



ESCOLA DE ENGENHARIA DE SÃO CARLOS

DEPARTAMENTO DE ENGENHARIA ELÉTRICA
E DE COMPUTAÇÃO

FUNDAMENTOS DE SEMICONDUTORES

Prof. Dr. Emiliano R. Martins

Prefácio

Como indicado pelo nome, semicondutores são materiais com propriedades intermediárias entre isolantes e condutores. Mais precisamente, semicondutores são materiais cuja condutividade pode ser controlada de forma relativamente simples.

Materiais semicondutores são amplamente utilizados na indústria de eletrônica. Componentes eletrônicos, por exemplo, são feitos de silício, que é o material semicondutor mais importante do ponto de vista tecnológico atualmente. Naturalmente, a indústria eletrônica impulsiona a pesquisa em semicondutores, pesquisa essa que resultou em uma verdadeira revolução na sociedade após o século XX, com o advento de computadores cada vez mais potentes e da Internet. O sucesso e o impacto da pesquisa em semicondutores servem como exemplos paradigmáticos de como o investimento em ciências e tecnologia pode transformar a sociedade. É importante notar que, apesar de esse ser um campo de pesquisa com mais de 100 anos, semicondutores continuam sendo intensamente investigados, causando inclusive mini-revoluções na indústria, como exemplificado pela recente descoberta do grafeno, que resultou no prêmio Nobel de Física em 2010.

Como é de se esperar, existe uma vastíssima literatura sobre semicondutores. De maneira geral, o estudo de semicondutores pode ser dividido em duas etapas: o estudo do material semicondutor em si, e o estudo dos dispositivos semicondutores. A primeira etapa fornece a base para a segunda etapa que, por sua vez, é base para toda eletrônica.

Qual é, então, o objetivo deste livro? Este livro é voltado para alunos de Engenharia Elétrica. Portanto, a principal motivação para estudar semicondutores aqui é derivada do interesse em dispositivos eletrônicos. Praticamente todos os dispositivos eletrônicos são feitos de junções **p-n**. Um diodo é uma junção **p-n**. Uma célula solar é uma junção **p-n**. Um fotodetector é uma junção **p-n**. Um transistor nada mais é que duas junções **p-n** em série. Portanto, é impossível compreender o funcionamento de dispositivos semicondutores sem possuir uma sólida base sobre junções **p-n**. Além disso, um conhecimento adequado de junções **p-n** já é mais que meio caminho andado para a compreensão da vasta maioria dos dispositivos eletrônicos e optoeletrônicos. Assim, fica claro que nosso principal objetivo aqui é lançar as bases teóricas necessárias para um entendimento adequado do que é e de como funciona uma junção **p-n**. No final deste curso, o aluno deverá entender com clareza de onde vem a relação tensão versus

corrente em um diodo. No caminho, o aluno também entrará em contato com noções centrais a respeito de baterias, fotodetectores, células solares e transistores.

O livro é dividido em quatro capítulos: 1-Noções de Física Estatística, 2 - Semicondutores, 3 - Dispositivos semicondutores e 4 - Aplicações de dispositivos semicondutores.

No primeiro capítulo, vamos estudar dois conceitos de Física Estatística que são de fundamental importância em eletrônica: o conceito de potencial eletroquímico e a distribuição de Fermi-Dirac. É impossível compreender o que é uma junção **p-n** sem compreender o que é o potencial eletroquímico. No caminho, temos que pincelar algumas noções relacionadas à mecânica quântica, mas não podemos nos aprofundar nelas pela óbvia falta de espaço. Assim, o primeiro capítulo será dedicado ao estudo do potencial eletroquímico e da distribuição de Fermi-Dirac, sendo que ambos são parte da disciplina Física Estatística (ou Física Térmica). É importante ressaltar o óbvio aqui: este não é um livro de Física Estatística. Em momento nenhum o aluno deve imaginar que o estudo do capítulo 1 por si só substitui um curso de Física Estatística ou termodinâmica. Ao contrário, no capítulo 1 serão tratados somente os conceitos indispensáveis para semicondutores. No segundo capítulo, iremos focar em alguns aspectos da física de semicondutores mais específicos, como a teoria de bandas e deduzir algumas importantes propriedades elétricas dos semicondutores. No terceiro capítulo, iremos utilizar todos os conceitos dos capítulos anteriores para estudar a junção **p-n** e deduzir suas propriedades elétricas. Finalmente, no quarto capítulo, serão mostrados mais dois exemplos de aplicações dos conhecimentos adquiridos nesse curso: dispositivos fotovoltaicos e transistores.

SUMÁRIO

Sumário

Prefácio	2
1–Noções de Física Estatística.....	5
1.1 Introdução	5
1.2 Equilíbrio Térmico.....	22
1.3 Função de partição - Parte I	29
1.4 Equilíbrio difusivo e potencial químico.....	36
1.5 Função de Partição, Parte II	50
1.6 Energia e número de elementos de um sistema	53
1.7 A distribuição de Fermi-Dirac	58
1.8 Analogia do sistema “caixa” com o sistema “moedas”.	72
1.9 Concentração de elétrons e nível de Fermi	74
1.10 Transporte	76
1.11 Relação entre corrente e densidade de portadores	85
1.12 Sugestões bibliográficas.....	93
1.13 Exercícios.....	94
2 – Semicondutores.....	102
2.1 Teoria de Bandas.....	102
2.2 Elétrons e buracos	111
2.3 Concentração de elétrons livres	113
2.4 Densidade de estados	121
2.5 Concentração de buracos e nível de Fermi	131
2.6 Semicondutores extrínsecos (dopagem)	134
2.7 Exercícios.....	141
3 – Introdução aos dispositivos semicondutores: a junção p-n.....	142
3.1 Junção p-n em equilíbrio.....	142
3.2 Sistemas fora do equilíbrio: níveis de quasi-Fermi.....	159
3.3 Relação tensão x corrente na junção p-n (Equação de Shockley).	162
3.5 Sugestões bibliográficas.....	181
3.6 Exercícios.....	181
4 – Efeito fotovoltaico (fotodetectores e células solares) e transistores.	184
4.1 Fotodetectores e células solares	184
4.2 Transistores	190
4.3 Exercícios.....	195
Apêndice A: Interpretação geométrica da energia livre	196

1–Noções de Física Estatística

1.1 Introdução

Um bom começo para o estudo da Física Estatística é a compreensão de um princípio probabilístico que pode ser enunciado como: *a propriedade do estado de um sistema macroscópico em equilíbrio é sempre a mais provável*. O objetivo desta seção introdutória é explicar esse princípio.

Porém, para compreender este enunciado, devemos primeiro entender qual é o significado dos termos “sistema”, “estado” e “equilíbrio”.

Sistema é uma daquelas palavras que, de tanto serem usadas, acabaram perdendo qualquer significado (tanto em inglês, quanto em português). Em Física Estatística, chamamos de sistema o conjunto de elementos em que estamos interessados. Por exemplo, se quisermos estudar a interação entre dois gases, então, os dois gases formarão o nosso sistema. Dizemos também que o sistema é isolado quando ele não interage com o mundo externo. No exemplo dos dois gases, o sistema estaria isolado se os dois gases interagissem somente entre si. Isso poderia ser feito colocando-os em um compartimento isolado do mundo externo. Poderíamos também chamar de sistema o conjunto dos gases junto com o compartimento, já que o volume do compartimento poderia ser de interesse. No fundo, “sistema” é só um nome pomposo que quer dizer a coisa ou coisas que queremos estudar.

Já o conceito de estado é um pouco mais sutil. Apesar de esse ser um termo utilizado em várias áreas da física, ele ganhou um significado particularmente importante com o advento da física quântica. Assim, é importante compreender tanto o que é um estado de maneira geral como também o que é um estado quântico. De maneira geral, “estado” é sinônimo de condição. Quando perguntamos qual é o estado de um sistema, estamos efetivamente perguntando qual é a condição do sistema. O que exatamente essa pergunta significa depende do contexto. Se eu estiver interessado em comprar seu carro e te perguntar qual é o estado ou condição do seu carro, você vai entender que eu quero saber se ele está novo ou velho, qual é a quilometragem, e por aí vai. Em física clássica, o que define o estado de um sistema é a posição e o momento (termo equivalente à “quantidade de movimento”) de todas as partículas do sistema. E essa definição vem do fato que, para prever a evolução temporal do sistema, você tem que saber a posição

e o momento de todas as partículas do sistema. Então, no exemplo dos gases, quando perguntamos qual é o estado do gás, estamos efetivamente perguntando qual é o momento e posição de cada elemento do gás. A cada estado, podemos associar propriedades desse estado. A propriedade mais comum, e que é de grande interesse aqui, é a energia.

Agora, imagine um sistema composto por uma bolinha de massa m em uma rampa, como ilustrado na Figura 1. Se a bolinha for solta de uma altura h , ela terá energia total inicial em forma de energia potencial gravitacional ($U = mgh$, onde g é aceleração gravitacional). Desprezando o atrito, sabemos que essa bolinha irá oscilar, transformando a energia potencial U em energia cinética e vice-versa. Se perguntarmos, então, qual é a energia do sistema, a resposta seria U . Note que, neste exemplo, a energia não depende do tempo. Porém, se perguntássemos qual é o estado completo do sistema, teríamos que informar a posição e o momento, que neste caso são funções periódicas do tempo. Esse é um exemplo de um sistema onde existem vários estados possíveis (cada par posição e momento representa um estado possível e, naturalmente, o estado do sistema muda com o tempo, à medida que a bola oscila), mas onde vários estados possuem a mesma propriedade (se você soltar a bolinha de uma determinada altura, todos os estados acessados pelo sistema enquanto a bolinha oscila vão ter a mesma energia).

E se perguntássemos quais são as energias possíveis? Neste caso, estaríamos perguntando quais valores que U pode assumir. Como podemos, a priori, soltar a bolinha de qualquer altura h , então, U pode assumir qualquer valor. Mesmo se fixássemos uma altura máxima em, digamos, 10 metros, ainda assim teremos infinitos valores de U , porque U poderia tomar qualquer valor entre 0 e $10mg$ (lembre-se que m é a massa e g é a aceleração gravitacional). Dizemos, então, que existe um contínuo de valores possíveis de U entre 0 e $10mg$. A palavra “contínuo” aqui enfatiza que todos os valores entre 0 e $10mg$ são possíveis, sem nenhuma região excluída (o que, por sua vez, causaria uma descontinuidade).

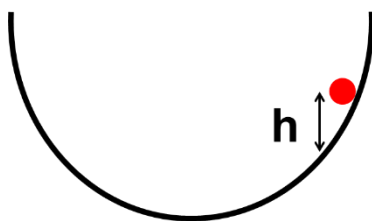


Figura 1 Um sistema que consiste em uma bolinha de massa m deslizando sem atrito em uma rampa pode apresentar qualquer estado de energia. Basta ajustar a altura h que a bolinha é solta para obter o estado desejado.

Em física clássica, as energias possíveis dos sistemas estão sempre em um contínuo. Basta pensar que, se você pegar uma massa, a priori, é possível ajustar a energia cinética dessa massa para qualquer valor: basta ajustar sua velocidade (a teoria da relatividade não altera esse fato, ela só impõe um limite à velocidade da massa, que é o da velocidade da luz; mas, mesmo assim, existe um contínuo de infinitas velocidades possíveis entre 0 e a velocidade da luz).

O conceito de estado em mecânica quântica é similar ao clássico. Por exemplo, o estado de um elétron é dado por uma função que carrega todas as informações a respeito das propriedades físicas do elétron. Mas existe uma diferença importante: é comum encontrar sistemas quânticos onde os estados possíveis não formam uma distribuição contínua, mas sim discreta. Utilizando o exemplo da bolinha, isso seria equivalente a dizer que uma bolinha quântica não pode ser solta de uma altura qualquer. Ela só pode ser solta de certas alturas bem determinadas, como ilustrado na Figura 2. Neste exemplo, existem três alturas possíveis (e, portanto, existe um número discreto, que você pode contar, de alturas possíveis) que chamamos de h_0 , h_1 e h_2 . Portanto, neste exemplo, temos apenas três energias possíveis, $U_0 = mgh_0$, $U_1 = mgh_1$ e $U_2 = mgh_2$.

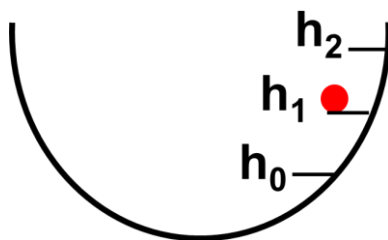


Figura 2 Em um sistema quântico, apenas alguns níveis discretos de energia são permitidos. Aqui, representamos essa propriedade delimitando três alturas h_0 , h_1 e h_2 das quais a bolinha pode ser solta. Portanto esse sistema possui apenas três estados de energia possíveis.

É importante explicar de onde vem essa restrição no estado do sistema. Na verdade, essa não é uma peculiaridade da mecânica quântica. Essa é, sim, uma peculiaridade de sistemas cujos elementos são ondas. Uma fibra óptica, por exemplo, é um sistema que confina uma onda eletromagnética e esse confinamento resulta em uma discretização dos estados possíveis, que chamamos de modos. Se você for a uma loja e pedir uma fibra óptica monomodo, você estaria pedindo uma fibra que admite apenas um estado possível, enquanto uma fibra multimodo pode admitir dois ou mais estados, mas sempre um número discreto de estados.

A origem dessa discretização é o confinamento da onda. O confinamento impõe certas relações de fase na onda. Por exemplo: suponha que uma onda eletromagnética seja confinada entre dois metais, como mostrado na Figura 3. Para que a onda possa existir dentro dessa cavidade (nome pomposo para caixa), é necessário que a fase acumulada quando a onda se propaga na horizontal e volta sobre si mesma seja um múltiplo de 2π . Caso contrário, cada vez que ela der uma volta completa ela vai sofrer interferência destrutiva consigo mesma, o que impedirá a sua existência.

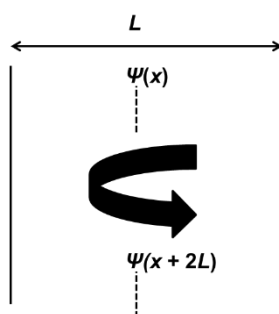


Figura 3 Uma onda confinada só pode existir se a fase acumulada em um giro completo for um múltiplo de 2π .

Podemos expressar essa restrição em termos matemáticos. Para tal, considere que a onda seja descrita pela função:

$$\psi(x) = \exp(ikx)$$

Equação 1

Onde i é o número imaginário puro e k é a constante de propagação da onda. A dependência temporal da onda foi suprimida aqui porque estamos interessados apenas na dependência espacial do acúmulo de fase, o que torna a dependência temporal irrelevante para nós neste exemplo.

Qual é a fase acumulada pela onda quando ela percorre uma volta inteira? Vale lembrar que cada vez que ela é refletida por uma parede metálica, ela ganha uma fase π . Como existem duas paredes, o acúmulo de fase na reflexão nas paredes metálicas é de 2π , ou seja, as duas paredes juntas não contribuem para o acúmulo de fase. Então, o acúmulo de fase vem só da propagação na cavidade. Como a cavidade possui comprimento L , então uma volta inteira corresponde a uma distância $2L$. Portanto, a relação entre a onda que sai de um ponto x e a onda que volta sobre o mesmo ponto depois de um giro completo ($x + 2L$) é:

$$\psi(x + 2L) = \exp[ik(x + 2L)] = \psi(x)\exp(ik2L)$$

Equação 2

Assim, vemos que o acúmulo de fase em um giro completo é igual a $k2L$. Mas como $\psi(x)$ e $\psi(x + 2L)$ referem-se ao mesmo ponto, então, obviamente $\psi(x) = \psi(x + 2L)$. Mas isso só pode ser verdade se o acúmulo de fase for um múltiplo de 2π . Portanto, a condição de existência da onda confinada por paredes metálicas é expressa pela relação:

$$k2L = q2\pi$$

Equação 3

Onde q é um número inteiro. O que a Equação 3 está efetivamente nos dizendo é que os únicos estados possíveis de uma onda confinada em uma caixa de comprimento L são aqueles que obedecem à condição:

$$k = \frac{q\pi}{L}$$

Equação 4

Assim, o primeiro estado é dado por $q = 1$, o segundo por $q = 2$, e por aí vai. Note que, embora tenhamos infinitos estados (q pode ser um número inteiro qualquer), eles são discretos, pois q é um número inteiro (não existem estados com $q = 1.5$, por exemplo).

E o que isso tem a ver com mecânica quântica? Tudo a ver, porque um dos postulados mais fundamentais da mecânica quântica é que a matéria possui propriedades ondulatórias. Você já deve ter ouvido falar da dualidade onda-partícula.

Um elétron não é como uma bolinha, que possui localização determinada. Um elétron é, na verdade, uma onda. Que a matéria é onda é uma das descobertas mais revolucionárias e malucas a que o homem já chegou (tão maluca e com consequências tão bizarras que Einstein morreu sem acreditar que a mecânica quântica pudesse ser uma descrição correta da natureza; mas a natureza é o que é).

Agora, suponha que a Equação 1 seja a função de onda de um elétron e que o elétron esteja preso em um poço (ou uma cavidade) com comprimento L (um poço que prende uma partícula quântica, no caso aqui o elétron, é chamado de poço quântico – o nome é bonito, mas continua sendo só um poço). Como a argumentação feita para encontrar os valores possíveis de k vale para qualquer onda, então a Equação 4 também é verdade para o elétron. Acontece que o momento do elétron, que vamos chamar de p , é proporcional à constante de propagação:

$$p = \frac{h}{2\pi} k$$

Equação 5

Onde h é a famosa constante de Planck (se você está se perguntando de onde saiu essa relação entre p e k , ela meio que faz parte dos postulados da mecânica quântica. Se ela é um postulado, então, nós não sabemos muito bem o porquê dessa relação ser verdadeira, apenas sabemos que é verdadeira). Diante disso, substituindo a Equação 4 na Equação 5 temos:

$$p = q \frac{h}{2L}$$

Equação 6

Ou seja, o elétron preso em um poço quântico só pode ter valores de momento que obedecem a Equação 6. Por outro lado, sabemos da mecânica clássica que a energia cinética de uma partícula é dada pela relação:

$$E = \frac{p^2}{2m}$$

Equação 7

Onde m é a massa da partícula (se você achou a relação acima meio obscura, lembre-se que $p = mv$, onde v é a velocidade, e que se você substituir p por mv acima você encontra $E = mv^2/2$. Ou seja, a relação acima é só o bom e velho $E = mv^2/2$, mas escrito em termos de p). Substituindo a Equação 6 na Equação 7, temos que as energias possíveis do elétron preso em um poço quântico são dadas por:

$$E = \frac{h^2 q^2}{8L^2 m}$$

Equação 8

Apesar da simplicidade, a Equação 8 descreve inúmeras propriedades de suma importância. Ela nos diz quais são os níveis de energia possíveis de um elétron aprisionado em um poço quântico. Por exemplo, se um material for reduzido a dimensões quânticas, ele passará a emitir e absorver luz de acordo com as diferenças de energia dadas pelos níveis da Equação 8. Hoje em dia, existem técnicas para formar materiais minúsculos, da ordem de algumas dezenas de angstroms, chamados de pontos quânticos. Esses materiais absorvem e emitem luz dependendo do tamanho deles (já que E depende de L na Equação 8), ou seja, o material efetivamente muda de cor dependendo somente do tamanho dele. Pontos quânticos são uma tecnologia emergente utilizada desde tratamento de câncer até a fabricação de LEDs (*Light Emitting Diodes*). Você já deve ter visto algum televisor com tecnologia QLEDs. O “Q” do QLED vem de *quantum dot*, que é o termo em inglês para ponto quântico.

Note também que, em nosso exemplo, cada número inteiro positivo define um estado (números negativos não entram porque definiriam o mesmo estado do número positivo correspondente). Além disso, cada número inteiro positivo define uma energia. Portanto, em nosso exemplo, as energias de estados diferentes serão diferentes. Mas nem sempre isso é verdade. Como veremos mais adiante, se o poço for tridimensional, existirão mais de um estado com a mesma energia. De maneira geral, estados diferentes podem ter a mesma energia (o caso do poço quântico unidimensional é um caso especial).

Bom, toda essa discussão foi para apresentar o conceito de estado e, particularmente, de estados quânticos. Tudo isso porque queremos entender o princípio enunciado no começo dessa seção. Se você rebobinar a fita e voltar para o começo da seção 1.1, vai ver que está faltando estabelecer a noção de equilíbrio. Agora que sabemos o que é um

sistema e o que é o estado do sistema, fica fácil estabelecer o que é equilíbrio: dizemos que um sistema está em equilíbrio quando as suas propriedades não são mais alteradas. Voltando ao exemplo dos dois gases interagindo, as propriedades de cada gás (que poderiam ser descritas por grandezas macroscópicas como volume, pressão e temperatura) vão alterar à medida que os gases interagirem. Porém, chegará uma hora em que o sistema como um todo atingirá o equilíbrio, de maneira que as propriedades dos dois gases passarão a ser fixas. Note que os gases continuarão interagindo, mas as propriedades deles não serão mais alteradas pela interação após atingirem o equilíbrio. Por quê? O que essas propriedades de equilíbrio têm de tão especial assim que faz com que, uma vez que o sistema as atinge, elas ficam lá para sempre? A resposta está no enunciado do princípio probabilístico com o qual iniciamos este capítulo. Vamos lembrar o que ele diz: *a propriedade do estado de um sistema macroscópico em equilíbrio é sempre a mais provável.*

Vamos pensar um pouco sobre este enunciado. Nessa frase, estamos efetivamente dizendo que a propriedade do estado de equilíbrio é a mais provável. Note que o princípio poderia ser enunciado da seguinte maneira: *um sistema macroscópico cuja propriedade não é mais alterada possui a propriedade mais provável.* Essa segunda maneira de enunciar o princípio é menos elegante, porém ela é mais útil para enfatizar que a palavra equilíbrio quer dizer simplesmente que a propriedade está fixa.

Então, o que exatamente o princípio está nos dizendo? Em primeiro lugar, lembre-se que o estado do sistema se refere à condição física do sistema. Se o sistema for um gás, seu estado é determinado pela posição e momento de todas as moléculas do gás. Assim, cada estado do gás possui certas propriedades como, por exemplo, a energia total do gás, que geralmente é um parâmetro de interesse (lembre-se também que estados diferentes podem ter a mesma energia). Para facilitar a resposta, vamos supor que estamos interessados apenas na energia total do sistema. Então, o princípio está nos dizendo que, quando dois sistemas interagirem entre si (ou com o universo exterior), uma hora eles vão adquirir um valor de energia igual ao valor de energia mais provável e, depois disso, a energia do sistema não altera mais.

Não sabemos como exatamente ocorre a interação (afinal de contas Física Estatística é uma ciência probabilística), sabemos apenas que alguns estados são possíveis como resultado dessa interação. Por exemplo, suponha que o sistema seja novamente constituído de dois gases isolados do mundo externo, um com energia $U1_0$ e o outro com energia $U2_0$ (aqui, novamente, a energia total de cada gás é a propriedade de

interesse). Se deixarmos os gases interagindo, não saberemos bem quais serão as suas energias. Mas porque energia é conservada, sabemos que todas as energias possíveis do sistema terão que satisfazer a condição: $U_1 + U_2 = U_{1_0} + U_{2_0}$, onde U_1 e U_2 são, respectivamente, as energias dos gases 1 e 2, em qualquer momento. Portanto, os estados possíveis são aqueles cujas energias satisfazem a condição $U_1 + U_2 = U_{1_0} + U_{2_0}$. E, de todas as energias possíveis (a energia é a propriedade, e você pode escolher tanto U_1 como U_2 como sendo a sua energia de interesse, já que, por conservação de energia, eles não são independentes), o princípio nos diz que o sistema terá a mais provável. Em outras palavras, o princípio nos diz: “pegue todas as energias possíveis e veja qual é a mais provável, chame essa energia mais provável de X . Então, seu sistema terá energia X depois de um tempo e não vai mais alterar”.

Toda essa elucubração leva a três perguntas: como posso achar as energias e os estados possíveis, o que o termo “propriedade mais provável” realmente significa e por que cargas d’água o sistema evolui para essa propriedade mais provável e lá fica? Essas três perguntas devem ser abordadas juntas e são melhores tratadas utilizando exemplos.

Vamos utilizar o bom e velho exemplo da moeda lançada e das probabilidades de dar cara ou coroa. Suponha que nosso sistema seja composto por N moedas que foram lançadas ao ar. **Vamos supor que a energia U do sistema seja definida pelo número de moedas que caíram com a coroa virada para cima. Assim, uma coroa para cima define $U = 1$, duas coroas $U = 2$ e assim em diante.** Com essa definição, quantas energias possíveis nós temos? Bom, isso depende de quantas moedas temos. Chamando o número de moedas de N , então se $N = 1$, só teremos duas energias possíveis: $U = 0$ (0 coroas para cima), ou $U = 1$ (1 coroa para cima). Se $N = 2$, então temos três energias possíveis: $U = 0$, $U = 1$ ou $U = 2$ (0, 1 ou 2 coroas viradas para cima, respectivamente). Mas qual é a energia mais provável? Bom, existem quatro estados possíveis: *cara-cara*, *cara-coroa*, *coroa-cara* e *coroa-coroa*. Mas dois desses estados possíveis possuem a mesma energia (tanto o estado *cara-coroa* como o estado *coroa-cara* possuem energia igual a 1).

Então, existe 25% de chance da energia ser zero ($U = 0$) - que corresponde ao estado *cara-cara*; 25% de chance da energia ser 2 ($U = 2$) - que corresponde ao estado *coroa-coroa*; e 50% de chance da energia ser 1 ($U = 1$) - que corresponde tanto ao estado *cara-coroa* quanto ao estado *coroa-cara*.

Portanto, a energia mais provável é 1 ($U = 1$), porque o sistema formado por duas moedas possui dois estados possíveis para essa mesma energia.

Vamos, então, generalizar esse raciocínio calculando qual é a probabilidade de termos M coroas em um sistema com N moedas. Primeiramente, quantos são os estados possíveis? Essa é fácil, temos 2^N estados possíveis de caras e coroas. Para calcular a probabilidade de saírem M coroas, basta calcularmos quantos estados possuem M coroas e dividir esse número por 2^N . O número de estados com M coroas viradas corresponde ao número de maneiras que podemos arranjar N moedas com M delas com a coroa virada e $N - M$ com cara virada. Para encontrar esse número, primeiro imagine que cada moeda tenha um pequeno marcador e que tenhamos N potinhos lado a lado, cada potinho para uma moeda. Esqueça, por enquanto, que temos cara e coroa e se pergunte de quantas maneiras podemos distribuir N moedas diferentes (pois o marcador as diferenciou) nos potinhos. Bom, aí é fácil: a resposta é $N!$ maneiras. Agora, note que na contagem $N!$ há diversas configurações de cara e coroa idênticas que foram contadas mais de uma vez, porque o marcador as havia diferenciado. Por exemplo, suponha que tínhamos duas coroas e três caras (cinco moedas). Vamos supor que marcamos as moedas com os marcadores M1, M2, M3, M4 e M5. Neste exemplo, suponha também que as moedas M1 e M2 estão com a cara virada e as M3, M4 e M5 estão com a coroa virada. Assim, existem $5! = 120$ diferentes estados de moedas **marcadas**. Mas note que o estado M1M2M3M4M5 foi contado como sendo diferente do estado M2M1M3M4M5. Mas sabemos que, na verdade, esses dois estados são idênticos (ambos correspondem ao mesmo estado *cara-cara-coroa-coroa-coroa*). Então, temos que dividir o número de estados de moedas marcadas (que neste exemplo é 120) por dois, já que contamos os estados M1M2 e M2M1 como distintos quando, na verdade, são iguais (note que todos os estados foram contados duas vezes, mesmo se M1 e M2 não forem adjacentes. Por exemplo, o estado M1M3M4M2M5 é o mesmo estado M2M3M4M1M5). Mas também temos que dividir por $3! = 6$ porque os estados M3M4M5, M3M5M4, M4M3M5, M4M5M3, M5M3M4 e M5M4M3 foram todos contados como distintos quando, na verdade, são iguais. Assim, definindo a função $A(M)$ como sendo o número de *estados* com M coroas viradas para cima em um total de N moedas, temos que:

$$A(M) = \frac{N!}{M!(N - M)!}$$

Equação 9

Então, a probabilidade de termos M coroas viradas para cima quando N moedas são lançadas é:

$$P(N, M) = \frac{A(M)}{2^N} = \frac{N!}{2^N M! (N - M)!}$$

Equação 10

Podemos utilizar a Equação 9 ou a Equação 10 para estudar o comportamento da probabilidade $P(N, M)$ em função do número de moedas N e do número de coroas viradas M . A Figura 4 ilustra alguns histogramas (normalizados) obtidos da Equação 9. **Note como o número de estados possíveis se concentra em torno de $M = N/2$ à medida que N cresce.** Para valores altos de N , podemos até quantificar qual é a largura do histograma. Se definirmos o limite do histograma como sendo dado pelo valor de M no qual a função A é e^{-1} vezes menor que o valor máximo de A , então a largura fracional $\Delta M/N$ do histograma será dada por:

$$\frac{\Delta M}{N} \approx 4 \sqrt{\frac{1}{2N}}$$

Equação 11

O que isso quer dizer? Esse comportamento quer dizer que quanto **maior** for o número de moedas, **menor** será a proporção de valores de M para os quais temos um valor apreciável de número de estados $A(M)$. Para os exemplos da Figura 4, temos que para $N = 10^2$ o desvio é da ordem de 10^{-1} , para $N = 10^4$ o desvio é da ordem de 10^{-2} e para $N = 10^6$ o desvio é da ordem de 10^{-3} . Portanto, quanto maior for o número de moedas, mais provável será que o resultado seja metade das moedas com coroa para cima. Utilizando uma linguagem um pouco mais precisa, concluimos que a chance de encontrarmos um número de coroas que seja apreciavelmente diferente da metade é muito baixa. Por exemplo, se tivéssemos um milhão de moedas, não ficaríamos surpresos se encontrássemos 500.005 (quinhentos mil e cinco) coroas viradas para cima, porque 500.005 é um número proporcionalmente muito próximo de 500.000 (quinhentos mil). Mas é virtualmente impossível encontrarmos 250 mil

coroas para cima. O que o desvio fracional nos diz é qual é a proporção (em relação ao ponto de máximo) de coroas a mais ou a menos que temos uma chance apreciável de encontrar. Para o caso de $N = 10^6$, o desvio fracional é da ordem de 10^{-3} , o que significa que, $\Delta M \sim 10^3$ ou seja, é provável que o número de coroas que encontraremos seja entre 499 mil e 501 mil. Qualquer valor fora dessa faixa é muito improvável. Portanto, a propriedade do sistema (que em nosso exemplo é a energia determinada pelo número de coroas para cima) estaria muito, muito próxima da propriedade mais provável, simplesmente porque o número de estados possíveis que resultam na propriedade mais provável é bem maior do que o número de estados que resultam nas demais propriedades (em outras palavras, temos muito mais estados com a propriedade $U = N/2$ do que estados com outras energias).

Agora, imagine que você chegue a uma sala com mil moedas espalhadas no chão. Se você passar o olho e vir mais ou menos o mesmo tanto de caras e o mesmo tanto de coroas, você vai pensar que as moedas foram espalhadas ao acaso, porque elas estão bagunçadas (bagunçadas no sentido do número de caras ou coroas, não no sentido de distribuição espacial). Agora, se todas as mil estiverem com a cara virada para cima, você vai deduzir que alguém virou as moedas propositadamente. Essa sua intuição é consequência do fato de você saber que a propriedade mais “desorganizada” é muito mais provável que a propriedade “organizada”. Por isso, dizemos que a tendência do universo é aumentar a desorganização, o que é uma das leis fundamentais da Física Estatística (esse princípio do aumento da desorganização é formulado como a segunda lei da Termodinâmica, que afirma que a entropia de sistemas sempre aumenta – ou, em alguns casos ideais, é mantida constante).

Em um sistema macroscópico, temos um número altíssimo de elementos (por exemplo, em um gás, líquido ou sólido, N seria o número de átomos no sistema). Portanto, em um sistema macroscópico, o desvio fracional em torno da propriedade mais provável é muito, muito pequeno. Por exemplo, um gás possui aproximadamente 10^{23} moléculas em 1 cm^3 em condições de temperatura e pressão ambientes. Portanto, temos N da ordem de 10^{23} , o que resulta em um desvio fracional da ordem de 10^{-12} !! Isso quer dizer, por exemplo, que só é provável que você encontre energias que variem em torno de $10^{-10}\%$ da energia mais provável (as energias prováveis estão na faixa $U_{\text{maximo}} \pm U_{\text{maximo}} \times 10^{-12}$). Como essa faixa de variação é absolutamente irrisória e fora de qualquer possibilidade de detecção experimental, então, para sistemas

macroscópicos, podemos efetivamente dizer que o sistema terá a propriedade mais provável.

Note que, apesar da Equação 9 só valer para sistemas binários (com dois estados por elementos possíveis, no exemplo acima cara ou coroa), o comportamento da função $P(N, M)$ é geral: em sistemas macroscópicos, a função densidade de probabilidade é praticamente uma função impulso situada na propriedade mais provável.

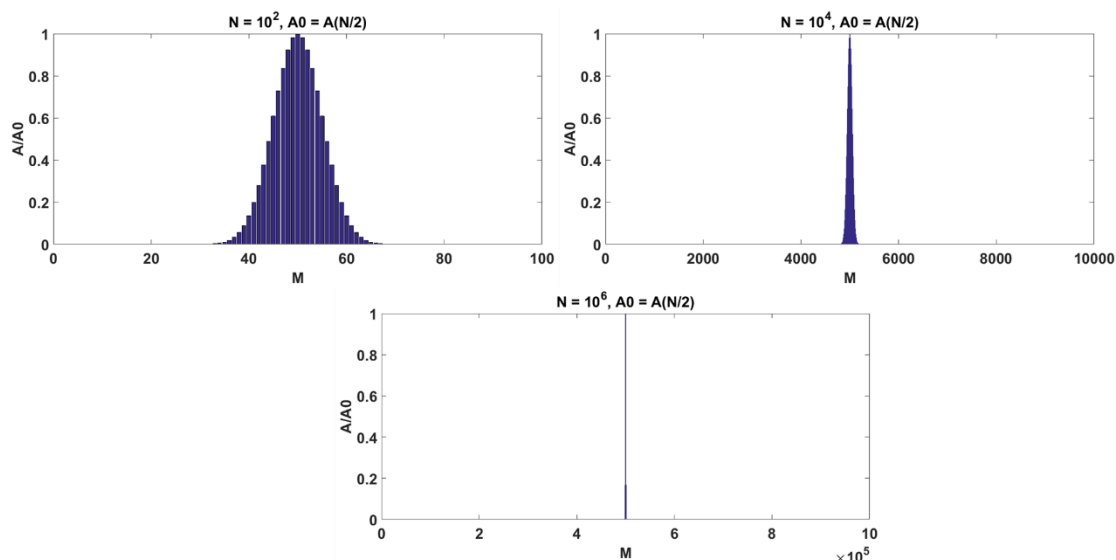


Figura 4 Comportamento de A em função de M para diferentes valores de N . Para facilitar a visualização, a função A foi normalizada. Note que à medida que N aumenta, a função A fica cada vez mais concentrada em torno de $M = N/2$.

Vamos recapitular o raciocínio: utilizando o exemplo das moedas, vimos que se o número de moedas N for muito grande, então, o número de estados nos quais metade das moedas estão com a coroa para cima é muito, muito maior que o número de estados com qualquer outro número de coroas viradas para cima. Isso significa que se você jogar um número muito alto de moedas para cima (10 milhões, por exemplo) então, é virtualmente garantido que o estado final apresentará metade de caras e metade de coroas viradas para cima (descontando variações percentuais insignificantes). Se o tanto de coroas viradas para cima definir a propriedade do sistema (no nosso exemplo a propriedade é a energia) então, é garantido que a propriedade final seja a mais provável (metade de coroas para cima). A razão pela qual é virtualmente garantido que teremos metade de caras e metade de coroas viradas para cima advém somente do fato que existe um número muito grande de estados que resultam em metade caras e metade coroas (no exemplo de duas moedas, temos dois estados com metade de caras e metade

de coroas: o estado *cara-coroa*, e o estado *coroa-cara*). Portanto, afirmar que é virtualmente garantido que o estado observado terá metade das coroas viradas para cima se N for muito grande é simplesmente consequência do fato de que o número de estados possíveis que resultam em metade de coroas viradas para cima é muito maior que o número de estados possíveis para outros números de coroas viradas para cima.

Praticamente toda a noção de equilíbrio em Física Estatística é construída em torno dessa observação: que a propriedade final de um sistema macroscópico é a mais provável. A razão pela qual isso é verdade é a mesma do exemplo das moedas: existe um número muito maior de estados possíveis que resultam nessa propriedade. E já sabemos também que essa é a propriedade associada ao maior grau de desorganização. Mas note um pequeno, porém importantíssimo, detalhe: essa conclusão só é válida se supormos que cada estado é igualmente provável. Se os estados não fossem igualmente prováveis, nosso raciocínio iria por água abaixo. Por exemplo, se as moedas fossem “viciadas” de maneira que a probabilidade de sair coroa fosse maior que a de sair cara, então, teríamos que reformular todo o cálculo probabilístico desenvolvido até aqui. Toda a teoria, portanto, depende dessa suposição, que de tão importante recebe o nobre nome de *suposição fundamental da Física Estatística*. Formalmente, **a suposição fundamental da Física Estatística é que todos os estados do sistema são igualmente prováveis**. A Física Estatística é construída em torno desta suposição.

Vamos ver outro exemplo, agora mais próximo de uma situação real, sem moedas. Suponha que você tenha um recipiente esférico lotado de certo gás. Vamos dizer que existem 10^{23} moléculas do gás. Suponha que no tempo t_0 todas as moléculas estão aprisionadas por um campo magnético em um cantinho do recipiente. Aí você desliga o campo magnético. E, agora, você quer estudar a evolução do gás. Digamos que você queira estudar a distribuição espacial do gás. Sendo muito bom em física newtoniana e eletrodinâmica, você resolve aplicar as leis de Newton juntamente com as equações de Maxwell para estudar a evolução de cada molécula, sabendo que elas têm certa velocidade inicial e que elas possuem repulsão elétrica entre si. Para determinar, então, o estado do gás (neste exemplo, suponha que podemos ignorar o momento das moléculas na determinação do estado do gás, e que assim o estado seja determinado apenas pela distribuição espacial) você teria que aplicar as leis de Newton e as equações de Maxwell em 10^{23} moléculas. É claro que você não vai dar conta. Você e nenhum computador vão dar conta. A física convencional, apesar de continuar sendo válida, fica

impraticável nesse problema. O que você faz? Você faz Física Estatística. O que isso exatamente quer dizer? Você supõe (sim supõe!) que, depois de um intervalo de tempo Δt , necessário para o sistema evoluir (e entrar em equilíbrio), uma única molécula tem a mesma probabilidade de estar em qualquer posição no recipiente, ou seja, nenhum ponto do recipiente é especial. Essa é a suposição fundamental da Física Estatística, expressa em termos específicos do nosso exemplo: cada molécula tem igual probabilidade de estar em qualquer canto do recipiente (todos os estados são igualmente prováveis). Portanto, em qualquer instante de tempo depois de Δt (ou seja, em qualquer instante de tempo após atingir equilíbrio) uma molécula tem a mesma chance de estar aqui ou de estar acolá: ela pode estar em qualquer lugar. Dado que você fez essa suposição, você poderia calcular qual é a probabilidade de se obter certa distribuição espacial de moléculas. Para isso você teria que calcular o número total de estados que resultam nessa certa distribuição espacial e dividir pelo número total de estados possíveis (lembre-se que, nesse exemplo, cada estado seria caracterizado pela coordenada espacial de cada molécula do gás). Daria mais trabalho fazer isso do que cancelar uma conta de celular.

Mas você, que não é bobo, observa que esse é um sistema macroscópico (muitas moléculas). Então, você sabe que existem muito mais estados possíveis que resultam na propriedade mais provável. Assim, você já sabe que a probabilidade de o sistema estar na propriedade mais provável é muito, muito, perto de um. E qual é a propriedade mais provável? É a relacionada ao maior grau de desordem, ou seja, é a propriedade caracterizada por uma distribuição uniforme de moléculas. Em outras palavras: existem muito, muito mais estados nos quais as moléculas enchem o recipiente de maneira uniforme do que estados resultando em uma distribuição não uniforme de moléculas. Portanto, essa é sua resposta: eu não sei exatamente onde cada molécula vai estar, mas eu sei que a distribuição espacial delas será uniforme, pois essa é a propriedade mais provável e como esse é um sistema macroscópico, então, a probabilidade da propriedade mais provável é virtualmente igual a um. Na seção seguinte, vamos utilizar esse princípio para quantificar algumas grandezas.

Talvez você esteja com uma pulga atrás da orelha com essa história de distribuição uniforme de moléculas ser a mais desorganizada, já que a noção de uniformidade lembra a noção de organização, e não o contrário. Mas esse conceito não tem nada de muito científico e misterioso. A noção de desorganização aqui (e dessa ser a propriedade mais provável) é bem intuitiva e poderia ser exemplificada da seguinte

forma: suponha que você tenha dois amigos viciados em sapatos. Um amigo é muito organizado, e o outro é muito desorganizado. Se você entrasse no quarto do amigo organizado, você esperaria encontrar todos os sapatos guardados em algum canto (portanto, uma distribuição espacial não-uniforme de sapatos). Agora, se você entrasse no quarto do amigo desorganizado, então, você esperaria encontrar sapato para tudo quanto é lado (portanto, uma distribuição espacial uniforme de sapatos). Você, com certeza, já ouviu dizer que a segunda lei da Termodinâmica é a lei do aumento da entropia. Apesar de chamar lei, e apesar de soar grandiosa, essa lei é simplesmente uma observação probabilística. Como veremos na seção 1.2, a entropia está relacionada ao número de estados possíveis do sistema. Dessa forma, o que a segunda lei da Termodinâmica afirma é simplesmente que o sistema evolui para a propriedade que possui mais estados possíveis (ou seja, a propriedade mais provável). E o número de estados possíveis para uma mesma propriedade é expresso em termos da entropia, mas a mensagem é exatamente essa: o sistema evolui para a propriedade mais provável, que é a que possui mais estados possíveis. O exemplo do sapato é um exemplo típico da segunda Lei da Termodinâmica: o rapaz desorganizado representa um sistema térmico e, por ser desorganizado, ele possui entropia alta, já que existe um número enorme de estados que resultam na distribuição uniforme de sapatos (existem vários jeitos de fazer bagunça), caracterizado por sapatos para tudo quanto é lado (sapato verde no canto esquerdo e vermelho no canto direito corresponde à mesma propriedade que sapato verde no canto direito e vermelho no canto esquerdo – ambos correspondem à propriedade bagunça!). A propriedade dos sapatos do rapaz organizado, por outro lado, possui um único estado possível: se você mudar um único sapato, o rapaz organizado vai chiar porque você alterou o estado, e conseqüentemente a propriedade, do sistema dele. Portanto, o rapaz organizado possui entropia baixa (apenas um estado possível para aquela propriedade).

Para finalizar esta seção introdutória, é importante enfatizar que a Física Estatística é uma disciplina que lida com falta de informação. Ela não sabe como o sistema interage. Ela simplesmente supõe que todos os estados possíveis do sistema têm igual probabilidade, ou seja, que o sistema pode evoluir para qualquer estado possível com igual probabilidade. Mas daí resulta que, como a propriedade mais provável possui um número muito maior de estados possíveis, então, a Física Estatística supõe que essa será a propriedade final. É por isso que a Física Estatística transforma uma observação

probabilística (a propriedade mais provável tem probabilidade virtualmente igual a um) em princípio, que é exatamente o princípio enunciado no começo dessa seção.

Portanto, a Física Estatística é intrinsecamente diferente da física newtoniana. Na verdade, a natureza segue as leis de Newton. Portanto, as leis de Newton é que são as leis fundamentais (na verdade as leis fundamentais são as da mecânica quântica). As leis da Física Estatística, por outro lado, são observações gerais que utilizam noções de probabilidade para afirmar coisas gerais sobre sistemas que, apesar de evoluírem de acordo com as leis de Newton (ou da mecânica quântica), não podemos saber exatamente como se dá essa evolução (geralmente não podemos saber simplesmente porque o sistema é muito grande, ou seja, por questões práticas). Por “coisas gerais”, eu quero dizer que a Física Estatística lida com propriedades estatísticas do sistema. Por exemplo, ao invés de tentar determinar a localização de cada molécula de um líquido ou gás, a Física Estatística busca leis que permitam encontrar a densidade do líquido ou do gás. A densidade aqui é a coisa geral, pois a densidade te dá uma informação estatística sobre a distribuição de massa em um sistema, em vez de te dizer onde cada molécula está exatamente. Densidade é um parâmetro típico de Física Estatística.

Na próxima seção, vamos estudar um caso típico onde o sistema é formado por dois subsistemas que interagem entre si. Para isso, vamos colocar o princípio que o sistema possui a propriedade mais provável em linguagem matemática, o que nos levará naturalmente à ideia de equilíbrio térmico.

1.2 Equilíbrio Térmico

Suponha que tenhamos dois sistemas macroscópicos: o sistema A e o sistema B. Suponha que a energia do sistema total seja U_0 (o sistema total é a união do sistema A com o B). No início, esses sistemas estão isolados, e o sistema A possui energia U_A , e o sistema B energia U_B . Portanto, temos $U_0 = U_A + U_B$. Agora, coloquemos os dois sistemas em contato e os deixemos interagir. O que podemos afirmar sobre a propriedade do sistema após atingir equilíbrio?

Primeiro, note que podemos falar dos estados do sistema A, dos estados do sistema B e dos estados do sistema completo A + B. Por exemplo, vamos supor que tanto o sistema A quanto o sistema B sejam formados por duas moedas cada. Então, o sistema total será formado por quatro moedas. Assim, se o sistema A estiver no estado *cara-cara*, e o sistema B tiver no estado *cara-coroa*, então, o sistema total estará no estado *cara-cara-cara-coroa*. E, nesse estado, a energia do sistema A será zero, a energia do sistema B será um e a energia do sistema total será um. Se os sistemas interagem, então, é possível que os estados mudem, respeitando sempre a conservação de energia. Assim, por exemplo, vamos supor que o estado inicial do sistema seja *cara-cara-cara-coroa*. Se os sistemas interagem, depois de certo tempo, eu posso encontrar o sistema em qualquer outro estado que conserve a energia total (que nesse exemplo é $U_0 = 1$). Eu posso, por exemplo, encontrar o sistema A no estado *coroa-cara* e o sistema B no estado *cara-cara*, ou seja, o sistema total no estado *coroa-cara-cara-cara*.

Além disso, podemos utilizar o princípio básico que foi apresentado na seção anterior para obtermos mais informações. Esse princípio nos diz que a propriedade final do sistema é a mais provável. Para explorá-lo de maneira mais quantitativa, temos que começar a definir alguns parâmetros. Estamos interessados aqui principalmente nas energias dos sistemas A e B após equilíbrio. Portanto, os estados serão caracterizados por suas respectivas energias. Vamos chamar o número de estados possíveis com energia U_A do sistema A de $g_A(U_A)$. Note que o fato de a função g_A ser uma função de U_A expressa a dependência do número de estados na energia em questão. De maneira análoga, definimos $g_B(U_B)$ como sendo o número de estados com energia U_B do sistema B. No nosso exemplo anterior, o sistema A e o sistema B possuem duas moedas cada, portanto teremos que $g_A(0) = g_B(0) = 1$ (porque só existe um estado com energia zero, que é o estado *cara-cara*), $g_A(1) = g_B(1) = 2$ (porque existem dois

estados com energia 1, que são os estados *cara-coroa* e *coroa-cara*), e $g_A(2) = g_B(2) = 1$ (porque só existe um estado com energia 2, que é o estado *coroa-coroa*).

Estamos interessados em expressar a função g_t do estado total A + B em termos das multiplicidades de g_A e g_B . Em princípio, podemos expressar g_t como função de U_0 , ou de U_A , ou de U_B (note que a conservação de energia faz com que U_A , U_B e U_0 não sejam independentes; de fato $U_0 = U_A + U_B$). Mas seria meio bocado expressar g_t como função de U_0 , já que, em razão da conservação de energia, todos os estados possíveis do sistema total sempre terão energia U_0 ; então U_0 é uma constante do sistema, e não uma variável. Assim, faz muito mais sentido expressar g_t como função ou de U_A ou de U_B . Tanto faz qual a gente escolhe, então, eu vou escolher expressar g_t como função de U_A , utilizando a notação $g_t(U_A)$. Assim, **$g_t(U_A)$ é o número de estados do sistema total onde o subsistema A possui energia U_A** . Por exemplo, o estado *cara-cara-cara-coroa* é um estado do sistema total onde o subsistema A possui energia $U_A = 0$ (já que o estado do subsistema A nesse caso é *cara-cara*).

Para não nos perdemos na notação, vamos voltar novamente no exemplo dos sistemas A e B com duas moedas cada e vamos supor novamente que $U_0 = 1$. Se $U_0 = 1$, então temos apenas quatro estados possíveis do sistema total, que são: *cara-cara-cara-coroa*, *cara-cara-coroa-cara*, *cara-coroa-cara-cara* e *coroa-cara-cara-cara*.

No estado *cara-cara-cara-coroa* temos $U_A = 0$ (porque A está no estado *cara-cara*) e $U_B = 1$ (porque B está no estado *cara-coroa*).

No estado *cara-cara-coroa-cara* temos $U_A = 0$ (porque A está no estado *cara-cara*) e $U_B = 1$ (porque B está no estado *coroa-cara*).

No estado *cara-coroa-cara-cara* temos $U_A = 1$ (porque A está no estado *cara-coroa*) e $U_B = 0$ (porque B está no estado *cara-cara*).

No estado *coroa-cara-cara-cara* temos $U_A = 1$ (porque A está no estado *coroa-cara*) e $U_B = 0$ (porque B está no estado *cara-cara*).

Quais são, então, os valores que $g_t(U_A)$ podem assumir? Como U_A , neste exemplo, só possui dois valores possíveis (0 ou 1) então, temos apenas $g_t(0)$ e $g_t(1)$. Como temos dois estados do sistema total onde $U_A = 0$ (os estados *cara-cara-cara-coroa* e *cara-cara-coroa-cara*), então $g_t(0) = 2$. Como também só temos dois estados do sistema total onde $U_A = 1$ (os estados *cara-coroa-cara-cara* e *coroa-cara-cara-cara*), então $g_t(1) = 2$.

Note que existe apenas um estado em A cuja energia é zero (em notação matemática: $g_A(0) = 1$), que é o estado *cara-cara*. Mas como existem dois estados em B cujas energias são 1 ($g_B(1) = 2$), então, o estado *cara-cara* de A pode combinar com esses dois estados de B, resultando em $g_t(0) = 2$. Uma observação análoga serve para explicar $g_t(1) = 2$.

Para ter certeza de que você entendeu a ideia, determine os valores de $g_t(U_A)$ para o sistema A + B com $U_0 = 2$.

Agora, vamos generalizar essa ideia. Como podemos expressar, de maneira geral, o número de estados do sistema total A + B onde o sistema A tenha energia U_A ? Oras, porque energia é conservada, se o sistema A tiver energia U_A , então, necessariamente o sistema B deverá ter energia $U_B = U_0 - U_A$. Assim, suponha que existem $g_A(U_A)$ estados para o sistema A cuja energia seja U_A e $g_B(U_0 - U_A)$ estados para o sistema B cuja energia seja $U_B = U_0 - U_A$. Como cada um desses estados do sistema A (com energia U_A) pode combinar com cada um desses estados do sistema B (com energia U_B), então, temos $g_A(U_A) g_B(U_0 - U_A)$ combinações possíveis (que conservam energia) entre estes estados; ou seja, o número de estados do sistema total cujo subsistema A está em um estado com energia U_A será:

$$g_t(U_A) = g_A(U_A)g_B(U_0 - U_A)$$

Equação 12

Como veremos a seguir, a expressão matemática da condição de equilíbrio é construída em cima da Equação 12. Dessa forma, a função g desempenha um papel central em Física Estatística. Por ser tão importante, ela recebe um nome especial: a função g é chamada de multiplicidade do sistema, já que ela diz qual é o número de estados possíveis para uma determinada propriedade.

Até agora não fizemos nada demais, apenas utilizamos noções básicas de combinatória para expressar a multiplicidade do sistema total em termos da multiplicidade dos seus subsistemas, que chamamos de A e B. De fato, o que a Equação 12 nos diz é simplesmente que o número de estados possíveis para o sistema total onde a energia do sistema A é U_A (lado esquerdo da equação) é o tanto de estados possíveis de A nessa energia vezes o tanto de estados possíveis de B na energia complementar $U_B = U_0 - U_A$, já que cada um desses estados de A pode combinar com cada um

desses estado de B. Por exemplo, suponha, em um caso qualquer, que $U_0 = 8$, $U_A = 3$ e que queremos saber quantos estados do sistema total possuem o estado do subsistema A com energia $U_A = 3$ (ou seja, queremos determinar $g_t(3)$). Suponha que, nesse exemplo qualquer, existem quatro estados do sistema A que possuem energia $U_A = 3$. Isso quer dizer que $g_A(3) = 4$. Cada um desses quatro estados pode combinar com qualquer outro estado do sistema B, desde que energia seja conservada. Em outras palavras, qualquer um desses quatro estados pode combinar com qualquer outro estado do sistema B que tenha energia $U_B = U_0 - U_A = 8 - 3 = 5$. Agora, suponha que o sistema B possua dez estados com energia $U_B = 5$. Isso quer dizer que $g_B(5) = 10$. Como cada um dos quatro estados de A pode combinar com qualquer um dos dez estados de B, então, teremos 40 estados totais onde a energia do estado A é 3. Portanto $g_t(3) = g_A(3)g_B(8 - 3) = 40$.

Agora, podemos afirmar algo a mais sobre as propriedades do sistema? Claro que sim, porque o princípio básico nos diz que a propriedade do sistema no equilíbrio vai ser a mais provável (lembre-se que o princípio faz referência ao sistema total!). E qual é a propriedade mais provável? Uai, é a que tem o maior número de estados. Em outras palavras, conte quantos estados do sistema total possuem o subsistema A com uma determinada energia U_A ; faça isso para todos os valores de U_A : a propriedade mais provável é valor de U_A com mais estados. Então, em termos de $g(U_A)$, a propriedade U_A mais provável é dada pelo valor de U_A que maximiza a função g . Mas sabemos que um ponto de máximo é um ponto onde a variação de g em termos de U_A é zero, ou seja, onde a derivada de g em relação a U_A é zero:

$$\frac{dg}{dU_A} \Big|_{(U_A=U_f)} = 0$$

Equação 13

Aqui, chamamos o valor de U_A que maximiza g de U_f . A Equação 13 por si só não tem nada demais. Ela só está dizendo que U_f é o ponto que maximiza g . **Agora, se o sistema for macroscópico, sabemos que a propriedade final será praticamente igual a U_f .** Utilizando este princípio, basta encontrarmos U_f para determinarmos a propriedade final (de equilíbrio). Assim, derivando g_t em relação à U_A na Equação 12 temos (lembre-se que só podemos aplicar o princípio no sistema total, por isso derivamos g_t , e não g_A ou g_B):

$$\frac{dg_t}{dU_A} = \frac{dg_A(U_A)}{dU_A} g_B(U_0 - U_A) + g_A(U_A) \frac{dg_B(U_0 - U_A)}{dU_A}$$

Equação 14

Da Equação 13 temos, então, que:

$$\left[\frac{dg_A(U_A)}{dU_A} \Big|_{(U_A=U_f)} \right] g_B(U_0 - U_f) + g_A(U_f) \left[\frac{dg_B(U_0 - U_A)}{dU_A} \Big|_{(U_A=U_f)} \right] = 0$$

Equação 15

Para limpar um pouco a notação, vamos omitir a notação que explicita que o valor de U_A deve ser avaliado no ponto $U_A = U_f$. Assim, a Equação 15 ficaria:

$$\left[\frac{dg_A(U_A)}{dU_A} \right] g_B(U_0 - U_A) + g_A(U_A) \left[\frac{dg_B(U_0 - U_A)}{dU_A} \right] = 0$$

Além disso, vamos expressar a derivada de g_B em relação à U_A em termos da derivada de g_B em relação à U_B . Reescrevendo a Equação 15 em termos de U_A e U_B temos:

$$\frac{dg_A(U_A)}{dU_A} g_B(U_B) - g_A(U_A) \frac{dg_B(U_B)}{dU_B} = 0$$

Equação 16

Note que o sinal do segundo termo foi trocado porque agora a derivada de g_B é em relação à U_B , e não mais em relação à U_A (para expressar a derivada em relação à U_A basta utilizar a regra da cadeia, junto com a igualdade $U_A + U_B = U_0$). Como suprimimos a notação que explicita que o estado de energia U que satisfaz essa condição é o estado final, devemos ter cuidado para não esquecer que essa igualdade só é válida para o U_A e U_B , mais prováveis, ou seja, para o $U_A = U_f$ e $U_B = U_f - U_0$. Rearranjando a Equação 16 temos:

$$\frac{1}{g_A(U_A)} \frac{dg_A(U_A)}{dU_A} = \frac{1}{g_B(U_B)} \frac{dg_B(U_B)}{dU_B}$$

Equação 17

Lembrando que:

$$\frac{d \ln(f(x))}{dx} = \frac{df(x)}{dx} \frac{d \ln(f)}{df} = \frac{df(x)}{dx} \frac{1}{f(x)}$$

Ou seja:

$$\frac{1}{f(x)} \frac{df(x)}{dx} = \frac{d \ln(f(x))}{dx}$$

Então, a Equação 17 pode ser reescrita como:

$$\frac{d \ln(g_A(U_A))}{dU_A} = \frac{d \ln(g_B(U_B))}{dU_B}$$

Equação 18

Para facilitar a notação, podemos definir a variável σ como:

$$\sigma = \ln(g)$$

Equação 19

E reexpressar a Equação 18 em termos de σ . Assim, concluímos que, no equilíbrio, os sistemas A e B possuem propriedades que satisfazem:

$$\frac{d\sigma_A}{dU_A} = \frac{d\sigma_B}{dU_B}$$

Equação 20

A Equação 20 é um dos principais resultados da Física Estatística. Ela nos diz que no equilíbrio a derivada de σ em relação a U é igual nos dois sistemas. Note que essa relação é consequência direta da conservação de energia, da suposição fundamental da Física Estatística e também do princípio probabilístico de que a propriedade final é a mais provável. De certa forma, a Equação 20 pode ser interpretada como a expressão matemática do princípio probabilístico incorporando a conservação de energia.

A parte de física propriamente dita dessa seção termina na Equação 20. Mas sabemos que uma característica inerente às sociedades civilizadas é a de dar nomes a coisas importantes. Como vivemos em um mundo semicivilizado, temos, então, que dar nomes para coisas e expressá-las em unidades reconhecidas pela maioria dos países semicivilizados. A variável σ não tem nenhum nome muito especial, mas ela é utilizada para definir a variável S da seguinte forma:

$$S = k_B \sigma$$

Equação 21

Onde k_B é a famosíssima constante de Boltzmann. Em vez de nomear σ , preferiu-se nomear S como definido pela Equação 21: chamamos S de **entropia**. Podemos, então, reexpressar a condição de equilíbrio em termos da entropia:

$$\frac{dS_A}{dU_A} = \frac{dS_B}{dU_B}$$

Equação 22

Se alguém te perguntar o que ocorre quando dois sistemas são postos em contato, a resposta seria que, depois de um tempo (até atingir equilíbrio), a variação da entropia em relação à energia passa a ser igual nos dois sistemas. Horrível! Então, para facilitar ainda mais, é melhor dar um nome também para a derivada da entropia em relação à energia. A convenção é definir a variável T como:

$$\frac{1}{T} = \frac{dS}{dU}$$

Equação 23

Assim, no equilíbrio temos que $T_A = T_B$. É claro que a variável T ganha um nome também: chamamos a variável T como definida na Equação 23 de **temperatura**. Com essas definições, podemos dizer que a propriedade mais provável é a propriedade na qual a temperatura do sistema A é igual à temperatura do sistema B. Então, se alguém te perguntar o que acontece quando dois sistemas são postos em contato, a resposta é que a temperatura dos dois será a mesma após atingirem equilíbrio. E se alguém te perguntar por que isso ocorre, a resposta correta é: pela mesma razão que, se você jogar

um trilhão de moedas para o alto, metade cairá com a coroa para cima, ou seja, porque isso é o mais provável que o ocorra.

A Equação 23 é a definição mais fundamental de temperatura.

Nesta seção, chegamos à conclusão que quando dois sistemas interagem, o equilíbrio é atingido quando as temperaturas dos sistemas se igualam. Vimos como essa condição corresponde à propriedade mais provável do sistema, a qual determinamos encontrando o ponto de máximo da multiplicidade do sistema total. Portanto, tudo o que fizemos até aqui foi determinar qual é a propriedade mais provável do sistema total. Embora esse seja um resultado de suma importância em Física Estatística e Termodinâmica, muitas vezes precisamos de mais detalhes sobre os sistemas que interagem. Em particular, é importante determinar também qual é a probabilidade de encontrarmos cada um dos estados dos subsistemas. Esse é o assunto da próxima seção.

1.3 Função de partição - Parte I

Na seção anterior, tratamos os subsistemas A e B no mesmo pé de igualdade. Física Estatística, entretanto, lida frequentemente com sistemas em contato com um reservatório, que é um sistema gigante do qual temos algumas informações (geralmente a temperatura) e que é pouco afetado (por ser gigante) pelo sistema de interesse. Por exemplo, se quero estudar as propriedades térmicas de um gás em um laboratório, posso tratar o laboratório como sendo o reservatório, que define a temperatura ambiente (temperatura ambiente é sinônimo de temperatura do reservatório). Sendo assim, o caso mais comum que encontramos na prática é o caso onde o sistema A é um sistema de interesse, que queremos estudar, e o sistema B é um sistema gigante, do qual já conhecemos algumas de suas propriedades (como a temperatura) e que chamamos de reservatório.

Suponha, então, que temos um sistema de interesse em contato com algum reservatório. Como disse no parágrafo anterior, geralmente conhecemos o reservatório. Então, por razões práticas, seria interessante expressar algumas propriedades do sistema em termos de propriedades do reservatório. Uma relação muito útil e muito importante que podemos obter diretamente utilizando apenas conservação de energia é a razão entre as probabilidades de o sistema de interesse estar em determinados estados. Para explicar melhor o que quero dizer, imagine que a energia do sistema total (sistema A + reservatório) seja U_0 . Qual é, então, a probabilidade de eu encontrar o sistema A em um

determinado estado, que vou chamar de estado β ? Vamos supor também que o estado β tenha energia ε_β . Note que eu estou usando uma notação diferente para a energia: troquei o U por ε . Fiz essa troca para expressar o fato de que agora os sistemas têm tamanhos diferentes, então as energias dos estados do sistema A (denotadas por ε) são tipicamente bem menores que as energias dos estados do reservatório, que denotaremos por U . Note também que aqui estou olhando para um único estado do sistema A (estado este que chamei de β) e perguntando qual é a probabilidade de eu encontrar o sistema A nesse estado quando o sistema A está interagindo com o reservatório.

Essa pergunta é diferente da que tratamos na seção 1.1, onde perguntamos quantos estados possuem a mesma energia (lembrando que esse número de estados foi expresso pela função multiplicidade). Para ilustrar a diferença entre essas duas perguntas, vamos usar de novo o exemplo das moedas. Vamos supor novamente que uma moeda com coroa virada para cima corresponda a uma unidade de energia e com cara para cima à energia zero. Vamos supor que tenhamos duas moedas. Na seção 1.1 perguntamos quantos estados resultariam em uma determinada energia. A resposta seria, um estado para a energia 0 (o estado *cara-cara*), dois estados para a energia 1 (estados *cara-coroa* e *coroa-cara*) e um estado para a energia 2 (estado *coroa-coroa*). Isso foi o que fizemos na seção 1.1. E o que estamos perguntando agora? Agora, estamos perguntando qual é a probabilidade de eu encontrar o sistema A em um determinado estado (por exemplo, o estado *cara-cara*) quando o sistema A interage com o reservatório. Como o sistema A só possui quatro estados, talvez o seu primeiro pensamento seja que a probabilidade é de 25% para cada estado. Mas isso só seria verdade se o sistema A estivesse isolado do universo, ou seja, se o sistema A fosse o sistema total. Mas no nosso caso de interesse isso não é verdade, e não é verdade porque o sistema total inclui o reservatório, e isso significa que existem inúmeros estados possíveis, que resultam da combinação dos estados do sistema A com os estados do reservatório. Mesmo no nosso exemplo anterior, do sistema A e B com duas moedas cada um, a resposta não seria 25% para cada estado. Por exemplo, vimos que se $U_0 = 1$, então os estados totais possíveis são: *cara-cara-cara-coroa*, *cara-cara-coroa-cara*, *cara-coroa-cara-cara*, *coroa-cara-cara-cara*. Então, neste exemplo, se perguntarmos qual é a chance de eu encontrar o sistema A no estado *cara-cara*, a resposta seria 50%, porque dois dos quatro estados totais (os dois primeiros) possuem o sistema A no estado *cara-cara*. Por outro lado, a chance de eu encontrar o sistema A no estado *cara-coroa* é de 25% (porque só o terceiro

estado total possui o sistema A no estado *cara-coroa*) e no estado *coroa-cara* também é de 25%. Outros exemplos poderiam até dar resultados onde um estado tem probabilidade zero. Por exemplo, se $U_0 = 3$, então a probabilidade de eu encontrar o sistema A no estado *cara-cara* é zero (prove essa afirmação).

Vamos, agora, generalizar o cálculo da probabilidade de encontrar o sistema A em um determinado estado β . Como sabemos que energia é conservada, temos que, se a energia do sistema total (sistema A + reservatório) for U_0 , e se o sistema A estiver no estado β , e se este estado possui energia ε_β , então o reservatório tem que ter energia $U_0 - \varepsilon_\beta$. E quantos estados do sistema total são possíveis nos quais o sistema A esteja no estado β ? Se existem M estados do reservatório que podem combinar com o estado β do sistema A, então β aparecerá em M estados totais possíveis e, portanto, teremos M estados totais nos quais o sistema A estará no estado β . No exemplo das moedas, o estado total envolve as duas moedas do sistema A mais todas as moedas do reservatório. Mas quem é M ? **Como M é o número de estados do reservatório que podem combinar com o estado β respeitando conservação de energia, então M nada mais é do que o número de estados do reservatório que possuem energia $U_0 - \varepsilon_\beta$. Como resolvemos chamar o número de estados de multiplicidade, M nada mais é do que a multiplicidade do reservatório para a energia $U_0 - \varepsilon_\beta$. Assim, chamando a multiplicidade do reservatório de g_R , temos que $M = g_R(U_0 - \varepsilon_\beta)$.**

Voltando ao exemplo das moedas, vamos supor que U_0 seja igual a 1000 ($U_0 = 1000$), e que o reservatório possua dez mil moedas, e que o sistema A possua duas moedas. Para facilitar, vamos chamar o estado *cara-cara* do sistema A de estado α , o estado *cara-coroa* do sistema A de estado β , o estado *coroa-cara* do sistema A de estado ζ e o estado *coroa-coroa* do sistema A de estado κ . O sistema total possui então 10 mil + 2 moedas. Como o estado β possui energia 1 (já que esse é o estado *cara-coroa*), então M corresponderá ao número de estados **do reservatório** com energia igual a 999 (respeitando, assim, a energia total igual a 1000). Neste exemplo, para calcular M , basta calcular o número de maneiras que podemos arranjar as 10 mil moedas do reservatório com 999 coroas viradas para cima.

Mas M não é exatamente o que queremos. Queremos saber qual é a probabilidade de encontrarmos o sistema A no estado β . Bom, se tivermos ao todo N estados possíveis (respeitando conservação de energia) do sistema total, e se o estado β aparece em M

desses N estados, então obviamente a probabilidade será dada por M/N . Expressando M em termos da multiplicidade do reservatório, teremos:

$$P(\beta) = \frac{M}{N} = \frac{g_R(U_0 - \varepsilon_\beta)}{N}$$

Equação 24

Note que a expressão acima é válida tanto para sistemas pequenos como para sistemas grandes. A informação que o reservatório é grande ainda não foi utilizada, portanto a relação acima não depende do tamanho do reservatório.

Note também que, no exemplo das moedas, se perguntássemos qual é a probabilidade de encontrar o sistema A no estado β , a resposta seria $P(\beta)$, mas se perguntássemos qual é a probabilidade de encontrar o sistema A com energia igual a 1, então a resposta seria $P(\beta) + P(\zeta)$, porque tanto o estado β como o estado ζ possuem energia igual a 1. É muito importante manter essa diferença em mente. Note também que a probabilidade depende somente da energia do estado (a única informação sobre o estado do sistema A que aparece na equação acima é a sua energia ε_β). Portanto, no exemplo das moedas teríamos também que $P(\beta) = P(\zeta)$ (essa igualdade é válida para quaisquer pares de estados com a mesma energia).

Mas temos um problema: é praticamente impossível calcular N diretamente, exatamente porque o reservatório vai ser um sistema gigantesco. Para nos livrarmos de N , vamos, então, calcular qual é a razão entre a probabilidade de encontrar o sistema em um estado β e a probabilidade de encontrar o sistema em um estado α :

$$\frac{P(\beta)}{P(\alpha)} = \frac{g_R(U_0 - \varepsilon_\beta)}{g_R(U_0 - \varepsilon_\alpha)}$$

Equação 25

A Equação 25 é particularmente útil porque estamos expressando propriedades do sistema A (lado esquerdo da equação) em termos de propriedades do reservatório (lado direito da equação). Mas podemos melhorar ainda mais. Já sabemos qual é a relação entre a multiplicidade e a entropia do sistema (Equação 19 e Equação 21). Podemos, então, reescrever a Equação 25 em termos da entropia:

$$\frac{P(\beta)}{P(\alpha)} = \frac{\exp[\sigma_R(U_0 - \varepsilon_\beta)]}{\exp[\sigma_R(U_0 - \varepsilon_\alpha)]} = \frac{\exp\left[\frac{S_R(U_0 - \varepsilon_\beta)}{k_B}\right]}{\exp\left[\frac{S_R(U_0 - \varepsilon_\alpha)}{k_B}\right]}$$

Rearranjando a relação acima:

$$\frac{P(\beta)}{P(\alpha)} = \exp\left[\frac{S_R(U_0 - \varepsilon_\beta) - S_R(U_0 - \varepsilon_\alpha)}{k_B}\right]$$

Equação 26

Para deixar o argumento da exponencial em uma forma mais agradável, podemos expandir S em uma série de Taylor em torno de U_0 :

$$S_R(U_0 - \varepsilon) = S_R(U_0) - \varepsilon \frac{dS_R}{dU} + \frac{\varepsilon^2}{2} \frac{d^2S}{dU^2} + \dots$$

Equação 27

Sabemos que a derivada da entropia é o inverso da temperatura (Equação 23). Então, o segundo termo da série é inversamente proporcional à temperatura. Já os termos de mais alta ordem envolvem derivadas da derivada da entropia, ou seja, envolvem derivadas da temperatura. Portanto, os termos de mais alta ordem dependem de variações da temperatura em relação à energia. Como aqui estamos tratando de um reservatório, que é gigante, temos que a variação da temperatura é desprezível (é bem mais fácil aquecer uma xícara de café do que o planeta Terra). Portanto, por se tratar de um reservatório, podemos descartar os termos de ordem mais alta que 1. Assim, temos que:

$$S_R(U_0 - \varepsilon) \cong S_R(U_0) - \varepsilon \frac{dS_R}{dU} = S_R(U_0) - \varepsilon \frac{1}{T}$$

Equação 28

Lembre-se que T é a temperatura do reservatório. Entretanto, como estamos supondo que o sistema A está em equilíbrio térmico com o reservatório, então T será também a temperatura do sistema A. Esse ponto não é muito enfatizado em livros textos, mas é importante lembrar que T é, a priori, a temperatura do reservatório e só podemos

considerá-la como sendo também a temperatura do sistema A se ambos estiverem em equilíbrio térmico.

Utilizando a expansão da Equação 28 temos:

$$S_R(U_0 - \varepsilon_\beta) - S_R(U_0 - \varepsilon_\alpha) = \left[S_R(U_0) - \varepsilon_\beta \frac{1}{T} \right] - \left[S_R(U_0) - \varepsilon_\alpha \frac{1}{T} \right] = -\frac{\varepsilon_\beta - \varepsilon_\alpha}{T}$$

Equação 29

Substituindo a Equação 29 na Equação 26, temos finalmente que:

$$\frac{P(\beta)}{P(\alpha)} = \exp \left[-\frac{\varepsilon_\beta - \varepsilon_\alpha}{k_B T} \right] = \frac{\exp \left[-\frac{\varepsilon_\beta}{k_B T} \right]}{\exp \left[-\frac{\varepsilon_\alpha}{k_B T} \right]}$$

Equação 30

Vamos parar um pouco para meditar sobre esse resultado. Primeiro, vamos lembrar quais foram as informações que utilizamos para chegar nesse resultado. A única lei da física que utilizamos nessa seção foi a lei de conservação de energia. Mas note que a Equação 24 só é verdadeira se todos os estados do sistema total (sistema A + reservatório) tiverem igual probabilidade. Nenhum estado do sistema total pode ser mais provável que outro (caso contrário, a Equação 24 não seria verdade e toda a dedução iria por água abaixo). Portanto, estamos usando novamente a **suposição fundamental da Física Estatística** (seção 1.1). Se ela é sempre verdade ninguém sabe. Mas como estamos lidando com sistemas dos quais não sabemos tudo, somos forçados a supor algumas coisas para podermos tirar conclusões gerais; e a suposição que todos os estados (do sistema total) são igualmente prováveis permite que toda a teoria seja desenvolvida. Como a Física Estatística é um dos campos do conhecimento mais completos e precisos ao descrever o comportamento de sistemas macroscópicos, podemos concluir, com certa confiança, que essa é uma boa suposição.

Deduzimos, portanto, que a razão entre as probabilidades de encontrar o sistema em um estado β (cuja energia é ε_β) e um estado α (cuja energia é ε_α) depende só dos valores de ε_β , ε_α , e da temperatura do reservatório. Como a temperatura do reservatório também é a temperatura do sistema quando ambos estão em equilíbrio térmico, tudo o

que precisamos saber para achar a razão entre as probabilidades é a temperatura do sistema (ou reservatório). Note que não especificamos quase nada sobre o sistema A; portanto, a Equação 30 é extremamente abrangente e vale para praticamente qualquer sistema em contato com um reservatório. De fato, a Equação 30 é uma obra de arte: ela nos permite tirar conclusões importantíssimas utilizando apenas parâmetros macroscópicos (neste caso, a temperatura) do sistema.

Bom, mas ainda falta o toque final. Conseguimos expressar a razão entre as probabilidades, mas seria ainda melhor termos uma expressão para a probabilidade mesmo. Mas não é muito difícil utilizar a informação que temos sobre a razão entre as probabilidades para encontrar a probabilidade absoluta. Pense o seguinte: apesar de termos deduzido a Equação 30 para os estados β e α , poderíamos ter escolhido qualquer par de estados. Portanto, a razão entre as probabilidades de quaisquer pares é sempre exponencial. Então, a probabilidade absoluta deve ser proporcional à exponencial, ou seja:

$$P(c) = \frac{\exp\left[-\frac{\epsilon_c}{k_B T}\right]}{Z}$$

Equação 31

Onde Z é um número e o índice c foi usado para designar um estado qualquer. Para calcular Z , basta lembrarmos que a soma das probabilidades sobre todos os estados possíveis deve ser igual a 1. Em linguagem matemática, temos que:

$$\sum_c P(c) = 1$$

Equação 32

Aqui, a soma a respeito do índice c corresponde a uma soma em relação a todos os estados do sistema A. No exemplo das duas moedas, c percorreria os estados α , β , ζ ou κ . Portanto haveria quatro termos na soma, ou seja, a Equação 32 ficaria: $P(\alpha) + P(\beta) + P(\zeta) + P(\kappa) = 1$.

Voltando ao caso geral, da Equação 31 temos, então, que:

$$\sum_c P(c) = \frac{\sum_c \exp\left[-\frac{\epsilon_c}{k_B T}\right]}{Z} = 1$$

Equação 33

Portanto, temos que:

$$Z = \sum_c \exp\left[-\frac{\epsilon_c}{k_B T}\right]$$

Equação 34

Como o número Z , na verdade, depende da temperatura, é mais conveniente expressá-lo como uma função dela. Fazendo isso, temos:

$$Z(T) = \sum_c \exp\left[-\frac{\epsilon_c}{k_B T}\right]$$

Equação 35

A função Z é de extrema importância em Física Estatística e recebe o nome de **função de partição**. Na verdade, ela não está ainda completa (por isso que o título dessa seção é “Função de Partição, parte I”). Como veremos a seguir, a função de partição também depende do potencial químico, que é o tópico da próxima seção.

1.4 Equilíbrio difusivo e potencial químico

Toda a teoria desenvolvida até aqui foi baseada na lei de conservação de energia e na suposição de que todos os estados do sistema total são igualmente prováveis. Mas se você notar bem, fizemos uma suposição adicional: supomos que a multiplicidade é função apenas da energia. E pudemos fazer isso porque, em nenhum momento, consideramos que pudesse haver troca de elementos entre os sistemas. De fato, até aqui consideramos que os sistemas podem trocar energia (desde que a energia total seja conservada), mas ainda não levamos em consideração um caso onde, além de energia, os sistemas pudessem também trocar elementos.

No exemplo das moedas, estávamos interessados na multiplicidade em função da energia (representada pelo número de coroas viradas para cima) e fizemos toda a análise para um número fixo de moedas. Demos exemplos com diferentes números de moedas, mas nunca consideramos um caso onde dois sistemas poderiam trocar moedas entre si. Mas situações onde há troca de elementos entre sistemas são bem comuns. Por exemplo, poderíamos ter dois gases da mesma molécula, em dois compartimentos diferentes, cada um com uma energia e número de moléculas diferentes. Aí, poderíamos deixar os sistemas interagirem termicamente e poderíamos também abrir uma porta entre os dois compartimentos para que houvesse, também, troca de moléculas. Assim, faria sentido associar aos estados do sistema A (gás no compartimento A) tanto a propriedade “energia”, como a propriedade “número de moléculas”. Claro que isso valeria também para o sistema B e para o reservatório.

O que muda nesse caso? Pouca coisa. Praticamente todo o raciocínio que desenvolvemos até aqui ainda é válido; só precisamos estendê-lo um pouquinho para incluir a variável “número de elementos”. Em linguagem matemática, a multiplicidade, agora, é função de duas variáveis: a energia U e o número de elementos N (esses elementos poderiam ser átomos, moléculas, elétrons, etc.). Todo o raciocínio mostrado nas seções 1.1 e 1.2 ainda é válido. Em particular, a **Equação 22** ainda é válida, porque ainda estamos procurando o ponto de máximo de g , e uma das condições que devem ser satisfeitas no ponto de máximo é que a derivada parcial em relação a U é zero, o que leva ao resultado da **Equação 22** (com a diferença que a derivada deve ser substituída pela derivada parcial). Mas agora, temos uma segunda condição que também deve ser satisfeita no ponto de máximo: a derivada parcial em relação à variável N também tem que ser zero, o que leva naturalmente a um resultado análogo ao da **Equação 22**, mas com N no lugar de U (lembre-se que também utilizamos conservação de energia para chegar na **Equação 22**; então, para chegarmos em uma equação análoga para N teremos que utilizar a conservação do número de elementos, e é óbvio que é verdade que o número de elementos é conservado: estamos supondo que há transferência de elementos de um sistema ao outro, portanto o número total de elementos é constante). Assim, a condição de máximo é satisfeita quando ambas as relações abaixo são satisfeitas:

$$\left(\frac{\partial S_A}{\partial U_A}\right)_{N_A} = \left(\frac{\partial S_B}{\partial U_B}\right)_{N_B} \quad e \quad \left(\frac{\partial S_A}{\partial N_A}\right)_{U_A} = \left(\frac{\partial S_B}{\partial N_B}\right)_{U_B}$$

Equação 36

Os símbolos N e U , colocados no canto direito inferior de cada derivada parcial, servem somente para lembrar que a derivada parcial é tomada com a outra variável fixa. Então, na primeira igualdade, a entropia é variada em relação à energia e essa variação se dá com o número de elementos fixo, enquanto na segunda igualdade, a entropia é variada em relação ao número de elementos, mas com a energia fixa. Portanto, estamos tratando a energia e o número de elementos como variáveis independentes. Uma das principais dificuldades em termodinâmica é identificar quais são as variáveis independentes, mas não podemos entrar nessa questão com detalhes aqui. Lembre-se que a entropia, agora, é função da energia e do número de elementos (U e N). Daqui em diante, vamos omitir o subíndice para limpar um pouco a notação.

Se demos um nome para a derivada parcial que aparece na primeira igualdade da Equação 36 (lembre-se que é essa derivada define a temperatura), é justo darmos um nome para a derivada que aparece na segunda igualdade; afinal de contas, essa derivada também é um parâmetro que é igualado quando os dois sistemas entram em equilíbrio. A convenção é definir o parâmetro μ como:

$$\mu = -T \frac{\partial S}{\partial N}$$

Equação 37

O parâmetro μ , definido na Equação 37, é chamado de **potencial químico**. Note que o potencial químico é função da temperatura T . A maneira mais simples de se ter uma intuição sobre o que exatamente esse parâmetro significa é utilizar a intuição que temos sobre a relação entre temperatura e energia. Note que o potencial químico foi definido em relação à variável N de maneira análoga à definição de temperatura: enquanto a temperatura foi definida em termos da derivada da entropia em relação à U , o potencial químico foi definido em termos da derivada em relação à N . Então, tudo o que sabemos intuitivamente sobre a relação entre temperatura e energia pode ser estendido para a relação entre o potencial químico e o número de elementos. Por exemplo, sabemos que se dois sistemas possuem temperaturas diferentes, então, energia irá fluir do sistema com maior temperatura para o sistema com menor temperatura (afinal de contas, as temperaturas têm que se igualar no equilíbrio). Analogamente, **se dois sistemas**

possuem potenciais químicos diferentes, então elementos irão difundir do sistema com potencial químico maior para o sistema com potencial químico menor.

No equilíbrio térmico, as temperaturas são igualadas. No equilíbrio difusivo, os potenciais químicos são igualados. Se tivermos tanto equilíbrio térmico como difusivo, então tanto a temperatura quanto o potencial químico são igualados. Além disso, sabemos que o equilíbrio térmico é atingido através da troca de energia (que muda a temperatura). Do mesmo modo, equilíbrio difusivo é atingido através da troca de elementos (que muda o potencial químico). Portanto, do mesmo jeito que a temperatura define quem envia energia para quem, o potencial químico define quem envia elementos para quem. Dessa forma, um material com potencial químico alto é um material que irá difundir elementos para outro material com potencial químico mais baixo. Com tudo isso, não é difícil de imaginar que o potencial químico de um gás deva depender da concentração do gás, já que sabemos que um gradiente de concentração resulta em difusão.

Agora, por que esse termo chama potencial? Ele recebe esse nome porque desempenha um papel praticamente idêntico ao de outros potenciais em física. Por exemplo, se houver uma diferença de potencial gravitacional, vai haver deslocamento de partículas (elas caem). Esse tipo de deslocamento é associado com o que chamamos de corrente de arrasto. Portanto, o potencial gravitacional gera uma corrente de arrasto. De modo análogo, o potencial químico também gera uma corrente, só que a corrente é de difusão (a corrente de difusão e a corrente de arrasto serão tratadas com mais detalhes na seção 1.8).

Não é difícil verificar que o potencial químico tem unidades de energia. Para isso, basta expressar a variação da entropia em termos de suas derivadas parciais; de cálculo, sabemos que, se S é função de U e N , então:

$$\Delta S = \frac{\partial S}{\partial U} \Delta U + \frac{\partial S}{\partial N} \Delta N$$

Equação 38

Utilizando as definições de temperatura e potencial químico, teremos:

$$\Delta S = \frac{1}{T} \Delta U - \frac{\mu}{T} \Delta N$$

Equação 39

Isolando a variação de energia, temos:

$$\Delta U = T\Delta S + \mu\Delta N$$

Equação 40

Como ΔN é adimensional (já que o número de partículas não tem dimensão), fica claro por essa relação que μ tem unidade de energia.

Agora, considere a seguinte situação: vamos supor que você tenha dois compartimentos idênticos em contato difusivo e você coloque uma quantidade de certo gás em um deles e espera um pouco. Essa situação está representada na Figura 5a. Você já sabe que, no equilíbrio, os potenciais químicos serão iguais (e, portanto, as concentrações serão iguais). Conseqüentemente, no equilíbrio, o fluxo (ou corrente) de elementos de um lado para o outro é zero.

Agora, vamos supor que você coloque os mesmos compartimentos, só que na vertical, como representado na Figura 5b. O que vai ocorrer? Agora, os elementos no compartimento superior possuem energia gravitacional maior que os elementos no compartimento inferior. Essa diferença de energia potencial gravitacional vai ocasionar um fluxo (uma corrente de arrasto) de elementos de cima para baixo, aumentando o potencial químico dos elementos do compartimento de baixo. Mas se o potencial químico dos elementos de baixo aumenta, aparece um fluxo por difusão (uma corrente de difusão) de baixo para cima. Dessa forma, o sistema entrará em equilíbrio quando a corrente de arrasto de cima para baixo (ocasionada pela diferença de potencial gravitacional) for igual à corrente de difusão de baixo para cima (ocasionado pela diferença de potencial químico), resultando em uma corrente total nula! Em outras palavras, o sistema entrará em equilíbrio quando o potencial total (químico + gravitacional) for igual nos dois compartimentos. **Portanto, quem define o fluxo total é o potencial total e o equilíbrio é atingido quando o potencial total for igualado!** Note que, no equilíbrio, nem o potencial gravitacional nem o químico são iguais nos dois compartimentos. De fato, o compartimento de cima tem potencial gravitacional maior e o de baixo tem potencial químico maior. Mas o potencial total (químico + gravitacional) é igual em ambos no equilíbrio.

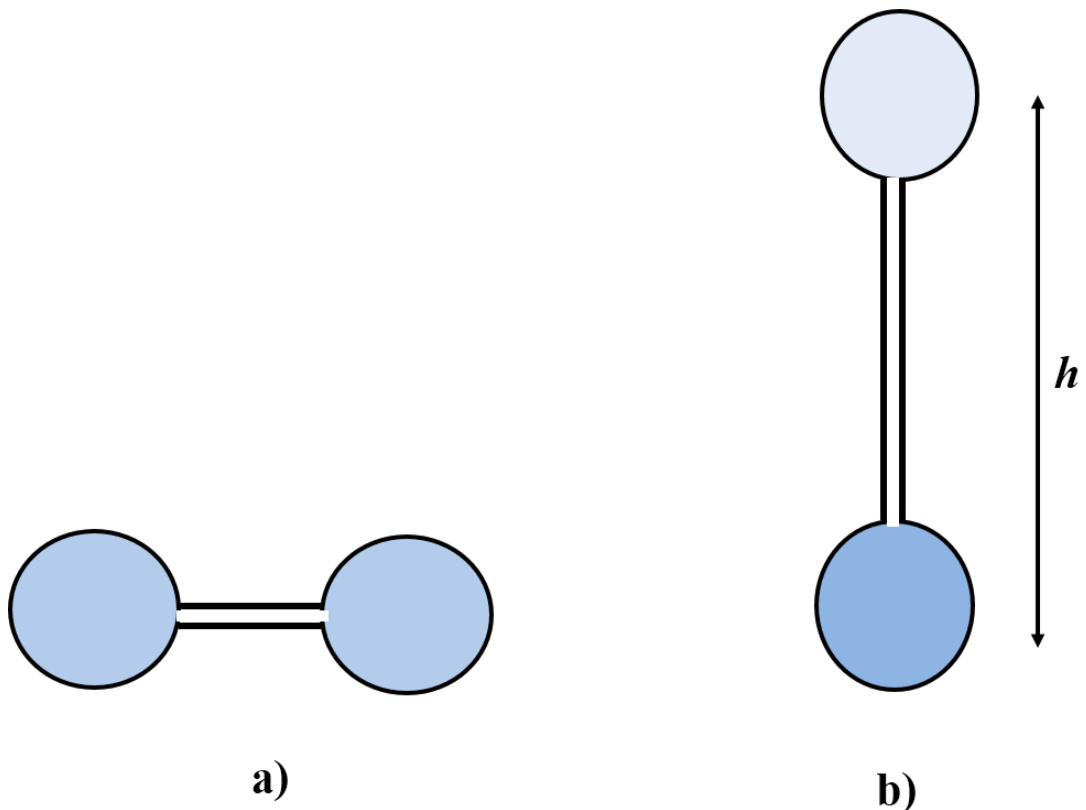


Figura 5 Relação do potencial químico com outros potenciais. a) dois compartimentos em equilíbrio difusivo terão a mesma concentração de gases. Portanto, o potencial químico em ambos os lados é o mesmo. b) Os mesmos dois compartimentos, agora, são colocados em alturas diferentes. Isso significa que um elemento no compartimento superior possui uma quantidade de energia mgh maior que um elemento no compartimento inferior. Essa diferença diminui a concentração do gás na parte superior e aumenta na parte inferior, causando uma diferença de potencial químico entre os dois compartimentos. Essa diferença de potencial químico equilibra a diferença de potencial gravitacional, de maneira que, agora, o potencial total é o mesmo nos dois compartimentos. Consequentemente, a corrente de arrasto causada pela força gravitacional é compensada pela corrente de difusão causada pela diferença de potencial químico e, no equilíbrio, a corrente total é nula.

Então, vamos recapitular. Utilizando apenas a condição de maximização da entropia, chegamos à conclusão que o potencial químico, expresso pela Equação 37, é constante em dois sistemas em contato difusivo no equilíbrio. Essa situação é exemplificada pela Figura 5a. Mas, na situação da Figura 5b, o potencial químico não é constante, apesar de o fluxo total ser nulo. Como reconciliar o exemplo da Figura 5b com o resultado geral de que a derivada da entropia em relação ao número de elementos (ou seja, o potencial químico) ter que ser constante no equilíbrio (lembre-se que essa é a condição que maximiza a multiplicidade, ou seja, é a condição mais provável)?

Na verdade, o potencial μ , que aparece na Equação 37, é sempre o potencial total (químico + gravitacional + eletrostático + nuclear e por aí vai.). Por isso, que a

Equação 37 descreve tanto a situação da Figura 5a quanto a situação da Figura 5b, mas na Figura 5a o potencial total é igual ao potencial químico, portanto, a derivada resulta no potencial químico, enquanto na Figura 5b o potencial total é o potencial químico mais o potencial gravitacional e, conseqüentemente, a derivada resulta no potencial químico + gravitacional.

Para mostrar isso de maneira um pouco mais formal, vamos dar nome aos bois. Vamos chamar o potencial químico de μ_q , o potencial gravitacional de μ_g e o potencial total de μ_t . Como no exemplo da Figura 5a não existe nenhuma força externa (ou seja, não existe nenhum potencial além do potencial químico), a Equação 40 já descreve a situação da Figura 5a, ou seja:

$$\Delta U = T\Delta S + \mu_q\Delta N$$

na Figura 5a

Rearranjando a fórmula acima para expressar a variação da entropia em termos das variações de energia e do número de partículas:

$$\Delta S = \frac{\Delta U}{T} - \frac{\mu_q}{T}\Delta N$$

Como por cálculo sabemos que:

$$\Delta S = \frac{\partial S}{\partial U}\Delta U + \frac{\partial S}{\partial N}\Delta N$$

Chegamos à conclusão que:

$$\mu_t = \mu_q = -T \frac{\partial S}{\partial N}$$

na Figura 5a

Como anteriormente (Equação 37).

Agora, como podemos adaptar a Equação 40 para incorporar o potencial gravitacional? A resposta é fácil: no compartimento inferior não muda nada (porque o

potencial gravitacional aí ainda é zero), mas no superior temos que incorporar o potencial gravitacional. Fazer isso é bem fácil, basta notar que se eu trazer ΔN partículas para o compartimento superior, então o potencial gravitacional vai contribuir com um aumento de energia igual à energia potencial gravitacional de cada uma dessas ΔN moléculas. Portanto, para o compartimento superior, temos:

$$\Delta U = T\Delta S + \mu_q\Delta N + mgh\Delta N = T\Delta S + (\mu_q + mgh)\Delta N = T\Delta S + (\mu_q + \mu_g)\Delta N$$

na Figura 5b

onde mgh é a energia potencial de uma única molécula no compartimento superior. Assim, se você fizer o mesmo procedimento de reorganizar a expressão acima para expressar a variação da entropia em relação às variações de energia e número de partículas, vai chegar facilmente à conclusão que:

$$\mu_t = \mu_q + \mu_g = -T \frac{\partial S}{\partial N}$$

na Figura 5b

Assim, sempre que algum potencial externo for incluído em sistemas em contato difusivo, o equilíbrio será atingido quando o potencial total for igualado nos sistemas. E esse é o resultado que já esperávamos pela análise da Figura 5.

Note que poderíamos ter mais de um elemento. Neste caso, cada elemento teria o seu próprio potencial químico e o equilíbrio seria atingido para cada elemento individualmente (e a entropia seria função da energia e dos números de cada elemento; por exemplo, se tivéssemos dois elementos diferentes, a entropia seria função de três variáveis, duas para os elementos e uma para a energia). Matematicamente, a entropia seria função da energia U , do número de elementos N_1 , do número do outro elemento N_2 , e por aí vai.

Agora, vamos ver outro exemplo, de peculiar importância para o funcionamento de dispositivos eletrônicos. De fato, o exemplo seguinte pode ser interpretado como sendo um protótipo de um diodo.

Imagine duas situações, representadas na Figura 6a e na Figura 6b. Em ambos os casos, o sistema total é composto de dois sistemas: o A e o B. Os sistemas são feitos do

mesmo material e têm as mesmas dimensões. Ambos estão conectados por um pequeno afunilamento.

Agora, imagine que você aumenta o potencial químico do sistema B. Você pode fazer isso injetando algum elemento ou molécula nele. Imagine que, neste caso, você aumentou o potencial químico aumentando a concentração de elétrons livres. Você não precisa carregar o material para fazer isso. Por exemplo, você pode simplesmente aumentar a concentração de elétrons livres através de alguma reação química que mantenha o sistema neutro. Vamos supor, então, que você aumentou a concentração de elétrons livres sem carregar o material (ele continua neutro). Então, você aumentou o potencial químico dos elétrons livres desse material (você passa, agora, a ter maior concentração de elétrons livres no material B do que no material A). Vamos supor, então, que o potencial químico do sistema A seja μ_A , e potencial químico do sistema B seja igual a $\mu_A + \mu_0$. O que vai acontecer? Obviamente, a diferença de potencial químico vai fazer com que elétrons difundam pelo sistema.

Então, agora você sabe que, como o sistema B tem potencial químico maior, então os elétrons vão difundir de B para A. Mas vamos supor que você observa a difusão no começo da interação, bem antes do potencial químico modificar significativamente em razão da difusão. Ou seja, para facilitar a discussão, vamos supor que o potencial químico não altere mesmo com a difusão. Então, temos que o potencial químico em A é constante e igual a μ_A , enquanto em B é constante e igual a $\mu_A + \mu_0$. Bom, os elétrons estão difundindo do material B para o material A. Portanto, existe um fluxo de elétrons de B para A. Uai, se existe um fluxo de cargas, então, existe uma corrente elétrica. Se você puser um voltímetro entre os dois materiais vai medir uma diferença de potencial eletrostático (uma tensão). Vamos supor que a tensão medida foi V_0 (medida com o terminal positivo do voltímetro no sistema A), como mostrado na Figura 6a.

Agora, congela esse experimento. Separe os materiais e os deixe quietos por enquanto. Pegue o segundo sistema, idêntico ao primeiro e representado na Figura 6b. Mas, ao invés de modificar o potencial químico de B, coloque uma diferença de potencial eletrostático entre A e B de $-V_0$. Então, agora, temos que tanto o potencial químico do sistema A, quanto o de B são idênticos (e iguais a μ_A); mas, agora, existe uma diferença de potencial eletrostático imposta externamente a eles. Óbvio que, se a diferença de potencial observada for $-V_0$, vai haver uma corrente de B para A e essa corrente terá a mesma magnitude da corrente que você mediu no experimento da Figura

6a, já que o módulo da tensão é o mesmo. Mas, agora, a corrente que você mediu vai estar na direção oposta da anterior (elétrons vão, agora, de A para B, ou seja, a corrente é de B para A). Então, as correntes do experimento da Figura 6a e da Figura 6b possuem os mesmos módulos, mas direções opostas. Note também que as correntes são de natureza diferente: enquanto a corrente da Figura 6a é uma corrente de difusão, a corrente da Figura 6b é uma corrente de arrasto (mais detalhes sobre as diferenças correntes de difusão e arrasto serão abordados na seção 1.8).

O que vai acontecer se você pegar essa bateria, ajustada em V_0 , e a colocar do mesmo jeito entre o sistema B e A, mas agora com o potencial químico do sistema B novamente igual a $\mu_A + \mu_0$, como mostrado na Figura 6c?

O que vai ocorrer é que a corrente de elétrons de B para A, devido à diferença de potencial químico (que gera a corrente de difusão), será compensada pela corrente induzida pela diferença de potencial eletrostático (que gera a corrente de arrasto). Assim, a corrente total será zero e, conseqüentemente, a tensão medida pelo voltímetro também será zero, mesmo com a bateria conectada (na verdade, por causa mesmo da bateria, já que sem ela teríamos a situação da Figura 6a).

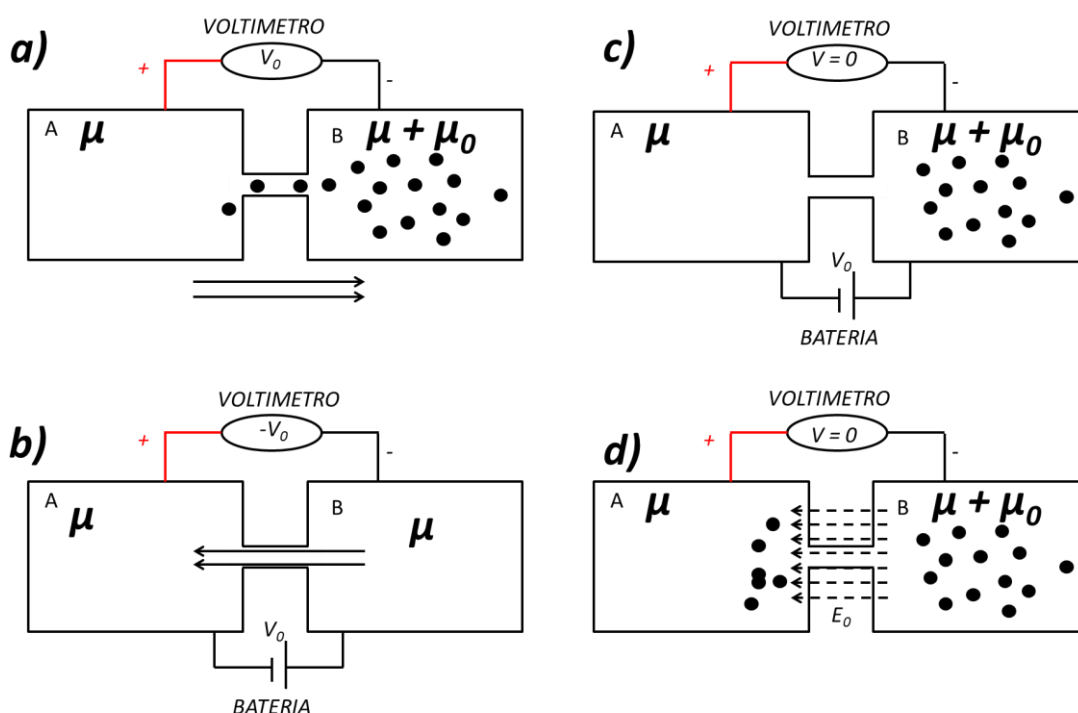


Figura 6 Situações ilustrativas do significado de potencial químico. Em a), aumentamos o potencial químico dos elétrons livres de B, o que induziu uma corrente e conseqüentemente uma diferença de potencial V_0 (medido com o terminal positivo em A). A direção da corrente é indicada pelas setas (lembre-se que a corrente é na direção contrária ao fluxo de elétrons). Em b), o potencial químico é o mesmo em ambos os sistemas mas aplicamos uma bateria com diferença de potencial $-V_0$ entre os terminais A e B. Como os sistemas são idênticos, a corrente induzida pela bateria é a mesma que a induzida pelo potencial químico, mas com direção oposta. Em c),

conectamos a bateria e mantivemos a diferença de potencial químico. Neste caso, a corrente será nula, assim como a tensão medida pelo voltímetro. Em d), voltamos ao caso inicial e deixamos a corrente fluir. A corrente carrega o sistema A com elétrons, ao mesmo tempo em que carrega o sistema B positivamente (pois agora faltam elétrons) gerando um campo elétrico que aponta de B para A (representado pelas flechas pontilhadas). Esse campo elétrico gera uma força nos elétrons (lembre-se que a força é contrária à direção do campo elétrico). Esse campo elétrico cresce até que o potencial eletrostático compense o potencial químico μ_0 , resultando no mesmo potencial eletroquímico para os sistemas A e B.

Que conclusão podemos tirar desse experimento *gedanken*? A conclusão é que a diferença de potencial químico μ_0 tem o mesmo efeito da diferença de potencial eletrostático V_0 . Os dois geraram a mesma intensidade de corrente (apenas com direções opostas) e a mesma diferença de potencial no voltímetro! Então, do ponto de vista elétrico (ou seja, no que concerne a corrente e a tensão que aparecem nos terminais), as situações da Figura 6a e Figura 6b são idênticas, a não ser pela direção da corrente (e pela natureza da corrente, uma de difusão e a outra de arrasto). Isso significa que o potencial químico tem exatamente o mesmo efeito da bateria (de fato, uma bateria química nada mais é que um componente que mantém uma diferença de potencial químico através de reações químicas). Temos só uma diferença de unidades: enquanto a bateria é expressa em volts, que é joules por coulomb, o potencial químico tem unidade de energia.

Portanto, a interpretação correta para esse experimento é que a bateria fornece a mesma energia que o potencial químico, o que podemos expressar matematicamente como:

$$\mu_0 = |q|V_0$$

Equação 41

Onde $|q|$ é o módulo da carga fundamental do elétron.

Se quisermos descrever a situação da Figura 6c, em termos dos potenciais totais, teremos que os potenciais totais nos sistemas A e B serão:

$$\mu_{At} = \mu_A \text{ e } \mu_{Bt} = \mu_A + \mu_0 - |q|V_0$$

$$\text{mas como } \mu_0 = |q|V_0$$

$$\therefore \mu_{At} = \mu_{Bt}$$

na Figura 6c

O sinal de negativo na expressão $-|q|V_0$ acima advém do fato de que estamos tratando da energia (em Joules) de elétrons, que tem carga negativa.

Onde μ_{At} e μ_{Bt} são os potenciais totais dos sistemas A e B (note que o zero de referência para a energia potencial eletrostática foi colocado no sistema A e lembre-se que o potencial positivo da bateria em B indica que a energia potencial eletrostática dos elétrons é menor em B que em A, já que a carga dos elétrons é negativa). A igualdade nos potenciais totais denota a situação de equilíbrio.

O experimento *gedanken* da Figura 6c é bem análogo ao experimento dos gases da Figura 5b e a conclusão que chegamos é a mesma (ainda bem): o fluxo (corrente) de elementos em ambos os lados vai cessar quando o potencial total for o mesmo em ambos os sistemas (ou seja, constante ao longo do sistema total). A diferença é que na Figura 5b tínhamos potenciais químico e gravitacional, enquanto na Figura 6c temos potenciais químico e eletrostático. Além disso, em ambos os casos o potencial que aparece além do potencial químico é externo aos sistemas: o gravitacional é resultado da presença do planeta Terra e o eletrostático é devido à bateria e, tanto Terra, quanto bateria, não fazem parte dos sistemas analisados (por isso dizemos que são potenciais externos).

Mas existe uma diferença importantíssima entre um exemplo com elementos sem carga (como na Figura 5) e um exemplo com elementos com carga (como na Figura 6). E essa diferença é ilustrada pelo experimento *gedanken* da Figura 6d. Na Figura 6d, retiramos a bateria e observamos o fluxo de elétrons. Agora, não existe mais nenhum potencial externo e, portanto, poderíamos esperar que o fluxo cessasse quando os potenciais químicos se igualarem. Mas não é bem isso que ocorre porque, à medida que elétrons vão do sistema B para o sistema A, o excesso de elétrons em A e a falta em B acarretarão uma concentração local de cargas negativas em A e positivas em B. Note que o sistema total continua neutro, mas o sistema A é carregado negativamente e o B positivamente. Este acúmulo de cargas locais, por sua vez, acarretará a formação de um campo elétrico interno, semelhante ao caso de um capacitor carregado, como indicado na Figura 6d. Como temos, agora, um campo elétrico interno, temos também uma diferença de potencial eletrostático interna! Portanto, temos, agora, que levar em consideração o fato de que o movimento de elétrons pode causar diferença de potencial eletrostático interna devido ao fato de elétrons possuírem cargas.

Note que a diferença de potencial eletrostático interna vai gerar uma corrente de arrasto na direção oposta à corrente de difusão (para verificar isso, basta notar que a

força que o campo elétrico impõe nos elétrons é no sentido de A para B). Essa corrente cresce à medida que mais cargas acumulam em ambos os lados (o que aumenta o campo elétrico), até que a corrente de arrasto iguale a corrente de difusão, fazendo com que a corrente total seja zero.

Em termos de energia, a corrente cessará quando a energia potencial eletrostática associada ao campo elétrico igualar a diferença entre os potenciais químicos. Agora, temos um caso bem parecido com o que tínhamos quando pusemos a bateria com a voltagem ajustada para V_0 (Figura 6c). A diferença é que, agora, é o campo elétrico interno (que apareceu entre os sistemas como consequência da própria difusão) que faz o papel que a bateria externa fazia na Figura 6c. Obviamente, o potencial eletrostático que aparece devido ao campo elétrico é igual a $-V_0$ (medido entre A e B). Entretanto, se você, agora, colocar um voltímetro entre os sistemas A e B, não vai medir nada, porque agora você tem $+V_0$ devido ao potencial químico, e $-V_0$, devido ao potencial eletrostático (associado ao campo elétrico). Portanto, a diferença de potencial total é zero, igualzinho ao caso da Figura 6c! Esse último exemplo, da Figura 6d, é importantíssimo para a eletrônica, porque ele fundamenta o mecanismo de equilíbrio em uma junção **p-n**, como veremos nos capítulos seguintes.

Assim, chegamos à conclusão que, caso o elemento seja carregado, a difusão acarretará em criação de um potencial eletrostático interno. Em todo caso, independente do potencial ser interno ou externo, o processo será sempre encerrado quando o potencial total for constante, como já sabíamos.

Talvez você esteja estranhando a situação da Figura 6d, afinal de contas temos uma diferença de potencial eletrostática devido ao campo elétrico interno, mas se eu puser um voltímetro entre os terminais, a leitura será zero! Por que o voltímetro não consegue ler essa diferença de potencial eletrostático? Simples. Porque o voltímetro sempre lê a diferença de potencial total, seja ela eletrostática ou química (e se elétrons possuíssem massa alta teríamos que levar o potencial gravitacional em consideração também). Note que, no exemplo da Figura 6a, o voltímetro lê uma diferença de potencial, mas ainda não existe nenhum campo elétrico, portanto o voltímetro está lendo uma diferença de potencial em um caso onde a diferença de potencial eletrostático é nula (não tem campo elétrico em lugar nenhum)! Isso ocorre porque existe uma diferença de potencial química. Por outro lado, na Figura 6d temos tanto uma diferença de potencial

eletrostático quanto uma diferença de potencial química, mas uma anula a outra, e a diferença de potencial total é zero!

Faz sentido, então, falar de potencial químico separado de potencial eletrostático e vice-versa? Do ponto de vista elétrico, não faz muito sentido, já que ambos agem juntos. Portanto, estaremos sempre interessados no potencial eletroquímico, que é o potencial químico e eletrostático juntos, ou seja, o potencial total. Podemos definir, então, o potencial eletroquímico para elétrons como:

$$\mu_{eq} = \mu - |q|V$$

Equação 42

Onde *eq* significa eletroquímico (sinal de negativo advém do fato da carga do elétron ser negativa). Note que, se a força da gravidade desempenhar um papel significativo no sistema, teremos que incorporá-la também:

$$\mu_{total} = \mu - |q|V + mgh$$

Equação 43

E por aí vai. Em alguns casos, o potencial eletrostático é nulo e o gravitacional é importante, como em difusão por osmose, por exemplo. Para nós, o gravitacional nunca vai ser relevante, por isso que podemos focar somente no potencial eletroquímico.

Um problema que causa enorme confusão na literatura de semicondutores é que os símbolos para potenciais químico e eletroquímico geralmente são os mesmos. Alguns livros chamam o potencial químico de potencial interno e o potencial eletroquímico de potencial total, mas isso também varia. Em todo caso, o importante é lembrar que o potencial químico desempenha o papel de um potencial no mesmo patamar do potencial eletrostático. Assim, estamos quase sempre interessados no potencial total porque, afinal de contas, é o potencial total que define a direção da corrente. De fato, no nosso exemplo, a difusão cessou quando a energia potencial eletrostática $-|q|V$ cancelou o potencial químico μ_0 , o que significa que o potencial total do sistema B passou a ser igual ao potencial total do sistema A.

Talvez você esteja se perguntado: por que então o potencial químico ganha um nome diferente do potencial eletrostático, que ganha um nome diferente do potencial gravitacional, já que todos eles agem juntos? A resposta é simples: porque a origem dos

potenciais é diferente. A origem do potencial gravitacional é a força gravitacional. A origem do potencial eletrostático é a força de Coulomb (ou seja, o campo elétrico). A origem do potencial químico é a “força” de difusão (força entre aspas porque não existe uma força física neste caso, apenas a agitação térmica, que causa difusão). A origem do potencial químico não é uma força física, mas sim a variação na concentração dos elementos, embora o efeito seja o mesmo. No fundo, o potencial químico está associado à energia cinética das partículas.

Note que nosso conceito e uso dos potenciais não têm absolutamente nada de diferente do conceito de potencial em física newtoniana. Por exemplo, se você tiver um corpo carregado, ele estará sujeito tanto ao potencial gravitacional quanto ao potencial eletrostático. Você sabe também que o corpo ficará em repouso se os dois potenciais se balancearem (como uma bola de boliche carregada que levita devido à presença de um campo elétrico). Ambos causam forças, ambos agem juntos, mas as origens deles são diferentes!

Resumindo: daqui para frente, lembre-se sempre que μ pode representar tanto o potencial químico como o potencial total do sistema, mas que é o potencial total que define a dinâmica de transporte. Portanto, daqui em diante, utilizaremos o símbolo μ para designar o potencial total.

1.5 Função de Partição, Parte II

Como agora os estados têm como propriedade tanto a energia como o número de elementos, a dedução da Função de Partição apresentada na seção 1.3 deve ser estendida para incluir a variável N . O raciocínio é praticamente idêntico ao da seção 1.3. Em particular, a razão entre as probabilidades de se encontrar um sistema A , quando este está em contato térmico e difusivo com um reservatório, em um estado β com energia ε_β e número de elementos n_β , e um estado α com energia ε_α e número de elementos n_α é:

$$\frac{P(\beta)}{P(\alpha)} = \frac{g_R(U_0 - \varepsilon_\beta, N_0 - n_\beta)}{g_R(U_0 - \varepsilon_\alpha, N_0 - n_\alpha)}$$

Equação 44

Lembre-se que g_R é a multiplicidade do reservatório, e não a do sistema. Além disso, do mesmo jeito que U_0 representa a energia do sistema total (sistema A +

reservatório), N_0 representa o número de elementos do sistema total (sistema A + reservatório).

O procedimento para encontrar a função de partição e a probabilidade $P(c)$ de um estado qualquer c é análogo ao da seção 1.3. Primeiro, reescrevemos Equação 44 em termos da entropia:

$$\begin{aligned} \frac{P(\beta)}{P(\alpha)} &= \frac{\exp\left[\frac{S_R(U_0 - \varepsilon_\beta, N_0 - n_\beta)}{k_B}\right]}{\exp\left[\frac{S_R(U_0 - \varepsilon_\alpha, N_0 - n_\alpha)}{k_B}\right]} \\ &= \exp\left[\frac{S_R(U_0 - \varepsilon_\beta, N_0 - n_\beta) - S_R(U_0 - \varepsilon_\alpha, N_0 - n_\alpha)}{k_B}\right] \end{aligned}$$

Equação 45

Agora, expandimos novamente S_R em uma série de Taylor, só que agora temos que utilizar a expansão em série de Taylor para duas variáveis. Retendo apenas os termos de primeira ordem temos:

$$S_R(U_0 - \varepsilon, N_0 - n) \cong S_R(U_0, N_0) - \varepsilon \frac{\partial S_R}{\partial U} - n \frac{\partial S_R}{\partial N}$$

Equação 46

Utilizando as definições da Equação 23 e da Equação 37, podemos reescrever a Equação 46 em termos da temperatura e potencial:

$$S_R(U_0 - \varepsilon, N_0 - n) = S_R(U_0, N_0) - \frac{\varepsilon}{T} + \mu \frac{n}{T}$$

Equação 47

Substituindo a Equação 47 na Equação 45 temos:

$$\frac{P(\beta)}{P(\alpha)} = \frac{\exp\left[\frac{-(\varepsilon_\beta - n_\beta \mu)}{k_B T}\right]}{\exp\left[\frac{-(\varepsilon_\alpha - n_\alpha \mu)}{k_B T}\right]}$$

Equação 48

Novamente, para encontrar o valor absoluto da probabilidade temos que achar a função de partição, que é a soma das exponenciais sobre todos os **estados** do sistema A (que é o sistema de interesse), que está em contato térmico e difusivo com o reservatório. Para diferenciar os estados, adotamos novamente o subscrito c . Dessa forma, a soma sobre todos os estados será:

$$\zeta(T, \mu) = \sum_c \exp \left[\frac{-(\varepsilon_c - n_c \mu)}{k_B T} \right]$$

Equação 49

Note que eu alterei o símbolo da soma. Antes era Z (veja Equação 34) e agora é ζ . Tive que mudar o símbolo porque a soma da Equação 49 recebe um nome diferente, só para complicar a vida mesmo. Apesar de continuar sendo a função de partição, a Equação 49 é chamada de soma de Gibbs. Não satisfeitas com o excesso de nomes, algumas pessoas também dão a essa soma a épica alcunha de “soma grandiosa” (*grand sum* em inglês), provavelmente porque a soma inclui todos os estados do sistema A . Note também que, enquanto a função Z era função apenas de T , a soma de Gibbs é função de T e μ , mas não esqueça que μ também depende de T (veja a Equação 37).

Finalmente, temos que a probabilidade de o sistema A estar em um dado estado c cuja energia é ε_c e cujo número de elementos é n_c é dada por:

$$P(c) = \frac{\exp \left[\frac{-(\varepsilon_c - n_c \mu)}{k_B T} \right]}{\zeta(T, \mu)}$$

Equação 50

Note, portanto, que a probabilidade de o sistema estar em um estado c depende da energia ε_c do estado, do número de elementos n_c do estado, e da temperatura e do potencial do reservatório (lembre-se que apesar de T e μ serem a temperatura e potencial do reservatório, no caso de equilíbrios térmico e difusivo teremos, por consequência direta, o fato de que T e μ serão também a temperatura e potencial do sistema). Finalmente, note que as informações sobre o sistema em si (É um gás? Um sólido? Qual gás? Qual sólido?) entram na soma de Gibbs, pois ela depende dos estados do sistema A e suas respectivas propriedades (energia e número de elementos).

1.6 Energia e número de elementos de um sistema

A Equação 50 é o resultado mais importante que obtivemos até agora. Ela nos diz que, se pegarmos um sistema em equilíbrio térmico e difusivo com um reservatório (lembre-se que o reservatório é qualquer sistema grande), então a probabilidade de encontrar o sistema em um estado com energia ε e número de elementos n é dada pela Equação 50. Mas qual é efetivamente o estado de um sistema em equilíbrio com um reservatório? A resposta da Física Estatística a essa pergunta é: não sei! Mas sei que a probabilidade de o sistema estar em um determinado estado c cuja energia é ε_c e cujo número de elementos é n_c é dada pela Equação 50.

Então, façamos desse limão uma limonada. Se não podemos determinar exatamente qual é o estado do sistema (e, conseqüentemente quais são as suas propriedades, como a energia), vamos usar o que sabemos para calcular parâmetros estatísticos do sistema (honremos, afinal de contas, o nome Física Estatística). O parâmetro mais comum em Física Estatística é a média.

Vamos supor que você tenha um sistema que seja composto por uma única partícula presa à uma caixa, e que essa partícula esteja em contato térmico com o reservatório. Esse exemplo “sistema uma partícula presa em uma caixa” é formalmente equivalente ao sistema “um gás monoatômico em um recipiente fechado e em equilíbrio térmico com o meio ambiente”. Como a partícula está em contato térmico com o reservatório, cada vez que você olhar para o sistema de interesse, você vai encontrar ele em um estado diferente. Como você poderia calcular média das energias que você observou no seu sistema? Vamos supor que você faça N observações no seu sistema “uma partícula na caixa”, e que você anote quantas vezes cada estado c foi observado nessas N observações. Denotando por N_c a quantidade de vezes que o estado c foi observado, e por ε_c a energia do estado c , então a soma $\sum_c \varepsilon_c N_c$, efetuada para todos os estados do sistema “uma partícula na caixa”, corresponde à soma de todas as energias observadas nessas N observações (note também que $\sum_c N_c = N$). Se eu quiser calcular a média da energia do meu sistema “uma partícula na caixa”, basta eu somar todas as energias observadas e dividir pelo número de observações, ou seja:

$$\langle U \rangle = \frac{\sum_c \varepsilon_c N_c}{N}$$

Na equação acima, os colchetes indicam que estamos pegando a média da variável entre os colchetes. Portanto, o lado esquerdo deve ser lido como “a média de U ”. Agora, quanto mais medições eu fizer, mais a razão N_c/N se aproxima da probabilidade de eu encontrar o estado c . Em termos matemáticos:

$$P(c) = \lim_{N \rightarrow \infty} \frac{N_c}{N}$$

Assim, se eu fizer muitas medidas, eu poderei expressar a média das energias que eu observei no sistema como:

$$\langle U \rangle = \sum_c \varepsilon_c P(c)$$

Equação 51

Precisamos então saber qual é a probabilidade de eu encontrar um determinado estado c de um sistema em contato com um reservatório. Mas nós já sabemos que essa probabilidade é dada pela Equação 31 (nesse exemplo, como a partícula está presa em uma caixa fechada, o sistema interage só termicamente com o reservatório; se o sistema também estivesse em contato difusivo, então $P(c)$ seria dado pela Equação 50). Assim, temos que:

$$\langle U \rangle = \sum_c \varepsilon_c P(c) = \frac{\sum_c \varepsilon_c \exp\left[-\frac{\varepsilon_c}{k_B T}\right]}{Z}$$

Equação 52

O que falta, então, para acharmos U ? Faltam duas coisas relacionadas: precisamos achar quais são os estados possíveis e precisamos saber quais são os seus níveis de energia. Se soubermos isso, poderemos calcular tanto a função de partição como a soma da Equação 52. O cálculo dos estados possíveis e níveis de energia é um problema típico em mecânica quântica e é completamente impossível nos aprofundarmos nessa questão fora de um curso dedicado à mecânica quântica. Felizmente, para o nosso exemplo “uma única partícula em uma caixa”, nós já temos a resposta: como a partícula é

representada por uma função de onda, então podemos utilizar a Equação 8 para determinar os estados e níveis de energia possíveis para o nosso sistema.

O único porém é que a Equação 8 foi deduzida para uma cavidade unidimensional. Uma cavidade propriamente dita, entretanto, tem que ser tridimensional. Felizmente, o raciocínio para obter a relação de energia para uma cavidade tridimensional é praticamente idêntico ao caso unidimensional. A única diferença é que requeremos que a fase acumulada em uma volta seja igual a um múltiplo de 2π para cada dimensão independentemente. Assim, temos que, se a função de onda tridimensional for dada por:

$$\psi(x, y, z) = \exp[i(k_x x + k_y y + k_z z)]$$

Equação 53

Então, as condições de acúmulo de fase para as três dimensões independentemente serão:

$$k_x = \frac{c_x \pi}{L_x}, k_y = \frac{c_y \pi}{L_y}, k_z = \frac{c_z \pi}{L_z}$$

Equação 54

Onde c_x, c_y e c_z , são números inteiros positivos, e L_x, L_y e L_z são as dimensões da caixa onde o elemento está preso. Para facilitar, vamos supor que a caixa seja um cubo, ou seja, $L_x = L_y = L_z = L$. Nesse caso, temos, obviamente:

$$k_x = \frac{c_x \pi}{L}, k_y = \frac{c_y \pi}{L}, k_z = \frac{c_z \pi}{L}$$

Equação 55

Como o momento continua sendo proporcional à constante de propagação, a única diferença aqui do que foi discutido para o caso unidimensional é que o momento, agora, é um vetor com três componentes, ou seja:

$$\vec{p} = \frac{hk_x}{2\pi} \hat{x} + \frac{hk_y}{2\pi} \hat{y} + \frac{hk_z}{2\pi} \hat{z}$$

Equação 56

Aqui, utilizamos a notação que flecha representa vetor e “^” representa vetor unitário. Independentemente da dimensão, a energia cinética é sempre relacionada ao módulo quadrado do vetor momento por:

$$E = \frac{|\vec{p}|^2}{2m}$$

Equação 57

De onde concluímos que:

$$E = \frac{h^2}{8L^2m} (c_x^2 + c_y^2 + c_z^2)$$

Equação 58

Que é a versão em três dimensões da Equação 8. Cada combinação c_x, c_y e c_z de números naturais (não inclui o zero) corresponde a um estado diferente. Note também que diferentes estados podem ter a mesma energia. Por exemplo, a combinação $c_x = 1, c_y = 1, c_z = 2$ corresponde a um estado com energia igual a $6h^2/(8L^2m)$, enquanto a combinação $c_x = 1, c_y = 2, c_z = 1$ corresponde a um outro estado, mas também com energia igual a $6h^2/(8L^2m)$.

Assim, temos que a função de partição será:

$$Z = \sum_c \exp \left[-\frac{\varepsilon_c}{k_B T} \right] = \sum_{c_x=1}^{\infty} \sum_{c_y=1}^{\infty} \sum_{c_z=1}^{\infty} \exp \left[-\frac{\frac{h^2}{8L^2m} (c_x^2 + c_y^2 + c_z^2)}{k_B T} \right]$$

Equação 59

Note que o somatório deve envolver todas as combinações possíveis de c_x, c_y e c_z , ou seja, todas as combinações de números inteiros positivos indo de um a infinito. Essa conta será feita mais adiante (seção 2.3, Equação 133 à Equação 134). Por enquanto, para não quebrar o fluxo do raciocínio, vou colocar o resultado direto:

$$Z = \sum_{c_x=1}^{\infty} \sum_{c_y=1}^{\infty} \sum_{c_z=1}^{\infty} \exp \left[-\frac{\frac{h^2}{8L^2m} (c_x^2 + c_y^2 + c_z^2)}{k_B T} \right] = \frac{L^3}{\left(\frac{h^2}{2\pi m k_B T} \right)^{3/2}}$$

Equação 60

Note que L^3 é o volume da caixa. Portanto, a função de partição para o sistema “um elemento na cavidade” depende só do volume da caixa, da massa do elemento e da temperatura.

Podemos, agora, calcular a média da energia do elemento:

$$\langle U \rangle = \frac{1}{Z} \sum_{c_x=1}^{\infty} \sum_{c_y=1}^{\infty} \sum_{c_z=1}^{\infty} \frac{h^2}{8L^2m} (c_x^2 + c_y^2 + c_z^2) \exp \left[-\frac{\frac{h^2}{8L^2m} (c_x^2 + c_y^2 + c_z^2)}{k_B T} \right]$$

Equação 61

Onde a função de partição Z é dada pela Equação 60. O resultado desse somatório é:

$$\langle U \rangle = \frac{3}{2} k_B T$$

Equação 62

Portanto, a média da energia de um átomo em um caixa é $3/2 k_B T$. Esse é um resultado famoso: você já deve ter ouvido falar em algum curso que a energia por átomo de um gás ideal é $3/2 k_B T$ (ou, equivalentemente, que a energia de um gás ideal é $3/2 k_B TN$, onde N é o número de átomos do gás). A prova é essa aí.

Podemos facilmente generalizar o raciocínio para incluir contato difusivo, além do térmico. Para isso basta utilizarmos a soma de Gibbs em vez da função de partição. Assim, se considerarmos um sistema em contato tanto térmico como difusivo com um reservatório, então as médias do número de elementos e da energia seriam dadas respectivamente por:

$$\langle N \rangle = \sum_c n_c P(c) = \sum_c n_c \frac{\exp \left[\frac{-(\epsilon_c - n_c \mu)}{k_B T} \right]}{\zeta(T, \mu)}$$

Equação 63

e

$$\langle U \rangle = \sum_c \varepsilon_c P(c) = \sum_c \varepsilon_c \frac{\exp \left[\frac{-(\varepsilon_c - n_c \mu)}{k_B T} \right]}{\zeta(T, \mu)}$$

Equação 64

Para encerrar essa seção, vamos fazer um resumo da nomenclatura que temos utilizado, conectando o problema fictício de sistemas compostos por moedas com o problema real de sistemas compostos por elétrons. É importante você não confundir estado do elétron com o estado do sistema. Nos exemplos das moedas, uma única moeda representa um único elétron e o estado da moeda é definido pelo fato de ela estar com cara ou coroa virada. Se tivermos mais de uma moeda, o estado do sistema vai ser definido pela combinação dos estados de todas as moedas. Por exemplo, se tivermos duas moedas, uma no estado cara e outra no estado coroa, então o estado **do sistema** será *cara-coroa*. Analogamente, se tivermos um elétron em um poço quântico, vimos que os seus estados são designados pela combinação de números naturais c_x, c_y, c_z . Cada trinca de números naturais define um estado do elétron (assim como cara ou coroa define o estado de uma moeda). Mas, se o sistema tiver mais de um elétron, o estado do sistema será definido pela combinação dos estados de cada elétron. Por exemplo, se o sistema tiver dois elétrons, então o estado do sistema será denotado pela combinação de duas trincas de números naturais. Por exemplo, se um elétron estiver no estado $c_x = 1, c_y = 1, c_z = 1$, e o outro elétron estiver no estado $c_x = 2, c_y = 1, c_z = 2$, então o estado do sistema será definido pela combinação 1, 1, 1, 2, 1, 2 (note que, nesse exemplo, o número de elementos do sistema é 2 e a energia do sistema é dada pela soma da energia do elétron no estado $c_x = 1, c_y = 1, c_z = 1$, com a energia do elétron no estado $c_x = 2, c_y = 1, c_z = 2$).

1.7 A distribuição de Fermi-Dirac

Agora, vamos focar nossa atenção em um problema de altíssima relevância para eletrônica. Nosso objetivo agora é determinar a probabilidade de um elétron ocupar um determinado estado, quando o sistema possui uma quantidade macroscópica de elétrons. Para não confundir estado do elétron com estado do sistema, vamos utilizar o termo “orbital” para designar o estado do elétron. Sendo assim, o termo “orbital” denota os estados possíveis dos elétrons, ou seja, cada uma das trincas (cada trinca é um orbital).

Para isso considere novamente o sistema poço quântico tridimensional. Como no fundo estamos interessados em dimensões macroscópicas (por exemplo, da ordem de centímetros), vou chamar o poço de “caixa”. Nosso principal interesse nesse momento é conseguir visualizar como a função multiplicidade desse sistema é montada. Lembre-se que a energia e número de partículas do sistema de interesse vão ser altas, de maneira que elas podem ser tratadas como funções contínuas. Mas, para pegarmos uma intuição de como a multiplicidade é montada, vamos considerar alguns casos exemplos onde a energia e o número de partículas são pequenos. Além disso, nessa análise, para simplificar, eu vou ignorar o *spin*.

Então lembrando: a função multiplicidade g é uma função de duas variáveis: da energia U e do número de partículas N : $g(U, N)$. Já sabemos também quais são os estados possíveis do sistema caixa: são os estados caracterizados pelas trinca de números naturais c_x, c_y, c_z . Como resolvemos chamar cada trinca de orbital, podemos dizer a energia do orbital c_x, c_y, c_z é dada por:

$$U_{c_x, c_y, c_z} = \frac{h^2}{8L^2m} (c_x^2 + c_y^2 + c_z^2)$$

Sendo assim, qual é o valor de $g(U_{1,1,1}, 1)$? Essa pergunta se reduz à pergunta: de quantas maneiras podemos encontrar o sistema com um único elétron e energia total igual à $U_{1,1,1}$. A resposta é: uma só, quando esse elétron estiver no orbital $c_x = 1, c_y = 1, c_z = 1$. Qualquer outro orbital em que o elétron estiver vai ter energia diferente. Portanto $g(U_{1,1,1}, 1) = 1$. Para ficar mais organizado, vou deixar explícito qual é o único estado do sistema que satisfaz a condição $U = U_{1,1,1}, N = 1$:

$$g(U_{1,1,1}, 1) = 1$$

Estados possíveis:

$$(1,1,1)$$

Analogamente, se tivermos $U = U_{1,1,2}, N = 1$, teremos três possibilidades. Explícitamente:

$$g(U_{1,1,2}, 1) = 3$$

Estados possíveis:

(1,1,2)

(1,2,1)

(2,1,1)

Agora vamos pegar um caso com dois elétrons. Por exemplo, temos que $g(U_{1,1,1}, 2) = 0$, porque não existe nenhum estado com dois elétrons e cuja energia total é a energia de um único elétron no orbital com nível mais baixo de energia. Analogamente, também temos que $g(2U_{1,1,1}, 2) = 0$, porque para ter a energia $2U_{1,1,1}$ teríamos que ter dois elétrons no orbital $c_x = 1, c_y = 1, c_z = 1$, o que, de acordo com princípio de exclusão de Pauli, não é possível (se estivéssemos considerando o spin, cada trinca seria associada à dois orbitais, e assim seria possível dois elétrons com a trinca 1, 1, 1; e, portanto, teríamos $g(2U_{1,1,1}, 2) = 1$).

Vamos dar mais um exemplo, agora para o caso $U = U_{1,1,1} + U_{1,1,2}, N = 2$. Nesse caso teremos:

$$g(U = U_{1,1,1} + U_{1,1,2}, N = 2) = 3$$

Estados possíveis:

(1,1,1; 1,1,2)

(1,1,1; 1,2,1)

(1,1,1; 2,1,1)

Onde o estado $(A; B)$ deve ser lido como “um elétron no orbital A e o outro elétron no orbital B ”. Por exemplo, o estado $(1,1,1; 1,1,2)$ é o estado “um elétron no orbital 1,1,1 e o outro elétron no orbital 1,1,2. Note que o estado $(A; B)$ é idêntico ao estado $(B; A)$. Isso é consequência do fato de elétrons serem indistinguíveis. Por exemplo, na lista colocamos o estado $(1,1,1; 1,1,2)$, mas não colocamos o estado $(1,1,2; 1,1,1)$, porque este último é idêntico ao primeiro.

Note que estamos sendo consistentes com a nomenclatura, e denotando o estado de um único elétron pelo termo “orbital”, e reservando o termo “estado” para o estado do sistema (ou seja, a combinação de trincas).

Para fixar a ideia, vamos dar mais um exemplo, agora para: $U = 2U_{1,1,2}$, $N = 2$:

$$g(2U_{1,1,2}, 2) = 6$$

Estados possíveis:

(1,1,2; 1,2,1)

(1,1,2; 2,1,1)

(1,2,1; 2,1,1)

(1,1,1; 1,2,2)

(1,1,1; 2,1,2)

(1,1,1; 2,2,1)

Confira que em todos os esses estados a energia é $U = 2U_{1,1,2}$. Note que, em razão do princípio de exclusão de Pauli, não incluímos estados com o mesmo par de orbitais (tipo (1,1,2; 1,2,1)).

Esses exemplos ilustram como montar os estados possíveis para cada caso e assim obter a função multiplicidade para todas as combinações de energia e número de partículas: basta fazer a lista para todos os estados possíveis para cada combinação de energia e número de partículas.

Agora, vamos modificar a notação um pouco. Quando tínhamos um elétron, colocamos o orbital desse elétron dentro do parêntese. Quando tínhamos dois elétrons, colocamos os dois orbitais, separados por ponto e vírgula. Se tivéssemos três elétrons, teríamos três orbitais separados por ponto em vírgula. Mas ao invés de fazer isso, vamos colocar um super parênteses e encher ele com todos os orbitais que existem; assim:

$$(1,1,1; 1,1,2; 1,2,1; 2,1,1; 1,2,2; 2,1,2; 2,2,1; 1,1,3; \dots)$$

Onde as reticências “...” devem ser lidas como “e por aí vai, até o infinito”. Dessa forma, podemos identificar o estado simplesmente colocando a letra o quando o orbital estiver ocupado, e a letra d quando o orbital estiver desocupado. Por exemplo, o estado (1,1,1) seria representado assim:

$$(1,1,1) \leftrightarrow \begin{pmatrix} o & d & d & d & d & d & d & d & d \\ 1,1,1 & 1,1,2 & 1,2,1 & 2,1,1 & 1,2,2 & 2,1,2 & 2,2,1 & 1,1,3 & \dots \end{pmatrix}$$

Outro exemplo:

$$(1,1,2; 1,2,1) \leftrightarrow \begin{pmatrix} d & o & o & d & d & d & d & d & d \\ 1,1,1 & 1,1,2 & 1,2,1 & 2,1,1 & 1,2,2 & 2,1,2 & 2,2,1 & 1,1,3 & \dots \end{pmatrix}$$

Sendo assim, com essa notação, a lista de estados possíveis para a combinação $U = 2U_{1,1,2}$, $N = 2$ ficaria:

$$g(2U_{1,1,2}, 2) = 6$$

Estados possíveis:

$$\begin{pmatrix} d & o & o & d & d & d & d & d & d \\ 1,1,1 & 1,1,2 & 1,2,1 & 2,1,1 & 1,2,2 & 2,1,2 & 2,2,1 & 1,1,3 & \dots \end{pmatrix}$$

$$\begin{pmatrix} d & o & d & o & d & d & d & d & d \\ 1,1,1 & 1,1,2 & 1,2,1 & 2,1,1 & 1,2,2 & 2,1,2 & 2,2,1 & 1,1,3 & \dots \end{pmatrix}$$

$$\begin{pmatrix} d & d & o & o & d & d & d & d & d \\ 1,1,1 & 1,1,2 & 1,2,1 & 2,1,1 & 1,2,2 & 2,1,2 & 2,2,1 & 1,1,3 & \dots \end{pmatrix}$$

$$\begin{pmatrix} o & d & d & d & o & d & d & d & d \\ 1,1,1 & 1,1,2 & 1,2,1 & 2,1,1 & 1,2,2 & 2,1,2 & 2,2,1 & 1,1,3 & \dots \end{pmatrix}$$

$$\begin{pmatrix} o & d & d & d & d & o & d & d & d \\ 1,1,1 & 1,1,2 & 1,2,1 & 2,1,1 & 1,2,2 & 2,1,2 & 2,2,1 & 1,1,3 & \dots \end{pmatrix}$$

$$\begin{pmatrix} o & d & d & d & d & d & o & d & d \\ 1,1,1 & 1,1,2 & 1,2,1 & 2,1,1 & 1,2,2 & 2,1,2 & 2,2,1 & 1,1,3 & \dots \end{pmatrix}$$

Vamos utilizar a lista acima para ilustrar a ideia principal dessa seção. Vamos supor que o sistema “caixa”, com energia $U = 2U_{1,1,2}$ e número de elétrons $N = 2$ seja o

nosso sistema total. Nosso objetivo é encontrar a probabilidade de algum orbital de interesse estar ocupado. Como exemplo, vamos utilizar o orbital 2,1,1.

Em um exemplo com poucos estados é trivial encontrar a probabilidade de algum orbital estar ocupado: basta contar quantos estados do sistema total possuem o orbital de interesse ocupado. Para esse sistema, existem dois estados onde o orbital 2,1,1 está ocupado: o segundo e o terceiro estados. Como existem seis estados possíveis, e partindo da suposição fundamental da Física Estatística que todos os estados são igualmente prováveis, então a probabilidade de eu encontrar o sistema com o orbital 2,1,1 ocupado é de 2/6.

Agora vamos rearranjar a nossa lista de maneira a expressar esse resultado de um jeito que vai ser bem útil quando o sistema for grande. Primeiro, vamos destacar o orbital de interesse colocando-o na primeira coluna de cada parênteses. Com essa notação, a nossa lista fica assim:

$$g(2U_{1,1,2}, 2) = 6$$

Estados possíveis:

$$\begin{pmatrix} d & d & o & o & d & d & d & d & d \\ 2,1,1 & 1,1,1 & 1,1,2 & 1,2,1 & 1,2,2 & 2,1,2 & 2,2,1 & 1,1,3 & \dots \end{pmatrix}$$

$$\begin{pmatrix} o & d & o & d & d & d & d & d & d \\ 2,1,1 & 1,1,1 & 1,1,2 & 1,2,1 & 1,2,2 & 2,1,2 & 2,2,1 & 1,1,3 & \dots \end{pmatrix}$$

$$\begin{pmatrix} o & d & d & o & d & d & d & d & d \\ 2,1,1 & 1,1,1 & 1,1,2 & 1,2,1 & 1,2,2 & 2,1,2 & 2,2,1 & 1,1,3 & \dots \end{pmatrix}$$

$$\begin{pmatrix} d & o & d & d & o & d & d & d & d \\ 2,1,1 & 1,1,1 & 1,1,2 & 1,2,1 & 1,2,2 & 2,1,2 & 2,2,1 & 1,1,3 & \dots \end{pmatrix}$$

$$\begin{pmatrix} d & o & d & d & d & o & d & d & d \\ 2,1,1 & 1,1,1 & 1,1,2 & 1,2,1 & 1,2,2 & 2,1,2 & 2,2,1 & 1,1,3 & \dots \end{pmatrix}$$

$$\begin{pmatrix} d & o & d & d & d & d & o & d & d \\ 2,1,1 & 1,1,1 & 1,1,2 & 1,2,1 & 1,2,2 & 2,1,2 & 2,2,1 & 1,1,3 & \dots \end{pmatrix}$$

Note que os estados são exatamente os mesmos, nada mudou: eu só troquei a posição do orbital de interesse, colocando-o na primeira coluna. Agora, eu vou reorganizar um pouco mais, colocando a lista de maneira que todos os estados ocupados apareçam no topo da lista. Fica assim:

$$g(2U_{1,1,2}, 2) = 6$$

Estados possíveis:

$$\begin{pmatrix} o & d & o & d & d & d & d & d & d \\ 2,1,1 & 1,1,1 & 1,1,2 & 1,2,1 & 1,2,2 & 2,1,2 & 2,2,1 & 1,1,3 & \dots \end{pmatrix}$$

$$\begin{pmatrix} o & d & d & o & d & d & d & d & d \\ 2,1,1 & 1,1,1 & 1,1,2 & 1,2,1 & 1,2,2 & 2,1,2 & 2,2,1 & 1,1,3 & \dots \end{pmatrix}$$

$$\begin{pmatrix} d & d & o & o & d & d & d & d & d \\ 2,1,1 & 1,1,1 & 1,1,2 & 1,2,1 & 1,2,2 & 2,1,2 & 2,2,1 & 1,1,3 & \dots \end{pmatrix}$$

$$\begin{pmatrix} d & o & d & d & o & d & d & d & d \\ 2,1,1 & 1,1,1 & 1,1,2 & 1,2,1 & 1,2,2 & 2,1,2 & 2,2,1 & 1,1,3 & \dots \end{pmatrix}$$

$$\begin{pmatrix} d & o & d & d & d & o & d & d & d \\ 2,1,1 & 1,1,1 & 1,1,2 & 1,2,1 & 1,2,2 & 2,1,2 & 2,2,1 & 1,1,3 & \dots \end{pmatrix}$$

$$\begin{pmatrix} d & o & d & d & d & d & o & d & d \\ 2,1,1 & 1,1,1 & 1,1,2 & 1,2,1 & 1,2,2 & 2,1,2 & 2,2,1 & 1,1,3 & \dots \end{pmatrix}$$

Continuamos exatamente com a mesma lista de estados possíveis, nada mudou. Ainda sem alterar absolutamente nada, vamos fazer uma divisão imaginária no nosso sistema. Pense o seguinte: o sistema “caixa” efetivamente é o conjunto de todos os orbitais da caixa. Sem mudar absolutamente nada, eu poderia muito bem afirmar que o meu sistema “caixa” é composto de dois sistemas: um que só possui o orbital de interesse 2,1,1 e o outro que possui o resto dos orbitais, ou seja, todos os outros, com exceção do 2,1,1. Essa divisão é puramente imaginária, não estamos mudando nada fisicamente. Vamos representar graficamente a divisão imaginária como mostrado abaixo, onde chamei o sistema com um único orbital de sistema **A** e o outro de sistema

B

$$g(2U_{1,1,2}, 2) = 6$$

Estados possíveis:

		$\left(\begin{array}{c} o \\ 2,1,1 \end{array} \right)$	$\left(\begin{array}{c} d \\ 1,1,1 \end{array} \right)$	$\left(\begin{array}{c} o \\ 1,1,2 \end{array} \right)$	$\left(\begin{array}{c} d \\ 1,2,1 \end{array} \right)$	$\left(\begin{array}{c} d \\ 1,2,2 \end{array} \right)$	$\left(\begin{array}{c} d \\ 2,1,2 \end{array} \right)$	$\left(\begin{array}{c} d \\ 2,2,1 \end{array} \right)$	$\left(\begin{array}{c} d \\ 1,1,3 \end{array} \right)$	$\left(\begin{array}{c} d \\ \dots \end{array} \right)$	
		$\left(\begin{array}{c} o \\ 2,1,1 \end{array} \right)$	$\left(\begin{array}{c} d \\ 1,1,1 \end{array} \right)$	$\left(\begin{array}{c} d \\ 1,1,2 \end{array} \right)$	$\left(\begin{array}{c} o \\ 1,2,1 \end{array} \right)$	$\left(\begin{array}{c} d \\ 1,2,2 \end{array} \right)$	$\left(\begin{array}{c} d \\ 2,1,2 \end{array} \right)$	$\left(\begin{array}{c} d \\ 2,2,1 \end{array} \right)$	$\left(\begin{array}{c} d \\ 1,1,3 \end{array} \right)$	$\left(\begin{array}{c} d \\ \dots \end{array} \right)$	
SISTEMA A		$\left(\begin{array}{c} d \\ 2,1,1 \end{array} \right)$	$\left(\begin{array}{c} d \\ 1,1,1 \end{array} \right)$	$\left(\begin{array}{c} o \\ 1,1,2 \end{array} \right)$	$\left(\begin{array}{c} o \\ 1,2,1 \end{array} \right)$	$\left(\begin{array}{c} d \\ 1,2,2 \end{array} \right)$	$\left(\begin{array}{c} d \\ 2,1,2 \end{array} \right)$	$\left(\begin{array}{c} d \\ 2,2,1 \end{array} \right)$	$\left(\begin{array}{c} d \\ 1,1,3 \end{array} \right)$	$\left(\begin{array}{c} d \\ \dots \end{array} \right)$	SISTEMA B
		$\left(\begin{array}{c} d \\ 2,1,1 \end{array} \right)$	$\left(\begin{array}{c} o \\ 1,1,1 \end{array} \right)$	$\left(\begin{array}{c} d \\ 1,1,2 \end{array} \right)$	$\left(\begin{array}{c} d \\ 1,2,1 \end{array} \right)$	$\left(\begin{array}{c} o \\ 1,2,2 \end{array} \right)$	$\left(\begin{array}{c} d \\ 2,1,2 \end{array} \right)$	$\left(\begin{array}{c} d \\ 2,2,1 \end{array} \right)$	$\left(\begin{array}{c} d \\ 1,1,3 \end{array} \right)$	$\left(\begin{array}{c} d \\ \dots \end{array} \right)$	
		$\left(\begin{array}{c} d \\ 2,1,1 \end{array} \right)$	$\left(\begin{array}{c} o \\ 1,1,1 \end{array} \right)$	$\left(\begin{array}{c} d \\ 1,1,2 \end{array} \right)$	$\left(\begin{array}{c} d \\ 1,2,1 \end{array} \right)$	$\left(\begin{array}{c} d \\ 1,2,2 \end{array} \right)$	$\left(\begin{array}{c} o \\ 2,1,2 \end{array} \right)$	$\left(\begin{array}{c} d \\ 2,2,1 \end{array} \right)$	$\left(\begin{array}{c} d \\ 1,1,3 \end{array} \right)$	$\left(\begin{array}{c} d \\ \dots \end{array} \right)$	
		$\left(\begin{array}{c} d \\ 2,1,1 \end{array} \right)$	$\left(\begin{array}{c} o \\ 1,1,1 \end{array} \right)$	$\left(\begin{array}{c} d \\ 1,1,2 \end{array} \right)$	$\left(\begin{array}{c} d \\ 1,2,1 \end{array} \right)$	$\left(\begin{array}{c} d \\ 1,2,2 \end{array} \right)$	$\left(\begin{array}{c} d \\ 2,1,2 \end{array} \right)$	$\left(\begin{array}{c} o \\ 2,2,1 \end{array} \right)$	$\left(\begin{array}{c} d \\ 1,1,3 \end{array} \right)$	$\left(\begin{array}{c} d \\ \dots \end{array} \right)$	

Agora note que podemos identificar a quantidade de estados do sistema total onde o orbital 2,1,1 está ocupado (os dois primeiros estados da lista acima) através de informações do sistema **B**. Note que, no nosso exemplo, se o orbital 2,1,1 está ocupado, então necessariamente o sistema **B** possui um único elétron (uma única letra *o*). Isso ocorre porque, no nosso exemplo, o número total de elétrons é dois, então se um elétron estiver no orbital 2,1,1 (ou equivalentemente, se um elétron estiver no sistema **A**), então o sistema **B** tem que conter o outro elétron. De maneira análoga, como a soma das energias dos orbitais ocupados tem que ser $U = 2U_{1,1,2}$, se o orbital 2,1,1 estiver ocupado, e tendo este orbital, neste exemplo, metade da energia total (lembre-se que $U_{2,1,1} = U_{1,1,2}$), então o sistema **B** tem que ter a outra metade. Então, identificamos que a quantidade de vezes que o orbital 2,1,1 aparece na lista de estados totais como “ocupado” corresponde à quantidade de estados possíveis do sistema **B** cuja energia e número de elétrons seja complementar à energia e número de elétrons do orbital 2,1,1.

Sendo assim, neste exemplo, quando o orbital 2,1,1 estiver ocupado, o sistema **B** deve ter um único elétron e a energia total do sistema **B** deve ser igual à $U_{1,1,2}$. Existem apenas dois estados do sistema **B** onde isso é possível (lembre-se que o sistema **B** exclui o orbital de interesse), a saber:

*Lista só do Sistema **B** para $U_B = 2U_{1,1,2} - U_{2,1,1}$ e $N_B = 1$*

$$\begin{pmatrix} d & o & d & d & d & d & d & d \\ 1,1,1 & 1,1,2 & 1,2,1 & 1,2,2 & 2,1,2 & 2,2,1 & 1,1,3 & \dots \end{pmatrix}$$

$$\begin{pmatrix} d & d & o & d & d & d & d & d \\ 1,1,1 & 1,1,2 & 1,2,1 & 1,2,2 & 2,1,2 & 2,2,1 & 1,1,3 & \dots \end{pmatrix}$$

Esses são, de fato, os estados do sistema **B** que, combinados com o orbital 2,1,1 ocupado, formam os dois primeiros estados do sistema total.

Utilizando o conceito de multiplicidade, podemos então afirmar que existem dois estados do sistema total onde o orbital 2,1,1 está ocupado porque existem dois estados do sistema **B** com energia e número de partículas complementares, ou seja, porque $g_B(U - U_{2,1,1}, N - 1) = 2$, onde, no nosso exemplo, $U = 2U_{1,1,2}$ e $N = 2$.

Quando o orbital 2,1,1 estiver desocupado, por outro lado, o sistema **B** terá que possuir dois elétrons e energia total igual a $2U_{1,1,2}$. Existem quatro estados do sistema **B** que satisfazem essas duas condições, a saber:

*Lista só do Sistema **B** para $U_B = 2U_{1,1,2}$ e $N_B = 2$*

$$\begin{pmatrix} d & o & o & d & d & d & d & d \\ 1,1,1 & 1,1,2 & 1,2,1 & 1,2,2 & 2,1,2 & 2,2,1 & 1,1,3 & \dots \end{pmatrix}$$

$$\begin{pmatrix} o & d & d & o & d & d & d & d \\ 1,1,1 & 1,1,2 & 1,2,1 & 1,2,2 & 2,1,2 & 2,2,1 & 1,1,3 & \dots \end{pmatrix}$$

$$\begin{pmatrix} o & d & d & d & o & d & d & d \\ 1,1,1 & 1,1,2 & 1,2,1 & 1,2,2 & 2,1,2 & 2,2,1 & 1,1,3 & \dots \end{pmatrix}$$

$$\begin{pmatrix} o & d & d & d & d & o & d & d \\ 1,1,1 & 1,1,2 & 1,2,1 & 1,2,2 & 2,1,2 & 2,2,1 & 1,1,3 & \dots \end{pmatrix}$$

Esses são, de fato, os quatro estados do sistema **B** que combinam com o orbital 2,1,1 desocupado para formar os quatro últimos estados da lista do sistema total. Portanto, existem quatro estados do sistema total onde o orbital 2,1,1 está desocupado, e isso

porque $g_B(U, N) = 4$, onde, novamente, $U = 2U_{1,1,2}$ e $N = 2$. Estas considerações estão ilustradas abaixo, onde utilizo a nomenclatura $N_{ocupado}$ e $N_{desocupado}$ para denotar o número de estados do sistema total onde o orbital 2,1,1 está ocupado e desocupado, respectivamente:

$$g(2U_{1,1,2}, 2) = 6$$

Estados possíveis:

$\left(\begin{array}{cccccccc} o & d & o & d & d & d & d & d & d \\ 2,1,1 & 1,1,1 & 1,1,2 & 1,2,1 & 1,2,2 & 2,1,2 & 2,2,1 & 1,1,3 & \dots \end{array} \right)$	}	$N_{ocupado} = g_B(2U_{1,1,2} - U_{2,1,1}, N - 1) = 2$
$\left(\begin{array}{cccccccc} o & d & d & o & d & d & d & d & d \\ 2,1,1 & 1,1,1 & 1,1,2 & 1,2,1 & 1,2,2 & 2,1,2 & 2,2,1 & 1,1,3 & \dots \end{array} \right)$		
$\left(\begin{array}{cccccccc} d & d & o & o & d & d & d & d & d \\ 2,1,1 & 1,1,1 & 1,1,2 & 1,2,1 & 1,2,2 & 2,1,2 & 2,2,1 & 1,1,3 & \dots \end{array} \right)$	}	$N_{desocupado} = g_B(2U_{1,1,2}, N) = 4$
$\left(\begin{array}{cccccccc} d & o & d & d & o & d & d & d & d \\ 2,1,1 & 1,1,1 & 1,1,2 & 1,2,1 & 1,2,2 & 2,1,2 & 2,2,1 & 1,1,3 & \dots \end{array} \right)$		
$\left(\begin{array}{cccccccc} d & o & d & d & d & o & d & d & d \\ 2,1,1 & 1,1,1 & 1,1,2 & 1,2,1 & 1,2,2 & 2,1,2 & 2,2,1 & 1,1,3 & \dots \end{array} \right)$		
$\left(\begin{array}{cccccccc} d & o & d & d & d & d & o & d & d \\ 2,1,1 & 1,1,1 & 1,1,2 & 1,2,1 & 1,2,2 & 2,1,2 & 2,2,1 & 1,1,3 & \dots \end{array} \right)$		

Sendo assim, concluímos que, para o nosso sistema exemplo, temos:

$$P\left(\begin{array}{c} o \\ 2,1,1 \end{array}\right) = \frac{N_{ocupado}}{N_{ocupado} + N_{desocupado}} = \frac{g_B(2U_{1,1,2} - U_{2,1,1}, N - 1)}{g(2U_{1,1,2}, 2)} = \frac{2}{6}$$

e

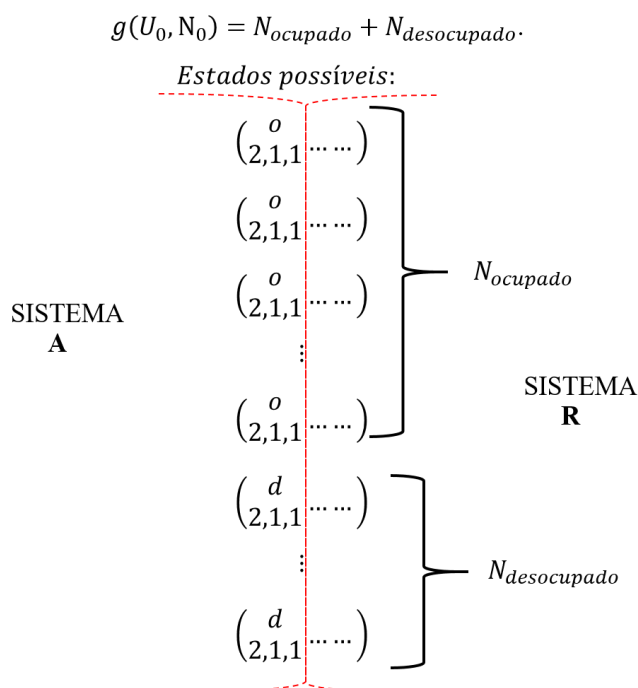
$$P\left(\begin{array}{c} d \\ 2,1,1 \end{array}\right) = \frac{N_{desocupado}}{N_{ocupado} + N_{desocupado}} = \frac{g_B(2U_{1,1,2}, N)}{g(2U_{1,1,2}, 2)} = \frac{4}{6}$$

onde $P\left(\begin{array}{c} o \\ 2,1,1 \end{array}\right)$ e $P\left(\begin{array}{c} d \\ 2,1,1 \end{array}\right)$ denotam a probabilidade do orbital 2,1,1 estar ocupado e desocupado, respectivamente.

Obviamente, temos também que:

$$P\left(\begin{array}{c} o \\ 2,1,1 \end{array}\right) + P\left(\begin{array}{c} d \\ 2,1,1 \end{array}\right) = 1$$

Agora vamos generalizar essa ideia para um sistema macroscópico. A caixa é a mesma que antes, mas agora imagine que queiramos levantar a lista para $U = U_0$, $N = N_0$, onde U_0 e N_0 são muito grandes, ou seja, macroscópicos (N_0 , por exemplo, poderia ser expresso em *mols*). A lista de todos os estados possíveis seria levantada de maneira idêntica aos exemplos anteriores, só que agora você precisaria de um supercomputador para passar a eternidade levantando a lista. Aí você escolhe novamente um orbital de interesse. Vamos escolher novamente o orbital 2,1,1 como exemplo. Então você rearranja a sua mega lista colocando o orbital de interesse na primeira coluna e colocando os estados onde o orbital de interesse está ocupado no topo da lista e os estados onde ele está desocupado na parte de baixo. A mega lista é representada de maneira meio rudimentar abaixo, onde os três pontinhos verticais indicam a presença de um monte de orbitais.



Note que agora eu chamei o outro sistema de sistema **R**. Eu fiz isso para denotar que, como agora estamos considerando energia e número de elétrons macroscópicos, tanto a energia como o número de elétrons do outro sistema será muito, muito, muito maior do que a energia e o número de elétrons do sistema **A**. De fato, ou o sistema **A** tem um único elétron e energia igual à $U_{2,1,1}$ (no caso do orbital estar ocupado), ou ele não tem

nenhuma energia e nenhum elétron. Isso significa que eu posso tratar o sistema **R** como um reservatório. Mas, independentemente do outro sistema ser grande ou não, o mesmo raciocínio que fizemos para o caso anterior continua válido e, portanto, temos que:

$$P\left(\begin{smallmatrix} o \\ 2,1,1 \end{smallmatrix}\right) = \frac{N_{ocupado}}{N_{ocupado} + N_{desocupado}} = \frac{g_R(U_0 - U_{2,1,1}, N_0 - 1)}{g(U_0, N_0)}$$

e, também:

$$P\left(\begin{smallmatrix} d \\ 2,1,1 \end{smallmatrix}\right) = \frac{N_{desocupado}}{N_{ocupado} + N_{desocupado}} = \frac{g_R(U_0, N_0)}{g(U_0, N_0)}$$

Agora vamos dividir uma probabilidade pela outra:

$$\frac{P\left(\begin{smallmatrix} o \\ 2,1,1 \end{smallmatrix}\right)}{P\left(\begin{smallmatrix} d \\ 2,1,1 \end{smallmatrix}\right)} = \frac{g_R(U_0 - U_{2,1,1}, N_0 - 1)}{g_R(U_0, N_0)}$$

Em termos da entropia, temos:

$$\frac{P\left(\begin{smallmatrix} o \\ 2,1,1 \end{smallmatrix}\right)}{P\left(\begin{smallmatrix} d \\ 2,1,1 \end{smallmatrix}\right)} = \exp\left[\frac{S_R(U_0 - U_{2,1,1}, N_0 - 1) - S_R(U_0, N_0)}{k_B}\right]$$

Agora, utilizando a informação que o sistema **R** é um reservatório, podemos expandir a sua entropia em série de Taylor e truncá-la nos termos de primeira ordem:

$$S_R(U_0 - U_{2,1,1}, N_0 - 1) \cong S_R(U_0, N_0) - U_{2,1,1} \frac{\partial S_R}{\partial U} - 1 \frac{\partial S_R}{\partial N}$$

Utilizando as definições de temperatura e potencial, temos:

$$S_R(U_0 - U_{2,1,1}, N_0 - 1) \cong S_R(U_0, N_0) - \frac{U_{2,1,1}}{T} + \frac{\mu}{T}$$

Com essa expansão, temos que:

$$\frac{P\left(\begin{smallmatrix} o \\ 2,1,1 \end{smallmatrix}\right)}{P\left(\begin{smallmatrix} d \\ 2,1,1 \end{smallmatrix}\right)} = \exp\left[\frac{S_R(U_0, N_0) - \frac{U_{2,1,1}}{T} + \frac{\mu}{T} - S_R(U_0, N_0)}{k_B}\right]$$

ou seja:

$$\frac{P\left(\begin{smallmatrix} o \\ 2,1,1 \end{smallmatrix}\right)}{P\left(\begin{smallmatrix} d \\ 2,1,1 \end{smallmatrix}\right)} = \exp\left[\frac{-\frac{U_{2,1,1}}{T} + \frac{\mu}{T}}{k_B}\right]$$

Isso nos permite expressar $P\left(\begin{smallmatrix} d \\ 2,1,1 \end{smallmatrix}\right)$ em termos de $P\left(\begin{smallmatrix} o \\ 2,1,1 \end{smallmatrix}\right)$

$$P\left(\begin{smallmatrix} d \\ 2,1,1 \end{smallmatrix}\right) = P\left(\begin{smallmatrix} o \\ 2,1,1 \end{smallmatrix}\right) \exp\left[\frac{U_{2,1,1} - \mu}{k_B T}\right]$$

Mas como $P\left(\begin{smallmatrix} o \\ 2,1,1 \end{smallmatrix}\right) + P\left(\begin{smallmatrix} d \\ 2,1,1 \end{smallmatrix}\right) = 1$, temos que:

$$P\left(\begin{smallmatrix} o \\ 2,1,1 \end{smallmatrix}\right) + P\left(\begin{smallmatrix} o \\ 2,1,1 \end{smallmatrix}\right) \exp\left[\frac{U_{2,1,1} - \mu}{k_B T}\right] = 1$$

e, portanto:

$$P\left(\begin{smallmatrix} o \\ 2,1,1 \end{smallmatrix}\right) = \frac{1}{1 + \exp\left[\frac{U_{2,1,1} - \mu}{k_B T}\right]}$$

Esse raciocínio poderia ser feito para qualquer orbital, portanto o resultado acima pode ser generalizado para um orbital c qualquer (onde c representa uma trinca qualquer), o que leva à:

$$P\left(\begin{smallmatrix} o \\ c \end{smallmatrix}\right) = \frac{1}{1 + \exp\left[\frac{U_c - \mu}{k_B T}\right]}$$

Equação 65

Esse importante resultado é conhecido como a distribuição de Fermi-Dirac, que é a probabilidade de um determinado orbital estar ocupado. Como a distribuição de Fermi-Dirac depende apenas da energia do orbital, ela é usualmente expressa como função da energia, como mostrado na Equação 66.

$$f(\varepsilon) = \frac{1}{1 + \exp\left[\frac{(\varepsilon - \mu)}{k_B T}\right]}$$

Equação 66

Essa fórmula é absolutamente idêntica à formula da Equação 65, só a notação que mudou. A notação da Equação 65 é melhor, porque ela te lembra que o cálculo é feito para um orbital específico, mas a notação da Equação 66 é mais comum na literatura.

Note que, como estamos efetivamente tratando o orbital de interesse como um sistema em contato térmico e difusivo com o reservatório, poderíamos ter aplicado a Equação 50 diretamente. Neste caso, temos apenas dois estados possíveis para o sistema “orbital”: ou o orbital está ocupado por um elétron ou está desocupado. Portanto, temos apenas duas opções para n (lembre-se que n denota o número de elementos do estado): $n = 0$ (para o estado “desocupado”) e $n = 1$ (para o estado “ocupado”). Mas, se o orbital estiver no estado “desocupado”, então a energia do orbital será $\varepsilon = 0$ (já que a energia vem do elétron no orbital, e não do orbital em si). Por outro lado, se o orbital estiver ocupado, terá energia $\varepsilon = \varepsilon_c$. Assim, temos apenas dois estados possíveis para o sistema “orbital”: o estado “desocupado”, para o qual temos $n = 0, \varepsilon = 0$; e o estado “ocupado”, para o qual temos $n = 1, \varepsilon = \varepsilon_c$. Assim, a soma de Gibbs para o sistema “orbital” será:

$$\zeta(T, \mu) = \exp\left(-\frac{0 - 0\mu}{k_B T}\right) + \exp\left(-\frac{\varepsilon_c - 1\mu}{k_B T}\right) = 1 + \exp\left(-\frac{\varepsilon_c - \mu}{k_B T}\right)$$

Com a soma de Gibbs, podemos calcular a probabilidade do sistema “orbital” estar no estado ocupado. Designando o orbital pela letra “c”, e utilizando a Equação 50 temos que:

$$P\left(\begin{smallmatrix} o \\ c \end{smallmatrix}\right) = \frac{\exp\left[\frac{-(\varepsilon_c - n_c \mu)}{k_B T}\right]}{\zeta(T, \mu)} = \frac{\exp\left[\frac{-(\varepsilon_c - \mu)}{k_B T}\right]}{1 + \exp\left[\frac{-(\varepsilon_c - \mu)}{k_B T}\right]}$$

$$\therefore P\left(\begin{smallmatrix} o \\ c \end{smallmatrix}\right) = \frac{1}{1 + \exp\left[\frac{(\varepsilon_c - \mu)}{k_B T}\right]}$$

Equação 67

Que é a distribuição de Fermi-Dirac, novamente (a Equação 67 é idêntica à Equação 65). Note que a temperatura e potenciais são, estritamente falando, a temperatura e os potenciais do reservatório, ou seja, do sistema fictício composto por todos os orbitais da caixa excluindo o orbital de interesse. Mas, exatamente pelo fato do reservatório ser um reservatório, ou seja, pelo fato de U_0 e N_0 serem macroscópicos, é que podemos identificar essa temperatura e potencial com a temperatura e potencial do sistema total, uma vez que esses parâmetros são virtualmente inalterados quando retiramos o orbital de interesse.

Procure lembrar de onde vem a distribuição de Fermi-Dirac: ela é obtida simplesmente fazendo a lista de todos os estados possíveis e contando quantos desses estados possuem o orbital de interesse ocupado.

1.8 Analogia do sistema “caixa” com o sistema “moedas”.

O sistema “caixa” é muito parecido com o sistema “moedas” que temos usado como exemplo. Nessa seção, eu comparo os dois sistemas, e mostro como o sistema “moedas” tem que ser modificado para que a analogia fique perfeita.

Primeiro, note que quem desempenha o papel das moedas não são os elétrons, mas sim os orbitais. Além disso, quem desempenha o papel de *cara* e *coroa* são as letras *d* e *o*.

Tanto as moedas como os orbitais são distinguíveis. Para ilustrar isso, suponha que temos duas moedas, que chamarei de $M1$ e $M2$; e, analogamente, dois orbitais (como exemplo, vou escolher os orbitais 1,1,1 e 1,1,2). As moedas são distinguíveis, isso quer dizer que o estado $\begin{pmatrix} cara & coroa \\ M1 & M2 \end{pmatrix}$ é diferente do estado $\begin{pmatrix} coroa & cara \\ M1 & M2 \end{pmatrix}$.

$$\begin{pmatrix} cara & coroa \\ M1 & M2 \end{pmatrix} \neq \begin{pmatrix} coroa & cara \\ M1 & M2 \end{pmatrix}$$

Analogamente, o estado $\begin{pmatrix} d & o \\ 1,1,1 & 1,1,2 \end{pmatrix}$ é diferente do estado $\begin{pmatrix} o & d \\ 1,1,1 & 1,1,2 \end{pmatrix}$:

$$\begin{pmatrix} d & o \\ 1,1,1 & 1,1,2 \end{pmatrix} \neq \begin{pmatrix} o & d \\ 1,1,1 & 1,1,2 \end{pmatrix}$$

Mas vimos que, porque elétrons são indistinguíveis, temos que o estado $\begin{pmatrix} d & o \\ 1,1,1 & 1,1,2 \end{pmatrix}$ e o estado $\begin{pmatrix} o & d \\ 1,1,2 & 1,1,1 \end{pmatrix}$ são absolutamente idênticos. Na verdade, temos apenas um estado, só mudamos o jeito de representá-lo. Ambas as representações representam o estado com um elétron no orbital 1,1,2 e nenhum elétron no orbital 1,1,1. Isso quer dizer que não importa a ordem com que você lista os orbitais (como já sabemos). Então:

$$\begin{pmatrix} d & o \\ 1,1,1 & 1,1,2 \end{pmatrix} \leftrightarrow \begin{pmatrix} o & d \\ 1,1,2 & 1,1,1 \end{pmatrix}$$

Leia o símbolo \leftrightarrow como “é o mesmo estado que”. O análogo da indistinguibilidade dos elétrons no exemplo das moedas seria a posição das moedas: não importa se eu ponho a moeda $M1$ na esquerda ou na direita, isso não altera o estado. De fato, o estado $\begin{pmatrix} cara & coroa \\ M1 & M2 \end{pmatrix}$ é absolutamente idêntico ao estado $\begin{pmatrix} coroa & cara \\ M2 & M1 \end{pmatrix}$. Se não fosse, teríamos quatro estados com uma *cara* e uma *coroa*, ao invés de apenas dois. Similarmente, se a posição fosse relevante, teríamos dois estados com duas *cara*, ao invés de um. Resumindo:

$$\begin{pmatrix} cara & coroa \\ M1 & M2 \end{pmatrix} \leftrightarrow \begin{pmatrix} coroa & cara \\ M2 & M1 \end{pmatrix}$$

Para não ter perigo de confundir, lembre-se também que:

$$\begin{pmatrix} \textit{cara} & \textit{coroa} \\ M1 & M2 \end{pmatrix} \neq \begin{pmatrix} \textit{coroa} & \textit{cara} \\ M1 & M2 \end{pmatrix}$$

Mas tem uma parte da analogia que tem que ser ajustada. No exemplo das moedas, quando nos perguntamos quantos estados com N_0 moedas possuem energia U_0 , vimos que esse problema se reduzia a encontrar de quantas formas poderíamos arranjar N_0 moedas com U_0 *coroas* viradas para cima. Isso foi verdade porque atribuímos a mesma energia $U = 1$ para todas as moedas. Mas isso não é verdade para os orbitais, uma vez que eles têm energias diferentes. Para a analogia ficar correta, teríamos que readaptar o exemplo das moedas de maneira que cada moeda teria um valor de energia diferente se a *coroa* estiver virada para cima. Por exemplo, a moeda $M1$ poderia ter energia $U = 1$ quando a *coroa* estiver para cima, enquanto a moeda $M2$ poderia ter energia $U = 1.5$ quando a *coroa* estiver para cima, e por aí vai.

Finalmente, note que o problema de encontrar os estados para um número N_0 de elétrons NÃO se reduz ao problema de encontrar permutações com N_0 letrinhas o e resto d . É verdade que, se temos N_0 elétrons, então necessariamente precisamos de N_0 letrinhas o em cada estado possível. Mas, além disso, a soma das energias dos orbitais onde se encontram as letrinhas o tem que ser igual à energia U_0 do sistema. Então não é qualquer combinação com N_0 letrinhas o que vai ser um estado possível

1.9 Concentração de elétrons e nível de Fermi

A maioria dos livros sobre semicondutores partem da distribuição de Fermi-Dirac para deduzir a relação tensão versus corrente em dispositivos semicondutores. Esse procedimento será mostrado nos capítulos seguintes. O importante, por enquanto, é saber interpretar corretamente o que a distribuição significa fisicamente. Já afirmamos, mais de uma vez, que ela corresponde à probabilidade de um determinado orbital cuja energia é ε estar ocupado, mas podemos antecipar aqui um exemplo particularmente importante em eletrônica a respeito de como utilizar a distribuição de Fermi-Dirac.

Primeiro, note que a caixa será um pedaço de semicondutor, por exemplo, de silício. A energia U_0 , por outro lado, vai depender da temperatura ambiente. De fato, a energia U_0 será aquela que satisfizer a condição:

$$\frac{\partial S_{\text{semicondutor}}}{\partial U} \Big|_{U=U_0} = \frac{1}{T}$$

Equação 68

onde T é a temperatura ambiente.

Como no exemplo da caixa, o parâmetro N_0 será o número de elétrons (como veremos mais adiante, estamos particularmente interessados nos elétrons livres, ou seja, elétrons que contribuem para a corrente). Sendo assim, a distribuição de Fermi-Dirac nos permite relacionar o número de elétrons N_0 (na verdade a concentração de elétrons) com o potencial μ . Como cada orbital c possui uma probabilidade $f(c)$ de estar ocupado, então:

$$N_0 = \sum_c P \binom{0}{c}$$

Equação 69

Onde a soma em “ c ” é feita sobre todos os orbitais possíveis (na seção 2.3 trataremos especificamente dessa soma e explicaremos com mais detalhes em que sentido ela é verdadeira). Esse é um típico e importante exemplo de como a distribuição de Fermi-Dirac é útil: ela relaciona o número de elétrons (lado esquerdo) com o potencial μ do sistema (que entra no lado direito pela distribuição de Fermi-Dirac). Portanto, sabendo μ podemos encontrar N_0 e vice-versa.

Antes de terminarmos esta seção, vamos explicitar duas propriedades da distribuição de Fermi-Dirac que ajudam a aumentar a intuição sobre o seu significado físico. Primeiro, imagine que temos um sistema com temperatura $T = 0 \text{ K}$, chamada de zero absoluto. Neste caso, a exponencial no denominador da Equação 66 será infinita para $\varepsilon > \mu$, o que resulta em $f(\varepsilon) = 0$ para $\varepsilon > \mu$. Por outro lado, se $\varepsilon < \mu$, então a exponencial é igual à zero, o que resulta em $f(\varepsilon) = 1$ para $\varepsilon < \mu$. Conclusão: no zero absoluto, todos os níveis de energia abaixo de μ estão ocupados, enquanto todos os níveis acima estão desocupados. Na verdade, o valor de μ quando a temperatura é zero recebe um nome especial: nível de Fermi. Atribuindo o símbolo ε_F para o nível de Fermi, temos por definição que:

$$\mu(T = 0) = \varepsilon_F$$

Equação 70

Infelizmente, existe uma lambança gigantesca a respeito do uso do termo “nível de Fermi”. A definição correta para nível de Fermi é a da Equação 70, mas muitos livros chamam o potencial, em qualquer temperatura, de nível de Fermi. Então, cuidado quando você encontrar esse termo na literatura e tente deduzir pelo contexto se o livro está falando realmente do nível de Fermi ou se está falando simplesmente de μ . Como diria o velho Rosa: viver é muito perigoso.

Finalmente, note que podemos interpretar o potencial μ do sistema como correspondendo ao nível de energia para o qual a probabilidade de ocupação é igual à 0.5. Procure na internet gráficos de $f(\varepsilon)$ para diferentes temperaturas e tente interpretá-los qualitativamente.

1.10 Transporte

A última seção deste capítulo trata do fenômeno de transporte, que é o estudo dos mecanismos de deslocamento de partículas. Em particular, estamos interessados no transporte de cargas, que geralmente é expresso em termos de correntes. Existem dois tipos de corrente, a corrente de arrasto e a corrente de difusão. A corrente de arrasto é a corrente associada à Lei de Ohm e é a forma mais simples de corrente. Vamos tratar rapidamente da corrente de arrasto e, a seguir, explicar a corrente de difusão, que é o assunto principal desta seção.

Vamos, então, começar fazendo uma revisão sobre corrente de arrasto. Não precisamos falar especificamente de corrente de elétrons, podemos generalizar o argumento para a corrente de um elemento qualquer. Para facilitar, vou chamar o elemento de partícula.

Por definição, corrente, de forma geral, é o número de partículas que atravessam uma determinada área por segundo. Muitas vezes, utilizamos a densidade de corrente, que é a corrente por unidade de área.

Vamos supor que temos um grupo de partículas percorrendo um tubo com área A e que todas as partículas possuem velocidade v , como mostrado na Figura 7. Queremos encontrar uma expressão para a corrente de partículas, ou seja, para o número de partículas que atravessam a área A por segundo. Para isso, escolha um intervalo de tempo Δt . Quantas partículas vão atravessar a área A nesse intervalo de tempo? Para acharmos a resposta, basta identificarmos qual é a distância que uma partícula percorre

no intervalo de tempo Δt . Obviamente, a distância é $\Delta L = v\Delta t$. Como todas as partículas que estão no volume delimitado por ΔL vão atravessar a área no intervalo Δt , temos que o número de partículas que atravessam A no intervalo Δt é exatamente o número de partículas que estão dentro do volume delimitado por ΔL , ou seja:

$$N_{\Delta t} = \rho \cdot \Delta L \cdot A = \rho \cdot v \cdot \Delta t \cdot A$$

Equação 71

Onde ρ é a densidade de partículas (número de partículas por unidade de volume) e $\Delta L \cdot A$ é o volume delimitado por ΔL . Portanto, a corrente I que passa por A será:

$$I = \frac{N_{\Delta t}}{\Delta t} = \rho \cdot v \cdot A$$

Equação 72

Consequentemente, a densidade de corrente J_{arr} será:

$$J_{arr} = \rho \cdot v$$

Equação 73

Onde colocamos o subíndice *arr* para lembrarmos que estamos tratando da corrente de arrasto. Essa é a versão mais simples da corrente de arrasto. Note que a origem física da corrente de arrasto é a velocidade v das partículas. Portanto, esse tipo de corrente é associado ao movimento conjunto de partículas (aqui, supomos que todas as partículas possuem a mesma velocidade, representada pelas flechinhas na Figura 7)

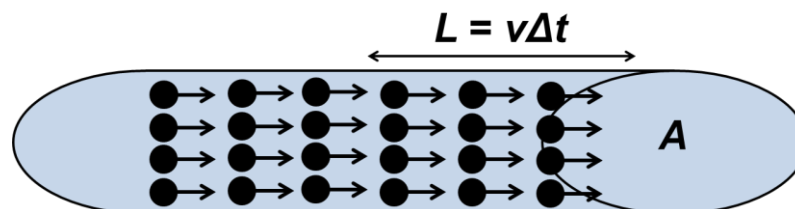


Figura 7 Corrente de arrasto em um tubo com área A .

Agora, vamos supor que exista certa variação na velocidade das partículas: umas estão mais rápidas e outras mais devagar. Por enquanto, vamos supor que tanto a magnitude como o sentido (direita ou esquerda) da corrente possam variar, mas, para

facilitar, vamos supor que a direção (horizontal) seja fixa. Qual será a densidade de corrente nesse caso? Para calcular a densidade de corrente total, vamos supor que temos N grupos de partícula e que cada grupo possua uma determinada velocidade. Designando por ρ_n a densidade de partículas com velocidade v_n , temos que a densidade de corrente total será:

$$J_{arr} = \sum_{n=1}^N \rho_n \cdot v_n$$

Equação 74

Agora, note o seguinte: se quiséssemos calcular qual é a média das velocidades das partículas no tubo, teríamos que somar todas as velocidades e depois dividir pelo número de partículas. Denotando o número de partículas do grupo n por N_n , teríamos, então, que:

$$\langle v \rangle = \frac{\sum_{n=1}^N N_n \cdot v_n}{\sum_{n=1}^N N_n}$$

Equação 75

Podemos reexpressar esse resultado em termos das densidades. Se V for o volume, então:

$$\langle v \rangle = \frac{\sum_{n=1}^N V \cdot \rho_n \cdot v_n}{\sum_{n=1}^N V \cdot \rho_n} = \frac{\sum_{n=1}^N \rho_n \cdot v_n}{\sum_{n=1}^N \rho_n}$$

∴

$$\langle v \rangle = \frac{\sum_{n=1}^N \rho_n \cdot v_n}{\rho}$$

Equação 76

Onde ρ é a densidade total. Utilizando a Equação 76, podemos reescrever a Equação 74 como:

$$J_{arr} = \rho \langle v \rangle$$

Equação 77

Essa é uma expressão mais geral para a corrente de arrasto. Note que se tivéssemos metade das partículas indo para a direita e metade para a esquerda com a mesma velocidade, então a média das velocidades seria zero e, conseqüentemente $J_{arr} = 0$. Essa é uma característica típica da corrente de arrasto: se a média das velocidades for zero, a corrente de arrasto também é zero.

Agora, suponha que um tubo, similar ao da Figura 7, contenha partículas cuja média das velocidades é zero. Neste caso, a corrente em qualquer área transversal dentro do tubo será realmente zero, porque teremos o mesmo número de partículas cruzando a área nos dois sentidos, resultando em uma corrente total nula. Mas, se o tubo for finito, então a corrente na extremidade do tubo não será zero: obviamente, partículas escaparão do tubo, mesmo se a média das velocidades das partículas dentro do tubo for zero. É claro que isso só ocorre na extremidade do tubo, porque temos partículas só de um lado da extremidade, enquanto o outro lado não tem partículas. Em linguagem mais técnica, existe um gradiente (uma variação), na densidade de partículas na extremidade, pois a densidade fora do tubo é zero, enquanto a densidade dentro do tubo é diferente de zero.

Assim, a corrente é zero em todos os pontos do tubo, menos na extremidade. Esse tipo de corrente, cuja origem é a variação na densidade de partículas, é chamada de corrente de difusão. Note que a dedução da Equação 77 supõe que a densidade seja constante ao longo do tubo, portanto a Equação 77 não leva em consideração a corrente de difusão. De fato, como já afirmamos anteriormente, a Equação 77 é a expressão para a corrente de arrasto somente.

Para podermos encontrar uma expressão para a corrente de difusão, podemos supor que a média das velocidades seja nula. Caso isso não seja verdade, podemos sempre separar a corrente entre uma componente de difusão e uma componente de arrasto e incluir o efeito da velocidade na corrente de arrasto; portanto não perdemos generalização com essa suposição.

Imagine que temos duas caixas com dimensões iguais, como mostrado na Figura 8. Em ambas as caixas, todas as partículas possuem velocidade com o mesmo módulo $|v|$, mas com direções e sentidos aleatórios, de maneira que a média das velocidades é zero. A única diferença entre as duas caixas é que as densidades de partículas são diferentes. Queremos determinar qual é a densidade de corrente de difusão, ou seja, quantas partículas passam de uma caixa para a outra (por unidade de área e tempo). Para isso,

vamos dividir o problema em duas partes: primeiro vamos calcular a corrente J_{e-d} , que é a contribuição de partículas que vão da caixa esquerda para a caixa direita. Em seguida, vamos calcular a contribuição de partículas que vão da caixa direita para a esquerda J_{d-e} , e depois subtrair uma da outra.

Vamos focar, agora, em determinar J_{e-d} . A priori, podemos usar o mesmo raciocínio que nos levou à Equação 73, mas com uma diferença. Antes, não importava qual era a distância ΔL que escolhemos para calcular o número de partículas que atravessa a área em um determinado instante de tempo Δt . Mas, aqui, como as partículas estão movendo aleatoriamente, elas trombarão com a estrutura atômica do material, alterando a direção da velocidade. Mas se a velocidade mudar, então não posso usar a Equação 73, porque o raciocínio que levou à Equação 73 supõe que as velocidades são fixas. Portanto, para utilizar o mesmo raciocínio, temos que escolher uma distância dentro da qual as partículas não trombem significativamente. O parâmetro estatístico que expressa a distância típica que uma partícula percorre entre duas colisões é chamado de caminho livre médio (*mean free path*). Uma escolha razoável para o comprimento ΔL é a componente z do caminho livre médio. Chamaremos o caminho livre médio de l e sua componente z de l_z (como na Figura 8), já que dentro dessa distância não haverá muitas colisões e, não havendo colisões, não haverá mudança de velocidade e poderei utilizar a expressão da Equação 73 para a corrente.

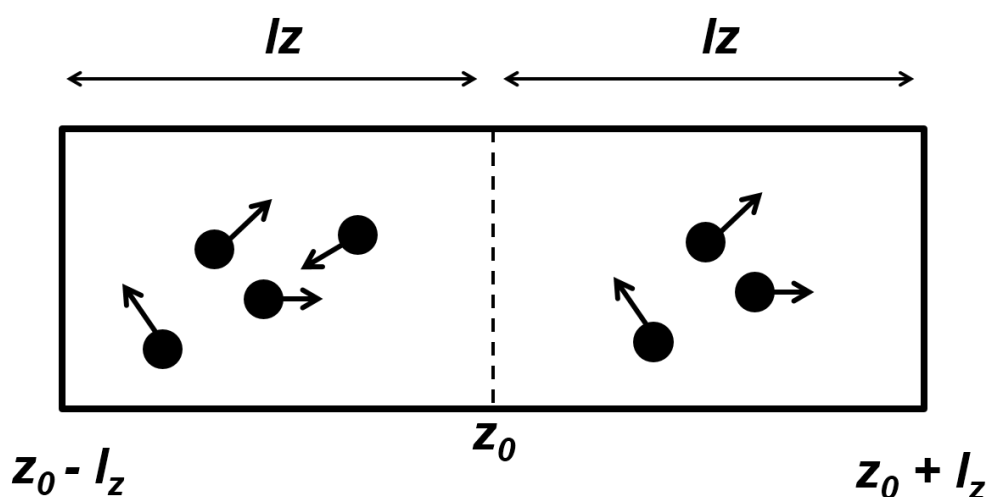


Figura 8 Esquema para o cálculo da corrente de difusão

Uma vez determinado o comprimento da caixa (que é igual à l_z , que é a componente z do caminho livre médio), temos que abordar um segundo problema: agora, temos partículas em todas as direções, portanto temos que pegar as contribuições para a corrente J_{e-d} de cada direção individualmente e depois somá-las. Para fazermos isso, precisamos pegar a projeção da velocidade no eixo z . Assim, vamos supor que estamos analisando o grupo de partículas cuja direção da velocidade faça um ângulo θ com o eixo z ; então a projeção da velocidade no eixo z será:

$$v_z = |v| \cos \theta$$

Equação 78

Além disso, como o tamanho da caixa é a projeção do caminho livre médio na direção z , e como essa é a projeção também do ângulo da velocidade, ou seja, como:

$$l_z = l \cos \theta$$

Equação 79

então, o próprio tamanho da caixa depende da direção θ da velocidade do grupo de partículas analisado.

Chamando a contribuição desse grupo de partículas para a corrente de $\Delta J_{e-d}(\theta)$, temos que a Equação 73 fica:

$$\Delta J_{e-d}(\theta) = \rho(z_0 - l_z, \theta) \cdot v_z$$

Equação 80

Onde $\rho(z_0 - l_z, \theta)$ é a densidade de partículas na caixa esquerda cuja velocidade faz um ângulo θ com o eixo z . A variável z denota a posição onde estamos avaliando a densidade (assim, $z_0 - l_z$ denota a posição da caixa esquerda).

O procedimento para encontrar $\Delta J_{d-e}(\theta)$ é o mesmo, mas agora temos que usar a densidade na caixa da direita. Portanto:

$$\Delta J_{d-e}(\theta) = \rho(z_0 + l_z, \theta) \cdot v_z$$

Equação 81

Assim, temos que a contribuição das partículas com direção θ para a corrente entre as duas caixas é:

$$\Delta J(\theta) = \Delta J_{e-d}(\theta) - \Delta J_{d-e}(\theta)$$

\therefore

$$\Delta J(\theta) = [\rho(z_0 - l_z, \theta) - \rho(z_0 + l_z, \theta)] \cdot v_z$$

Equação 82

Mas

$$\rho(z_0 - l_z, \theta) - \rho(z_0 + l_z, \theta) \approx -\frac{\partial \rho}{\partial z} \cdot 2 \cdot l_z$$

Equação 83

Portanto:

$$\Delta J(\theta) = -\frac{\partial \rho(z, \theta)}{\partial z} 2 \cdot v_z \cdot l_z$$

Equação 84

Note que a Equação 83 subentende que a derivada é avaliada no centro da caixa, ou seja, em $z = z_0$, que é exatamente onde estamos calculando a corrente. Substituindo a Equação 78 e Equação 79 na Equação 84 temos:

$$\Delta J(\theta) = -\frac{\partial \rho(z, \theta)}{\partial z} 2 \cdot |v| \cdot l \cdot \cos^2 \theta$$

Equação 85

Como as partículas possuem direções aleatórias, a densidade de partículas que apontam para um diferencial de ângulo sólido $d\Omega$ será:

$$\rho(z, \theta) = \rho(z) \frac{d\Omega}{4\pi}$$

Equação 86

Onde $\rho(z)$ é a densidade total de partículas, ou seja, a soma de todas as partículas (independente da direção da velocidade) em um volume em torno de z , dividido pelo volume. Se você achou a Equação 86 misteriosa, lembre-se que 4π nada mais é do que a integral de $d\Omega$ sobre toda a esfera. Portanto, a Equação 86 está simplesmente afirmando que a proporção de partículas que apontam para o ângulo sólido $d\Omega$ é igual à fração que $d\Omega$ ocupa em relação a todos os ângulos sólidos (já que não existe uma direção preferencial). Substituindo a Equação 86 na Equação 85, temos:

$$\Delta J(\theta) = -\frac{\partial \rho(z)}{\partial z} 2 \cdot |v| \cdot l \cdot \cos^2 \theta \cdot \frac{d\Omega}{4\pi}$$

Equação 87

Para encontrarmos a corrente total, basta somarmos a contribuição de cada ângulo, ou seja:

$$J_{dif} = \int \Delta J(\theta) = \frac{-\frac{\partial \rho(z)}{\partial z} 2 \cdot |v| \cdot l}{4\pi} \int \cos^2 \theta d\Omega$$

\therefore

$$J_{dif} = \int \Delta J(\theta) = \frac{-\frac{\partial \rho(z)}{\partial z} 2 \cdot |v| \cdot l}{4\pi} \int_0^{2\pi} \int_0^{\pi/2} \cos^2 \theta \sin \theta d\theta d\phi$$

Equação 88

Onde colocamos o subíndice *dif* para lembrarmos que estamos tratando da corrente de difusão. Note que a integral é feita sobre a metade superior da esfera (a integral em θ vai de 0 a $\pi/2$), já que as componentes da metade inferior possuem componente z negativa, portanto não contribuem para a corrente no centro da caixa (na verdade, estritamente falando, temos que as partículas com componente z negativa no lado esquerdo não contribuem, enquanto as partículas com z positivo do lado direito não contribuem; mas como a integral é simétrica tanto faz pegarmos o hemisfério superior ou inferior). A integral não oferece grandes dificuldades e o resultado é:

$$J_{dif} = -\frac{1}{3} \cdot |v| \cdot l \cdot \frac{\partial \rho}{\partial z}$$

Equação 89

Portanto, chegamos à conclusão de que a corrente de difusão é proporcional à variação na densidade de partículas, o que não é surpreendente, uma vez que a origem dessa corrente é exatamente a variação na densidade de partículas. A constante de proporcionalidade recebe o nome de difusividade ou coeficiente de difusão D . De maneira geral, a corrente de difusão é expressa em termos de D como:

$$J_{dif} = -D \cdot \frac{\partial \rho}{\partial z}$$

Equação 90

Portanto, para o nosso modelo, temos:

$$D = \frac{1}{3} \cdot |v| \cdot l$$

Equação 91

Lembre-se que a Equação 90 se refere à componente z da corrente de difusão. Poderíamos ter feito o mesmo raciocínio para as direções y e x também, e teríamos chegado exatamente no mesmo resultado (mas com a derivada em relação à y e x , respectivamente).

Generalizando para três dimensões, em linguagem matemática temos que o vetor corrente de difusão é definido por:

$$\vec{J}_{dif} = J_x \hat{x} + J_y \hat{y} + J_z \hat{z}$$

Equação 92

Onde as suas componentes são:

$$J_x = -D \cdot \frac{\partial \rho}{\partial x}, J_y = -D \cdot \frac{\partial \rho}{\partial y}, J_z = -D \cdot \frac{\partial \rho}{\partial z}$$

Equação 93

Note que podemos escrever a corrente de forma mais compacta utilizando o gradiente da densidade de partículas:

$$\vec{J}_{dif} = -D\nabla\rho$$

Equação 94

A Equação 94 é a forma mais geral, mas para esse curso a versão unidimensional basta (Equação 90). É importante notar que, apesar de termos chegado a uma expressão fechada para D , outros modelos poderiam levar a expressões diferentes. **O importante é observar que a corrente de difusão é proporcional à variação espacial da densidade de partículas.** Qual é exatamente a constante de proporção pode variar caso a caso.

Finalmente, note que o sinal negativo entre a corrente de difusão e o gradiente da corrente advém do fato de que, se tivermos mais partículas no lado esquerdo da caixa, então a corrente será positiva (vai da caixa esquerda para a direita), mas o gradiente da densidade é negativo (uma vez que a densidade diminui à medida que andamos para a direita, o que é o equivalente a afirmar que a densidade na caixa da esquerda é maior que na da direita).

1.11 Relação entre corrente e densidade de portadores

No último tópico deste capítulo, vamos encontrar algumas expressões extremamente importantes para o transporte de cargas em dispositivos semicondutores.

Nosso objetivo é relacionar o fluxo de partículas carregadas (ou corrente) que entra em uma região (um volume) com a concentração de cargas nessa região. Essa é uma conta muito simples, que frentistas fazem toda vez que relacionam a quantidade de carros dentro de um estacionamento com a quantidade de carros que entrou e saiu; a única diferença é que a linguagem matemática que vamos usar é um pouco mais sofisticada.

Para facilitar, primeiro vamos fazer a conta para o número de partículas (sem considerar a carga), e depois incluímos a carga.

Suponha, então, que temos uma caixa de comprimento Δz e área A , como mostrado na Figura 9. Suponha, também, que chegam N partículas por segundo na caixa, e saem M partículas por segundo. Queremos relacionar a quantidade de partículas dentro da caixa com esses fluxos N e M . Vamos chamar a quantidade de partículas dentro da caixa de $P(t)$ (P é função do tempo t , porque o número total de partículas irá alterar se o

fluxo total for diferente de zero). Se chegam N partículas por segundo, então, em um intervalo de tempo Δt , haverá um aumento de $N\Delta t$ partículas na caixa. Mas, como também saem M partículas por segundo, então haverá uma redução de $M\Delta t$. Assim, em um intervalo de tempo Δt , haverá um aumento líquido de $(N - M)\Delta t$ de partículas na caixa. Portanto:

$$(N - M)\Delta t = P(t + \Delta t) - P(t)$$

Equação 95

onde $P(t + \Delta t) - P(t)$ é o aumento de partículas no intervalo Δt . Podemos deixar essa equação um pouco mais formal. Para isso, em vez de chamar o fluxo que chega de N e o fluxo que sai de M , vamos colocar um eixo z no sistema e descrever o fluxo como função de z . Vamos chamar esse fluxo, que pode depender de z , de $F(z)$. Assim, teremos que, se a caixa começa em z , e termina em $z + \Delta z$, então $N = F(z)$ e $M = F(z + \Delta z)$. Utilizando essa notação, podemos reescrever a Equação 95 como:

$$[F(z) - F(z + \Delta z)]\Delta t = P(t + \Delta t) - P(t)$$

Equação 96

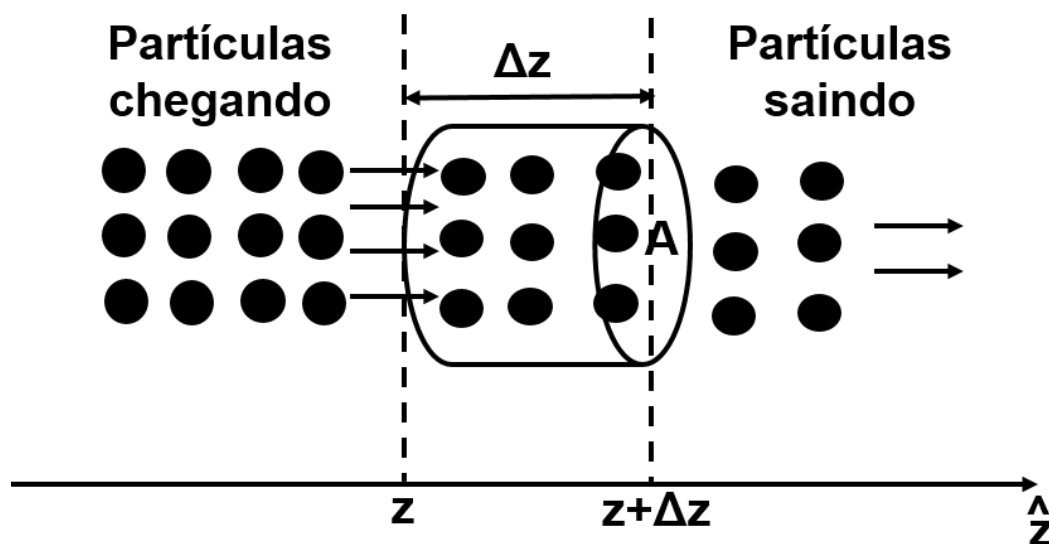


Figura 9 Uma caixa com comprimento Δz e área A , aonde chegam N partículas por segundo e saem M partículas por segundo.

Podemos melhorar ainda mais essa relação se expressarmos tudo em termos de densidades (ou concentração). Definindo a densidade volumétrica de partículas na caixa, $d = P/V$, onde $V = A\Delta z$ é o volume da caixa, temos que a Equação 96 pode ser reescrita como:

$$[F(z) - F(z + \Delta z)]\Delta t = d(t + \Delta t)V - d(t)V = [d(t + \Delta t) - d(t)]A\Delta z$$

Para ficar ainda mais profissional, vamos expressar o fluxo $F(z)$ em termos de densidade de corrente de partículas J . Isso é fácil: se a densidade de corrente for $J(z)$, isso quer dizer que passam $J(z)$ partículas por unidade de área e, conseqüentemente, $F(z) = J(z)A$. Assim:

$$[J(z) - J(z + \Delta z)]A\Delta t = [d(t + \Delta t) - d(t)]A\Delta z$$

Rearranjando:

$$\frac{J(z) - J(z + \Delta z)}{\Delta z} = \frac{d(t + \Delta t) - d(t)}{\Delta t}$$

Equação 97

No limite de uma caixa com comprimento infinitesimal e um intervalo de tempo infinitesimal, a relação acima se reduz à:

$$-\frac{\partial J}{\partial z} = \frac{\partial d}{\partial t}$$

Equação 98

Note que utilizamos o símbolo de derivada parcial porque J e d são funções de t e z . Note também o sinal de negativo que acompanha a derivada espacial. Ele aparece porque, na Equação 97, temos $[J(z) - J(z + \Delta z)]$, que é o oposto da subtração infinitesimal que define a derivada. Fisicamente, o sinal de negativo na derivada de J aparece porque uma derivada positiva implica que $J(z + \Delta z)$ é maior que $J(z)$, ou seja, uma derivada positiva implica que o fluxo que sai - $J(z + \Delta z)$ - é maior que o fluxo

que entra - $J(z)$; o que implica que a densidade dentro da caixa diminui (derivada temporal de d passa a ser negativa).

Não se assuste com a Equação 98. Lembre-se que ela é simplesmente uma versão meio invocada da Equação 95, que, por sua vez, nada mais é que a afirmação em linguagem matemática que “o que entra, menos o que sai, é o que fica”.

Agora, podemos facilmente utilizar a Equação 98 para o caso onde as partículas são cargas e , conseqüentemente, a corrente J é uma corrente de cargas. Temos dois casos possíveis: caso 1 – as partículas são carregadas positivamente, caso 2 – as partículas estão carregadas negativamente. Chamaremos a densidade de partículas para o caso 1 de p e para o caso 2 de n . Como J é geralmente expresso em coulombs por segundo por metro quadrado, precisamos que a densidade d no lado direito da Equação 98 seja a densidade volumétrica de coulombs (ou seja, coulombs/m³). Se temos p partículas por unidade de volume e cada uma dessas partículas carrega uma carga $|q|$, então obviamente teremos $|q|p$ coulombs por unidade de volume, ou seja, $d = |q|p$. Portanto, para cargas positivas, temos:

$$-\frac{\partial J}{\partial z} = |q| \frac{\partial p}{\partial t}$$

Equação 99

Analogamente, se temos n partículas e cada uma carrega uma carga $-|q|$, teremos:

$$-\frac{\partial J}{\partial z} = -|q| \frac{\partial n}{\partial t}$$

Equação 100

Se as partículas forem elétrons, usamos, então, a Equação 100. Como veremos mais adiante, existem dois portadores de cargas em semicondutores: elétrons e buracos. Buracos são portadores com carga positiva. Portanto, se as partículas forem buracos, usamos a Equação 99.

Note que não fizemos absolutamente nenhuma referência ao tipo de corrente: em momento nenhum especificamos se a corrente é de arrasto, ou de difusão, ou uma mistura das duas. Tudo o que dissemos é que existe uma corrente. Portanto, tanto a Equação 99 como a Equação 100 valem para qualquer tipo de corrente: de arrasto, de difusão ou uma combinação das duas.

Agora, em eletrônica estamos geralmente interessados em partículas que carregam cargas, mas que também podem contribuir para a corrente. Em outras palavras, estamos geralmente interessados em portadores de carga livres (elétrons livres ou buracos). Acontece que um portador de carga livre pode sofrer o processo de recombinação. Existem vários tipos de processos de recombinação com diversas origens físicas. Aqui, basta saber que esse processo de recombinação constitui uma perda da carga, porque ela passa a não ser mais móvel. Não é que a carga some: ela simplesmente deixa de ser livre; o que, para a corrente, é como se ela sumisse.

Precisamos, então, incorporar esses processos de perda na Equação 99 e na Equação 100. Para isso, vamos voltar à Equação 97, que vou reescrever abaixo em uma forma com interpretação mais óbvia:

$$[J(z) - J(z + \Delta z)]A\Delta t = [d(t + \Delta t) - d(t)]A\Delta z$$

Então, vamos lembrar o que essa equação nos diz: **o lado direito da equação é o aumento do número de partículas dentro da caixa em um intervalo de tempo Δt** (lembre-se que d é a densidade e que $A\Delta z$ é o volume, portanto estamos subtraindo o número de partículas dentro da caixa no tempo $t + \Delta t$ do número de partículas dentro da caixa no tempo t). O lado esquerdo, por sua vez, é o número de partículas que entram menos o número de partículas que saem nesse intervalo de tempo Δt .

Agora, vamos supor que, além de partículas entrando e saindo, temos também algum processo de perda, caracterizado por uma frequência γ . Essa frequência expressa a probabilidade de uma partícula se perder por unidade de tempo (γ tem unidade de $1/s$). Então, se tivermos P partículas no volume, teremos uma perda de $P\gamma$ partículas por segundo. Assim, o número de partículas que foi perdida em um intervalo de tempo Δt será $P\gamma\Delta t$. Lembrando que o número de partículas no volume é $P = dA\Delta z$, então o número de partículas perdidas em um intervalo de tempo Δt será $(dA\Delta z)\gamma\Delta t$. Como obviamente a variação do número de partículas é o número de partículas que entra, menos o que sai, menos o que é perdido, teremos, então, a seguinte relação:

$$[J(z) - J(z + \Delta z)]A\Delta t - dA\Delta z\gamma\Delta t = [d(t + \Delta t) - d(t)]A\Delta z$$

Dividindo ambos os lados por $A\Delta z\Delta t$:

$$\frac{[J(z) - J(z + \Delta z)]}{\Delta z} - d\gamma = \frac{[d(t + \Delta t) - d(t)]}{\Delta t}$$

No limite infinitesimal:

$$-\frac{\partial J}{\partial z} - d\gamma = \frac{\partial d}{\partial t}$$

Equação 101

Assim, a Equação 101 é a versão da Equação 98 incluindo o processo de perdas.

A frequência é geralmente expressa em termos do tempo de vida τ , definido como:

$$\tau = \frac{1}{\gamma}$$

Equação 102

O tempo de vida τ expressa o tempo típico que demora para uma partícula sofrer recombinação, daí o nome “tempo de vida”.

Podemos, agora, facilmente reescrever a Equação 101 para a concentração de portadores positivos:

$$-\frac{\partial J}{\partial z} - |q| \frac{p}{\tau_p} = |q| \frac{\partial p}{\partial t}$$

Equação 103

e negativos

$$-\frac{\partial J}{\partial z} + |q| \frac{n}{\tau_n} = -|q| \frac{\partial n}{\partial t}$$

Equação 104

Novamente, não fizemos nenhuma menção ao tipo de corrente, então as equações acima valem tanto para corrente de arrasto, como de difusão, ou mesmo uma combinação das duas.

Agora, vamos focar em um caso específico. Vamos supor que exista certa corrente qualquer até um ponto e depois desse ponto a corrente seja só de difusão, como ilustrado

na Figura 10. Portanto, para $z < 0$ temos uma corrente qualquer de cargas, **mas para $z > 0$ a corrente é só de difusão**. Queremos saber qual é a concentração de cargas para $z > 0$ em regime estacionário. Vou resolver para cargas positivas, mas o cálculo para cargas negativas é praticamente o mesmo (só trocando o sinal da carga).

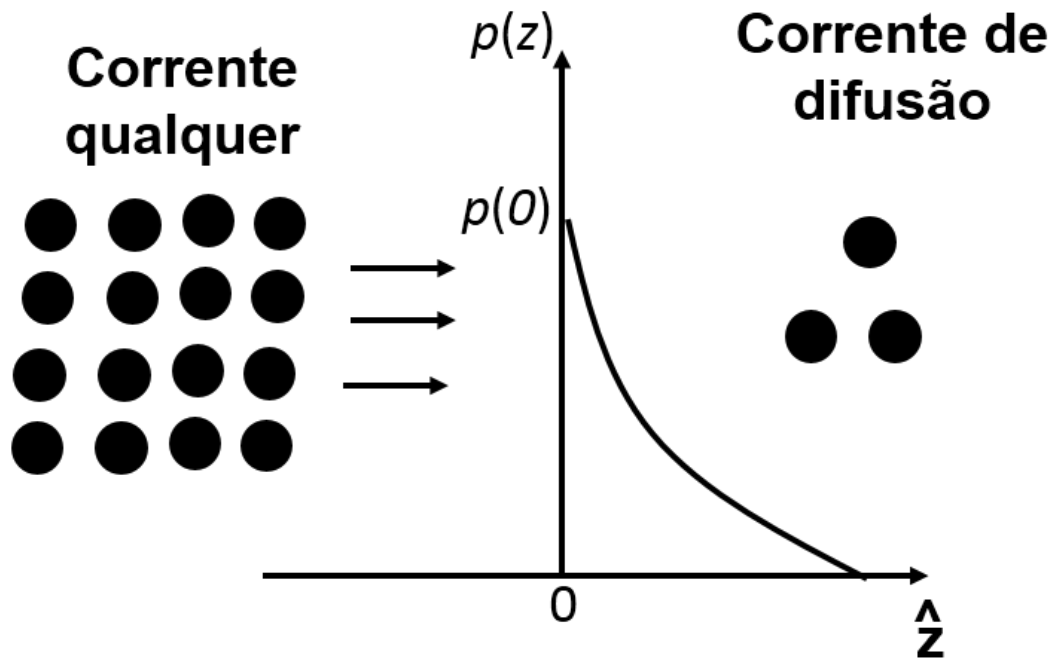


Figura 10 Ilustração do comportamento da densidade de portadores quando o transporte é por difusão. Neste exemplo, o transporte é por difusão para $z > 0$, o que implica em uma queda exponencial na densidade de portadores.

Em regime estacionário, a concentração de cargas não depende mais do tempo, então a derivada temporal na Equação 103 passa a ser zero. Assim temos que:

$$-\frac{\partial J}{\partial z} = |q| \frac{p}{\tau_p}$$

Equação 105

Agora, como sabemos que a corrente para $z > 0$ é só de difusão, então, pela Equação 90, temos que:

$$J = J_{dif} = -D_p \frac{\partial p}{\partial z} = -|q| D_p \frac{\partial p}{\partial z}$$

para $z > 0$

Equação 106

O índice p no coeficiente de difusão serve para diferenciar o coeficiente de difusão para buracos do coeficiente de difusão para elétrons. Substituindo a Equação 106 na Equação 105:

$$\frac{\partial^2 p}{\partial z^2} = \frac{p}{D_p \tau_p}$$

Equação 107

A solução geral dessa equação diferencial é:

$$p(z) = A \cdot \exp\left(-\frac{z}{L_p}\right) + B \cdot \exp\left(+\frac{z}{L_p}\right)$$

Equação 108

Onde L_p é chamado de comprimento de difusão e é dado por:

$$L_p = \sqrt{D_p \tau_p}$$

Equação 109

Como sabemos que $p(z)$ tem que ir para zero quando z vai para infinito, temos que $B = 0$. Além disso, temos também que $p(0) = A$. Portanto:

$$p(z) = p(0) \cdot \exp\left(-\frac{z}{L_p}\right)$$

Equação 110

Se você repetir o processo para as cargas negativas (faça isso sozinho), vai encontrar também que:

$$n(z) = n(0) \cdot \exp\left(-\frac{z}{L_n}\right)$$

Equação 111

Esses são dois resultados importantíssimos: **eles nos dizem que, se o transporte for por difusão, então a concentração de cargas decairá exponencialmente.** Por isso

que L é chamado de comprimento de difusão: ele é o comprimento típico que as cargas difundem antes de recombinarem. Utilizaremos essas expressões nos capítulos seguintes.

1.12 Sugestões bibliográficas

Para o propósito desse curso, o livro que considero mais adequado é Charles Kittel: *Thermal Physics*. Este é um livro utilizado no mundo todo, então é bem fácil de encontrá-lo. Caso o aluno queira estudar a disciplina para além do estritamente necessário nesse curso, recomendo também Ashley Carter: *Classical and Statistical Thermodynamics*. Caso o aluno estude o livro do Kittel ou do Carter e queira aprofundar ainda mais os seus conhecimentos, a referência mundial em Física Estatística a nível mais avançado é Reif: *Fundamentals of Statistical and Thermal Physics*.

Em relação aos tópicos de mecânica quântica, recomendo o primeiro capítulo do livro do Cohen-Tannoudji: *Quantum Mechanics*, Vol. 1. Os dois volumes desse livro do Cohen são excelentes e o livro é dividido de maneira a cobrir tanto um curso de graduação, quanto um de pós. Para o nosso propósito, só o primeiro capítulo já seria de grande ajuda.

1.13 Exercícios

Exercício 1

a) Um sistema é composto por um único elétron em um poço quântico de comprimento L . Utilizando o requerimento de acúmulo de fase múltiplo de 2π em um giro, deduza os valores possíveis de níveis de energia do elétron.

b) Suponha que o elétron esteja no nível de energia mais baixo. Se luz for incidente no poço, o elétron poderá absorver um fóton e ser excitado do nível de energia mais baixo até os níveis superiores. Utilizando conservação de energia, deduza uma expressão para o comprimento de onda do fóton que excite o elétron do nível de energia mais baixo até o segundo nível. Repita o procedimento para a transição entre o nível mais baixo e o terceiro nível. Para resolver esse exercício, você tem que relacionar a energia do fóton com o seu comprimento de onda. Para isso, lembre-se que a energia do fóton é igual a constante de Planck vezes a frequência.

c) Agora utilize os resultados do item b) para encontrar os valores de L (comprimento da caixa) que resultam em absorção de fótons vermelhos, o valor de L que resulta em absorção de fótons verdes e o valor de L que resulta em absorção de fótons azuis. Para resolver esse problema, consulte os valores das constantes na Internet.

Exercício 2

Um sistema é constituído por um elétron em um poço quântico tridimensional. Suponha que o poço seja um cubo de lado L . Se $g(E)$ é a multiplicidade da energia E , encontre:

a) $g\left(\frac{3h^2}{8L^2m}\right)$

b) $g\left(\frac{6h^2}{8L^2m}\right)$

Exercício 3

Dois sistemas, A e B, possuem temperaturas e entropias T_a , T_b , S_a e S_b , respectivamente. Os sistemas são postos em contato térmico, ou seja, os sistemas

podem trocar energia. Mostre que a entropia total do sistema vai aumentar se energia fluir do sistema com temperatura mais alta para o sistema com temperatura mais baixa.

Exercício 4

Um sistema está em equilíbrio térmico com um reservatório a uma temperatura T . Deduza uma expressão para a probabilidade de o sistema estar em um determinado estado com energia U .

Exercício 5

Suponha que o seu pai seja muito preocupado com a sua educação. Sabendo que semicondutores é uma das disciplinas mais importantes que você vai cursar, seu pai te pede para dar uma olhada em suas notas de aula. Após algumas horas, seu pai te chama para conversar. Abatido e consternado, seu pai te diz: *“Filho, tenho uma péssima notícia para te dar: seu professor é burro! Primeiro, ele fala que sistemas macroscópicos terão a propriedade mais provável. Lembre-se que você aprendeu em cálculo que um ponto de máximo possui derivada zero. Mas o seu professor te fala a seguir que a temperatura de um sistema é o inverso da derivada da entropia em relação à energia. Olha a conclusão absurda que vem dessas duas informações: de cálculo você aprendeu que a propriedade mais provável será aquele onde a derivada da entropia em relação à energia é zero! Mas, se a temperatura é o inverso da derivada da entropia, então o ponto de máximo corresponde à uma temperatura infinita! Então, de acordo com o seu professor, sistemas macroscópicos estão sempre na propriedade mais provável, o que equivale a dizer que sempre terão temperatura infinita!! Isso é um absurdo.*

Mas não para por aí. Seu professor te diz que toda a Física Estatística é construída com base na suposição que todos os estados de um sistema são igualmente prováveis. Mas daí, algumas linhas depois, seu professor te diz que, se existem N estados possíveis, então a probabilidade de um sistema estar no estado c é dado por $P(c) = g_R(U_0 - \varepsilon_c)/N$. Como assim? Qualquer um sabe que, se todos os estados são igualmente prováveis e se existem N estados possíveis, então a probabilidade de encontrar qualquer estado deveria ser $1/N$, e não $g_R(U_0 - \varepsilon_c)/N$. Isso é uma flagrante contradição”.

Defenda o seu professor e explique como essas duas aparentes contradições são esclarecidas.

Exercício 6

Considere um sistema A qualquer em contato com um reservatório. Suponha que a energia mínima do sistema A seja ε_0 , e que apenas um único estado, chamado de *ground-state*, tenha a energia mínima ε_0 . Chame de ε_1 a energia imediatamente acima de ε_0 . Prove que, se $\varepsilon_1 - \varepsilon_0 \gg k_B T$, então é praticamente certo que o sistema A estará no *ground-state*. Interprete seu resultado.

Exercício 7

A multiplicidade de um gás ideal é dada por $g(U) = CU^{3N/2}$, onde U é a energia do gás, C é uma constante e N é o número de partículas do gás. Prove que a energia do gás a uma temperatura T é dada por $U = (3/2)Nk_B T$.

Exercício 8

Temperatura é um problema sério na indústria de semicondutores. Por exemplo, suponha que você projete um poço quântico para absorver radiação em um certo comprimento de onda, ajustando a diferença de energia entre o *ground-state* e a energia imediatamente superior. Para que o seu dispositivo funcione, você precisa que o elétron esteja no *ground-state* para receber a radiação. Mas, dependendo das energias e da temperatura, ele pode pular para estados com energias mais altas. Esse “pulo” é chamado de excitação térmica e pode limitar muito a performance dos dispositivos. No nosso exemplo, a performance seria prejudicada porque se o elétron não estiver no *ground-state*, a radiação com comprimento de onda de projeto não será absorvida.

Nesta questão, vocês vão analisar se excitação térmica é um problema para pontos quânticos operando na região visível.

Considere então um sistema formado por um único elétron preso à um poço quântico tri-dimensional. Primeiro, você deve encontrar a largura do poço (considere que é um cubo) de maneira que a probabilidade de se encontrar um elétron FORA do *ground-state* seja negligenciável para temperatura ambiente. Nesse problema, você pode ignorar a presença do spin, uma vez que isso ainda não foi discutido em aula. A

seguir, você deve utilizar o seu resultado para concluir se excitação térmica é ou não um problema para poços quânticos operando no visível.

Seguem as dicas/passos para resolver o problema:

a) Primeiro precisamos resolver o que é uma probabilidade aceitável. Vamos escolher, de maneira meio arbitrária, que a maior probabilidade aceitável de se encontrar o elétron FORA do *ground-state* seja $P = 0.01$ (ou seja, 1%).

b) Podemos também simplificar o problema assumindo que, a probabilidade de encontrar o elétron em um estado com energia diferente da do *ground-state*, ou diferente da energia imediatamente superior à do *ground-state*, seja zero. Em outras palavras, além do *ground-state*, somente os estados que possuem a energia imediatamente superior à do *ground-state* precisam ser considerados: os outros podem ser ignorados. Vamos chamar a energia do *ground-state* de ε_0 e a energia imediatamente superior de ε_1 . Então, repetindo, além do *ground-state*, somente estados com energia ε_1 precisam ser considerados.

c) Com essas considerações, expresse o comprimento L do poço quântico tridimensional em termos da probabilidade P de se encontrar o elétron FORA do *ground-state*. Nessa parte, **deixe tudo em termos de símbolos** e explique o seu raciocínio. Dica: primeiro expresse P , e posteriormente L , em termos da diferença $\varepsilon_1 - \varepsilon_0$.

d) Agora substitua os números e encontre L , assumindo temperatura ambiente (especifique o valor da temperatura que você utilizou).

e) Qual é o comprimento de onda da radiação que o seu poço quântico absorveria? Faz parte do visível?

f) Se você reduzir o tamanho do poço, você aumenta ou diminui a probabilidade de se encontrar o elétron FORA do *ground-state*? Em outras palavras, reduzir o tamanho do poço piora ou melhora o problema da excitação térmica? Explique.

g) Qual é a conclusão? Excitação térmica é um problema para pontos quânticos operando no visível? Explique.

Exercício 9

O potencial químico de um gás ideal depende da concentração do gás de acordo com a relação:

$$\mu = k_B T \ln \left(\frac{n}{n_Q} \right)$$

Onde n é a concentração do gás (número de moléculas por unidade de volume) e n_Q é a concentração quântica, dada por:

$$n_Q = \left(\frac{2 \cdot \pi \cdot m \cdot k_B \cdot T}{h^2} \right)^{3/2}$$

No livro *Thermal Physics*, do Charles Kittel, tem um exemplo muito elegante do uso do potencial químico para determinar a concentração atmosférica: tratando o ar como um gás ideal, sabemos que o potencial total do gás deve ser constante no equilíbrio. Em outras palavras, o potencial total não pode depender da altura. Mas o potencial total é igual ao potencial químico mais o potencial gravitacional:

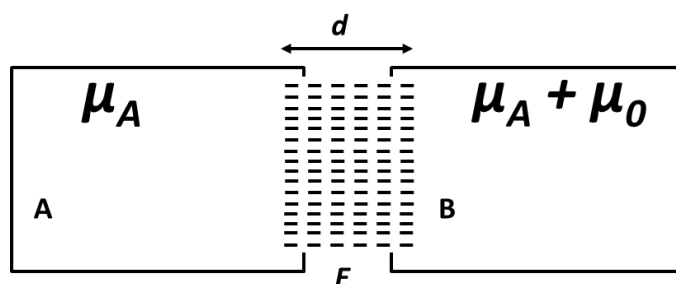
$$\mu_{total} = k_B T \ln \left(\frac{n}{n_Q} \right) + mgy$$

Utilizando o fato de que o potencial total deve ser constante (ou seja, independente de y), descubra como a concentração depende da altura.

Exercício 10

Suponha novamente que dois sistemas, A e B, possuem potenciais químicos diferentes: o potencial químico para elétrons no sistema B é maior que o potencial

químico no sistema A, como mostrado na figura abaixo. Os sistemas são inicialmente neutros, ou seja, a quantidade de cargas total é nula. Quando os dois sistemas são postos em contato difusivo, elétrons vão migrar do sistema B para o sistema A, causando acúmulo de cargas negativas no sistema A e positivas no sistema B. Suponha que o campo elétrico resultante seja constante e concentrado em uma região de comprimento d , na interface entre os sistemas, como mostrado na figura abaixo. Se, após a difusão cessar, o potencial químico do sistema A for μ_A e o potencial químico no sistema B for $\mu_A + \mu_0$, encontre o módulo do campo elétrico resultante em função dos potenciais químicos e da distância d . Determine também o sentido do campo elétrico (de A para B ou de B para A?).



Exercício 11

Quando uma bolinha é solta de certa altura e ganha velocidade na queda, dizemos que energia potencial gravitacional foi transformada em energia cinética. Analogamente, quando temos um resistor em um circuito D.C., dizemos que energia potencial eletrostática está sendo transformada em energia térmica (resistor está aquecendo).

Agora, considere novamente o problema do exercício 8, onde temos dois materiais com potenciais químicos diferentes e vamos considerar o regime transitório, onde o potencial eletrostático ainda não cancelou a difusão. No regime transitório, as cargas estão se movendo, portanto, existe uma corrente. Por outro lado, se o material tiver uma resistência R , haverá dissipação de energia ($P = RI^2$). De onde está vindo essa energia?

Exercício 12

Partindo da soma de Gibbs, deduza a distribuição de Fermi-Dirac e explique o seu significado físico.

Exercício 13

Considere o sistema “um elétron preso à um poço quântico” em contato com um reservatório com temperatura T . Se eu te perguntar qual é a probabilidade de eu encontrar esse elétron no estado caracterizado pela trinca $c_x = 1, c_y = 1, c_z = 1$; a resposta é dada Equação 31, com energia dado por $\varepsilon_{1,1,1} = \frac{h^2}{8L^2m}$ e função de partição dado pela Equação 59. Mas a pergunta “qual é a probabilidade de eu encontrar esse elétron no estado $c_x = 1, c_y = 1, c_z = 1$ ” é idêntica à pergunta “qual é a probabilidade de eu encontrar o orbital $c_x = 1, c_y = 1, c_z = 1$ ocupado”. Então por que a resposta não é dada pela distribuição de Fermi-Dirac?

Exercício 14

Suponha que um sistema A possua apenas quatro orbitais. Chamaremos esses orbitais de O_1, O_2, O_3 e O_4 . Os orbitais O_1 e O_2 possuem energia ε_1 , enquanto os orbitais O_3 e O_4 possuem energia ε_2 . O sistema A está em contato térmico (mas não difusivo) com um reservatório cuja temperatura é T . Se o sistema A possui dois elétrons, qual é a probabilidade do orbital O_1 estar ocupado?

Exercício 15

Suponha que um gás ideal de elétrons possua energia média $\langle U \rangle$ e potencial químico μ . Expresse a probabilidade de se encontrar um determinado orbital ocupado em função de $\langle U \rangle, \mu$ e da energia do orbital.

Exercício 16

Vimos que uma das ideias centrais em Física Estatística é identificar a propriedade de fato de um sistema macroscópico em equilíbrio com a propriedade mais provável. Vimos também que a propriedade mais provável é expressa matematicamente pela Equação 13. Por outro lado, vimos que a probabilidade de encontrar um sistema (que está em equilíbrio com um reservatório) em um determinado estado é dada pela Equação 31. Se o sistema ao qual se refere a Equação 31 for microscópico, poderemos

afirmar que ele está na sua propriedade mais provável? E se ele for macroscópico, poderemos afirmar que ele está na sua propriedade mais provável? E como a Equação 31 poderia ser utilizada para determinar qual é a propriedade mais provável do sistema em contato com o reservatório?

2 – Semicondutores

Neste capítulo vamos utilizar as noções apreendidas até aqui para podermos estudar as propriedades elétricas de semicondutores. Mas antes disso, precisamos definir melhor o que é um semicondutor, tarefa que exige uma introdução à importante teoria de bandas.

2.1 Teoria de Bandas

Vimos, no capítulo anterior, que os estados de partículas quânticas são discretos. Como exemplo, calculamos os níveis de energia de um elétron preso em um poço quântico. Esse exemplo, apesar de simples, é suficiente para descrever qualitativamente o comportamento energético de um elétron preso ao núcleo de um átomo: em vez do poço quântico, o elétron está preso ao núcleo em razão da energia potencial eletrostática imposta pelo núcleo, o que resulta em estados discretos. O cálculo dos estados neste último caso é bem mais complicado e requer a solução da equação de onda do elétron (chamada de Equação de Schrödinger) submetido a este potencial em particular; mas o resultado é qualitativamente o mesmo: o elétron preso ao núcleo de um átomo também tem níveis de energia discretos. Estes níveis de energia são os orbitais químicos *s*, *p*, *d* e *f*.

Estamos interessados em estudar semicondutores, que são compostos por muitos, muitos átomos. Semicondutores são cristais, ou seja, são materiais que apresentam uma distribuição periódica de átomos. Apesar de essa propriedade facilitar (e muito) o tratamento matemático do problema, não temos espaço aqui para formalizar esse tratamento (para quem quiser saber mais, façam um *Google* do termo *Bloch theorem*). Entretanto, podemos utilizar um problema análogo para ganharmos uma intuição sobre o que ocorre quando juntamos vários átomos. Vamos começar, então, identificando e justificando o problema análogo.

Suponha que tenhamos um átomo de hidrogênio (esse é o caso mais simples, pois o átomo de hidrogênio possui apenas um elétron). O elétron está preso pelo núcleo e no orbital de energia mais baixo. Pergunta: podemos encontrar um sistema clássico que seja matemático e fisicamente análogo ao átomo de hidrogênio?

Vamos pensar na analogia física, primeiro. Queremos encontrar um sistema clássico onde a partícula esteja “presa” no sistema (como o elétron está preso no poço quântico ou no átomo de hidrogênio) e que ela tenha alguma frequência de oscilação própria (no caso do elétron, a frequência de oscilação é proporcional à energia do estado). O caso

mais simples que atende a esses requisitos é o sistema massa-mola. Lembre-se que esse sistema é chamado de oscilador harmônico. Oscilador porque a mola oscila em torno do ponto de repouso; e harmônico porque a frequência de oscilação é bem definida (a oscilação acontece com uma única frequência). A analogia, então, é: enquanto o elétron está preso no átomo de hidrogênio pelo potencial elétrico, a massa está presa pelo potencial da mola (ou potencial de *Hooke*). Além disso, o elétron oscila em uma frequência bem definida determinada pela energia do orbital, enquanto a massa oscila na frequência de ressonância, que depende do valor da massa e da constante da mola.

E o que podemos dizer em relação à analogia matemática? Todo problema quântico se reduz à solução de uma equação de onda (a Equação de Schrödinger). A equação de onda, por outro lado é uma equação diferencial. A solução de uma equação diferencial pode ser elegantemente expressa em termos de autovetores e autovalores. No caso da equação de Schrödinger, os autovalores são os níveis de energia, e os autovetores são as funções de onda correspondentes. Por outro lado, o sistema massa-mola também é descrito por uma equação diferencial (como todo sistema clássico que obedece à segunda lei de Newton), onde o autovalor é a frequência de ressonância e o autovetor é a função do tempo que descreve a posição da massa.

Devido às analogias física e matemática entre os problemas, podemos, com certa confiança, postular que o comportamento de dois sistemas massa-mola acoplados será qualitativamente semelhante ao comportamento de dois átomos que interagem. Mas vamos por partes. Primeiro vamos lembrar como é a formulação matemática do oscilador harmônico. A Figura 11 mostra um sistema massa mola onde a mola possui constante k_0 e o ponto de repouso (onde a força é zero) é dado por $x = a$.

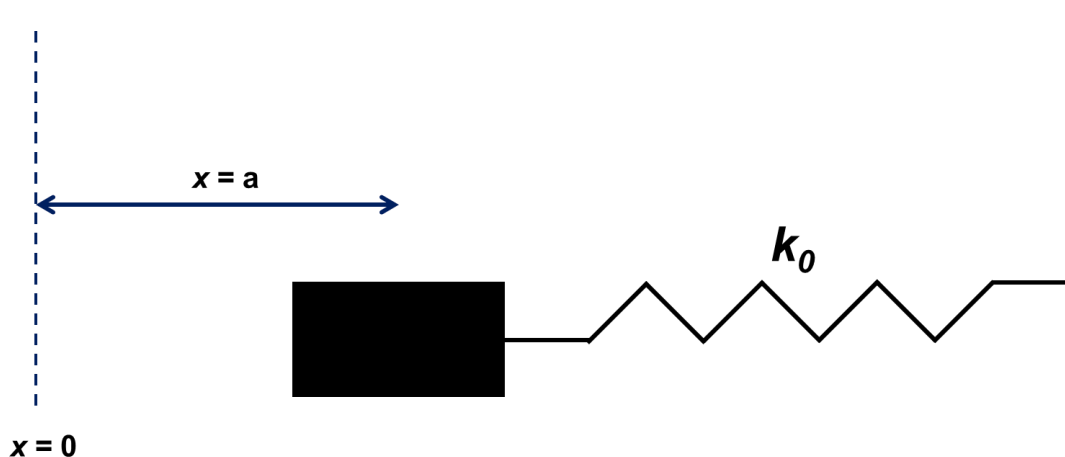


Figura 11 Oscilador harmônico exemplo. O ponto $x = a$ corresponde ao ponto de repouso (a mola não está nem esticada nem comprimida: a força total neste ponto é zero).

De acordo com a lei empírica de *Hooke*, a força exercida pela mola é proporcional à k_0 e ao deslocamento em relação ao ponto de repouso. Assim, a segunda lei de Newton toma a forma:

$$\frac{d^2x}{dt^2} + \frac{k_0}{m}(x - a) = 0$$

Equação 112

Onde m é a massa da mola. A solução desta equação é uma senoide:

$$x = A \cdot \cos(\omega_0 t + \varphi)$$

Equação 113

A fase e amplitude da senoide são determinadas pelas condições iniciais. O parâmetro mais importante, entretanto, é a frequência de ressonância, que é dada por:

$$\omega_0 = \sqrt{\frac{k_0}{m}}$$

Equação 114

Se você já sabia tudo isso de cabeça, ótimo. Se, por outro lado, você não se lembra muito bem da teoria do oscilador harmônico, então sugiro que você substitua a Equação 113 na Equação 112 para verificar se ela realmente é uma solução, com a frequência de ressonância dada pela Equação 114.

O oscilador harmônico sozinho é o sistema análogo ao átomo isolado. Dois átomos, portanto, são análogos a dois osciladores harmônicos. Além disso, dois átomos interagindo (ou seja, dois átomos perto o suficiente para que os potenciais elétricos se sobreponham) são análogos a dois osciladores harmônicos que interagem, ou seja, que estão acoplados. A Figura 12 ilustra o sistema completo, onde o acoplamento é dado pela mola em vermelho, com constante Ω .

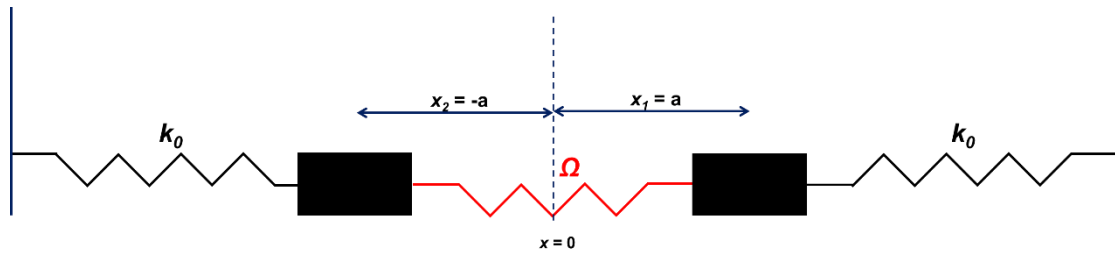


Figura 12 Dois osciladores harmônicos acoplados. A mola em vermelho, com constante Ω , representa o acoplamento

Não existe muito segredo a respeito da formulação matemática do oscilador harmônico acoplado. Como temos que aplicar a segunda lei de Newton em todas as massas do sistema, teremos duas equações diferenciais acopladas, cada uma correspondendo à lei de Newton em cada uma das massas:

$$\frac{d^2 x_1}{dt^2} + \frac{k_0}{m}(x_1 - a) + \frac{\Omega}{m}(x_1 - x_2 - b) = 0 \quad (1)$$

$$\frac{d^2 x_2}{dt^2} + \frac{k_0}{m}(x_2 + a) - \frac{\Omega}{m}(x_1 - x_2 - b) = 0 \quad (2)$$

Equação 115

A Equação 115(1) corresponde à segunda lei de Newton aplicada à massa da direita, cuja coordenada chamamos de x_1 , enquanto a Equação 115(2) corresponde à segunda lei de Newton aplicada na massa da esquerda, com coordenada x_2 . Note que o parâmetro b corresponde ao comprimento de repouso da mola de acoplamento (quando a distância entre as duas massas for igual a b , não haverá força advinda dessa mola). Confira sozinho se os sentidos das forças na Equação 115 estão corretos.

Para limpar a notação um pouco, vamos reescrever a Equação 115 como:

$$\frac{d^2 x_1}{dt^2} + \omega_0^2(x_1 - a) + \Upsilon^2(x_1 - x_2 - b) = 0 \quad (1)$$

$$\frac{d^2 x_2}{dt^2} + \omega_0^2(x_2 + a) - \Upsilon^2(x_1 - x_2 - b) = 0 \quad (2)$$

Equação 116

Onde ω_0 é dado pela Equação 114 e $\Upsilon = \sqrt{\Omega/m}$.

As equações acima estão acopladas (a primeira equação depende de x_2 e a segunda equação depende de x_1), portanto as soluções delas não serão iguais às dos sistemas isolados. O método geral de solução de equações acopladas envolve encontrar os autovetores e autovalores do sistema, o que equivale a diagonalizar as matrizes que representam o sistema. Mas, neste caso em particular, podemos tomar um atalho para resolver o problema. Note que podemos nos livrar do termo de acoplamento somando as duas equações, que resulta em:

$$\frac{d^2(x_1 + x_2)}{dt^2} + \omega_0^2(x_1 + x_2) = 0$$

Equação 117

Como as massas são iguais, o centro de massa é dado por:

$$x_{cm} = \frac{x_1 + x_2}{2}$$

Equação 118

Substituindo a Equação 118 na Equação 117 temos:

$$\frac{d^2 x_{cm}}{dt^2} + \omega_0^2 x_{cm} = 0$$

Equação 119

O que a Equação 119 nos diz? Ela é idêntica à equação de um oscilador harmônico simples (compare com a Equação 112) acoplado a uma mola com ponto de repouso em $x = 0$ e cuja frequência de ressonância é ω_0 . Isso quer dizer que o centro de massa do sistema acoplado pode oscilar com frequência de ressonância igual à frequência do sistema isolado. Portanto, somando as equações diferenciais acopladas, concluímos que o sistema como um todo (representado pelo centro de massa) pode oscilar com frequência igual à frequência de ressonância de um único oscilador.

E se subtraímos as equações acopladas em vez de somarmos? Neste caso, teremos:

$$\frac{d^2(x_1 - x_2)}{dt^2} + \omega_0^2(x_1 - x_2) + 2\Upsilon^2(x_1 - x_2) - 2(\omega_0^2 a + \Upsilon^2 b) = 0$$

Equação 120

Definindo a distância relativa entre as massas como:

$$x_R = x_1 - x_2$$

Equação 121

Temos:

$$\frac{d^2 x_R}{dt^2} + (\omega_0^2 + 2\Upsilon^2)x_R - 2(\omega_0^2 a + \Upsilon^2 b) = 0$$

Equação 122

Como o último termo da Equação 122 é uma constante, essa também é a equação de um oscilador harmônico, cuja frequência de ressonância é igual a:

$$\omega_R = \sqrt{\omega_0^2 + 2\Upsilon^2}$$

Equação 123

Confirme sozinho que a solução da Equação 122 é dada por:

$$x_R = A \cdot \cos(\omega_R t + \phi) + \frac{2(\omega_0^2 a + \Upsilon^2 b)}{\omega_R^2}$$

Equação 124

Portanto, existe um segundo modo de oscilação possível, onde a distância relativa entre as massas oscila com frequência ω_R . A solução mais geral é a combinação linear do modo do centro de massa com o modo da distância relativa. Mas, o que nos interessa aqui é que, em vez de termos uma única frequência de ressonância, agora temos duas: ω_0 e ω_R . Podemos interpretar esse resultado da seguinte forma: se tivéssemos as duas massas independentes (ou seja, se não tivéssemos a mola de acoplamento), teríamos uma única frequência de ressonância: ω_0 . Mas como temos duas massas, poderíamos dizer que temos dois modos de oscilação (um para cada massa desacoplada), compartilhando a mesma frequência de oscilação (a propósito, note, pela Equação 123, que a frequência de ressonância ω_R tende à ω_0 quando o acoplamento Υ tende a zero).

Ou seja, temos dois estados com a mesma frequência (cada estado corresponde à oscilação de uma massa). Mas, quando colocamos o acoplamento, essas duas frequências idênticas foram divididas em duas frequências diferentes: ω_0 e ω_R , como representado na Figura 13a. Note também que se tivéssemos três sistemas massa-molas acoplados, então, teríamos três equações diferenciais acopladas, três modos de oscilação independentes e, conseqüentemente, três frequências de ressonância. Assim, nossa observação pode ser generalizada como: se houverem n osciladores acoplados, então existirão n modos de oscilação (que são análogos aos estados) e n frequências de ressonâncias (que são análogas à energia dos estados).

Analogamente, se tivermos dois átomos idênticos separados, seus orbitais (que são os estados) e respectivas energias serão idênticas. Mas se os dois átomos interagirem, para cada energia dos orbitais independentes dos átomos isolados, aparecerão duas energias diferentes (porque, efetivamente, teremos dois novos orbitais). Se três átomos interagirem, cada energia de cada orbital se repartiria em três, e assim por diante, como representado na Figura 13b (note que estamos representando apenas dois orbitais nessa figura). Mas um sólido possui muitos, muitos átomos. Por exemplo, se tivermos 1 mol de átomos, teremos 1 mol de níveis de energia para cada orbital do átomo isolado. Mas esses níveis estarão tão próximos uns dos outros que poderemos tratá-los como um contínuo de níveis, como representado na Figura 13b. Cada um desses contínuos de energia é chamado de banda de energia.

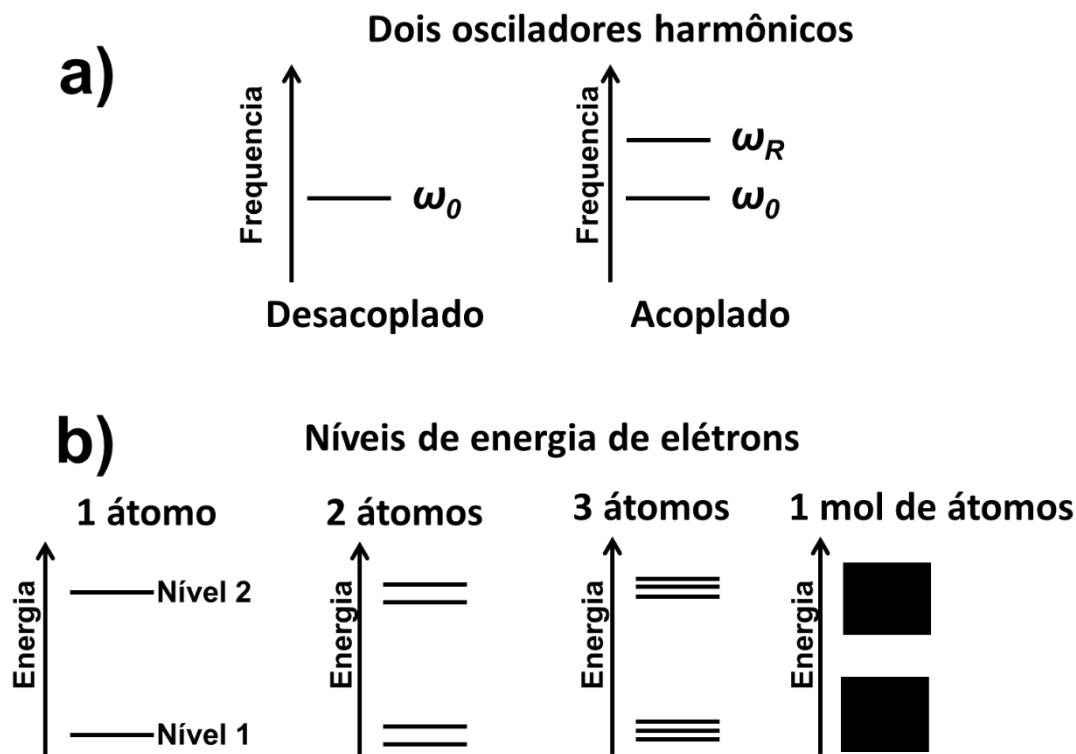


Figura 13 a) separação da frequência de ressonância do oscilador harmônico acoplado. b) efeito análogo na interação entre dois ou mais átomos: quando dois átomos interagem, a energia de cada orbital é separada em dois níveis com energias diferentes. Quando três átomos interagem, teremos três níveis de energia para cada orbital do átomo isolado. Quando 1 mol de átomos interage, temos tantos níveis de energia, tão próximos uns dos outros, que podemos tratá-los como um contínuo de níveis, formando assim as bandas de energia.

Na Figura 13b representamos um exemplo onde duas bandas de energia foram formadas, mas não falamos nada sobre como as bandas são preenchidas. Por exemplo, vamos supor que o átomo isolado tenha um único elétron, e este elétron ocupe o orbital com energia mais baixa, (a energia mais baixa é representada pelo nível 1 na Figura 13b). Se colocarmos dois átomos juntos, teremos dois elétrons, e eles estarão nos níveis de energia mais baixos (níveis estes que surgem da separação do nível 1 em dois níveis distintos). Se tivermos n átomos, então teremos n níveis de energia associados à separação do nível 1. Como teríamos também n elétrons, então todos esses n níveis estariam preenchidos. Assim, para o nosso exemplo do átomo com um único elétron, se tivermos muitos, muitos átomos, de maneira que os níveis possam ser tratados como bandas, então teremos a primeira banda da Figura 13b totalmente preenchida e a segunda totalmente desocupada. Esse é um exemplo entre várias possibilidades: dependendo do tipo de átomo e do tipo de ligação química, podemos ter bandas totalmente preenchidas, parcialmente preenchidas, ou totalmente vazias.

É melhor representar o preenchimento das bandas “colorindo” a parte que está ocupada, como mostrado na Figura 14. Nessa figura, temos duas bandas representadas, sendo que no primeiro exemplo as duas estão preenchidas; no segundo, a de cima está parcialmente preenchida; no terceiro, a de baixo está parcialmente preenchida e a de cima está vazia; e no quarto exemplo a de cima está vazia e a primeira está totalmente preenchida.

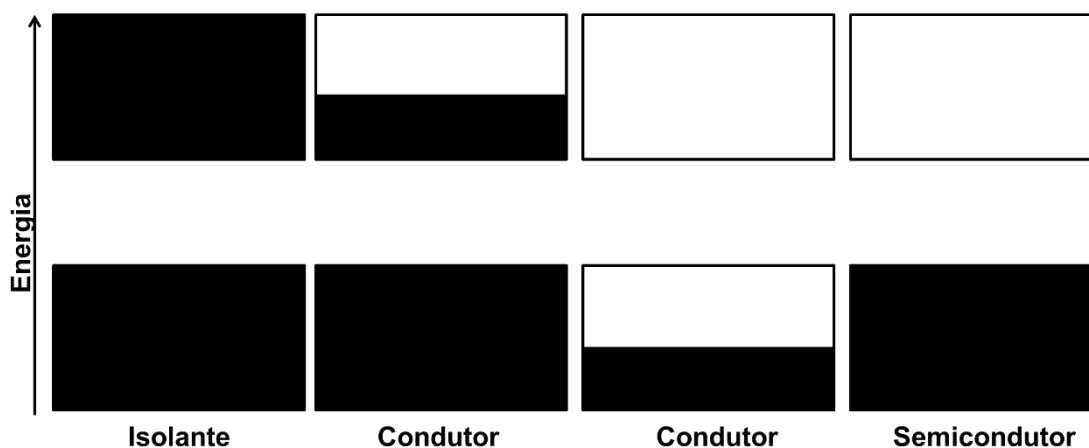


Figura 14 A ocupação das bandas de energia define as propriedades de condutividade do sólido.

Como já indicado pela Figura 14, é o preenchimento das bandas que determina as propriedades de condutividade do sólido. No primeiro caso, temos um isolante, porque todas as bandas estão preenchidas. Não é difícil entender o porquê de o primeiro caso ser um isolante: se aplicarmos um campo elétrico no material, o campo acelerará as cargas, ou seja, aumentará a energia delas; mas como não existem níveis de energia desocupados, não tem como as cargas ganharem energia, ou seja, não tem como as cargas serem aceleradas. Por isso, elas não respondem à aplicação de um campo elétrico externo. Se uma das bandas estiver parcialmente preenchida, por outro lado, existirão níveis disponíveis para os elétrons serem acelerados e o material, portanto, será um condutor. O caso de mais interesse para nós, por outro lado, é o caso onde a banda de baixo, que chamaremos de banda de valência, está totalmente preenchida, mas a banda de cima, que chamaremos de banda de condução, está vazia. Neste caso temos um semicondutor. Como você deve ter notado, o semicondutor é, na verdade, um material isolante (já que uma banda está completamente preenchida), mas como veremos a seguir, podemos facilmente alterar as propriedades de condução deste material através de reações químicas com outros átomos (dopagem).

2.2 Elétrons e buracos

Dissemos na seção anterior que o semicondutor é caracterizado por duas bandas de energia (a banda de valência e a banda de condução), com a banda inferior completamente preenchida e a banda superior completamente vazia. Essa condição, entretanto, só é perfeitamente satisfeita no limite de temperatura igual a zero kelvins (zero absoluto). De fato, alguns portadores são excitados termicamente da banda de valência para a banda de condução, induzindo certo grau de condutividade ao semicondutor, embora a condutividade seja tipicamente bem baixa.

Sabemos que os elétrons na banda de condução podem ser acelerados por um campo elétrico ou podem difundir caso exista um gradiente de concentração. Em outras palavras, os elétrons na banda de condução podem contribuir para a corrente no semicondutor. Dizemos, então, que esses são elétrons “livres”, pois existem níveis de energia para os quais os elétrons podem ser acelerados e também porque tipicamente eles apresentam fraca interação com a rede cristalina do semicondutor.

E os elétrons da banda de valência? Como a banda de valência só está totalmente preenchida no zero absoluto, então, em qualquer outra temperatura, a banda de valência apresentará estados disponíveis para que os elétrons ganhem energia com a aplicação de um campo elétrico externo. Esses estados disponíveis na banda de valência aparecem exatamente porque o elétron que estava nesses estados no zero absoluto foi excitado termicamente, deixando então esses estados da banda de valência para ocupar estados na banda de condução.

Além disso, apesar dos elétrons na banda de valência estarem geralmente bem confinados na estrutura cristalina, eles podem “pular” de um átomo para o outro, desde que o átomo vizinho possua estados disponíveis na banda de valência para receber o elétron.

Agora, imagine a seguinte situação, ilustrada na Figura 15: imagine que, por excitação térmica, um elétron de um átomo deixe um estado da banda de valência e ocupe um estado na banda de condução. Enquanto o elétron estava na banda de valência, ele estava confinado ao átomo. Isso significa que a função de onda do elétron estava localizada em torno do átomo. Mas, quando ele passa para a banda de condução, sua função de onda “delocaliza”, ou seja, o elétron se espalha pelo semicondutor inteiro, deixando de ser localizado. Portanto, esse salto de energia da banda de valência para a banda de condução resulta no surgimento de uma carga positiva local no átomo que

perdeu o elétron (a carga positiva é local, mas o semicondutor, como um todo, continua neutro, já que o elétron ainda está no semicondutor, só não está mais perto do átomo). Assim, o átomo que perdeu um elétron terá uma carga positiva e também um orbital de energia vazio (que era o orbital que o elétron ocupava): é como se o orbital vazio tivesse uma carga positiva (mas a carga positiva mesmo está no núcleo). Além disso, um elétron do átomo vizinho poderá pular para este átomo com um orbital vazio. Se isso ocorrer, o átomo vizinho passará a ter um orbital vazio e uma carga local positiva, como ilustrado na Figura 15. O que efetivamente ocorreu, então, é um deslocamento da carga positiva (e do orbital vazio). Como quem define a corrente é o deslocamento de cargas, o movimento da carga positiva associada ao movimento do orbital vazio induz a mesma corrente que seria induzida se uma partícula com carga positiva estivesse deslocando. Essa “partícula” é chamada de buraco, termo este que denota o fato de ela não ser bem uma partícula, mas sim uma falta de partícula, que resulta em uma carga positiva. É importante notar que o buraco pode ser tratado no mesmo patamar que o elétron: ele tem massa igual ao do elétron (associada, claro, à massa do elétron que desloca na direção contrária) e carga fundamental igual à do elétron, mas com sinal positivo. Finalmente, note que a presença do buraco não significa que o material como um todo ficou carregado: a carga positiva do buraco é compensada pela carga negativa do elétron livre na banda de condução. Lembre-se que foi a excitação de um elétron da banda de valência para a banda de condução que gerou o buraco, e não a retirada de um elétron do material. Mas, do mesmo jeito que o elétron na banda de condução é livre para se mover no material, o buraco que ele deixou também é. Existem, portanto, dois portadores de cargas que contribuem para a corrente em semicondutores: elétrons na banda de condução e buracos na banda de valência.

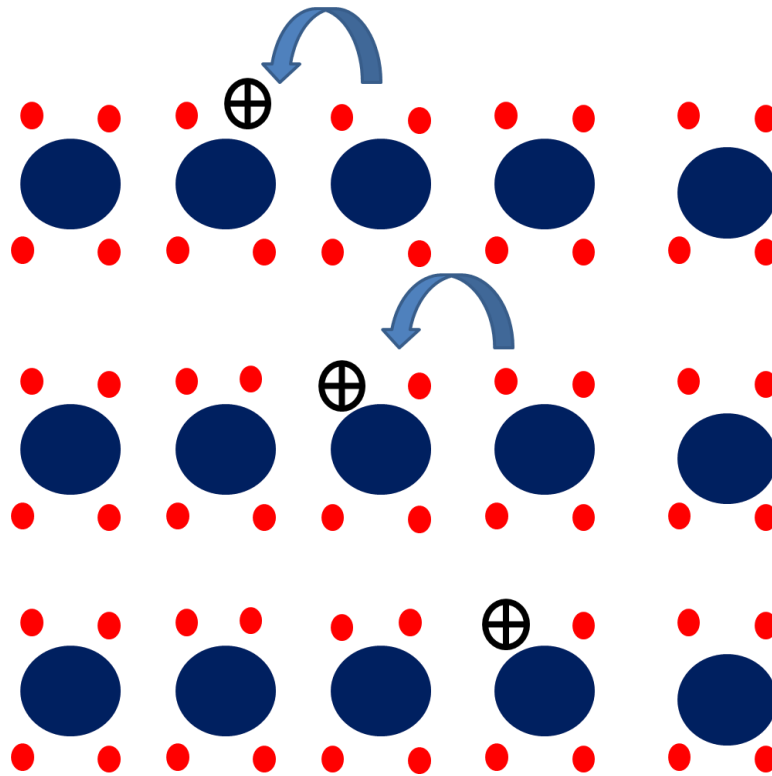


Figura 15 Representação do movimento de buracos. Representamos quatro elétrons por átomo (pontos vermelhos) preenchendo a banda de valência. Na linha de cima, o segundo átomo da esquerda para a direita perdeu um elétron na banda de valência por agitação térmica (foi excitado para a banda de condução). A falta do elétron constitui uma carga positiva (o buraco), que desloca para a direita à medida que os elétrons dos átomos vizinhos pulam para o átomo desocupado. Este movimento é representado pelas linhas de baixo. Note que o semiconductor como um todo não está carregado, pois o elétron que foi excitado para a banda de condução (e gerou o buraco) ainda está no semiconductor. Mas do mesmo jeito que esse elétron na banda de condução é livre para se mover, o buraco que ele deixou na banda de valência também é.

2.3 Concentração de elétrons livres

Como a condutividade dos semicondutores depende da concentração de portadores de carga livres, e já identificamos que os portadores livres são os elétrons na banda de condução e os buracos na banda de valência, temos que agora focar nossa atenção em como encontrar a concentração desses portadores. Vamos começar pelo cálculo de elétrons livres, ou seja, com o cálculo da concentração de elétrons que estão na banda de condução (a concentração é o número total de elétrons na banda de condução dividido pelo volume do sólido).

Na verdade, já sabemos como fazer isso: na seção 1.7 calculamos a distribuição de Fermi-Dirac, que é a probabilidade de um orbital (um estado) estar ocupado por um elétron. Então, já sabemos qual é a probabilidade de cada um dos orbitais da banda de condução estarem ocupados. Para sabermos qual é a concentração total, basta somarmos a probabilidade de cada um desses orbitais estar ocupados. Antes de fazer a

conta, entretanto, vamos dar um exemplo para não perdermos de vista o que estamos realmente fazendo.

O exemplo é o seguinte: suponha que eu tenha dois orbitais na banda de condução, cada um com probabilidade de ocupação igual a 0.5. Se eu somar as duas probabilidades, terei o número 1. Então, meu resultado seria 1. O que isso significa? Significa que eu identifiquei que dois orbitais com probabilidade 0.5 de ocupação correspondem a um orbital efetivamente ocupado. Em outras palavras, essa situação corresponde a um elétron na banda de condução. Note que fiz um salto entre a probabilidade e a realidade: como eu sabia que cada orbital tinha metade de chance de estar ocupado, considere que um deles está ocupado e o outro vazio. Se eu tiver quatro orbitais com probabilidade 0.5 cada, vou considerar que tenho dois orbitais ocupados, ou seja, que tenho dois elétrons. É claro que um raciocínio desses só funciona quando temos muitos, muitos orbitais, o que é o caso de sólidos.

Um último comentário antes de partirmos para o cálculo: talvez você esteja desconfortável com o fato de que estamos somando probabilidades e obtendo números maiores que 1. Mas lembre-se novamente o que é a distribuição de Fermi-Dirac: ela se refere à probabilidade de um único orbital estar ocupado. O que tem que dar 1 é a probabilidade do orbital estar ocupado mais a probabilidade do orbital estar desocupado. Quando somamos a probabilidade de um orbital estar ocupado com a probabilidade de outro orbital estar ocupado, estamos somando probabilidades independentes; por isso que o resultado pode ser, e quase sempre é, um número maior que 1, o qual associamos com o número total de portadores de cargas livres.

Então, vamos colocar em linguagem matemática a tarefa que temos que realizar. Como acabamos de estabelecer, o número total de elétrons livres N_e será a soma das probabilidades de ocupação de todos os orbitais da banda de condução. Em termos matemáticos:

$$N_e \cong \sum_{bc} f(\varepsilon_{bc}) = \sum_{bc} \frac{1}{1 + \exp\left(\frac{\varepsilon_{bc} - \mu}{k_B T}\right)}$$

Equação 125

Note que eu troquei o índice c pelo índice bc . Fiz isso para enfatizar que estamos somando somente os orbitais da banda de condução, já que estamos interessados no número total de elétrons **livres**. O subscrito e de N_e também serve para lembrarmos que

estamos calculando o número total de elétrons na banda de condução (mais para frente calcularemos N_p , que é o número total de buracos).

Vamos dedicar um bom espaço agora ao cálculo de N_e , inclusive trocando a notação para adequá-la à maioria dos livros texto de semicondutores. Mas é de suma importância que você não esqueça o que estamos fazendo. É de suma importância que você compreenda o porquê da Equação 125 estar correta (na verdade, como isso não é um cálculo exato, a Equação 125 ficaria mais correta se trocássemos o sinal de “igual” pelo sinal de “aproximado”); e também qual é o raciocínio por trás dela (como tentei explicar no exemplo que atribuí uma carga a dois orbitais com probabilidade de ocupação de 0.5). É mais importante você entender de onde vem a Equação 125 e o que ela expressa do que entender toda a formulação matemática que vem a seguir. Mas como a formulação matemática também é importante, vamos a ela.

Para começar, vamos trocar a nomenclatura para nos adequar à maioria dos livros de semicondutores. Como falei na seção 1.7, apesar de o nome correto para μ ser potencial eletroquímico, quase todos os livros de semicondutores chamam esse termo de nível de Fermi, e trocam o símbolo para ε_F . Não sei por que trocaram o nome, mas o símbolo foi trocado porque, no contexto de semicondutores, μ é geralmente utilizado para o parâmetro “mobilidade”. Então, vamos nos dobrar ao peso da sociedade e chamar o potencial eletroquímico de nível de Fermi, trocando também o símbolo. Assim, a Equação 125 fica:

$$N_e = \sum_{bc} f(\varepsilon_{bc}) = \sum_{bc} \frac{1}{1 + \exp\left(\frac{\varepsilon_{bc} - \varepsilon_F}{k_B T}\right)}$$

Equação 126

Além disso, precisamos definir mais três termos, bem simples, e que estão representados na Figura 16: ε_C é o nível de energia mais baixo da banda de condução, ε_V é o nível de energia mais alto da banda de valência, e ε_G é a separação entre os dois níveis anteriores, chamado de *band-gap*. Assim:

$$\varepsilon_G = \varepsilon_C - \varepsilon_V$$

Equação 127

O *band-gap*, portanto, é a região entre a banda de valência e a banda de condução, onde não existe nenhum nível de energia disponível (ou seja, nenhum orbital possui energia dentro do *band-gap*). Alguns textos em português chamam o *band-gap* de banda proibida, mas vou manter o termo em inglês.

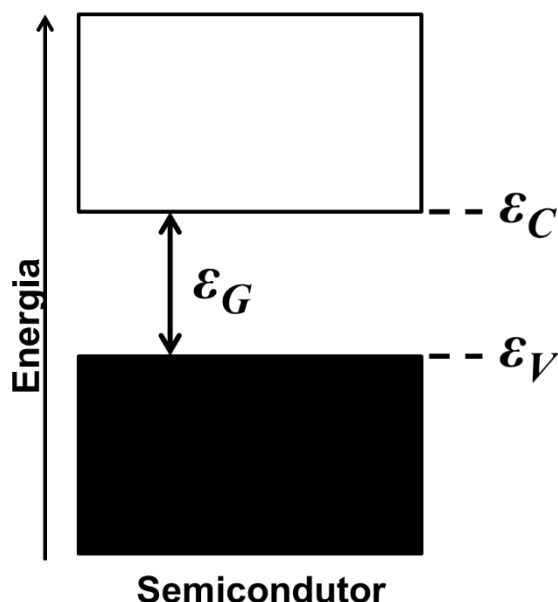


Figura 16 Definições de ϵ_V , ϵ_C e ϵ_G

Definidos os termos, podemos passar para o cálculo de N_e . À primeira vista, essa é uma tarefa monumental: teríamos que calcular todos os orbitais da banda de condução para podermos achar N_e . Felizmente, podemos fazer uma aproximação incrivelmente simples: podemos tratar os elétrons livres como se fossem um “gás” de elétrons. Podemos fazer isso exatamente porque os elétrons livres interagem pouco com a estrutura atômica. Nessa aproximação, a interação do elétron com a estrutura cristalina entra como uma modificação da massa do elétron, agora chamada de massa efetiva. O conceito de massa efetiva é fascinante: porque o elétron é uma onda que interage com uma rede cristalina, o movimento dele depende tanto da força externa como da própria rede cristalina e essa dependência é expressa pela massa efetiva. Em alguns casos mais exóticos, dependendo da fase acumulada em cada elemento do cristal, a rede poderá refletir o elétron quando se tenta acelerá-lo para dentro da rede. Isso significa que, se a força for aplicada em uma direção, o elétron seria acelerado na direção oposta, resultando assim em uma massa efetiva negativa! O tratamento matemático da massa efetiva requer mecânica quântica um pouco avançada e, por isso, usaremos a massa

efetiva como um parâmetro, que depende do material. O importante, aqui, é saber qualitativamente o que a massa efetiva expressa (a interação do elétron com a rede).

Então, nosso problema se reduz ao cálculo do número de elétrons em um gás ideal, com a massa substituída pela massa efetiva. **Assim, os orbitais da soma na Equação 126 correspondem aos estados do gás ideal.**

Já calculamos a energia associada a cada um desses orbitais, ela é dada pela Equação 58, que vou copiar aqui embaixo para facilitar:

$$E = \frac{h^2}{8L^2m} (c_x^2 + c_y^2 + c_z^2)$$

Equação 58

Lembre-se que cada trinca de números naturais c_x , c_y e c_z corresponde a um orbital diferente. Mas temos que fazer duas modificações na Equação 58 para podermos utilizá-la como sendo a energia dos elétrons na banda de condução.

Primeiro, precisamos trocar a massa do elétron pela massa efetiva. Chamando a massa efetiva dos elétrons livres de m_e^* , temos então:

$$E = \frac{h^2}{8L^2m_e^*} (c_x^2 + c_y^2 + c_z^2)$$

Segundo, precisamos ajustar a energia de referência: a Equação 58 tem como referência a energia $E = 0$, mas nossa referência agora é a menor energia da banda de condução, ou seja, ε_C . Portanto, na banda de condução, cada trinca corresponde à energia da Equação 58 mais a energia ε_C . Assim, as energias ε_{bc} da banda de condução serão:

$$\varepsilon_{bc} = \frac{h^2}{8L^2m_e^*} (c_x^2 + c_y^2 + c_z^2) + \varepsilon_C$$

Equação 128

Além dessas duas modificações na energia, aqui temos que levar em conta o fato de que estamos tratando especificamente de elétrons, que também são caracterizados pelo *spin*. Isso significa que cada orbital é caracterizado pelo conjunto de números inteiros

c_x , c_y e c_z mais o spin do elétron, que pode ser *up* ou *down*. Em outras palavras, cada trinca de números naturais corresponde a dois orbitais eletrônicos (uma com spin *up* e outra com spin *down*). Por essa razão, temos que multiplicar a soma por dois. Assim, substituindo a Equação 128 na Equação 126 e levando em consideração o *spin*, temos:

$$N_e = \sum_{bc} \frac{1}{1 + \exp\left(\frac{\varepsilon_{bc} - \varepsilon_F}{k_B T}\right)}$$

$$= 2 \sum_{c_x} \sum_{c_y} \sum_{c_z} \frac{1}{1 + \exp\left[\frac{\left(\frac{h^2}{8L^2 m_e^*} (c_x^2 + c_y^2 + c_z^2) + \varepsilon_C\right) - \varepsilon_F}{k_B T}\right]}$$

Equação 129

Podemos facilitar ainda mais o cálculo com uma observação de ordem prática: geralmente, as energias da banda de condução, mesmo quando subtraídas do nível de Fermi, possuem energia consideravelmente maior que a energia térmica dos elétrons (esta última pode ser interpretada como sendo igual à $k_B T$), o que significa que a exponencial é bem maior que 1, ou seja:

$$\exp\left[\frac{\left(\frac{h^2}{8L^2 m_e^*} (c_x^2 + c_y^2 + c_z^2) + \varepsilon_C\right) - \varepsilon_F}{k_B T}\right] \gg 1$$

Para se ter uma ideia dos valores envolvidos, vamos dar um exemplo. As energias em semicondutores são, geralmente, expressas na unidade elétron-volts ($e.V.$) - por exemplo, o *band-gap* do silício é de $1.14 e.V.$ Por outro lado, a energia térmica na temperatura ambiente ($T = 300 K$) é $k_B T = 0.0258 e.V.$ Isso significa que uma separação de $0.2 e.V.$ entre a energia na banda de condução e o nível de Fermi já resulta na exponencial com valor igual à 2326.

Assim, podemos aproximar o denominador na Equação 129 para:

$$1 + \exp \left[\frac{\left(\frac{h^2}{8L^2 m_e^*} (c_x^2 + c_y^2 + c_z^2) + \varepsilon_C \right) - \varepsilon_F}{k_B T} \right]$$

$$\approx \exp \left[\frac{\left(\frac{h^2}{8L^2 m_e^*} (c_x^2 + c_y^2 + c_z^2) + \varepsilon_C \right) - \varepsilon_F}{k_B T} \right]$$

Equação 130

Semicondutores que satisfazem essa condição são chamados de não-degenerados. Utilizando, então, a aproximação para semicondutores não-degenerados, temos:

$$N_e = 2 \sum_{c_x} \sum_{c_y} \sum_{c_z} \frac{1}{\exp \left[\frac{\left(\frac{h^2}{8L^2 m_e^*} (c_x^2 + c_y^2 + c_z^2) + \varepsilon_C \right) - \varepsilon_F}{k_B T} \right]}$$

$$\therefore$$

$$N_e = 2 \sum_{c_x} \sum_{c_y} \sum_{c_z} \exp \left[- \frac{\left(\frac{h^2}{8L^2 m_e^*} (c_x^2 + c_y^2 + c_z^2) + \varepsilon_C \right) - \varepsilon_F}{k_B T} \right]$$

Equação 131

Podemos isolar os termos da soma que não dependem dos índices c_x , c_y e c_z :

$$N_e = 2 \exp \left(- \frac{\varepsilon_C - \varepsilon_F}{k_B T} \right) \sum_{c_x} \sum_{c_y} \sum_{c_z} \exp \left[- \frac{\frac{h^2}{8L^2 m_e^*} (c_x^2 + c_y^2 + c_z^2)}{k_B T} \right]$$

Equação 132

Agora, só precisamos resolver a soma. Na verdade, já lidamos com esse problema antes: a soma na Equação 132 é idêntica à soma da Equação 59. Como foi afirmado naquela ocasião, a soma é resolvida aproximando-a de uma integral. Isso é válido

quando a variação do termo $\frac{h^2}{8L^2m_e^*}(c_x^2 + c_y^2 + c_z^2)$ é pequena em relação à variação nos coeficientes c_x , c_y e c_z . Neste caso, podemos aproximar a soma pela seguinte integral:

$$\sum_{c_x} \sum_{c_y} \sum_{c_z} \exp \left[-\frac{h^2}{8L^2m_e^*}(c_x^2 + c_y^2 + c_z^2) \right] \\ \approx \int_0^\infty \int_0^\infty \int_0^\infty \exp \left[-\frac{h^2}{8L^2m_e^*}(c_x^2 + c_y^2 + c_z^2) \right] dc_x dc_y dc_z$$

Equação 133

Essa integral é tabelada e o resultado, é claro, é o mesmo da Equação 60:

$$\int_0^\infty \int_0^\infty \int_0^\infty \exp \left[-\frac{h^2}{8L^2m_e^*}(c_x^2 + c_y^2 + c_z^2) \right] dc_x dc_y dc_z = \frac{L^3}{\left(\frac{h^2}{2\pi m_e^* k_B T}\right)^{3/2}}$$

Equação 134

Assim, temos que o número total de elétrons livres será:

$$N_e = 2 \exp \left(-\frac{\varepsilon_C - \varepsilon_F}{k_B T} \right) \frac{L^3}{\left(\frac{h^2}{2\pi m_e^* k_B T}\right)^{3/2}}$$

Equação 135

Como L^3 é o volume do semiconductor, podemos reexpressar este resultado em termos da concentração de elétrons livres n :

$$n = \frac{N_e}{L^3} = 2 \exp \left(-\frac{\varepsilon_C - \varepsilon_F}{k_B T} \right) \frac{1}{\left(\frac{h^2}{2\pi m_e^* k_B T}\right)^{3/2}} = n_c \exp \left(-\frac{\varepsilon_C - \varepsilon_F}{k_B T} \right)$$

$$\therefore n = n_c \exp \left(-\frac{\varepsilon_C - \varepsilon_F}{k_B T} \right)$$

Equação 136

Onde:

$$n_c = \frac{2}{\left(\frac{h^2}{2\pi m_e^* k_B T}\right)^{3/2}}$$

Equação 137

A Equação 136 nos conta qual é a concentração de elétrons livres. Ela depende do parâmetro n_c , que por sua vez depende da temperatura, de constantes, e da estrutura cristalina (esta última entra na massa efetiva). Além disso, a concentração depende da diferença entre ε_C e ε_F . Não vamos calcular ε_C , mas precisamos encontrar ε_F para completar a história. Isso será feito ainda neste capítulo, mas antes vamos recalculer a Equação 136 utilizando o importante conceito de densidade de estados.

2.4 Densidade de estados

Para encontrar a Equação 136, utilizamos, por razões pedagógicas, um método mais direto, mas que é pouco útil em problemas mais complexos. Por isso, vou agora voltar atrás e recalculer a Equação 136 utilizando outro método que envolve um conceito importantíssimo: o conceito de densidade de estados. É importante que o estudante consiga visualizar a equivalência entre os dois métodos (na verdade, uma das razões pelas quais vamos recalculer a Equação 136 é exatamente para que a comparação entre os métodos ajude a ilustrar o conceito de densidade de estados).

O conceito de densidade de estados não é complicado. Primeiro, lembre o que é densidade (em geral) e como ela pode ser medida (ou calculada). Por exemplo, se você tiver uma corda e quiser saber qual é a densidade de massa da corda, você vai pegar um pedaço pequeno da corda (pequeno o suficiente para que o pedaço possa ser tratado como uniforme), vai medir a massa desse pedaço de corda, e dividir pelo comprimento da corda. O resultado será a densidade linear da corda, expressa em kg/m .

A ideia da densidade de estados é análoga à densidade linear da corda, mas em vez de massa, temos número de estados; e, em vez de comprimento, temos energia. Portanto, se eu quiser saber qual é a densidade de estados em uma determinada faixa de energia, basta calcular o número de estados que existem dentro dessa faixa de energia

e dividir pela largura da faixa. Assim, obtenho a densidade de estados naquela faixa, expressa em unidades de $1/\text{joules}$.

Vamos pôr essa ideia em linguagem matemática. Definindo $N_{\varepsilon \rightarrow \varepsilon + \Delta\varepsilon}$ como sendo o número de orbitais cuja energia situa-se na faixa entre ε e $\varepsilon + \Delta\varepsilon$, então, por definição, densidade de estados será dada por:

$$D(\varepsilon) = \frac{N_{\varepsilon \rightarrow \varepsilon + \Delta\varepsilon}}{\Delta\varepsilon}$$

Equação 138

Onde $D(\varepsilon)$ é a densidade de estados (se fossemos seguir à risca a nomenclatura que adotamos aqui, que reserva o termo “estado” para o estado do sistema completo, e “orbital” para o estado de um único elétron, então a nomenclatura mais natural seria “densidade de orbitais”, mas vou utilizar o termo densidade de estados, por ser o termo utilizado na literatura). A noção de densidade de estados é em si uma aproximação, porque trata as energias como sendo contínuas. Mas é uma excelente aproximação na maioria dos casos e facilita muito os cálculos. De fato, podemos facilmente expressar a soma da Equação 126 em termos da densidade de estados:

$$N_e = \sum_{bc} f(\varepsilon_{bc}) = \sum_{\varepsilon} f(\varepsilon) N_{\varepsilon \rightarrow \varepsilon + \Delta\varepsilon} = \sum_{\varepsilon} f(\varepsilon) D(\varepsilon) \Delta\varepsilon$$

Equação 139

A interpretação da Equação 139 é direta. Para entender a equivalência entre as somas, lembre-se que a primeira soma ($\sum_{bc} f(\varepsilon_{bc})$) é uma soma sobre todos os orbitais possíveis. Você pega orbital por orbital, calcula sua energia, joga na distribuição de Fermi-Dirac e vai somando. Esse é o processo que fizemos na Equação 129: note que a soma na Equação 129 é feita sobre os orbitais possíveis (a soma é feita em c_x , c_y e c_z)

A segunda e a terceira somas na Equação 139, por outro lado, são efetuadas sobre as energias na banda de condução. Nessas somas, os orbitais são agrupados em pequenas faixas de energia $\Delta\varepsilon$. Então, em vez de calcular a distribuição de Fermi-Dirac orbital por orbital, nós a calculamos uma única vez para a faixa de energia entre ε e $\varepsilon + \Delta\varepsilon$ e a seguir multiplicamos pelo número de orbitais que existem nessa faixa de energia, que é o termo $N_{\varepsilon \rightarrow \varepsilon + \Delta\varepsilon}$. (claro que $\Delta\varepsilon$ tem que ser pequeno o suficiente para

que possamos desprezar a variação de energia dentro dessa faixa). Assim, a contribuição de todos os orbitais dentro da faixa situada entre ε e $\varepsilon + \Delta\varepsilon$ é obtida de uma vez só, e tudo o que resta é somar a contribuição de todas as faixas que juntas formam a banda de condução.

Como afirmado no parágrafo anterior, precisamos que $\Delta\varepsilon$ seja pequeno o suficiente para que $f(\varepsilon)$ represente a probabilidade de ocupação de todos os orbitais dentro da faixa de energia entre ε e $\varepsilon + \Delta\varepsilon$. Queremos, então, que $\Delta\varepsilon$ seja infinitesimal, o que transforma a última soma na Equação 139 em uma integral:

$$N_e = \lim_{\Delta\varepsilon \rightarrow 0} \sum_{\varepsilon} f(\varepsilon)D(\varepsilon)\Delta\varepsilon = \int_{\varepsilon_C}^{\infty} f(\varepsilon)D(\varepsilon)d\varepsilon$$

Equação 140

Note que a integral é realizada na banda de condução, como esperado. Assim, fica claro que, para calcular N_e , tudo o que precisamos saber é a densidade de estados na banda de condução.

Vamos, então, fazer um parêntesis para calcular a densidade de estados para o gás de elétrons ideal, ou seja, para o caso de elétrons em um poço. Primeiro, vamos fazer o cálculo para um caso onde os elétrons estão em um poço bidimensional e, depois, vamos generalizar o caso para três dimensões. Em seguida, vamos adaptar o resultado para o caso de elétrons na banda de condução de um sólido.

Começemos, então, tratando do caso “gás de elétrons bidimensional”. Neste caso, os orbitais serão definidos por pares de números naturais, em vez de trincas. Além disso, a energia será dada pela versão bidimensional da Equação 58, que é:

$$\varepsilon = \frac{h^2}{8L^2m} (c_x^2 + c_y^2)$$

Equação 141

Para visualizar como o cálculo é feito, imagine um sistema de coordenadas euclidiano com eixos c_x e c_y . Agora, imagine que marcamos um ponto em todas as coordenadas formadas por números inteiros (incluindo, por enquanto, os números negativos), como mostrado na Figura 17a. Por exemplo, marcamos as coordenadas (0, 1), (1, 0), (2,1), e assim por diante. Marcamos com um ponto todas as coordenadas

formadas por números inteiros. Note que cada uma dessas coordenadas corresponde a um par c_x, c_y . Portanto, cada um desses pontos representa um orbital (por enquanto, vamos nos permitir envolver os números inteiros negativos também, depois consertamos o erro causado por essa suposição).

Então, temos que cada orbital está representado por um pontinho no espaço euclidiano, como mostrado na Figura 17a. Agora, note que a distância de cada ponto até a origem define a energia do orbital correspondente. Para enxergar isso, basta lembrar que a distância do ponto até a origem, que chamaremos de R , é dada por:

$$R = \sqrt{c_x^2 + c_y^2}$$

Equação 142

Portanto, R é proporcional à raiz quadrada da energia. De fato, substituindo a Equação 141 na Equação 142, obtemos:

$$R = \sqrt{\frac{8L^2m}{h^2}} \sqrt{\varepsilon}$$

Equação 143

Lembre-se que, para calcular a densidade de estados, tudo o que precisamos fazer é calcular o número de orbitais em uma determinada faixa de energia e dividir pelo comprimento da faixa, como mostrado na Equação 138. Agora, imagine um anel que começa em R e termina em $R + \Delta R$, como mostrado na Figura 17b. Nessa figura, estamos mostrando o espaço euclidiano de longe, onde nem dá para ver mais os pontinhos independentemente. Como a energia do orbital só depende da distância do ponto até a origem, então todos os pontos dentro do anel vermelho da Figura 17b estarão em uma faixa de energia que começa em ε e termina em $\varepsilon + \Delta\varepsilon$, onde a energia e o raio estão relacionados pela Equação 143. Assim, para calcular a densidade de estados, precisamos contar quantos pontos estão dentro do anel vermelho da Figura 17b, que começa em R e termina em $R + \Delta R$. Seguindo a notação da Equação 138, vamos chamar esse número de pontos de $N_{R \rightarrow R+\Delta R}$ e, depois, reexpressaremos o resultado em termos da energia, utilizando a Equação 143.

Para calcular $N_{R \rightarrow R+\Delta R}$, vamos primeiro calcular quantos orbitais possuem distância até a origem menor que R . Em outras palavras, vamos calcular primeiro quantos orbitais estão dentro do disco vermelho na Figura 17c. Vamos chamar o número de orbitais com distância até a origem menor que R de $\eta(R)$. Calcular $\eta(R)$ é fácil: note que, como a distância entre cada pontinho é 1, então poderíamos desenhar um quadrado de lado 1, centrado em cada um dos pontinhos. Um exemplo desse quadrado imaginário é mostrado em torno do ponto na coordenada (2,2) na Figura 17a. O número de quadrados de lado 1 corresponderia, obviamente, ao número de pontinhos, ou seja, ao número de orbitais. Portanto, tudo o que precisamos fazer é calcular quantos quadrados de área 1 cabem dentro do disco de raio R da Figura 17c. Isso é fácil: é a área do disco dividido pela área de cada quadrado. Mas lembre-se que só podemos pegar os orbitais que correspondem ao primeiro quadrante, que é o quadrante dos números naturais. Por isso, temos que dividir o resultado por 4. Além disso, temos que multiplicar tudo por 2 para levar em consideração o *spin*. Assim:

$$\eta(R) = \frac{\pi R^2}{1} \cdot \frac{2}{4}$$

Equação 144

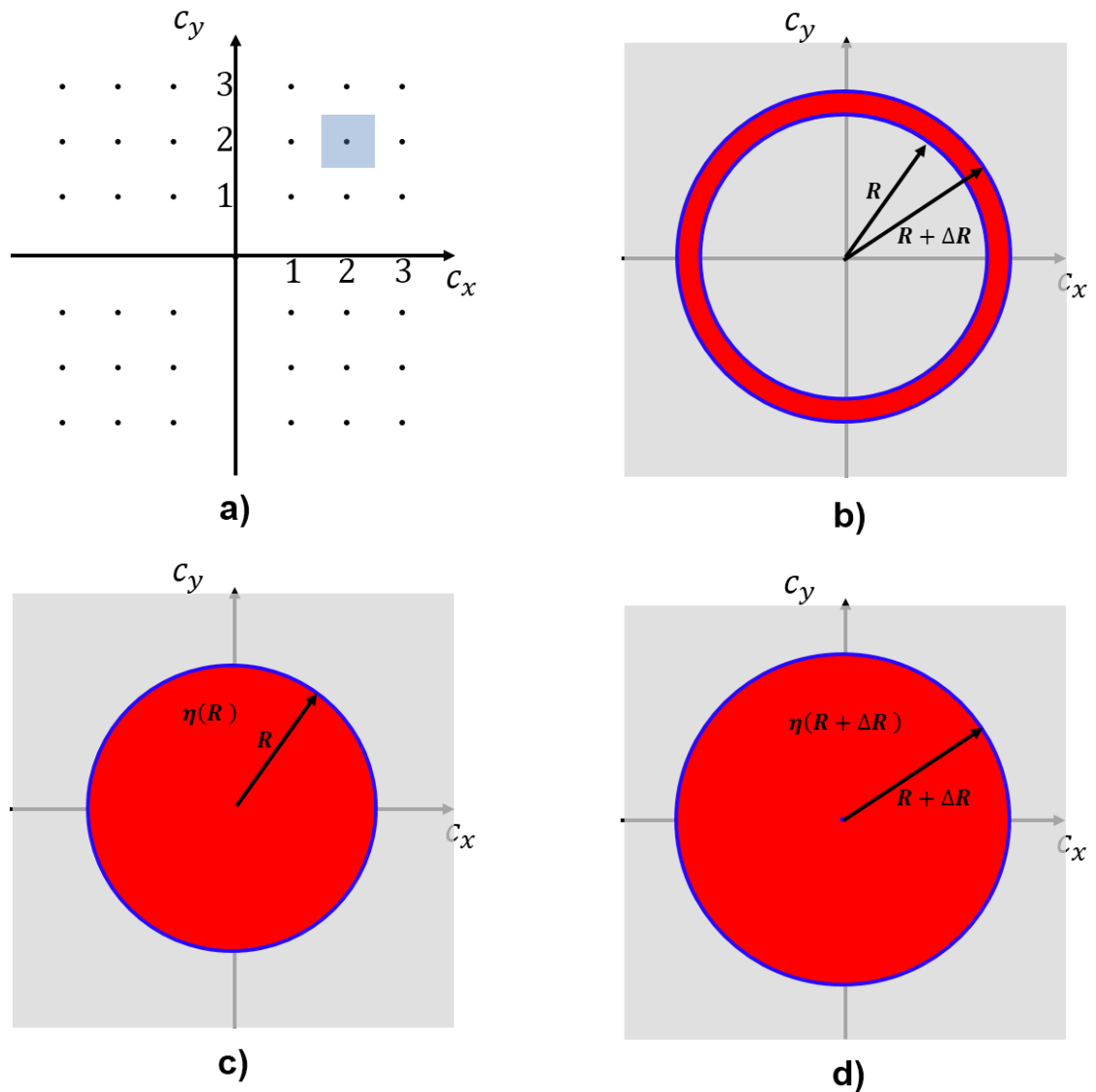


Figura 17 Representação de espaço euclidiano para o cálculo da densidade de estados de um gás de elétrons bidimensional. a) Cada coordenada formada por combinações de números inteiros representa um estado e a distância até a origem define a energia do estado. b) Queremos o número de estados dentro do anel. c) Para isso, calculamos o número de estados dentro do disco de raio R e d) o número de estados dentro do disco de raio $R + \Delta R$. Assim, o número de estados dentro do anel será o número de estados dentro do disco de raio $R + \Delta R$ menos o número de estados dentro do disco de raio R . As figuras b) até d) devem ser interpretadas como sendo uma visão de longe da figura a).

Agora, ficou fácil calcular $N_{R \rightarrow R+\Delta R}$. Para isso, basta pegar o número de pontos que cabem dentro do disco de raio $R + \Delta R$ (como mostrado na Figura 17d) e subtrair do número de pontos que cabem dentro do disco de raio R , ficando assim com o número de pontos que cabem dentro do anel que começa em R e termina em $R + \Delta R$, ou seja:

$$N_{R \rightarrow R+\Delta R} = \eta(R + \Delta R) - \eta(R)$$

Equação 145

Mas não queremos $N_{R \rightarrow R+\Delta R}$. O que realmente queremos é $N_{\varepsilon \rightarrow \varepsilon+\Delta\varepsilon}$. Mas adaptar de uma para a outra é muito fácil: basta transformarmos $\eta(R)$ em $\eta(\varepsilon)$ através da Equação 143. Note que $\eta(\varepsilon)$ é o número de estados com energia menor que a energia ε . Uma vez que encontrarmos $\eta(\varepsilon)$, ficará fácil encontrar $N_{\varepsilon \rightarrow \varepsilon+\Delta\varepsilon}$, já que, analogamente à Equação 145, teremos:

$$N_{\varepsilon \rightarrow \varepsilon+\Delta\varepsilon} = \eta(\varepsilon + \Delta\varepsilon) - \eta(\varepsilon)$$

Equação 146

A interpretação da Equação 143 é direta e praticamente idêntica à interpretação da Equação 145: como $\eta(\varepsilon + \Delta\varepsilon)$ é o número de orbitais com energia menor que $\varepsilon + \Delta\varepsilon$; e $\eta(\varepsilon)$ é o número de estados com energia menor que ε , então o número de estados com energia entre ε e $\varepsilon + \Delta\varepsilon$ é dado por $\eta(\varepsilon + \Delta\varepsilon) - \eta(\varepsilon)$.

Para encontrar $\eta(\varepsilon)$, basta substituir a Equação 143 na Equação 144:

$$\eta(\varepsilon) = \frac{2}{4} \cdot \frac{\pi R^2}{1} = \frac{2}{4} \cdot \pi \frac{8L^2 m}{h^2} \varepsilon$$

Equação 147

Agora, ficou fácil encontrar a densidade de estados. Da definição de densidade de estados (Equação 138), temos que:

$$D(\varepsilon) = \frac{N_{\varepsilon \rightarrow \varepsilon+\Delta\varepsilon}}{\Delta\varepsilon}$$

Portanto:

$$D(\varepsilon) = \frac{N_{\varepsilon \rightarrow \varepsilon+\Delta\varepsilon}}{\Delta\varepsilon} = \frac{\eta(\varepsilon + \Delta\varepsilon) - \eta(\varepsilon)}{\Delta\varepsilon}$$

Equação 148

Como queremos $\Delta\varepsilon$ muito pequeno, teremos:

$$D(\varepsilon) = \lim_{\Delta\varepsilon \rightarrow 0} \frac{\eta(\varepsilon + \Delta\varepsilon) - \eta(\varepsilon)}{\Delta\varepsilon} = \frac{d\eta}{d\varepsilon}$$

Equação 149

Substituindo a Equação 147 na Equação 149, temos, finalmente, que:

$$D(\varepsilon) = \frac{2}{4} \cdot \pi \frac{8L^2 m}{h^2} = \frac{4\pi L^2 m}{h^2}$$

Equação 150

A Equação 150 é a densidade de estados de um gás de elétrons bidimensional.

Adaptar o raciocínio para o caso de um gás de elétrons tridimensional, que é o caso de maior interesse, é fácil: a área vira um volume, os discos viram esferas, o anel vira uma casca da esfera e os quadradinhos de lado 1 viram cubinhos de lado 1. Além disso, enquanto para o caso bidimensional tínhamos que $1/4$ do espaço euclidiano era formado por coordenadas de números inteiros, para o caso tridimensional temos que $1/8$ do espaço será formado por números naturais (existem $2^3 = 8$ combinações de sinais para as três coordenadas; assim podemos dividir a esfera em oito pedaços iguais, sendo que apenas um desses pedaços terá c_x , c_y e c_z positivos). Além disso, ainda temos que multiplicar o resultado final por dois para levar em conta o *spin*. Com essas considerações, temos que o número de pontos dentro da esfera de volume R passa a ser:

$$\eta_{3D}(R) = \frac{4\pi R^3}{3} \cdot \frac{2}{8}$$

Equação 151

Onde $4\pi R^3/3$ é o volume da esfera. Note que coloquei o índice 3D para diferenciar do caso anterior. No caso tridimensional, temos:

$$R = \sqrt{c_x^2 + c_y^2 + c_z^2}$$

E

$$\varepsilon = \frac{h^2}{8L^2 m} (c_x^2 + c_y^2 + c_z^2)$$

Portanto:

$$\eta_{3D}(\varepsilon) = \frac{2}{8} \cdot \frac{4\pi R^3}{3} = \frac{2}{8} \cdot \frac{4\pi \left(\frac{8L^2 m}{h^2} \varepsilon\right)^{3/2}}{3}$$

Equação 152

E, portanto:

$$D_{3D}(\varepsilon) = \frac{d\eta_{3D}}{d\varepsilon} = \frac{\pi \left(\frac{8L^2 m}{h^2}\right)^{3/2}}{2} \sqrt{\varepsilon}$$

Rearranjando:

$$D_{3D}(\varepsilon) = 4\pi L^3 \left(\frac{2m}{h^2}\right)^{3/2} \sqrt{\varepsilon}$$

Equação 153

Note que a densidade de estados em três dimensões é função da energia, enquanto a densidade de estados em duas dimensões é constante.

Agora precisamos adaptar esse resultado para o nosso sistema de interesse, que é o gás de elétrons na banda de condução. A modificação mais óbvia que temos que fazer é substituir a massa pela massa efetiva, como discutido anteriormente. Além disso, o nosso novo zero de referência é a energia ε_C , mas a Equação 153 foi deduzida para uma referência de energia igual a zero (que corresponde à origem do sistema de coordenadas e ao centro da esfera). Portanto, para utilizarmos a mesma equação, temos que pegar a diferença entre a energia do gás ideal e a energia na banda de condução (assim, a energia que entra na Equação 153 continua tendo o zero como referência). Assim, chamando a densidade de estados na banda de condução de D_e , temos que:

$$D_e(\varepsilon) = 4\pi L^3 \left(\frac{2m_e^*}{h^2}\right)^{3/2} \sqrt{\varepsilon - \varepsilon_C}$$

Equação 154

Se você estiver desconfortável em relação ao fato que eu subtraí a energia por ε_C , lembre-se que a densidade de estados depende do valor de energia em relação ao zero (que agora é a energia da banda de condução). Portanto, agora é a energia $\varepsilon - \varepsilon_C$ que corresponde ao raio da esfera (o ponto $\varepsilon = \varepsilon_C$ corresponde agora ao centro da esfera, portanto $\varepsilon = \varepsilon_C$ possui raio zero e, conseqüentemente, a densidade de estados para a energia ε_C é zero).

Agora já temos a faca e o queijo na mão: basta substituímos $D(\varepsilon)$ na integral da Equação 140. Utilizando novamente a aproximação de semicondutores não-degenerados (Equação 130), temos que:

$$\begin{aligned} N_e &= \int_{\varepsilon_C}^{\infty} f(\varepsilon)D_e(\varepsilon)d\varepsilon = \int_{\varepsilon_C}^{\infty} \exp\left[-\frac{(\varepsilon - \varepsilon_F)}{k_B T}\right] D_e(\varepsilon)d\varepsilon \\ &= \int_{\varepsilon_C}^{\infty} \exp\left[-\frac{(\varepsilon - \varepsilon_F)}{k_B T}\right] 4\pi L^3 \left(\frac{2m_e^*}{h^2}\right)^{3/2} \sqrt{\varepsilon - \varepsilon_C} d\varepsilon \end{aligned}$$

Equação 155

Isolando os termos que não dependem da energia:

$$N_e = 4\pi L^3 \left(\frac{2m_e^*}{h^2}\right)^{3/2} \exp\left(\frac{\varepsilon_F}{k_B T}\right) \int_{\varepsilon_C}^{\infty} \exp\left(-\frac{\varepsilon}{k_B T}\right) \sqrt{\varepsilon - \varepsilon_C} d\varepsilon$$

Equação 156

E o resultado é:

$$N_e = L^3 \frac{2}{\left(\frac{h^2}{2\pi m_e^* k_B T}\right)^{3/2}} \exp\left(-\frac{\varepsilon_C - \varepsilon_F}{k_B T}\right)$$

Equação 157

Como L^3 é o volume, encontramos novamente que a concentração de elétrons na banda de condução (elétrons livres) é dada por:

$$n = \frac{N_e}{L^3} = n_c \exp\left(-\frac{\varepsilon_C - \varepsilon_F}{k_B T}\right)$$

Equação 158

Onde n_c é dado pela Equação 137. A Equação 158 é idêntica à Equação 136, como esperado.

2.5 Concentração de buracos e nível de Fermi

Agora, vamos voltar nossa atenção ao problema do cálculo do nível de Fermi. Para isso, vamos utilizar o seguinte fato: já sabemos que a concentração de elétrons na banda de condução (elétrons livres), n , tem que ser igual à concentração de buracos na banda de valência (que chamaremos de p), já que cada elétron excitado para a banda de condução deixa um buraco na banda de valência. Mas podemos calcular p diretamente se soubermos a densidade de estados na banda de valência. Do mesmo jeito que os elétrons na banda de condução se comportam como um gás, os buracos na banda de valência também se comportam como um gás, portanto a densidade de estados é a mesma da Equação 154, apenas trocando a massa efetiva do elétron pela massa efetiva do buraco e os níveis de referência. Assim, chamando a densidade de estados para buracos de $D_b(\varepsilon)$, temos:

$$D_b(\varepsilon) = 4\pi L^3 \left(\frac{2m_b^*}{h^2}\right)^{3/2} \sqrt{\varepsilon_V - \varepsilon}$$

Equação 159

Onde m_b^* é a massa efetiva do buraco. E a distribuição de Fermi-Dirac para buracos? Como a distribuição de Fermi-Dirac nos diz qual é a probabilidade de um estado estar ocupado, e o buraco é um estado desocupado, então a probabilidade de um estado estar ocupado por um buraco é igual à probabilidade de o estado estar desocupado pelo elétron, ou seja:

$$f_b(\varepsilon) = 1 - f(\varepsilon)$$

Equação 160

Onde $f_b(\varepsilon)$ é a probabilidade de o estado estar ocupado por um buraco. Assim:

$$f_b(\varepsilon) = 1 - \frac{1}{1 + \exp\left(\frac{\varepsilon - \varepsilon_F}{k_B T}\right)} = \frac{1}{1 + \exp\left(\frac{\varepsilon_F - \varepsilon}{k_B T}\right)}$$

Equação 161

Podemos utilizar novamente a aproximação para semicondutores não degenerados:

$$f_b(\varepsilon) \approx \frac{1}{\exp\left(\frac{\varepsilon_F - \varepsilon}{k_B T}\right)}$$

Equação 162

Assim, o número total de buracos na banda de valência será:

$$N_b = \int_{-\infty}^{\varepsilon_V} D_b(\varepsilon) f_b(\varepsilon) d\varepsilon = \int_{-\infty}^{\varepsilon_V} 4\pi L^3 \left(\frac{2m_b^*}{h^2}\right)^{3/2} (\varepsilon_V - \varepsilon)^{1/2} \exp\left(-\frac{\varepsilon_F - \varepsilon}{k_B T}\right) d\varepsilon$$

Equação 163

E o resultado para a concentração de buracos, que chamaremos de p , é:

$$p = \frac{N_b}{L^3} = n_d \exp\left(-\frac{\varepsilon_F - \varepsilon_V}{k_B T}\right)$$

Equação 164

Onde n_d é dado por:

$$n_d = \frac{2}{\left(\frac{h^2}{2\pi m_b^* k_B T}\right)^{3/2}}$$

Equação 165

Note que a única diferença entre n_c e n_d é que o primeiro envolve a massa efetiva do elétron e o segundo envolve a massa efetiva do buraco.

Antes de calcularmos o nível de Fermi, note também o seguinte fato: se multiplicarmos a Equação 158 pela Equação 164, teremos:

$$n \cdot p = n_c n_d \exp\left(-\frac{\varepsilon_C - \varepsilon_F}{k_B T}\right) \exp\left(-\frac{\varepsilon_F - \varepsilon_V}{k_B T}\right) = n_c n_d \exp\left(-\frac{\varepsilon_C - \varepsilon_V}{k_B T}\right)$$

Equação 166

Mas a diferença entre a energia da banda de condução e de valência é igual ao *band-gap*, portanto:

$$n \cdot p = n_c n_d \exp\left(-\frac{\varepsilon_G}{k_B T}\right)$$

Equação 167

Esse é um resultado notável: de acordo com a Equação 167, o produto entre a concentração de elétrons livres n e a concentração de buracos p não depende do nível de Fermi. Por enquanto, estamos estudando o caso do semiconductor intrínseco (não dopado), onde sabemos que n e p são iguais. Mas, como veremos mais adiante, se doparmos o semiconductor isso não será mais verdade. Mas o efeito da dopagem é uma mudança no nível de Fermi e, como o produto entre n e p não depende do nível de Fermi, o produto não é alterado com a dopagem. Isso significa que se a dopagem aumenta o número de elétrons, necessariamente ela diminui o número de buracos (e vice-versa), de maneira que o produto entre os dois é mantido constante. Voltaremos a essa questão mais adiante.

Podemos, agora, calcular o nível de Fermi do semiconductor intrínseco com facilidade. Para tanto, basta utilizarmos o fato de que, em um semiconductor intrínseco, $n = p$. Portanto:

$$n_c \exp\left(-\frac{\varepsilon_C - \varepsilon_F}{k_B T}\right) = n_d \exp\left(-\frac{\varepsilon_F - \varepsilon_V}{k_B T}\right)$$

Equação 168

Daí temos que:

$$\exp\left(\frac{-2\varepsilon_F + \varepsilon_C + \varepsilon_V}{k_B T}\right) = \frac{n_c}{n_d} = \left(\frac{m_e^*}{m_b^*}\right)^{3/2}$$

Equação 169

Portanto:

$$-2\varepsilon_F + \varepsilon_C + \varepsilon_V = \frac{3}{2}k_B T \ln\left(\frac{m_b^*}{m_e^*}\right)$$

Equação 170

Rearranjando:

$$\varepsilon_F = \frac{\varepsilon_C + \varepsilon_V}{2} + \frac{3}{4}k_B T \ln\left(\frac{m_b^*}{m_e^*}\right)$$

Equação 171

Como a interação com a rede cristalina é tipicamente fraca, a massa efetiva do elétron e do buraco são muito próximas, o que significa que o segundo termo do lado direito da igualdade é praticamente zero. Portanto:

$$\varepsilon_F \approx \frac{\varepsilon_C + \varepsilon_V}{2}$$

Equação 172

A Equação 172 é um importante resultado. Essa relação nos diz que, em um semicondutor não dopado, o nível de Fermi está praticamente no meio do *band-gap*.

2.6 Semicondutores extrínsecos (dopagem)

Um semicondutor intrínseco *per se* é um material virtualmente inútil. Em praticamente todas as tecnologias relacionadas à eletrônica, os semicondutores são dopados. Dopagem se refere à inserção de átomos de impureza (isto é, átomos que não fazem parte do cristal) na estrutura cristalina do semicondutor. Tipicamente, dependendo do tipo de átomo utilizado, as ligações químicas entre os átomos de impureza e a estrutura cristalina resultam na liberação de um elétron para a banda de condução (dopagem tipo **n**), ou na liberação de um buraco na banda de valência (dopagem do tipo **p**). No caso do silício, um exemplo de dopante tipo **n** é o fósforo; e um exemplo de dopante tipo **p** é o boro.

Um semicondutor dopado é chamado de extrínseco. Utilizando uma nomenclatura mais direta, podemos chamar semicondutores dopados de semicondutores tipo **n** ou semicondutores tipo **p**, dependendo do tipo de dopagem. Existem inúmeros exemplos de dopagem que podem ser encontrados facilmente na Internet, inclusive detalhando as

reações químicas envolvidas. No nosso caso, estamos particularmente interessados no que ocorre com os níveis de energia de um semicondutor quando ele é dopado.

A Figura 18 ilustra as bandas de energia de um semicondutor intrínseco, um semicondutor tipo **n** e um semicondutor tipo **p**. Na dopagem do tipo **n**, os átomos de impureza criam estados de energia (orbitais) dentro do *band-gap* e próximo da banda de condução. Esse novo estado de energia (orbital) é comumente chamado de “estado de impureza”. Os elétrons que ocupam o estado de impureza podem ser facilmente excitados (termicamente) para a banda de condução, devido à proximidade entre esses níveis de energia. Portanto, a dopagem tipo **n** “popula” a banda de condução, isto é, aumenta a concentração de elétrons livres na banda de condução. Como a probabilidade de ocupação ainda é dada pela distribuição de Fermi-Dirac (lembre-se que a distribuição de Fermi-Dirac é um resultado geral, válido para qualquer material em equilíbrio térmico, seja ele dopado ou não), o aumento da ocupação na banda de condução corresponde ao aumento no nível de Fermi (ou seja, o nível de Fermi se aproxima da banda de condução). O aumento no nível de Fermi expressa a maior probabilidade de ocupação na banda de condução, já que, de acordo com a distribuição de Fermi-Dirac, quando o nível de Fermi se aproxima da banda de condução, a probabilidade de ocupação aumenta (confirme esta afirmação inspecionando a fórmula da distribuição de Fermi-Dirac).

O aumento da concentração de elétrons na banda de condução e a elevação do nível de Fermi não devem ser entendidos como causa ou consequência. O entendimento mais apropriado é que estas são duas manifestações equivalentes do mesmo efeito: uma elevação do nível de Fermi corresponde ao aumento da concentração de elétrons na banda de condução e vice-versa. A causa tanto do aumento de um quanto da elevação do outro é a dopagem tipo **n**.

De maneira análoga, a dopagem tipo **p** cria um estado de energia (estado de impureza) próximo da banda de valência, o que permite que elétrons da banda de valência sejam termicamente excitados para o estado de impureza. Isso faz com que a banda de valência seja “populada” por buracos e com que o nível de Fermi abaixe (ou seja, se aproxime da banda de valência).

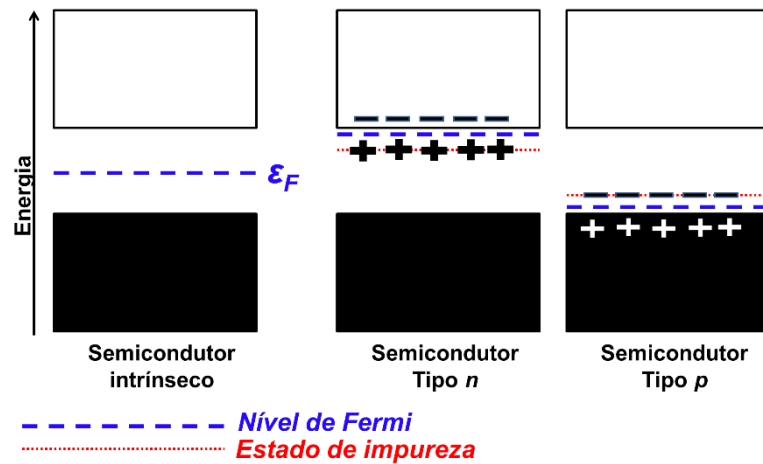


Figura 18 Diagrama de energia do semicondutor intrínseco, semicondutor tipo n e semicondutor tipo p .

É muito importante ressaltar que o processo de dopagem não altera a carga total do semicondutor. Tanto o semicondutor tipo n quanto o semicondutor tipo p continuam sendo materiais neutros, ou seja, com o mesmo número de cargas negativas e positivas. O que ocorre é simplesmente um aumento de portadores de cargas livres: elétrons para dopagem tipo n e buracos para dopagem tipo p .

Note que, em um semicondutor dopado, não é mais verdade que o número de elétrons livres é igual ao número de buracos. De fato, um semicondutor tipo n possui muito mais elétrons livres do que buracos, ou seja, $n \gg p$. Por outro lado, em um semicondutor tipo p temos $p \gg n$. Mas, como afirmado na seção anterior, o produto entre p e n (Equação 166) depende apenas do *band-gap* e, portanto, é inalterado pela dopagem.

Precisamos, agora, calcular as concentrações de portadores livres (elétrons livres e buracos) em semicondutores extrínsecos. Mas essa, agora, é uma tarefa bem simples. Primeiramente, note que a Equação 158 e a Equação 164 ainda são válidas. De fato, elas foram deduzidas levando em consideração apenas os estados de energia nas bandas de condução e valência, respectivamente (ou, equivalentemente, as densidades de estados). Mas esses níveis não são afetados pela dopagem: a dopagem cria um nível no *band-gap*, mas não altera significativamente os níveis das bandas de condução e valência. A única coisa que muda nessas expressões são os níveis de Fermi, que agora **não são mais dados pela** Equação 171 (Por quê? O que foi utilizado na dedução da Equação 171 que não é mais verdade agora?).

Então, nosso problema se reduz a encontrar o nível de Fermi. É claro que o nível de Fermi depende do grau de dopagem. Tipicamente, a concentração de dopantes é bem maior que a concentração de portadores livres do semiconductor intrínseco; portanto, podemos supor que todos os portadores livres em um semiconductor extrínseco vêm da dopagem. **Mas se cada átomo de impureza contribui para um portador livre, então a concentração de portadores livres será praticamente igual à concentração de átomos de impureza** (a concentração de portadores livres só não é exatamente igual à concentração de átomos de impureza porque alguns portadores livres são excitados termicamente). Vamos chamar a concentração de átomos de impureza em um semiconductor tipo **n** de N_d (o subscripto d se refere ao fato de o átomo de impureza ser um doador de elétrons). Assim, temos que, para um semiconductor tipo **n**:

$$n \approx N_d$$

Equação 173

Mas, da Equação 158, temos também que:

$$n = n_c \exp\left(-\frac{\varepsilon_C - \varepsilon_F}{k_B T}\right)$$

Equação 158

Portanto:

$$N_d \approx n_c \exp\left(-\frac{\varepsilon_C - \varepsilon_F}{k_B T}\right)$$

Equação 174

Podemos, agora, facilmente encontrar o nível de Fermi para um semiconductor tipo **n**. Para isso, basta isolar ε_F na Equação 174, o que resulta em:

$$\varepsilon_F = \varepsilon_C + k_B T \ln\left(\frac{N_d}{n_c}\right)$$

Equação 175

Podemos fazer ainda melhor: podemos rearranjar a Equação 158 para expressá-la em termos de parâmetros do semiconductor intrínseco, que já calculamos na seção anterior.

Primeiro, vamos dar nome aos bois: vamos chamar o nível de Fermi do semiconductor intrínseco (dado pela Equação 171 ou Equação 172) de ε_{FI} , onde o subscrito I serve para lembrarmos que o termo designa somente o nível de Fermi do semiconductor intrínseco. Vamos chamar também os portadores livres do semiconductor intrínseco de n_i e p_i (respectivamente: Equação 158 e Equação 164, ambas com $\varepsilon_F = \varepsilon_{FI}$).

Agora, vamos rearranjar a Equação 158 para expressá-la em termos de n_i e ε_{FI}

$$n = n_c \exp\left(-\frac{\varepsilon_C - \varepsilon_F}{k_B T}\right) = n_c \exp\left(-\frac{\varepsilon_C - \varepsilon_{FI} + \varepsilon_{FI} - \varepsilon_F}{k_B T}\right)$$

$$\therefore n = n_c \exp\left(-\frac{\varepsilon_C - \varepsilon_{FI}}{k_B T}\right) \exp\left(-\frac{\varepsilon_{FI} - \varepsilon_F}{k_B T}\right)$$

Equação 176

Comparando o lado direito da Equação 176 com a expressão para n_i , concluímos que:

$$n = n_i \exp\left(-\frac{\varepsilon_{FI} - \varepsilon_F}{k_B T}\right)$$

Equação 177

A Equação 177 expressa a concentração de elétrons livres em termos do nível de Fermi intrínseco (que está no meio do *band-gap*) e o nível de Fermi do semiconductor dopado. Claro que, se $\varepsilon_F = \varepsilon_{FI}$, então $n = n_i$. Lembre-se que, tanto Equação 158 como a Equação 177 são gerais: valem para qualquer semiconductor com qualquer dopagem (inclusive para semicondutores intrínsecos, já que se $\varepsilon_F = \varepsilon_{FI}$, então $n = n_i$).

Agora, se estivermos tratando especificamente de um semiconductor com dopagem tipo **n**, onde a concentração de dopantes é muito maior que a concentração de portadores intrínsecos, então a Equação 173 também é verdade. Portanto, neste caso, temos que:

$$N_d \approx n_i \exp\left(-\frac{\varepsilon_{FI} - \varepsilon_F}{k_B T}\right)$$

Equação 178

E, portanto:

$$\varepsilon_F = \varepsilon_{FI} + k_B T \ln\left(\frac{N_d}{n_i}\right)$$

Equação 179

A Equação 178 e a Equação 174 são equivalentes: ambas expressam a mesma coisa, mas utilizando parâmetros diferentes. Tanto uma quanto a outra podem ser utilizadas. De maneira análoga, a Equação 179 e a Equação 175 também são equivalentes.

Note que quanto maior for N_d em relação a n_i , maior será o deslocamento do nível de Fermi em relação ao nível de Fermi intrínseco (ou seja, em relação ao centro do *band-gap*). Não se esqueça que a Equação 174, a Equação 175, a Equação 178 e a Equação 179 só valem para semicondutores do tipo **n** (por que? Qual condição restringe a aplicação dessas equações para semicondutores do tipo **n**?).

Podemos repetir o mesmo raciocínio para encontrarmos uma expressão para a densidade de buracos. Começamos reescrevendo a Equação 164 em termos de p_i e ε_{FI} :

$$p = n_d \exp\left(-\frac{\varepsilon_F - \varepsilon_V}{k_B T}\right) = n_d \exp\left(-\frac{\varepsilon_F - \varepsilon_{FI} + \varepsilon_{FI} - \varepsilon_V}{k_B T}\right)$$

$$\therefore p = n_d \exp\left(-\frac{\varepsilon_{FI} - \varepsilon_V}{k_B T}\right) \exp\left(-\frac{\varepsilon_F - \varepsilon_{FI}}{k_B T}\right)$$

Equação 180

De onde concluímos que:

$$p = p_i \exp\left(-\frac{\varepsilon_F - \varepsilon_{FI}}{k_B T}\right)$$

Equação 181

Esse é um resultado geral e válido para qualquer semicondutor com qualquer dopagem.

Se estivermos tratando especificamente de um semicondutor tipo **p** onde a concentração de dopantes é N_a (o subscrito a vem do inglês *acceptor*, pois o átomo de impureza recebe um elétron), e se essa concentração for bem maior que a concentração de portadores intrínsecos, então também é verdade que:

$$p \approx N_a$$

Equação 182

De onde concluímos que para um semicondutor tipo **p**:

$$N_a = p_i \exp\left(-\frac{\varepsilon_F - \varepsilon_{FI}}{k_B T}\right)$$

Equação 183

Portanto, o nível de Fermi do semicondutor tipo **p** será:

$$\varepsilon_F = \varepsilon_{FI} - k_B T \ln\left(\frac{N_a}{p_i}\right)$$

Equação 184

Note que quanto maior for N_a em relação a p_i , mais deslocado o nível de Fermi estará em relação ao centro do *band-gap*, mas agora o deslocamento é para baixo.

2. 7 Exercícios

Exercício 1

Explique como as bandas de energia são formadas.

Exercício 2

Explique como o preenchimento das bandas define as propriedades elétricas de um material.

Exercício 3

- a) Explique qualitativamente o conceito de densidade de estados.
- b) Deduza uma expressão matemática para a densidade de estados de um gás ideal.
- c) Explique a passagem da expressão da densidade de estados de um gás ideal para a densidade de estados do “gás” de elétrons livres na banda de condução.

Exercício 4

Por que a Equação 171 não é válida para semicondutores extrínsecos? O que foi utilizado na dedução dessa equação que não é válido para semicondutores extrínsecos:

Exercício 5

Prove que o produto entre a concentração de elétrons livres e buracos não é alterado pelo processo de dopagem.

Se esse produto não é alterado, o que ocorrerá com a concentração de buracos quando o semicondutor for dopado com impurezas doadoras (dopagem tipo **n**)? E o que ocorrerá com a concentração de elétrons livres quando o semicondutor for dopado com impurezas receptoras (dopagem tipo **p**)?

Exercício 6

Suponha que um semicondutor intrínseco esteja a uma temperatura T . Qual é a média da energia dos elétrons na banda de condução desse semicondutor? Discuta o significado físico de cada termo da energia média (qual parte é cinética e qual parte potencial).

Exercício 7

A energia cinética média dos elétrons livres em um certo semicondutor de silício intrínseco é de 0.025 e.V. (elétrons-volts). Qual é a concentração de elétrons livres nesse semicondutor?

Exercício 8

Encontre uma expressão para o potencial químico de um gás ideal de elétrons livres, em termos da sua concentração. Compare seu resultado com a expressão dada no Exercício 9 do capítulo 1. Qual a origem física da diferença?

3 – Introdução aos dispositivos semicondutores: a junção p-n

Neste capítulo, vamos utilizar todos os conceitos desenvolvidos nos capítulos anteriores para estudar a junção **p-n**, que forma a base de toda a eletrônica.

Começaremos estudando as propriedades da junção **p-n** em equilíbrio, ou seja, sem nenhuma tensão externa aplicada. Em seguida, estudaremos as características de tensão e corrente quando uma tensão externa é aplicada, o que corresponde a uma situação de desequilíbrio.

3.1 Junção p-n em equilíbrio

Como sugerido pelo nome, a junção **p-n** é formada pela união de um semicondutor tipo **p** com um semicondutor tipo **n**. Na prática, junções **p-n** são formadas através da dopagem seletiva de um único substrato semicondutor; mas por razões pedagógicas, vamos supor, aqui, que a junção seja formada pela união física de dois semicondutores em substratos independentes, como ilustrado na Figura 19. O que vai ocorrer quando um semicondutor tipo **n** for colocado em contato com um semicondutor tipo **p**? Sabemos que existe uma concentração maior de elétrons livres no tipo **n** do que no tipo **p**. Essa concentração maior de elétrons livres corresponde ao nível de Fermi mais alto no semicondutor tipo **n** (Figura 19a); e já sabemos o que ocorre quando o nível de Fermi (ou seja, o potencial eletroquímico) é mais alto de um lado do que de outro: os elétrons livres difundirão da região de maior potencial (maior nível de Fermi) para a região de

menor potencial (menor nível de Fermi), ou seja, elétrons difundirão do semiconductor tipo **n** para o semiconductor tipo **p**. Estudamos um caso muito similar no Capítulo 1 (Figura 6). Essa difusão resultará no acúmulo local de cargas que, por sua vez, resulta na formação de um campo elétrico que age de maneira a inibir a corrente de difusão. O potencial eletrostático decorrente da formação do campo elétrico altera o potencial eletroquímico até que este seja constante ao longo do dispositivo, resultando, assim, em equilíbrio e encerramento da corrente (Figura 19b). Note que nem o potencial eletrostático nem o potencial químico são individualmente constantes: o que é constante é a soma dos dois, ou seja, o potencial eletroquímico (que é o nível de Fermi). Esse mesmo fenômeno pode ser descrito em termos de correntes: a corrente de arrasto (que é resultante do aparecimento do campo elétrico) cancela a corrente de difusão (que é resultante da diferença de potencial químico). Leia a legenda da Figura 19 com atenção e tente compreender a relação entre as bandas de energia e o potencial eletrostático em unidade de volts $V(x)$.

Assim, o que aprendemos no capítulo 1 é suficiente para explicar qualitativamente o que ocorre com os elétrons livres na formação da junção **p-n**. Mas existe uma importante diferença entre a junção **p-n** e o exemplo do Capítulo 1 (Figura 6): na junção **p-n**, além de elétrons livres, temos também buracos. Intuitivamente, sabemos que os buracos difundirão da região **p** para a região **n**, já que região **p** tem uma concentração maior de buracos. Mas note que a região **p** possui nível de Fermi mais baixo, portanto, os buracos vão do potencial eletroquímico (nível de Fermi) menor para o potencial maior, o que é o oposto do que ocorre para elétrons livres (e para qualquer outra partícula). Essa aparente esquizofrenia dos buracos é facilmente compreendida lembrando que, na verdade, são os elétrons da banda de valência que estão migrando da região **n** para a região **p**. Portanto, buracos indo de **p** (menor nível de Fermi) para **n** (maior nível de Fermi) correspondem a elétrons da banda de valência indo de **n** para **p**. **Lembre-se, então, que buracos vão do nível de Fermi mais baixo para o mais alto.** Naturalmente, o fluxo total (resultante da corrente de difusão e da corrente de arrasto), tanto de elétrons quanto de buracos, cessará quando o nível de Fermi equalizar, como representado na Figura 19b.

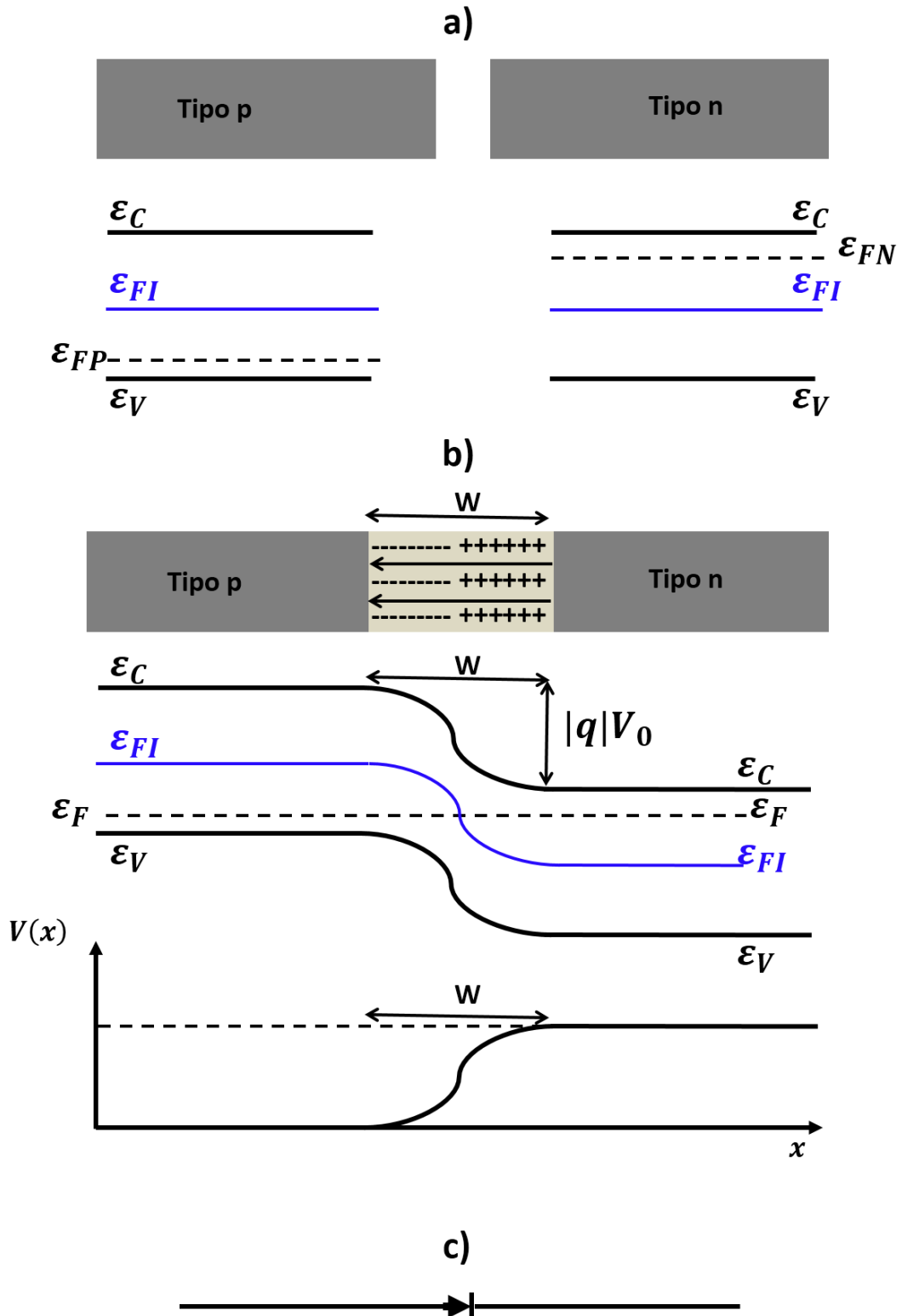


Figura 19 Experimento “mental” mostrando a formação de uma junção p-n através da união de dois semicondutores dopados. a) Semicondutores e níveis de energia antes da formação da junção: ϵ_{FI} é o nível de Fermi intrínseco e ϵ_{FP} e ϵ_{FN} são os níveis de Fermi do semiconductor tipo p e do tipo n, respectivamente. b) (topo) semiconductor após a formação da junção; (meio) níveis de energia (em joules) do semiconductor após a formação da junção; (base) potencial eletrostático em unidades de volts ao longo da junção (lembre-se que um potencial eletrostático em unidade de volts $V(x)$ corresponde a um potencial eletrostático em unidade de joules $\epsilon = -|q|V(x)$). Esse sinal de negativo na carga do elétron faz com que o diagrama de energia caia quando $V(x)$ sobe. A região de

comprimento W é chamada de região de depleção. As setas na região de depleção indicam a direção do campo elétrico. Note que o nível de Fermi é constante ao longo da junção, o que corresponde à condição de equilíbrio. c) símbolo da junção p-n (diodo) utilizado em circuitos.

Analisando a Figura 19b com atenção, podemos notar que o potencial eletrostático resultante do campo elétrico, ao equalizar os níveis de Fermi, também entorta as bandas de energia. Note, também, que o nível de Fermi intrínseco, por estar sempre no centro do *band-gap*, também é entortado. A origem física do entortamento das bandas é muito simples: o potencial eletrostático está desempenhando um papel análogo ao potencial gravitacional. De fato, o entortamento nas bandas de energia advinda da presença do potencial eletrostático é análogo à variação na energia total de uma pessoa que deita em um morro: se você comparar a energia total de uma pessoa deitada em um plano (nesse caso a energia total seria apenas a energia interna dela) com a energia total dessa mesma pessoa deitada em um morro (que agora seria a energia interna mais a energia gravitacional), então obviamente a energia potencial advinda da altura do morro entortará o diagrama de energia dessa pessoa. Essa analogia está representada de maneira meio tosca na Figura 20.

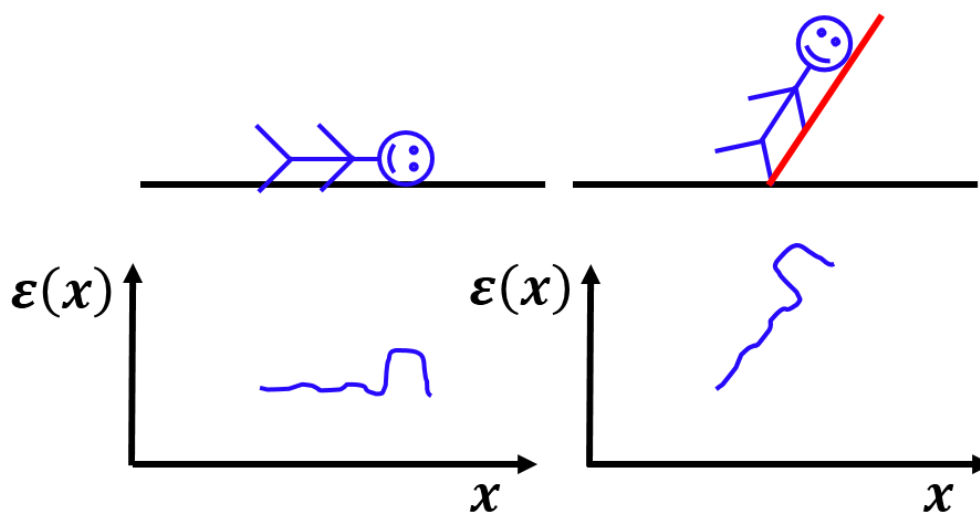


Figura 20 Analogia gravitacional do efeito da energia potencial eletrostática nas bandas de energia da junção p-n. Lado esquerdo: energia da pessoa deitada em um plano (esse caso é análogo às bandas antes da formação da junção). Lado direito: energia da mesma pessoa deitada em um morro (esse caso é análogo às bandas após a formação da junção, e consequente aparecimento do campo elétrico). A energia potencial gravitacional do morro entorta o diagrama de energia, de maneira análoga ao que ocorre com as bandas de energia do semiconductor em razão da energia potencial eletrostática.

Voltando à Figura 19b, note a região onde há acúmulo de cargas na junção. Atribuímos a largura W a ela. Essa região é chamada de região de depleção, termo este que denota o fato dela não possuir cargas livres, uma vez que os portadores de carga

livres que estavam ali presentes difundiram para o outro lado da junção. É importante notar que as cargas na região de depleção são formadas primordialmente pelos íons dos átomos dopantes (e que não são nada livres). Lembre-se também que os átomos dopantes são, a princípio, neutros, mas após a dopagem, eles são ionizados porque perdem um elétron para a banda de condução por excitação térmica – se for dopagem tipo **n**, ou ganham um elétron da banda de valência por excitação térmica, - se for dopagem tipo **p**. Voltaremos a comentar sobre a distribuição de cargas na região de depleção mais adiante, mas antes, vamos revisitar o problema do cálculo do potencial eletrostático (denotado na Figura 19 por V_0) resultante do campo elétrico na região de depleção.

No capítulo 1, deduzimos que $|q|V_0$ deve ser igual à diferença de potencial químico que existe antes da interação entre os sistemas. Sabemos também que os níveis de Fermi correspondem ao potencial total (eletroquímico); mas como antes da formação da junção o campo elétrico ainda não existe, então o potencial total nesse caso será o próprio potencial químico. Em outras palavras, os níveis de Fermi dos semicondutores **n** e **p** antes da formação da junção correspondem aos seus potenciais químicos. Assim, esperamos que $|q|V_0$ seja igual à diferença entre os níveis de Fermi antes da formação da junção, fato este que já está representado na Figura 19b. Em termos matemáticos, esperamos, então, que $|q|V_0 = \varepsilon_{FN} - \varepsilon_{FP}$. Entretanto, precisamos provar isso de maneira um pouco mais formal, porque a dedução do capítulo 1 foi feita de maneira fenomenológica. Esse é um termo sofisticado para descrever uma dedução informal, que parte do bom senso e de intuições, mas que não é muito rigorosa. É necessário e instrutivo, portanto, revisitar esse problema de maneira mais formal, uma vez que já ganhamos (espero) a intuição física do problema.

A dedução formal da relação entre V_0 e os potenciais químicos utiliza a condição de equilíbrio no transporte de portadores de cargas. Como afirmado anteriormente, uma maneira de visualizar a condição de equilíbrio é interpretá-la como sendo resultante do equilíbrio entre corrente de arrasto e de difusão: o campo elétrico induz uma corrente de arrasto contrária à corrente de difusão, de maneira que a corrente total seja nula. Esse equilíbrio é atingido independentemente tanto para elétrons livres como para buracos, mas podemos escolher apenas um desses portadores para fazer o cálculo. Vou escolher os buracos, por terem carga positiva.

Para expressar a condição de equilíbrio para buracos, em termos de suas correntes, temos que primeiro lembrar quais são as expressões matemáticas relevantes. Vimos na seção sobre transporte (seção 1.10) que a densidade de corrente de arrasto (que chamaremos novamente de J_{arr}) e a densidade de corrente de difusão (que chamaremos novamente de J_{dif}) são expressas, respectivamente, por:

$$J_{arr} = \rho \cdot v = |q| \cdot p \cdot v$$

Equação 185

e

$$J_{dif} = -D \cdot \frac{\partial \rho}{\partial x} = -|q| \cdot D \cdot \frac{\partial p}{\partial x}$$

Equação 186

Onde ρ é a concentração de cargas, v é a velocidade das cargas, D é o coeficiente de difusão e p é a concentração de buracos. Vamos reescrever essas duas relações de maneira um pouco diferente. Primeiro, note que, obviamente, a concentração de cargas ρ está relacionada à concentração de buracos p ($\rho = |q|p$). Além disso, de acordo com a lei de Ohm, a velocidade das cargas na corrente de arrasto é proporcional ao campo elétrico. A constante de proporção é chamada de mobilidade e representada pelo símbolo μ (esse é o mesmo símbolo utilizado para potencial eletroquímico no capítulo 1, então cuidado para não confundir). Assim, a corrente de arrasto pode ser reexpressa como:

$$J_{arr} = |q| \cdot p(x) \cdot v = |q| \cdot p(x) \cdot \mu \cdot E(x)$$

Equação 187

Onde $E(x)$ é o campo elétrico. Note que já coloquei explicitamente quais parâmetros dependem da posição ao longo da junção, denotada pela variável x . Podemos, também, encontrar uma relação mais conveniente para a corrente de difusão. Einstein mostrou (e você também vai ter que mostrar na lista de exercícios) que o coeficiente de difusão está relacionado à mobilidade e à temperatura por:

$$\frac{D}{\mu} = \frac{k_B \cdot T}{|q|}$$

Equação 188

Utilizando essa relação, podemos reexpressar a corrente de difusão como:

$$J_{dif} = -|q| \cdot D \cdot \frac{\partial p}{\partial x} = -\mu \cdot k_B \cdot T \cdot \frac{\partial p}{\partial x}$$

Equação 189

A condição de equilíbrio, obviamente, é expressa matematicamente por:

$$J_{arr} + J_{dif} = 0$$

Equação 190

Portanto, temos que:

$$|q| \cdot p(x) \cdot \mu \cdot E(x) - \mu \cdot k_B \cdot T \cdot \frac{\partial p}{\partial x} = 0$$

Equação 191

Ou seja:

$$E(x) = \frac{k_B \cdot T}{|q|} \cdot \frac{1}{p(x)} \cdot \frac{\partial p}{\partial x}$$

Equação 192

Vamos parar um pouco para lembrar o que queremos. Queremos uma relação entre V_0 e os níveis de Fermi do semiconductor antes da formação da junção, denotados por ε_{FP} e ε_{FN} , na Figura 19. Vimos, no capítulo anterior, que as concentrações de portadores de cargas (elétrons livres e buracos) dependem desses valores (Equação 176 e Equação 181). Portanto, o lado direito da Equação 192 é função dos níveis de Fermi por meio de $p(x)$. Por outro lado, o potencial eletrostático pode ser encontrado através da integral do campo elétrico. Então, estamos no caminho certo, pois se integrarmos os dois lados da Equação 192 teremos o potencial eletrostático do lado esquerdo e uma expressão para $p(x)$ (que depende dos níveis de Fermi) do lado direito. Para fazer a integral, é

conveniente colocar um eixo e uma origem explícitos. O eixo escolhido é representado na Figura 21.

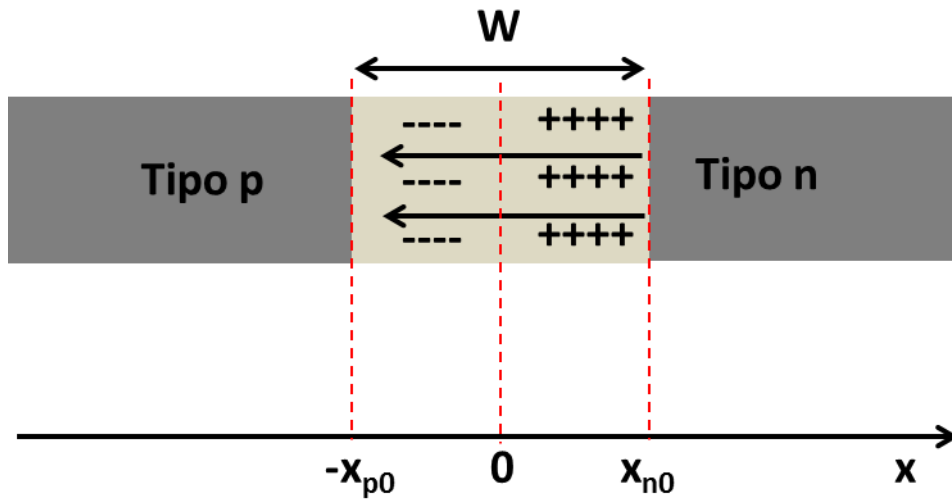


Figura 21 Representação do eixo de coordenadas. A origem é colocada no centro da região de depleção. O ponto x_{n0} denota o início da região de depleção na parte n enquanto o ponto $-x_{p0}$ denota o início da região da depleção na parte p.

Como o campo elétrico está limitado à região de depleção, a diferença de potencial eletrostático entre a parte **n** e a parte **p** será dada por (veja também a Figura 19b):

$$V(x_{n0}) - V(-x_{p0}) = - \int_{-x_{p0}}^{x_{n0}} E(x) dx$$

Equação 193

Como V_0 é a diferença entre o potencial na parte **n** e o potencial na parte **p**, (ou seja, $V_0 = V(x_{n0}) - V(-x_{p0})$), veja novamente a Figura 19b) temos que:

$$V_0 = - \int_{-x_{p0}}^{x_{n0}} E(x) dx$$

Equação 194

Substituindo $E(x)$ da Equação 192, temos:

$$V_0 = - \int_{-x_{p0}}^{x_{n0}} \frac{k_B \cdot T}{|q|} \cdot \frac{1}{p(x)} \cdot \frac{\partial p}{\partial x} dx$$

$$\therefore V_0 = -\frac{k_B \cdot T}{|q|} [\ln p(x_{n0}) - \ln p(-x_{p0})]$$

Equação 195

Assim, chegamos à importante relação:

$$V_0 = \frac{k_B \cdot T}{|q|} \ln \left[\frac{p(-x_{p0})}{p(x_{n0})} \right]$$

Equação 196

Lembre-se que poderíamos ter utilizado a condição de equilíbrio para as correntes de elétrons livres. Se tivéssemos feito isso, chegaríamos à seguinte relação:

$$V_0 = \frac{k_B \cdot T}{|q|} \ln \left[\frac{n(x_{n0})}{n(-x_{p0})} \right]$$

Equação 197

O que implica que:

$$n(-x_{p0}) \cdot p(-x_{p0}) = n(x_{n0}) \cdot p(x_{n0})$$

Equação 198

Esse resultado, na verdade, poderia ter sido deduzido utilizando apenas as expressões do capítulo 2. Note que a Equação 198 afirma que o produto entre as concentrações dos portadores no semiconductor tipo **p** (lado esquerdo da equação) é igual ao produto para o semiconductor tipo **n** (lado direito da equação). Isso significa que o produto $n \cdot p$ não depende da dopagem. Portanto, o produto $n \cdot p$ também é igual ao produto $n_i \cdot p_i$ do semiconductor intrínseco (o que não é de se surpreender, já que vimos na Equação 167 que o produto $n_i \cdot p_i$ não depende do nível de Fermi).

Voltando para o nosso objetivo, que é expressar V_0 em termos dos níveis de Fermi antes da formação da junção, podemos utilizar tanto a Equação 196 como a Equação 197. É claro que temos que utilizar também as relações deduzidas no capítulo 2 para expressar as concentrações em termos dos níveis de Fermi (Equação 176 ou Equação

181). Escolhendo a concentração de buracos, podemos utilizar a Equação 181. Substituindo-a na Equação 196, temos:

$$\begin{aligned}
 V_0 &= \frac{k_B \cdot T}{|q|} \ln \left[\frac{p(-x_{p0})}{p(x_{n0})} \right] = \frac{k_B \cdot T}{|q|} \ln \left[\frac{p_i \exp \left[-\frac{\varepsilon_F - \varepsilon_{FI}(-x_{p0})}{k_B T} \right]}{p_i \exp \left[-\frac{\varepsilon_F - \varepsilon_{FI}(x_{n0})}{k_B T} \right]} \right] \\
 &= \frac{k_B \cdot T}{|q|} \ln \left[\frac{\exp \left[-\frac{\varepsilon_F - \varepsilon_{FI}(-x_{p0})}{k_B T} \right]}{\exp \left[-\frac{\varepsilon_F - \varepsilon_{FI}(x_{n0})}{k_B T} \right]} \right]
 \end{aligned}$$

Equação 199

Lembre-se que o nível de Fermi é constante, mas o nível de Fermi intrínseco varia ao longo da junção, como mostrado na Figura 19b. Por isso, ε_F não depende de x , mas ε_{FI} depende.

Mas note também o seguinte: o que aparece na Equação 199 é a diferença entre o nível de Fermi e o nível de Fermi intrínseco. Fora da região de depleção, essa diferença ($\varepsilon_F - \varepsilon_{FI}$) é a mesma que existe no semiconductor dopado antes da formação da junção (compare a Figura 19a com a Figura 19b).

Em termos matemáticos, isso significa que na região do semiconductor tipo **p** (ou seja, para $x \leq -x_{p0}$) temos:

$$(\varepsilon_F - \varepsilon_{FI})_{JUNÇÃO \text{ para } x \leq -x_{p0}} = (\varepsilon_{FP} - \varepsilon_{FI})_{\text{tipo p ANTES DA JUNÇÃO}}$$

Equação 200

Onde o lado esquerdo da igualdade se refere à região **p** da junção (região **p** da Figura 19b) e o lado direito se refere ao semiconductor tipo **p** antes da formação da junção (Figura 19a).

Similarmente, na região do semiconductor tipo **n** (ou seja, para $x \geq x_{n0}$), temos:

$$(\varepsilon_F - \varepsilon_{FI})_{JUNÇÃO \text{ para } x \geq x_{n0}} = (\varepsilon_{FN} - \varepsilon_{FI})_{\text{tipo n ANTES DA JUNÇÃO}}$$

Equação 201

Portanto, podemos reexpressar as diferenças na Equação 199 (que se referem à junção) em termos dos níveis de Fermi antes da formação da junção, ou seja:

$$\begin{aligned}
 V_0 &= \frac{k_B \cdot T}{|q|} \ln \left[\frac{\exp \left[-\frac{\varepsilon_F - \varepsilon_{FI}(-x_{p0})}{k_B T} \right]_{JUNÇÃO}}{\exp \left[-\frac{\varepsilon_F - \varepsilon_{FI}(x_{n0})}{k_B T} \right]_{JUNÇÃO}} \right] \\
 &= \frac{k_B \cdot T}{|q|} \ln \left[\frac{\exp \left[-\frac{\varepsilon_{FP} - \varepsilon_{FI}}{k_B T} \right]_{\text{tipo } p \text{ ANTES DA JUNÇÃO}}}{\exp \left[-\frac{\varepsilon_{FN} - \varepsilon_{FI}}{k_B T} \right]_{\text{tipo } n \text{ ANTES DA JUNÇÃO}}} \right]
 \end{aligned}$$

Equação 202

Como o nível de Fermi intrínseco antes da formação da junção é o mesmo para o semiconductor tipo **p** e tipo **n**, temos o mesmo ε_{FI} no numerador e no denominador da segunda igualdade na Equação 202. Portanto:

$$\begin{aligned}
 V_0 &= \frac{k_B \cdot T}{|q|} \ln \left[\frac{\exp \left[-\frac{\varepsilon_{FP} - \varepsilon_{FI}}{k_B T} \right]_{\text{tipo } p \text{ ANTES DA JUNÇÃO}}}{\exp \left[-\frac{\varepsilon_{FN} - \varepsilon_{FI}}{k_B T} \right]_{\text{tipo } n \text{ ANTES DA JUNÇÃO}}} \right] \\
 &= \frac{k_B \cdot T}{|q|} \ln \left[\exp \left[-\frac{\varepsilon_{FP} - \varepsilon_{FN}}{k_B T} \right] \right]
 \end{aligned}$$

Equação 203

Daí, concluímos que, como esperado:

$$V_0 = \frac{\varepsilon_{FN} - \varepsilon_{FP}}{|q|}$$

Equação 204

A Equação 204 é análoga à Equação 41. A vida é mesmo uma incansável repetição. Para finalizar essas elucubrações a respeito da junção **p-n** em equilíbrio, precisamos encontrar o campo elétrico na região de depleção e quantificar sua largura. Para fazer isso, precisamos, primeiro, identificar como as cargas são distribuídas na região de depleção (lembre-se que a região de depleção não possui cargas livres, o que resulta em acúmulo de cargas não livres). Os elétrons que migraram da região **n** deixaram para trás

os átomos dopantes. Como esses átomos perderam elétrons (pois os elétrons migraram), eles ficaram ionizados. Em linguagem estranha, porém técnica, diz-se que os elétrons descobriram os íons (descobrir no sentido contrário a cobrir, não no sentido de achar). Alguns livros, sem o menor senso do ridículo, utilizam a expressão “íons pelados” para denotar os íons “descobertos”. As cargas positivas na região **n** advêm principalmente desses íons “descobertos”. De maneira análoga, as cargas negativas na região **p** advêm principalmente dos íons descobertos na região **p**. Eu disse principalmente porque as cargas livres que difundiram sofrem recombinação (deixando, então, de ser livres) e também contribuem para a formação da região de depleção, mas essas cargas geralmente estão mais “diluídas” ao longo do semicondutor, enquanto os íons estão fixos na região de depleção. Portanto, a concentração de íons costuma ser bem mais relevante. Para facilitar as contas, vamos negligenciar a contribuição das cargas recombinadas (o que é, na prática, uma ótima aproximação).

Se a dopagem for uniforme, então os íons estarão uniformemente distribuídos ao longo da junção. Portanto, a densidade de cargas será uniforme ao longo da junção, apresentando transições abruptas nos limites da junção, como mostrado na Figura 22, onde N_d é a concentração de íons doadores e N_a é a concentração de íons aceitadores. Note que a densidade de cargas é zero fora da região de depleção, uma vez que os íons fora da região de depleção não estão descobertos (outra forma de dizer que o material é neutro fora da região de depleção; não se esqueça que a junção como um todo também é neutra).

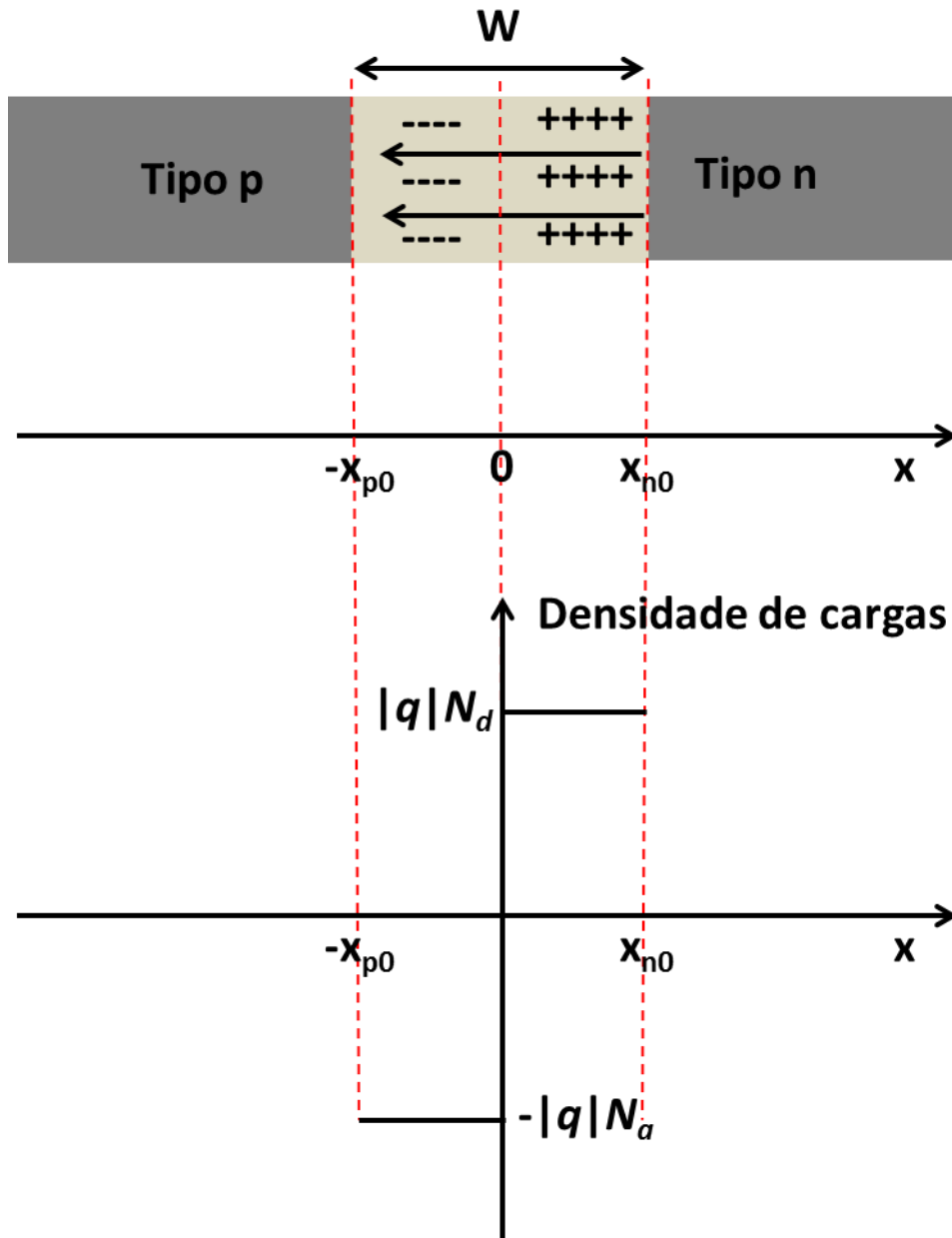


Figura 22 Distribuição especial da densidade de cargas.

Antes de prosseguirmos, é bom enfatizar que a distribuição de íons não precisa ser necessariamente uniforme. Existem várias maneiras de se dopar um semiconductor e, a priori, qualquer tipo de distribuição pode ser obtido, dependendo da dopagem. Mas vamos tratar só da distribuição uniforme por duas razões: ela é simples e ela é comum.

Já encontramos a diferença de potencial entre os extremos da junção, mas ainda não dissemos nada sobre o campo elétrico em si, apenas que ele é confinado na região de depleção. Por outro lado, acabamos de identificar como a densidade de cargas é distribuída ao longo da junção (estamos assumindo distribuição uniforme). Podemos utilizar essa informação para encontrarmos o campo elétrico na junção? Sabemos que

o campo elétrico surge exatamente do acúmulo de cargas locais na região de depleção, mas qual é a relação matemática entre essas cargas e o campo elétrico? Você deve se lembrar, de eletromagnetismo, que a relação entre campo elétrico e densidade de cargas é dada pela importante Lei de Gauss:

$$\nabla \cdot \vec{E} = \frac{\rho}{\epsilon}$$

Equação 205

Onde ϵ é a constante dielétrica do material (no nosso caso, é a constante dielétrica do semiconductor). Como nosso problema é unidimensional, a Lei de Gauss é reduzida a:

$$\frac{\partial E}{\partial x} = \frac{\rho}{\epsilon}$$

Equação 206

Mas para distribuição uniforme, temos que (Figura 22):

$$\rho = -|q|N_a \quad \text{para} \quad -x_{p0} \leq x < 0$$

e

$$\rho = |q|N_d \quad \text{para} \quad 0 < x \leq x_{n0}$$

Equação 207

Portanto:

$$\frac{\partial E}{\partial x} = \frac{-|q|N_a}{\epsilon} \quad \text{para} \quad -x_{p0} \leq x < 0$$

e

$$\frac{\partial E}{\partial x} = \frac{|q|N_d}{\epsilon} \quad \text{para} \quad 0 < x \leq x_{n0}$$

Equação 208

Integrando a Equação 208, temos:

$$E(x) = E(0) - \frac{|q|N_a}{\epsilon}x \text{ para } -x_{p0} \leq x < 0$$

e

$$E(x) = E(0) + \frac{|q|N_d}{\epsilon}x \text{ para } 0 < x \leq x_{n0}$$

Equação 209

Sabendo que o campo elétrico está confinado na região de depleção, a Equação 209 resulta em uma dependência espacial do campo elétrico como representada na Figura 23 (note que $E(0) < 0$).

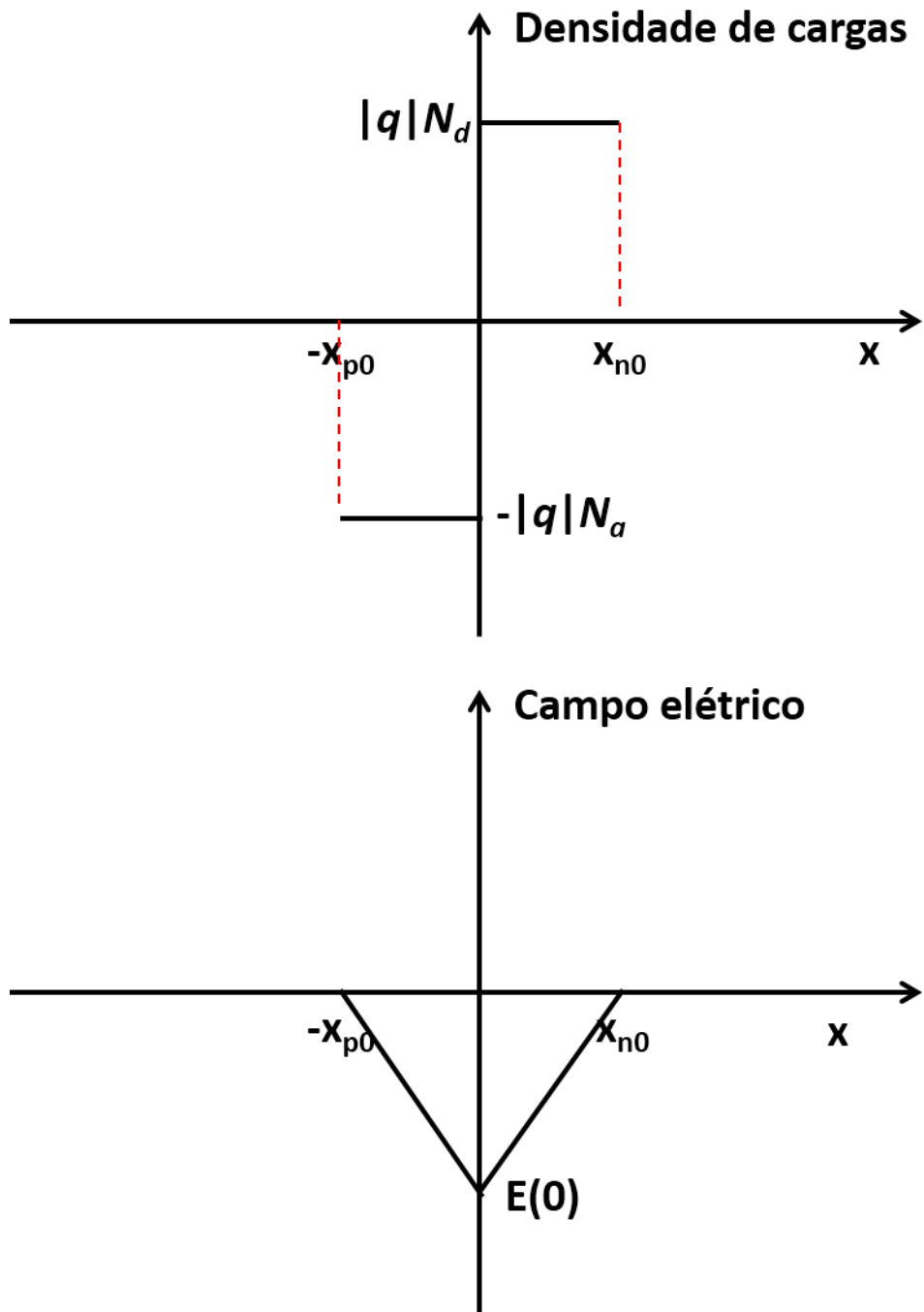


Figura 23 Dependência especial da densidade de cargas e do campo elétrico.

Não é difícil deduzir (portanto deduza sozinho) da Equação 209 e do fato de que o campo está confinado na região de depleção que o valor máximo do módulo do campo é dado por:

$$|E(0)| = \frac{|q|N_a}{\epsilon} x_{p0} = \frac{|q|N_d}{\epsilon} x_{n0}$$

Equação 210

O que implica que $N_a x_{p0} = N_d x_{n0}$. Como a largura da região de depleção é dada por $W = x_{n0} + x_{p0}$, temos também:

$$W = x_{n0} + x_{p0} = x_{n0} \left(1 + \frac{N_d}{N_a} \right)$$

Equação 211

Note também que como:

$$V_0 = - \int_{-x_{p0}}^{x_{n0}} E(x) dx$$

Então, V_0 coincide com área do triângulo da Figura 23, ou seja:

$$V_0 = \frac{(x_{n0} + x_{p0})|E(0)|}{2} = \frac{W|E(0)|}{2}$$

Assim, pela Equação 210 temos que:

$$V_0 = \frac{W}{2} |E(0)| = \frac{W}{2} \frac{|q|N_a}{\epsilon} x_{p0}$$

ou

$$V_0 = \frac{W}{2} |E(0)| = \frac{W}{2} \frac{|q|N_d}{\epsilon} x_{n0}$$

Equação 212

Juntando a Equação 211 com a Equação 212, temos, finalmente, que:

$$W = \sqrt{\frac{2 \cdot \epsilon \cdot V_0}{|q|} \left(\frac{N_a + N_d}{N_a \cdot N_d} \right)}$$

Equação 213

Note que todos os parâmetros físicos ($E(x)$, V_0 e W) foram descritos em termos dos parâmetros de dopagem N_a e N_d , que são os parâmetros de controle do engenheiro projetista (lembre-se também que esses últimos definem os níveis de Fermi ε_{FN} e ε_{FP} pela Equação 179 e Equação 184; portanto V_0 na Equação 204 pode ser facilmente expresso em termos de N_a e N_d).

Assim, já estabelecemos as principais propriedades da junção **p-n** em equilíbrio. Vimos como o potencial eletrostático e a largura da região de depleção dependem da diferença inicial entre os níveis de Fermi do semiconductor tipo **n** e do semiconductor tipo **p**.

Mas não podemos parar por aqui, porque quando a junção **p-n** for utilizada como um dispositivo eletrônico, precisaremos saber o que ocorre quando uma tensão externa for aplicada. Essa já não é mais uma situação de equilíbrio, uma vez que a tensão externa constitui uma fonte de energia externa, tirando assim a junção **p-n** da sua condição de equilíbrio.

Antes de tratarmos especificamente das relações de tensão e corrente na junção **p-n**, é necessário discutir o que ocorre com os níveis de Fermi quando o sistema é perturbado para fora de seu estado de equilíbrio. Esse é o assunto da próxima seção.

3.2 Sistemas fora do equilíbrio: níveis de quasi-Fermi.

Vimos no capítulo 2 que em um semiconductor, dopado ou não, as concentrações de elétrons livres e de buracos são dadas, respectivamente, por:

$$n = n_c \exp\left(-\frac{\varepsilon_C - \varepsilon_F}{k_B T}\right) \text{ ou } n = n_i \exp\left(-\frac{\varepsilon_{FI} - \varepsilon_F}{k_B T}\right)$$

Equação 214

e

$$p = n_d \exp\left(-\frac{\varepsilon_F - \varepsilon_V}{k_B T}\right) \text{ ou } p = p_i \exp\left(-\frac{\varepsilon_F - \varepsilon_{FI}}{k_B T}\right)$$

Equação 215

Partindo dessas duas equações, não é difícil provar (mas você tem que provar) que o produto $n \cdot p$ é independente do nível de Fermi, ou seja, $n \cdot p = n_i \cdot p_i$ (lembre-se que n_i e p_i denotam as concentrações do semiconductor intrínseco). Portanto, o processo de

dopagem não altera o produto $n \cdot p$: se você aumentar o nível de Fermi através de dopagem tipo **n**, você aumentará a concentração n de elétrons livres, mas a concentração de buracos p diminuirá.

Agora, considere a seguinte situação: suponha que você tenha um semicondutor intrínseco ($n = n_i$ e $p = p_i$). Como visto no capítulo 2, a concentração n e a concentração p do semicondutor intrínseco dependem somente da temperatura e do *band-gap*. Portanto, para uma dada temperatura, a concentração de elétrons e buracos é fixa (e, no caso do semicondutor intrínseco, temos também que elas são iguais, ou seja, $n_i = p_i$). Agora, suponha que você ilumine o semicondutor. O ato de iluminá-lo injeta uma energia extra, porque a absorção da luz faz com que elétrons sejam excitados da banda de valência para a banda de condução. Portanto, no momento em que o semicondutor for iluminado, a energia dos portadores de carga não será mais apenas energia térmica, o que significa que o semicondutor não estará mais em equilíbrio térmico. O que vai ocorrer, então, com os níveis de Fermi?

Para responder a essa pergunta, note que a concentração de elétrons depende de $\varepsilon_C - \varepsilon_F$ (Equação 214): quanto **menor** for $\varepsilon_C - \varepsilon_F$, maior será n . Portanto, como já sabemos, se a concentração de elétrons aumenta, o nível de Fermi sobe (se aproxima da banda de condução). Por outro lado, a concentração de buracos depende de $\varepsilon_F - \varepsilon_V$ (Equação 215): se a concentração de buracos aumenta, o nível de Fermi desce. Então, o nível de Fermi sobe ou desce quando o semicondutor é iluminado? Como a iluminação excita elétrons da banda de valência para a banda de condução, a iluminação aumenta, ao mesmo tempo, tanto a concentração de elétrons livres quanto a de buracos. Portanto, temos uma situação meio paradoxal: o aumento em n faz o nível de Fermi subir, enquanto o aumento em p faz o nível de Fermi descer. Isso só será possível se o nível de Fermi se separar em dois: um nível de Fermi para elétrons e outro nível de Fermi para buracos. Essa separação do nível de Fermi é a marca de um sistema fora do equilíbrio. E chamamos os níveis que surgem da separação de níveis de *quasi-Fermi* (o prefixo “quasi” - que é latim e não português mal feito - advém do fato de que, estritamente falando, a ideia de nível de Fermi é desenvolvida para sistemas em equilíbrio, como vimos amplamente no Capítulo 1).

Note também que quando o semicondutor for iluminado, o produto $n \cdot p$ não será mais igual ao produto $n_i \cdot p_i$ do semicondutor intrínseco em equilíbrio, já que a absorção de luz aumenta tanto a concentração de elétrons como a concentração de

buracos. Essa alteração do produto $n \cdot p$ é consequência direta do fato de estarmos injetando energia externa no sistema.

O nível de *quasi-Fermi* desempenha o mesmo papel do nível de Fermi, mas ele é utilizado em condições de *quasi-equilíbrio*. Esse termo “*quasi-equilíbrio*” denota uma condição fora do equilíbrio, mas que não altera rápido no tempo, como, por exemplo, uma iluminação constante. Em sistemas fora do equilíbrio, é necessário atribuir um nível de *quasi-Fermi* para os elétrons e um nível de *quasi-Fermi* para os buracos. Só assim o produto $n \cdot p$ pode desviar do produto $n_i \cdot p_i$. É claro que os valores dos níveis de *quasi-Fermi* dependem do grau de desequilíbrio. No exemplo do semiconductor iluminado, os níveis de *quasi-Fermi* dependeriam da taxa de absorção de fótons no semiconductor.

Utilizando, então, o conceito de níveis de *quasi-Fermi*, podemos estender a Equação 214 e a Equação 215 para incluir situações fora do equilíbrio. Assim, temos:

$$n = n_c \exp\left(-\frac{\varepsilon_C - \varepsilon_{QFN}}{k_B T}\right) \quad \text{ou} \quad n = n_i \exp\left(-\frac{\varepsilon_{FI} - \varepsilon_{QFN}}{k_B T}\right)$$

Equação 216

e

$$p = n_d \exp\left(-\frac{\varepsilon_{QFP} - \varepsilon_V}{k_B T}\right) \quad \text{ou} \quad p = p_i \exp\left(-\frac{\varepsilon_{QFP} - \varepsilon_{FI}}{k_B T}\right)$$

Equação 217

Onde ε_{QFN} e ε_{QFP} são os níveis de *quasi-Fermi* para elétrons e para buracos, respectivamente. A Figura 24 ilustra a separação dos níveis de Fermi em níveis de *quasi-Fermi* quando o semiconductor sai do equilíbrio. Finalmente, note também que podemos interpretar a Equação 216 e a Equação 217 como incluindo também a condição de equilíbrio, desde que os níveis de *quasi-Fermi* nesse caso se reduzam ao nível de Fermi, como mostrado na Figura 24.

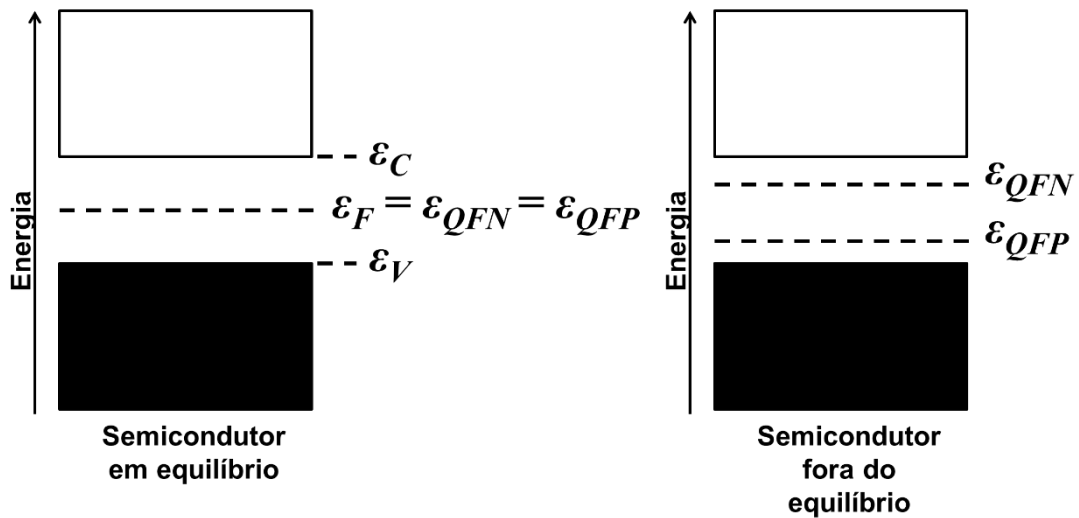


Figura 24 Representação da separação do nível de Fermi quando o semicondutor sai da condição de equilíbrio.

3.3 Relação tensão x corrente na junção p-n (Equação de Shockley).

Chegamos, finalmente, ao último tópico deste capítulo, que é a determinação da relação tensão versus corrente em uma junção **p-n**. Mas antes de calcular essa relação, vamos entender qualitativamente o que vai ocorrer quando aplicarmos uma tensão externa entre os lados **p** e **n** da junção.

Existem duas maneiras de desequilibrar o sistema através da aplicação de uma tensão externa. Essas duas maneiras estão representadas na Figura 25. Se a tensão maior for aplicada do lado **p**, como mostrado na Figura 25a, diz-se então que a junção foi polarizada diretamente. Se, por outro lado, a tensão maior for aplicada no lado **n**, como mostrado na Figura 25b, diz-se então que a junção foi polarizada inversamente (ou reversamente).

Não é difícil visualizar o que vai ocorrer quando a junção for polarizada diretamente ou inversamente. Lembre-se que, quando **nenhuma** tensão externa é aplicada, o equilíbrio é atingido porque a corrente de difusão cancela a corrente de arrasto. Mas se uma tensão externa for aplicada, o campo elétrico na região de depleção, que é responsável pela corrente de arrasto, diminuirá ou aumentará, conforme a polarização, desequilibrando assim o balanço de correntes.

Como mostrado na Figura 25a, se a junção for polarizada diretamente, então o campo elétrico externo (advindo da tensão externa) será contrário ao campo elétrico interno da região de depleção. Dessa forma, o campo elétrico total na região de depleção será reduzido, reduzindo assim também o potencial eletrostático e a corrente de arrasto

entre as junções. Portanto, a corrente de arrasto não será mais suficiente para cancelar a corrente de difusão e, conseqüentemente, vai aparecer uma corrente total diferente de zero, advindo desse “excesso” da corrente de difusão. Como a corrente de difusão é de buracos indo de **p** para **n** e de elétrons indo de **n** para **p**, então a corrente de difusão vai de **p** para **n**. Essa corrente de difusão será alta, porque os gradientes de concentração são muito altos (existe uma alta concentração de buracos em **p**, e baixa em **n**; por outro lado, existe uma alta concentração de elétrons em **n**, e de baixa de elétrons em **p**). Portanto, a corrente em polarização direta, para ambos os portadores, tipicamente é alta.

Por outro lado, se a junção for polarizada inversamente, como mostrado na Figura 25b, então tanto o campo elétrico total como o potencial eletrostático aumentarão, aumentando assim a corrente de arrasto. Portanto, vai haver um excesso de corrente de arrasto, resultando em uma corrente total diferente de zero. Mas a corrente de arrasto, neste caso, é de buracos indo de **n** para **p** e de elétrons livres indo de **p** para **n**. Mas existem pouquíssimos buracos na região **n** e pouquíssimos elétrons livres na região **p**. Essa falta de portadores manifesta-se como uma resistência ôhmica altíssima e, conseqüentemente, a corrente de arrasto é tipicamente bem baixa. Essa é a origem física da propriedade retificadora de junções **p-n**: quando buracos vão de **p** para **n** e elétrons de **n** para **p** (polarização direta), temos uma corrente de difusão alta, porque o gradiente de concentração de portadores é alto, portanto a corrente em polarização direta é alta (resistência baixa). Quando buracos vão de **n** para **p** e elétrons de **p** para **n** (polarização inversa), temos uma corrente de arrasto baixa, porque temos poucos portadores (elétrons em **p** e buracos em **n**) contribuindo para a corrente, portanto a corrente em polarização inversa é baixa (resistência alta). Em outras palavras, a junção **p-n** apresenta resistividade bem diferente dependendo da direção da polarização.

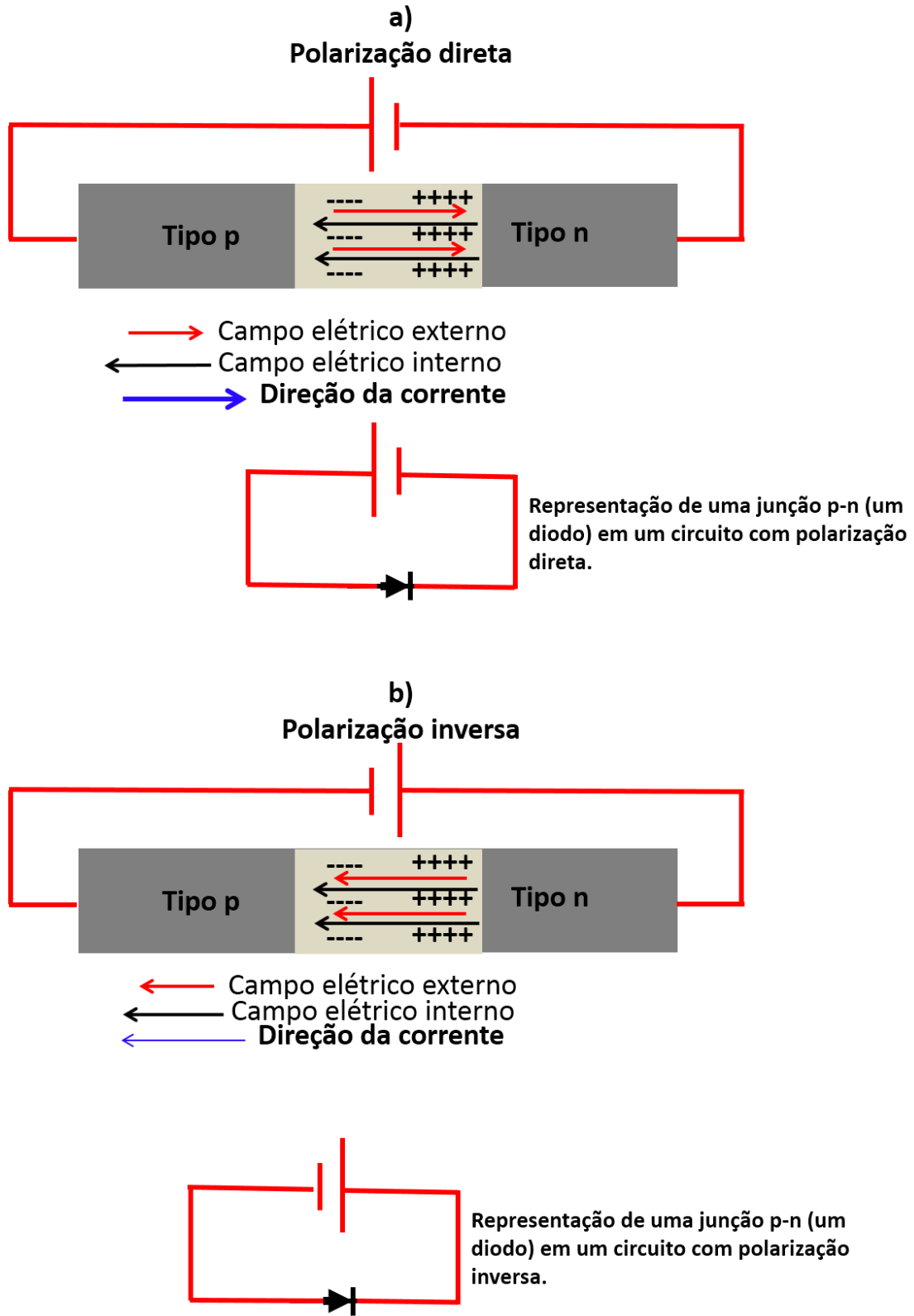


Figura 25 Tipos de polarização da junção p-n. Note que na representação da junção em circuitos, a corrente em polarização direta vai na direção da seta que representa o diodo, enquanto a corrente reversa vai na direção contrária.

Antes de passarmos para a descrição quantitativa da relação entre tensão e corrente na junção **p-n**, temos que entender o que ocorre com os níveis de Fermi quando a junção é polarizada. Vamos focar no caso de polarização direta, mas o raciocínio para polarização inversa é bem semelhante: para passar da descrição de polarização direta para inversa basta inverter o sinal da tensão externa.

A Figura 26 mostra o diagrama de energia para a junção em repouso (sem polarização), para a junção diretamente polarizada e para a junção inversamente polarizada. No caso da polarização direta (figura do meio), a tensão externa, denominada por V_{ext} , reduz o potencial eletrostático, o que faz com que as bandas entortem menos. Além disso, como discutido na seção anterior, quando o sistema sai do equilíbrio, ocorre uma separação entre os níveis de Fermi de elétrons e buracos, que passam, agora, a ser denominado níveis de *quasi-Fermi*. O comportamento dos níveis de *quasi-Fermi* com polarização direta é mostrado na Figura 26 (banda do meio). O nível de *quasi-Fermi* dos elétrons, ε_{QFN} , é constante ao longo de toda a região **n** e, também, ao longo da região de depleção. Ao atingir a região **p**, entretanto, o nível de *quasi-Fermi* cai até atingir o nível de *quasi-Fermi* dos buracos, ε_{QFP} .

Esse comportamento pode ser explicado da seguinte forma: na região **n**, as cargas não “percebem” a perturbação, ou seja, a concentração de elétrons é praticamente inalterada pela aplicação do potencial externo. Isso ocorre exatamente porque a concentração de elétrons já é muito alta nessa região e, portanto, os elétrons que advêm da polarização não alteram significativamente a concentração na parte **n**. Como é a diferença entre a banda de condução ε_C e o nível de *quasi-Fermi* ε_{QFN} que define a concentração de elétrons (Equação 216), essa diferença tem que ser virtualmente idêntica à diferença antes da aplicação da tensão externa (no equilíbrio, topo da Figura 26). Em termos matemáticos, na região **n**, a diferença $\varepsilon_C - \varepsilon_{QFN}$ é praticamente idêntica à diferença $\varepsilon_C - \varepsilon_F$ (ou seja, $\varepsilon_C - \varepsilon_{QFN} \approx \varepsilon_C - \varepsilon_F$). Na região **p**, por outro lado, a concentração de elétrons é baixa. Quando a tensão externa for aplicada, haverá uma injeção de elétrons nessa região (elétrons são injetados pela região de depleção: são os elétrons de **n** que difundiram para **p**; lembre-se que agora eles podem difundir porque a barreira de potencial foi reduzida) e, portanto, a variação na concentração de elétrons é proporcionalmente muito grande. Esse aumento drástico na concentração de elétrons resulta em $\varepsilon_C - \varepsilon_{QFN} \ll \varepsilon_C - \varepsilon_F$ na região **p** (note pela Equação 216 que, se a diferença entre as energias diminui, então a concentração de elétrons n aumenta). Além disso,

estes elétrons em excesso, que adentram a região **p**, recombinarão com os buracos, causando assim uma forte queda na concentração de elétrons ao longo da região **p**. Conseqüentemente, a diferença $\varepsilon_C - \varepsilon_{QFN}$ aumenta ao longo da região **p**, o que explica a queda em ε_{QFN} . É claro que, como consequência do processo de recombinação, quanto mais longe da região de injeção de cargas (que é a fronteira entre a região **p** e a região de depleção), menor será a concentração desse excesso de elétrons. Em outras palavras, na região **p**, longe da região de depleção, o nível de *quasi*-Fermi deve cair para o nível em equilíbrio $\varepsilon_C - \varepsilon_{QFN} \approx \varepsilon_C - \varepsilon_F$).

Assim, a condição $\varepsilon_C - \varepsilon_{QFN} \approx \varepsilon_C - \varepsilon_F$ é satisfeita em toda a região **n** (porque a concentração de elétrons é alta) e também na região **p** longe da região de depleção (porque o processo de recombinação não deixa os elétrons chegarem).

O raciocínio para explicar o comportamento do nível de *quasi*-Fermi para buracos em polarização direta é análogo: na região **p**, a concentração de buracos é alta e, conseqüentemente, pouco alterada pela polarização. Portanto, na região **p** temos $\varepsilon_{QFP} - \varepsilon_V \approx \varepsilon_F - \varepsilon_V$, como mostrado na Figura 26. Além disso, essa diferença aumenta ao longo da região **n** em razão do processo de recombinação de buracos.

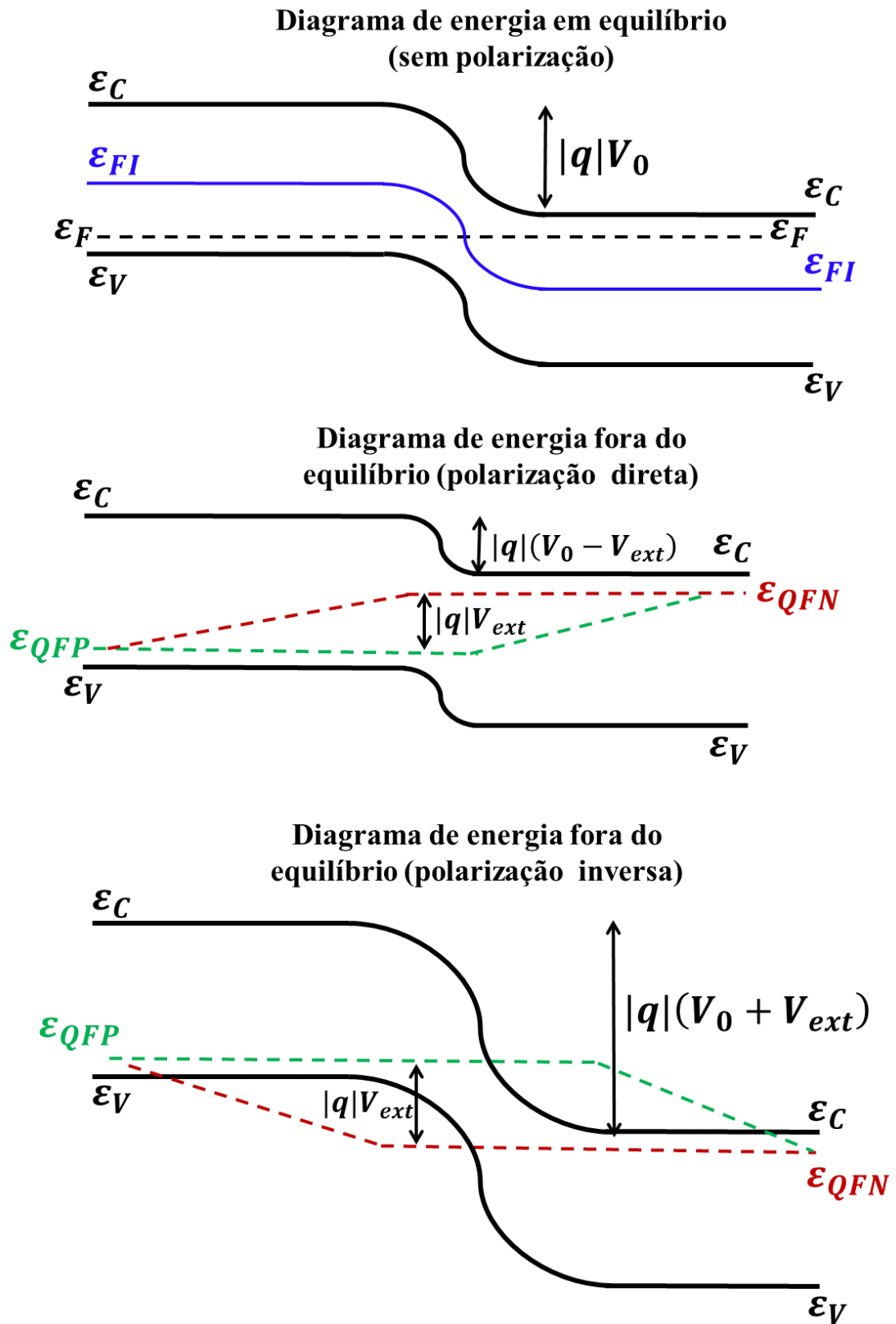


Figura 26 Diagramas de energia dentro e fora do equilíbrio

Note, na Figura 26, que o potencial eletrostático total para polarização direta agora é dado pelo potencial inicial V_0 menos o potencial externo V_{ext} , como esperado. Note

também que, como é a tensão externa que fornece energia para separar os níveis de *quasi-Fermi*, temos que:

$$|q|V_{ext} = \varepsilon_{QFN} - \varepsilon_{QFP} \quad \text{para} \quad -x_{p0} \leq x \leq x_{n0}$$

Equação 218

Se a relação acima ficou obscura para você, pense o seguinte: o que significa dizer que tem uma diferença de potencial V_{ext} entre os terminais da junção? Isso quer dizer que a energia potencial total dos elétrons de um lado da junção é maior do que a do outro lado por um tanto $|q|V_{ext}$ (similar à interpretação da queda de tensão em um resistor). Mas o potencial total é o potencial eletroquímico, ou seja, é o nível de Fermi. Portanto, a afirmação “existe uma diferença de potencial V_{ext} ” na junção é idêntica à afirmação “o nível de *quasi-Fermi* em um lado da junção é maior que do outro lado por um tanto $|q|V_{ext}$ ”. Se você coloca um voltímetro nos terminais da junção e o voltímetro lê V_{ext} , o que você está medindo na verdade é a diferença de potencial eletroquímico, ou seja, a diferença entre os níveis de *quasi-Fermi*.

O comportamento dos níveis de *quasi-Fermi* e das concentrações de portadores motiva a classificação dos portadores em portadores minoritários e portadores majoritários. Os elétrons livres na região **n** são portadores majoritários e os elétrons livres na região **p** são portadores minoritários. Por outro lado, buracos na região **n** são portadores minoritários e buracos na região **p** são portadores majoritários. A corrente total é formada pela contribuição tanto de portadores majoritários como de portadores minoritários.

Um exemplo da distribuição de corrente é mostrado qualitativamente na Figura 27. Lembre-se que a corrente total ao longo da junção deve ser constante. Na região **n**, longe da região de depleção, praticamente não há buracos; portanto, a corrente é predominantemente de portadores majoritários (nesse caso, elétrons). Na região de depleção as contribuições das correntes de elétrons e buracos são da mesma ordem de magnitude (não faz sentido falar de portadores majoritários e minoritários na região de depleção). Já na região **p**, longe da região de depleção, a corrente é novamente predominantemente de portadores majoritários (que, agora, são os buracos). Note que a corrente ao longo da região de depleção é praticamente constante. Isso ocorre porque, em geral, as taxas de recombinação na região de depleção são muito baixas e podem

ser negligenciadas. Essa também é uma das razões pelas quais a queda dos níveis de *quasi-Fermi* na região de depleção pode ser negligenciada.

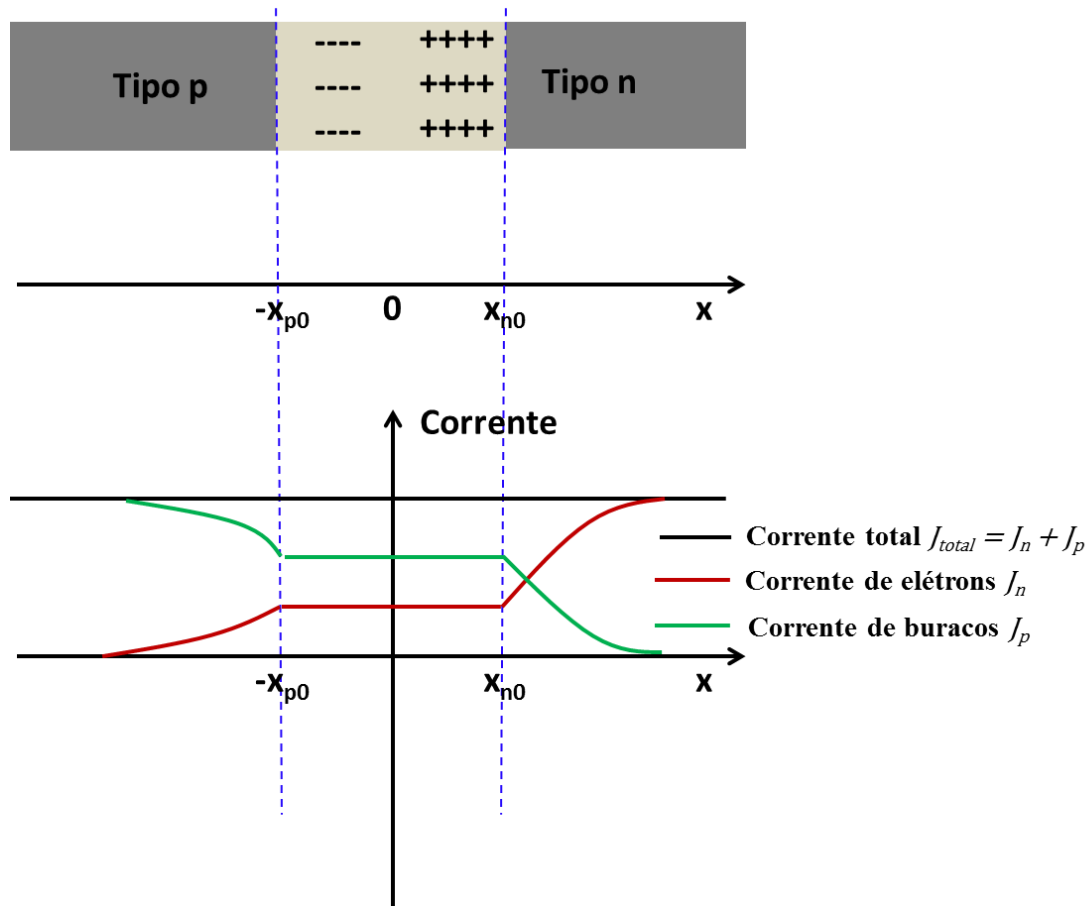


Figura 27 Contribuições das correntes de elétrons e buracos para a corrente total

Existem várias maneiras de se calcular a corrente total na junção **p-n**, já que a princípio podemos calcular a corrente em qualquer ponto da junção (já que a corrente total é constante). Então, podemos escolher o ponto onde é mais fácil calcular a corrente. Para isso, vamos fazer uma breve análise semi-quantitativa das propriedades das correntes ao longo da junção.

Vamos, primeiro, focar nossa atenção na região **n** da junção **p-n** (no eixo da Figura 27, a região $x \geq x_{n0}$). Já sabemos que temos dois portadores, minoritários e majoritários, e que, na região **n**, os portadores majoritários são elétrons livres e os minoritários são buracos. Quais são os tipos de correntes que temos nos dois casos? Para responder a essa pergunta, note que o campo elétrico está todo confinado na região de depleção. Isso é, na verdade, uma aproximação, mas é uma aproximação muito boa, exatamente porque existe uma concentração alta de portadores majoritários fora da região de depleção. Essa concentração alta de portadores majoritários faz com que o campo elétrico seja baixo fora da região de depleção, pela mesma razão que, quando

fazemos análise de circuitos, consideramos o campo elétrico em um fio metálico como nulo. Para enxergar isso de maneira mais clara, basta analisar a lei de Ohm para os portadores majoritários. A forma mais comum da lei de Ohm expressa a corrente em termos da condutividade do material. Assim, a corrente de arrasto de elétrons livres na região **n** pode ser expressa como:

$$J_{arr_n} = \sigma_n^{\text{região n}} \cdot E \quad \text{para } x \geq x_{n0}$$

Equação 219

Onde $\sigma_n^{\text{região n}}$ é a condutividade de elétrons livres na região **n**. Note que coloquei o subíndice *arr_n* na corrente porque a lei de Ohm se refere à corrente de arrasto; e também porque estamos tratando, nesse momento, da corrente de elétrons livres (portadores majoritários na região **n**). Pela lei de Ohm, fica claro porque o campo elétrico é muito pequeno na região **n**: como a concentração de portadores majoritários (elétrons livres) é alta, a condutividade também é alta, o que significa que o campo elétrico é baixo (como J_{arr_n} é finito, se $\sigma_n^{\text{região n}}$ vai para infinito, então E tem que ir para zero). Além da corrente de arrasto de portadores majoritários, temos também a corrente de difusão:

$$J_{dif_n} = |q|D_N \frac{\partial n}{\partial x} \quad \text{para } x \geq x_{n0}$$

Equação 220

Considerações semelhantes podem ser feitas para os portadores majoritários na região **p** (que são buracos). Assim, o transporte de portadores majoritários é formado tanto por corrente de arrasto como por corrente de difusão.

E o transporte de portadores minoritários? Nesse caso, o transporte é dominado pela corrente de difusão. E isso ocorre porque o campo elétrico é baixo. De fato, a corrente de arrasto de portadores minoritários na região **n** (buracos na região **n**) é:

$$J_{arr_p} = \sigma_p^{\text{região n}} \cdot E \quad \text{para } x \geq x_{n0}$$

Equação 221

Mas, como estamos tratando de portadores minoritários, temos uma concentração muito baixa de portadores. Conseqüentemente, a condutividade para buracos na região **n**, $\sigma_p^{\text{região n}}$, é baixa. Como tanto $\sigma_p^{\text{região n}}$ como E são baixos, então J_{arr_p} na região **n** é desprezível. Portanto:

$$J_{arr_p} = \sigma_p^{\text{região n}} \cdot E \approx 0 \quad \text{para } x \geq x_{n0}$$

Equação 222

Resumindo: o E da Equação 221 é o mesmo E da Equação 219, e ambos são muito pequenos; além disso, na região **n**, J_{arr_p} é desprezível porque $\sigma_p^{\text{região n}}$ é pequeno, enquanto J_{arr_n} não é desprezível porque $\sigma_n^{\text{região n}}$ é grande.

Como considerações análogas podem ser feitas para os portadores minoritários na parte **p** (elétrons livres na parte **p**), temos também que:

$$J_{arr_n} = \sigma_n^{\text{região p}} \cdot E \approx 0 \quad \text{para } x \leq -x_{p0}$$

Equação 223

Note que o J_{arr_n} da Equação 223 não é o mesmo J_{arr_n} da Equação 219, pois eles se referem a regiões diferentes e, conseqüentemente, a condutividades diferentes (mas ambos são correntes de arrasto de elétrons livres).

Para facilitar, vamos então colocar todas essas informações juntas:

Região n ($x \geq x_{n0}$)

$$J_n(x) = J_{arr_n} + J_{dif_n} = \sigma_n^{\text{região n}} \cdot E + |q|D_N \frac{\partial n}{\partial x} \quad (1)$$

$$J_p(x) = J_{dif_p} = -|q|D_P \frac{\partial p}{\partial x} \quad (2)$$

Equação 224

Região p ($x \leq -x_{p0}$)

$$J_n(x) = J_{dif_n} = |q|D_N \frac{\partial n}{\partial x} \quad (1)$$

$$J_p(x) = J_{arr_p} + J_{dif_p} = \sigma_p^{\text{região p}} \cdot E - |q|D_P \frac{\partial p}{\partial x} \quad (2)$$

Equação 225

Como, para portadores minoritários, só temos uma forma de corrente (corrente de difusão), é mais fácil calcular a corrente para portadores minoritários. O problema é que nenhuma região possui apenas portadores minoritários. Assim, se escolhermos fazer o cálculo na região **n**, teremos que fazê-lo tanto para elétrons livres como para buracos, mas só a de buracos é que vai ser fácil. Por outro lado, se escolhermos fazer o cálculo na região **p**, teremos que fazê-lo, novamente, tanto para elétrons livres como para buracos, mas só a de elétrons livres é que vai ser fácil. Dançamos? Na verdade, não, porque podemos tirar vantagem do fato de que as correntes dentro da região de depleção são constantes. Por isso, podemos fazer o cálculo da corrente de portadores minoritários em cima do ponto x_{n0} e em cima do ponto $-x_{p0}$ e somar as duas. Note que isso só é possível porque as correntes são constantes ao longo da região de depleção. Assim, a corrente total pode ser calculada como:

$$J_{total} = J_p(x_{n0}) + J_n(-x_{p0})$$

Equação 226

Dessa forma, para obter a corrente total, basta calcular a corrente de portadores minoritários na região **n**, a corrente de portadores minoritários na região **p**, e avaliá-las nas interfaces entre essas regiões e a região de depleção (as interfaces são os pontos x_{n0} e $-x_{p0}$, respectivamente).

Agora, nossa vida ficou bem mais fácil, porque já estudamos (na seção 1.10) a situação em que temos transporte dominado por corrente de difusão. Vimos que, nesse caso, a concentração de portadores decai exponencialmente devido ao processo de recombinação (Equação 110 e Equação 111 para buracos e elétrons, respectivamente). Para facilitar, vou repetir a relação para buracos (Equação 110) aqui embaixo (apenas trocando a dependência em z por x , já que agora adotamos o eixo x):

$$p(x) = p(0) \cdot \exp\left(-\frac{x}{L_p}\right)$$

Podemos já utilizar essa relação para encontrarmos J_p na Equação 224? Quase. Precisamos, primeiro, levar em conta um detalhe importante: o processo de

recombinação faz referência apenas ao excesso de portadores (excesso aqui é definido em relação à concentração em equilíbrio). São os portadores “extras”, que aparecem como consequência da tensão externa, que sofrem o processo de recombinação. Os portadores em equilíbrio, por outro lado, exatamente por estarem em equilíbrio, possuem uma taxa líquida de recombinação igual à zero. Isso ocorre porque a recombinação sofrida pelos portadores em equilíbrio é compensada pela excitação térmica. Assim, chamando a concentração de buracos no equilíbrio de p_0 , o “excesso” de buracos de p_e e a concentração total de p , temos:

$$p = p_0 + p_e$$

Equação 227

Então, temos que o excesso de portadores obedecerá à relação da Equação 110, ou seja:

$$p_e(x) = p_e(x_{n0}) \cdot \exp\left(-\frac{x - x_{n0}}{L_p}\right) \quad \text{para } x \geq x_{n0}$$

Equação 228

Onde já adaptamos as condições de contorno para o eixo da Figura 27. Considerações análogas levam à concentração de portadores minoritários na região **p**:

$$n_e(x) = n_e(-x_{p0}) \cdot \exp\left(\frac{x + x_{p0}}{L_N}\right) \quad \text{para } x \leq -x_{p0}$$

Equação 229

Onde n_e é o excesso de elétrons livres. E pronto: para achar a corrente total, basta inserir a Equação 228 na Equação 224; a Equação 229 na Equação 225, e depois substituir os resultados na Equação 226. Esse é o caminho das pedras para calcular a corrente total na junção.

Mas, antes de fazer calcular a corrente total, vou mostrar com um pouco mais de rigor porque que é o excesso de portadores que decai exponencialmente. Vou fazer isso só para buracos, porque o raciocínio para elétrons é praticamente idêntico.

Para chegar na Equação 228 rigorosamente, basta utilizarmos a relação para conservação de portadores, que é análoga à Equação 103, mas incorporando a taxa de

excitação térmica, que aqui chamaremos de geração térmica. Assim, a relação para conservação de buracos fica:

$$-\frac{\partial J}{\partial x} - R + G = |q| \frac{\partial p}{\partial t}$$

Equação 230

Onde R e G são as taxas de recombinação e geração, respectivamente. Já sabemos que a corrente é de difusão, então:

$$-|q|D_p \frac{\partial^2 p}{\partial x^2} - R + G = |q| \frac{\partial p}{\partial t}$$

Equação 231

Onde D_p é o coeficiente de difusão para buracos. Em regime estacionário a derivada temporal é zero e, portanto:

$$|q|D_p \frac{\partial^2 p}{\partial x^2} = R - G$$

Equação 232

Lembre-se também da seção 1.10 que a taxa de recombinação é proporcional à concentração de portadores. Assim, chamando o tempo de vida para buracos de τ_p , temos também que:

$$R = |q| \frac{p}{\tau_p}$$

Agora, vamos supor que estamos na situação de equilíbrio. Neste caso, teríamos:

$$|q|D_p \frac{\partial^2 p_0}{\partial x^2} = \frac{p_0}{\tau_p} - G$$

Equação 233

Mas, como no equilíbrio não existe corrente de difusão, a concentração de portadores tem que ser constante ao longo do material, portanto a derivada espacial (de primeira e segunda ordens) é zero, ou seja,

$$\frac{\partial^2 p_0}{\partial x^2} = 0$$

E, portanto, a taxa de geração térmica é igual à:

$$G = \frac{p_0}{\tau_P}$$

Equação 234

Assim, voltando para a situação fora do equilíbrio, temos:

$$|q|D_p \frac{\partial^2 p}{\partial x^2} = R - G = |q| \frac{p}{\tau_P} - |q| \frac{p_0}{\tau_P}$$

Equação 235

Expressando a concentração total em termos da concentração em equilíbrio, como na Equação 227, temos:

$$|q|D_p \frac{\partial^2 (p_e + p_0)}{\partial x^2} = |q| \frac{p_e + p_0}{\tau_P} - |q| \frac{p_0}{\tau_P}$$

Como:

$$\frac{\partial^2 p_0}{\partial x^2} = 0$$

Temos, portanto:

$$D_p \frac{\partial^2 p_e}{\partial x^2} = \frac{p_e}{\tau_P}$$

Equação 236

Essa relação é idêntica à que encontramos na seção 1.10 (Equação 107) e a solução, para o sistema de coordenadas da Figura 27, é a Equação 228.

Vamos, então, finalizar o cálculo. Lembrando: para achar a corrente total, basta inserir a Equação 228 na Equação 224; a Equação 229 na Equação 225, e depois substituir os resultados na Equação 226.

Primeiro, inserindo a Equação 228 na Equação 224:

$$J_p(x) = J_{dif_p}(x) = -|q|D_P \frac{\partial(p_0 + p_e)}{\partial x} = -|q|D_P \frac{\partial p_e}{\partial x}$$

$$\therefore J_p(x) = \frac{|q|D_P}{L_P} \cdot p_e(x_{n0}) \cdot \exp\left(-\frac{x - x_{n0}}{L_P}\right) \quad \text{para } x \geq x_{n0}$$

Equação 237

Agora, inserindo a Equação 229 na Equação 225:

$$J_n(x) = J_{dif_n}(x) = |q|D_N \frac{\partial(n_0 + n_e)}{\partial x} = |q|D_N \frac{\partial n_e}{\partial x}$$

$$\therefore J_n(x) = \frac{|q|D_N}{L_N} \cdot n_e(-x_{p0}) \cdot \exp\left(\frac{x + x_{p0}}{L_N}\right) \quad \text{para } x \leq -x_{p0}$$

Equação 238

Finalmente, substituindo a Equação 237 e Equação 238 na Equação 226:

$$J_{total} = J_p(x_{n0}) + J_n(-x_{p0}) = \frac{|q|D_P}{L_P} \cdot p_e(x_{n0}) + \frac{|q|D_N}{L_N} \cdot n_e(-x_{p0})$$

Equação 239

A princípio terminamos nosso trabalho. Mas ele ainda não está em uma forma satisfatória, porque, em eletrônica, estamos geralmente interessados na relação entre a corrente e a tensão externa. Portanto, precisamos reexpressar o lado direito da Equação 239 em termos da tensão externa. Para isso, basta utilizarmos a Equação 216 e a Equação 217.

Vamos primeiramente expressar $n_e(-x_{p0})$ em termos da tensão externa. Da Equação 216, temos que:

$$n(-x_{p0}) = n_i \exp\left[-\frac{\varepsilon_{FI}(-x_{p0}) - \varepsilon_{QFN}(-x_{p0})}{k_B T}\right]$$

Equação 240

Por outro lado, da Equação 217, temos:

$$p(-x_{p0}) = p_i \exp \left[-\frac{\varepsilon_{QFP}(-x_{p0}) - \varepsilon_{FI}(-x_{p0})}{k_B T} \right]$$

Equação 241

Multiplicando a concentração de elétrons e buracos no ponto $(-x_{p0})$:

$$n(-x_{p0}) \cdot p(-x_{p0}) = n_i \cdot p_i \exp \left[\frac{\varepsilon_{QFN}(-x_{p0}) - \varepsilon_{QFP}(-x_{p0})}{k_B T} \right]$$

Equação 242

Substituindo a Equação 218 na Equação 242, temos:

$$n(-x_{p0}) \cdot p(-x_{p0}) = n_i \cdot p_i \exp \left(\frac{|q|V_{ext}}{k_B T} \right)$$

Equação 243

Podemos reexpressar esse resultado em termos das concentrações em equilíbrio. Como o produto $n \cdot p$ é inalterado pela dopagem, temos que $n_i \cdot p_i = n_0 \cdot p_0$, onde o lado direito se refere ao produto $n \cdot p$ no equilíbrio. Portanto, a Equação 243 pode ser reexpressa como:

$$n(-x_{p0}) \cdot p(-x_{p0}) = n_0(-x_{p0}) \cdot p_0(-x_{p0}) \exp \left(\frac{|q|V_{ext}}{k_B T} \right)$$

Equação 244

Note também que $p(-x_{p0})$ é a concentração de buracos na região **p**, ou seja, é a concentração de portadores majoritários na região **p**. Mas a concentração de portadores majoritários é pouco alterada pela aplicação da tensão externa (em outras palavras, a distância entre $\varepsilon_{QFP}(-x_{p0})$ e $\varepsilon_{FI}(-x_{p0})$ ou $\varepsilon_V(-x_{p0})$ para a junção polarizada é praticamente igual à distância entre ε_F e $\varepsilon_{FI}(-x_{p0})$ ou $\varepsilon_V(-x_{p0})$ para a junção em equilíbrio – compare a Figura 26 do meio com a Figura 26 de cima). Assim, a

concentração de portadores majoritários da junção fora do equilíbrio é praticamente igual à concentração de portadores majoritários da junção no equilíbrio, ou seja:

$$p(-x_{p0}) \approx p_0(-x_{p0})$$

Equação 245

Portanto, a Equação 244 fica:

$$n(-x_{p0}) = n_0(-x_{p0}) \exp\left(\frac{|q|V_{ext}}{k_B T}\right)$$

Equação 246

Finalmente, como $n(-x_{p0}) = n_e(-x_{p0}) + n_0(-x_{p0})$, temos que:

$$\begin{aligned} n_e(-x_{p0}) &= n(-x_{p0}) - n_0(-x_{p0}) \\ \therefore n_e(-x_{p0}) &= n_0(-x_{p0}) \left[\exp\left(\frac{|q|V_{ext}}{k_B T}\right) - 1 \right] \end{aligned}$$

Equação 247

Lembre-se que $n_0(-x_{p0})$ é a concentração de elétrons na região **p** quando a junção está em equilíbrio, e que aprendemos a calcular essa concentração no Capítulo 2 (basta utilizar a Equação 177 com o nível de Fermi do semiconductor tipo **p**, ou seja, com $\varepsilon_F = \varepsilon_{FP}$, onde ε_{FP} é mostrado na Figura 19). Um procedimento análogo (tente sozinho) leva à:

$$p_e(x_{n0}) = p_0(x_{n0}) \left[\exp\left(\frac{|q|V_{ext}}{k_B T}\right) - 1 \right]$$

Equação 248

Substituindo a Equação 247 e a Equação 248 na Equação 239, temos, finalmente, que:

$$J_{total} = \frac{|q|D_P}{L_P} \cdot p_e(x_{n0}) + \frac{|q|D_N}{L_N} \cdot n_e(-x_{p0})$$

$$\therefore J_{total} = |q| \left[\frac{D_P}{L_P} p_0(x_{n0}) + \frac{D_N}{L_N} n_0(-x_{p0}) \right] \left[\exp\left(\frac{|q|V_{ext}}{k_B T}\right) - 1 \right]$$

Equação 249

Se, em vez de densidade de corrente, quisermos a corrente mesmo, basta multiplicar a densidade de corrente pela área transversal da junção. Assim:

$$I = A|q| \left[\frac{D_P}{L_P} p_0(x_{n0}) + \frac{D_N}{L_N} n_0(-x_{p0}) \right] \left[\exp\left(\frac{|q|V_{ext}}{k_B T}\right) - 1 \right]$$

Equação 250

O termo entre parênteses depende de parâmetros do semiconductor, por meio de D_N , D_P , L_N e L_P , e também de parâmetros de dopagem, por meio de $p_0(x_{n0})$ e $n_0(-x_{p0})$. Esse termo é chamado de corrente de saturação e denominado pelo símbolo I_0 (ou também I_s):

$$I_0 = A|q| \left[\frac{D_P}{L_P} p_0(x_{n0}) + \frac{D_N}{L_N} n_0(-x_{p0}) \right]$$

Equação 251

Assim, temos que:

$$I = I_0 \left[\exp\left(\frac{|q|V_{ext}}{k_B T}\right) - 1 \right]$$

Equação 252

A Equação 252 é a importantíssima equação de Shockley, deduzida pelo incansável e ubíquo William B. Shockley Jr., provavelmente o sujeito que mais contribuiu para o

desenvolvimento da eletrônica na história. Um engenheiro electricista não saber quem é William B. Shockley Jr. é tão ruim quanto um músico não saber quem é Mozart, ou um político não saber quem foi Churchill, ou um filósofo nunca ter ouvido falar de Aristóteles, etc...

A equação de Shockley expressa a relação entre corrente I e tensão V_{ext} em uma junção **p-n**. Apesar de termos deduzido essa relação para polarização direta, nenhum passo da dedução requer que este seja o caso. Portanto, essa relação vale tanto para polarização direta ($V_{ext} > 0$) quanto para polarização inversa ($V_{ext} < 0$). Note que, para polarização inversa com $|V_{ext}| < k_B T / |q|$, a exponencial passa a ser bem menor que 1 e, portanto, $I \sim I_0$. Como I_0 é tipicamente bem pequeno (da ordem de 10^{-9} A para diodos de silício), a junção praticamente não conduz para polarização inversa. Para polarização direta, por outro lado, a corrente aumenta exponencialmente com a tensão. Procure na Internet um gráfico da relação corrente versus tensão em um diodo como ilustração do comportamento da equação de Shockley.

Como afirmado no Capítulo 1, o objetivo principal deste curso é compreender a operação da junção **p-n**. Por problemas de espaço e tempo, foi necessário omitir o cálculo de alguns parâmetros como, por exemplo, a capacitância de junções **p-n**. Os tópicos incluídos foram escolhidos não só pela importância, mas também porque servem de base para a compreensão dos demais tópicos. Por exemplo, acredito que o aluno que conseguiu entender a lógica por trás do cálculo da largura da região de depleção não terá dificuldades em entender a lógica por trás do cálculo da capacitância.

Nesse curso, você aprendeu as noções que levam à equação de Shockley. Você deve, então, ser capaz de responder à pergunta: “Por que, em um diodo, a corrente apresenta uma dependência exponencial em relação à tensão?” (lembre-se que diodos nada mais são que junções **p-n**). Se você rebobinar a fita, vai perceber que a origem da dependência exponencial é a distribuição de Fermi-Dirac. Você deve ser capaz também de responder à pergunta: “Por que o potencial eletrostático interno de um diodo não gera corrente externa e não pode ser medido com um voltímetro colocado em seus terminais?” Você deve lembrar que essa pergunta só pode ser respondida utilizando o conceito de potencial eletroquímico. Portanto, não é possível compreender de maneira adequada o funcionamento de um diodo sem entender os conceitos de potencial eletroquímico e a distribuição de Fermi-Dirac.

Para finalizar este capítulo, algumas palavras sobre o que vem a seguir. O curso de semicondutores é base para os cursos de eletrônica. Nos cursos de eletrônica, você vai estudar como diodos se comportam em circuitos, como calcular corrente e tensão em circuitos com diodos e por aí vai. Além disso, nos cursos de eletrônica você vai estudar o elemento mais importante em eletrônica: o transistor. Como veremos no capítulo seguinte, transistores nada mais são que duas junções **p-n** concatenadas. Portanto, o entendimento adequado do comportamento de junções **p-n** serve não apenas para o estudo de diodos, mas também serve como base fundamental para o adequado estudo de transistores. Em suma: o aluno que absorveu as principais noções de semicondutores de maneira adequada já caminhou passos essenciais para um sóbrio estudo de eletrônica.

O capítulo seguinte trata de dois importantes exemplos de aplicações dos conhecimentos adquiridos até aqui. O primeiro exemplo lida com os mecanismos de conversão de luz em eletricidade, que é o fundamento de fotodetectores e células solares. O segundo exemplo é uma breve introdução aos transistores.

3.5 Sugestões bibliográficas

Para um primeiro estudo, sugiro o livro *Solid State Electronic Devices*, dos autores Ben G. Streetman e Sanjay K. Banerjee. Para um segundo estudo ou aprofundamento, sugiro o livro *Physics of Semiconductor Devices*, de S. M. Sze.

3.6 Exercícios

Exercício 1

Utilize um argumento convincente para demonstrar que:

$$|q|E = \frac{\partial \varepsilon_{FI}}{\partial x}$$

Onde q é a carga fundamental, E é o campo elétrico na região de depleção da junção **p-n** e ε_{FI} é o nível de Fermi intrínseco.

Exercício 2

Utilizando a condição de equilíbrio em junções **p-n**, prove a relação de Einstein:

$$\frac{D}{\mu} = \frac{k_B T}{|q|}$$

Onde D é o coeficiente de difusão, μ é a mobilidade, k_B é a constante de Boltzmann e q é a carga fundamental. Dica: utilize a condição de equilíbrio de corrente na região de depleção de uma junção **p-n**, a relação para **p** (ou **n**, tanto faz) em função do nível de Fermi intrínseco e utilize a relação do Exercício 1. Se você não conseguir sozinho consulte o capítulo 4 do livro do Ben G. Streetman (*Solid State Electronic Devices, Sixth Edition*).

Exercício 3

Expresse a condutividade de um material em termos da concentração de portadores e da mobilidade.

Exercício 4

Encontre uma expressão para a resistência de um diodo quando a condição $V_{ext} \gg \frac{k_B T}{|q|}$ for satisfeita. A resistência, nessas condições, será alta ou baixa? Repita o problema para o caso onde $V_{ext} \ll \frac{k_B T}{|q|}$.

Exercício 5

Explique qualitativamente a origem física da propriedade de retificação de junções **p-n**.

Exercício 6

Rascunhe um gráfico dos diagramas de energia da junção **p-n** para o caso da junção em equilíbrio (sem polarização externa). Repita o exercício para o caso de polarização direta. Compare os dois casos e explique as diferenças.

Exercício 7

Partindo das expressões para as concentrações de portadores em termos dos níveis de Fermi e/ou quase-Fermi, deduza a equação de Shockley.

Exercício 8

Explique o significado físico de cada um dos termos que aparecem na definição de I_0 .

Exercício 9

A equação de Shockley é expressa em termos de parâmetros da junção **p-n** em equilíbrio, a saber: $p_0(x_{n0})$ e $n_0(-x_{p0})$. Lembre-se que $p_0(x_{n0})$ nada mais é do que a concentração de buracos no lado **n** da junção em equilíbrio, enquanto $n_0(-x_{p0})$ é a concentração de elétrons no lado **p** da junção em equilíbrio. Sendo assim, tanto $p_0(x_{n0})$ como $n_0(-x_{p0})$ são as concentrações de portadores minoritários. Expresse essas concentrações explicitamente em termos do *band-gap* do material, da temperatura e da concentração de dopantes em cada lado da junção.

Exercício 10

Sabemos, de eletromagnetismo, que a diferença de potencial eletrostático V entre dois pontos é a integral de linha do campo elétrico entre esses dois pontos.

Aqui, você aprendeu que, em uma junção **p-n**, existe um campo elétrico entre o material **p** e o material **n**.

Então, se você juntar o que você aprendeu em eletromagnetismo com o que aprendeu em semicondutores, você deduz que existe uma diferença de potencial entre **p** e **n**.

Mas, se você for a um laboratório, pegar um diodo, pegar um voltímetro e medir a diferença de potencial entre os terminais do diodo (que equivale a medir a diferença de potencial entre **p** e **n**), vai ver que a diferença de potencial é 0.

Explique o paradoxo.

4 – Efeito fotovoltaico (fotodetectores e células solares) e transistores.

4.1 Fotodetectores e células solares

Na seção 3.2, vimos que quando um semicondutor é iluminado, a absorção de luz excita elétrons da banda de valência para a banda de condução, aumentando assim simultaneamente as concentrações de elétrons livres e de buracos. Vimos também que esse aumento simultâneo de portadores é descrito por uma cisão do nível de Fermi em dois níveis distintos, que agora passam a ser chamados de níveis de *quasi*-Fermi. Além disso, vimos que essa cisão do nível de Fermi em níveis de *quasi*-Fermi é a marca de um sistema fora de equilíbrio.

Na seção 3.3, estudamos com detalhes a dinâmica de uma junção **p-n** fora do equilíbrio. Naquela ocasião, o equilíbrio foi perturbado pela aplicação de um potencial eletrostático externo, resultando assim na cisão do nível de Fermi. Vimos que se o potencial for aplicado de maneira a reduzir a barreira de potencial eletrostática, então a junção conduzirá corrente, que é essencialmente a corrente de difusão advinda da diferença de potencial químico entre as partes **p** e **n**.

Então, vamos juntar esses dois fatos. O fato 1 é que a iluminação gera a cisão dos níveis de *quasi*-Fermi. O fato 2 é que uma tensão externa gera a cisão dos níveis de *quasi*-Fermi. Pergunta: a cisão dos níveis de *quasi*-Fermi gerada pela iluminação pode causar uma diferença de potencial nos terminais da junção **p-n**?

Para responder a essa pergunta, vamos primeiro supor que temos um diodo em aberto, ou seja, a junção **p-n** não está conectada a nenhum circuito. E vamos supor também que iluminamos essa junção **p-n**. Essa situação está ilustrada na Figura 28. Agora, já sabemos o que vai ocorrer: o nível de Fermi vai se dividir em níveis de *quasi*-

Fermi, como mostrado na Figura 28b. Além disso, os elétrons livres e os buracos gerados na região de depleção pela iluminação são ambos portadores livres submetidos a um campo elétrico (que é o campo da região de depleção). Esse campo vai, então, arrastar os elétrons para a parte **n** e os buracos para a parte **p**, gerando assim uma corrente interna que vai de **n** para a **p** (ou seja, uma corrente interna reversa, que na Figura 28b chamei de corrente fotoelétrica I_{lum}). Mas a corrente total tem que ser zero, já que o circuito está aberto (o diodo não está conectado a nada). Para que a corrente total seja 0, tem que haver alguma corrente na direção contrária que cancele essa corrente de arrasto. Podemos esperar que exista alguma outra corrente? Claro que sim: a divisão do nível de Fermi implica que o nível de Fermi de elétrons é maior no lado **n** do que no lado **p**, e já sabemos que um desnível no nível de Fermi gera uma corrente. Esse desnível no nível de Fermi causa, então, uma corrente de difusão que vai de **p** para **n**, que é exatamente a mesma corrente que fluiria pelo diodo se a cisão do nível de Fermi fosse provocada por uma polarização direta externa. Assim, as duas correntes se cancelam e a corrente total é zero.

Mas não respondi à pergunta: vai haver uma diferença de potencial nos terminais do diodo? Você já sabe a resposta: é claro que sim. Vai haver uma diferença de potencial porque os níveis de Fermi não são iguais na parte **n** e na parte **p**, devido à iluminação. A frase “O nível de Fermi na parte **n** é diferente do nível de Fermi na parte **p**” é uma frase absolutamente idêntica à frase “Existe uma diferença de potencial entre a parte **n** e a parte **p**”. Por quê? No capítulo 1, você aprendeu o porquê: porque o nível de Fermi é o potencial eletroquímico dos portadores de carga, e o potencial eletroquímico é a energia potencial total dos portadores cargas (químico + eletrostático) e uma diferença de potencial é sempre a diferença de potencial total (química + eletrostática. Só não incluímos a gravitacional porque ela é desprezível, nesse caso). Se você levitar uma bola com um campo elétrico, você estará equilibrando o potencial gravitacional com o potencial eletrostático e, por isso, a bola não cai. É a mesma coisa aqui, só que com o potencial químico no lugar do potencial gravitacional.

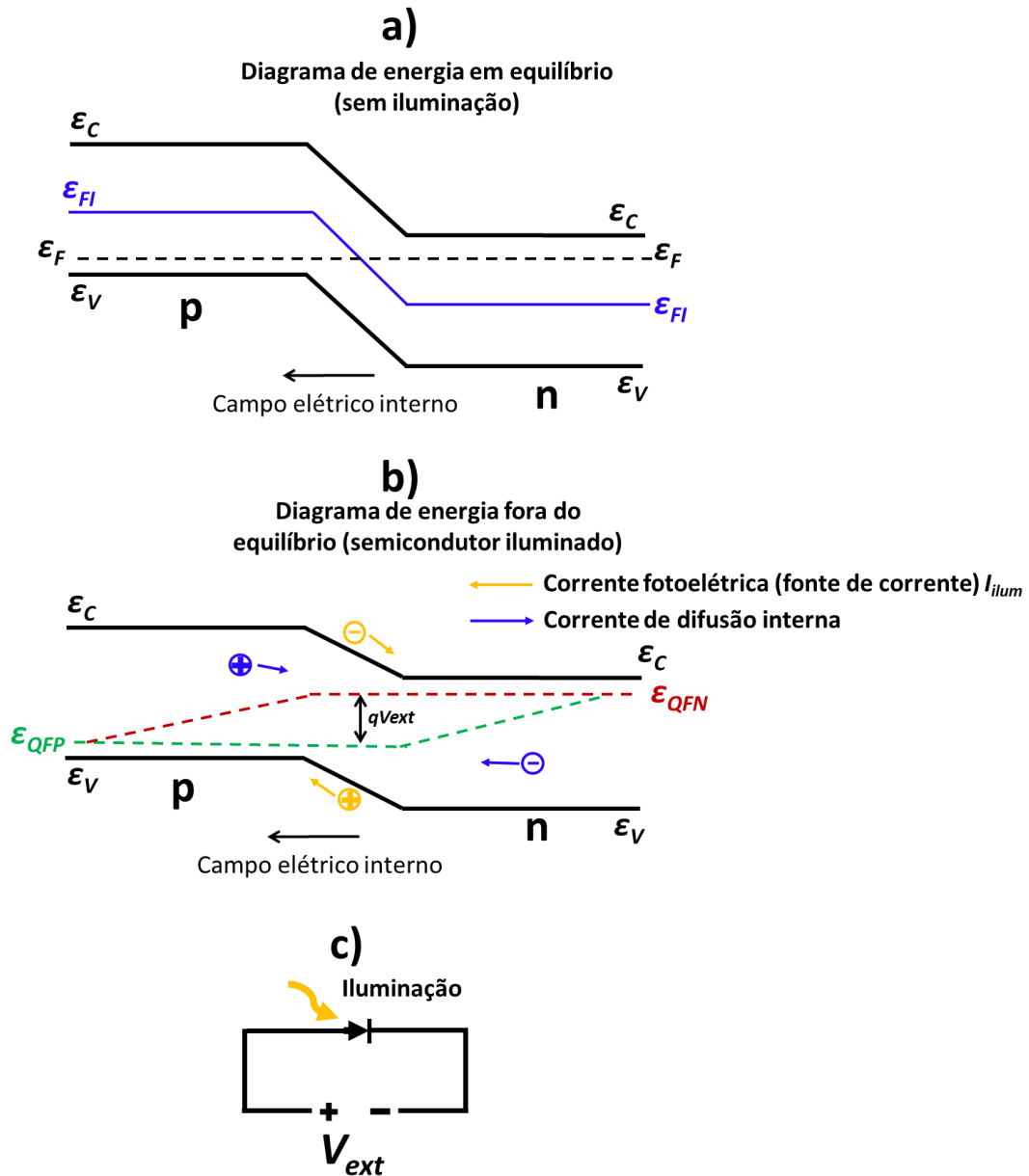


Figura 28 a) Diagrama de energia da junção p-n em equilíbrio. b) Diagrama de energia da junção p-n fora do equilíbrio em razão de iluminação externa. c) representação da junção iluminada em um circuito.

Portanto, aparece sim uma diferença de potencial nos terminais da junção, e essa é exatamente a diferença de potencial que você precisaria colocar externamente para gerar a corrente de difusão, que em nosso sistema iluminado cancela a corrente de arrasto fotoelétrica I_{ilum} . Para melhor exemplificar essa dinâmica, podemos fazer um circuito equivalente, como representado na Figura 29b. Temos uma fonte de corrente I_{ilum} , que é a corrente resultante da geração de portadores por iluminação (representada em amarelo na Figura 28b). Essa corrente passa pelo diodo em polarização direta, e essa corrente que passa pelo diodo é exatamente a corrente de difusão resultante da cisão

dos níveis de *quasi-Fermi*. Note que a Figura 29a é a mesma que a Figura 28c, apenas enfatizando uma questão de nomenclatura: quando a junção está em um circuito aberto, a tensão externa que aparece devido à iluminação é chamada de V_{oc} (do inglês *open circuit voltage*).

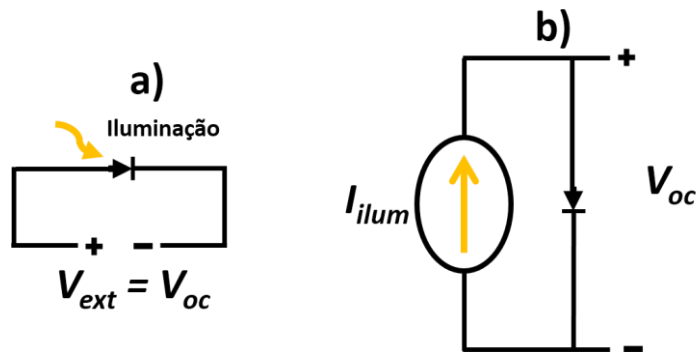


Figura 29 a) representação da junção iluminada em um circuito aberto. b) circuito equivalente

Agora, imagine que você coloque uma resistência alta nos terminais da junção, de maneira que uma pequena corrente passe a circular pelo circuito, como mostrado na Figura 30. Agora, o circuito está fechado e uma pequena corrente circula pelo circuito. Essa corrente é a corrente I_R da Figura 30. Qual é a origem dessa corrente? É claro que a origem dessa corrente é a corrente I_{ilum} . Aliás, lembre-se que I_{ilum} é uma corrente inversa (veja as setas amarelas na Figura 28b) e note que I_R também é uma corrente reversa na Figura 30a. E por que I_R não é igual à I_{ilum} ? Porque parte de I_{ilum} é cancelada pela corrente I_D , que é a corrente de difusão do diodo. No circuito equivalente Figura 30b, expressamos esse cancelamento através da divisão da fonte de corrente I_{ilum} : parte desce pelo diodo (que é a corrente de difusão) e parte vai para o resistor.

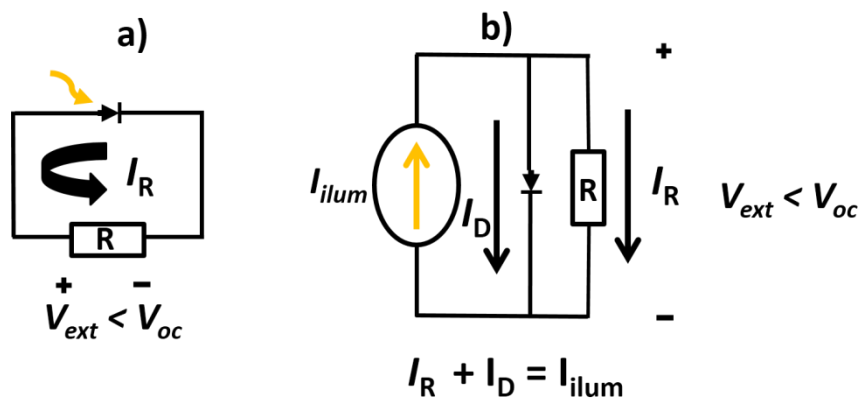


Figura 30 a) Representação do circuito fechado com a junção iluminada. Note que a circulação da corrente é na direção da corrente reversa da junção. b) Circuito equivalente

Mas agora compare a situação do circuito aberto (Figura 29) com a do circuito fechado (Figura 30). No circuito aberto, toda a corrente da fonte de corrente passava pelo diodo ($I_D = I_{lumin}$). Mas no circuito fechado, parte da fonte de corrente é desviada para a resistência e, portanto, agora $I_D < I_{lumin}$. Então, a corrente I_D no circuito aberto é maior do que a corrente I_D no circuito fechado. Mas como a tensão no diodo depende da corrente (e vice-versa), essa queda em I_D acarreta uma queda na tensão de saída. No limite que R vai para zero, toda a corrente é desviada e a tensão passa a ser zero, como mostrado na Figura 31. A corrente que circula agora é chamada de I_{sc} (do inglês *short circuit current*).

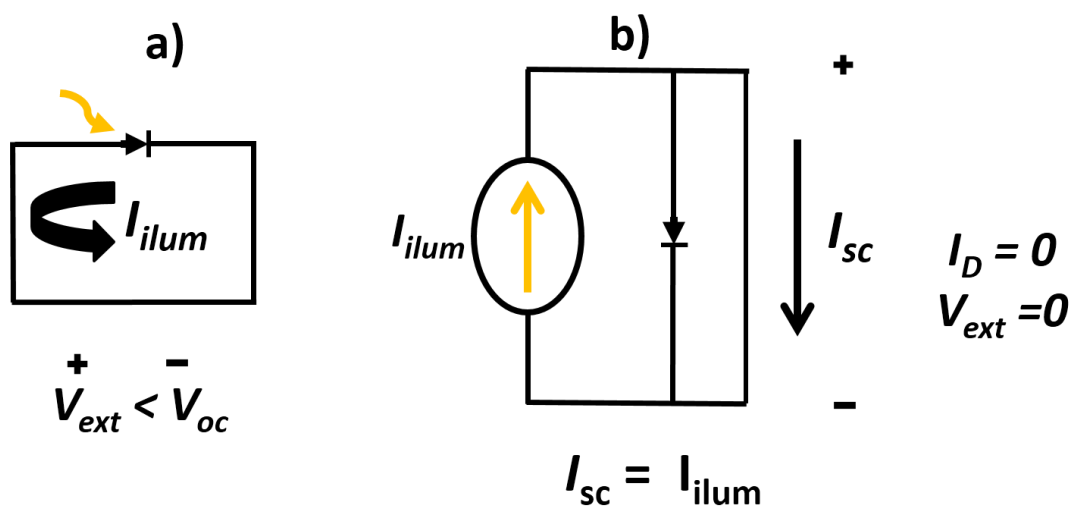


Figura 31 Situação onde a junção é curto circuitada. Nesse caso, a corrente que passa pelo curto é chamada de I_{sc} , do inglês *short circuit current*.

Voltemos para um caso intermediário (Figura 30). Veja a bizarrice: o diodo está polarizado diretamente (a parte **p** tem tensão mais alta que a parte **n**), mas a corrente que circula pelo diodo é inversa!!! Então, temos uma situação onde a tensão é positiva, mas a corrente é negativa, resultando em potência $P = VI$ negativa! Essa é a marca de um sistema que gera energia: a tensão e a corrente estão com polarizações contrárias e a potência “dissipada” pelo sistema é negativa, ou seja, o sistema está gerando energia. Eis a célula solar!

Essa é a essência tanto de células solares como de fotodetectores. Mas o sistema é um pouco mais complicado que o que foi apresentado porque sempre existem perdas ôhmicas (associadas à resistividade do semiconductor) e perdas na corrente gerada pela iluminação (*current leakage*). Um circuito equivalente mais completo, que inclui essas perdas, é mostrado na Figura 32. A resistência em série R_s modela as perdas ôhmicas e

a resistência em paralelo R_{SH} (do inglês *shunt resistance*) modela perdas associadas ao chamado *current leakage*, que é geralmente associado a defeitos de fabricação. A junção ideal possui $R_S = 0$ e $R_{SH} = \infty$.

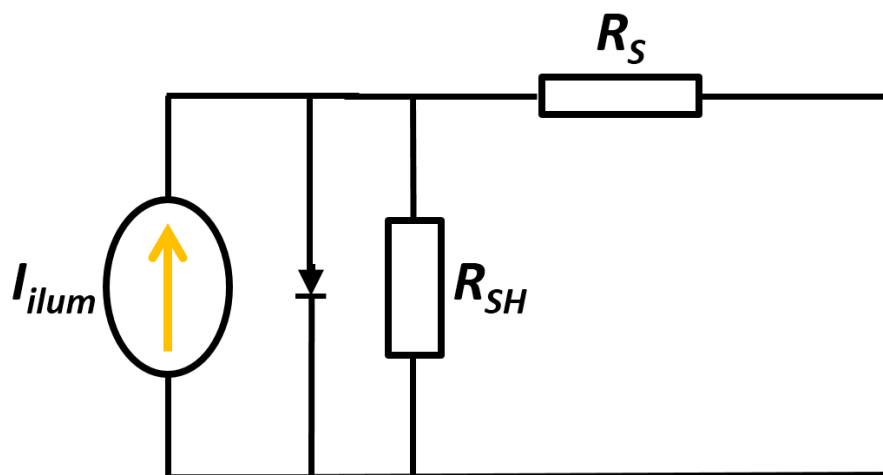


Figura 32 Circuito equivalente levando em consideração perdas ôhmicas e perdas por recombinação.

Para finalizar essa introdução, imagine a seguinte situação. Imagine que você compre duas células solares idênticas, que dão uma certa tensão V . Mas você não quer V , você quer $2V$. Então, você, como bom engenheiro, vai lá e coloca as células em série, como representado na Figura 33a. Você desce do telhado feliz da vida que você é um bom engenheiro e resolveu o problema. Enquanto ambas estão sendo iluminadas, a vida está ótima. Aí passa um passarinho e faz caca em uma das células. A caca, agora, impede que a luz chegue em uma das células, como mostrado Figura 33b. Agora, veja a encrenca que você arrumou: a célula premiada pelo passarinho virou um simples diodo com polarização reversa. E você já sabe que o diodo em polarização reversa possui resistência altíssima. Resultado: a célula sem iluminação vira praticamente uma resistência infinita, o que impede a circulação de corrente e mata seu sistema. Agora, seu chuveiro (que é a resistência R_L , da Figura 33) não esquenta mais. Por isso, colocar painéis solares em série é sempre problemático: se um falhar, leva todos os outros juntos.

Quem gostar do assunto e quiser aprender mais, sugiro o sempre muito útil *website pveducation.org* como ponto de partida.

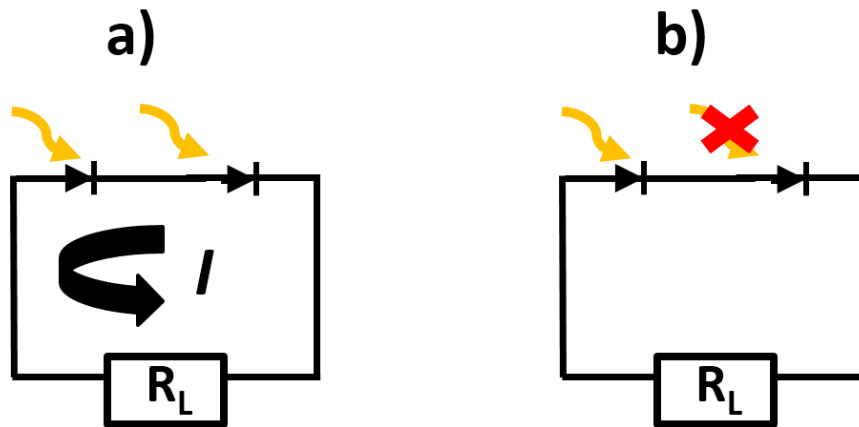


Figura 33 Conexão de painéis solares em série. a) quando ambos estão iluminados, a tensão total é a soma das tensões de cada um. b) quando um deles passa a não ser mais iluminado, ele vira um simples diodo, só que com polarização reversa, o que resulta em uma resistência altíssima do diodo não iluminado, matando a corrente no circuito e impedindo a transferência de energia da célula para a carga.

4.2 Transistores

O último exemplo de aplicações dos conceitos de semicondutores trata de transistores. Existem vários tipos de transistores, mas os principais são concatenações de junções **p-n**.

Um transistor é um dispositivo de três terminais, onde um terminal é utilizado para controlar a condução nos outros dois terminais. Os tipos mais importantes são o MOSFET (*Metal Oxide Semiconductor Field Effect Transistor*) e o BJT (*Bipolar Junction Transistor*). Por razões puramente pedagógicas, vou apresentar o princípio de operação do BJT. Além disso, existem dois tipos de BJT: um formado por uma junção **p-n-p** e o outro formado por uma junção **n-p-n**. A operação de um é análoga à do outro, alterando apenas as polaridades. Por isso, vou tratar apenas do **n-p-n**.

A estrutura de um transistor BJT **n-p-n** é mostrada na Figura 34a. Existem três terminais, chamados de emissor, base e coletor (a razão da nomenclatura ficará clara mais adiante). O terminal da base é conectado na parte **p** da junção e essa parte usualmente é mais estreita que as demais. O transistor é projetado de maneira que a dopagem na parte **p** seja bem menor que a dopagem nas partes **n**. Dessa maneira, o transporte é dominado por elétrons livres (no caso de transistores **p-n-p**, o transporte é predominantemente de buracos).

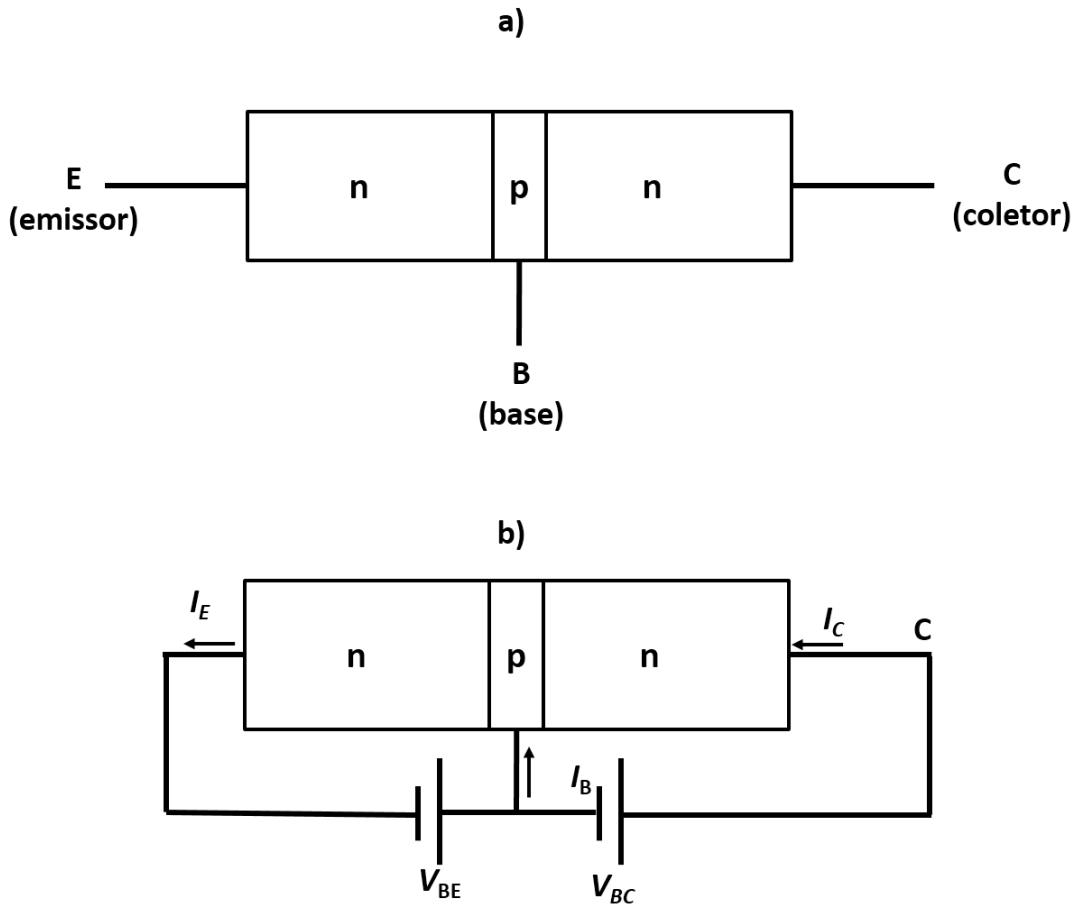


Figura 34 a) Estrutura de um BJT n-p-n. b) BJT polarizado no modo ativo.

O transistor polarizado no modo ativo (que é o modo do transistor quando este é utilizado como um amplificador) é mostrado na Figura 34b. A junção **n-p** formada entre emissor e base é polarizada diretamente, enquanto a junção entre base e coletor tem polarização inversa. A ideia desse esquema de polarização é a seguinte: quando a junção emissor-base é polarizada diretamente, elétrons vão migrar do emissor para a base. Quando esses elétrons atravessarem a região de depleção entre **n** e **p** (emissor e base), eles vão difundir na base (região **p**) até que alcancem a região de depleção da junção base-coletor. Quando esses elétrons alcançarem a região de depleção entre a junção base e coletor (que está com polarização inversa), o campo elétrico dessa região “coleta” os elétrons rapidamente, pois esse campo exerce uma força que arrasta os elétrons na direção do coletor. Note que se todos os elétrons que saem do emissor (ou seja, que são emitidos pelo emissor) forem coletados pelo coletor, a corrente de elétrons na base seria nula, o que é o caso ideal. Os mecanismos de transporte, então, são: difusão de portadores minoritários na região **p** até atingir a região de depleção entre

base e coletor, onde os elétrons são arrastados (e, portanto, coletados) pelo campo elétrico nessa junção. O processo está ilustrado na Figura 35 (leia a legenda da figura).

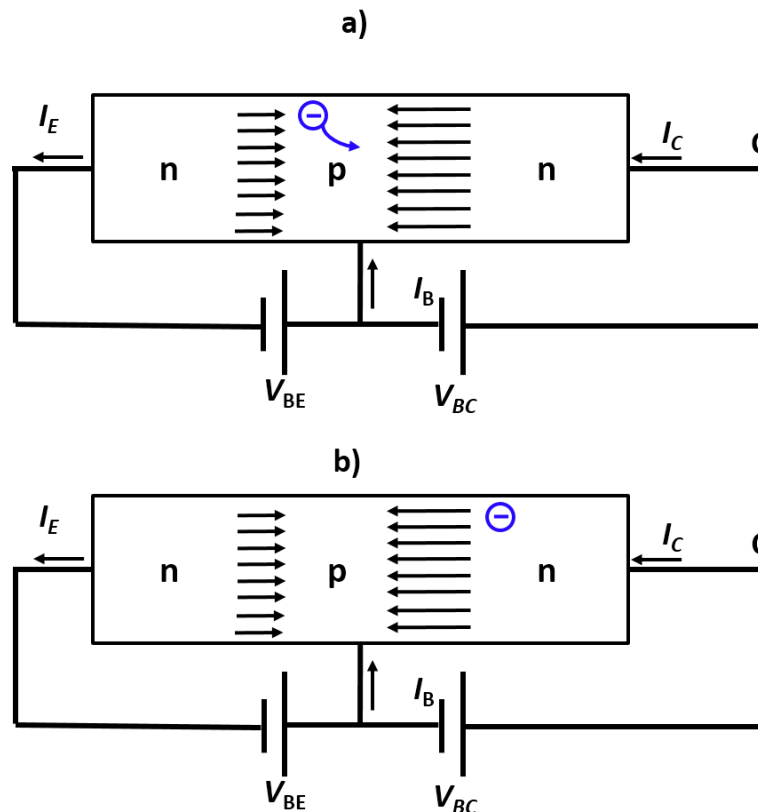


Figura 35 Representação do transporte de cargas no BJT n-p-n polarizado no modo ativo. As setas pretas mostram a direção do campo elétrico nas regiões de depleção a) A junção emissor-base está polarizada diretamente, então elétrons vão migrar do emissor para a base. Esses elétrons são portadores minoritários na parte p. Esses portadores minoritários difundem até alcançar a região de depleção entre base e coletor. b) quando os elétrons chegam na região de depleção, eles são rapidamente coletados, pois o campo elétrico arrasta esses elétrons até o coletor.

Como mencionado acima, a ideia do transistor é controlar a corrente entre coletor e emissor através da tensão na base. Então, precisamos expressar a relação entre essa corrente e a tensão na base. Já temos todas as ferramentas conceituais para isso: já sabemos que o transporte de portadores minoritários é difusão: os elétrons na região **p** difundem até o coletor. E já sabemos que a concentração de portadores, nesse caso, segue uma exponencial (Equação 228 ou Equação 229). Tudo o que precisamos saber, então, é qual é a concentração de portadores no início da região **p** (imediatamente após a região de depleção entre emissor e base), que é onde começa a difusão.

Para ficar mais clara a situação, vamos colocar um eixo x no transistor, como mostrado na Figura 36a. Nesse eixo, a origem está no começo da região **p**, onde o

transporte é dado por difusão, e a distância que as cargas têm que difundir é dada por W . A Figura 36b mostra a distribuição de cargas real (vermelho pontilhado) e a distribuição ideal (azul sólido), que é a distribuição para o caso onde a recombinação é zero. Nesse último caso, a exponencial se aproxima de uma reta (por quê?). Note que, em ambos os casos, a concentração de portadores minoritários $n_e(x)$ vai para zero em $x = W$. Isso ocorre porque os elétrons, nesse ponto, são coletados pela segunda região de depleção (lembre-se também que o subíndice e nos lembra que estamos tratando de **excesso** de portadores minoritários).

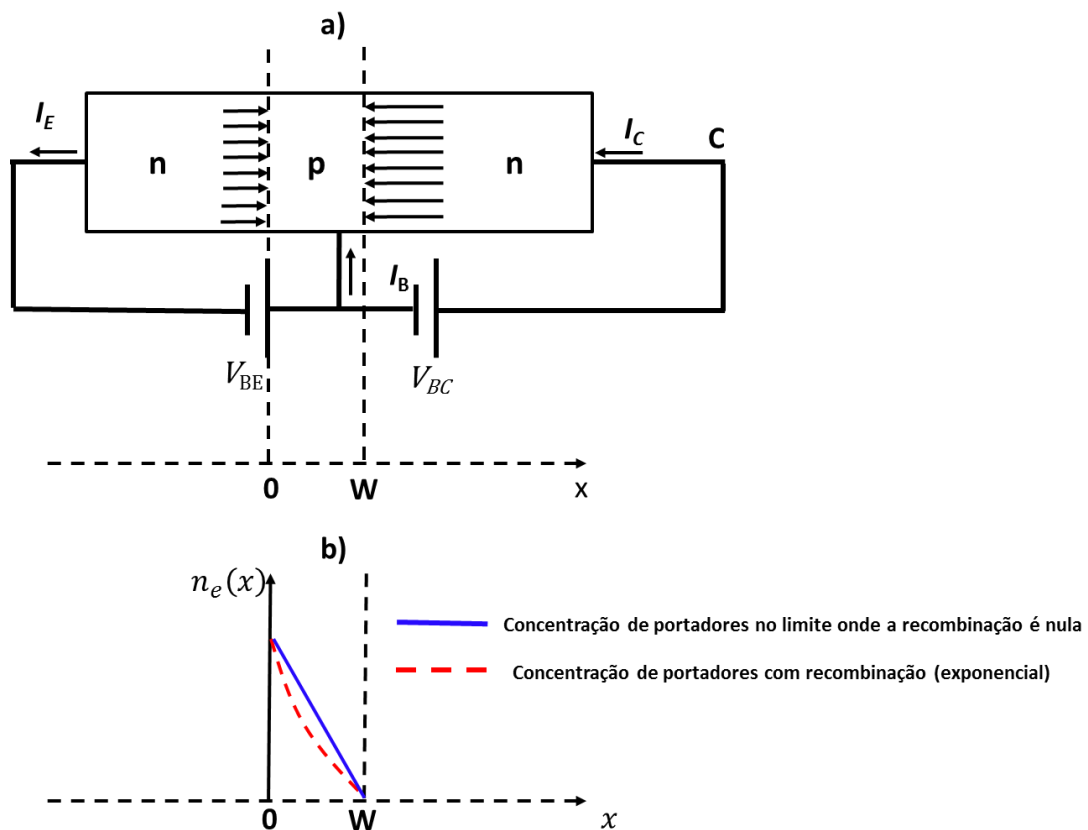


Figura 36 Distribuição de portadores minoritários (elétrons) na região p.

Se quisermos trabalhar com o caso real, teremos que $n_e(x)$ será uma exponencial que tem que satisfazer as condições de contorno para $x = 0$ e $x = W$. Mas como a recombinação é muito baixa, podemos aproximar essa exponencial por uma reta. Assim, a distribuição de portadores de carga será a reta que satisfaz as condições de contorno em para $x = 0$ e $x = W$, ou seja:

$$n_e(x) = -\frac{n_e(0)}{W}x + n_e(0)$$

Equação 253

Mas a corrente de difusão é:

$$J = |q|D_N \frac{dn}{dx} = |q|D_N \frac{dn_e}{dx}$$

Equação 254

Portanto:

$$J = -\frac{|q|D_N}{W} n_e(0)$$

Equação 255

Note que a corrente é negativa porque elétrons movem na direção de x positivo. Tudo o que precisamos, agora, é encontrar $n_e(0)$ para finalizar a história. Mas $n_e(0)$ é a concentração de portadores minoritários no início da parte **p**. Já determinamos qual é essa concentração: a resposta está na Equação 247. Podemos simplesmente repetir esse resultado aqui (adaptando para o novo eixo de coordenadas):

$$n_e(0) = n_0(0) \left[\exp\left(\frac{|q|V_{ext}}{k_B T}\right) - 1 \right]$$

Equação 256

Lembre-se que $n_0(0)$ é a concentração de elétrons na região **p** no equilíbrio. Portanto, esse é um parâmetro do material. E V_{ext} é a tensão na junção **p-n** que, nesse caso, chamamos de V_{BE} (Figura 36). Assim:

$$n_e(0) = n_0(0) \left[\exp\left(\frac{|q|V_{BE}}{k_B T}\right) - 1 \right]$$

Equação 257

Finalmente, substituindo a Equação 257 na Equação 255 temos:

$$J = -\frac{|q|D_N n_0(0)}{W} \left[\exp\left(\frac{|q|V_{BE}}{k_B T}\right) - 1 \right]$$

Equação 258

Está aí a essência do transistor: a corrente entre dois terminais (que é a corrente entre o emissor e o coletor) é controlada pela tensão na base (que é V_{BE}). Nos cursos de eletrônica você vai aprender como fazer um monte de coisas com o transistor: amplificadores, chaves, portas lógicas. E a essência está aí. Esse é mais um exemplo de como um estudo mais detalhado de junções **p-n** é fundamental para entender conceitos mais avançados de eletrônica.

4.3 Exercícios

Exercício 1

Explique a origem física do efeito fotovoltaico

Exercício 2

Utilize a equação de Shockley para encontrar uma expressão para a tensão em circuito aberto de uma célula solar, em termos da corrente de iluminação.

Exercício 3

Rascunhe a relação de corrente (eixo vertical) contra tensão (eixo horizontal) de uma célula solar, mostrando os principais parâmetros.

Exercício 4

Na Figura 36, a concentração de portadores é aproximada por uma reta. Mas sabemos que, quando o transporte é por difusão (como é o caso da Figura 36), a concentração é uma exponencial. Prove que no limite que a recombinação vai para zero (ou seja, o tempo de vida vai para infinito), a exponencial tende a uma reta.

Apêndice A: Interpretação geométrica da energia livre

É relativamente comum encontrar a afirmação que calor (calor é o termo $T\Delta S$) é a forma de energia que carrega entropia, enquanto o termo $\mu\Delta N$ é o termo livre de entropia, também chamado de energia livre. Nesse apêndice, eu proponho uma interpretação geométrica dessas afirmações.

Temos considerado a entropia como função da energia e do número de partículas. Nesse caso, a energia e o número de partículas desempenham o papel de variáveis livres. Mas podemos muito bem inverter a relação e considerar a entropia e o número de partículas como variáveis livres, enquanto a energia depende de ambos. A interpretação geométrica fica mais fácil considerando a energia como função da entropia e do número de partículas.

Vamos supor que temos uma curva da energia contra a entropia para um valor fixo de N , e outra curva da energia contra a entropia para um valor fixo de $N + \Delta N$. Aqui ΔN deve ser interpretado como uma variação infinitesimal. Se não houver descontinuidades na função, e se ΔN for infinitesimal, então as duas curvas são virtualmente idênticas, mas uma estará deslocada verticalmente da outra, como representado na figura abaixo:

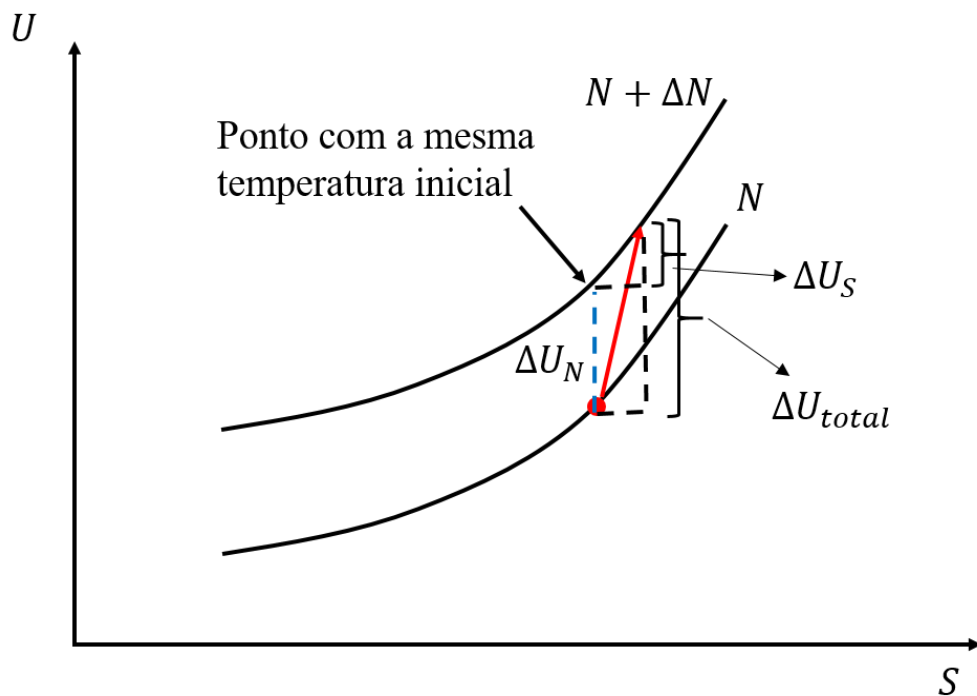


Figura 37 Representação da energia em função da entropia e do número de partículas, destacando duas curvas de U contra S para valores fixos de N , infinitesimalmente próximos.

Eu exagerei no deslocamento vertical para que a gente possa enxergar o significado geométrico dos termos mais facilmente. Para isso, considere que o sistema está no ponto marcado pela bolinha vermelha. Agora considere que o sistema recebeu ΔN partículas, migrando assim para o novo ponto representado pela flecha. Podemos representar a variação total da energia como:

$$\Delta U_{total} = \Delta U_S + \Delta U_N$$

Onde ΔU_S é a variação da energia devido à variação da entropia, e ΔU_N é a variação da energia devido à variação no número de partículas. Ambos termos estão ilustrados na figura. Note que essa relação nada mais é que a expressão do cálculo:

$$\Delta U_{total} = \frac{\partial U}{\partial S} \Delta S + \frac{\partial U}{\partial N} \Delta N$$

onde

$$\Delta U_S = \frac{\partial U}{\partial S} \Delta S \quad \text{e} \quad \Delta U_N = \frac{\partial U}{\partial N} \Delta N$$

Note também que as expressões acima nada mais são que a relação geométrica “o cateto oposto é igual à tangente vezes o cateto adjacente”.

Em um processo isolado do universo, a variação da energia quando o sistema recebeu ΔN partículas seria o ΔU_{total} . Mas, se o sistema estiver em contato térmico com um reservatório, a temperatura do sistema será fixa pela temperatura do reservatório. Como a temperatura é a tangente da curva da energia contra a entropia, o ponto de mesma temperatura na curva $N + \Delta N$ será o ponto com mesma tangente, ou seja, o ponto imediatamente acima do ponto de onde o sistema partiu, como representado na figura acima (lembre-se que as curvas são virtualmente idênticas, mas deslocadas verticalmente). Isso significa que o sistema vai descer a curva do ponto onde está a flecha, para o ponto de mesma temperatura. Em outras palavras, o sistema vai devolver ΔU_S para o reservatório, sobrando apenas ΔU_N . Sendo assim, o termo

$$\Delta U_S = \frac{\partial U}{\partial S} \Delta S = T \Delta S$$

representa a parte da energia que é devolvida ao reservatório em razão da temperatura estar fixa. Esse termo a gente chama de calor e, como ilustrado na figura acima, ele realmente é o termo que alterou a entropia inicialmente. Note também que, depois de perder essa parte da energia para o reservatório, a entropia volta para a condição inicial, já que o ponto com mesma temperatura na curva $N + \Delta N$ está na mesma linha vertical do ponto inicial (mesma entropia).

Agora eu preciso mostrar que o termo ΔU_N , que já mostramos que é a energia que sobra depois que o sistema voltou à temperatura inicial, realmente coincide com o termo $\mu\Delta N$. Para isso, considere as mesmas curvas do exemplo anterior, mas agora vamos desenhar um triângulo retângulo cujo cateto oposto seja o termo ΔU_N , como ilustrado abaixo.

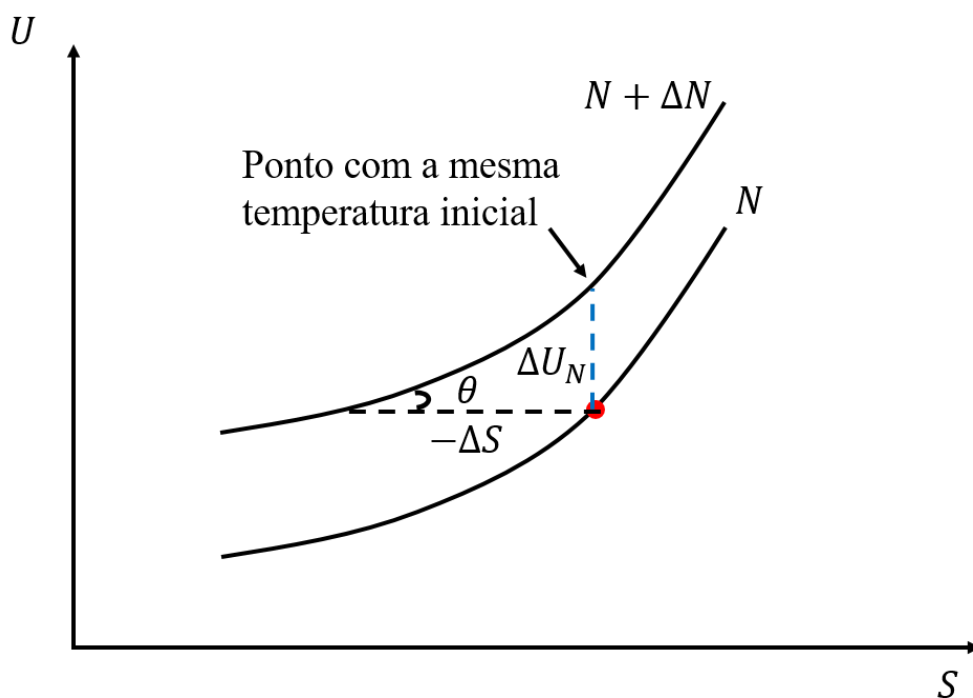


Figura 38 Representação geométrica do potencial.

O cateto adjacente ao ângulo θ na figura acima possui comprimento igual à $-\Delta S$. Eu preciso do sinal de negativo porque comprimento é um número positivo, mas ΔS aqui é negativo quando fomos da curva N para a curva $N + \Delta N$ (S foi para a esquerda). Sendo assim, temos que:

$$\tan \theta = \frac{\Delta U_N}{-\Delta S}$$

mas

$$\tan \theta = \frac{\partial U}{\partial S} = T$$

portanto:

$$T = \frac{\Delta U_N}{-\Delta S}$$

Por outro lado, temos que:

$$\Delta U_N = \frac{\partial U}{\partial N} \Delta N$$

portanto:

$$T = \frac{\frac{\partial U}{\partial N} \Delta N}{-\Delta S}$$

O que nos leva à conclusão que:

$$\frac{\partial U}{\partial N} = -T \frac{\Delta S}{\Delta N}$$

Como os deslocamentos são infinitesimais, temos efetivamente que:

$$\frac{\partial U}{\partial N} = -T \frac{\partial S}{\partial N}$$

Sendo assim, temos que:

$$\Delta U_N = \frac{\partial U}{\partial N} \Delta N = -T \frac{\partial S}{\partial N} \Delta N$$

Como sabemos, o potencial é definido como:

$$\mu = -T \frac{\partial S}{\partial N}$$

Concluimos então que, realmente:

$$\Delta U_N = \mu \Delta N$$

Um último comentário: o potencial poderia ser definido como a derivada parcial da energia em relação ao número de partículas, mas a energia geralmente utilizada é a chamada “energia livre de Helmholtz”, definida como:

$$F = U - TS$$

Onde F é a energia de Helmholtz. Dessa definição, temos que:

$$\Delta F = \Delta U - \Delta TS - T\Delta S = \mu \Delta N - S\Delta T$$

Sendo assim, a energia livre é função do número de partículas e da temperatura, onde:

$$\frac{\partial F}{\partial N} = \mu \quad \text{e} \quad \frac{\partial F}{\partial T} = -S$$

Em um processo com temperatura fixa ($\Delta T = 0$), a variação da energia livre de Helmholtz se reduz à:

$$\Delta F = \mu \Delta N$$

Por isso que o termo $\mu \Delta N$ é também chamado de energia livre.

O que seria do mundo sem Pitágoras?