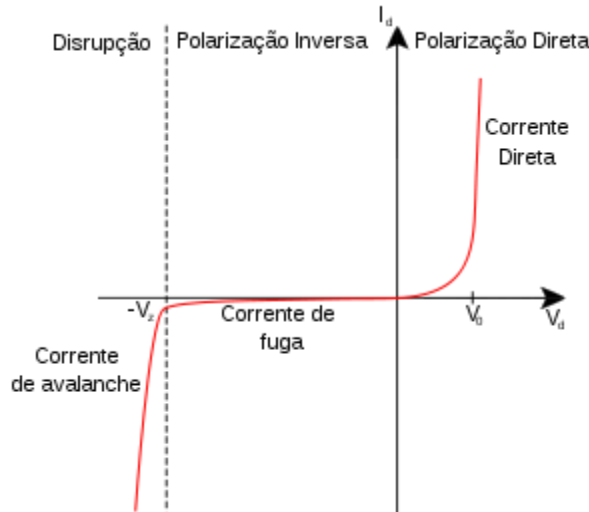


Figura 8.14 – Curva característica I-V da junção pn ou diodo. Na polarização reversa, a partir de uma voltagem alta, ocorre a disrupção do diodo (“breakdown”), que não chegamos a discutir neste curso.

A Eq. (8.14) é a equação de Shockley, ou equação do diodo ideal. A Fig. 8.14 mostra a curva característica I-V da junção pn ou diodo. Note que a junção atua como um retificador de corrente: o diodo deixa a corrente passar com facilidade na polarização direta e praticamente não deixa a corrente passar na polarização reversa.

A junção pn é o ponto de partida para o entendimento de outros dispositivos semicondutores. Vamos discutir brevemente alguns deles. A Fig. 8.15 (esquerda) mostra como funciona o **diodo emissor de luz** ou **LED**. De forma simplificada, é a mesma estrutura da junção pn (operando em polarização direta), utilizando-se um material de gap direto. Desta maneira, elétrons e buracos são levados à camada de depleção e aí ocorre a recombinação com a emissão de luz. O **laser** de semiconductor opera de forma semelhante, mas neste caso deve-se otimizar a arquitetura para promover a emissão estimulada, com o uso de altas dopagens, a fabricação de uma cavidade ótica e o uso de muitas camadas (heteroestruturas semicondutoras) ao invés de apenas uma junção.

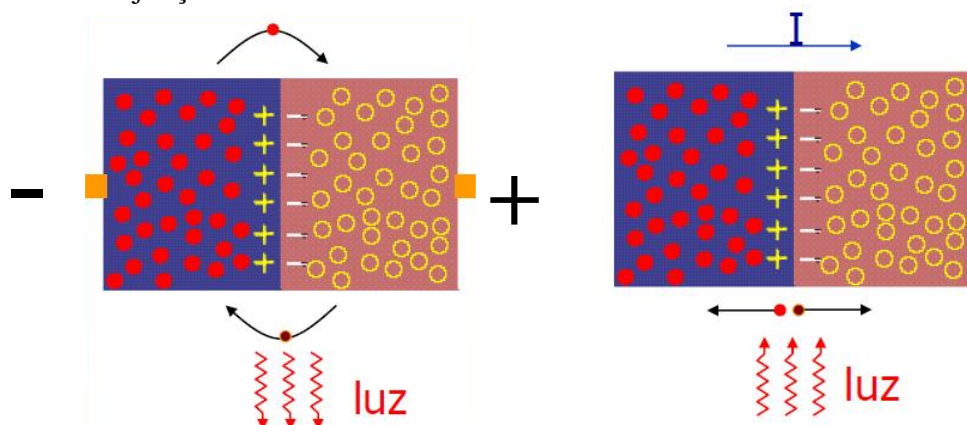


Figura 8.15 – Funcionamento do diodo emissor de luz (LED).

Também na Fig. 8.15 (direita) ilustramos o funcionamento da **célula solar**. Seu comportamento é essencialmente o inverso do LED: quando a luz incide sobre a camada de depleção, pares elétron-buraco são gerados e são rapidamente acelerados pelo campo elétrico da camada para regiões opostas da junção, dando origem a uma corrente elétrica. A Fig. 8.16 mostra a evolução da eficiência das células solares, feitas com diferentes materiais semicondutores, ao longo do tempo.

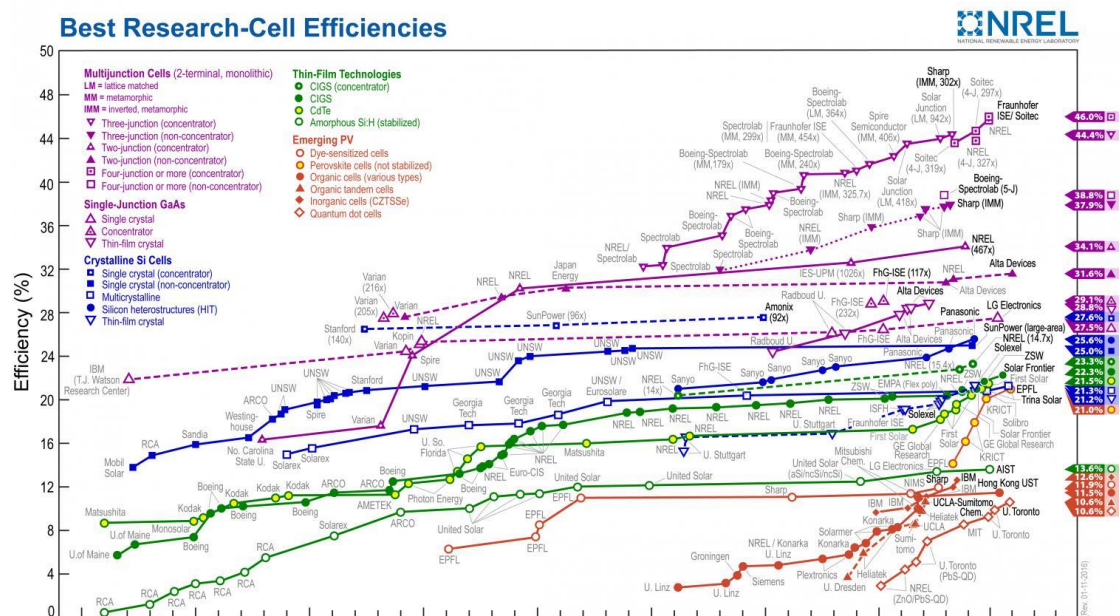


Figura 8.16 – Evolução da eficiência das células solares ao longo dos anos.

Finalmente, mostramos na Fig. 8.17 a arquitetura e a curva característica do **MOSFET** (“metal-oxide-semiconductor field-effect transistor”), o tipo de transistor mais comum nos computadores. O MOSFET é um dispositivo com 3 contatos metálicos: fonte, dreno e porta. A corrente flui entre a fonte e o dreno, controlada pelo eletrodo da porta. A curva característica ilustra este comportamento. A corrente percorre o “canal”, localizado entre duas junções pn, como mostra a figura. O transistor é a unidade básica da computação binária: o 0 e 1 da “lógica booleana” dos computadores são realizados fisicamente pela passagem ou bloqueio da corrente pelo canal, controlada pela porta.

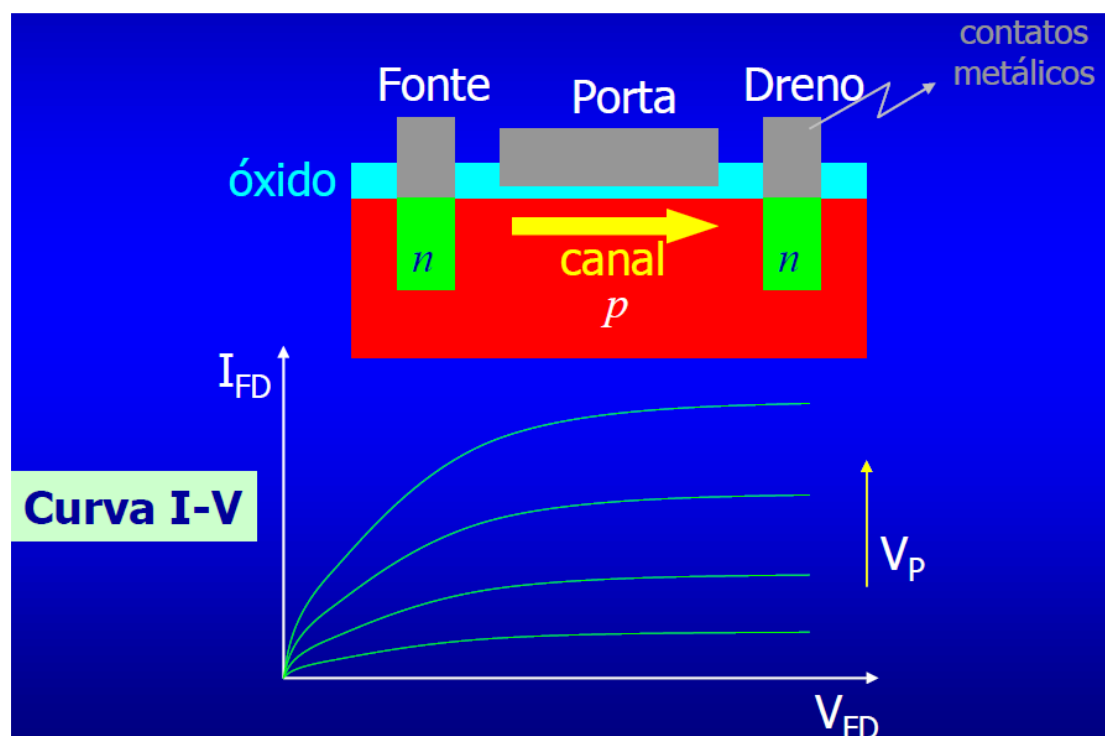


Figura 8.17 – Arquitetura e curva característica do MOSFET.

Referências:

- “Early history of the physics and chemistry of semiconductors - from doubts to fact in a hundred years”, G. Busch, Eur. J. of Phys. 10, 254 (1989).
- Ashcroft e Mermin, Capítulos 28 e 29.
- “Materiais e Dispositivos Eletrônicos”, Sérgio M. Rezende, Editora Livraria da Física.