



ESCOLA DE ENGENHARIA DE SÃO CARLOS

DEPARTAMENTO DE ENGENHARIA ELÉTRICA E DE
COMPUTAÇÃO

SEL 0450 – SEMICONDUTORES

Notas de Aula

Prof. Dr. Emiliano R. Martins

1 – Introdução e contextualização

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 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 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 faz parte de uma área do conhecimento chamada de Física do Estado Sólido, que é uma disciplina que requer uma sólida base em Mecânica Quântica e em Física Estatística. Os conceitos de Física do Estado Sólido são então utilizados como base para a segunda etapa, que constitui o estudo dos dispositivos semicondutores principais: diodos, transistores, dispositivos optoeletrônicos, etc... Podemos estimar, portanto, que um curso de semicondutores demandaria por volta de 3 anos ou, no mínimo, 20 créditos. Como essas notas de aula são para um curso de 2 créditos, nosso primeiro desafio é encontrar uma estratégia que permita condensar um curso de 20 créditos em 2. Isso só é possível se definirmos objetivos claros e razoáveis para o curso. Se quisermos abranger a disciplina inteira, vamos naturalmente tropeçar na falta de tempo e acabar nos perdendo.

Qual é então o objetivo deste curso? Este é um curso oferecido para o departamento de Engenharia Elétrica, portanto a principal motivação para estudar semicondutores aqui é derivada do interesse em dispositivos eletrônicos. Praticamente

todos os dispositivos eletrônicos, por outro lado, 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 base sólida 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 x corrente em um diodo. No caminho, o aluno também entrará em contato com noções centrais a respeito de baterias e células solares.

O curso (e estas notas de aula) será dividido em três partes. Primeiro 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. No caminho, temos que pincelar algumas noções relacionadas à mecânica quântica, mas não podemos nos aprofundar nelas pela óbvia falta de tempo e espaço. Assim, o segundo capítulo destas notas de aula 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 curso de Física Estatística e essas não são notas de aula de Física Estatística. Em momento nenhum o aluno deve imaginar que o estudo do capítulo 2 substitui um curso de Física Estatística ou termodinâmica. Ao contrário, no capítulo 2 serão tratados somente os conceitos indispensáveis para semicondutores. No terceiro capítulo iremos focar em alguns aspectos da física de semicondutores mais específicos, como a teoria de bandas. Aqui, teremos que apresentar brevemente mais alguns conceitos de mecânica quântica e juntá-los com os conceitos do capítulo 2 para podemos deduzir algumas propriedades elétricas dos semicondutores. No quarto e último capítulo iremos utilizar todos os conceitos dos capítulos anteriores para estudar a junção p-n e deduzir suas propriedades elétricas.

2 – Noções de Física Estatística

2.1 Introdução

Um bom começo para o estudo da Física Estatística é a compreensão de um princípio probabilístico que pode ser enunciado como: *um sistema macroscópico em equilíbrio sempre está no seu estado mais provável*. O objetivo desta seção introdutória é explicar esse princípio.

Para compreender este princípio, 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 que estamos interessados. Por exemplo, se queremos estudar a interação entre dois gases, então os dois gases formam o nosso sistema. Dizemos também que o sistema é isolado quando não há interação do sistema com o mundo externo. No exemplo dos dois gases, o sistema estaria isolado se os dois gases interagirem somente entre si. Isso poderia ser feito colocando os gases 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. Esse termo ficou popularizado com o advento da mecânica quântica, mas é importante definir o que é um estado propriamente dito e 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 o 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. Ou seja, estado se refere a certas propriedades de interesse. Em física, as propriedades de interesse geralmente são a energia, a posição e a quantidade de movimento (que daqui em diante vou chamar de momento) dos elementos. Então no exemplo dos gases, quando perguntamos qual é o estado dos gases, estamos efetivamente perguntando qual é o momento, energia e posição de cada elemento do gás.

Agora imagine um sistema composto por uma bolinha de massa m em uma rampa, como ilustrado na Figura 1. Se eu soltar a bolinha de uma altura h , ela terá energia total inicial em forma de energia potencial ($U = mgh$, 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 o estado de energia do sistema, a resposta será U . Note que neste exemplo o estado de energia não depende do tempo. 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. Mas se perguntarmos qual é o estado de energia do sistema, sabemos que a resposta será U para qualquer instante de tempo. E se perguntarmos quais são os estados de energia possíveis? Neste caso, estamos perguntando quais valores que U pode tomar. Como podemos a priori soltar a bolinha de qualquer altura h , então U pode tomar qualquer valor. Mesmo se você fixar uma altura máxima, digamos de 10 metros, ainda assim teremos infinitos valores de U , porque U poderá tomar qualquer valor entre 0 e $10mg$. 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 causaria uma descontinuidade).

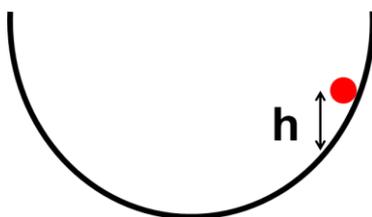


Figura 1 Um sistema que consiste de uma bolinha de massa m deslizando sem atrito em uma rampa pode apresentar qualquer estado de energia. Basta ajustar a altura h que a bolinha é solta para obter o estado desejado.

Em física clássica, o estado de energia dos sistemas é sempre um contínuo. Basta pensar que se você pegar uma massa, a priori você pode ajustar a energia cinética dessa massa para qualquer valor, basta ajustar a velocidade da massa (a teoria da relatividade não altera esse fato, ela só põe um limite na velocidade da massa, que não pode ultrapassar a velocidade da luz; mas mesmo assim, existe um contínuo de infinitas velocidades possíveis entre 0 e a velocidade da luz). Em mecânica quântica, entretanto, isso não é mais verdade. 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, existe um número inteiro (e, portanto, discreto, que você pode contar), de alturas possíveis, neste caso 3 alturas possíveis, que chamamos de h_0 , h_1 e h_2 . Portanto temos apenas três estados de energia possíveis, $U_0 = mgh_0$, $U_1 = mgh_1$ e $U_2 = mgh_2$.

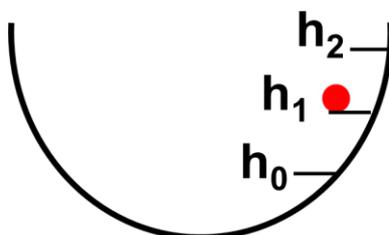


Figura 2 Em um sistema quântico, apenas alguns níveis discretos de energia são permitidos. Aqui representamos esta propriedade delimitando três alturas h_0 , h_1 e h_2 das quais a bolinha pode ser solta. Portanto esse sistema possui apenas três estados de energia possíveis.

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 está pedindo uma fibra que admite apenas um estado possível, enquanto uma fibra multimodo pode admitir 2 ou mais estados, mas sempre um número discreto de estados.

A origem dessa discretização é o confinamento da onda. Toda vez que nós confinamos uma onda, esse 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 . Se ela não for um múltiplo de 2π , então cada vez que ela der uma volta ela vai sofrer interferência destrutiva consigo mesma, o que impede a sua existência.

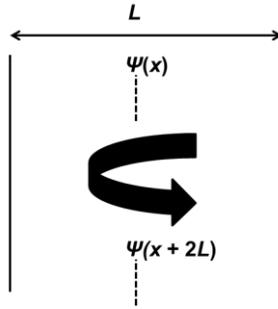


Figura 3 Uma onda confinada só pode existir se a fase acumulada em um giro completo for um múltiplo de 2π .

Podemos expressar essa restrição em termos matemáticos. Para tal, suponhamos 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 a condição de acúmulo de fase depende só da dependência espacial, portanto a dependência temporal é irrelevante para nós neste exemplo. Qual é a fase acumulada pela onda quando ela percorre uma volta inteira? Cada vez que ela tromba em uma parede, 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. Se a cavidade tem comprimento L , uma volta inteira corresponde à uma distância $2L$. Portanto a relação entre a onda que sai de um ponto e a onda que volta sobre o mesmo ponto depois de um giro completo é:

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

Equação 2

Assim, vemos que o acúmulo de fase em um giro completo é igual a $k2L$. Mas para a onda existir na cavidade, é necessário que o acúmulo de fase em um giro completo seja igual à um múltiplo de 2π . Portanto, a condição de existência 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 uma onda confinada em uma caixa de comprimento L possui estados possíveis caracterizados pela 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), estes 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 tem propriedades ondulatórias. Você já deve ter ouvido falar da dualidade onda-partícula. Um elétron não é um ponto fixo no espaço, um elétron é uma onda. Que a matéria é onda é uma das descobertas mais revolucionárias e malucas 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, se a Equação 1 for a função de onda de um elétron e se o elétron estiver 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). 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 é proporcional à constante de propagação:

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

Equação 5

Onde h é uma constante chamada de constante de Planck (se você está se perguntando da 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 essa relação ser verdadeira, apenas sabemos que é verdadeira). 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. Como 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 $p = mv$ 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 Equação 6 em 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 em alguns materiais. 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 a diferença de energia dada pelos níveis da Equação 8. Hoje em dia existem técnicas para formar materiais minúsculos, da ordem de nanômetros, chamados de pontos quânticos. Esses materiais absorvem e emitem luz dependo 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).

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 do seção 2.1, 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 o seu estado não é mais alterado. Voltando ao exemplo dos dois gases interagindo, o estado de cada gás (que poderia ser descrito por grandezas macroscópicas como volume, pressão e temperatura) irá alterar à medida que os gases interagem, mas chega uma hora que o sistema como um todo atinge equilíbrio e o estado dos dois gases passa a se manter fixo. Note que os gases continuam interagindo, mas o estado deles não é mais alterado pela interação após atingir o equilíbrio. Por que? O que o estado de equilíbrio tem de tão especial assim que uma vez que o sistema o atinge ele fica lá pra sempre? A resposta está no enunciado do princípio. Vamos lembrar o que diz o princípio: *um sistema macroscópico em equilíbrio sempre está no seu estado mais provável.*

Vamos pensar um pouco sobre qual é a informação que está aí. Nessa frase, estamos efetivamente dizendo que o estado de equilíbrio é o estado mais provável. Note que o princípio poderia ser enunciado da seguinte maneira: *um sistema macroscópico cujo estado não é mais alterado está sempre no seu estado mais provável.* Essa segunda maneira de enunciar o princípio é menos elegante, porém ela é útil para enfatizar que a palavra equilíbrio quer dizer simplesmente que o estado está fixo. Então o que exatamente o princípio está nos dizendo? Ele está nos dizendo que quando os elementos de um sistema interagem entre si (ou com o universo exterior), uma hora eles vão atingir um estado, que é o estado mais provável, e depois disso eles não saem mais deste estado. 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 dois gases isolados do mundo externo, um com energia $U1_0$ e o outro com energia $U2_0$. Se deixarmos os gases interagindo, não saberemos bem qual será a energia do gás 1 e do gás 2, mas porque energia é conservada, então sabemos que todos os estados possíveis do sistema são tais que satisfazem a condição $U1 + U2 = U1_0 + U2_0$, onde $U1$ e $U2$ são as energias do gás 1 e 2 depois ou durante a interação. E de todos os estados possíveis, o princípio está nos dizendo que o sistema estará no mais provável. O princípio nos diz: pegue todos os estados possíveis e veja qual é o mais provável, chame esse estado mais provável de X. Então o seu sistema vai estar no estado X depois de um tempo. É isso que o princípio está afirmando.

Toda essa elucubração leva a três perguntas: como posso achar os estados possíveis, o que o termo estado mais provável realmente significa, e por que cargas d'água o sistema evolui para o estado 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 o nosso sistema seja composto por N moedas que foram lançadas ao ar. *Vamos supor que o estado do sistema seja definido pelo número de moedas que caíram com a coroa virada pra cima.* Com essa definição de estado, quantos estados possíveis nós temos? Bom, isso depende do tanto de moedas N que nós temos. Se $N = 1$, então só temos dois estados possíveis: 0 coroas para cima, ou 1 coroa para cima. Se $N = 2$, então temos 3 estados possíveis: 0 coroas, 1 coroa e 2 coroas. Mas qual é o estado mais provável? Bom, existem 4 configurações possíveis: cara-cara, cara-coroa, coroa-cara e coroa-coroa. Então existe 25% de chance do estado ser 0 coroas (que corresponde à configuração cara-cara), 25% de chance do estado ser 2 coroas (que corresponde à configuração coroa-coroa) e 50% de chance do estado ser 1 coroa (que corresponde tanto à configuração cara-coroa quanto à configuração coroa-cara). Portanto o estado mais provável é o estado 1 coroa, porque temos duas configurações possíveis para esse mesmo estado. Note a distinção entre configuração e estado. Configuração se refere à distribuição exata de caras e coroas, enquanto estado se refere à uma propriedade dessa distribuição (quantas coroas estão para cima).

Vamos então generalizar esse raciocínio calculando qual a probabilidade de termos M coroas em um sistema com N moedas. Primeiro, quantas são as configurações possíveis? Essa é fácil, temos 2^N configurações possíveis de caras e coroas. Para calcular a probabilidade de sair M coroas, basta calcularmos quantas configurações possuem M coroas e dividir esse número por 2^N . O número de configurações com M coroas viradas corresponde ao número de maneiras que podemos arranjar N moedas com M delas com a coroa virada e $M - N$ com a 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 1 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 esse número $N!$ contou diversas configurações de cara e coroa idênticas mais de uma vez, porque o marcador as havia diferenciado.

Por exemplo, suponha que tínhamos 2 coroas e 3 caras (5 moedas). Vamos supor que marcamos a moedas com os marcadores M1, M2, M3, M4 e M5. Neste caso as moedas M1 e M2 são cara e as M3, M4 e M5 são coroas. Assim, existem $5! = 120$ diferentes configurações de moedas marcadas. Note que a configuração M1M2M3M4M5 foi contada diferente da configuração M2M1M3M4M5. Mas sabemos que, para o nosso propósito, essas duas configurações são na verdade idênticas (ambas correspondem à configuração Coroa-Coroa-Cara-Cara-Cara): elas só foram diferenciadas pelo marcador. Então temos que cada dividir 120 por 2 porque contamos as configurações M1M2 e M2M1 como distintas quando na verdade essa é uma configuração só (note que todas as configurações foram contadas duas vezes, mesmo se M1 e M2 não forem adjacentes; por exemplo, a configuração M1M3M4M2M5 é a mesma da configuração M2M3M4M1M5). Mas também temos que dividir por $3! = 6$ porque as configurações M3M4M5, M3M5M4, M4M3M5, M4M5M3, M5M3M4 e M5M4M3 foram todas contadas como distintas quando na verdade são iguais. Assim, definindo a função $A(M)$ como sendo o número de configurações com M coroas para N moedas, temos que:

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

Equação 9

Então a probabilidade de termos M moedas nada mais é que:

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

Equação 10

Podemos utilizar a Equação 9 ou Equação 10 para estudar o comportamento da probabilidade $P(N, M)$ em função do número de moedas N e do número de coroas M . A Figura 4 ilustra alguns histogramas (normalizados) obtidos da Equação 9. **Note como o número de configurações 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 é dada por:

$$\frac{\Delta M}{N} \approx 2 \sqrt{\left(\frac{1}{2N}\right)}$$

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 configurações $A(M)$. 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, quando maior for o número de moedas, mais provável será que o resultado seja metade das moedas com coroa pra cima. Utilizando uma linguagem um pouco mais precisa, concluímos que a chance de encontrarmos um número de coroas que seja apreciavelmente diferente da metade é muito baixa. Por exemplo, se tivermos 1 milhão de moedas, não ficaremos surpresos se encontrarmos 500 mil e 5 coroas viradas para cima, porque 500 mil e 5 é um número proporcionalmente muito próximo de 500 mil (ou seja, o desvio fracional é muito pequeno: $5/N = 5/10^6$). Mas é virtualmente impossível encontrarmos 250 mil coroas para cima. O que o desvio fracional nos diz é qual o número 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 será entre 499 mil e 501 mil. Qualquer valor fora dessa faixa é muito improvável. Portanto, o estado do sistema estará muito, muito próximo do estado mais provável, simplesmente porque o número de configurações possíveis para o estado mais provável é bem maior do que para os outros estados.

Agora imagine que você chegue em uma sala com mil moedas espalhadas no chão. Se você passar o olho e ver 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 de número de caras ou coroas, não no sentido de distribuição espacial). Agora se todas as mil estiverem com a cara virada pra cima, você vai deduzir que alguém virou as moedas propositadamente. Essa sua intuição é consequente do fato que você sabe que o estado mais “desorganizado” é muito mais provável que o estado “organizado”. Por isso que 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, na qual a entropia de sistemas sempre aumenta – ou em alguns casos ideais é mantida constante). Como então que foi possível que o homem, que é um sistema ultra-organizado, tenha surgido puramente por processos naturais é um problema que escapa meu entendimento de Física Estatística.

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 do estado mais provável é muito, muito pequeno. Por exemplo, um gás possui aproximadamente 10^{23} moléculas em 1cm^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} !! Assim, para sistemas macroscópicos podemos efetivamente dizer que o sistema estará no estado 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 no estado mais provável.

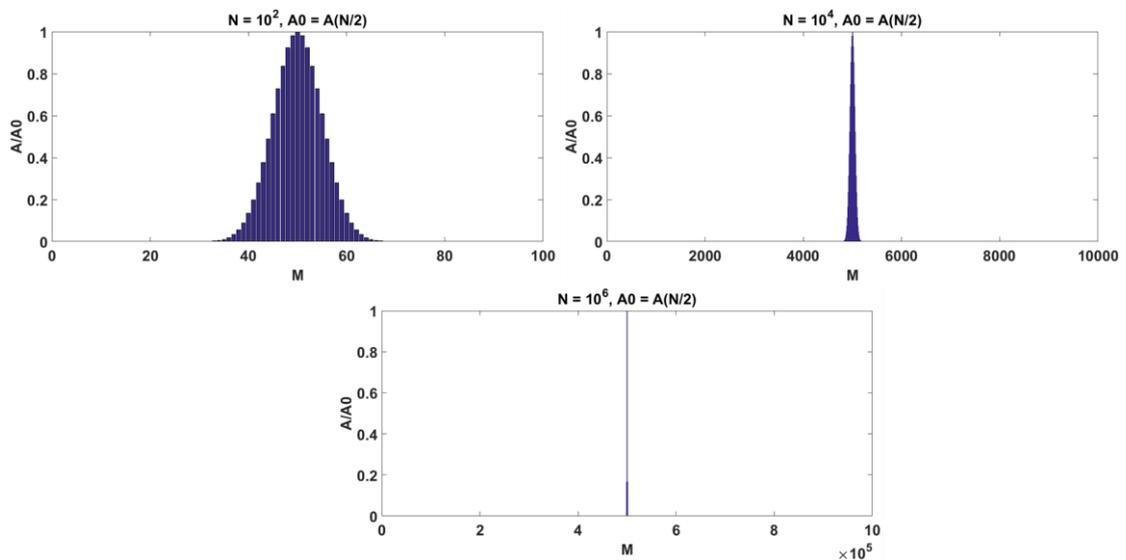


Figura 4 Comportamento de A em função de M para diferentes valores de N . Para facilitar a visualização, a função A foi normalizada. Note que à medida que N aumenta, a função A fica cada vez mais concentrada em torno de $M = N/2$.

Vamos recapitular o raciocínio: utilizando o exemplo das moedas, vimos que se o número de moedas N for muito grande, então o número de configurações que resultam em metade das moedas com coroa pra é muito, muito maior que o número de

configurações para qualquer outro número de coroas viradas para cima. Isso significa que se você jogar um número muito alto de moedas para cima, 10 milhões de moedas, por exemplo, então é virtualmente garantido que a configuração final apresentará metade caras e metade coroas para cima. Se o tanto de coroas virado para cima definir o estado, então é garantido que o estado final é o estado mais provável (metade de coroas para cima). A razão pela qual é virtualmente garantido que teremos metade caras e metade coroas advém só e somente só do fato de que existe um número muito grande de configurações que resultam em metade caras e metade coroas (no exemplo de duas moedas, temos 2 configurações com metade cara e metade coroa: Cara-coroa, coroa-cara). Portanto, afirmar que é virtualmente garantido que o estado final terá metade das coroas viradas para cima se N for muito grande é simplesmente consequência do fato de que o número de configurações possíveis que resultam em metade de coroas viradas para cima é muito maior que o número de configurações possíveis para outros números de coroas viradas para cima.

Praticamente toda a Física Estatística é construída em torno dessa observação: que a configuração final de um sistema macroscópico será o estado mais provável. A razão pela qual isso é verdade é a mesma do exemplo das moedas: existe um número muito maior de configurações possíveis que resultam no estado mais provável. E já sabemos também que esse é o estado mais desorganizado. Mas note um pequeno, porém importantíssimo detalhe: essa conclusão só é válida se supormos que cada configuração é igualmente provável. Se as configurações não forem igualmente prováveis, nosso raciocínio iria por água abaixo. Por exemplo, se as moedas fossem “viciadas” de maneira que a probabilidade de sair coroa é 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 todas as configurações 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 um 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, você resolve aplicar as leis de Newton 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 o estado é determinado pela distribuição espacial do gás) você teria que aplicar a lei de Newton em 10^{23} moléculas. É claro que você não vai dar conta. Você e nenhum computador vai dar conta. A física convencional, apesar de ser verdadeira, 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, não tem nenhum ponto do recipiente que 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. 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 uma certa distribuição espacial de moléculas. Para isso você teria que calcular o número total de configurações que resultam nessa certa distribuição espacial e dividir pelo número total de configurações possíveis. Daria um trabalho infinito para fazer isso. Mas você observa que esse é um sistema macroscópico (muitas moléculas). Então você sabe que existem muito mais configurações possíveis para o estado mais provável, ou seja, que a probabilidade do sistema estar no estado mais provável é muito perto de 1. E qual é o estado mais provável? É o estado mais bagunçado. E qual é o estado mais bagunçado? É o estado caracterizado por uma distribuição uniforme de moléculas, ou seja, o estado no qual as moléculas enchem o recipiente de maneira uniforme. Portanto essa é a sua resposta: eu não sei exatamente onde cada molécula vai estar, mas eu sei que a distribuição espacial delas será uniforme, pois esse é o estado mais provável e como esse é um sistema macroscópico então a probabilidade do estado mais provável é virtualmente igual a 1. Na seção seguinte, vamos utilizar esse princípio para quantificar algumas grandezas.

Talvez você esteja com uma pulga atrás da orelha porque apesar de ter certeza que se você soltar um gás em um recipiente então o gás vai preencher o recipiente como

um todo, essa história de distribuição uniforme de moléculas ser a mais desorganizada pode parecer meio estranha, já que a noção de uniformidade lembra a noção de organização. Mas esse conceito não tem nada de muito científico e misterioso não. A noção de desorganização aqui (e desse ser o estado 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 entrar no quarto do amigo organizado, você espera encontrar todos os sapatos guardados em algum canto (portanto uma distribuição espacial não-uniforme de sapatos). Agora se você entrar no quarto do amigo desorganizado, então você espera 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 não é bem uma lei. É simplesmente uma observação probabilística. Como veremos mais para frente, a entropia está relacionada ao número de configurações possíveis do estado do sistema. Dessa forma, o que a segunda lei da Termodinâmica afirma é simplesmente que o sistema evolui para o estado que possui mais configurações possíveis. Essa propriedade de ter mais configurações possíveis é formulada em termos do conceito de entropia, mas a mensagem é exatamente essa: o sistema evolui para o estado mais provável, que é o que possui mais configurações 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 configurações que resultam no estado bagunça, caracterizado por sapato pra tudo quanto é lado (sapato verde no canto esquerdo e vermelho no canto direito corresponde ao mesmo estado que sapato verde no canto direito e vermelho no canto esquerdo – ambos correspondem ao estado bagunça!). O estado dos sapatos do rapaz organizado, entretanto, possui uma única configuração possível: se você mudar um único sapato, o rapaz organizado via chiar porque você alterou o estado do sistema dele. Portanto o rapaz organizado possui entropia baixa (apenas uma configuração possível).

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 todas as configurações possíveis do sistema têm igual probabilidade, ou seja, que o sistema pode evoluir para qualquer configuração possível com igual probabilidade. Mas daí resulta que como o estado mais provável

possui um número muito maior de configurações possíveis, então a Física Estatística supõe que esse será o estado final. É por isso que a Física Estatística transforma uma observação probabilística (o estado mais provável tem probabilidade virtualmente igual à 1) 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, mas não precisamos entrar aí por enquanto). 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 buscar encontrar onde cada molécula de um líquido ou gás está, 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, ao invés de te dizer onde cada molécula está exatamente. Densidade é um parâmetro típico de Física Estatística.

2.2 Equilíbrio Térmico

Suponha que tenhamos dois sistemas macroscópicos: sistema A e sistema B. Suponha que a energia total do sistema seja U_0 . No início, estes sistemas estão isolados, e o sistema A possui energia U_a e o sistema B energia U_b , de maneira que $U_0 = U_a + U_b$. Agora colocamos os dois sistemas em contato e deixamos interagir. O que podemos afirmar sobre o estado do sistema após atingir equilíbrio?

Para responder essa pergunta, podemos utilizar o princípio básico que foi apresentado na seção anterior. Esse princípio nos diz que o estado final do sistema é o estado mais provável. Para explorar esse princípio de maneira mais quantitativa, temos que começar a definir alguns parâmetros. Estamos interessados aqui principalmente no estado de energia dos sistemas após equilíbrio. Portanto, os estados serão caracterizados pela energia dos sistemas. Vamos chamar o número de configurações possíveis do estado com energia U do sistema A de $g_a(U)$. Note que o

fato da função g_a ser uma função de U expressa a dependência do número de configurações no estado em questão. De maneira análoga, definimos $g_b(U)$ como sendo o número de configurações do estado com energia U do sistema B. Qual será então o número de configurações do estado do sistema total A + B caracterizado pelo sistema A no estado U ? Oras, porque energia é conservada, se o sistema A estiver no estado U , então necessariamente o estado B deverá estar no estado $U_0 - U$. Quantas configurações serão possíveis então para o sistema nesse estado? Se eu tenho $g_a(U)$ configurações para o sistema A e $g_b(U_0 - U)$ configurações para o sistema B, e todas as combinações dessas configurações de A e B resultam no sistema total com estado de energia U em A, então o número de configurações do sistema total no estado caracterizado pelo sistema A com energia U é:

$$g(U) = g_a(U) \cdot g_b(U_0 - U)$$

Equação 12

A função g é geralmente chamada de multiplicidade do sistema, já que ela diz qual é o número de configurações possíveis para um determinado estado. Até agora não fizemos nada demais, simplesmente demos um nome para o número de configurações possíveis de um estado e utilizamos a conservação de energia para expressar o número de configurações do sistema total em termos do número de configurações do sistema A e do sistema B. O que a Equação 12 nos diz é simplesmente que o número de configurações possíveis para o estado do sistema total caracterizado pela energia U do sistema A é o tanto de configurações possíveis de A nesse estado vezes o tanto de configurações possíveis de B no estado $U - U_0$.

Agora, podemos afirmar algo a mais sobre o estado do sistema? Claro que sim, porque o princípio básico nos diz que o sistema vai estar no estado mais provável. E o que é o estado mais provável? Uai, é o que tem o maior número possível de configurações. Então, em termos de $g(U)$, o estado mais provável é dado pelo valor de U que maximiza a função g . Mas sabemos que um ponto de máximo é um ponto aonde a variação de g em termos de U é zero, ou seja, a derivada de g em relação a U avaliada no ponto onde U que maximiza g é zero:

$$\left. \frac{\partial g}{\partial U} \right|_{(U=U_f)} = 0$$

Equação 13

Aqui chamamos o valor de U 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 o estado final será praticamente igual à U_f . Utilizando este princípio basta encontrarmos U_f para encontrarmos o estado final. Derivando g em relação à U temos:

$$\frac{\partial g}{\partial U} = \frac{\partial ga(U)}{\partial U} \cdot gb(U_0 - U) + ga(U) \cdot \frac{\partial gb(U_0 - U)}{\partial U}$$

Equação 14

Da Equação 13 temos então que:

$$\left. \frac{\partial ga(U)}{\partial U} \right|_{U=U_f} \cdot gb(U_0 - U_f) + ga(U_f) \cdot \left. \frac{\partial gb(U_0 - U)}{\partial U} \right|_{U=U_f} = 0$$

Equação 15

Para limpar um pouco a notação, vamos chamar a energia do estado B de U_B e a energia do estado A de U_A . Assim temos que $U = U_A$ e $U_0 - U = U_B$. Reescrevendo a Equação 15 em termos de U_A e U_B temos:

$$\frac{\partial ga(U_A)}{\partial U_A} \cdot gb(U_B) - ga(U_A) \cdot \frac{\partial gb(U_B)}{\partial U_B} = 0$$

Equação 16

Note que o sinal do segundo termo foi trocado porque agora a derivada de gb é em relação à U_B e não mais em relação à U . Note também que suprimimos a notação que explicita que o U que satisfaz essa condição é o U final. Portanto devemos ter cuidado para não esquecer que essa igualdade só é válida para o U mais provável, ou seja, para o U final. Re-arranjando a Equação 16 temos:

$$\frac{1}{ga(U_A)} \frac{\partial ga(U_A)}{\partial U_A} = \frac{1}{gb(U_B)} \frac{\partial gb(U_B)}{\partial U_B}$$

Equação 17

Lembrando que a derivada da função $\ln(x)$ em relação à x é $1/x$, a relação acima pode ser reescrita como:

$$\frac{\partial \ln(ga(U_A))}{\partial U_A} = \frac{\partial \ln(gb(U_B))}{\partial U_B}$$

Equação 18

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

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

Equação 19

E re-expressar a Equação 18 em termos de σ . Assim, concluímos que, no equilíbrio, os sistemas A e B estão em estados que satisfazem:

$$\frac{\partial \sigma_A}{\partial U_A} = \frac{\partial \sigma_B}{\partial U_B}$$

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 termodinâmica e também do princípio probabilístico de que o estado final estará na configuração 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 semi-civilizado, temos então que dar nomes para coisas e expressá-las em unidades reconhecidas pela maioria dos países semi-civilizados. A variável σ não tem nenhum nome muito especial não, 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 re-expressar a condição de equilíbrio em termos da entropia:

$$\frac{\partial S_A}{\partial U_A} = \frac{\partial S_B}{\partial U_B}$$

Equação 22

Se alguém te perguntar o que ocorre quando dois sistemas são postos em contato, a resposta seria que 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 pra derivada da entropia em relação à energia. A convenção é definir a variável T como:

$$\frac{1}{T} = \frac{\partial S}{\partial U}$$

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 o estado mais provável é o estado no 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 atingir equilíbrio. E se alguém te perguntar por que isso ocorre, a resposta correta é: pela mesma razão que se você jogar 1 trilhão de moedas para o alto, metade cairá com a coroa para cima.

A Equação 23 é a definição mais fundamental de temperatura.

2.3 Função de partição, Parte I

Física Estatística lida frequentemente com sistemas em contato com um reservatório. Reservatório nada mais é do que 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).

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 do sistema estar em determinadas configurações. Para explicar melhor o que quero dizer, imagine que a energia total do sistema + reservatório seja U_0 . Qual é então a probabilidade de eu encontrar o sistema em uma determinada configuração, que vou chamar de configuração β ? Vamos supor também que a configuração β tenha energia ε_B . Note que aqui eu fixei a configuração, estou olhando para uma única configuração do sistema e perguntando qual é a probabilidade de eu encontrar o sistema naquela configuração. Essa pergunta é diferente da que tratamos na seção 2.1, onde perguntamos quantas configurações possuem a mesma energia ε . Para ilustrar a diferença entre essas duas perguntas, vamos usar de novo o exemplo das moedas. Vamos supor que uma moeda com coroa virada para cima corresponda a um nível 1 de energia e com cara para cima a um nível 0. Vamos supor que tenhamos 2 moedas. Na seção 2.1 perguntamos quantas configurações resultariam nos estados de energia possíveis. A resposta seria, 1 configuração para o estado com 0 unidades de energia (cara-cara), 2 configurações para o estado com 1 unidade de energia (cara-coroa e coroa-cara) e 1 configuração para o estado com 2 unidades de energia (coroa-coroa). Isso foi o que fizemos na seção 2.1. E o que estamos perguntando agora? Agora estamos olhando para um sistema em contato com o reservatório. Vamos supor que o sistema ainda seja 2 moedas com o esquema tal que 1 coroa corresponda a uma unidade de energia e que o reservatório seja um sistema com 1 trilhão de moedas e com o mesmo esquema de energia. Como o sistema possui 2 moedas, temos novamente as mesmas quatro configurações possíveis. Para facilitar, vamos chamar a configuração cara-cara de configuração α , a configuração cara-coroa de configuração β , a configuração coroa-cara de configuração ζ e a configuração coroa-coroa de configuração κ . A pergunta agora é: dado que o sistema está em contato com o reservatório, qual é a probabilidade de eu encontra-lo na configuração β ? (Claro que β aqui é só um exemplo, a pergunta poderia ser feita para qualquer uma das 4 configurações). Você deve estar se perguntando porque tanto fuzuê, já que é óbvio que a resposta seria que a probabilidade é de 25% para qualquer uma das configurações. Mas isso não é verdade, a probabilidade não é a mesma para todas as configurações. Lembre-se que a suposição fundamental da Física Estatística se refere às configurações do sistema completo, que no nosso exemplo é o sistema com 4

moedas mais o reservatório. Então, para eu calcular a probabilidade de eu encontrar o sistema na configuração β (ou α , ou ζ , ou κ), eu tenho que ver quantas vezes a configuração β aparece nas configurações totais (que incluem o reservatório). E essa contagem de quantas vezes β aparece utiliza conservação de energia e a multiplicidade do reservatório.

Vamos focar então agora no cálculo da probabilidade de achar o sistema em uma determinada configuração β . Como sabemos que energia é conservada, temos que, se a energia total (sistema + reservatório) for U_0 , se o sistema estiver na configuração β , e se a configuração β possui energia ε_β , então o reservatório terá energia $U_0 - \varepsilon_\beta$. E quantas configurações totais são possíveis nas quais o sistema esteja na configuração β ? Se existem M configurações do reservatório que podem combinar com o sistema na configuração β , então obviamente teremos M configurações totais (sistema + reservatório) nas quais o sistema está na configuração β . No exemplo das moedas, a configuração total envolve as duas moedas do sistema mais todas as moedas do reservatório, e teremos M configurações possíveis nas quais as moedas do sistema estão na configuração cara-coroa (que corresponde à configuração β). Mas quem é M ? M são todas as configurações do reservatório que podem combinar com a configuração β . Mais especificamente, isso quer dizer que M são todas as configurações do reservatório que podem combinar com a configuração β respeitando conservação de energia. Portanto, M nada mais é do que o número de configurações do reservatório que possui energia $U_0 - \varepsilon_\beta$. Voltando ao exemplo das moedas, vamos supor que U_0 seja igual 1000 e que o reservatório possua 10 mil moedas. O sistema total possui então 10 mil + 2 moedas. Como a configuração β possui energia 1 (já que ela é cara-coroa), então M corresponderá ao número de maneiras que podemos arranjar as 10 mil moedas com 999 coroas viradas para cima. Ou seja, M é o número de configurações do reservatório com energia igual à 999 (respeitando assim a energia total igual à 1000). Como resolvemos chamar o número de configurações possíveis de multiplicidade, M nada mais é do que a multiplicidade do reservatório para energia $U_0 - \varepsilon_\beta$. Assim, chamando a multiplicidade do reservatório de g_R , temos que $M = g_R(U_0 - \varepsilon_\beta)$.

Mas M não é exatamente o que queremos. Queremos saber qual é a probabilidade de encontrarmos o sistema na configuração β . Bom, se tivermos ao todo N configurações (do sistema + reservatório), e se a configuração β aparece em M dessas

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

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

Equação 24

Note, que no exemplo das moedas, se perguntássemos qual é a probabilidade de encontrar o sistema na configuração β , a resposta seria $P(\beta)$, mas se perguntarmos qual é a probabilidade de encontrar o sistema com energia igual à 1, então a resposta seria $P(\beta) + P(\zeta)$. É muito importante manter essa diferença em mente. Note também que a probabilidade depende somente da energia do sistema (a única informação sobre o sistema que aparece na equação acima é a sua energia ε_β). Portanto, no exemplo das moedas teríamos também que $P(\beta) = P(\zeta)$.

Mas temos um problema: é praticamente impossível calcular N diretamente, exatamente porque o reservatório é um sistema gigantesco. Para nos livrarmos de N , vamos calcular qual é a razão entre a probabilidade de encontrar o sistema em uma configuração β e a probabilidade de encontrar o sistema em um configuração 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 (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 re-escrever a Equação 25 em termos da entropia:

$$\frac{P(\beta)}{P(\alpha)} = \frac{\exp[\sigma_R(U_0 - \varepsilon_\beta)]}{\exp[\sigma_R(U_0 - \varepsilon_\alpha)]} = \frac{\exp\left[\frac{S_R(U_0 - \varepsilon_\beta)}{k_B}\right]}{\exp\left[\frac{S_R(U_0 - \varepsilon_\alpha)}{k_B}\right]} = \exp\left[\frac{1}{k_B} S_R(U_0 - \varepsilon_\beta) - \frac{1}{k_B} S_R(U_0 - \varepsilon_\alpha)\right]$$

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 :

$$S_R(U_0 - \varepsilon) = S_R(U_0) - \varepsilon \frac{\partial S_R}{\partial U} + \varepsilon^2 \frac{\partial^2 S_R}{\partial U^2} + \dots$$

Equação 27

Sabemos que a derivada parcial 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 são proporcionais à 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 um. Assim, temos que:

$$S_R(U_0 - \varepsilon) \cong S_R(U_0) - \varepsilon \frac{\partial S_R}{\partial U} = S_R(U_0) - \frac{\varepsilon}{T}$$

Equação 28

Lembre-se que T é a temperatura do reservatório. Entretanto, na grande maioria dos casos de interesse, esse sistema estará em equilíbrio térmico com o reservatório, portanto T será também a temperatura do sistema. Esse ponto não é muito enfatizado nos livros, mas é importante lembrar que T é, a priori, a temperatura do reservatório e só podemos considerá-la como a temperatura do sistema também se ambos estiverem em equilíbrio térmico.

Utilizando a expansão da Equação 28 temos:

$$S_R(U_0 - \varepsilon_\beta) - S_R(U_0 - \varepsilon_\alpha) = \frac{\varepsilon_\beta - \varepsilon_\alpha}{T}$$

Equação 29

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

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

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 verdade se for verdade que o conjunto sistema + reservatório tem igual probabilidade de estar em qualquer uma das configurações. Nenhuma configuração é mais provável que a outra. Portanto estamos usando novamente a **suposição fundamental da Física Estatística** (seção 2.1). 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 todas as configurações 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.

Então utilizando apenas a suposição de igual probabilidade para todas as configurações e a lei de conservação de energia, deduzimos que a razão entre as probabilidades de encontrar o sistema em uma configuração β (cuja energia é ε_β) e uma configuração 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). 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 as configurações β e α , poderíamos ter escolhido qualquer par de configurações. Portanto, a razão entre as probabilidades de quaisquer pares é sempre exponencial. Então a probabilidade absoluta deve ser proporcional à exponencial, ou seja:

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

Equação 31

Aonde Z é um número e o índice c foi usado para designar uma configuração qualquer. Para calcular Z , basta lembrarmos que a soma das probabilidades sobre todas as configurações possíveis deve ser igual à 1. Em linguagem matemática temos que:

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

Equação 32

Aqui, a soma a respeito do índice c corresponde à uma soma em relação à todas as configurações possíveis do sistema. Note que estamos tratando apenas das configurações do sistema sozinho (sem o reservatório). A Equação 32 está simplesmente nos dizendo que a probabilidade de eu encontrar um dos estados do sistema na configuração total é de 100%, o que é óbvio, já que a configuração total é composta pelas configurações do sistema combinadas com as do reservatório. No exemplo das 2 moedas, c corresponderia às configurações α , β , ζ ou κ . Portanto haveriam 4 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:

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

Equação 33

Portanto temos que:

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

Equação 34

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

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

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, que é o tópico das próximas seções.

2.4 Potencial químico

Toda a teoria desenvolvida até aqui foi baseada na lei de conservação de energia e na suposição de que todas as configurações são igualmente prováveis. Mas se você notar bem, fizemos uma suposição adicional: supomos que a multiplicidade é função apenas da energia. No exemplo das moedas, estávamos interessados na multiplicidade em função da energia (representada pelo número de coroas viradas pra 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 troca de moléculas também. Assim, faria sentido definir o estado do sistema A (gás no compartimento A) como sendo definido pela energia do sistema A e pelo número de moléculas no compartimento A. Claro que isso valeria também para o sistema B.

O que mudaria nesse caso? Pouca coisa. Praticamente todo o raciocínio que desenvolvemos até aqui anda é válido: só precisamos estendê-lo um pouquinho para incluir a variável número de moléculas. 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 das seções 2.1 e 2.2 ainda são válidos. 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 à U é zero, o que leva ao resultado da Equação 22. Mas agora, temos uma segunda condição que também deve ser satisfeita no ponto de máximo: a derivada parcial em relação à N também tem que ser zero, o que leva naturalmente a um resultado análogo à Equação 22 mas com N no lugar de U (lembre-se que também utilizamos conservação de energia para chegar na Equação 22, então para chegar em uma equação análoga para N temos que utilizar a conservação do número de elementos, que é óbvia: 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 servem somente para lembrar que a derivada parcial é tomada com a outra variável fixa. Então a primeira igualdade significa que a entropia é variada em relação à energia e essa variação se dá com o número de elementos fixo. Enquanto a segunda igualdade significa que 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 variáveis são 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).

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:

$$\mu = -T \left(\frac{\partial S}{\partial N}\right)_U$$

Equação 37

O parâmetro μ definido como 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 N de maneira análoga à definição de temperatura: a temperatura foi definida em termos da derivada da entropia em relação à U e 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, energia irá fluir do sistema com maior temperatura para o sistema com menor temperatura (afinal de contas a temperatura tem que 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 equilíbrio térmico é atingido através da troca de energia (para mudar a temperatura). Do mesmo modo, equilíbrio difusivo é atingido através da troca de elementos (para mudar 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. Portanto um material com potencial químico alto é um material que irá difundir elementos para outro material com potencial químico mais baixo.

Mas existem algumas diferenças. Por exemplo, poderíamos ter mais de um elemento. Neste caso, cada elemento teria o seu próprio potencial químico (e a entropia seria função da energia e do número de cada elemento; por exemplo, se tivéssemos dois elementos diferentes, a entropia seria função de 3 variáveis, duas para os elementos e uma para a energia).

Agora vamos ver outro exemplo peculiar ao potencial químico, e de suma importância para o funcionamento de junções p-n. Imagine duas situações, representadas na Figura 5a e Figura 5b. Representamos os mesmos sistemas em ambas as figuras. 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 tem as mesmas dimensões. Ambos estão conectados por um pequeno afunilamento (que pode ser feito do mesmo material, para facilitar).

Agora imagine o seguinte. 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 algum elemento que possua carga. Você não precisa necessariamente carregar o

material em si para fazer isso. Por exemplo, você pode simplesmente aumentar a concentração de elétrons livres, mas mantendo o sistema neutro. De qualquer forma, você foi e aumentou o potencial químico dos elétrons desse material (você passa agora a ter mais 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 à $\mu_A + \mu_0$. O que vai acontecer? Esses elétrons vão difundir pelo sistema. Você não precisava saber nada de potencial químico para saber que se você espalhar elétrons em um canto de um material eles vão difundir e se espalhar pelo material. Mas agora você sabe que 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. Ou seja, para facilitar a discussão vamos supor que o potencial químico não altera mesmo com a difusão. Então temos que o potencial químico em A é quase constante e igual à μ_A enquanto em B é quase constante e igual à $\mu_A + \mu_0$. Bom, os elétrons estão difundindo do material B para o material A. Portanto existe um fluxo de elétrons de B para A. Uai, se tem fluxo de cargas tem corrente. Se você puser um voltímetro entre os dois materiais vai medir uma tensão que é igual a essa corrente vezes a resistência do sistema. Vamos supor que a tensão medida foi V_0 (medida com o terminal positivo do voltímetro no sistema A), como mostrado na Figura 5a.

Agora congela esse experimento. Separa os materiais e deixa eles quietos por enquanto. Pega o segundo sistema, idêntico ao primeiro e representado na Figura 5b. Mas ao invés de modificar o potencial químico de B, coloca uma diferença de potencial entre A e B de $-V_0$. Então agora temos que tanto o potencial químico do sistema A quanto o de B são μ_A , mas existe uma diferença de potencial 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 será igual à corrente que você mediu antes, já que o módulo da tensão é a mesma. A única diferença é que 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).

O que vai acontecer se você pegar essa bateria, ajustada em V_0 e colocar do mesmo jeito entre o sistema B e A, mas agora com o sistema B com o potencial químico maior, novamente igual à $\mu_A + \mu_0$, como mostrado na Figura 5c? O que vai ocorrer é que a corrente de elétrons de B para A devido à diferença de potencial químico será compensada pela corrente induzida pela diferença de potencial. Assim, a

corrente total será 0 e, conseqüentemente, a tensão medida pelo voltímetro também será 0, mesmo com a bateria conectada (na verdade por causa mesmo da bateria, já que sem a bateria teríamos a situação da Figura 5a).

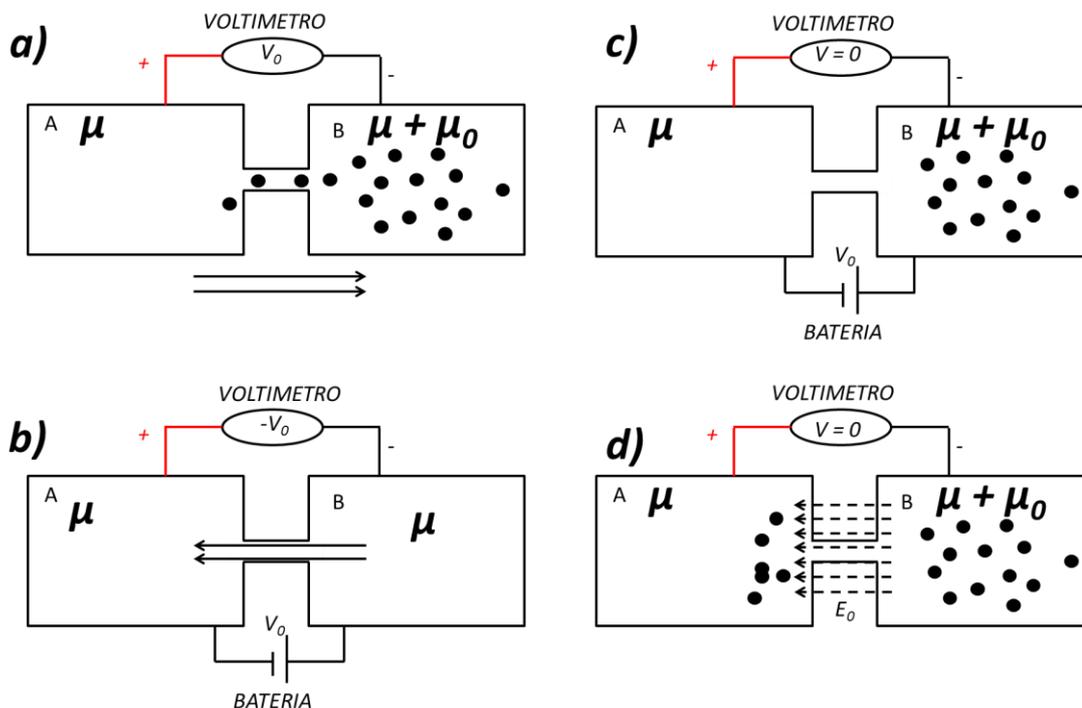


Figura 5 Situações ilustrativas do significado de potencial químico. Em a) aumentamos o potencial químico dos elétrons livres de B, o que induziu uma corrente e conseqüentemente uma diferença de potencial V_0 (medido com o terminal positivo em A). A direção da corrente é indicada pelas setas. Em b), mantemos o potencial químico constante 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 mantemos 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 que carrega o sistema B positivamente (pois agora falta 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 que podemos tirar desse experimento mental? A diferença de potencial químico μ_0 tem o mesmo efeito do potencial elétrico V_0 . Os dois geraram a mesma 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 5a e Figura 5b são idênticas, a não ser pela direção da corrente. 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). 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 (lembre-se que pela definição da Equação 21 a entropia tem a mesma unidade da constante de Boltzmann, que é joules/kelvin; daí podemos deduzir pela definição da Equação 37 que potencial químico tem unidade de energia). Portanto, a interpretação correta para esse experimento é que a bateria fornece a mesma energia para um coulomb que o potencial químico, o que podemos expressar matematicamente como:

$$\mu_0 = -qV_0$$

Equação 38

Aonde q é o módulo da carga fundamental do elétron e o sinal de negativo é consequência do fato de que a potencial eletrostático gera uma corrente com direção oposta ao potencial químico.

Mas o que aconteceria se deixássemos o fluxo de elétrons continuar no primeiro experimento? Retiramos a bateria e observamos o fluxo de elétrons. Afirmamos anteriormente que o fluxo vai continuar até os potenciais químicos se igualarem. Mas não é bem isso que acontece porque aqui temos um novo efeito, advindo do fato de elétrons possuírem cargas. O que vai ocorrer é que começará a ter falta de elétrons no material B e excesso no material A. Esse excesso em A e falta em B cria um campo elétrico entre A e B, como indicado na Figura 5d. Esse campo elétrico vai então exercer uma força na direção de B, o que irá inibir o fluxo de elétrons. Portanto o fluxo vai ocorrer até que a força do campo elétrico iguale a “força” difusiva. Em termos de energia, a corrente irá parar quando o potencial eletrostático induzido pelo campo elétrico iguala 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 , como na Figura 5c. A única diferença é que agora é o campo elétrico interno, que apareceu entre os materiais como consequência da própria difusão, que faz o papel de bateria. Obviamente, o potencial que aparece devido ao campo elétrico é igual à $-V_0$ (medido entre A e B). Se você colocar um voltímetro agora 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 campo elétrico. Portanto a diferença de potencial total é 0, igualzinho ao caso da Figura 5c!

Mas agora temos um problema: temos que o potencial químico de B continua maior que o potencial químico de A, mas não temos difusão (por causa da barreira

imposta pelo campo elétrico). Por outro lado, temos um potencial eletrostático (que vem do campo elétrico), mas também não temos corrente. Isso significa que o que aprendemos sobre potencial químico está errado e o que aprendemos na primeira aula de eletricidade (tensão gera corrente) está também errado? De jeito nenhum, ambos estão corretos, mas ambos causam o mesmo efeito. É como duas forças atuando juntas. 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. É daí que nasce a noção de potencial eletroquímico, que é o potencial químico e elétrico juntos, ou seja, o potencial total. Podemos definir então o potencial eletroquímico como:

$$\mu_{eq} = \mu + qV$$

Equação 39

Onde *eq* significa eletroquímico. Mas não precisamos parar aí. Poderíamos também incluir o potencial gravitacional:

$$\mu_{total} = \mu + qV + mgh$$

Equação 40

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 potencial químico ou potencial 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 elétrico ou gravitacional. 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 o potencial eletrostático qV passou a cancelar 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. Em termos matemáticos, temos que:

$$\mu_A = \mu$$

$$\mu_B = \mu + \mu_0 + qV = \mu$$

Onde μ_A é o potencial total do sistema A, μ_B é o potencial total do sistema B e utilizamos a Equação 38.

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. A origem do potencial químico é a “força” de difusão (força entre aspas porque não existe uma força física neste caso, apenas agitação térmica que causa difusão). A origem do potencial químico não é uma força física, mas sim a variação na concentração dos elementos, mas o efeito é o mesmo. Não tem nada de diferente aqui do conceito de potencial em física newtoniana. Se você tiver um corpo carregado, ele estará sujeito tanto ao potencial gravitacional quanto ao potencial eletrostático. Você sabe também que o corpo entrará em repouso quando os dois potenciais se balancearem. Ambos causam forças, ambos agem juntos, mas as origens deles são diferentes!

Resumindo: daqui pra 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 (no caso do potencial químico, o tipo de transporte é a difusão). Tudo o que dissemos até aqui é válido para o potencial total do sistema. Portanto, daqui em diante, utilizaremos o símbolo μ para designar o potencial total também.

2.5 Função de Partição, Parte II

Como agora estamos caracterizando um estado tanto pela energia quanto pelo número de elementos, a dedução da Função de Partição apresentada na seção 2.3 deve ser estendida para incluir a variável N . O raciocínio é praticamente idêntico ao da seção 2.3. Em particular, a razão entre as probabilidades de se encontrar um sistema em uma configuração β com energia ε_β e número de elementos n_β e uma configuração α com energia ε_α e número de elementos n_α é:

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

Equação 41

Lembre-se que g_R é a multiplicidade do reservatório, e não do sistema. Além disso, do mesmo jeito que U_0 representa a energia total (sistema + reservatório), N_0 representa o número de elementos total (sistema + reservatório).

O procedimento para encontrar a função de partição e a probabilidade $P(c)$ de uma configuração qualquer c é análogo ao da seção 2.3. Primeiro reescrevemos Equação 41 em termos da entropia:

$$\frac{P(\beta)}{P(\alpha)} = \frac{\exp\left[\frac{S_R(U_0 - \varepsilon_\beta, N_0 - n_\alpha)}{k_B}\right]}{\exp\left[\frac{S_R(U_0 - \varepsilon_\alpha, N_0 - n_\alpha)}{k_B}\right]} = \exp\left[\frac{1}{k_B} S_R(U_0 - \varepsilon_\beta, N_0 - n_\beta) - \frac{1}{k_B} S_R(U_0 - \varepsilon_\alpha, N_0 - n_\alpha)\right]$$

Equação 42

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:

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

Equação 43

Utilizando as definições da Equação 23 e da Equação 37, podemos reescrever a Equação 43 em termos da temperatura e potencial:

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

Equação 44

Substituindo a Equação 44 na Equação 42 temos:

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

Equação 45

Novamente, para encontrar o valor absoluto da probabilidade temos que achar a função de partição, que é a soma das exponenciais sobre todas as **configurações** possíveis. Para diferenciar as configurações, adotamos novamente o subscrito c . Assim, cada configuração é determinada por uma combinação única de c . Dessa forma, a soma sobre todas as configurações será:

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

Equação 46

Note que eu alterei o símbolo da soma. Antes era Z (veja Equação 34) e agora é a ζ . Tive que mudar o símbolo porque a soma da Equação 46 recebe um nome diferente, só para complicar a vida mesmo. Apesar de continuar sendo a função de partição, a soma da Equação 46 é chamada de soma de Gibbs. Não satisfeita 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 todas as configurações possíveis (incluindo configurações com zero de energia e/ou zero elementos). 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 do sistema estar em uma dada configuração c com energia ε_c e número de elementos n_c é dada por:

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

Equação 47

Note, portanto, que a probabilidade do sistema estar em uma configuração c depende da energia ε_c , do número de elementos n_c , 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 esta soma depende das configurações e estados possíveis, que são propriedades do sistema.

2.6 Energia e número de elementos de um sistema

A Equação 47 é o resultado mais importante que obtivemos até agora. Ela nos diz que se pegarmos um sistema em equilíbrio com um reservatório (lembre-se que o reservatório é qualquer sistema grande, que pode inclusive ser o próprio planeta Terra), então a probabilidade de encontrar o sistema em uma configuração com energia ε e número de elementos n é dada pela Equação 47. Mas qual é efetivamente a energia e o número de elementos de um sistema em equilíbrio com um reservatório? A resposta da Física Estatística a essa pergunta é: não sei!! Mas eu sei que a probabilidade do sistema ter energia ε e número de elementos n é dada pela Equação 47. Então façamos desse limão uma limonada. Se não podemos determinar exatamente qual é a energia e o número de elementos, vamos usar a Equação 47 para calcular parâmetros estatísticos do sistema (honremos, afinal de contas, o nome Física Estatística). O parâmetro mais comum em física estatística é a média. Por exemplo, a média da energia do sistema:

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

Equação 48

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 ”. Lembre-se que a função P na Equação 48 é dada pela Equação 47. O que estamos fazendo é multiplicando cada energia possível pela sua probabilidade e somando tudo, para achar a média. Talvez você esteja decepcionado de não poder calcular U , mas somente $\langle U \rangle$. Mas lembre-se que estamos tratando de um sistema macroscópico, portanto a função P vai estar extremamente concentrada em torno do valor mais provável, o que significa que a média de U é praticamente igual ao valor de U mais provável. Como a probabilidade do valor mais provável de U é quase igual à 1, então efetivamente estamos calculando o valor de U do sistema: a Física Estatística se torna virtualmente exata!!!

Para tornar o procedimento padrão em Física Estatística um pouco mais palpável, tomemos um exemplo simples: um gás monoatômico em um recipiente fechado e em equilíbrio térmico com o meio ambiente. Suponhamos que queremos calcular $\langle U \rangle$ para este sistema. Neste caso, não há difusão entre o gás e o reservatório (pois o recipiente está fechado). Portanto n é fixo neste exemplo. É importante notar que n é

fixo não porque os potenciais do gás e do reservatório são iguais. Eles não são iguais. Mas n é fixo porque o sistema não está em equilíbrio difusivo com o reservatório (pois o compartimento está fechado): apenas em equilíbrio térmico. Se você abrir o compartimento, aí os potenciais do sistema e do reservatório serão igualados. O fato de que os sistemas não estão em equilíbrio difusivo significa que não podemos determinar o número de elementos do sistema utilizando Física Estatística, pois os elementos do sistema não estão interagindo com o reservatório. Portanto temos que saber a priori qual é o número de elementos do sistema (em geral, podemos deduzir isso pela temperatura, pressão e volume do gás)!! Isso significa que temos que jogar fora tudo o que estudamos até aqui? Claro que não. Isso só significa que, como a única interação é térmica, podemos utilizar a Equação 31 ao invés da Equação 47 para a probabilidade de U . Portanto neste caso a média seria dada por:

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

Equação 49

O que falta então para acharmos U ? Faltam duas coisas relacionadas: precisamos achar quais são as configurações e seus níveis de energia possíveis. Se soubermos isso, podemos calcular tanto a função de partição como a soma da Equação 49. O cálculo das configurações e níveis de energia é um problema típico em mecânica quântica e é completamente impossível aprofundar nessa questão fora de um curso dedicado a mecânica quântica. Podemos, entretanto, dar um exemplo simples: o exemplo de um único átomo em uma cavidade. Primeiro vamos lembrar que um único átomo não é um sistema macroscópico. Isso significa que a distribuição de probabilidade não vai estar tão concentrada em torno do valor mais provável, então temos que levar em conta todas as configurações possíveis.

Como o átomo é representado por uma função de onda, então podemos utilizar a Equação 8 para determinar os níveis de energia possíveis para a cavidade. Mas a Equação 8 foi deduzida para uma cavidade unidimensional. Uma cavidade propriamente dita, entretanto, tem que ser tri-dimensional. Felizmente, o raciocínio para obter a relação de energia para uma cavidade tri-dimensional é praticamente idêntico ao caso uni-dimensional. A única diferença é que requeremos que a fase acumulada em uma volta seja igual à um múltiplo de 2π para cada dimensão

independentemente. Assim, temos que, se a função de onda tri-dimensional for dada por:

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

Equação 50

Então a condição de acúmulo de fase para as três dimensões independentemente será:

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

Equação 51

Onde c_x , c_y e c_z , são números inteiros e L_x , L_y e L_z são as dimensões da caixa onde o átomo 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 52

O momento continua sendo proporcional à constante de propagação. A única diferença aqui do que foi discutido para o caso uni-dimensional é que o momento 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 53

Aqui utilizamos a notação que flecha representa vetor e $\hat{}$ representa vetor unitário. Em qualquer dimensão, a energia cinética é relacionada ao módulo quadrado do vetor momento por:

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

Equação 54

De onde concluímos que:

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

Equação 55

Que é a versão em três dimensões da Equação 8. Cada combinação de c_x , c_y e c_z corresponde a uma configuração diferente. Por exemplo, a combinação $c_x = 1$, $c_y = 0$, $c_z = 0$ corresponde a uma configuração com energia igual à $h^2/(8L^2m)$, enquanto a combinação $c_x = 0$, $c_y = 1$, $c_z = 0$ corresponde a uma outra configuração, mas também com energia igual à $h^2/(8L^2m)$. Já a combinação $c_x = 1$, $c_y = 1$, $c_z = 0$ corresponde a uma configuração com energia $2h^2/(8L^2m)$, e assim por diante.

Assim, temos que a função de partição será:

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

Equação 56

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 zero a infinito (a combinação $c_x = 0$, $c_y = 0$, $c_z = 0$, entretanto, não entra no somatório, pois ela corresponderia a um nível de energia igual à zero, o que por sua vez corresponderia a nenhum átomo na cavidade). Em geral, essa conta é feita aproximando o somatório a uma integral. Como não vamos ganhar muita intuição fazendo essa conta, já vou colocar o resultado direto (quem quiser ver detalhes do procedimento de uma olhada na bibliografia sugerida no final deste capítulo):

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

Equação 57

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

Podemos agora calcular a média da energia do átomo:

$$\langle U \rangle = \sum_c \varepsilon_c P(c) = \frac{\sum_{c_x} \sum_{c_y} \sum_{c_z} \frac{h^2}{8L^2 m} (c_x^2 + c_y^2 + c_z^2) \exp\left(-\frac{\frac{h^2}{8L^2 m} (c_x^2 + c_y^2 + c_z^2)}{k_B T}\right)}{Z}$$

Equação 58

Onde a função de partição Z é dada pela Equação 57. Se você fizer essa conta, vai encontrar que:

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

Equação 59

Portanto, a média da energia para 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 $(3/2)k_B T$. A prova é essa aí.

Esse é o exemplo mais simples de como calculamos grandezas macroscópicas em Física Estatística.

Podemos facilmente generalizar o raciocínio para incluir contato difusivo, além de térmico. Para isso basta utilizarmos a soma de Gibbs ao invés 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:

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

Equação 60

e

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

Equação 61

2.7 Distribuição de Fermi-Dirac

Agora vamos focar nossa atenção em um problema mais específico e de relevância direta em semicondutores e eletrônica. Vamos supor novamente que temos um sistema, pode ser um gás, um líquido ou um sólido, tanto faz. Agora estamos interessados nos elétrons do sistema.

Como mencionado na seção 2.1, elétrons ocupam estados discretos de energia. Vamos voltar nesse problema no capítulo 3, mas aqui só precisamos lembrar que cada elétron no sistema estará em um determinado estado (não esquecendo que diferentes estados podem ter a mesma energia). Pelo princípio de exclusão de Pauli, sabemos também que cada estado só poderá ser ocupado por um único elétron (alguns livros usam o termo orbital ao invés de estado para enfatizar que agora estamos tratando especificamente de estados de energia para elétrons). O que queremos aqui é calcular a probabilidade de um determinado estado estar ocupado ou não.

Existem várias maneiras de calcular a probabilidade de um único estado sujeito ao princípio de exclusão de Pauli estar ocupado. Uma das maneiras mais simples (veja as sugestões bibliográficas no final desse capítulo) utiliza os resultados da seção 2.5. Para tal, basta interpretarmos o estado em questão como sendo o sistema de interesse, ou seja, o estado quântico passou a ser o sistema de interesse. Todos os outros estados passam a ser o reservatório. Então estamos para um determinado estado. Vamos supor que a energia deste estado seja ε_0 (lembre-se que outros estados podem ter a mesma energia, mas não precisamos nos preocupar com isso aqui). Quais são as configurações possíveis desse estado (ou, equivalentemente, quais são as configurações possíveis do sistema)? Pelo princípio de exclusão de Pauli, temos apenas duas configurações possíveis: ou o estado está ocupado com um elétron ou está desocupado. Portanto temos apenas duas opções para n : $n = 0$ ou $n = 1$. Mas se o estado estiver desocupado, então teremos $\varepsilon = 0$ (já que a energia vem do elétron, e não do estado em si). Por outro lado, se o estado estiver ocupado, terá energia $\varepsilon = \varepsilon_0$. Assim, temos apenas duas configurações possíveis: a configuração correspondendo ao par $n = 0$, $\varepsilon = 0$ e a configuração correspondendo ao par $n = 1$, $\varepsilon = \varepsilon_0$. Assim, a soma de Gibbs será:

$$\zeta(T, \mu) = \exp\left(-\frac{0-0\mu}{k_B T}\right) + \exp\left(-\frac{\varepsilon_0 - 1\mu}{k_B T}\right) = 1 + \exp\left(-\frac{\varepsilon_0 - \mu}{k_B T}\right)$$

Equação 62

Com a soma de Gibbs, podemos calcular a probabilidade do estado estar ocupado:

$$f(\varepsilon = \varepsilon_0, n = 1) = \frac{\exp\left(-\frac{\varepsilon_0 - \mu}{k_B T}\right)}{1 + \exp\left(-\frac{\varepsilon_0 - \mu}{k_B T}\right)} = \frac{1}{1 + \exp\left(\frac{\varepsilon_0 - \mu}{k_B T}\right)}$$

Equação 63

Aqui, mudamos o símbolo de probabilidade de P para f porque queremos enfatizar que essa é uma probabilidade específica, ou seja, que f é a probabilidade de um estado estar ocupado por um elétron.

Fizemos o cálculo para um estado com uma determinada energia ε_0 , onde eu mantive o subscrito 0 para enfatizar que estamos olhando para uma única configuração. Mas como poderíamos ter feito o mesmo cálculo para qualquer configuração com qualquer energia, não precisamos do índice 0 para expressar o resultado. Além disso, como qualquer estado ocupado terá $n = 1$, também não precisamos explicitar que a probabilidade se refere a um estado com $n = 1$, basta lembrarmos que estamos calculando a probabilidade do estado estar ocupado (que é o mesmo que dizer que estamos calculando f para $n = 1$). Portanto concluímos que a probabilidade de um estado com energia ε estar ocupado é:

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

Equação 64

Esse importante resultado é a distribuição de Fermi-Dirac.

A maioria dos livros sobre semicondutores partem da distribuição de Fermi-Dirac para deduzir a relação tensão x corrente em dispositivos semicondutores. Esse procedimento será mostrado nos capítulos seguintes dessas notas de aula. 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 estado com energia ε estar ocupado, mas podemos dar aqui um exemplo particularmente importante em eletrônica a respeito de como utilizar a Equação 64. Um dos parâmetros mais relevantes em eletrônica é a condutividade dos materiais. Sabemos que a condutividade depende do número de portadores de cargas (como elétrons) livres no material. Portanto, para calcularmos a condutividade, precisamos primeiro achar o número de portadores livres. Se soubermos que um material possui $N_\varepsilon(\varepsilon)$ estados quânticos com energia ε para os quais os portadores são livres, e que a probabilidade de cada um desses $N_\varepsilon(\varepsilon)$ estados estarem ocupados é $f(\varepsilon)$, então podemos concluir que o número total de portadores de carga livre será:

$$N = \sum_{\varepsilon_c} N_\varepsilon(\varepsilon_c) f(\varepsilon_c)$$

Equação 65

Esse é um típico e importante exemplo de como a distribuição de Fermi-Dirac é útil. Em particular, a Equação 65 relaciona o número de portadores livres (lado esquerdo) com o potencial μ do sistema (que entra no lado direito pela distribuição de Fermi-Dirac). Portanto, sabendo μ podemos encontrar N e vice-versa. Como veremos nos capítulos seguintes, na verdade essa relação é geralmente expressa em termos de densidade de portadores e a soma é aproximada por uma integral, mas a ideia é a mesma: relacionar número ou densidade de portadores com o potencial μ .

Antes de terminarmos essa seção, vamos explicitar duas propriedades da distribuição de Fermi-Dirac que ajudam a aumentar a intuição sobre o seu significado. Primeiro, imagine que temos um sistema com temperatura $T = 0$, chamada de zero absoluto. Neste caso a exponencial no denominador 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 a 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 (lembre-se que μ é função da temperatura) 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 66

Infelizmente, existe uma lambança gigante a respeito do uso desse termo “nível de Fermi”. A definição correta para nível de Fermi é a da Equação 66, 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 !! Viver é muito perigoso.

Finalmente, note que podemos interpretar o potencial μ do sistema como correspondendo a o nível de energia para o qual a probabilidade de ocupação é igual à 0.5.

2.8 Transporte

A última seção desse 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 mais simples e inclusive vamos utilizar a noção de corrente de arrasto para explicar a corrente de difusão, que é o assunto principal dessa seção.

A maioria dos livros começa o tratamento sobre difusão partindo da chamada lei de Fick, que na verdade não é bem uma lei, mas uma expressão semi-empírica para a corrente de difusão. Não existe uma dedução formal para a lei de Fick, mas vou tentar chegar nessa expressão por meio de argumentos semi-quantitativos para podermos entender melhor de onde ela vem e o que ela expressa. Para o nosso propósito, basta chegarmos na lei de Fick, portanto o final dessa seção é na verdade o começo do tratamento de transporte em livros de Física Estatística e Termodinâmica. Nos próximos capítulos vamos utilizar a lei de Fick para calcular a corrente em uma junção p-n.

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

Por definição, a corrente é o número de partículas que atravessam uma determinada área por segundo e a densidade de corrente é o número de partículas por segundo e por unidade de área.

Vamos supor que nós temos um grupo de partículas percorrendo um tubo com área A e todas as partículas possuem velocidade v , como mostrado na Figura 6. Queremos encontrar uma expressão 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 sabermos a resposta, basta identificarmos qual é a distância máxima que a partícula pode percorrer 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 67

Onde ρ é a densidade volumétrica de partículas. Portanto, a corrente I que passa por A será:

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

Equação 68

Consequentemente, a densidade de corrente J será:

$$J = \rho \cdot v$$

Equação 69

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.

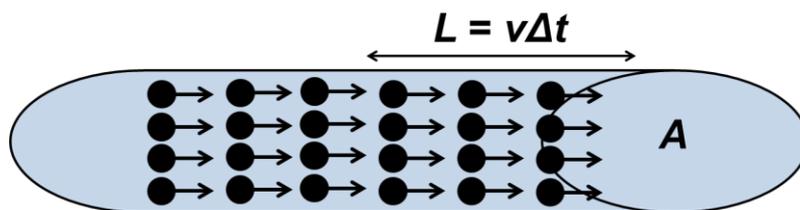


Figura 6 Corrente de arrasto em um tubo com área A .

Agora vamos supor que exista uma 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 da corrente possa variar, mas para facilitar vamos supor que a direção 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 cada

grupo possui uma determinada velocidade. Designando por ρ_n a densidade de partículas com velocidade v_n , temos que a densidade de corrente total será:

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

Equação 70

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 , temos então que:

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

Equação 71

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

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

Equação 72

Onde ρ é a densidade total. Utilizando a Equação 72, podemos reescrever a Equação 70 como:

$$J = \rho \langle v \rangle$$

Equação 73

Essa é uma expressão mais geral para a corrente de arrasto. Note que se tivermos metade das partículas indo pra direita e metade pra esquerda com a mesma velocidade, então a média das velocidades será zero e consequentemente J também será 0. Essa é uma característica típica da corrente de arrasto: se a média das velocidades for zero, a corrente de arrasto também é zero. Mas note o seguinte: se a

média das velocidades for zero, a corrente em qualquer área dentro do tubo será realmente zero, mas a corrente na boca do tubo não vai ser zero. Obviamente, partículas irão escapar do tubo, mesmo se a média das velocidades das partículas dentro do tubo for zero. É claro que isso só ocorre na boca do tubo, porque temos partícula só de um lado da boca (que é a área onde estamos calculando a corrente), enquanto o outro lado não tem partículas. Em linguagem mais técnica, existe um gradiente (uma variação), na densidade de partículas bem em cima da boca, pois fora do tubo a densidade é zero. Então a corrente é zero em todos os pontos do tubo menos na boca. 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 73 supõe que a densidade é constante ao longo do tubo, portanto a Equação 73 não leva em consideração a corrente de difusão. De fato, como já afirmamos anteriormente, a Equação 73 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 é nula. Caso isso não seja verdade, podemos sempre dividir 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 7. 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 somar as duas.

Vamos focar agora então em determinar J_{e-d} . A priori, podemos usar o mesmo raciocínio que nos levou à Equação 69, 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 irão trombar entre si, alterando a direção da velocidade. Entretanto, o raciocínio que levou à Equação 69 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, que chamaremos de l_z , como na Figura 7, já que dentro dessa distância não haverá muitas colisões.

Uma vez determinado o comprimento da caixa, temos que abordar um segundo problema: agora temos partículas em todas as direções, portanto temos que pegar a contribuição de cada uma dessas componentes para a corrente J_{e-d} . Se uma partícula possui direção com ângulo θ em relação ao eixo z , então a projeção da velocidade no eixo z será:

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

Equação 74

De maneira análoga, a projeção do caminho médio no eixo z será:

$$l_z = l \cos \theta$$

Equação 75

Onde l é caminho médio livre.

Assim, a contribuição para a corrente ΔJ_{e-d} das partículas com direção θ em relação ao eixo z será:

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

Equação 76

Aqui, $\rho(z_0 - l_z, \theta)$ é a densidade de partículas na caixa da esquerda com velocidade na direção θ . O procedimento para encontrar ΔJ_{d-e} é o mesmo, mas agora temos que usar a densidade na caixa da direita. Portanto:

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

Equação 77

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

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

Equação 78

Mas

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

Equação 79

Portanto:

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

Equação 80

Note que a Equação 79 subentende que a derivada é avaliada no centro da caixa, ou seja, em $z = z_0$, que é exatamente o ponto onde estamos calculando a corrente. Substituindo a Equação 74 e Equação 75 na Equação 80 temos:

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

Equação 81

Como as partículas possuem direção aleatória, a densidade de partículas que apontam para um determinado ângulo sólido $d\Omega$ será:

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

Equação 82

Onde $\rho(z)$ é a densidade total de partículas, ou seja, a soma de todas as partículas, independente da direção de velocidade. Se você achou a Equação 82 misteriosa, lembre-se que 4π nada mais é do que a integral de $d\Omega$ sobre toda a esfera. Portanto a Equação 82 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. Substituindo a Equação 82 na Equação 81 temos:

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

Equação 83

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

$$J = -\frac{2 \int \frac{\partial \rho}{\partial z} \cdot |v| \cdot l \cdot \cos^2 \theta \cdot d\Omega}{4\pi} = -\frac{2 \frac{\partial \rho}{\partial z} \cdot |v| \cdot l \cdot \int_0^{2\pi} \int_0^{\frac{\pi}{2}} \cos^2 \theta \cdot \sin \theta \cdot d\theta \cdot d\phi}{4\pi}$$

Equação 84

Note que a integral é feita sobre a metade superior da esfera, uma vez 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:

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

Equação 85

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 :

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

Equação 86

Assim, a corrente de difusão pode ser expressa em termos de D como:

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

Equação 87

Lembre-se que a Equação 87 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). Em linguagem matemática temos que o vetor corrente de difusão é definido por:

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

Equação 88

Onde as suas componentes são:

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

Equação 89

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

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

Equação 90

A Equação 90 será utilizada nos próximos capítulos. 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 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 a medida que andamos para a direita, o que é o mesmo que afirmar que a densidade na caixa esquerda é maior que na direita).

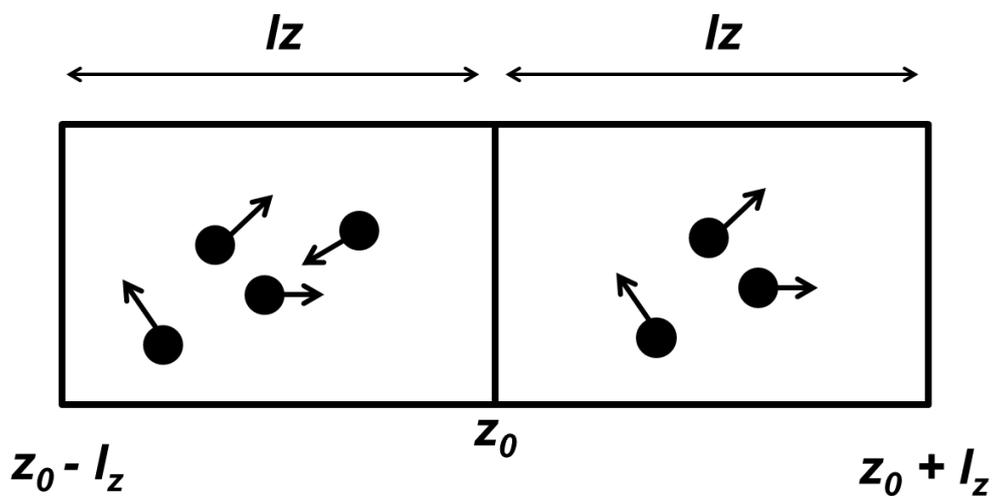


Figura 7 Esquema para o cálculo da corrente de difusão

2.9 Sugestões bibliográficas

Para o propósito desse curso, o livro que considero mais adequado e do qual procurei seguir a notação na medida do possível é Charles Kittel: *Thermal Physics*. Esse 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, eu recomendo também Ashley Carter: *Classical and Statistical Thermodynamics*. Em minha opinião, esse último é o livro mais claro e bem escrito sobre esse assunto. Entretanto, ele não é tão conhecido assim, então é um pouco mais difícil de encontrar. Caso o aluno estude ou 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, eu recomendo o primeiro capítulo do livro 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.

3 – Semicondutores

Neste capítulo vamos utilizar as noções do capítulo anterior 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.

3.1 Teoria de Bandas

Vimos no capítulo anterior que os estados de energia de partículas quânticas são discretos. Como exemplo, calculamos os níveis de energia de um único 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: ao invés do poço quântico, o elétron está preso devido à energia potencial elétrica imposta pelo núcleo, o que resulta em estados de energia discretos. O cálculo dos estados de energia neste último caso é bem mais complicado e requer a solução da função de onda do elétron para este potencial em particular (a função de onda é chamada de Equação de Schrodinger), mas o resultado é qualitativamente o mesmo: o elétron terá níveis de energia discretos. Estes níveis de energia são os orbitais químicos *s*, *p*, *d* e *f*.

Mas aqui 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 dessa 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*), mas 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 temos 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 no orbital de energia mais baixo. Pergunta: podemos encontrar um sistema clássico que seja matematicamente e fisicamente análogo ao átomo de hidrogênio?

Vamos pensar no caso da 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 à estes requisitos é caso do 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 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 Hook). Além disso, o elétron oscila em uma frequência bem definida determinada pela energia do orbital, enquanto a massa oscila na frequência de ressonância.

E o que podemos dizer em relação à analogia matemática? Todo problema quântico se reduz à solução de uma equação de onda (Equação de Schrodinger). 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 Schrodinger, 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 a 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 à analogia 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 8 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 $x = a$.

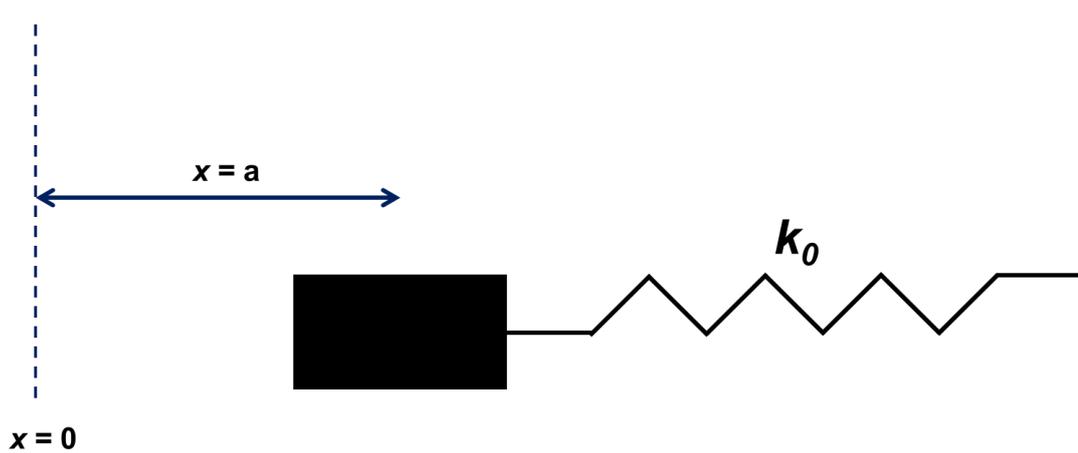


Figura 8 Oscilador harmônico exemplo. O ponto $x = a$ corresponde ao ponto de repouso (a mola não está nem esticada nem comprimida e a força total neste ponto é zero).

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

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

Equação 91

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

$$x = A \cos(\omega_0 t + \phi)$$

Equação 92

A fase e amplitude da senóide 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 93

Se você já sabia tudo isso de cabeça, ótimo. Se, por outro lado, você não lembra muito bem da teoria do oscilador harmônico, então eu sugiro que você substitua a Equação 92 na Equação 91 para verificar se ela realmente é uma solução, com a frequência de ressonância dada pela Equação 93.

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 9 ilustra o sistema completo, onde o acoplamento é dado pela mola com constante Ω .

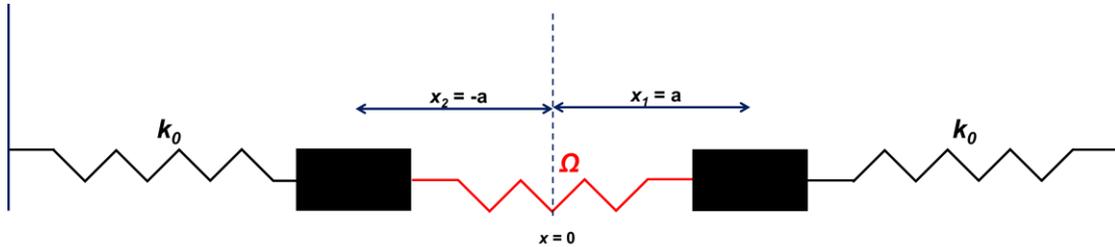


Figura 9 Dois osciladores harmônicos acoplados. A mola em vermelho com constante Ω representa o acoplamento

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:

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

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

Equação 94

A Equação 94(1) corresponde à segunda lei de Newton aplicada à massa da direita, cuja coordenada chamamos de x_1 , enquanto a Equação 94(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 à b , não haverá força advinda dessa mola). Confira sozinho se os sentidos das forças na Equação 94 estão corretos.

Para limpar a notação um pouco, vamos re-escrever a Equação 94 como:

$$\frac{\partial^2 x_1}{\partial t^2} + \omega_0^2(x_1 - a) + \gamma^2(x_1 - x_2 - b) = 0 \quad (1)$$

$$\frac{\partial^2 x_2}{\partial t^2} + \omega_0^2(x_2 + a) - \gamma^2(x_1 - x_2 - b) = 0 \quad (2)$$

Equação 95

Onde ω_0 é dado pela Equação 93 e $\gamma^2 = (\Omega/m)$. As equações acima estão acopladas (a primeira equação depende de x_2 e a segunda equação depende de x_1), portanto a solução delas não será igual à do sistema isolado. O método geral de solução de equações acopladas envolve encontrar os autovetores e autovalores do sistema, o que equivale à 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:

$$\frac{\partial^2(x_1 + x_2)}{\partial t^2} + \omega_0^2(x_1 + x_2) = 0$$

Equação 96

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

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

Equação 97

Substituindo a Equação 97 na Equação 96 temos:

$$\frac{\partial^2(x_{CM})}{\partial t^2} + \omega_0^2(x_{CM}) = 0$$

Equação 98

O que a Equação 98 nos diz? Ela é idêntica à equação de um oscilador harmônico simples (compare com a Equação 91) 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 ao invés de somarmos? Neste caso teremos:

$$\frac{\partial^2(x_1 - x_2)}{\partial t^2} + \omega_0^2(x_1 - x_2) - 2\omega_0 a + 2\gamma^2(x_1 - x_2 - b) = 0$$

Equação 99

Definindo a distância relativa entre as massas como:

$$x_R = x_1 - x_2$$

Equação 100

Temos:

$$\frac{\partial^2(x_R)}{\partial t^2} + (2\gamma^2 + \omega_0^2)x_R - (2\omega_0^2 a + 2\gamma^2 b) = 0$$

Equação 101

Como o último termo da Equação 101 é uma constante, essa também é a equação de um oscilador harmônico, com frequência de ressonância igual à:

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

Equação 102

De fato, confirme sozinho que a solução da Equação 101 é dada por:

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

Equação 103

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 do modo do centro de massa com o modo da distancia relativa, mas o que nos interessa aqui é que ao invés de termos uma única frequência de ressonância, agora temos duas frequências de ressonância: 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), compartilhando a mesma frequência de oscilação (a propósito, note pela Equação 102 que ω_R tende à ω_0 quando o acoplamento tende à zero). Ou seja, temos duas configurações com a mesma frequência (cada configuração 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 10a. Analogamente, se temos dois átomos separados, cada um terá a sua energia individual, e elas serão idênticas. Mas se os dois átomos

interagirem, para cada orbital independente do átomo isolado, irão aparecer dois orbitais diferentes. Se três átomos interagirem, cada orbital se repartiria em três, e assim por diante, como representado na Figura 10b. Mas um sólido possui muitos, muitos átomos. Por exemplo, se tivermos um mol de átomos, teremos 1 mol de níveis de energia para cada orbital do átomo isolado. Mas esses níveis estarão tão próximos um dos outros que podemos tratá-los como um contínuo de níveis, como representado na Figura 10b. Cada um desses contínuos é chamado de banda de energia.

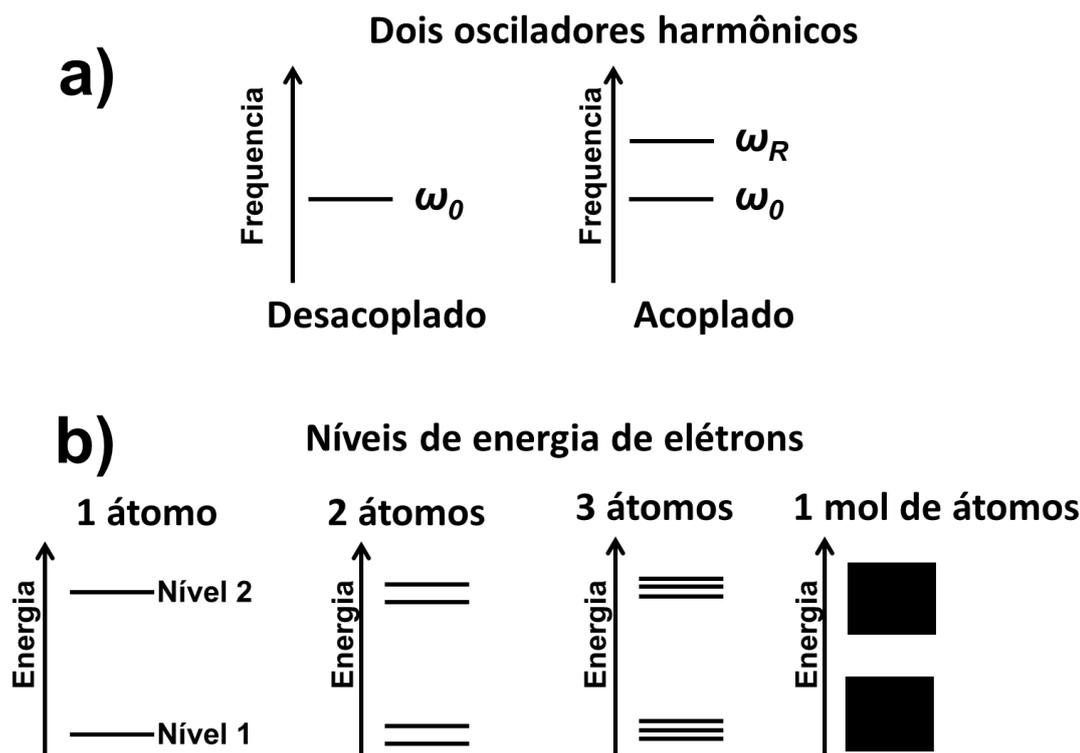


Figura 10 a) separação da frequência de ressonância do oscilador harmônico acoplado. b) efeito análogo na interação entre 2 ou mais átomos: se dois átomos interagem, cada nível de energia é separado em 2 níveis com energias diferentes. Se três átomos interagem, teremos três níveis de energia para cada nível do átomo isolado. Se 1 mol de átomos interagem, temos tantos níveis de energia tão próximos um dos outros que podemos tratá-los como um contínuo de níveis, formando bandas de energia.

Na Figura 10b representamos um exemplo onde 2 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 no nível 1. Se colocarmos dois átomos juntos, teremos dois elétrons, que estarão nos níveis de energia mais baixo, que estão relacionados com o nível 1. Se tivermos muitos, muitos átomos, de maneira que os níveis possam ser tratados como bandas, então teremos a primeira banda da Figura 10b preenchida e a segunda desocupada. É melhor, então, representar

o preenchimento das bandas “colorindo” a parte que está ocupada, como mostrado na Figura 11. Nesta 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.

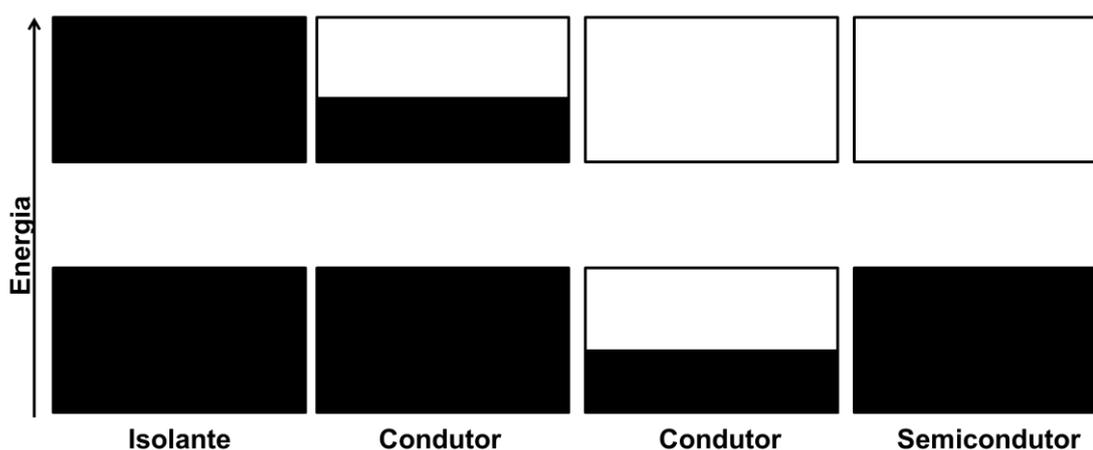


Figura 11 A ocupação das bandas de energia define as propriedades de condutividade do sólido.

Como já indicado pela Figura 11, 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 a razão do primeiro caso ser um isolante: se aplicarmos um campo elétrico no material, o campo iria acelerar as cargas, ou seja, iria aumentar a energia delas, mas como não existem níveis de energia desocupados, não tem como as cargas serem aceleradas e por isso elas não respondem ao campo elétrico. 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 a banda está completamente preenchida), mas como veremos a seguir podemos facilmente alterar as propriedades de condução deste material através de reações químicas (dopagem) com outros átomos.

3.2 Elétrons e buracos

Dissemos na seção anterior que o semicondutor é caracterizado por duas bandas de energia (banda de valência e a banda de condução), com a banda inferior completamente preenchida e banda superior completamente vazia. Essa condição, entretanto, só é perfeitamente satisfeita no limite de temperatura igual à zero kelvins (zero absoluto). De fato, alguns portadores são excitados termicamente da banda de valência para a banda de condução, induzindo um 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 seja, eles contribuem para a corrente no semicondutor. Dizemos então que esses são elétrons “livres”, pois existem níveis de energia para os quais o elétron pode ser acelerado e também porque tipicamente eles apresentam fraca interação com a rede cristalina do semicondutor. E os elétrons da banda de valência? Esses elétrons estão geralmente bem confinados na estrutura cristalina, mas eles podem “pular” de um átomo para o outro, desde que o átomo possua estados disponíveis na banda de valência para receber esse elétron extra. A banda de valência, entretanto, estará totalmente preenchida, a não ser que um elétron tenha sido excitado para a banda de condução. Se isso ocorrer, o átomo que perdeu o elétron terá uma carga positiva e um nível de energia disponível. O elétron do átomo vizinho poderá, portanto, pular para o átomo com um espaço vazio. Se isso ocorrer, o átomo vizinho terá a carga positiva, como ilustrado na Figura 12. O que efetivamente ocorreu, então, é um deslocamento da carga positiva. Como quem define a corrente é o deslocamento de cargas, o movimento do espaço vazio corresponde ao movimento de uma partícula carregada positivamente. Essa “partícula” é chamada de buraco, para lembrar que ela não é bem uma partícula, mas sim uma falta de partícula, que resulta em uma carga positiva no átomo. 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 pra 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.

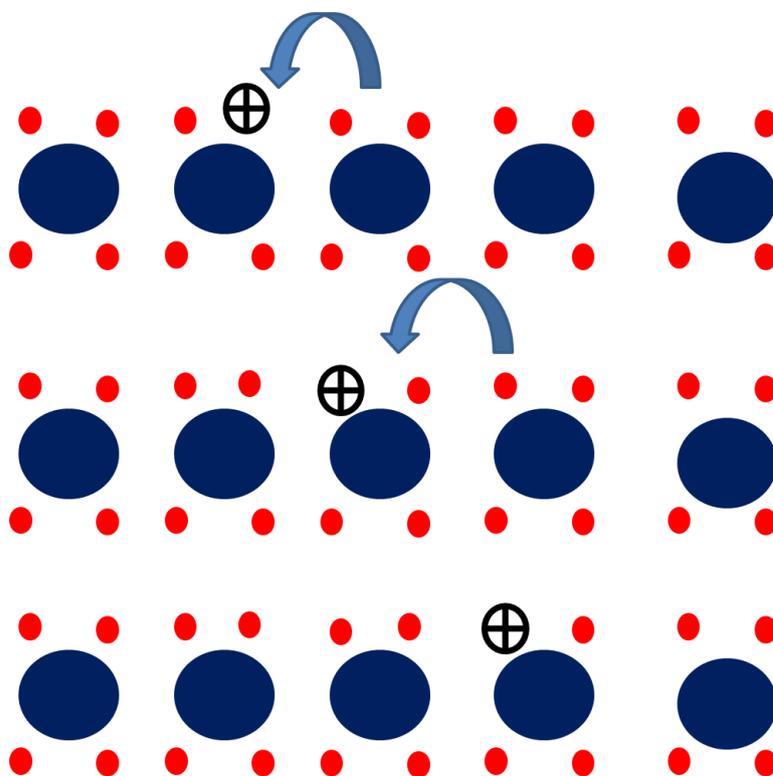


Figura 12 Representação do movimento de buracos. Representamos 4 elétrons (pontos vermelhos) preenchendo a banda de valência. O segundo átomo da esquerda para a direita perdeu um elétron na banda de valência por agitação térmica (foi excitado para a banda de condução). A falta do elétron constitui uma carga positiva (o buraco), que desloca para a direita à medida que os elétrons dos átomos vizinhos pulam para o átomo desocupado. Note que o semicondutor como um todo não está carregado, pois o elétron que foi excitado para a banda de condução (e gerou o buraco) ainda está no semicondutor. Mas do mesmo jeito que este elétron na banda de condução é livre para se mover, o buraco que ele deixou na banda de valência também é.

3.3 Band-gap e concentração de portadores em semicondutores intrínsecos

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.

Na verdade, já sabemos como fazer isso: na seção 2.7 nós calculamos a distribuição de Fermi-Dirac, que é a probabilidade de um estado de energia estar ocupado por um elétron (lembre-se que esses estados de energia são os orbitais atômicos). Então nós já sabemos qual é a probabilidade de cada um dos níveis de energia da banda de condução estar ocupado. Para saber qual é a concentração total,

basta somarmos a probabilidade de cada um dos níveis. 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 níveis de energia na banda de condução, cada um com probabilidade igual à 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 estados com probabilidade 0.5 de ocupação corresponde a um estado efetivamente ocupado. Em outras palavras, essa situação corresponde a um elétron na banda de condução. Note que eu fiz um salto entre a probabilidade e a realidade: como eu sabia que cada estado tinha metade de chance de estar ocupado, considerar que um deles está ocupado e o outro vazio. Se eu tiver 4 estados com probabilidade 0.5 cada, vou considerar que tenho 2 estados ocupados. É claro que um raciocínio desses só funciona quando temos muito, muitos estados, 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 do que é a distribuição de Fermi-Dirac: ela se refere à probabilidade de um único estado estar ocupado. O que tem que dar 1 é a probabilidade do estado estar ocupado mais a probabilidade do estado estar desocupado. Quando somamos a probabilidade de um estado estar ocupado com a probabilidade de outro estado 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 portadores na banda de condução N_e será a soma das probabilidades de ocupação de todos os níveis de energia na banda de condução. Em termos matemáticos:

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

Equação 104

Note que eu troquei o índice c pelo índice bc . Fiz isso para enfatizar que estamos somando somente os níveis 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 pra frente calcularemos N_p , que é o número total de buracos).

Nós 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 104 estar correta; e também qual é o raciocínio por trás dela (como eu tentei explicar no exemplo que atribui uma carga à dois estados com probabilidade 0.5). É mais importante você entender de onde vem a Equação 104 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 textos de semicondutores. Como falei na seção 2.7, apesar do 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 . Eu não sei por que trocaram o nome, mas o símbolo foi trocado porque geralmente μ é utilizado para o parâmetro “mobilidade” no contexto de semicondutores, que veremos mais adiante. Então vamos nos dobrar ao peso da sociedade e chamarmos o potencial eletroquímico de nível de Fermi, trocando também o símbolo. Assim, a Equação 104 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 105

Além disso, precisamos definir mais três termos, bem simples, e que estão representados na Figura 13: o nível da banda de condução ε_C , o nível da banda de valência ε_V , e o *band-gap* ε_G . Por definição, ε_C é o nível de energia mais baixo da banda de condução, ε_V é o nível mais alto da banda de valência e ε_G é a diferença entre eles:

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

Equação 106

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. Alguns textos em português chamam o *band-gap* de banda proibida, mas eu vou me dar ao luxo de manter o termo em inglês, mesmo porque os melhores livros textos estão em inglês.

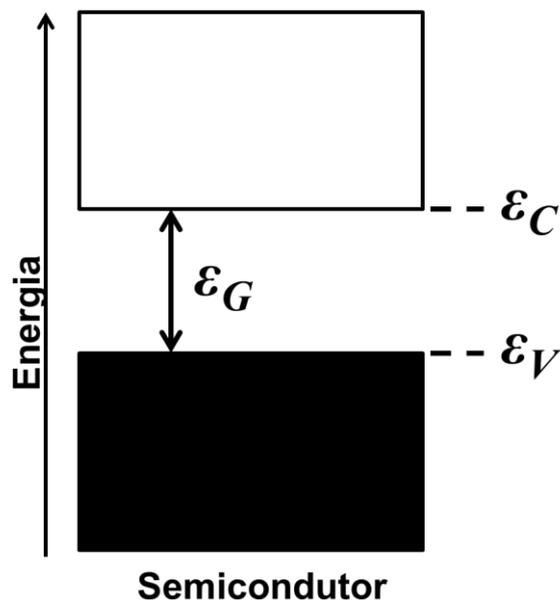


Figura 13 Definições de ϵ_V , ϵ_C e ϵ_G

Definidos os termos, podemos passar para o cálculo de N_e . À primeira vista, essa é uma tarefa monumental: teríamos que calcular todos os níveis de energia da banda de condução para podermos achar N_e . Felizmente, podemos fazer uma aproximação incrivelmente simples: podemos tratar os elétrons livres como se fosse 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 no modelo como uma modificação da massa do elétron, 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 rede e essa dependência é expressa pela massa efetiva. Em alguns casos, 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 o nosso problema se reduz ao cálculo do número de partículas em um gás de elétrons, com a massa substituída pela massa efetiva. Assim, os níveis quânticos da soma na Equação 105 são os mesmo do gás ideal e nós já inclusive calculamos esses níveis de energia (veja a Equação 55). Lembre-se que cada conjunto de números inteiros c_x , c_y e c_z corresponde a uma configuração, ou seja, a um estado diferente. Na verdade, 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 configuração é caracterizada 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 par corresponde a duas configurações eletrônicas (uma com spin *up* e outra com spin *down*). Por essa razão, temos que multiplicar a soma por 2. Assim temos:

$$N_e = \sum_{bc} f(\varepsilon_{bc}) = \sum_{bc} \frac{1}{1 + \exp\left(\frac{\varepsilon_{bc} - \varepsilon_F}{k_B T}\right)} = 2 \sum_{c_x} \sum_{c_y} \sum_{c_z} \frac{1}{1 + \exp\left[\frac{(\varepsilon_{c_x, c_y, c_z} + \varepsilon_C) - \varepsilon_F}{k_B T}\right]}$$

Equação 107

Onde a energia $\varepsilon_{c_x, c_y, c_z}$ é dada pela Equação 55 com a massa substituída pela massa efetiva. Você deve ter notado que a energia ε_{bc} foi expressa como $\varepsilon_{bc} = \varepsilon_{c_x, c_y, c_z} + \varepsilon_C$. O termo $\varepsilon_{c_x, c_y, c_z}$ vem pelo fato de que estamos tratando os elétrons como um gás e essa é a energia do gás. Mas tivemos que somar pela energia mínima na banda de condução porque a energia do gás que calculamos na Equação 55 tem como referencia um ponto de energia 0, enquanto no nosso caso agora o ponto de referencia é a energia mínima da banda de condução. Isso quer dizer que o nível mais baixo de energia na banda de condução do semiconductor é o nível mais baixo de energia da Equação 55 mais a energia ε_C .

Podemos facilitar ainda mais o cálculo com uma observação de ordem prática: geralmente, a banda de condução, mesmo quando subtraída do nível de Fermi, possui 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 e podemos aproximar o denominador como:

$$1 + \exp\left[\frac{(\varepsilon_{c_x, c_y, c_z} + \varepsilon_C) - \varepsilon_F}{k_B T}\right] \approx \exp\left[\frac{(\varepsilon_{c_x, c_y, c_z} + \varepsilon_C) - \varepsilon_F}{k_B T}\right]$$

Equação 108

Semicondutores que satisfazem essa condição são chamados de não-degenerados.

Utilizando essa aproximação, temos:

$$N_e = 2 \sum_{c_x} \sum_{c_y} \sum_{c_z} \frac{1}{\exp\left[\frac{(\varepsilon_{c_x, c_y, c_z} + \varepsilon_C) - \varepsilon_F}{k_B T}\right]} = 2 \sum_{c_x} \sum_{c_y} \sum_{c_z} \exp\left[-\frac{(\varepsilon_{c_x, c_y, c_z} + \varepsilon_C) - \varepsilon_F}{k_B T}\right]$$

Equação 109

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

$$N_e = 2 \exp\left[-\left(\frac{\varepsilon_C - \varepsilon_F}{k_B T}\right)\right] \sum_{c_x} \sum_{c_y} \sum_{c_z} \exp\left(-\frac{\varepsilon_{c_x, c_y, c_z}}{k_B T}\right)$$

Equação 110

Agora só precisamos resolver a soma. Na verdade, já lidamos com esse problema antes: a soma na Equação 110 é idêntica à soma da Equação 56. Como foi afirmado naquela ocasião, a soma é resolvida aproximando-a de uma integral. Isso é válido quando a variação de $\varepsilon_{c_x, c_y, c_z}$ é pequena em relação à variação nos coeficientes c_x , c_y , c_z . Neste caso, podemos aproximar a soma pela seguinte integral:

$$\sum_{c_x} \sum_{c_y} \sum_{c_z} \exp\left(-\frac{\varepsilon_{c_x, c_y, c_z}}{k_B T}\right) \approx \int_0^\infty \int_0^\infty \int_0^\infty \exp\left(-\frac{\varepsilon_{c_x, c_y, c_z}}{k_B T}\right) dc_x dc_y dc_z$$

Equação 111

Substituindo a expressão para energia (Equação 55), integral é expressa como:

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

Equação 112

E o resultado, é claro, é o mesmo da Equação 57:

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

Equação 113

Note que a massa do elétron foi substituída pela massa efetiva, simbolizada por m_e^* . Assim temos que o número total de elétrons livres será:

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

Equação 114

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

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

Equação 115

Onde V é o volume e

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

Equação 116

A Equação 115 nos conta qual é a concentração de portadores 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 , esse é um trabalho de mecânica quântica completamente fora do escopo do nosso estudo. Mas precisamos encontrar ε_F para completar a história. Isso será feito ainda nesta seção, mas antes vamos voltar na Equação 115.

Para encontrar a Equação 115 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 115 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 115 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. Como o nome sugere, por definição, a quantidade de estados N_ε entre os níveis de energia ε e $\varepsilon + d\varepsilon$ é dado por:

$$N_\varepsilon = D(\varepsilon)d\varepsilon$$

Equação 117

Onde $D(\varepsilon)$ é a densidade de estados. Cuidado para não confundir N_ε com N_e . 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, se soubermos a densidade de estados na banda de condução, podemos substituir a soma da Equação 105 por uma integral envolvendo a densidade de estados:

$$N_e = \sum_{bc} f(\varepsilon_{bc}) = \int_{\varepsilon_c}^{\infty} D(\varepsilon)f(\varepsilon)d\varepsilon$$

Equação 118

A interpretação da Equação 118 é direta: invés de somar a distribuição de Fermi-Dirac estado por estado, você pega a distribuição de Fermi-Dirac para uma determinada energia, e multiplica pelo tanto de estados que tem aquela energia e depois soma (na verdade integra) sobre todas as energias possíveis.

Vamos então fazer um parêntesis para calcular a densidade de estados para um gás de elétrons livres. Para isso, imagine um sistema de coordenadas euclidiano com eixos x , y e z . Agora imagine que marcamos um ponto para todas as coordenadas formadas por números inteiros. Por exemplo, marcamos as coordenadas $(0, 0, 1)$, $(1, 0, 0)$, $(2,0,1)$ e assim por diante. Marcamos com um ponto todas as coordenadas formadas por números inteiros. Note que cada uma dessas coordenadas corresponde a um conjunto c_x , c_y e c_z . Por enquanto, vamos nos permitir envolver os números inteiros negativos também, depois consertamos o erro causado por essa suposição. Agora imagine uma esfera, de raio R , centrada na origem dos três eixos. Pergunta: quantos

pontos existem dentro da esfera? Essa pergunta não é difícil de responder, mas antes de respondê-la, note que essa pergunta equivale à pergunta “quantas combinações possíveis de c_x , c_y e c_z estão dentro da esfera?”. Para responder essa pergunta, note que podemos associar cada ponto (cada coordenada formada por números inteiros) com um volume igual à 1. Isto é, se colocarmos um cubinho com lado igual à 1 centrado em cada um dos pontos, iremos preencher todo o espaço, e o número de pontos dentro da esfera será igual ao número de cubinhos. Mas o número de cubinhos é praticamente igual ao volume ocupado pela esfera dividido pelo volume dos cubinhos (praticamente igual pois só seria matematicamente idêntico se dividíssemos cubos por cubos, mas isso acarretaria uma complicação desnecessária). Chamando o número de pontos de P , e lembrando que cada cubinho possui volume igual à 1, temos:

$$P = \frac{\frac{4}{3}\pi R^3}{1} = \frac{4}{3}\pi R^3$$

Equação 119

Vamos supor agora que o raio R possua coordenadas (c_{x0} , c_{y0} e c_{z0}). Por trigonometria básica temos que:

$$R = \sqrt{c_{x0}^2 + c_{y0}^2 + c_{z0}^2}$$

Equação 120

Mas, de acordo com a Equação 55, podemos associar cada coordenada marcada por um ponto com um determinado estado (cuja energia é dada pela Equação 55). No caso particular do estado associado à coordenada (c_{x0} , c_{y0} e c_{z0}), o nível de energia, que chamaremos de ε_0 , é:

$$\varepsilon_0 = \frac{h^2}{8L^2m} (c_{x0}^2 + c_{y0}^2 + c_{z0}^2)$$

Equação 121

O que isso tudo quer dizer? Como dada combinação c_x , c_y e c_z virou um ponto no sistema de coordenadas, e como cada ponto dentro da esfera corresponde a um estado com nível de energia menor que ε_0 (Equação 121), então o número de pontos P dentro da esfera de raio R efetivamente corresponde ao número de estados com energia menor que ε_0 .

Na verdade, temos que dividir P por 8, porque só podemos incluir números positivos para c_x , c_y e c_z . Como existem $2^3 = 8$ combinações de sinais para as três coordenadas, podemos dividir a esfera em 8 pedaços iguais, sendo que apenas um desses pedaços terá c_x , c_y e c_z positivos. Todos os outros pedaços terão pelo menos um desses números negativos. Além disso, temos que multiplicar o resultado final por 2 pois cada estado corresponde na verdade à dois estados, um com spin *up* e outro com spin *down*.

Para generalizar a ideia, vamos chamar então a quantidade **total de estados com energia menor que ε de $\eta(\varepsilon)$** . Temos então que:

$$\eta(\varepsilon) = \frac{2P}{8} = \frac{1}{3} \pi R^3 = \frac{1}{3} \pi (c_x^2 + c_y^2 + c_z^2)^{3/2} = \frac{1}{3} \pi \left(\frac{8L^2 m}{h^2} \varepsilon \right)^{3/2}$$

Equação 122

Reorganizado:

$$\eta(\varepsilon) = \frac{8\pi L^3}{3} \left(\frac{2m}{h^2} \right)^{3/2} \varepsilon^{3/2}$$

Equação 123

Agora podemos encontrar a densidade de estados com certa facilidade. Primeiro, identificamos que, como $\eta(\varepsilon)$ é o número de estados com energia menor que ε , então o número de estado com energia entre ε e $\varepsilon + d\varepsilon$ é:

$$N_\varepsilon = \eta(\varepsilon + d\varepsilon) - \eta(\varepsilon)$$

Equação 124

Mas da Equação 117 temos:

$$D(\varepsilon)d\varepsilon = \eta(\varepsilon + d\varepsilon) - \eta(\varepsilon)$$

Equação 125

Portanto:

$$D(\varepsilon) = \frac{\eta(\varepsilon + d\varepsilon) - \eta(\varepsilon)}{d\varepsilon} = \frac{\partial \eta}{\partial \varepsilon}$$

Equação 126

A derivada da Equação 123 não oferece nenhuma dificuldade e o resultado é:

$$D(\varepsilon) = 4\pi L^3 \left(\frac{2m}{h^2} \right)^{3/2} \varepsilon^{1/2}$$

Equação 127

Note que a densidade de energia aumenta com a energia. Isso é esperado uma vez que, quanto maior for a energia, mais combinações possíveis de c_x , c_y e c_z resultam no mesmo valor de energia.

Como queremos utilizar esse resultado para o gás de elétrons na banda de condução, temos que substituir a massa pela massa efetiva, como discutido anteriormente. Além disso, o nosso novo zero de referencia é a energia da banda de condução, mas a Equação 127 foi deduzida para uma referencia de energia igual à 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 real e a energia na banda de condução (assim, a energia que entra na Equação 127 continua tendo o zero como referencia). Portanto, para o gás na banda de condução temos:

$$D(\varepsilon) = 4\pi L^3 \left(\frac{2m_e^*}{h^2} \right)^{3/2} (\varepsilon - \varepsilon_C)^{1/2}$$

Equação 128

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, portanto, densidade de estados zero).

Agora já temos a faca e o queijo na mão: basta substituírmos $D(\varepsilon)$ na integral da Equação 118. Utilizando novamente a aproximação de semicondutores não-degenerados (Equação 108), temos que:

$$N_e = \int_{\varepsilon_C}^{\infty} D(\varepsilon) \exp\left[-\frac{(\varepsilon - \varepsilon_F)}{k_B T}\right] d\varepsilon = \int_{\varepsilon_C}^{\infty} 4\pi L^3 \left(\frac{2m_e^*}{h^2} \right)^{3/2} (\varepsilon - \varepsilon_C)^{1/2} \exp\left[-\frac{(\varepsilon - \varepsilon_F)}{k_B T}\right] d\varepsilon$$

Equação 129

Isolando os termos que não dependem da energia:

$$N_e = 4\pi L^3 \left(\frac{2m_e^*}{h^2} \right)^{3/2} \exp \left[\frac{(\mathcal{E}_F)}{k_B T} \right] \int_{\mathcal{E}_C}^{\infty} (\mathcal{E} - \mathcal{E}_C)^{1/2} \exp \left[-\frac{\mathcal{E}}{k_B T} \right] d\mathcal{E}$$

Equação 130

E o resultado é:

$$N_e = L^3 \frac{2}{\left(\frac{h^2}{2\pi m_e^* k_B T} \right)^{3/2}} \exp \left[-\left(\frac{\mathcal{E}_C - \mathcal{E}_F}{k_B T} \right) \right]$$

Equação 131

Como L^3 é o volume, encontramos novamente que a densidade de cargas é dada por:

$$n = \frac{N_e}{V} = n_c \exp \left[-\left(\frac{\mathcal{E}_C - \mathcal{E}_F}{k_B T} \right) \right]$$

Equação 132

Onde n_c é dado pela Equação 116. A Equação 132 é idêntica à Equação 115, como esperado.

Agora vamos voltar nossa atenção ao problema do cálculo do nível de Fermi. Para isso, vamos utilizar o seguinte fato: nós já sabemos que a densidade de elétrons livres na banda de condução, n_e , tem que ser igual à densidade de buracos livres na banda de valência (que chamaremos de n_p) já que cada elétron excitado para a banda de condução deixa um buraco na banda de valência. Mas poderíamos calcular n_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 128, apenas trocando a massa efetiva do elétron pela massa efetiva do buraco e os níveis de referência. Assim, chamando a densidade de estados de buracos de $D_b(\mathcal{E})$, temos:

$$D_b(\mathcal{E}) = 4\pi L^3 \left(\frac{2m_b^*}{h^2} \right)^{3/2} (\mathcal{E}_V - \mathcal{E})^{1/2}$$

Equação 133

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 do estado estar desocupado pelo elétron, ou seja:

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

Equação 134

Onde f_b é a probabilidade do 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 135

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 136

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

$$N_p = \int_{-\infty}^{\varepsilon_v} D_p(\varepsilon) f_p(\varepsilon) d\varepsilon = \int_{-\infty}^{\varepsilon_v} 4\pi L^3 \left(\frac{2m_b^*}{h^2}\right)^{3/2} (\varepsilon_v - \varepsilon)^{1/2} \exp\left(\frac{-(\varepsilon_F - \varepsilon)}{k_B T}\right) d\varepsilon$$

Equação 137

Essa integral é bem parecida com a anterior; já vamos colocar então resultado para a densidade de buracos:

$$p = \frac{N_p}{V} = n_d \exp\left[-\left(\frac{\varepsilon_F - \varepsilon_v}{k_B T}\right)\right]$$

Equação 138

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 139

Note que a única diferença entre n_c e n_d é que n_c envolve a massa efetiva do elétron e n_d envolve a massa efetiva do buraco. Antes de calcularmos o nível de Fermi, note também o seguinte fato: se multiplicarmos a Equação 138 pela Equação 132, teremos:

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

Equação 140

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

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

Equação 141

Esse é um resultado notável: de acordo com a Equação 141, o produto entre a concentração de buracos p e a concentração de elétrons livres n não depende do nível de Fermi. Por enquanto, estamos estudando o caso do semicondutor intrínseco (não dopado), onde sabemos que p e n são iguais. Mas, como veremos mais adiante, se doparmos o semicondutor isso não será mais verdade. Mas o efeito da dopagem é uma mudança no nível de Fermi e como a produto entre p e n não depende do nível de Fermi, o produto será constante independentemente do semicondutor estar dopado ou não. Isso significa que, se a dopagem aumenta o número de elétrons, necessariamente ela diminui o número e buracos, 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 semicondutor intrínseco com facilidade, basta utilizarmos o fato de que em um semicondutor intrínseco $n = p$. Portanto:

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

Equação 142

Daí temos que:

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

Equação 143

Portanto:

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

Equação 144

Rearranjando:

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

Equação 145

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 146

A Equação 146 é um importante resultado. Essa relação nos diz que em um semiconductor não dopado, o nível de Fermi está praticamente no meio do *band-gap*.

3.4 Semicondutores extrínsecos (dopagem)

Um semiconductor intrínseco *per se* é um material virtualmente inútil. Em praticamente todas as tecnologias relacionadas à eletrônica, os semicondutores utilizados são dopados. Dopagem se refere à inserção de átomos de impureza (isto é,

átomos que não fazem parte do cristal) na estrutura cristalina do semiconductor. Tipicamente, dependendo do tipo de átomo utilizado, as ligações químicas entre os átomos de impureza e a estrutura cristalina resultam, ou 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 semiconductor dopado é chamado de extrínseco. Utilizando uma nomenclatura mais direta, chamamos 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 mais interessados no que ocorre com os níveis de energia do semiconductor quando ele é dopado.

A Figura 14 ilustra as bandas de energia de um semiconductor intrínseco, um semiconductor tipo n e um semiconductor tipo p . Na dopagem do tipo n , os átomos de impureza criam estados de energia dentro do *band-gap* e próximo da banda de condução. Os elétrons que ocupam estes estados podem ser facilmente excitados (termicamente) para a banda de condução, devido à proximidade entre os níveis de energia. Portanto, o processo de dopagem do 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, seja ele dopado ou não), o aumento da ocupação na banda de condução corresponde ao aumento na energia de Fermi (ou seja, o nível de Fermi se aproxima da banda de condução). O aumento da energia de Fermi expressa a maior probabilidade de ocupação na banda de condução, já que, de acordo com a distribuição de Fermi-Dirac, quando a energia 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 população na banda de condução e o aumento na energia de Fermi não devem ser entendidos como causa ou consequência. O entendimento mais apropriado é que estas são duas manifestações do mesmo efeito: um aumento no nível de Fermi corresponde ao aumento na concentração de elétrons na banda de condução e vice-versa. A causa tanto do aumento da população quanto do aumento do nível de Fermi aqui é a dopagem tipo n . De maneira análoga, a dopagem tipo p cria um estado de energia próximo da banda de valência, o que permite que

elétrons da banda de valência sejam termicamente excitados para o estado criado pelos átomos 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). É muito importante ressaltar que o processo de dopagem não altera a carga total do semiconductor. Tanto o semiconductor tipo n quanto o semiconductor tipo p continuam sendo materiais neutros, ou seja, com mesmo número de cargas negativas e positivas. O que ocorre é simplesmente um aumento de cargas livres: elétrons para dopagem tipo n e buracos para dopagem tipo p .

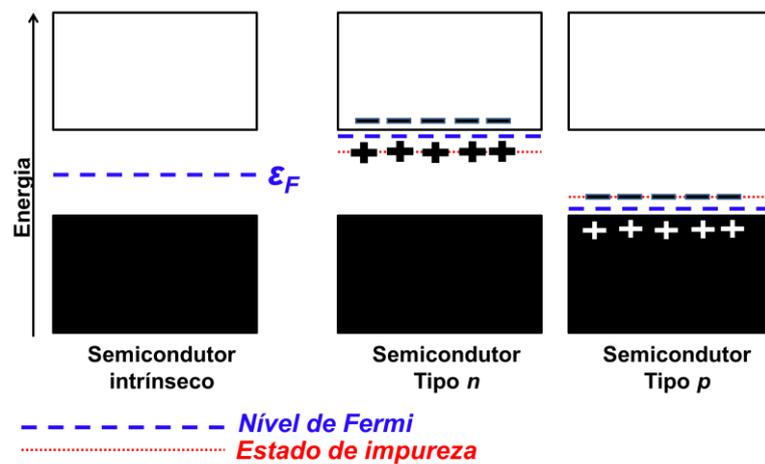


Figura 14 Diagrama de energia do semiconductor intrínseco, semiconductor tipo n e semiconductor tipo p .

Note que, em um semiconductor dopado, não é mais verdade que o número de elétrons livres é igual ao número de buracos. De fato, um semiconductor tipo n possui muito mais elétrons livres do que buracos, ou seja, $n \gg p$. Por outro lado, em um semiconductor tipo p temos $p \gg n$. Mas, como afirmado na seção anterior, o produto entre p e n (Equação 141) depende apenas do *band-gap* e, portanto, é o mesmo em semicondutores intrínsecos e extrínsecos.

Precisamos agora calcular as concentrações de portadores livres em semicondutores extrínsecos. Mas essa agora é uma tarefa bem simples. Primeiramente, note que a Equação 132 e a Equação 138 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 muda nessas expressões são os níveis de Fermi,

que agora **não são mais dados pela** Equação 145 (por que? O que foi utilizado na dedução da Equação 145 que não é mais verdade agora?).

Então o 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. Vamos chamar a concentração de átomos de impureza em um semiconductor tipo n de N_d (o subscrito d se refere ao fato do átomo de impureza ser um doador de elétrons). Assim, temos que, para um semiconductor tipo n :

$$n \approx N_d$$

Equação 147

Mas, da Equação 132, temos também que:

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

Equação 148

Portanto:

$$N_d \approx n_c \exp\left[-\left(\frac{\varepsilon_C - \varepsilon_F}{k_B T}\right)\right]$$

Equação 149

Podemos agora facilmente encontrar o nível de Fermi para um semiconductor tipo n isolando ε_F na Equação 149, o que resulta em:

$$\varepsilon_F = \varepsilon_C + k_B T \ln\left(\frac{N_d}{n_c}\right)$$

Equação 150

Podemos fazer ainda melhor: podemos re-arranjar a Equação 148 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

semicondutor intrínseco (dado pela Equação 145 ou Equação 146) de ε_{FI} , onde o subscrito I é para lembrarmos que o termo designa somente ao semicondutor intrínseco. Vamos chamar também os portadores livres do semicondutor intrínseco de n_i e p_i (respectivamente: Equação 132 e Equação 138, ambas com $\varepsilon_F = \varepsilon_{FI}$). Agora, vamos re-arranjar a Equação 148 para expressá-la em termos de n_i e ε_{FI} :

$$n = n_c \exp\left[-\left(\frac{\varepsilon_C - \varepsilon_F}{k_B T}\right)\right] = n_c \exp\left[-\left(\frac{\varepsilon_C - \varepsilon_{FI} + \varepsilon_{FI} - \varepsilon_F}{k_B T}\right)\right] = n_c \exp\left[-\left(\frac{\varepsilon_C - \varepsilon_{FI}}{k_B T}\right)\right] \exp\left[-\left(\frac{\varepsilon_{FI} - \varepsilon_F}{k_B T}\right)\right]$$

Equação 151

Comparando o lado direito da Equação 151 com a expressão para n_i , concluímos que:

$$n = n_i \exp\left[-\left(\frac{\varepsilon_{FI} - \varepsilon_F}{k_B T}\right)\right]$$

Equação 152

A Equação 152 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 semicondutor dopado. Claro que, se $\varepsilon_F = \varepsilon_{FI}$, então $n = n_i$. Lembre-se que tanto a Equação 148 como a Equação 152 são gerais: valem para qualquer semicondutor 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 semicondutor com dopagem tipo n , onde a concentração de dopantes é muito maior que a concentração de portadores intrínsecos, então a Equação 147 também é verdade. Neste caso:

$$N_D = n_i \exp\left[-\left(\frac{\varepsilon_{FI} - \varepsilon_F}{k_B T}\right)\right]$$

Equação 153

E portanto:

$$\varepsilon_F = \varepsilon_{FI} + k_B T \ln\left(\frac{N_D}{n_i}\right)$$

Equação 154

A Equação 153 e a Equação 149 são equivalentes: ambas expressam a mesma coisa, mas utilizando parâmetros diferentes. Tanto uma quanto a outra pode ser utilizada. De maneira análoga, a Equação 154 e a Equação 150 são equivalentes. Note

que, quanto maior for N_D em relação à n_i , maior será o deslocamento do nível de Fermi em relação ao centro do *band-gap*. Não se esqueça que a Equação 149, a Equação 150, a Equação 153 e a Equação 154 só valem para semicondutores do tipo n .

Podemos repetir o mesmo raciocínio para encontrarmos uma expressão para a densidade de buracos. Começamos re-escrevendo a Equação 138 em termos de p_i e ε_{FI} :

$$p = n_d \exp\left[-\left(\frac{\varepsilon_F - \varepsilon_V}{k_B T}\right)\right] = n_d \exp\left[-\left(\frac{\varepsilon_F - \varepsilon_{FI} + \varepsilon_{FI} - \varepsilon_V}{k_B T}\right)\right] = n_d \exp\left[-\left(\frac{\varepsilon_{FI} - \varepsilon_V}{k_B T}\right)\right] \exp\left[-\left(\frac{\varepsilon_F - \varepsilon_{FI}}{k_B T}\right)\right]$$

Equação 155

De onde concluímos que:

$$p = p_i \exp\left[-\left(\frac{\varepsilon_F - \varepsilon_{FI}}{k_B T}\right)\right]$$

Equação 156

Este 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 N_A (o subscripto A vem do inglês *acceptor*, pois o átomo de impureza recebe um elétron) for bem maior que a concentração de portadores intrínsecos, então também é verdade que:

$$p \approx N_A$$

Equação 157

De onde concluímos que, para um semicondutor tipo p:

$$N_A = p_i \exp\left[-\left(\frac{\varepsilon_F - \varepsilon_{FI}}{k_B T}\right)\right]$$

Equação 158

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 159

Note que, quanto maior for N_A em relação à p_i , mais deslocado o nível de Fermi estará em relação ao centro do *band-gap*, mas agora o deslocamento é para baixo.

4 – Introdução aos dispositivos semicondutores: a junção p-n

No último capítulo do curso, vamos utilizar todos os conceitos dos 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 à uma situação de desequilíbrio.

4.1 Junção p-n em equilíbrio

Como sugerido pelo nome, a junção p-n é formada pela união de um semicondutor p com um semicondutor 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 15. O que vai ocorrer quando um semicondutor tipo n for colocado em contato com um semicondutor tipo p? Existe uma concentração maior de elétrons livres no semicondutor tipo n do que no semicondutor tipo p. Essa concentração maior de elétrons livres corresponde ao nível de Fermi mais alto no semicondutor tipo n (Figura 15a). 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. Estudamos um caso quase idêntico no Capítulo 2 (Figura 5): os elétrons irão difundir da região de maior potencial (maior nível de Fermi) para a região de menor potencial (menor nível de Fermi), ou seja, elétrons irão difundir do semicondutor tipo n para o semicondutor tipo p. 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. No equilíbrio, 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 15b). Portanto, o que aprendemos no capítulo 2 é

suficiente para explicar 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 da Figura 5: na junção p-n, além de elétrons livres, temos também buracos. Intuitivamente, sabemos que os buracos irão difundir da região p para a região n, já que região p tem uma concentração maior de buracos. Mas note que a junção 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 contrário do que ocorre para elétrons livres. Essa aparente esquizofrenia dos buracos é facilmente compreendida se lembrarmos 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) corresponde à 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, tanto a difusão de elétrons quanto de buracos irá cessar quando o nível de Fermi equalizar, como representado na Figura 15b.

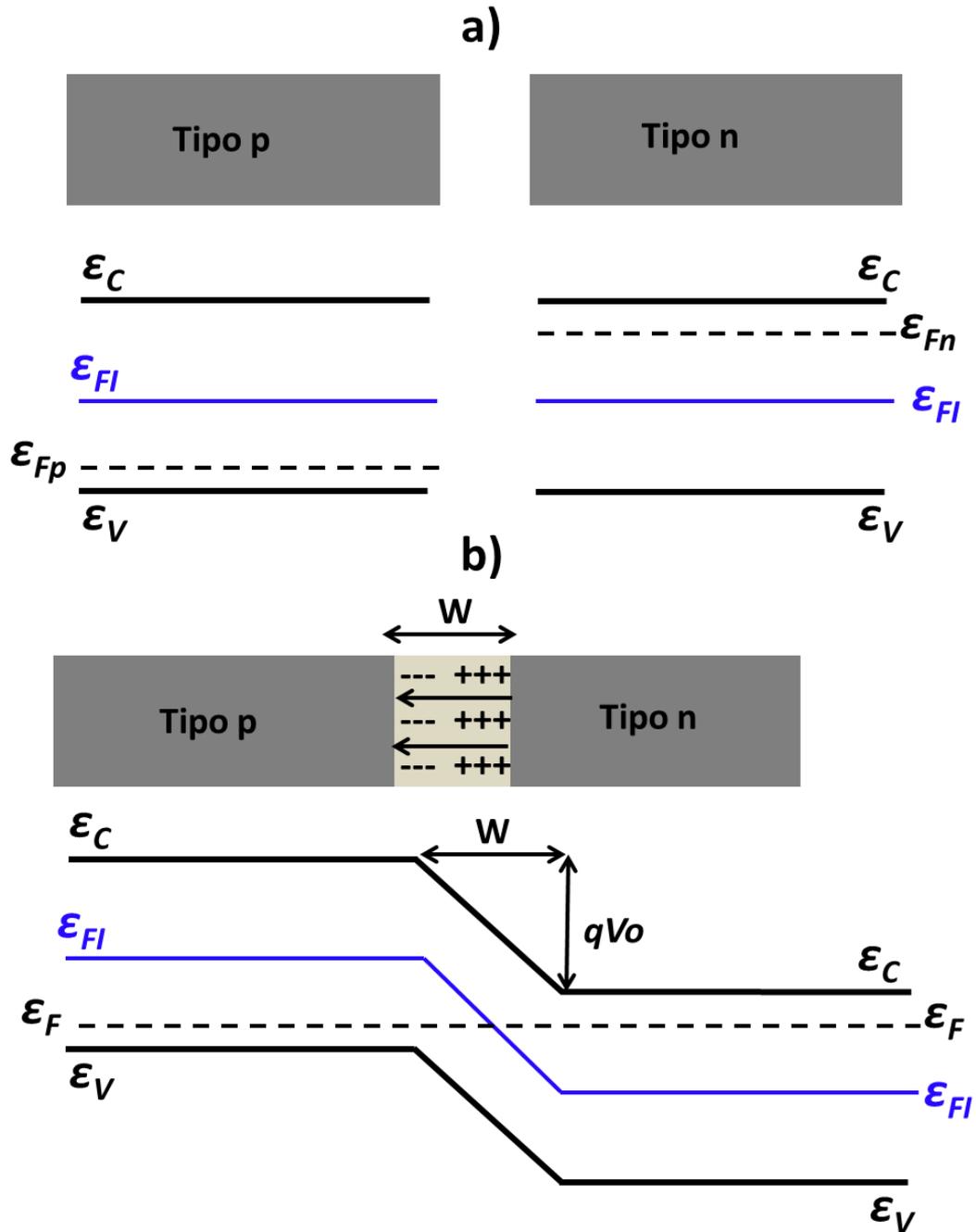


Figura 15 Experimento “mental” mostrando a formação de uma junção p-n através da união de dois semicondutores dopados. a) Semicondutores e níveis de energia antes da formação da junção: ϵ_{FI} é o nível de Fermi intrínseco e ϵ_{Fp} e ϵ_{Fn} são os níveis de Fermi do semiconductor p e do semiconductor n, respectivamente. b) Semiconductor e níveis de energia após a formação da junção. A região de comprimento W é chamada de região de depleção e 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.

Analisando a Figura 15b com atenção, podemos notar que o potencial eletrostático resultante do campo elétrico equaliza os níveis de Fermi, entortando assim as bandas de energia. Note também que o nível de Fermi intrínseco, por estar sempre no centro

do *band-gap*, também é entortado. Esse fenômeno não tem nenhum segredo: o potencial eletrostático está desempenhando aí um papel análogo ao potencial gravitacional. A Figura 16 mostra um sistema mecânico análogo: para que ocorra equilíbrio, as duas prateleiras têm que estar na mesma altura, o que faz com que as caixas estejam em alturas diferentes. No exemplo, a prateleira é o nível de Fermi e o topo da caixa é a banda de condução (claro que a base da caixa pode ser a banda de valência). A diferença é que no exemplo mecânico a distorção é descontínua enquanto no caso da junção p-n a distorção é contínua, já que o potencial eletrostático cresce gradativamente ao longo do centro junção.

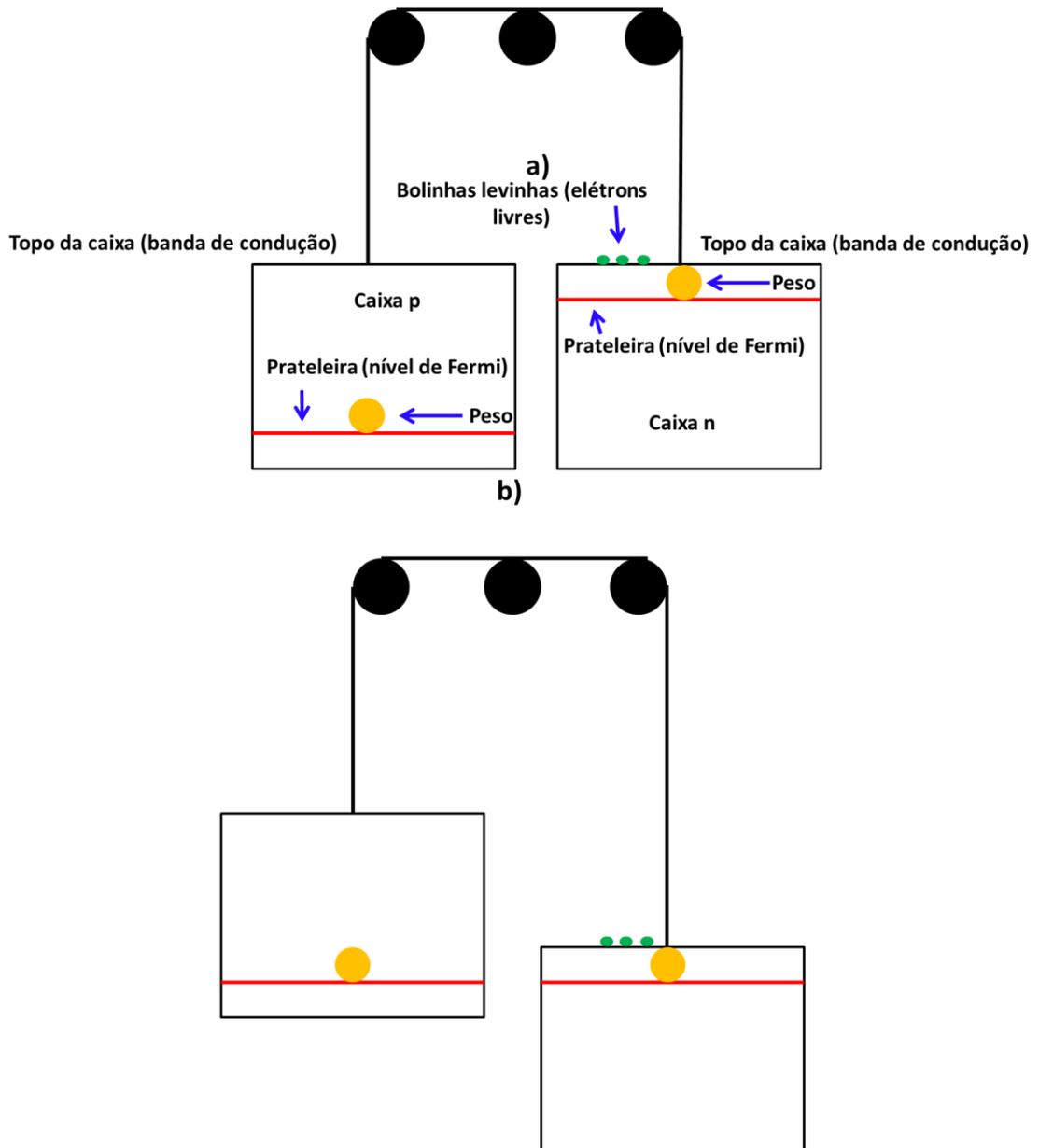


Figura 16 Analogia mecânica da junção p-n. O exemplo supõe que apenas os pesos das bolinhas amarelas sejam significativos a) Caixas antes da interação. A caixa n tem um potencial maior pelo fato do peso estar em uma prateleira mais alta. Após a interação b), as prateleiras são igualadas, o que significa que as caixas estão em alturas desiguais, de maneira que o potencial total seja constante.

Voltando à Figura 15, note a região onde há acúmulo de cargas. Atribuímos o comprimento W à essa região. Essa região é chamada de região de depleção, termos este que denota o fato dessa ser uma região sem cargas livres, uma vez que as cargas que estavam ali presentes difundiram para o outro lado da junção. É importante lembrar que as cargas na região de depleção são formadas primordialmente pelos íons dos átomos dopantes. Voltaremos a comentar sobre a distribuição de cargas na região de depleção mais adiante, mas antes vamos revisar o problema do cálculo do

potencial eletrostático resultante do campo elétrico da região de depleção, denotado na Figura 15 por V_0 .

No capítulo 2 deduzimos que $|q|V_0$ deve ser igual à diferença de potencial químico. Sabemos que a diferença de potencial químico corresponde à diferença entre os níveis de Fermi dos semicondutores n e p antes da formação da junção, portanto esperamos que $|q|V_0$ seja igual à essa diferença, fato este que já está representado na Figura 15b. Entretanto, a dedução no capítulo 2 foi feita de maneira fenomenológica. Este é um termo sofisticado para descrever uma dedução informal, que parte do bom senso e de intuições, mas que não segue uma linha de dedução formal. É necessário e instrutivo, portanto, revisitar esse problema de uma maneira um pouco 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 cargas. Rigorosamente falando, o campo elétrico anula a difusão porque o campo elétrico induz uma corrente de arrasto contrária à corrente de difusão. Esse equilíbrio é atingido tanto para elétrons 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 das correntes para buracos, temos que primeiro lembrar quais são as expressões para as correntes. Vimos no capítulo de transporte que a densidade de corrente de arrasto (que chamaremos de J_{arr}) e a densidade de corrente de difusão (que chamaremos de J_D) são dadas, respectivamente, por:

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

Equação 160

e

$$J_D = -|q|D \frac{\partial p}{\partial x}$$

Equação 161

Onde ρ é a densidade de cargas, v é a velocidade das cargas, D é o coeficiente de difusão e p é a densidade de portadores (no caso aqui, buracos). Vamos re-escrever essas duas relações de maneira um pouco diferente. Note primeiro que, obviamente, a

densidade de portadores está relacionada à densidade de cargas ($\rho = |q|p$). 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 2, então cuidado para não confundir). Assim, a corrente de arrasto pode ser re-expressa como:

$$J_{arr} = |q| \cdot p(x) \cdot \mu \cdot E(x)$$

Equação 162

Onde E 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 163

Utilizando essa relação, podemos re-rexpressar a corrente de difusão como:

$$J_D = -\mu \cdot k_B \cdot T \frac{\partial p}{\partial x}$$

Equação 164

Pronto, já temos as expressões para as correntes de arrasto e difusão. A condição de equilíbrio, obviamente, é expressa matematicamente por:

$$J_{arr} + J_D = 0$$

Equação 165

Portanto temos que:

$$|q| \cdot p(x) \cdot \mu \cdot E(x) - \mu \cdot k_B \cdot T \frac{\partial p}{\partial x} = 0$$

Equação 166

Ou seja:

$$E(x) = \frac{k_B \cdot T}{|q|} \frac{1}{p(x)} \frac{\partial p}{\partial x}$$

Equação 167

Vamos parar um pouco para lembrar o que queremos. Queremos uma relação entre o potencial eletrostático e os níveis de Fermi do semiconductor antes da formação da junção, denotados por ε_{Fp} e ε_{Fn} na Figura 15. Vimos no capítulo anterior que a concentração de cargas depende desses valores, portanto podemos expressar p em termos dos níveis de Fermi. Por outro lado, o potencial eletrostático pode ser encontrado através do campo elétrico. Então estamos no caminho certo, pois se integrarmos os dois lados da Equação 167 teremos o potencial eletrostático de um lado e uma expressão para as cargas (que dependem dos níveis de Fermi) do outro. Para fazer a integral, é conveniente colocar um eixo e uma origem explícitos. O eixo escolhido é representado na Figura 17.

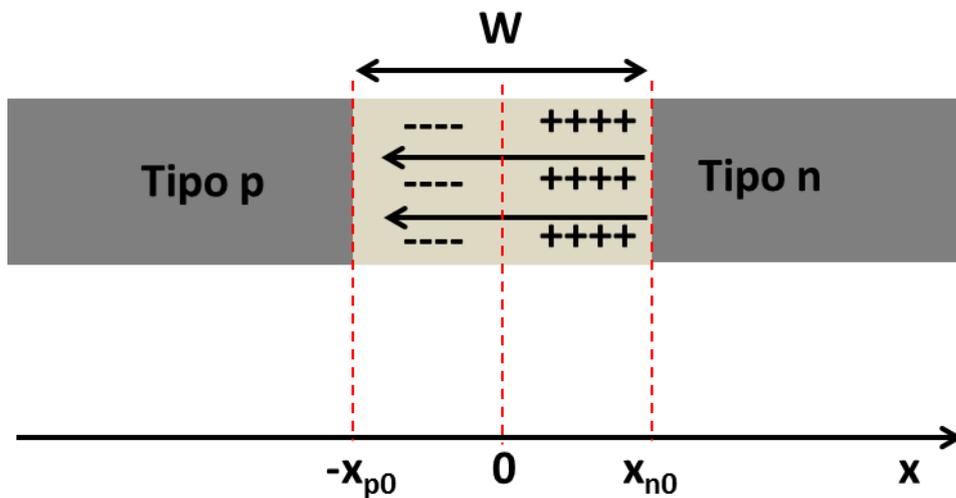


Figura 17 Representação do eixo de coordenadas. A origem é colocada no centro da região de depleção. O ponto x_{n0} denota o início da região de depleção na parte n enquanto o ponto $-x_{p0}$ denota o início da região da depleção na parte p.

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:

$$V(x_{n0}) - V(-x_{p0}) = - \int_{-x_{p0}}^{x_{n0}} E(x) dx$$

Equação 168

Como V_o é a diferença entre o potencial na parte n e o potencial na parte p, temos que:

$$V_o = - \int_{-x_{p0}}^{x_{n0}} E(x) dx$$

Equação 169

Substituindo $E(x)$ da Equação 167, temos:

$$V_o = - \int_{-x_{p0}}^{x_{n0}} \frac{k_B \cdot T}{|q|} \frac{1}{p(x)} \frac{\partial p}{\partial x} dx = \frac{k_B \cdot T}{|q|} [\ln p(-x_{p0}) - \ln p(x_{n0})]$$

Equação 170

Assim chegamos na importante relação:

$$V_o = \frac{k_B \cdot T}{|q|} \left[\ln \frac{p(-x_{p0})}{p(x_{n0})} \right]$$

Equação 171

Lembre-se que poderíamos ter utilizado a condição de equilíbrio para elétrons. Se tivéssemos feito isso, chegaríamos na seguinte relação:

$$V_o = \frac{k_B \cdot T}{|q|} \left[\ln \frac{n(x_{n0})}{n(-x_{p0})} \right]$$

Equação 172

O que implica que:

$$p(-x_{p0})n(-x_{p0}) = p(x_{n0})n(x_{n0})$$

Equação 173

Esse resultado, na verdade, poderia ser deduzido utilizando apenas as expressões do capítulo 3. Na lista de exercícios você terá que provar essa relação utilizando as expressões do capítulo 3. Note que ela diz que o produto entre as concentrações dos portadores no semicondutor p é igual ao produto para o semicondutor n. Esse produto,

na verdade, também é igual para o semiconductor intrínseco (como você tem que provar na lista), ou seja, dopagem não muda altera o produto np .

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 171 como a Equação 172. É claro que temos que utilizar também as relações deduzidas no capítulo 3 para expressar as concentrações em termos dos níveis de Fermi. Escolhendo a concentração de buracos, podemos utilizar a Equação 156. Substituindo esta na Equação 171, temos:

$$V_0 = \frac{k_B \cdot T}{|q|} \left[\ln \frac{p(-x_{p0})}{p(x_{n0})} \right] = \frac{k_B \cdot T}{|q|} \left[\ln \frac{p_i \exp \left[- \left(\frac{\varepsilon_F - \varepsilon_{FI}(-x_{p0})}{k_B T} \right) \right]}{p_i \exp \left[- \left(\frac{\varepsilon_F - \varepsilon_{FI}(x_{n0})}{k_B T} \right) \right]} \right]$$

Equação 174

Lembre-se que o nível de Fermi é constante, mas o nível de Fermi intrínseco varia ao longo da junção, como mostrado na Figura 15b. Por isso ε_F não depende de x , mas ε_{FI} depende. Mas note também o seguinte: o que aparece na Equação 174 é a diferença entre o nível de Fermi e o nível de Fermi intrínseco. Fora da região de depleção, essa diferença é a mesma do semiconductor dopado antes da formação da junção, do mesmo jeito que na analogia mecânica a diferença entre a prateleira e o topo ou fundo da caixa não muda quando a caixa é levantada ou abaixada como um todo. Portanto, podemos re-expressar essa diferença em termos dos níveis originais, ou seja:

$$V_0 = \frac{k_B \cdot T}{|q|} \left[\ln \frac{p_i \exp \left[- \left(\frac{\varepsilon_F - \varepsilon_{FI}(-x_{p0})}{k_B T} \right) \right]}{p_i \exp \left[- \left(\frac{\varepsilon_F - \varepsilon_{FI}(x_{n0})}{k_B T} \right) \right]} \right] = \frac{k_B \cdot T}{|q|} \left[\ln \frac{p_i \exp \left[- \left(\frac{\varepsilon_{Fp} - \varepsilon_{FI0}}{k_B T} \right) \right]}{p_i \exp \left[- \left(\frac{\varepsilon_{Fn} - \varepsilon_{FI0}}{k_B T} \right) \right]} \right]$$

Equação 175

Onde o índice 0 foi adicionado à ε_{FI0} para lembrarmos que esse é o valor do nível intrínseco antes da formação da junção.

Note que, como o nível de Fermi intrínseco é o mesmo para o semiconductor tipo p ou tipo n antes da formação da junção, temos o mesmo ε_{F10} no numerador e no denominador da última igualdade da Equação 175. Portanto, temos que:

$$V_O = \frac{k_B \cdot T}{|q|} \left[\ln \frac{\exp\left[-\left(\frac{\varepsilon_{Fp} - \varepsilon_{F1}}{k_B T}\right)\right]}{\exp\left[-\left(\frac{\varepsilon_{Fn} - \varepsilon_{F1}}{k_B T}\right)\right]} \right] = \frac{k_B \cdot T}{|q|} \left[\ln \exp\left[\left(\frac{\varepsilon_{Fn} - \varepsilon_{Fp}}{k_B T}\right)\right] \right] = \frac{k_B \cdot T}{|q|} \left[\frac{\varepsilon_{Fn} - \varepsilon_{Fp}}{k_B \cdot T} \right]$$

Equação 176

Daí concluímos que, como esperado:

$$V_O = \frac{\varepsilon_{Fn} - \varepsilon_{Fp}}{|q|}$$

Equação 177

A Equação 177 é idêntica à Equação 38. A vida é mesmo uma eterna repetição.

Para finalizar essas elucubrações a respeito da junção p-n em equilíbrio, precisamos encontrar o campo elétrico na região de depleção e quantificar a largura da região de depleção. 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 íons das impurezas dopantes. 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 advém principalmente desses íons “descobertos”. De maneira análoga, as cargas negativas na região p advém principalmente dos íons descobertos na região p. Eu disse principalmente porque as cargas livres que difundiram também recombinaem e contribuem para a formação da região de depleção, mas essas cargas geralmente estão mais “diluídas” ao longo do semiconductor, 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, de fato, uma boa aproximação). Se a dopagem for uniforme, então os íons estarão uniformemente distribuídos ao longo da junção, o que significa que os íons descobertos também terão distribuição uniforme. Portanto, a densidade de cargas será uniforme ao longo da junção, apresentando transições abruptas nos limites da junção como mostrado na Figura 18, onde N_d é a concentração de íons doadores e N_a é a concentração de íons aceptadores. 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).

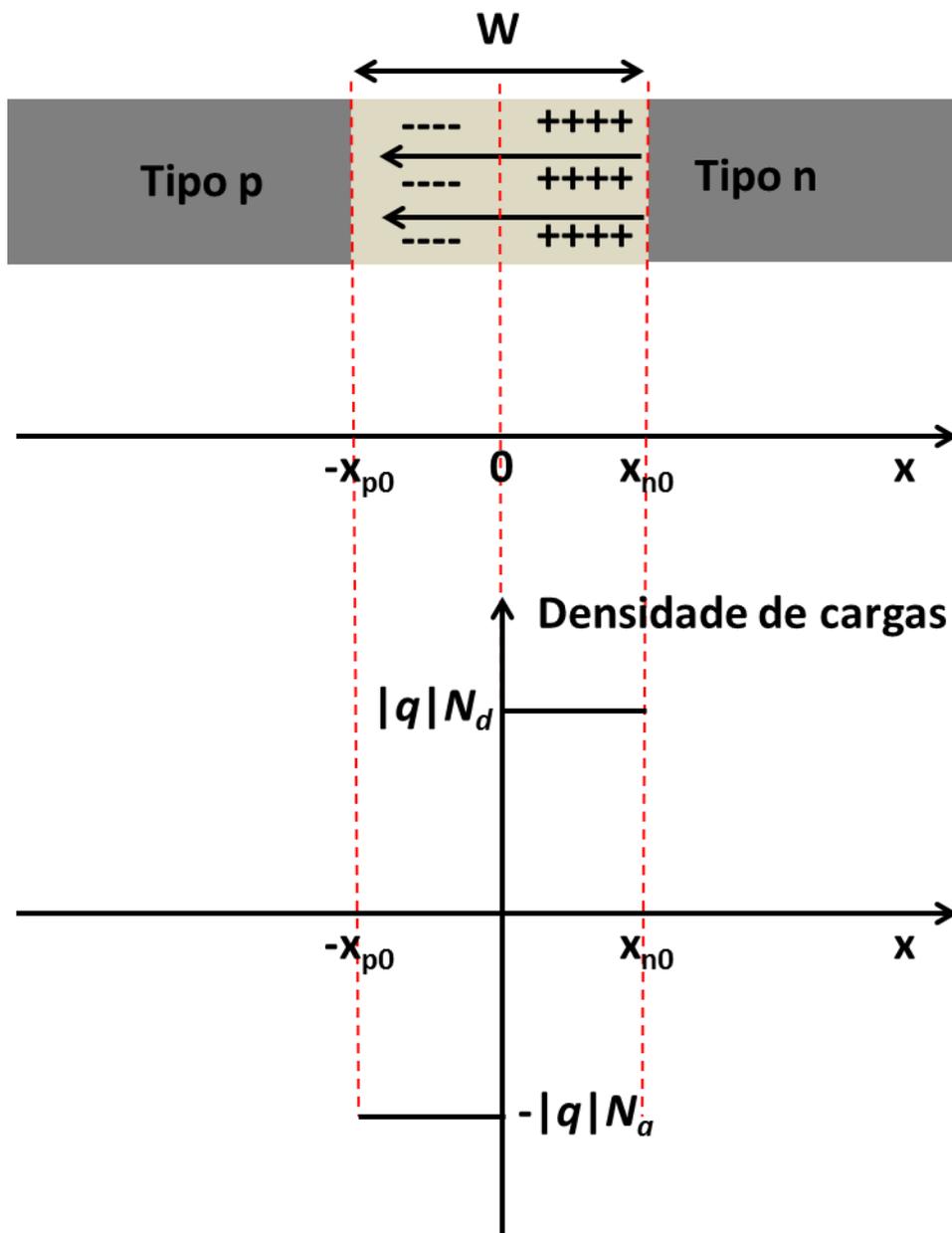


Figura 18 Distribuição especial da densidade de cargas.

Nós 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. Podemos utilizar essa informação para encontrarmos o campo elétrico na junção? Sabemos que o campo elétrico surge exatamente do acúmulo das cargas, mas qual é a relação matemática entre as cargas e o campo elétrico? Você deve lembrar-se (espero) do curso de eletromagnetismo que a relação entre campo elétrico e densidade de cargas é dada pela importante Lei de Gauss:

$$\nabla \cdot E = \frac{\rho}{\varepsilon}$$

Equação 178

Onde ε é a constante dielétrica (no nosso caso, é a constante dielétrica do semiconductor). Como o nosso problema é uni-dimensional, a Lei de Gauss é reduzida à:

$$\frac{\partial E}{\partial x} = \frac{\rho}{\varepsilon}$$

Equação 179

Mas já identificamos que:

$$\begin{aligned} \rho &= -|q|N_a \quad \text{para} \quad -x_{p0} \leq x < 0 \\ &e \\ \rho &= |q|N_d \quad \text{para} \quad 0 < x \leq x_{n0} \end{aligned}$$

Equação 180

Portanto:

$$\begin{aligned} \frac{\partial E}{\partial x} &= -\frac{|q|N_a}{\varepsilon} \quad \text{para} \quad -x_{p0} \leq x < 0 \\ &e \\ \frac{\partial E}{\partial x} &= \frac{|q|N_d}{\varepsilon} \quad \text{para} \quad 0 < x \leq x_{n0} \end{aligned}$$

Equação 181

Integrando a Equação 181, temos:

$$E(x) = -|E_0| - \frac{|q|N_a}{\varepsilon}x \quad \text{para} \quad -x_{p0} \leq x < 0$$

e

$$E(x) = -|E_0| + \frac{|q|N_d}{\varepsilon}x \quad \text{para} \quad 0 < x \leq x_{n0}$$

Equação 182

Onde $E_0 = E(x = 0)$. Sabendo que o campo elétrico está confinado na região de depleção, a Equação 182 resulta em uma dependência espacial do campo elétrico como representada na Figura 19.

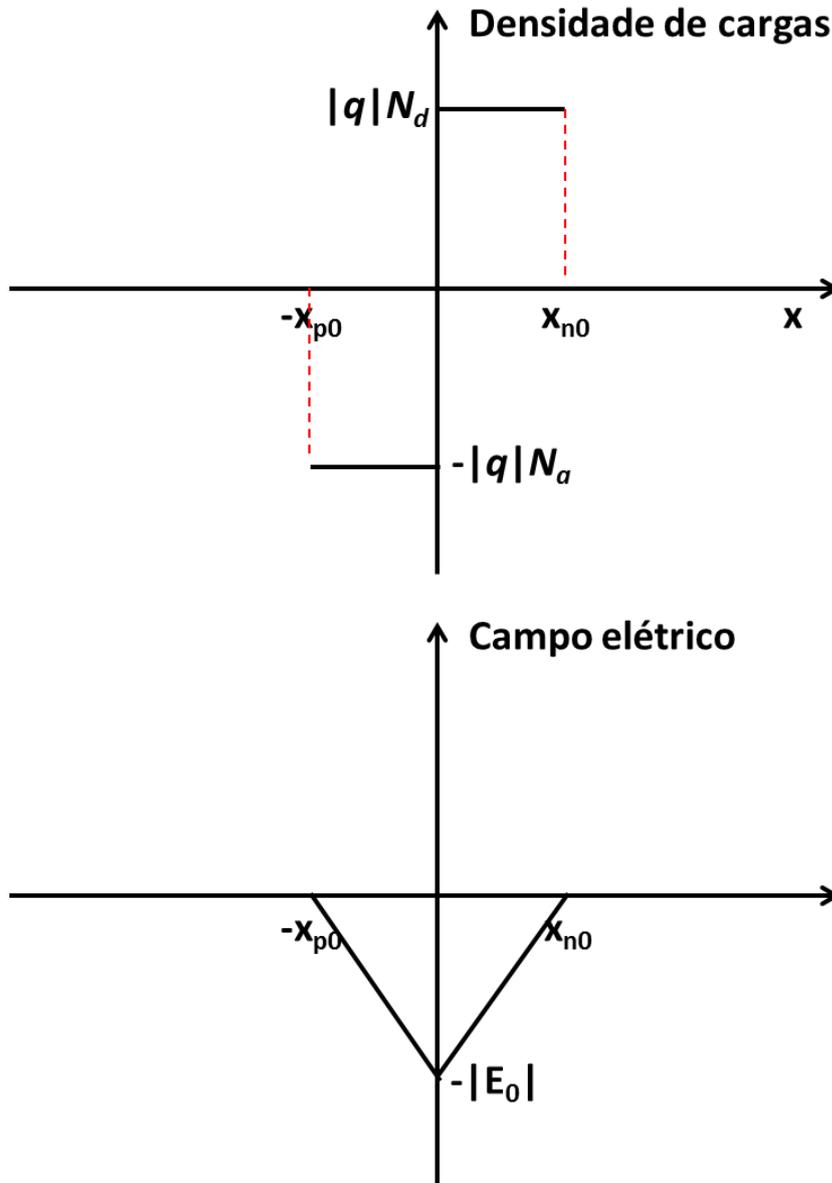


Figura 19 Dependência especial da densidade de cargas e do campo elétrico.

Não é difícil deduzir da Equação 181 e do fato de que o campo está confinado na região de depleção que o valor máximo do módulo do campo é dado por:

$$|E_0| = \frac{|q|N_a x_{p0}}{\varepsilon} = \frac{|q|N_d x_{n0}}{\varepsilon}$$

Equação 183

Note também que, como

$$V_0 = - \int_{-x_{p0}}^{x_{n0}} E(x) dx$$

Então V_0 coincide com área do triângulo da Figura 19, ou seja :

$$V_0 = \frac{(x_{n0} + x_{p0})|E_0|}{2} = \frac{W|E_0|}{2} = \frac{W|q|N_a x_{p0}}{2 \cdot \varepsilon}$$

Equação 184

Note também que, pela Equação 183, temos que:

$$N_a x_{p0} = N_d x_{n0}$$

Portanto:

$$W = x_{p0} + x_{n0} = x_{p0} \left(1 + \frac{N_a}{N_d} \right) = x_{p0} \left(\frac{N_d + N_a}{N_d} \right)$$

Ou seja:

$$x_{p0} = W \left(\frac{N_d}{N_a + N_d} \right)$$

Equação 185

Substituindo a Equação 185 na Equação 184 e isolando W temos finalmente que:

$$W = \sqrt{\frac{2 \cdot \varepsilon \cdot V_0}{|q|} \left(\frac{N_a + N_d}{N_a N_d} \right)}$$

Equação 186

Note que, como tanto as concentrações de dopantes (visto no Capítulo 3) como o potencial eletrostático podem ser expressos em termos dos níveis de Fermi originais, W é função apenas desses níveis de Fermi, da temperatura e da constante dielétrica do semiconductor.

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 n e do semiconductor p. Mas não podemos parar aqui. Se a junção p-n for utilizada como um dispositivo eletrônico, precisamos saber o que ocorre quando uma tensão externa for aplicada. Esta já não é mais uma situação de equilíbrio difusivo, uma vez que a tensão externa modifica o potencial eletrostático, podendo assim induzir o fluxo de cargas, ou seja,

induzir corrente. 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 do seu estado de equilíbrio.

4.2 Sistemas fora do equilíbrio: níveis de quasi-Fermi.

Vimos no capítulo 3 que, para um semiconductor, dopado ou não, as concentrações de elétrons livres e buracos são dadas por:

$$n = n_c \exp\left[-\left(\frac{\mathcal{E}_C - \mathcal{E}_F}{k_B T}\right)\right] \quad \text{ou} \quad n = n_i \exp\left[-\left(\frac{\mathcal{E}_{FI} - \mathcal{E}_F}{k_B T}\right)\right]$$

Equação 187

$$p = n_d \exp\left[-\left(\frac{\mathcal{E}_F - \mathcal{E}_V}{k_B T}\right)\right] \quad \text{ou} \quad p = p_i \exp\left[-\left(\frac{\mathcal{E}_F - \mathcal{E}_{FI}}{k_B T}\right)\right]$$

Equação 188

Partindo dessas duas equações, não é difícil provar (mas você tem que provar) que o produto np é independente do nível de Fermi, ou seja, $np = n_i p_i$ (lembre-se que n_i e p_i denotam as concentrações do semiconductor intrínseco). Portanto, o processo de dopagem não altera o produto np : se você aumentar o nível de Fermi através de dopagem tipo n, você irá aumentar a concentração n de elétrons livres, mas a concentração de buracos p irá diminuir na mesma proporção.

Agora considere a seguinte situação: suponha que você tenha um semiconductor intrínseco ($n = n_i$ e $p = p_i$). Como visto no capítulo 3 (Equação 141), a concentração n e a concentração p do semiconductor 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 semiconductor intrínseco, temos também que elas são iguais, ou seja, $n_i = p_i$). Agora, imagine que você ilumine o semiconductor. 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 semiconductor foi iluminado, a sua energia não é mais apenas térmica, o que significa que o semiconductor não está mais em equilíbrio térmico. Estritamente falando, não podemos mais utilizar as equações deduzidas até aqui, já que todas elas partem da distribuição de Fermi-Dirac, que por sua vez supõe equilíbrio térmico. Note também que, quando o semiconductor é iluminado, o produto np não é mais igual ao produto $n_i p_i$ do semiconductor intrínseco em equilíbrio, já que a

absorção de luz aumenta tanto a concentração de elétrons como a concentração de buracos. Essa variação do produto np não é surpreendente, já que a priori não podemos mais utilizar a Equação 187 e a Equação 188 para expressar a concentração de portadores quando não estamos mais na condição de equilíbrio.

O que fazer José? Jogar fora 100 páginas de notas de aula e um curso inteiro? Não né. Podemos salvar a Equação 187 e a Equação 188 definindo o nível de quasi-Fermi. 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. A motivação para a definição do nível de quasi-Fermi vem da observação que, para que a Equação 187 e a Equação 188 ainda sejam válidas fora do equilíbrio, é necessário atribuir um nível de Fermi para os elétrons e um nível de Fermi para os buracos. Só assim o produto np pode desviar do produto $n_i p_i$. Em termos técnicos, dizemos que, quando o semiconductor sai da condição de equilíbrio, ocorre uma separação entre os níveis de Fermi de buracos e elétrons, que passam a ser descritos pelos níveis de quasi-Fermi. É claro que o valor dos níveis de quasi-Fermi dependem do grau de desequilíbrio. No exemplo do semiconductor iluminado, os níveis de quase-Fermi dependeriam da taxa de absorção de fótons no semiconductor.

Utilizando então o conceito de níveis de quase-Fermi, podemos estender a Equação 187 e a Equação 188 para incluir situações fora do equilíbrio:

$$n = n_c \exp\left[-\left(\frac{\varepsilon_C - \varepsilon_{QFN}}{k_B T}\right)\right] \quad \text{ou} \quad n = n_i \exp\left[-\left(\frac{\varepsilon_{FI} - \varepsilon_{QFN}}{k_B T}\right)\right]$$

Equação 189

$$p = n_d \exp\left[-\left(\frac{\varepsilon_{QFP} - \varepsilon_V}{k_B T}\right)\right] \quad \text{ou} \quad p = p_i \exp\left[-\left(\frac{\varepsilon_{QFP} - \varepsilon_{FI}}{k_B T}\right)\right]$$

Equação 190

Onde ε_{QFN} e ε_{QFP} são os níveis de quasi-Fermi para elétrons e buracos, respectivamente. A Figura 20 ilustra a separação dos níveis de Fermi quando o semiconductor sai do equilíbrio.

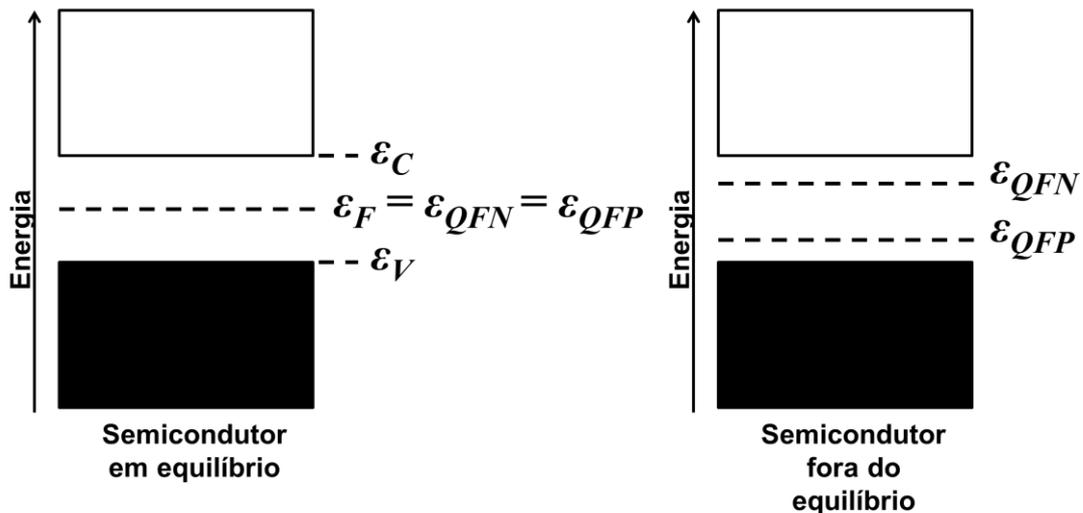


Figura 20 Representação da separação do nível de Fermi quando o semicondutor sai da condição de equilíbrio.

4.3 Relação tensão x corrente na junção p-n (Equação de Shockley).

Chegamos finalmente no último tópico dessas notas de aula, que é a determinação da relação tensão x corrente em uma junção p-n. Mas antes de calcular essa relação, vamos entender qualitativamente o que vai ocorrer quando aplicarmos uma tensão 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 21. Se a tensão maior for aplicada do lado p, como mostrado na Figura 21a, 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 21b, diz-se então que a junção foi polarizada inversamente.

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 for aplicada externamente, o campo elétrico responsável pela corrente de arrasto irá diminuir ou aumentar, conforme a polarização.

Se a junção for polarizada diretamente, então o campo elétrico externo (advindo da tensão externa) é contrário ao campo elétrico interno da região de depleção. Dessa forma, o campo elétrico total na região de depleção é reduzido, reduzindo assim também o potencial eletrostático entre as junções. Mas se o campo e o potencial na região de depleção forem reduzidos, a corrente de arrasto também será reduzida.

Portanto, a corrente de arrasto não será mais suficiente para cancelar toda 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, ou seja, vai do potencial externo maior para o menor. Como tanto a densidade de buracos em p como a densidade de elétrons em n é alta, a corrente de difusão será alta também.

Por outro lado, se a junção for polarizada inversamente, então o campo elétrico e o potencial eletrostático irão aumentar, 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 indo de p para n. Mas existem pouquíssimos buracos na região n e pouquíssimos elétrons na região p. Essa falta de portadores manifesta-se como uma resistência ôhmica altíssima e, conseqüentemente, espera-se que a corrente de arrasto seja 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, temos muitos portadores contribuindo para a corrente, portanto a corrente em polarização direta é alta (resistência baixa). Quando buracos vão de n para p e elétrons de p para n, temos poucos portadores 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.

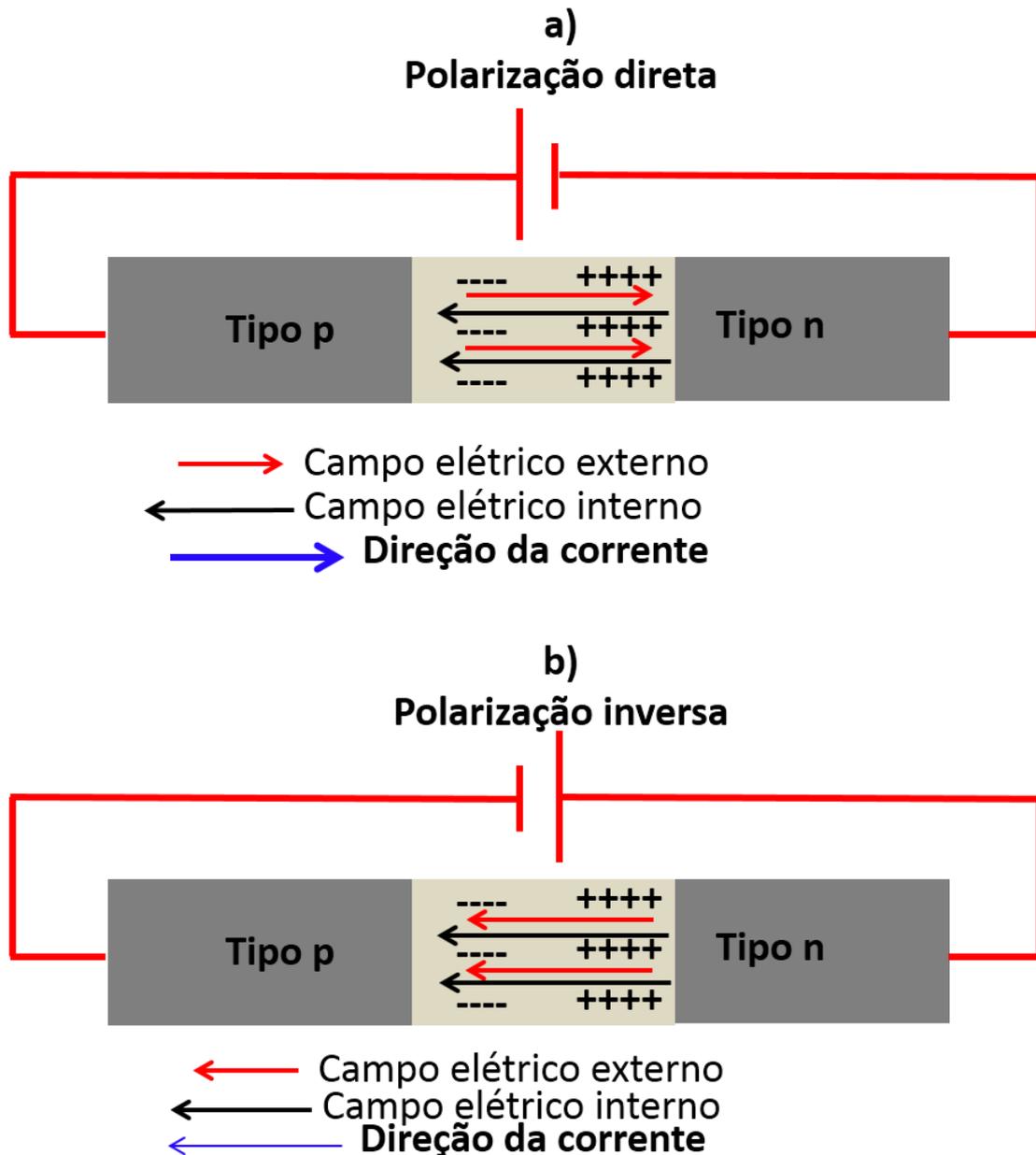


Figura 21 Tipos de polarização da junção p-n

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 22 mostra o diagrama de energia para a junção em repouso (sem polarização) e para a junção diretamente polarizada. 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 é mostrado na Figura 22. 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. Esse comportamento pode ser explicado da seguinte forma: na região n, fora da região de depleção, as cargas não “percebem” a perturbação, ou seja, a densidade de elétrons é praticamente inalterada pela aplicação do potencial externo. Isso ocorre exatamente porque a densidade de elétrons é muito alta nessa região. Como é a diferença entre o nível de quasi-Fermi e a banda de condução que define a densidade de elétrons (Equação 189), essa diferença tem que ser virtualmente idêntica à diferença que ocorria antes da aplicação da tensão externa, ou seja, o nível de quasi-Fermi é praticamente idêntico ao nível de Fermi no equilíbrio (em relação à banda de condução). Na região p, por outro lado, a densidade de elétrons é baixa. Quando a tensão externa for aplicada, haverá uma injeção de elétrons nessa região e, portanto, a variação na densidade de elétrons é proporcionalmente muito maior e, por isso, o nível de quasi-Fermi é bem diferente do nível de Fermi no equilíbrio. Além disso, estes elétrons em excesso irão recombinar com os buracos, causando assim uma forte variação espacial na densidade de elétrons. É claro que, quanto mais longe da região de injeção de cargas, menor será a densidade de elétrons e, portanto, o nível de quasi-Fermi deve cair para o nível em equilíbrio, que é exatamente o nível de quasi-Fermi para buracos, como mostrado na Figura 22. Na região de depleção existe uma leve queda no nível de Fermi, mas que pode ser negligenciada em um primeiro estudo. Um raciocínio análogo explica a variação no nível de quasi-Fermi para buracos.

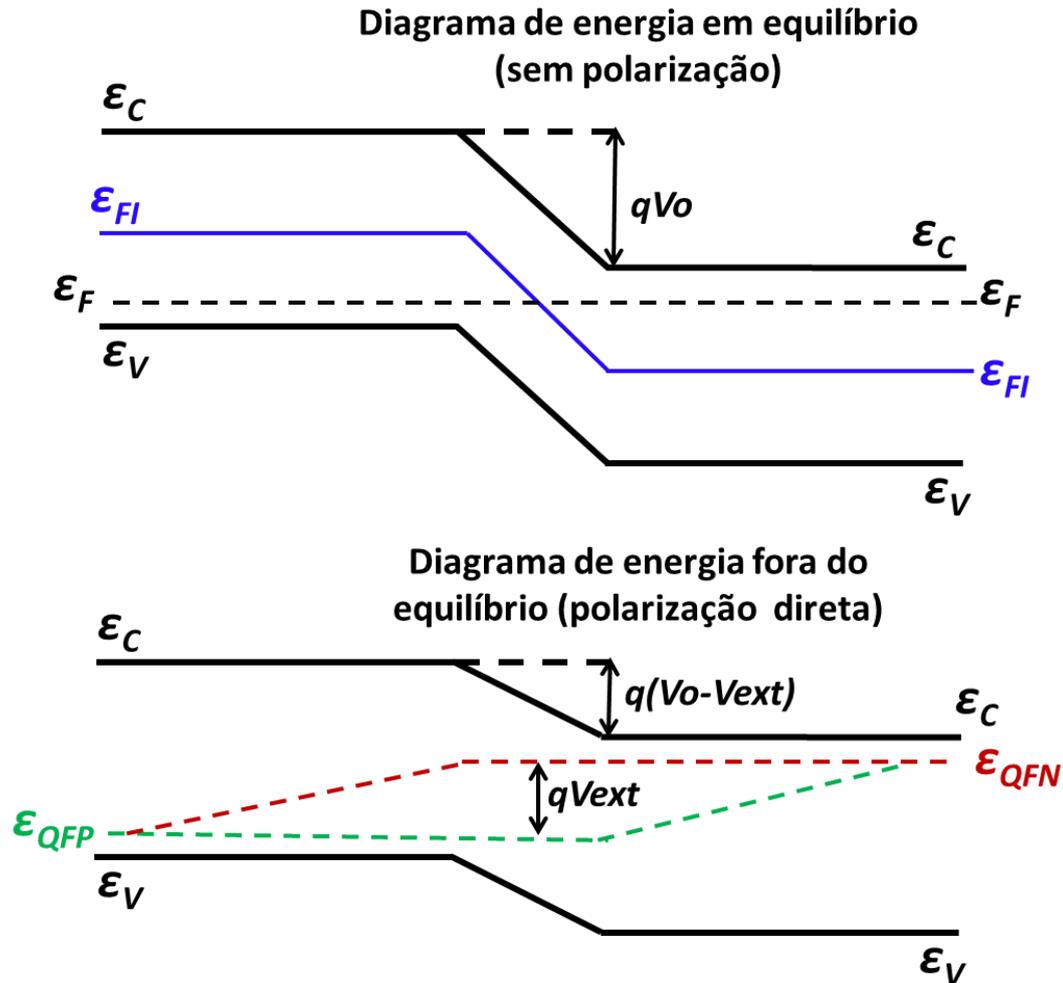


Figura 22 Diagramas de energia dentro e fora do equilíbrio

Note na Figura 22 que o potencial eletrostático total 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 a 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 191

Ou seja, o potencial externo vezes a carga fundamental do elétron é igual à diferença entre os níveis de quasi-Fermi na região de depleção (claro que isso só é estritamente válido quando negligenciamos a variação dos níveis dentro da região de depleção).

O comportamento dos níveis de quasi-Fermi e das densidades de portadores motiva a distinção entre dois regimes de portadores: minoritários e majoritários. Os elétrons

na região n são os portadores majoritários e os elétrons 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. A contribuição de cada portador é mostrada na Figura 23. 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 (neste 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 quase-Fermi na região de depleção podem ser negligenciadas.

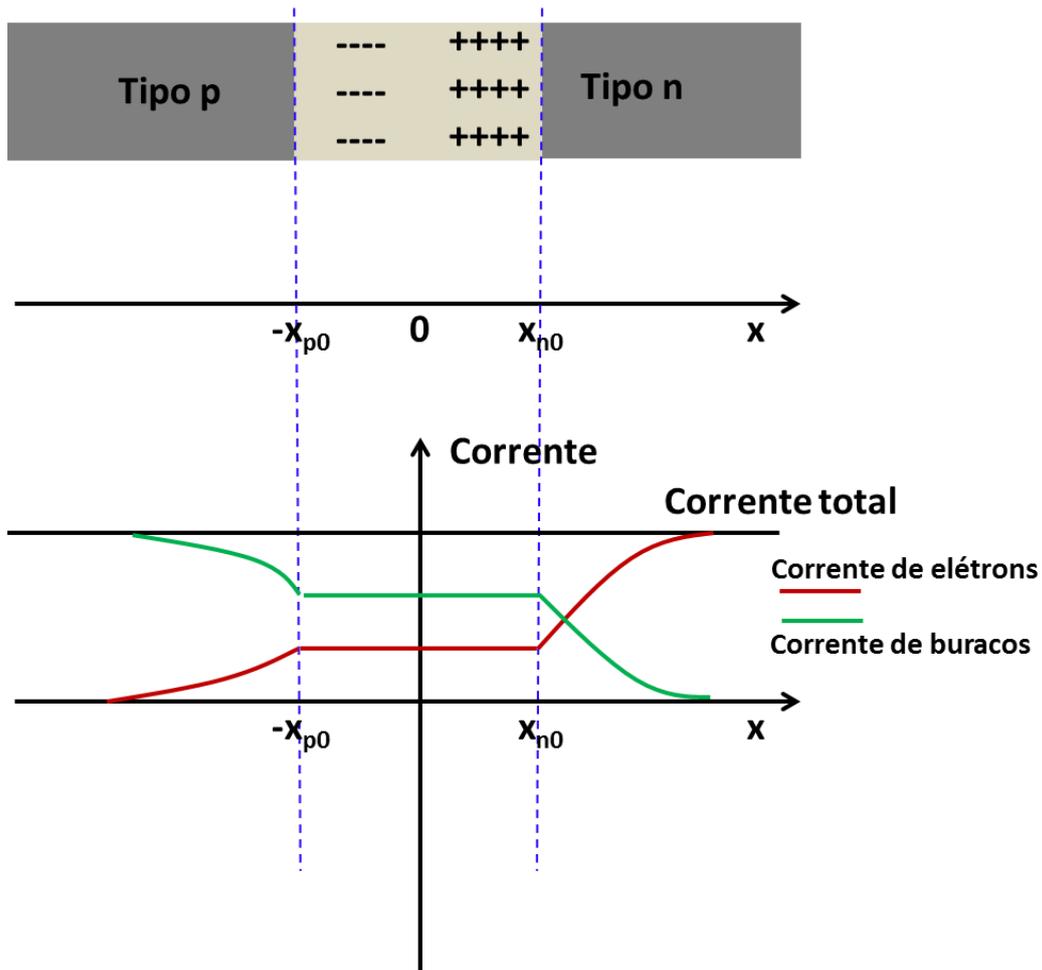


Figura 23 Contribuições das correntes de elétrons e buracos para a corrente total

Existem várias maneiras de se calcular a corrente total na junção p-n. Uma das mais simples e elegantes utiliza o fato de que a corrente de difusão para portadores minoritários é dominada pelo processo de recombinação e, por isso, tem uma expressão muito simples. Você teve que deduzir na lista de exercícios de transporte que a densidade de excesso de portadores decai exponencialmente devido ao processo de recombinação. Em particular, a densidade de excesso de portadores minoritários na região n (ou seja, buracos) é dada por:

$$p_e(x) = p_e(x_{n0}) \exp\left[-\frac{(x - x_{n0})}{L_p}\right] \text{ para } x \geq x_{n0}$$

Equação 192

Onde p_e é o excesso de buracos (onde excesso é definido como na lista de transporte, ou seja, excesso em relação à densidade no equilíbrio) e L_p é o comprimento de difusão de buracos (cuja expressão em termos do coeficiente de

difusão e tempo de vida também está na lista de transporte). Analogamente, a densidade de excesso de portadores minoritários na região p é dada por:

$$n_e(x) = n_e(-x_{p0}) \exp\left[\frac{(x+x_{p0})}{L_n}\right] \quad \text{para } x \leq x_{p0}$$

Equação 193

Como a corrente de difusão depende somente do coeficiente de difusão e da densidade de portadores, temos que a corrente de difusão de portadores minoritários (J_{DM}) é dada por:

$$J_{DMP}(x) = -|q|D_p \frac{\partial p(x)}{\partial x} = -|q|D_p \frac{\partial (p_e(x) + p_0)}{\partial x} = -|q|D_p \frac{\partial p_e(x)}{\partial x} \quad \text{para } x \geq x_{n0}$$

∴

$$J_{DMP}(x) = \frac{|q|D_p}{L_p} p_e(x_{n0}) \exp\left[-\frac{(x-x_{n0})}{L_p}\right] \quad \text{para } x \geq x_{n0}$$

e

$$J_{DMN}(x) = +|q|D_n \frac{\partial n(x)}{\partial x} = |q|D_n \frac{\partial (n_e(x) + n_0)}{\partial x} = |q|D_n \frac{\partial n_e(x)}{\partial x} \quad \text{para } x \leq -x_{p0}$$

∴

$$J_{DMN}(x) = \frac{|q|D_n}{L_n} n_e(x_{n0}) \exp\left[\frac{(x-x_{n0})}{L_n}\right] \quad \text{para } x \leq -x_{p0}$$

Equação 194

Onde J_{DMP} e J_{DMN} são as correntes de difusão para buracos e elétrons, respectivamente. Note que p_0 e n_0 denotam a concentração de buracos e elétrons em equilíbrio (lembre-se que essas concentrações não dependem de x , por isso que suas derivadas são zero).

Sabemos que a corrente ao longo da junção tem que ser constante; portanto, a corrente total é igual à em corrente total em qualquer ponto ao longo do eixo x . Assim, para determinar a corrente total, podemos determinar a corrente em qualquer ponto. Em particular, a corrente total na região de depleção é constante e, portanto, temos que:

$$I = A(J_{DMN}(x = -x_{p0}) + J_{DMP}(x = x_{n0}))$$

Equação 195

Onde A é área transversal do junção p-n. Utilizando as expressões da Equação 194, temos que:

$$I = A|q| \left(\frac{D_n}{L_n} n_e(-x_{p0}) + \frac{D_p}{L_p} p_e(x_{n0}) \right)$$

Equação 196

Tudo o que precisamos fazer para encontrar a corrente, portanto, é achar o excesso de elétrons e buracos nas fronteiras da região de depleção. Para isso, basta utilizarmos a Equação 189 e a Equação 190. Dessas relações, temos que:

$$n(-x_{p0}) = n_i \exp \left[- \left(\frac{\varepsilon_{FI}(-x_{p0}) - \varepsilon_{QFN}(-x_{p0})}{k_B T} \right) \right]$$

Equação 197

e

$$p(x_{n0}) = p_i \exp \left[- \left(\frac{\varepsilon_{QFP}(x_{n0}) - \varepsilon_{FI}(x_{n0})}{k_B T} \right) \right]$$

Equação 198

Podemos colocar essas expressões em uma forma mais útil. Note primeiro que:

$$n(x_{n0}) = n_i \exp \left[- \left(\frac{\varepsilon_{FI}(x_{n0}) - \varepsilon_{QFN}(x_{n0})}{k_B T} \right) \right]$$

Equação 199

Portanto

$$p(x_{n0})n(x_{n0}) = p_i n_i \exp \left(\frac{\varepsilon_{QFN}(x_{n0}) - \varepsilon_{QFP}(x_{n0})}{k_B T} \right)$$

Equação 200

Substituindo a Equação 191 na Equação 200, temos:

$$p(x_{n0})n(x_{n0}) = p_i n_i \exp\left(\frac{|q|V_{ext}}{k_B T}\right)$$

Equação 201

Note também que $n(x_{n0})$ é a concentração de elétrons na região n, ou seja, é a concentração de portadores majoritários na região n. Mas a concentração de portadores majoritários é pouco alterada pela aplicação da tensão externa. Assim, a concentração de portadores majoritários é praticamente igual à concentração de portadores majoritários quando o semiconductor está em equilíbrio, ou seja:

$$n(x_{n0}) \approx n_0(x_{n0})$$

Equação 202

Onde n_0 é a concentração de elétrons na região n quando o semiconductor está em equilíbrio (sem tensão externa aplicada). Mas como vimos anteriormente o produto np em equilíbrio é constante, o que permite re-escrever a Equação 201 como:

$$p(x_{n0})n(x_{n0}) = p_0(x_{n0})n_0(x_{n0}) \exp\left(\frac{|q|V_{ext}}{k_B T}\right)$$

Equação 203

Substituindo a Equação 202 na Equação 203, temos finalmente que:

$$p(x_{n0}) = p_0(x_{n0}) \exp\left(\frac{|q|V_{ext}}{k_B T}\right)$$

Equação 204

Lembre-se que $p_0(x_{n0})$ é a concentração de buracos no ponto x_{n0} para a junção em equilíbrio, e que aprendemos a calcular essa concentração no capítulo 3 (basta utilizar a Equação 188 com o nível de Fermi do semiconductor tipo n, ou seja, com $\epsilon_F = \epsilon_{FN}$, onde ϵ_{FN} é mostrado na Figura 15). Um procedimento análogo (tente sozinho) leva à:

$$n(-x_{p0}) = n_0(-x_{p0}) \exp\left(\frac{|q|V_{ext}}{k_B T}\right)$$

Equação 205

Estamos quase lá. Já temos as expressões para os portadores minoritários nas fronteiras da região de depleção. Mas o que aparece na Equação 196 são os excessos de portadores. Mas:

$$p_e(x_{n0}) = p(x_{n0}) - p_0(x_{n0}) = p_0(x_{n0}) \left[\exp\left(\frac{|q|V_{ext}}{k_B T}\right) - 1 \right]$$

e

$$n_e(-x_{p0}) = n(-x_{p0}) - n_0(-x_{p0}) = n_0(-x_{p0}) \left[\exp\left(\frac{|q|V_{ext}}{k_B T}\right) - 1 \right]$$

Equação 206

Substituindo a Equação 206 na Equação 196, temos finalmente que:

$$I = A|q| \left[\frac{D_n}{L_n} n_0(-x_{p0}) + \frac{D_p}{L_p} p_0(x_{n0}) \right] \left[\exp\left(\frac{|q|V_{ext}}{k_B T}\right) - 1 \right]$$

Equação 207

O termo entre parênteses depende do semiconductor (através de D_n , D_p , L_n e L_p) e também das dopagens (através de $n_0(-x_{p0})$ e $p_0(x_{n0})$). Esse termo é chamado de corrente de saturação e denominado pelo símbolo I_0 (às vezes I_S):

$$I_0 = A|q| \left[\frac{D_n}{L_n} n_0(-x_{p0}) + \frac{D_p}{L_p} p_0(x_{n0}) \right]$$

Equação 208

Assim temos que:

$$I = I_0 \left[\exp\left(\frac{|q|V_{ext}}{k_B T}\right) - 1 \right]$$

Equação 209

A Equação 209 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 eletricitista não saber quem é William B. Shockley é tão mau 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 reversa com $|V_{ext}| < k_B T / |q|$, a exponencial passa a ser bem menor que um e , portanto, $I \sim -I_0$. Como I_0 é tipicamente bem pequeno, a junção praticamente não conduz para polarização reversa. Para polarização direta, por outro lado, a corrente aumenta exponencialmente com a tensão. Procure na internet um gráfico da relação corrente x tensão em um diodo como ilustração do comportamento da equação de Shockley.

Como afirmado no capítulo 1, o objetivo deste curso é deduzir a equação de Shockley. Por problemas de espaço e tempo, foi necessário omitir algumas noções, como por exemplo a capacitância de junções p-n. A escolha do que incluir e do que omitir, entretanto, foi feita levando em consideração o grau de dificuldade que o aluno teria em compreender os conceitos. Os tópicos incluídos, portanto, 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, este último tendo sido omitido desse texto e do curso.

Nesse curso, você aprendeu as noções que levam à equação de Shockley. Você deve então ser capaz de responder a pergunta “por que em um diodo a corrente apresenta uma dependência exponencial em relação à tensão?” (lembre-se diodos nada mais são que junções p-n). Se você rebobinar a fita, você 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 e não pode ser medido com um voltímetro colocado nos seus terminais?” Você deve lembrar-se 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, 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. Transistores nada mais são que duas junções p-n em série. Portanto, o entendimento adequado da física 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.

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

FIM

