

UNIVERSIDADE DE BRASÍLIA
INSTITUTO DE FÍSICA

ALICE AMARAL DE SOUZA

**MEDIDA DE INFORMAÇÃO NA DINÂMICA
MOLECULAR**

BRASÍLIA
05 DE NOVEMBRO DE 2021

Alice Amaral de Souza

Medida de informação na dinâmica molecular

Monografia apresentada ao Instituto de Física da Universidade de Brasília como parte dos requisitos necessários à obtenção do título de Bacharel em Física.

Orientador: Bernardo de Assunção Melo

Universidade de Brasília – UnB

Instituto de Física

Brasília

05 de novembro de 2021

Alice Amaral de Souza

Medida de informação na dinâmica molecular/ Alice Amaral de Souza . –
Brasília, 05 de novembro de 2021-
101 p. : il. (algumas color.) ; 30 cm.

Orientador: Bernardo de Assunção Melo

Monografia – Universidade de Brasília – UnB
Instituto de Física, 05 de novembro de 2021.

1. Entropia. 2. Informação. I. Bernardo de Assunção Melo. II. Universidade de
Brasília. III. Instituto de Física. IV. Medida de informação na dinâmica molecular.

CDU 02:141:005.7

Alice Amaral de Souza

Medida de informação na dinâmica molecular

Monografia apresentada ao Instituto de Física da Universidade de Brasília como parte dos requisitos necessários à obtenção do título de Bacharel em Física.

Trabalho aprovado. Brasília, 05 de novembro de 2021:

Bernardo de Assunção Melo
Orientador

Olavo Leopoldino da Silva Filho
Convidado 1

Annibal Dias de Figueiredo
Convidado 2

Brasília
05 de novembro de 2021

Agradecimentos

De forma bem sucinta, agradeço ao tempo.

Agradeço ao tempo, que foi o motivo pelo qual decidi me aprofundar no tema que o envolve. Ao tempo que me permite crescer e evoluir a cada dia, mês, ano e a cada segundo... ou microsegundo. Ao tempo que levei pra compreender e concluir esse trabalho, desde o momento que refleti sobre a entropia pela primeira vez há 3 anos atrás. O tempo, que é uma questão tão presente dentro das minhas indagações, me ensina e me mostra, quotidianamente, que o segredo está na frequência.

Agradeço ao meu tempo de vida. Ao tempo passado, presente e futuro e, sobretudo, ao tempo que me é dado. Tempo dado no sentido de troca. Ao tempo que tenho de ser e de estar - ao lado de tantas pessoas que me agregam um tanto e contribuem para minha formação como ser humana.

Um agradecimento mais que especial aos meus pais, que me conceberam o tempo! Também agradeço ao tempo que dedicaram à mim e à minha educação, por seguirem me apoiando e fortalecendo todos os dias para que eu conclua meus projetos e realize meus sonhos. Sem eles, eu não seria nada. Agradeço às minhas irmãs, Catarina e Mariana, que me permitem ser a minha versão mais verdadeira em tempo integral! Ao tempo que passamos e rimos juntas e que nos proporciona momentos de reflexão e aprendizados. Vocês me inspiram e me fazem querer ser uma pessoa melhor no mundo!!!

Agradeço à minha vó Tereza, a matriarca da família, que me mostra o quanto o tempo é precioso, finito e deve ser bem aproveitado. Agradeço também aos meus tios, primos e a todos aqueles que fazem ou fizeram parte da minha árvore genealógica e que compartilham de um mesmo saber ancestral. Agradeço ainda ao meu companheiro de vida, Davi, que me deu todo o suporte emocional necessário durante o desenvolvimento desse trabalho e com o qual eu passo grande parte do meu tempo, amando e sendo amada.

Lembro do meu primeiro teste de mecânica no primeiro semestre da faculdade, há 5 anos atrás. Só havia uma pergunta: o que é tempo? Eu mal poderia imaginar que essa questão desencadearia várias outras dentro da minha cabeça e que isso viria a se tornar o tema do meu trabalho final de curso; graças à professora Vanessa, que inspira muito como mulher e física. Agradeço também ao professor Olavo, que me acompanhou no início, meio e fim do curso; que me orientou no PIBIC e, agora, me avalia no TCC, juntamente com o professor Annibal, o qual também sou grata. Agradeço ao professor Antony, que me abriu a cabeça diversas vezes durante as aulas de teoria eletromagnética A e contribuiu com o meu trabalho através de comentários pertinentes a respeito do tema escolhido. Agradeço ao professor Hugo Nazareno, o qual tive a honre de conhecer durante o processo de

amadurecimento do tema e incentivou o prosseguimento do estudo. Agradeço também ao professor Cássio, que embora não tenha participado de forma direta do trabalho, fomentou várias discussões e mostrou a importância de se entender a evolução histórica dos conceitos da física durante disciplina ministrada por ele. Por último, e não menos importante, um agradecimento ao professor Bernardo, que me orientou e possibilitou o desenvolvimento desse trabalho, além de ter se mostrado um verdadeiro amigo. Sem ele eu jamais teria conseguido. Obrigada pela confiança e paciência! Também agradeço aos demais professores que tive durante o curso e que me apresentaram o magnífico universo da física.

Por fim, agradeço à todos os amigos, dentro e fora da UnB, que estão e somam comigo. Sem esses, a vida com certeza seria mais difícil. Um agradecimento especial àqueles que me acompanharam durante toda a jornada dentro da física e me deram coragem para continuar. Àqueles que compartilharam da dor mas também dos momentos de glória pós prova. Àqueles que tornaram os momentos de desespero um pouco menos sacrificantes. Agradeço aos meninos, João Valeriano, Pedro Cintra e João Augusto, que me acolheram no curso e me ensinaram um tanto, admiro-os por tamanha genialidade. À Mariana Casement, amiga que admiro muito e que foi minha companheira de laboratório por vários semestres, juntamente com o Victor Porto e o Caio Lima; amigos queridos que tive o prazer de dividir relatórios. Agradeço ao Marquinho, que foi meu grande ombro amigo, que dividiu tardes de estudo comigo e me explicou diversas vezes a matéria. Ao Victor Vendramini, que também é um amigo que admiro muito e que me compreende como ninguém, compartilhando das mesmas inquietações. Agradeço às meninas Debora, Carol, Tábata, Amanda, Graciele e, principalmente, à Lorena Reis e à Lorena Carmo por serem mulheres incríveis dentro da física e promoverem um espaço acadêmico mais democrático, gratidão por me inspirarem tanto! Serei eternamente grata às risadas, brincadeiras, lamúrias e surpresas que vivemos juntos e que tornaram tudo mais leve, meus amigos.

Resumo

O presente trabalho tem como objetivo investigar as interpretações existentes a respeito do conceito de entropia a fim de correlacioná-las a partir de um método analítico, buscando também encontrar pontos de desacordo entre elas. Após realizar uma análise histórica conceitual no âmbito da termodinâmica e física estatística, discute-se uma nova interpretação de entropia proposta pela Teoria da Informação de Shannon. Nesse sentido, associaremos a medida de informação a quantidade de bits necessários para armazenar computacionalmente a posição e a velocidade de todas as partículas de um sistema contido em uma região do espaço de configuração. Para isso, assumiremos a existência de um algoritmo de compressão que se beneficie das simetrias, correlações, periodicidades e indistinguibilidade das partículas para validar o primeiro Teorema da Codificação desenvolvido por Shannon. Por fim, comparando a quantidade média de informação referente às posições de partículas distinguíveis e indistinguíveis, percebe-se que há realmente uma economia de bits utilizados para armazenamento desses dados. Além disso, verifica-se uma certa similaridade entre a medida de informação obtida e a expressão equivalente à entropia termodinâmica do gás ideal, concluindo que há uma conexão entre os dois conceitos.

Palavras-chaves: Entropia. Informação. Tempo. Shannon.

Abstract

This study investigated the many interpretations regarding the definition of entropy, aiming to compare and find discrepancies between them. After the historical analysis of the entropy's definition throughout time, from a thermodynamic and statistical physics perspective, it is discussed a theory that correlates entropy and the information measure, using the same mathematical expression, named Shannon's Information. We will associate the measure of information to the amount of bits needed to store the position and velocity of all the particles of a system in a given phase space. We will assume the existence of a compression algorithm that benefits from symmetries, correlations, cycles, the indistinguishable aspect of the particle, so we can validate the first Code Theorem developed by Shannon. After comparing the average information available of the particles position, distinguishably and indistinguishably, it is possible to notice that there is a "bits economy" on the storage of this data. Besides, it is observed similarities between the amount of information gathered and the equivalent formula of the thermodynamic entropy of the ideal gas, which leads to a connection between this two concepts.

Key-words: entropy. information. time. Shannon.

Lista de ilustrações

| | |
|---|-----|
| Figura 1 – O problema central da termodinâmica consiste em determinar o estado de equilíbrio de um sistema depois de remover uma membrana interna em um sistema composto. Fonte: (CALLEN, 1998) | 26 |
| Figura 2 – Mudanças de estado adiabáticas. | 31 |
| Figura 3 – Diagrama de um sistema de comunicação. Fonte: (SHANNON, 1948) | 54 |
| Figura 4 – Formato de armazenamento em bits, bytes e word. | 73 |
| Figura 5 – Estrutura de ponto flutuante de acordo com a IEEE-754. | 77 |
| Figura 6 – Em um sistema bidimensional, o espaço quadrado de comprimento L é dividido em cubículos de comprimento δ^r , utilizadas para definir as posições das partículas com a precisão necessária para a evolução numérica da dinâmica. As caixas de dimensão Δ^r são utilizadas pelo algoritmo de compressão. Os índices originais i são mostrados em preto. Os índices α definidos pelo algoritmo de ordenação descritos em 7.10 são mostrados em vermelho. | 86 |
| Figura 7 – Diagrama de Venn e o vínculo entre as medidas de informação de Shannon. | 100 |

Lista de tabelas

| | |
|---|----|
| Tabela 1 – Tabela ASCII - Código Padrão Americano para o Intercâmbio de Informação. Disponível em < https://www.asciitable.com/ > | 75 |
| Tabela 2 – Números reais de 1.0 a 5.0 armazenados no formato de ponto flutuante com precisão simples. | 78 |
| Tabela 3 – Correspondência entre os índices i e α das partículas contidas na caixa de dimensão L , mostradas na Fig.6. | 86 |

Lista de símbolos

| | |
|----------|-----------------------------|
| S | entropia |
| T | temperatura |
| U | energia interna |
| V | volume |
| μ | potencial eletroquímico |
| Q | calor |
| Γ | número de estados quânticos |
| τ | tempo de relaxamento |
| \hbar | constante de Planck |
| \prec | precedido |

Sumário

| | | |
|-----------|--|-----------|
| | Introdução | 19 |
| I | INTERPRETAÇÕES DO CONCEITO DE ENTROPIA | 23 |
| 1 | DEFINIÇÃO TERMODINÂMICA | 25 |
| 1.1 | Contexto histórico e origem do conceito entropia | 25 |
| 1.2 | Entropia como função termodinâmica | 26 |
| 1.3 | Irreversibilidade e a seta do tempo | 29 |
| 1.4 | Entropia do gás ideal | 34 |
| 2 | ENTROPIA ESTATÍSTICA | 37 |
| 2.1 | História e contexto da entropia estatística | 37 |
| 2.2 | O espaço de configurações | 38 |
| 2.3 | Entropia como uma medida da desordem? | 42 |
| 2.4 | Entropia de Gibbs | 44 |
| 2.5 | Entropia do gás ideal | 49 |
| 3 | ENTROPIA INFORMACIONAL | 53 |
| 3.1 | História e contexto da teoria da informação | 53 |
| 3.2 | Entropia como medida de informação | 55 |
| 3.3 | Unidade de Informação em bits | 56 |
| 3.4 | Entropia de sistemas contínuos | 58 |
| 3.5 | Teoremas da codificação | 59 |
| 3.5.1 | Primeiro Teorema da Codificação | 59 |
| 3.5.2 | Segundo Teorema da Codificação | 61 |
| 4 | COMPARAÇÃO DAS ENTROPIAS | 65 |
| II | MEDIDA DA INFORMAÇÃO | 71 |
| 5 | COMO MEDIR INFORMAÇÃO | 73 |
| 5.1 | Variáveis inteiras | 73 |
| 5.2 | Variáveis de ponto flutuante | 74 |
| 5.3 | Algoritmos de compressão | 78 |
| 6 | INFORMAÇÃO DE UMA PARTÍCULA | 81 |

| | | |
|------------|---|-----------|
| 7 | INFORMAÇÃO DE UM SISTEMA DE PARTÍCULAS | 83 |
| 7.1 | Partículas distinguíveis não correlacionadas | 83 |
| 7.2 | Partículas distinguíveis em equilíbrio termodinâmico | 84 |
| 7.3 | Partículas indistinguíveis | 85 |
| | Conclusão | 89 |
| | REFERÊNCIAS | 93 |
| | ANEXOS | 95 |
| | ANEXO A – CONCEITOS USADOS NA TEORIA DE SHANNON | 97 |
| A.1 | Desigualdade de informação | 97 |
| A.2 | Entropia Relativa ou Divergência | 97 |
| A.3 | Informação Mútua e Entropia Condicional | 98 |
| A.4 | A relação entre as medidas de informação | 99 |

Introdução

A definição de *entropia* é, até hoje, uma questão em aberto e de difícil compreensão para vários físicos devido às inúmeras interpretações para o mesmo conceito. Entretanto, apesar de possuir diversas interpretações, é uma quantidade que tem a propriedade de ser única e, portanto, sua medida deve ser a mesma independente do método utilizado para quantificá-la, computacionalmente ou não (LIEB; YNGVASON, 2000).

Na teoria da comunicação, *informação* é a medida de liberdade de escolha ao selecionar uma mensagem e pode ser calculada utilizando mecanismos semelhantes aos da física estatística. Por isso está relacionada ao conceito de entropia que conhecemos da termodinâmica. Pelas palavras do professor Antony Polito, do Instituto de Física da UnB, “informação é simplesmente um padrão - construído a partir de algum tipo de entidade física, mas estritamente independente dela - que, uma vez produzido, é capaz de desencadear algum processo em um sistema que seja capaz de reconhecer o padrão” (comentário feito por e-mail).

Se por um lado, a entropia da termodinâmica tem uma definição elementar fundamental e fenomenológica, por outro, a física estatística procura explicar o porquê da 2ª lei ter bons resultados através de uma formulação matemática rigorosa (LIEB; YNGVASON, 2000). Segundo esses autores, alguns físicos defendem a ideia de que as leis da termodinâmica são incompletas e podem ser derivadas da física estatística apenas estudando a mecânica aplicada a um grande número de partículas.

A mecânica estatística é de fato uma poderosa ferramenta para a compreensão de vários fenômenos físicos que lidam com grandes populações e aproximações, em especial em sistemas no, ou próximo do, equilíbrio, embora ainda seja bastante complicado obter valores tão precisos (de entropia, por exemplo) mesmo com uma implementação computacional (LIEB; YNGVASON, 2000). No entanto, a existência da segunda lei, tida como uma das leis mais incontestáveis da física, deve possuir uma lógica fundamental independente de modelos matemáticos e do fato de que a matéria é composta de partículas que interagem entre si. No trecho retirado¹ do artigo (LIEB; YNGVASON, 2000))

Uma teoria é tanto mais impressionante quanto maior a simplicidade de suas premissas, quanto mais diferentes tipos de coisas ela relaciona e quanto mais extensa sua área de aplicabilidade. Portanto, a impressão profunda que a termodinâmica clássica causou em mim. É a única teoria física de conteúdo universal a respeito do qual estou convencido de que, no quadro da aplicabilidade de seus conceitos básicos, nunca será derrubada. (Albert Einstein)

¹ Essa e outras traduções do texto foram realizadas por mim.

Tendo em vista que a definição de entropia envolve problemas de contagem de possibilidades, percebe-se que seu conceito pode ser amplamente aplicado em diversas áreas do conhecimento, assim como o conceito de informação também o é. Por esse motivo, além de estudar individualmente as interpretações de entropia, analisando a evolução histórica conceitual sofrida ao longo da história, este trabalho tem como intuito concatená-la ao conceito de *informação* dentro do contexto da dinâmica molecular.

O conceito de entropia surge, então, na termodinâmica, cuja definição é dada pela segunda lei. Descrevemos resumidamente o contexto que possibilitou seu aparecimento e introduzimos a definição termodinâmica de entropia, pela formulação de Callen, no primeiro capítulo desse trabalho. Também discutimos a relação entre entropia e irreversibilidade a partir das ideias de Robin Giles, que derivou logicamente o conceito de entropia através do método da comparação, o qual estabelece uma conexão profunda com o conceito de tempo. Por último, na seção 1.4, demonstraremos brevemente a expressão de entropia no caso do gás ideal.

No segundo capítulo, fazemos uma abordagem similar à do primeiro. Explicamos rapidamente a evolução do conceito de entropia no âmbito da física estatística, juntamente com uma elucidação histórica. Em seguida, demonstramos a relação entre o espaço de configuração e a entropia de um sistema, seguindo a derivação de Landau. Também discute-se a entropia como medida de desordem, onde verificam-se que as implicações estatísticas da evolução temporal se qualifica a reversibilidade como fenômeno improvável na natureza. Na seção 2.4, apresentamos a interpretação de entropia dada por Gibbs e por último, a interpretação de Boltzmann para o caso do gás ideal na seção 2.5.

Seguindo a linha cronológica, nos anos 50, Claude Elwood Shannon surge com um novo significado de entropia que, embora tenha uma interpretação diferente das outras duas, possui uma mesma representação matemática. Aqui a entropia aparece como uma quantidade média de informação, estabelecendo o quanto de incerteza ainda se tem sobre um sistema. Shannon busca quantificar a informação sem aventurar-se a oferecer uma definição universal *do que é* a informação, mas esclarecer *como* ocorre sua transmissão.

Assim, no terceiro capítulo procuramos explicar, primeiramente, o contexto histórico em que se desenvolve a teoria da informação e como ela opera. Em seguida, demonstra-se como Shannon deriva matematicamente a fórmula de entropia através de uma aproximação axiomática, associando à entropia o conceito de informação. Dessa correlação, define *bit* como a unidade de medida da informação, palavra que foi, posteriormente, apropriada com um significado de “dígito binário” na ciência da computação. Considerando ainda, que a informação é representada por códigos e que estes geralmente estão associados a variáveis discretas, definimos então a entropia de Shannon para variáveis contínuas. Por último, na seção 3.5, apresentamos os teoremas da codificação de Shannon, em que se discutem os principais resultados obtidos por ele e os quais revolucionaram a forma como

lidamos com dados no universo digital.

Após compreender a entropia sob três perspectivas diferentes, uma análise conceitual de entropia é feita no capítulo 4, no qual destaca-se os pontos em comum e os pontos de divergência entre as definições.

No quinto capítulo nos preocupamos em esclarecer como a medida de informação será feita. Aqui, nos referimos à informação como a quantidade de dados necessários para armazenamento de dados em sistemas computacionais. Por isso, além de expor os tipos de dados e a notação usada para representá-los, comenta-se brevemente como esses dados são armazenados, explicitando o formato de ponto fixo e de ponto flutuante para variáveis inteiras e reais.

Portanto, na última parte do trabalho realizamos medidas de informação dentro da dinâmica molecular com a finalidade de compará-la à expressão de entropia dos gases ideais prevista pela termodinâmica. Assim, no capítulo 6, abordamos o caso isolado de uma partícula para, depois, estender para o caso geral envolvendo mais partículas no capítulo 7. Nesse último, assumiremos a existência de um algoritmo de compressão que reduza a quantidade de informação, em bits, necessária para armazenar os dados de um sistema de partículas distinguíveis não correlacionadas se elas passarem a compor um sistema de partículas indistinguíveis.

Com esse trabalho concluímos que a entropia é um conceito amplo, capaz de definir não só a direção que o tempo corre, como também estabelece o conceito de temperatura, a partir da definição do fluxo de calor. Além disso, observa-se que a quantidade de informação que se tem sobre um sistema está diretamente ligada à sua entropia e, dada uma distribuição de probabilidade, é possível inferir se o sistema está em um estado de menor ou maior entropia.

Finalmente, ao se relacionar a definição estatística e termodinâmica à teoria da informação, observa-se que a entropia é capaz não só de medir a quantidade de informação, mas que também pode ser usada como uma forma de caracterizar nossa ignorância à respeito de um sistema. Percebe-se uma grande correspondência entre as entropias proposta por Gibbs, Boltzmann e Shannon, no que diz respeito ao seu caráter probabilístico. Existem inúmeras formas de se observar e caracterizar um sistema, e o conhecimento de quantidades macroscópicas reduz o grau de incerteza que temos a respeito de detalhes do seu estado microscópico.

Parte I

Interpretações do conceito de entropia

1 Definição termodinâmica

A concepção de entropia surge inicialmente no contexto da termodinâmica em 1850 por Rankine. Começamos nosso estudo fazendo uma análise histórica e contextual para entender de onde vieram as motivações e os caminhos que levaram os pensadores da época a tais conclusões. Neste primeiro capítulo, apresentamos a entropia como um conceito que transcende a simples definição da Segunda Lei da termodinâmica.

1.1 Contexto histórico e origem do conceito entropia

Noções básicas de calor e temperatura existem desde os anos 1600, quando cientistas supuseram corretamente uma relação entre calor e o movimento dos constituintes microscópicos da matéria (WOLFRAM, 2002). A curiosidade em saber o que é calor, energia e trabalho levou pensadores da época a estudar a mecânica de sistemas macroscópicos tomando como base as suas propriedades médias e fundamentais fazendo, assim, nascer o que chamamos de termodinâmica ou física térmica. Mas o que impulsionou de fato os estudos da termodinâmica foi a vontade de acelerar os processos industriais, desenvolver máquinas a vapor e diminuir o tempo de transporte de bens de consumo até os mercados consumidores, época marcada pela primeira revolução industrial - entre a segunda metade do século XVIII e meados de 1840 (NASCIMENTO; BRAGA; FABRIS, 2004). Nicolas Léonard Sadi Carnot introduz a idéia de transformações cíclicas e prova que a mais eficiente das máquinas termodinâmicas é aquela em que todas as operações são reversíveis. Posto isso, nota-se que a termodinâmica é uma área da física fundamentada nas experimentações daquela época, razão pela qual os resultados obtidos são tão acurados e precisos.

Por volta de 1850, Rudolf Clausius e William Thomson (Kelvin) estabeleceram tanto a primeira lei - da conservação de energia - como a segunda lei da termodinâmica, onde baseiam suas interpretações no livro de Carnot “*Réflexions sur la puissance motrice du feu*”, lançado em 1824 (NASCIMENTO; BRAGA; FABRIS, 2004). Dando continuidade ao trabalho de Carnot de se criar uma máquina de movimento perpétuo, Clausius e Kelvin tentam entender como funciona a transferência dessa energia de um corpo para outro. De forma quase simultânea, afirmam como argumento básico da Segunda Lei o seguinte (apud (FERMI, 1956)):

É impossível uma transformação cujo único resultado final seja transformar calor extraído de uma fonte à mesma temperatura em trabalho. (Lord Kelvin)

É impossível uma transformação cujo único resultado final é transferir calor de um corpo a uma determinada temperatura para um corpo a uma temperatura mais alta. (Clausius)

Com esses depoimentos, é evidente que os fundadores dessa teoria tinham em mente transições entre estados de equilíbrio quando chegaram a tais conclusões.

O termo entropia, como sendo a razão entre calor e temperatura, é introduzido pela primeira vez em 1865 por Clausius, quando este compreende que a segunda lei, na verdade, implica num crescimento contínuo dessa propriedade, definida supostamente para sistemas fora do equilíbrio.

Uma curiosidade é que, cronologicamente, a Segunda Lei foi estruturada antes mesmo que a Primeira fosse formulada (NASCIMENTO; BRAGA; FABRIS, 2004). No entanto, só é possível formular a Segunda Lei tendo algum conhecimento a respeito dos fundamentos da conservação da energia. Dessa forma, entende-se que Carnot não só colaborou para a formulação da Segunda Lei como deve ter, implícita e adequadamente, equacionado a Primeira Lei. Logo, as contribuições pioneiras de Carnot para a fundamentação da Primeira e Segunda Lei fazem dele o verdadeiro “pai da termodinâmica”.

1.2 Entropia como função termodinâmica

Mesmo que a ideia de entropia tenha se estendido a várias interpretações, a simplicidade com que ela foi abordada dentro da teoria termodinâmica é interessante, ao passo que engloba o entendimento de todas as outras quantidades termodinâmicas.

A noção do que é entropia foi desenvolvida a partir da solução do *problema central da termodinâmica*, representado na figura 1. O problema envolve calcular os parâmetros que definem o estado de um sistema: energia, volume, número de mols, etc.

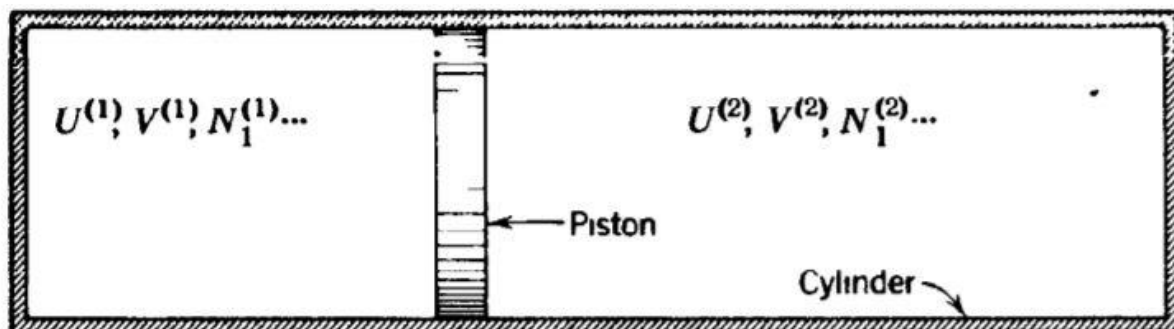


Figura 1 – O problema central da termodinâmica consiste em determinar o estado de equilíbrio de um sistema depois de remover uma membrana interna em um sistema composto. Fonte: (CALLEN, 1998)

Vale informar que esse problema central, cujos resultados compreendem toda a formulação da teoria termodinâmica, cumpre apenas uma parcela dos casos possíveis termodinamicamente. É um problema no qual, estritamente falando, não se pode usar o método tradicional da termodinâmica, que é assumir estados intermediários próximos ao equilíbrio.

O exemplo consiste em dois sistemas simples, contidos em um cilindro fechado, separados um do outro por um pistão fixo interno. Tanto a parede do cilindro quanto o pistão são rígidos, impermeáveis à matéria e adiabáticos (não permitem troca de calor). Ao se liberar o pistão, observa-se um deslocamento em busca de uma nova posição de equilíbrio. Se o revestimento adiabático for removido do pistão fixo, de modo que o calor possa fluir entra os dois sistemas, verifica-se uma nova redistribuição não só de matéria como de energia entre os dois sistemas.

O sistema composto é denominado fechado se for cercado por uma parede restritiva em relação à energia, volume e número total de mols de cada substância desse sistema. Se um sistema fechado está em equilíbrio com relação às restrições internas, e se algumas dessas restrições são removidas, certos processos anteriormente não permitidos tornam-se permitidos. A solução do problema é dada em termos dos parâmetros U, V e N após a redistribuição de matéria, ou seja, quando o sistema atingir o chamado *equilíbrio termodinâmico*.

Segundo Callen (1998), toda a termodinâmica pode ser formulada a partir da definição da função entropia, e do estabelecimento de 3 postulados. Infere-se, então, a existência de uma relação fundamental que é conhecida justamente por estabelecer a entropia como uma função dos parâmetros extensivos, cujos valores serão tais que maximizem a função entropia. Nesse caso, em termos da energia U , o volume V e o número de mols N_i :

$$S = S(U, V, N_1, \dots, N_r) \quad (1.1)$$

Se a relação fundamental de um sistema particular é estabelecida, então toda a informação termodinâmica concebível sobre o sistema pode ser verificável a partir da análise desse problema. Inclusive, todas as quantidades termodinâmicas de todos os subsistemas podem ser encontradas a partir dessa relação.

- ★ Postulado I: *Existe uma função de parâmetros extensivos (chamada entropia S) de um sistema composto, definida para todos os estados de equilíbrio e com a seguinte propriedade: os valores assumidos pelos parâmetros extensivos na ausência de uma restrição interna são aqueles que maximizam a entropia sobre a variedade de estados de equilíbrio restritos.*

★ Postulado II: *A entropia de um sistema composto é aditiva sobre todos os subsistemas constituintes. A entropia é contínua, diferenciável e é uma função monotonicamente crescente de energia.*

Entende-se, então, que a entropia total S do sistema é dada pela soma das entropias $S^{(\alpha)}$, cuja função é definida em termos dos parâmetros extensivos de cada subsistema isolado, $S^{(\alpha)} = S^{(\alpha)}(U^{(\alpha)}, V^{(\alpha)}, N_1^{(\alpha)}, \dots, N_r^{(\alpha)})$:

$$S = \sum_{\alpha} S^{(\alpha)}. \quad (1.2)$$

A propriedade aditiva aplicada separadamente a cada subsistema requer que *a entropia de um sistema simples seja uma função homogênea de primeira ordem dos parâmetros extensivos* (CALLEN, 1998). Para uma constante arbitrária λ , isso implica na propriedade extensiva:

$$S(\lambda U, \lambda V, \lambda N_1, \dots, \lambda N_r) = \lambda S(U, V, N_1, \dots, N_r). \quad (1.3)$$

Subentende-se ainda, pela propriedade da monotonicidade, que a derivada parcial da entropia é uma quantidade positiva,

$$\left(\frac{\partial S}{\partial U} \right)_{V, N_1, \dots, N_r} > 0. \quad (1.4)$$

A função pode ser invertida e a energia se torna, então, uma função, contínua e diferenciável, escrita em termos de S, V, N_1, \dots, N_r . O último postulado que define a entropia é, na verdade, uma extensão do postulado de Nernst da Terceira Lei da Termodinâmica proposto por Plank, e diz assim:

★ Postulado III: *A entropia de qualquer sistema vai a zero em um estado no qual a derivada da energia é zero.*

$$\left(\frac{\partial U}{\partial S} \right)_{V, N_1, \dots, N_r} = 0 \quad (1.5)$$

Assim, escrevendo a equação de energia na sua forma diferencial, é possível obter os parâmetros intensivos que descrevem o sistema

$$dU = \left(\frac{\partial U}{\partial S} \right)_{V, N_1, \dots, N_r} dS + \left(\frac{\partial U}{\partial V} \right)_{S, N_1, \dots, N_r} dV + \sum_{j=1}^r \left(\frac{\partial U}{\partial N_j} \right)_{S, V, \dots, N_r} dN_j, \quad (1.6)$$

$$dU = T dS - P dV + \mu_1 dN_1 + \dots + \mu_r dN_r. \quad (1.7)$$

onde μ representa o potencial eletroquímico, T é a temperatura e P é a pressão. Portanto, infere-se do Postulado III que a entropia de um sistema é nula para um estado em que a temperatura absoluta T também o seja.

1.3 Irreversibilidade e a seta do tempo

Existe uma relação profunda entre a entropia e o tempo. Embora não se saiba definir ou explicar exatamente “o que é” o tempo, sabe-se que a entropia é a única entidade capaz de apontar a direção em que o tempo corre.

Carnot foi o primeiro a relacionar processos cíclicos à reversibilidade e sua principal motivação era construir uma máquina térmica de motor perpétuo, ou seja, uma máquina que conseguisse transformar toda a energia térmica (calor) em trabalho. Desde então, vários estudos sobre o que é entropia surgem com objetivo de explicar processos reversíveis e irreversíveis, a fim de entender o que exatamente essa quantidade é capaz de medir.

A questão da irreversibilidade é um importante problema ainda irresoluto na mecânica estatística. Introduziremos, então, uma noção de processos reversíveis e irreversíveis do ponto de vista da termodinâmica. Uma lei tão precisa como a Segunda Lei possui uma lógica fundamental que independe de informação sobre os átomos que compõem o sistema e pode ser entendida sem saber a fonte da irreversibilidade propriamente dita (LIEB; YNGVASON, 2000).

Com o surgimento da física estatística, alguns físicos rebaixaram as leis da termodinâmica a uma posição inferior, com o argumento de que elas “apenas” retratariam a estatística aplicada à mecânica de um grande número de átomos. Entretanto, H. Lieb e Jakob Yngvason argumentam que “a existência da entropia, e o seu crescimento, podem ser entendidas sem fazer nenhuma referência à mecânica estatística ou máquinas térmicas”. No artigo, os autores tentam reafirmar a segunda lei como um dos pilares da física, resgatando um movimento pouco notado, que iniciou nos anos 1950, com os trabalhos de Peter Landsberg, Hans Buchdahl, Gottfried Falk, Hebert Jung e Robin Giles, nos quais respaldam suas ideias. Foi a partir desses trabalhos que o conceito de “comparação” surge como um dos principais alicerces da segunda lei. E é com base nesse artigo que demonstraremos a relação entre entropia e tempo.

Consideraremos inicialmente que a entropia (S) seja uma função da transição entre estados de equilíbrio e, portanto, só depende dos estados inicial e final do sistema estudado, independentemente do meio pelo qual ocorreu o processo. Tomando como exemplo o problema central da termodinâmica, o conjunto de partículas no recipiente que encontra-se em equilíbrio num primeiro momento, desorganiza-se até se estabelecer em um novo estado de equilíbrio termodinâmico, com novos parâmetros de pressão, volume e temperatura e, conseqüentemente, um novo valor de entropia. Mas ao tomar medidas a respeito dessa variação (ΔS) nota-se que esta deve ser uma quantidade positiva de acordo com a segunda lei e, portanto, implica num processo que é irreversível. Isto é, o sistema não consegue se organizar de tal maneira a estabelecer um estado de equilíbrio com as mesmas condições iniciais.

De acordo com a primeira lei, a quantidade de calor fornecida ao sistema pode ser parcialmente convertida em trabalho e o restante acarreta num acréscimo da energia interna desse sistema: $\Delta Q = \Delta U + \Delta W$. Quando um sistema busca o reequilíbrio sem que ocorra troca de calor entre as partículas e o mundo externo, o processo é considerado adiabático. Ou seja, toda variação de energia interna de um gás durante esse processo deve ser igual ao trabalho realizado pelo ou sobre o gás, podendo causar uma variação de temperatura ou não. Uma compressão ou expansão adiabática podem acontecer tão rapidamente a ponto de não haver tempo para trocas de calor com o meio. Mas se processos adiabáticos ocorrem de forma quase estática, ou seja, que variam suficientemente devagar, a entropia pode permanecer quase inalterada (CALLEN, 1998). Esses processos possuem o crescimento da entropia consideravelmente pequeno a ponto de ser tomada como constante, ($\Delta S \approx 0$), o que implica num processo que pode ocorrer nas duas direções, caracterizando um processo reversível.

Vale ressaltar que processos reversíveis nem sempre são adiabáticos e vice e versa. Um exemplo é a troca de calor entre dois corpos à mesma temperatura. A condição para que o processo seja reversível é que a entropia total do sistema fechado seja constante, sem levar em conta as partes individuais que o compõem. O requisito necessário para configurar um processo como adiabático é que a entropia de todo o subsistema deve permanecer constante durante todo o processo (LANDAU; LIFSHITZ, 1958). No entanto, processos quase estáticos são conceitos idealizados que correspondem a uma sucessão ordenada de estados em equilíbrio. Processos reais sempre envolvem processos intermediários sem equilíbrio que não possuem representação no espaço de configuração termodinâmico, e são definidos por uma sucessão *temporal* de estados em equilíbrio e em desequilíbrio (CALLEN, 1998). Assim, processos reversíveis em sistemas fechados acabam por ser casos limites de processos quase-estáticos, em que o crescimento da entropia é desprezível.

Se transformações quase-estáticas são caracterizadas por uma sucessão ordenada de eventos, existe um *tempo de relaxamento* τ que define a frequência com que cada restrição pode “relaxar” para que um sistema possa atingir um equilíbrio de forma reversível ou não. Tomando como exemplo um gás que se expande em um cilindro com um pistão, no momento em que este se desloca, diminui a densidade do meio e cria-se uma certa rarefação. O tempo de deslocamento para que o sistema atinja o novo estado equilibrado de forma contínua, deve ser correspondente ao tempo necessário para que essa “onda de rarefação” se propague pelo sistema, ou seja,

$$\tau \simeq \frac{V^{\frac{1}{3}}}{c} \quad (1.8)$$

A raiz cúbica de V representa o comprimento do sistema e c a velocidade do som em um meio rarefeito (CALLEN, 1998). Assim, se a expansão adiabática ocorre com um tempo maior que o tempo de relaxamento previsto, essa expansão é tida como reversível e isentrópica ($\Delta S = 0$). No caso da expansão ocorrer com o tempo equivalente ou menor

que o tempo de relaxamento, o processo sofrerá um aumento irreversível da entropia e a expansão, embora adiabática, não será isentropica.

Dito isso, seguimos para uma demonstração retirada do artigo de Lieb e Yngvason, no qual empenham-se principalmente em estabelecer entropia para estados equilibrados. Os autores assumem que já se entende os significados de “sistema termodinâmico” e “estado de equilíbrio”. Também definem brevemente alguns conceitos básicos como “sistema simples” e “sistema composto” e seus estados, que são denotados por letras maiúsculas como X, X', Y, \dots . Além disso estabelecem uma noção central de acessibilidade adiabática, em que *estados finais de um sistemas macroscópico podem ser alcançados a partir de um determinado estado inicial sem deixar uma marca no resto do universo* (LIEB; YNGVASON, 2000). Isto é, um estado Y é dito adiabaticamente acessível a partir do estado X se o único efeito líquido da ação, além da mudança de estado do sistema, é a produção de um certo tipo de trabalho, e escreve-se $X \prec Y$.

Os autores descrevem essas transições adiabáticas no exemplo da figura 2, onde indicam três estados possíveis de se acessar a partir de um estado inicial de um sistema composto por dois copos separados que se configuram inicialmente: um com 50g de whisky a 20°C e o outro com 20g de gelo a 0°C. Os estados (Y_1, Y_1') , (Y_2, Y_2') e Y_3 são todos adiabaticamente acessíveis de (X, X') . O primeiro estado (Y_1, Y_1') é alcançado através de um agitador (que não está sendo mostrado) atuando em cada copo. O segundo, (Y_2, Y_2') , é obtido quando os dois copos (subsistemas) entram em contato térmico. O estado Y_3 só é possível quando se coloca as duas substâncias em um único copo, onde ocorre um processo de mistura que altera o sistema. O estado original (X, X') não é adiabaticamente acessível a partir dos outros estados, (Y_1, Y_1') , (Y_2, Y_2') e Y_3 .

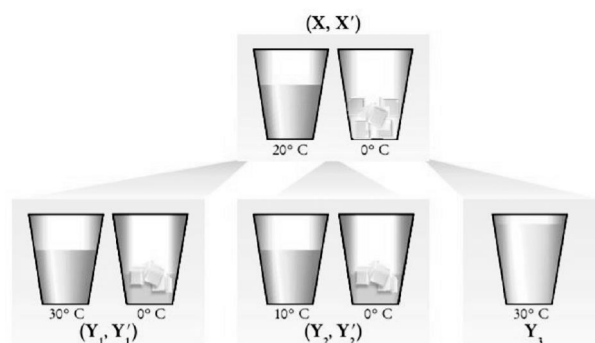


Figura 2 – Mudanças de estado adiabáticas.

A existência e significado da entropia então, não se baseiam apenas em uma fórmula que envolva probabilidades de microestados como $S = -\sum_i p_i \ln p_i$, como é definida na mecânica estatística. Na verdade, o que se nota geralmente é que a entropia é a única grandeza física capaz de diferenciar o passado do futuro e, por essa razão, é conhecida como “seta do tempo”, uma orientação da flecha que consegue definir uma direção para a ocorrência de eventos. Sendo assim, a possibilidade de definir a entropia inteiramente em termos de uma relação \prec foi vislumbrada pela primeira vez por Robin Giles¹, apud

¹ A definição dada por Giles é diferente da explicitada no artigo, embora seja bem similar na sua essência.

(LIEB; YNGVASON, 2000), e de forma simples: $X \prec Y$ expressa que um estado X precede um estado Y , ou que um estado Y advém de um estado anterior X . Isso significa que, analisando dois estados X e Y de um sistema, a entropia é obtida através da correlação:

$$X \prec Y \quad \text{se e somente se} \quad S(X) \leq S(Y). \quad (1.9)$$

Além disso, a função de entropia pode ser escolhida levando-se em consideração que se X e X' são estados de dois sistemas (idênticos ou diferentes), então a entropia do sistema composto por um par de estados é dado pela soma das entropias individuais:

$$S(X, X') = S(X) + S(X'). \quad (1.10)$$

A propriedade aditiva é, no entanto, uma afirmação não trivial. Como já foi colocado anteriormente, em sistemas termodinâmicos isolados a entropia só pode aumentar ou permanecer constante, nunca diminuir.

Em sistemas compostos, como o do exemplo anterior, todos os estados (Y, Y') nos quais $X \prec Y$ e $X' \prec Y'$, são adiabaticamente acessíveis a partir do estado inicial (X, X') . Como foi visualizado no primeiro caso do exemplo: ao acrescentar um mixer nos dois copos o whiskey esquentava e o gelo derrete em processos individuais. Todavia, é importante notar que estados (Y, Y') podem ser adiabaticamente acessíveis a partir de estados (X, X') sem que Y seja adiabaticamente acessível a partir de X exatamente. Isto é, é possível que a entropia de um sistema diminua em troca do aumento da entropia do meio, de maneira que a entropia líquida total seja crescente,

$$S(X) + S(X') \leq S(Y) + S(Y'). \quad (1.11)$$

Ao colocar os dois copos em contato e separá-los novamente, como foi o segundo caso do exemplo, observa-se um processo adiabático do ponto de vista do sistema composto, mas o resfriamento do whiskey não representa um processo adiabático para o sistema sozinho. A desigualdade acima reflete justamente essa possibilidade de que transições adiabáticas são possíveis mesmo que ocorra de $S(X) > S(Y)$ e, portanto, “é suficiente conhecer a entropia de cada parte do sistema composto para decidir quais transições devido às interações entre as partes são possíveis” (LIEB; YNGVASON, 2000).

Além da aditividade, existe outra propriedade semelhante chamada de *extensividade*, a qual diz que a entropia de uma massa arbitrária de uma substância pode ser escrita em termos de uma referência padrão, sendo multiplicada por uma constante quando os parâmetros extensivos também o forem:

$$S(\lambda X) = \lambda S(X). \quad (1.12)$$

Percebe-se que essas propriedades são equivalentes às já abordadas pelo postulado I e II no tópico anterior. No entanto, a definição de entropia como foi proposto pelo artigo não

depende de conceitos termodinâmicos ou estatísticos como calor, temperatura, energia, etc. Na verdade, sua derivação consiste basicamente na comparação entre listas de pares $X \prec Y$, que envolvam algumas propriedades que permitirão que essa relação seja descrita pela própria entropia.

Além de algumas condições “óbvias” como a da *reflexividade* e a *transistividade*, as demais propriedades resultantes da relação dada pela estrutura \prec funcionam bem com respeito à composição e escala dos estados, são elas:

- ▷ Reflexividade: $X \prec X$, para todo X ;
- ▷ Transitividade: $X \prec Y$ e $Y \prec Z$ implicam em $X \prec Z$;
- ▷ Processos adiabáticos são consistentes com a composição de estados, isso significa que a relação $X \prec Y$ e $Z \prec W$ implica em $(X, Z) \prec (Y, W)$;
- ▷ A acessibilidade adiabática também não se altera quando a quantidade analisada é multiplicada por um fator λ , ou seja, se $X \prec Y$, então $\lambda X \prec \lambda Y$;
- ▷ Os sistemas podem ser subdivididos adiabaticamente em duas partes: se $0 < \lambda < 1$ e $X \prec ([1 - \lambda]X, \lambda X)$, logo a recombinação das partes é também adiabática, $([1 - \lambda]X, \lambda X) \prec X$;
- ▷ A acessibilidade adiabática envolvendo pequenas perturbações é estável e, por isso, se $(X, \varepsilon Z) \prec (Y, \varepsilon W)$ para um valor pequeno e arbitrário $\varepsilon > 0$, então $X \prec Y$.

Essas propriedades, entretanto, não são requisitos suficientes para definir a entropia desses estados; é necessário estabelecer uma hipótese de comparação que garanta uma longa lista de pares de estados $X \prec Y$ possíveis. Tal hipótese pressupõe que estados de equilíbrio, de um estado simples ou composto, podem ser agrupados em classes. Uma classe é constituída de estados com a mesma massa e composição química, de maneira que se X e Y pertencem à mesma classe, então a relação $X \prec Y$ é válida bem como $Y \prec X$.

Os autores acreditam que a existência e singularidade da entropia sejam uma consequência da hipótese da comparação e das premissas dispostas acima. E, de fato, supondo que um sistema possui três estados distintos, X_0 , X e X_1 , e que λ seja um fator de escala entre 0 e 1 então, pela hipótese da comparação, tanto $X \prec ([1 - \lambda]X_0, \lambda X_1)$ é possível como seu reverso $([1 - \lambda]X_0, \lambda X_1) \prec X$. Se ambas as alternativas se sustentam, as propriedades de entropia resultam na relação:

$$S(X) = (1 - \lambda)S(X_0) + \lambda S(X_1). \quad (1.13)$$

Se X_0 e X_1 são estados de referência, então a entropia de X só dependerá do parâmetro λ . Se $S(X_0) \neq S(X_1)$, essa igualdade valerá para pelo menos um valor de λ . Logo, a partir das propriedades da estrutura \prec , conclui-se que existirá sempre um $0 \leq \lambda \leq 1$ que fornecerá uma relação $X_0 \prec X \prec X_1$.

1.4 Entropia do gás ideal

Antes mesmo que as leis da termodinâmica fossem postuladas, a curiosidade em compreender o comportamento dos gases já existia. Em 1738 Daniel Bernoulli levanta a ideia de que os gases consistiam em moléculas em movimento. Mas foi somente depois de 100 anos de adormecimento que a discussão a respeito do movimento dos gases foi retomada por Clausius (WOLFRAM, 2002), embora outros conceitos tenham sido desenvolvidos nesse meio tempo. Até meados do século XIX havia uma carência muito grande por trabalhos cientificamente bem fundamentados. Foi então que, em 1824, Carnot revolucionou a história da ciência quando publicou seu trabalho com um altíssimo rigor científico (NASCIMENTO; BRAGA; FABRIS, 2004). Na busca por um melhor funcionamento das máquinas a vapor, Carnot questiona a potência do fogo e discute a possibilidade de haver um equivalente mecânico para o calor em seu primeiro e único livro “Réflexions sur la puissance motrice du feu”.

Carnot já estava consciente da conservação da energia quando deixa claro em seu trabalho que “o calor originário da fonte de maior temperatura é transformado em trabalho e o restante do calor é absorvido pela fonte de menor temperatura” (NASCIMENTO; BRAGA; FABRIS, 2004). Ele acreditava ser possível construir uma máquina térmica ideal a partir da premissa de que se o processo pode ser realizado no sentido inverso, usando-se as mesmas quantidades, então, as condições iniciais do sistema são restabelecidas. Em sua obra, Carnot descreve a primeira lei de forma explícita ao estabelecer uma relação $s = e + U$, utilizada para representar a grandeza s como a quantidade de calor necessária para ocorrer uma expansão, e e a potência motriz produzida. Em seguida, Carnot calcula $e = T \log V$ para o trabalho no processo isotérmico, reescrevendo a equação anterior como $s = T \log V + U$.

Percebe-se que há uma grande similaridade entre a relação encontrada por Carnot e a expressão que define a primeira lei: $\partial Q = dU - dW$. Comparando com a equação 1.7, observa-se que se $dW = -P dV$ é o termo que indica o trabalho realizado pelo ou sobre o gás, então, o termo que descreve o comportamento do calor no meio é dado por $T dS$. Ou seja, a noção intuitiva do que é *temperatura* é dada pela própria definição de entropia,

$$dS = \frac{\partial Q}{T}. \quad (1.14)$$

Através da equação 1.6, nota-se que, em ambientes controlados com volume e massa molar constantes, a temperatura é $(\partial U / \partial S)_{V,N} = T$. A partir dessa relação, e das propriedades

da entropia e conservação de energia, é possível verificar o comportamento de um fluxo de calor, que parte de um corpo quente para um corpo mais frio.

Com objetivo de criar uma máquina que tivesse a máxima eficiência possível, Carnot desenvolveu um ciclo perfeito cujo circuito de operação deveria cumprir com algumas características: a máquina deveria ser apta para operar em ciclos sucessivos; a substância de operação deveria conseguir transportar calor de uma fonte quente para uma fonte fria; e as máquinas deveriam ser capazes de operar no sentido inverso (LARANJEIRAS; PORTELA, 2016). O ciclo é composto basicamente por uma expansão e compressão isotérmica e 2 transformações adiabáticas. Verifica-se que Carnot previa não só a primeira, como a segunda lei da termodinâmica.

Embora Carnot não compreendesse muito bem a performance de uma máquina irreversível, ele já notava que a conservação do calórico era condição necessária para que ocorresse a reversibilidade nos processos. Calórico, diferente de calor (no sentido de energia térmica), era considerada uma “energia de transformação” e será associada à ideia de entropia posteriormente com Clausius (CALLENDAR, 1910).

Mas, apesar de todas essas descobertas, foi Clapeyron quem interpretou o ciclo de Carnot e traduziu os processos em equações matemáticas. A partir da análise do ciclo, Clapeyron calculou a variação da pressão com a temperatura numa substância que se encontrava em equilíbrio entre duas fases. Em seguida, derivou sua importante equação, na qual estabelece uma relação entre as variáveis de um sistema, conhecida como equação de estado dos gases perfeitos:

$$PV = NRT. \quad (1.15)$$

Aqui, pressão e volume se relacionam por $PV^\gamma = cte$, onde γ representa o coeficiente politrópico de uma transformação termodinâmica, caracterizando o processo como isobárico ($\gamma = 0$), isotérmico ($\gamma = 1$), adiabático (e isentrópico quando $\gamma = C_p/C_v$) e isocórico ($\gamma \rightarrow \infty$). Outra expressão utilizada para indicar a quantidade de energia presente no gás ideal simples é dada pela equação fundamental:

$$U = cNRT, \quad (1.16)$$

onde c é uma constante e N, R e T representam, a massa molar do gás, a constante universal do gás ($R = N_A k_B = 8.3144 J/molK$) e a temperatura, respectivamente. Em particular, o calor específico a volume constante, $c = C_v$, é uma grandeza sem dimensão que dispõe de diferentes valores a depender do gás, sendo $3/2$ para gases monoatômicos, $5/2$ para diatômicos e 3 para moléculas não lineares se desprezarmos as contribuições quânticas vibracionais (LANDAU; LIFSHITZ, 1958).

Todo e qualquer gás pode ter seu comportamento aproximado ao ideal se considerarmos uma densidade de partículas suficientemente baixa, a ponto de as moléculas apresentarem pouca interação entre si (CALLEN, 1998). Seguindo essa lógica, consegui-

mos encontrar a fórmula de entropia para os gases ideais a partir da representação de entropia proposta por Euler

$$S = \left(\frac{1}{T}\right)U + \left(\frac{P}{T}\right)V - \sum_{k=1}^r \left(\frac{\mu_k}{T}\right)N_k, \quad (1.17)$$

e da relação de Gibbs-Duhem

$$d\left(\frac{\mu}{T}\right) = u d\left(\frac{1}{T}\right) + v d\left(\frac{P}{T}\right) \quad (1.18)$$

onde $u = U/N$ e $v = V/N$. O resultado da integral de 1.18 é escrita em termos de parâmetros de referência fixos u_0 e v_0 dado por

$$\frac{\mu}{T} - \left(\frac{\mu}{T}\right)_0 = -cR \ln \frac{u}{u_0} - R \ln \frac{v}{v_0}. \quad (1.19)$$

Substituindo a expressão (1.19) na relação de Euler (1.17), obtemos, finalmente, a entropia do gás ideal escrita na sua forma resumida

$$S = Ns_0 + NR \ln \left[\left(\frac{U}{U_0}\right)^c \left(\frac{V}{V_0}\right) \left(\frac{N}{N_0}\right)^{-(c+1)} \right] \quad (1.20)$$

onde

$$s_0 = (c+1)R - \left(\frac{\mu}{T}\right)_0 \quad (1.21)$$

Uma melhor demonstração da derivação de entropia pode ser encontrada em (CALLEN, 1998),66-70. A generalização para o caso em que o gás ideal tenha multicomponentes, ou seja, uma mistura de dois ou mais gases ideais simples, é caracterizada pela equação fundamental escrita na sua forma paramétrica:

$$S = \sum_j N_j s_{j0} + \left(\sum_j N_j c_j \right) R \ln \frac{T}{T_0} + \sum_j N_j R \ln \left(\frac{V}{V_j v_0} \right). \quad (1.22)$$

2 Entropia estatística

2.1 História e contexto da entropia estatística

Logo que as duas leis da termodinâmica foram postuladas, James Clerk Maxwell (1860) compreendeu a mecânica de uma colisão molecular individual e, a partir dela, derivou a distribuição de velocidades de moléculas de um gás (WOLFRAM, 2002). Nos anos seguintes, a teoria cinética dos gases se desenvolveu bastante e várias propriedades macroscópicas sobre o comportamento de gases em equilíbrio foram computadas.

Em 1872, Ludwig Boltzmann desenvolveu uma equação que descreve, em função do tempo, o comportamento de um gás detalhadamente, estando ele em equilíbrio ou não. Em 1860, assim que Clausius introduziu a ideia de entropia, Boltzmann demonstrou que sua equação resultaria em um teorema que confirmava a propriedade de maximização da entropia. Em um primeiro momento, a declaração de Boltzmann pareceu ter provado a Segunda Lei. Mas considerando que as colisões moleculares eram entendidas como reversíveis naquela época, sua equação também deveria ocorrer de forma inversa, o que acabaria invalidando a Segunda Lei. Muito depois, descobriu-se que a equação original de Boltzmann assumia implicitamente a interdependência das moléculas antes de cada colisão, mas não depois do choque, levantando a hipótese de assimetria temporal (WOLFRAM, 2002). Ainda, preocupado em justificar a questão da reversibilidade, Boltzmann percebeu, por volta de 1876, que havia muito mais estados alcatórios do que ordenados, apresentando pela primeira vez a ideia de ergodicidade. Essa descoberta levou Boltzmann a deduzir que a entropia é proporcional ao logaritmo da quantidade de estados possíveis de um sistema.

Um novo campo de estudo emergia no final do século XIX. Maxwell e Kelvin, no começo dos anos 1870, tentam derivar a Segunda Lei a partir da física desse novo mundo microscópico, sem muito sucesso. Segundo Wolfram, poderia ser uma limitação daquela época lidar com números grandes de moléculas. A mecânica estatística de sistemas de partículas foi expressa de forma mais generalizada por Willard Gibbs, no início de 1900. Gibbs tentou explicar qualitativamente que a entropia de um sistema deveria crescer caso fosse medida de forma "granulada", indicando que os estados próximos não fossem distinguidos. A partir do início do século XX os estudos a respeito de conceitos fundamentais da termodinâmica entram em decadência e dão espaço para uma teoria que ganhou maior notoriedade, a mecânica quântica. O estudo da entropia em um sistema quântico começa com Von Neumann em seu trabalho de 1932, o qual faz uma extensão do conceito da entropia de Gibbs no campo da mecânica quântica.

2.2 O espaço de configurações

Com objetivo de conectar as propriedades macroscópicas a sistemas de natureza microscópica, a física estatística é aplicada à mecânica de grandes números de átomos e todas as relações envolvendo grandezas termodinâmicas podem ser, de fato, derivadas da mecânica estatística (LIEB; YNGVASON, 2000). Nesse contexto, a entropia é interpretada como uma função, e também propriedade de um corpo, que está diretamente relacionada às diferentes configurações de estados disponíveis que os subsistemas podem assumir. Assim, podemos definir o espaço de configuração tanto no âmbito quântico quanto no clássico.

Para fins de simplificação, o presente capítulo e o próximo seguem, em geral, parte da formulação de Landau e Lifshitz, que podem ser encontradas em (LANDAU; LIFSHITZ, 1958), seções 7 e 8.

Admite-se, então, um sistema macroscópico fechado e em equilíbrio estatístico, composto por várias partes macroscópicas, onde $\Gamma(E)$ corresponde ao número de estados quânticos com energia menor ou igual a E . Considere ainda que $w_n = w(E_n)$ seja uma função da energia de um subsistema. A função w_n permite calcular a distribuição de probabilidades de valores de energia que um subsistema pode assumir. Ou seja, para obter a probabilidade $W(E) dE$ de um subsistema com um valor de energia entre E e $E + dE$, deve-se multiplicar a função w_n pelo número de estado quânticos que possuem energia dentro desse intervalo ($\frac{d\Gamma(E)}{dE} dE$). Logo, a distribuição de probabilidade de energia é

$$W(E) = \frac{d\Gamma}{dE} w_n(E), \quad (2.1)$$

que, pela condição de normalização,

$$\int W(E) dE = 1. \quad (2.2)$$

Essa função possui um valor máximo bem marcante no valor médio de energia $E = \bar{E}$, de maneira que pode-se definir um intervalo ΔE como a largura de um retângulo cuja altura é o valor máximo de $W = W(E)$, e a área retangular abaixo da curva seja igual a 1,

$$W(\bar{E}) \Delta E = 1. \quad (2.3)$$

Aplicando a expressão 2.1 na condição de normalização 2.2, a função $w(E)$ pode ser escrita em termos de

$$w(\bar{E}) \Delta \Gamma = 1, \quad (2.4)$$

onde

$$\Delta \Gamma = \frac{d\Gamma(\bar{E})}{dE} \Delta E \quad (2.5)$$

corresponde ao número de microestados existentes dentro desse intervalo de energia ΔE , que representa a flutuação de energia média de um subsistema.

Na estatística clássica, $w(E)$ é substituído pela função de distribuição clássica ρ , enquanto $\Delta\Gamma$ é equivalente ao volume $\Delta p\Delta q$, que representa o tamanho da região do espaço de fase em que o subsistema pode ser encontrado e, portanto:

$$\rho(\bar{E})\Delta p\Delta q = 1. \quad (2.6)$$

Todas essas definições podem ser aplicadas na estatística clássica apenas fazendo as substituições necessárias. No caso quase clássico, uma correspondência entre o volume de uma região no espaço de fase e os estados quânticos ali presentes é estabelecida pela relação $\Delta\Gamma = \Delta p\Delta q/(2\pi\hbar)^s$, em que uma célula de volume $(2\pi\hbar)^s$ corresponde a cada estado quântico no espaço de fase, e s ao número de graus de liberdade desse sistema. O volume $\Delta p\Delta q$, assim como $\Delta\Gamma$, representa o número de estados acessíveis à partícula, indicando “o tamanho da região do espaço de fase em que um subsistema pode quase sempre ser encontrado” (LANDAU; LIFSHITZ, 1958). Sendo assim, a entropia de uma partícula pode ser definida como o logaritmo do peso estatístico ($\Delta\Gamma$) do estado macroscópico do subsistema

$$S = \log \Delta\Gamma. \quad (2.7)$$

A expressão correspondente no caso da estatística clássica pode ser escrita na forma:

$$S = \log \frac{\Delta p\Delta q}{(2\pi\hbar)^s}. \quad (2.8)$$

A entropia também pode ser estabelecida em termos da função distribuição de energia w_n . Segundo Boltzmann, a distribuição de probabilidade $w(E)$ representa a fração de tempo gasto pelo sistema em torno de uma dada energia E (RIOUL, 2018). Levando em conta que o peso estatístico é estipulado em torno de um valor médio de energia, a entropia $S = \log \Delta\Gamma = -\log w(\bar{E})$ pode ser escrita como o valor médio

$$S = -\langle \log w(E_n) \rangle \quad (2.9)$$

que, por definição, equivale a

$$S = -\sum_n w_n \log w_n. \quad (2.10)$$

A expressão acima pode ser escrita utilizando a notação matricial proposta por Von Neumann, em que \hat{w} corresponde ao operador densidade:

$$S = -Tr(\hat{w} \log \hat{w}). \quad (2.11)$$

Aqui, $\log \hat{w}$ é definido como um operador cujos autovalores são iguais aos logaritmos dos autovalores do operador \hat{w} , e cujas autofunções são iguais às de \hat{w} .

Fazendo as devidas alterações em 2.9, a definição de entropia na sua representação clássica é dada por:

$$\begin{aligned} S &= -\langle \log[(2\pi\hbar)^s \rho] \rangle \\ &= -\int \rho \log[(2\pi\hbar)^s \rho] dp dq. \end{aligned} \quad (2.12)$$

Levando em consideração que o sistema fechado é subdividido em partes, cada subsistema terá peso estatístico. Se cada subsistema encontra-se em um estado quântico $\Delta\Gamma_a$, o número de diferentes estados presentes no sistema como um todo será dado pelo produto

$$\Delta\Gamma = \prod_a \Delta\Gamma_a. \quad (2.13)$$

A propriedade aditiva da entropia é introduzida pela função logarítmica, e consiste em somar as entropias individuais das partes para obter a entropia total do sistema. Assim como na termodinâmica, escreve-se

$$S = \sum_a S_a. \quad (2.14)$$

Apesar disso, um sistema fechado que se encontra em completo equilíbrio termodinâmico, e cuja energia total do sistema é denotada por E_0 , pode ter sua entropia definida sem se preocupar em subdividi-lo em partes. Isso pode ser feito apenas considerando que o próprio sistema seja uma pequena parte de um sistema maior (fictício). O termostato é um exemplo de meio no qual o sistema pode estar inserido que, se mantido em completo equilíbrio, garantirá que a energia média do sistema (que agora é um subsistema não fechado) permaneça igual à energia atual dele, E_0 . Nesse caso, a função distribuição do sistema, assim como para qualquer subsistema, pode ser obtida da mesma forma. O peso estatístico e a entropia também podem ser obtidos da mesma forma como foi demonstrado pelas equações 2.4 até 2.12.

As definições também podem ser generalizadas para sistemas em estados macroscópicos arbitrários fora do equilíbrio. Supõe-se que um sistema esteja em equilíbrio parcial, e considera-se um intervalo de tempo Δt menor que o tempo de relaxamento τ (1.8) necessário para se atingir o completo equilíbrio. Em seguida, o sistema é subdividido em partes tão pequenas, a ponto de haver um tempo de relaxamento τ' menor que o intervalo Δt (tendo em vista que o tempo de relaxamento decresce com o tamanho do sistema). Pode-se considerar que, durante o intervalo de tempo Δt , esses pequenos subsistemas encontram-se em um estado de equilíbrio particular. Logo, as definições de peso estatístico $\Delta\Gamma_a$ e entropia S_a podem ser aplicadas a cada subsistema, de modo que satisfaçam as condições em 2.13 e 2.14, para o peso estatístico total do sistema ($\Delta\Gamma$) e sua respectiva entropia (S).

No entanto, essa generalização para sistemas fora do equilíbrio, diferente do exemplo do termostato, não permite calcular a entropia sem fragmentá-los. Essa é uma ideia

ambígua, pois ao dividir os subsistemas em partes ainda menores, a entropia não se altera em nada já que esses subsistemas se encontrarão em certo “equilíbrio” interno. Ou seja, se esse tempo de relaxamento for menor que um tempo observável Δt , o conceito de entropia se torna irrelevante; por isso, se diz que a entropia não possui valor instantâneo.

Em razão disso, utilizam-se as definições de uma distribuição microcanônica para encontrar as propriedades estatísticas de um sistema fechado:

$$dw = \alpha \times \delta(E - E_0) \cdot \prod_a d\Gamma_a, \quad (2.15)$$

em que α corresponde a uma constante arbitrária. Considerando que $d\Gamma_a$ representa o diferencial da função $\Gamma_a(E_a)$ e que ele corresponde ao número de estados quânticos de um subsistema com energia $\leq E_a$, a equação acima pode ser reescrita como:

$$dw = \alpha \times \delta(E - E_0) \cdot \prod_a (d\Gamma_a/dE_a) dE_a. \quad (2.16)$$

“A função $\Delta\Gamma_a$ é, por definição, uma função do valor médio de energia \bar{E}_a de um subsistema; o mesmo se aplica a $S_a = S_a(\bar{E}_a)$ ” (LANDAU; LIFSHITZ, 1958). No entanto, assume-se que $\Delta\Gamma_a$ e S_a sejam funções da energia atual E_a do sistema. A derivada $d\Gamma_a(E_a)/dE_a$ é substituída por $\Delta\Gamma_a/\Delta E_a$, onde $\Delta\Gamma_a$ é uma função de E_a e ΔE_a o intervalo de energia correspondente. Substituindo $\Delta\Gamma_a$ por $e^{S_a(E_a)}$ na equação 2.16, se obtém a relação:

$$dw = \alpha \times \delta(E - E_0) e^S \prod_a dE_a/\Delta E_a, \quad (2.17)$$

onde $S = \sum_a S_a(E_a)$ corresponde à entropia total do sistema fechado, obtida a partir das funções de energia exata de cada subsistema. O termo $\prod_a 1/\Delta E_a$ é ínfimo e pode ser desconsiderado nas contas, permanecendo uma relação ainda com bastante acurácia:

$$dw = \alpha \times \delta(E - E_0) e^S \prod_a dE_a. \quad (2.18)$$

A função distribuição, expressa como produto de todos os diferenciais dE_a , sugere a probabilidade de que todos os subsistemas estejam com energias no intervalo entre E_a e $E_a + dE_a$. O fator $\delta(E - E_0)$ garante que um sistema fechado possui a energia total equivalente à soma das energias individuais dos subsistemas, $\sum E_a = E_0$.

Compreende-se por valores mais prováveis de energia os seus respectivos valores médios $E_a = \bar{E}_a$. Isso significa que a função da entropia, dependente das energias dos subsistemas $S(E_1, E_2, \dots)$ (com a condição $\sum E_a = E_0$), tem seu valor máximo possível quando $E_a = \bar{E}_a$. Se esses valores \bar{E}_a correspondem às energias dos subsistemas quando estes atingem o equilíbrio, a conclusão a que se chega é de que: a entropia de um sistema fechado em um estado de completo equilíbrio estatístico está em seu maior valor possível.

Como já dito anteriormente, o peso estatístico $\Delta\Gamma = e^{S(E)}$ representa a quantidade de estados acessíveis, ou melhor, o número de níveis de energia no intervalo ΔE ,

que descreve de certa forma a largura da distribuição de probabilidade de energias. Ao dividir ΔE por $\Delta \Gamma$, o que se observa é uma separação média entre os níveis adjacentes de energia dentro do espectro de energia do sistema considerado. Assim, a distância entre esses níveis de energia é dada por:

$$D(E) = \Delta E \cdot e^{-S(E)}. \quad (2.19)$$

Nesse sentido, a entropia $S(E)$ é encarregada pela densidade de níveis no espectro de energia de um sistema macroscópico. Portanto, pelo fato de ser aditiva, a separação média entre os níveis de energia de um corpo macroscópico decresce exponencialmente com o crescimento do tamanho do corpo, isto é, com o aumento do número de partículas nele.

2.3 Entropia como uma medida da desordem?

De forma geral, a entropia é entendida como um grau de desordem atribuído a um sistema, uma tendência natural ao caos ou “desorganização” desses sistemas. Apesar de não totalmente errada, essa costuma ser uma ideia rasa sobre a grandeza dado que sua compreensão é ainda mais complexa e abstrata que isso. Além disso, existe toda uma análise filosófica a respeito de suas aplicabilidades que também pode ser feita.

Como foi dito na seção anterior, a discussão a seguir foi baseada nas interpretações de Landau e Lifshitz, apresentadas na seção 8 do primeiro capítulo do livro (LANDAU; LIFSHITZ, 1958).

De acordo com a segunda lei da termodinâmica (na formulação de Clausius de 1856), um sistema fechado que ainda não atingiu seu valor máximo de entropia em algum instante, terá sua entropia aumentada nos momentos posteriores, ou então permanecerá constante, mas de modo algum irá diminuir. Na verdade, pela interpretação da física estatística estabelecida por Boltzmann em 1870, a probabilidade de se ter a entropia aumentada é tão grande em comparação com seu decréscimo, que ignora-se este último fato - que aconteceria devido a flutuações - alegando que este fenômeno não pode ser observado na natureza.

Naturalmente, se um sistema fechado não estiver em um estado de equilíbrio estatístico, seu estado macroscópico irá variar ao longo do tempo até atingir um estado de completo equilíbrio. Logo, se os estados macroscópicos desse sistema são descritos pela distribuição de energia nos subsistemas, a sequência de estados sucessivos pela qual o sistema deve transitar corresponderá às distribuições de energia mais e mais prováveis. Isso quer dizer que os sistemas fechados fora do equilíbrio devem transitar, continuamente, estados de menor entropia para aqueles de maior entropia, até que se atinja, finalmente, seu valor máximo, caracterizando o equilíbrio estatístico completo.

Mas, se tentamos aplicar a física estatística no universo como um todo, tratando-o como um único sistema fechado, encontraremos uma grande discordância entre teoria e experiência. Com a dimensão e idade do universo, é de se esperar que qualquer subespaço dele tenha um tempo de relaxamento finito no qual configura um estado de equilíbrio local. Apesar disso, “as experiências sugerem que as propriedades da natureza não se assemelham às de um sistema em equilíbrio”; e resultados astronômicos confirmam o mesmo ao longo de todo universo observável.

A resposta para essa contradição encontra-se na teoria geral da relatividade. A razão é que, quando grandes regiões do universo são consideradas, considera-se também a presença de campos gravitacionais. Quando as propriedades estatísticas dos corpos estão sendo discutidas, as propriedades métricas do espaço-tempo são vistas como “condições externas” às quais os corpos estão sujeitos. No entanto, o argumento de que um sistema fechado atinge o estado de equilíbrio depois de um longo período se aplica apenas para casos nos quais as condições externas são estáveis. A expansão cosmológica do universo sugere que sua métrica depende essencialmente do tempo e, nesse caso, essas condições externas não são estáveis. Os próprios campos gravitacionais são responsáveis pela mudança dessa métrica. Por esse motivo, o universo não pode ser considerado um sistema fechado, mas um sistema de campo gravitacional variável. Em consequência disso, a lei do aumento da entropia não prova necessariamente a existência de um equilíbrio estatístico.

Outra questão importante é que a lei de aumento da entropia não pode ser derivada com base na mecânica clássica devido à invariância do tempo proposta por ela. Uma vez que a mecânica clássica é simétrica com respeito às duas direções do tempo e, portanto, as equações se comportam invariavelmente ao se substituir t por $-t$, a estatística baseada na mecânica clássica preserva essa simetria. Ou seja, todo evento que ocorre em uma dada direção do tempo deve, teoricamente, acontecer de forma equivalente ao “voltar o tempo”. Contudo, se um processo envolvendo o crescimento da entropia de um sistema fechado é possível, então seu processo inverso também o será, o que implicaria em um decrescimento da entropia.

Mas a lei que se refere ao crescimento da entropia apenas admite que esse seja o resultado mais provável de um estado macroscópico, não contradizendo de fato essa simetria. Isto posto, é necessário definir uma direção para o tempo na qual a entropia cresça monotonicamente, o que equivale a estabelecer uma “seta do tempo”.

Nesse sentido, o problema se reduz a provar que a lei de aumento da entropia, definida pela termodinâmica, é equivalente à definição da mecânica quântica, fundamentada na então reconhecida equação de Schrödinger. Essa última permanece simétrica sob inversão temporal, provida pelo fato de que a função de onda ψ é substituída por seu conjugado complexo ψ^* . Isso significa que se, num dado instante $t = t_1$, a função de onda $\psi = \psi(t_1)$ é dada, e no segundo instante a equação de Schrödinger retorna $\psi(t_2)$, então

a situação reversa deve ser admissível; ou seja, $\psi = \psi^*(t_2)$ num momento inicial t_1 e $\psi = \psi^*(t_1)$ em t_2 .

Apesar da simetria, “a mecânica quântica envolve uma importante não-equivalência das duas direções do tempo” (LANDAU; LIFSHITZ, 1958). Isso ocorre devido às interações de objetos quânticos com sistemas que obedecem as leis da mecânica clássica com uma certa precisão. Significa dizer que, se duas interações quânticas A e B ocorrem sucessivamente e a probabilidade de qualquer resultado oriundo do processo B depende do resultado obtido de A , então o procedimento A deve ocorrer antes de B para que esse resultado seja válido.

Assim, não existe equivalência física nas duas direções do tempo na mecânica quântica e, teoricamente, a lei de aumento de entropia deve ser uma expressão macroscópica. Deve existir, nesse caso, uma desigualdade que envolva a constante de Plank, \hbar , que garanta a validade da lei de aumento da entropia e satisfaça o mundo real ao mesmo tempo. Contudo, até o presente momento nenhuma dessas relações foi estabelecida.

A questão sobre os fundamentos físicos da segunda lei permanece em aberto até hoje. Segundo os autores, a origem pode ser cosmológica ou relacionada ao problema geral das condições iniciais da cosmologia; a violação da simetria sob inversão do tempo pode estar relacionada a algumas interações fracas entre partículas elementares, além de outras hipóteses.

Enfim, a formulação geral de crescimento da entropia considera que, em todos os sistemas fechados que ocorrem na natureza, a entropia nunca decresce; ou ela cresce ou, no mínimo, permanece constante. Assim, de acordo com essas duas possibilidades, os processos que ocorrem em sistemas macroscópicos podem ser classificados como *reversíveis* ou *irreversíveis*. O caso reversível corresponde à constância da entropia total do sistema, independente do que acontece com as partes individuais que o compõem. Processos isentrópicos permitem que os eventos ocorram na direção reversa, e são passíveis de se encontrar na natureza apenas com um certo grau de aproximação (processos de quase equilíbrio). A maioria dos processos são, na verdade, irreversíveis e estão associados ao aumento monotônico da entropia. O processo reverso e na direção oposta do tempo, ocasionaria num decrescimento da entropia total do sistema que, apesar de improvável, não é impossível de acontecer, apenas levaria um tempo equivalente à idade do universo para que se pudesse observar tal evento.

2.4 Entropia de Gibbs

Como dito na seção 2.1, a mecânica estatística aplicada em sistemas de partículas foi estabelecida de forma mais geral por Willard Gibbs no começo do século XX. Ele foi o responsável por introduzir a ideia de *ensemble* como uma coleção de vários siste-

mas igualmente preparados, cada estado com sua respectiva probabilidade (WOLFRAM, 2002). Gibbs argumentava que, se ao longo da evolução temporal de um único ponto de um sistema, dinâmico ou estocástico, ele transitar por todos os estados do conjunto, ele deve se comportar de maneira “típica do conjunto”. Ou seja, segundo a hipótese da ergodicidade, durante um tempo suficientemente longo, o comportamento médio de um sistema pode ser deduzido à partir da trajetória de um “ponto típico” do sistema, equivalendo à média de todos os microestados do ensemble microcanônico, no qual a energia é bem definida.

Enquanto Boltzmann descreve a distribuição de probabilidade de um único subsistema, para sistemas fechados dentro e fora do equilíbrio termodinâmico, Gibbs propõe uma generalização desse resultado. Ele sugere um método mais conveniente de aproximação da solução desse problema baseado na aplicação da distribuição microcanônica de todo o sistema, como demonstrado brevemente na seção 2.2.

As demonstrações a seguir foram retiradas das interpretações de Zubarev, em que ele procura examinar conceitos básicos da mecânica estatística clássica aplicados a sistemas de partículas em equilíbrio usando o método de Gibbs.

Na mecânica estatística, utilizam-se interpretações probabilísticas para descrever processos dinâmicos de partículas interagindo. Gibbs não se contenta em retratar apenas um sistema, mas sim um grande número de cópias desse sistema com condições idênticas e, por isso, introduz o conceito de ensemble estatístico com a ideia de “representar” esse estado macroscópico do sistema. Condições externas idênticas, no sentido macroscópico, significa que “todos os membros desse ensemble são caracterizados pelo mesmo valor dos parâmetros macroscópicos e pelo mesmo tipo de contato com os corpos em volta”(ZUBAREV, 1974).

Assim, define-se o ensemble estatístico pela função distribuição $f(p, q, t)$, que representa a densidade de probabilidade da distribuição de sistemas no espaço de fase. A probabilidade de encontrar o sistema no tempo t dentro do elemento do espaço de fase $dpdq$, próximo ao ponto $(p, q) = (p_1, \dots, p_N; q_1, \dots, q_N)$, é dada por

$$dw = f(p, q, t)dpdq. \quad (2.20)$$

Mas, levando em consideração que a estatística clássica é o caso limite da estatística quântica para temperaturas suficientemente elevadas, quando os efeitos quânticos podem ser desconsiderados, a condição de normalização em 2.20 não corresponde a essa passagem da estatística quântica para o limite clássico(ZUBAREV, 1974).

Com certa aproximação, sabe-se que a dimensão mínima para o movimento unidimensional de uma partícula é definida pela constante de Plank $h = 2\pi\hbar$,

$$\Delta q_i \Delta p_i \geq h. \quad (2.21)$$

Ou seja, a dimensão mínima de uma célula no espaço de fase ocupada por uma partícula é h^3 , e no espaço com N partículas o volume é h^{3N} . Além disso, considerando que a permutação entre partículas idênticas do ensemble não altera o seu estado, o elemento do volume de fase deve ser reduzido por um fator $N!$ e, portanto, a função distribuição é definida, finalmente, pela relação:

$$dw = f(p, q, t) \frac{dp dq}{N! h^{3N}}. \quad (2.22)$$

Nesse caso, se estabelece a condição de normalização na forma

$$\int f(p, q, t) d\Gamma = 1, \quad (2.23)$$

onde

$$d\Gamma = \frac{dp dq}{N! h^{3N}} \quad (2.24)$$

simboliza o elemento de volume de fase adimensional, e a integral corresponde à somatória sobre todos diferentes estados.

Para se entender a formulação de entropia por Gibbs, é preciso compreender a evolução de uma função de distribuição no tempo a partir do teorema de Liouville. A equação de Liouville assume que a função distribuição é constante ao longo da trajetória no espaço de fase e pode ser escrita em termos dos *brackets de Poisson*¹. Ou, escrita na escrita na sua forma reduzida

$$\frac{\partial f}{\partial t} = \{H, f\}. \quad (2.25)$$

Essa é a fórmula básica para se construir um ensemble estatístico, estando ele em equilíbrio ou não. A partir dessa equação, é possível calcular o valor de f para qualquer tempo t , se for conhecido o valor de f em um tempo $t = t_0$. Se o caso for de equilíbrio estatístico, f e H não devem depender explicitamente do tempo, e a equação de Liouville toma a forma:

$$\{H, f\} = 0. \quad (2.26)$$

Define-se, então, o logaritmo da função distribuição com sinal negativo

$$\eta = -\ln f(p, q, t), \quad (2.27)$$

onde $-\eta$ é chamado por Gibbs de expoente de fase. A partir dessa relação, pode-se observar que η , assim como f , satisfazem a equação de Liouville

$$\frac{\partial \eta}{\partial t} = \{H, \eta\}. \quad (2.28)$$

¹ Os brackets de Poisson são definidos pela expressão:

$$\{H, f\} = \sum_k \left(\frac{\partial H}{\partial q_k} \cdot \frac{\partial f}{\partial p_k} - \frac{\partial H}{\partial p_k} \cdot \frac{\partial f}{\partial q_k} \right) \equiv \frac{\partial H}{\partial q} \frac{\partial f}{\partial p} - \frac{\partial H}{\partial p} \frac{\partial f}{\partial q},$$

onde H é a função Hamiltoniana, q a coordenada espacial e p o momentum.

O valor médio do logarítmo da função distribuição, representado pelo expoente de fase, corresponde à *entropia de Gibbs* que, com a condição de normalização 2.24, pode ser expressa como

$$S = -\langle \eta \rangle = - \int f(p, q, t) \ln f(p, q, t) \frac{dp dq}{N! h^{3N}}. \quad (2.29)$$

“Para um gás diluído, os estados das diferentes partículas podem ser tomados como estatisticamente independentes”(ZUBAREV, 1974) e, conseqüentemente, a função distribuição de um sistema completo pode ser obtida a partir do produto das funções de distribuição individuais de cada partícula:

$$f(p, q, t) = \frac{N!}{N^N} \prod_{i=1}^N f_i(p_i, q_i, t), \quad (2.30)$$

que por sua vez, respeitam a condição de normalização dada por

$$\int f_i(p_i, q_i, t) \frac{dp_i dq_i}{h^3} = N. \quad (2.31)$$

Tendo em vista que $\ln N! \approx N \ln \frac{N}{e}$, o fator $N!/N^N$ é incluído em 2.30 para que haja consistência em ambas as normalizações, 2.23 e 2.31.

Para sistemas descritos por distribuições como 2.30, a entropia 2.29 equivale à entropia proposta por Boltzmann, $S = S_B$, onde

$$S = - \int f_1(p_1, q_1, t) \frac{\ln f_1(p_1, q_1, t)}{e} \frac{dp_1 dq_1}{h^3}. \quad (2.32)$$

$f_1(p_1, q_1, t)$ representa a função distribuição para uma única partícula.

A entropia de Boltzmann se comporta bem com relação às leis termodinâmicas no sentido de que, para sistemas isolados, a entropia tende a crescer com o tempo ou permanece constante se estiverem em equilíbrio estatístico. Todavia, a definição 2.32 nos dá a expressão correta de entropia como uma função termodinâmica em estado de equilíbrio apenas para gases ideais. A entropia de Boltzmann não deve ser empregada de forma generalizada.

A definição de entropia de Gibbs satisfaz totalmente os casos em que o sistema encontra-se em equilíbrio. Mas se o sistema é isolado, é fácil perceber que a entropia não depende do tempo e, por isso, não pode crescer. Pelo teorema de Liouville, a função distribuição é invariante no tempo, e para tempos diferentes (t_0 e t , por exemplo), tem-se $f(p^0, q^0, 0) = f(p, q, t)$. Aplicando a definição de Gibbs 2.29 e de acordo com o teorema de Liouville, obtem-se que os pontos de fase ocupando uma região inicial no tempo t_0 permanecem constantes no tempo t , ou seja, $dp dq = dp^0 dq^0$. Em outras palavras, “o movimento do ponto de fase que representa o sistema no espaço de fase é semelhante ao movimento de um fluxo incompressível”(ZUBAREV, 1974).

Ainda que a definição seja mais abrangente que a de Boltzmann, percebe-se que a entropia proposta por Gibbs para sistemas isolados não varia com o tempo. Uma maneira de ajustar a equação de Gibbs para que a entropia de sistemas isolados aumente com o tempo é através do método de “granulação”, onde a função distribuição $f(p, q, t)$ (denominada densidade de grão-fino) é substituída por uma outra (densidade de grão-grosso)

$$\tilde{f}(p, q, t) = \frac{1}{\omega} \int_{\omega} f(p, q, t) dp dq, \quad (2.33)$$

onde a integração é realizada sobre as pequenas células (com valor fixo) de um espaço de fase. Esse ajuste corresponde ao fato de que todas as quantidades observáveis são sempre medidas médias com relação a uma certa região do espaço. A mecânica quântica impõe um limite mínimo para o tamanho das células, não podendo ser menor do que h^{3N} . Uma melhor demonstração desse método pode ser encontrada no livro de Zubarev.

A *distribuição canônica* estabelecida por Gibbs por volta de 1901 na estatística quântica é dada por

$$w_n = Ae^{-E_n/T} \quad (2.34)$$

onde A é a constante de normalização independente de E_n . A relação 2.34 indica a distribuição estatística de qualquer sistema macroscópico que seja uma pequena parte de um grande sistema isolado (LANDAU; LIFSHITZ, 1958). No entanto, se estendermos para um caso em que a distribuição seja dependente do tempo, essa relação precisa de ajustes.

Se o número de partículas de um corpo varia entre os diferentes subsistemas, a distribuição de Gibbs 2.34 é generalizada de maneira que N denotará o número de partículas (ainda idênticas) dentro de um dado volume, referente ao volume do subsistema. Assim, a função distribuição depende não só da energia do estado quântico como do número de partículas presente no corpo. A probabilidade que o subsistema contenha N partículas no n -ésimo estado é representado por w_{nN} , onde os níveis de energia são expressos por E_{nN} para cada diferente valor de N :

$$w_{nN} = Ae^{(\mu N - E_{nN})/T}. \quad (2.35)$$

Pela condição de equilíbrio, o potencial químico μ é o mesmo para todo o corpo e meio em que encontra inserido (LANDAU; LIFSHITZ, 1958). A partir da condição de normalização, conseguimos descrever a constante A em termos das quantidades termodinâmicas. Pela relação 2.9, a entropia do corpo é

$$S = -\langle \log w_{nN} \rangle = l \log A - \frac{\mu \bar{N}}{T} + \frac{\bar{E}}{T} \quad (2.36)$$

e, portanto,

$$T \log A = \bar{E} - TS - \mu \bar{N}. \quad (2.37)$$

Mas como $\bar{E} - TS = F$ é a energia livre de Gibbs, e a diferença $F - \mu\bar{N}$ é o potencial termodinâmico (Ω), então $\Omega = T \log A$ e a função distribuição pode ser reescrita como

$$w_{nN} = e^{(\Omega + \mu N - E_{nN})/T} \quad (2.38)$$

Se considerarmos um ensemble de um sistema energeticamente isolado e fechado dentro de um volume constante V , podemos descrever a função de distribuição como sendo constante dentro de um intervalo ΔE e zero fora desse intervalo:

$$f(p, q) = \begin{cases} [\Omega(E, N, V)]^{-1}, & \text{para } E \leq H(p, q) \leq E + \Delta E, \\ 0 & \text{fora do intervalo} \end{cases} \quad (2.39)$$

Essa distribuição é reconhecida como *distribuição microcanônica* e a constante $\Omega(E, N, V)$ é chamada de *peso estatístico* e é determinada pelos três parâmetros extensivos E, V e N . A entropia no caso de um ensemble microcanônico é estabelecida como o logaritmo natural do peso estatístico, que representa o número de estados em ΔE ,

$$S(E, N, V) = \ln \Omega(E, N, V), \quad (2.40)$$

onde observa-se que, definida dessa forma, a entropia estatística possui as mesmas propriedades da entropia termodinâmica.

2.5 Entropia do gás ideal

Boltzmann foi o primeiro a deduzir das leis da termodinâmica uma fórmula de entropia para o caso dos gases ideais. Na verdade, quase todas as aplicações detalhadas já feitas da segunda lei trataram de apenas uma área específica: o comportamento dos gases (WOLFRAM, 2002). A célebre fórmula de entropia proposta por Boltzmann em 1876:

$$S = k_B \log W \quad (2.41)$$

sugere que a entropia S é proporcional ao logaritmo do número de microestados de um sistema em equilíbrio termodinâmico representado por W , cuja constante de proporcionalidade é a constante de Boltzmann, $k_B = 1.3806485 \dots 10^{-23}$ joules por kelvin (J/K) (RI-OUL, 2018).

Um gás é considerado ideal quando as interações entre as partículas são tão fracas a ponto de serem desconsideradas (LANDAU; LIFSHITZ, 1958). Isso pode ocorrer devido à natureza das interações ou devido a ele ser rarefeito. Dependendo do distanciamento entre as moléculas de um gás, a influência de uma partícula sobre outra pode ser considerada irrelevante. Assim, a função distribuição derivada por Gibbs para corpos relativamente pequenos em sistemas fechados pode ser empregada no caso dos gases ideais, tendo em vista que cada molécula separada pode ser considerada um sistema independente quase fechado.

“A ausência de interação entre partículas de um gás ideal permite que o problema da mecânica quântica em determinar os níveis de energia E_n de um gás como um todo, se reduza à caracterização dos níveis de energia de uma única molécula” (LANDAU; LIFSHITZ, 1958), seção 37. Os autores sugerem que esses níveis são denotados por ϵ_k ; o sufixo k representa o conjunto de números quânticos que define o estado da molécula e a energia E_n é dada pela soma das energias de várias moléculas. Além disso, a mecânica quântica estabelece alguns efeitos mútuos de partículas que estão no mesmo estado quântico, chamados *efeitos de troca*.

Seja n_k o número de partículas de um gás que ocupam o k -ésimo estado quântico dentre os vários possíveis. Para caracterizar um gás que seja suficientemente rarefeito, como é o caso de um gás ideal, o valor médio de partículas, calculado para todos os k estados, deve ser:

$$\bar{n}_k \ll 1. \quad (2.42)$$

Essa condição sugere que a ocupação média não deve exceder 1 partícula por estado quântico a cada instante. Aplicando-se a relação 2.34, a probabilidade de se encontrar uma molécula no estado quântico k é proporcional a $\exp^{-\epsilon_k/T}$, e o número médio \bar{n}_k de partículas nesse mesmo estado é expresso por

$$\bar{n}_k = a \exp^{-\epsilon_k/T} \quad (2.43)$$

onde a é a constante obtida a partir da condição de normalização,

$$\sum_k \bar{n}_k = N, \quad (2.44)$$

e N é o número total de partículas contido no gás. A distribuição de moléculas de um gás ideal sobre os estados possíveis dada pela relação 2.43 foi descoberta pela primeira vez em 1877 por Boltzmann dentro da estatística clássica e é chamada de *distribuição de Boltzmann*.

A constante a pode ser representada em termos das quantidades termodinâmicas desse gás. Para isso, aplicamos a distribuição derivada por Gibbs 2.35 ao conjunto de partículas do gás que se encontra em um dado estado quântico, uma vez que os *efeitos de troca* acontecem somente entre partículas que estão no mesmo estado. Estabelecendo $E = n_k \epsilon_k$, $N = n_k$ e adicionando o sufixo k ao potencial de Gibbs em 2.38, definimos a distribuição de probabilidade para um número variável de partículas como:

$$w_{n_k} = e^{[\Omega_k + n_k(\mu - \epsilon_k)]/T}. \quad (2.45)$$

Por essa relação, $w_0 = e^{\Omega_k/T}$ corresponde à probabilidade de não se encontrar nenhuma partícula no estado escolhido e $w_1 = e^{\Omega_k + (\mu - \epsilon)}$ à probabilidade de se encontrar uma partícula no estado k . Como o caso de interesse é para $\bar{n}_k \ll 1$, a probabilidade w_0 é

considerada próxima de 1, e $w_1 = e^{(\mu - \epsilon_k)/T}$. Por fim, podemos escrever

$$\bar{n}_k = \sum_k w_{n_k} n_k = w_1 \cdot 1 \quad (2.46)$$

obtendo, finalmente, a distribuição de Boltzmann para o gás ideal:

$$\bar{n}_k = e^{(\mu - \epsilon_k)/T}, \quad (2.47)$$

onde o coeficiente em 2.43 é escrito em termos do potencial químico do gás.

3 Entropia informacional

3.1 História e contexto da teoria da informação

Nos anos 50 foi reconhecida uma relação matemática equivalente à entropia, capaz de medir a quantidade de informação transmitida em um sistema. Essa quantidade é uma das protagonistas dentro da teoria desenvolvida por Shannon.

Claude Elwood Shannon nasce em 1916 em Michigan, nos Estados Unidos. Com 16 anos ele ingressa na Universidade de Michigan, onde cursa matemática e engenharia elétrica simultaneamente. Na sua tese de mestrado, aos 21 anos, ele revoluciona o uso de circuitos lógicos ao desenvolver um esboço da teoria de circuitos digitais e ganha o prêmio Alfred Nobel, em 1940, pelo reconhecimento de seu trabalho. Nesse mesmo ano, após defender sua tese de doutorado na MIT, ele começa a trabalhar nos Laboratórios de Telefonia Bell, onde um de seus principais interesses consistia em compreender o que ele chamava de “transmissões de inteligência”, o que se tornará mais tarde a teoria da informação (RIOUL, 2018).

Segundo Rioul, foi durante a Segunda Guerra que Claude Shannon passou aproximadamente dez anos de sua vida (1939-1948) refletindo e explorando a concepção do que é informação. Durante esse período, ele desenvolvia uma nova teoria que ele intitulava como *teoria da comunicação* e utilizava o termo “incerteza” para descrever o que entendeu-se depois como ‘entropia’. Apesar disso, Shannon deixava claro que entendia informação como um termo probabilístico e não semântico, como é possível encontrar no trecho retirado de sua obra (SHANNON, 1948 apud RIOUL, 2018):

O problema fundamental da comunicação é reproduzir em um ponto exatamente ou aproximadamente uma mensagem selecionada em outro ponto. Frequentemente, as mensagens têm significado; isto é, eles se referem ou são correlacionados de acordo com algum sistema com certos aspectos físicos ou entidades conceituais. Esses aspectos semânticos da comunicação são irrelevantes para o problema de engenharia. O aspecto significativo é que a mensagem real é selecionado de um conjunto de mensagens possíveis. O sistema deve ser projetado para operar cada seleção possível, não apenas aquela que realmente for escolhida, pois isso é desconhecido no momento da elaboração do sistema.

No ano seguinte da publicação de seu artigo, Shannon juntamente com seu chefe no Laboratórios Bell, Warren Weaver, republicaram seu trabalho em formato de livro. Weaver foi responsável por difundir as ideias propostas por Shannon com uma abordagem

mais filosófica para o público geral ao explicar, logo nos primeiros capítulos do livro, o funcionamento de um sistema de comunicação. Além disso, ele amplia a visão dos leitores para possíveis aplicações dessa teoria em outras áreas da ciência que envolvem o problema da comunicação, como a linguística e as ciências sociais, por exemplo.

O trabalho de Shannon foi capaz de solucionar os problemas de compressão e transmissão de dados que surgiam naquela época. No primeiro capítulo do livro, Weaver cita questões que aparecem no amplo tema da comunicação e que seriam abordadas matematicamente por Shannon em três níveis:

- NÍVEL A: O quão acurado os símbolos de comunicação podem ser transmitidos? (Problema técnico)
- NÍVEL B: Com que precisão os símbolos transmitidos podem comunicar o significado desejado? (Problema semântico)
- NÍVEL C: Com que eficácia o significado recebido afeta a conduta da maneira desejada? (Problema efetivo)

Já no segundo capítulo, Weaver apresenta o diagrama mais importante do livro, cuja representação esquematiza simbolicamente um sistema de comunicação (Figura 3). Pela primeira vez na história foi possível distinguir fonte, canal e destinatário; emissor e receptor; sinal e ruído (RIOUL, 2018).

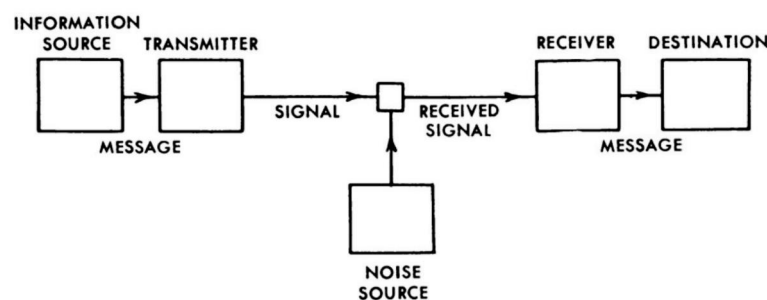


Figura 3 – Diagrama de um sistema de comunicação. Fonte: (SHANNON, 1948)

A fonte de informação seleciona uma dentre as várias mensagens possíveis. O transmissor codifica a mensagem desejada em um sinal e envia para o receptor através de um canal de comunicação. O receptor decodifica o sinal recebido e o transfere para o destinatário. Nesse processo podem haver alterações no sinal enviado; essas distorções ocorrem no momento que o sinal passa pelo canal e são conhecidas como ruídos.

3.2 Entropia como medida de informação

A palavra *comunicação* na teoria desenvolvida por Shannon deve ser entendida *em um sentido muito amplo para incluir todos os procedimentos pelos quais uma mente pode afetar outra* (SHANNON, 1948). E o termo *informação* na teoria da comunicação não deve ser confundido com o sentido usual de “significado”. De acordo com Weaver, o sentido de informação não é exatamente sobre *o que* você diz, mas o que você *pode* dizer. O conceito de informação, então, não se aplica a uma única mensagem, mas ao grau de liberdade que algo ou alguém tem ao fazer a escolha de uma mensagem dentre as várias possibilidades.

A concepção de entropia derivada por Shannon visa oferecer uma certa “intuição” sobre a medida da informação. Primeiro ele propõe uma aproximação axiomática, onde enuncia algumas propriedades, e depois estende-se na formulação matemática (RIOUL, 2018). A quantidade de informação $i(p)$ é tomada como uma medida de “surpresa”, “inesperado”, “incerteza”, onde sua representação indica uma função da probabilidade p de um evento cujas propriedades fundamentais são:

- (i) $i(p) \geq 0$ é uma função decrescente de p ;
- (ii) Para quaisquer eventos independentes com probabilidades p e q , $i(pq) = i(p) + i(q)$.

Como demonstrado no artigo, nota-se que a quantidade de informação $i(p)$ deve ser proporcional a $\log(1/p)$ já que essa função satisfaz as condições (i) e (ii), e como a escolha da constante de proporcionalidade é arbitrária, pode-se escrever:

$$i(p) = \log \frac{1}{p}. \quad (3.1)$$

A escolha da base do logaritmo define a unidade escolhida para medida da informação.

Nesse sentido, se a fonte de informação é capaz de transmitir uma mensagem, deve-se levar em conta as probabilidades de se escolher cada elemento que a compõe. Assumindo que essa mensagem seja expressa por uma variável aleatória X e tenha distribuição de probabilidade discreta $p(x)$, a quantidade de informação de um único evento, $X = x$, é $\log \frac{1}{p(x)}$. Então a quantidade média de informação sobre uma série de eventos de X será expressa pela soma dos valores médios de informação de cada evento, ou seja:

$$H(X) = \sum_x p(x) \log \frac{1}{p(x)} = - \sum_x p(x) \log p(x), \quad (3.2)$$

onde $H(X)$ é a notação utilizada para indicar a entropia da variável medida X ; mesmo sendo uma função de $p(x)$, e não de X . Sendo assim, é natural que a informação seja medida pela entropia dado que, na teoria da comunicação, a informação nada mais é que a liberdade de escolha que temos ao construir uma mensagem (SHANNON, 1948).

Conhecendo a entropia de uma fonte de informação, é possível encontrar a entropia máxima que ela pode alcançar apenas impondo a condição de que ela utilize os mesmos símbolos para compor a mensagem. A razão entre a entropia em uma situação específica e o valor máximo de uma fonte de informação é conhecida como *entropia relativa* e indica a liberdade de escolha dessa fonte diante dos mesmos símbolos (ver anexos A.1 e A.2). A fração restante, ou seja, um menos a entropia relativa, representa parte da mensagem que foi determinada por alguma lei estatística imposta na configuração da fonte e é chamada de *redundância*. Isso significa dizer que essa estrutura é de alguma forma desnecessária ou repetitiva, podendo ser desconsiderada sem que a compreensão da mensagem seja afetada. Para exemplificar, Weaver cita que a redundância na língua inglesa é de aproximadamente 50%. Isso significa que apenas metade das letras ou palavras que usamos ao falar ou escrever, em inglês, está sob nossa liberdade de escolha, a outra metade está sob controle da estrutura estatística da língua.

3.3 Unidade de Informação em bits

Levando em consideração que a informação é a medida da liberdade de escolha, e também de incerteza que uma fonte possui ao selecionar símbolos para compor uma mensagem, é preciso saber que unidade de medida usar para quantificar essa escolha. Nessa seção mostraremos que essa unidade está relacionada com a base escolhida para o logaritmo. Além disso, será possível visualizar como a escolha das funções de probabilidade está associada ao nível de incerteza sobre um sistema.

Inicia-se com uma situação simples, em que a fonte precisa escolher uma dentre duas mensagens possíveis. Para isso, é conveniente que se utilize a base 2 no logaritmo. Como $\log_2 2 = 1$, considera-se então que ela possui apenas um grau de liberdade de escolha entre uma ou outra e, portanto, utiliza-se esse caso para caracterizar uma quantidade unitária ou padrão (SHANNON, 1948).

Seguindo com essa lógica, entenderemos o uso da unidade de informação para outras bases, como é abordado no artigo (RIOUL, 2018), seção 10. É disponibilizado para uma fonte de informação uma distribuição discreta X que tenha M elementos equiprováveis, $x = 0, 1, 2, \dots, M - 1$, ou seja, cada elemento com probabilidade $p(x) = 1/M$. Como a informação tem propriedade aditiva, a entropia do conjunto de elementos correspondente terá valor igual ao logaritmo do número total de valores possíveis da sequência:

$$H(X) = \log M. \quad (3.3)$$

Se a quantidade de elementos dessa sequência é múltipla de 10, por exemplo, uma sequência será composta por números entre 0 e $10^n - 1$, onde $M = 10^n$. Escolhendo o logaritmo na base 10, a entropia indicará o número $m = \log_{10} M$ de dígitos decimais

necessários para configurar essa sequência. De forma similar, um número de m dígitos na base 2 pode ser escolhido aleatoriamente entre 0 e $2^m - 1$ ($M = 2^m$), de forma a fornecer uma entropia de $m = \log_2 M$ dígitos binários. O mesmo pode ser feito com qualquer base escolhida, embora a base 2 seja a mais utilizada na linguagem computacional atualmente. Normalmente, a entropia também será expressa com o logaritmo na base 2, e a unidade utilizada para indicar a quantidade de informação na sequência medida será o *bit*, contração de *binary digit*.

Enquanto um dígito binário (bit) pode ser 0 ou 1, a entropia $H(X)$ também pode ser expressa em *bits*, nesse caso como unidade de grandeza *bit = binary unit*, e pode ter qualquer valor positivo. Equivalentemente, a unidade de informação na base 10 é o *dit = decimal unit*, indicando a quantidade de informação existente em uma sequência com M valores expressos em m dígitos decimais. Em logaritmos naturais, em que a base é e , a unidade empregada é *nat = natural unit*.

Para demonstrar a diferença entre dígito binário e unidade binária, considere-se uma distribuição discreta de espaço amostral $X \in [0, 1]$. Essa distribuição segue a distribuição de Bernoulli, na qual um parâmetro p significa a probabilidade de sucesso e $q = 1 - p$ a probabilidade de falha. A entropia nesse caso, além de possuir qualquer valor entre 0 e 1, pode ser expressa, em bits, por

$$H(X) = p \log_2 \frac{1}{p} + (1 - p) \log_2 \frac{1}{1 - p}. \quad (3.4)$$

A partir dessa relação, o autor conclui que o valor de entropia atinge seu valor máximo igual a 1 na situação em que $x = 0$ e $x = 1$ forem equiprováveis, ou seja, quando a fonte é completamente livre para escolher entre uma ou outra mensagem. Caso contrário, se uma mensagem é mais provável que outra, a entropia de 1 dígito binário é, na verdade, menor que 1 bit.

No caso limite em que $p = 1$ e $q = 0$, temos certeza que um evento ocorre (certeza) e o outro não (impossibilidade), ou seja, não há incerteza, não há liberdade de escolha sobre uma mensagem e, portanto, não há entropia ($H = 0$).

Nota-se, então, que “a entropia (H) tem seu valor maximizado quando as probabilidades de seus eventos são iguais, isto é, quando a fonte é completamente livre e imparcial na hora de escolher a mensagem, e reduz a zero quando sua liberdade de escolha é nenhuma” (SHANNON, 1948).

O sistema internacional (SI) sugere a utilização de *shannon* (Sh) como unidade, ao invés de bit, para distinguir a quantidade de informação da quantidade de dados utilizados para representar essa informação. Nesse caso, retomando o exemplo acima, a entropia $H(X)$ de um dígito binário (bit) deve medir entre 0 e 1 Sh.

3.4 Entropia de sistemas contínuos

Até então, nos preocupamos em definir a informação contida numa mensagem que é formada por uma sequência de símbolos. No entanto, a teoria se estende para os casos de informação representada por sinais contínuos, "como a voz falada com sua variação contínua de tom e energia", como exemplifica Weaver (SHANNON, 1948). Assim, o conceito de entropia de Shannon pode ser generalizado para variáveis contínuas por analogia. Nesse intuito, a equação 3.2 pode ser reescrita como um valor esperado

$$H(X) = \mathbb{E} \log \frac{1}{p(X)}, \quad (3.5)$$

onde $p(x)$ é a distribuição discreta de X e $\mathbb{E} = \sum_x p(X)$ o somatório sobre todas as probabilidades. A generalização para variáveis contínuas pode ser feita utilizando exatamente a mesma fórmula, onde agora $p(x)$ está ligada a uma distribuição contínua,

$$h(X) = \mathbb{E} \log \frac{1}{p(X)}, \quad (3.6)$$

onde, nesse caso, a soma discreta é substituída pela integral (soma contínua),

$$h(X) = \int p(x) \log \frac{1}{p(x)} dx. \quad (3.7)$$

Assim como na física estatística, $p(x)$ aqui não se refere mais a uma probabilidade, mas a uma densidade de probabilidade, que se comporta de forma diferente. Quando uma variável aleatória U é contínua e uniformemente distribuída sobre o intervalo (a, b) , por exemplo, para $p(u) = 1/(b - a)$ a equação 3.6 se torna

$$h(U) = \mathbb{E} \log \frac{1}{p(U)} = \log(b - a) \quad (3.8)$$

Enquanto a entropia para sistemas discretos é sempre positiva, a expressão acima se torna negativa quando o intervalo considerado é menor que 1. E no limite desse intervalo indo a zero, a entropia tenderá a infinito para uma variável aleatória *determinística* (constante), $X = c$. Isto é,

$$h(c) = -\infty. \quad (3.9)$$

Esse resultado é distinto da entropia discreta para um sistema sem incerteza, que é simplesmente $H(c) = 0$. Ou seja, a entropia contínua não define uma quantidade de informação visto que ela é capaz de ter valores negativos. Shannon foi capaz de contornar esses problemas e desenvolver formulações que definem os limites máximos e mínimos da capacidade de um canal e a taxa de transmissão para mensagens contínuas, como veremos adiante.

3.5 Teoremas da codificação

Uns dos principais tópicos abordados por Shannon em seu artigo foi o da compressão e transmissão de informação. Analisando o diagrama da figura 3, observa-se que, enquanto a fonte de informação seleciona a mensagem desejada dentre várias possibilidades, o transmissor (como o caso de um telefone, por exemplo) é aquele que modifica a mensagem recebida em um *signal* e a envia para o receptor (outro telefone) através do *canal* de comunicação (o ar por onde passam as ondas eletromagnéticas). “Quando eu falo com você, meu cérebro é a fonte de informação, e o seu, o destino; meu sistema vocal é o transmissor, e o seu ouvido e o oitavo nervo associado a ele são receptores” (SHANNON, 1948). O transmissor, então, é responsável pela *codificação* e o receptor pela *decodificação* da mensagem.

Compreenderemos adiante que a entropia (ou informação) associada ao processo de formação de uma mensagem ou sinal é determinada pelo caráter estatístico do processo. A natureza estatística na concepção de uma mensagem é atribuída à fonte de informação, mas no caso da reprodução de um sinal a entropia é estabelecida com base no que o canal de comunicação recebe de informação relacionado com a sua capacidade de lidar com as diferentes situações que um sinal pode produzir (SHANNON, 1948).

3.5.1 Primeiro Teorema da Codificação

Começamos com o cenário em que o canal é isento de ruído, para entender melhor como funciona a codificação. O objetivo é encontrar a maior compressão da informação para que ela seja transmitida da fonte e alcance o canal de forma segura usando o mínimo de dados necessário, mas sem perder a essência da mensagem. Mas primeiro, entende-se como Shannon chegou na fórmula de entropia. A mensagem a ser transmitida pelo emissor é codificada na fonte de informação e pode ser expressa por uma longa sequência de elementos independente e identicamente distribuídos (i.i.d.),

$$\underline{x} = (x_1, x_2, \dots, x_n). \quad (3.10)$$

Considerando que $p(x)$ seja a probabilidade de que cada elemento x_i resulte em x , dada uma variável aleatória X , a probabilidade $p(\underline{x})$ é dada pelo produto das probabilidades individuais dos elementos de X , já que estes são independentes:

$$p(\underline{x}) = \prod_x p(x)^{n(x)}, \quad (3.11)$$

onde $n(x)$ é a quantidade de elementos x_i igual a x . O número n , referente à quantidade de elementos da sequência, tende a ser um número grande. A frequência de vezes que x aparece em relação ao número de elementos possíveis pode ser assimilada com sua probabilidade,

$$\frac{n(x)}{n} \approx p(x). \quad (3.12)$$

Introduzindo essa expressão 3.12 em 3.11 chega-se a um importante resultado deduzido por Shannon, chamada *propriedade da equipartição assintótica* (PEA):

$$p(\underline{x}) \approx \left(\prod_x p(x)^{p(x)} \right)^n = \exp(-nH(X)). \quad (3.13)$$

na qual a probabilidade decresce exponencialmente a medida que $n \rightarrow \infty$. A partir desse resultado obtém-se a entropia de Shannon, cuja expressão é dada pelo decaimento exponencial, com $H(X) \geq 0$.

A PEA (3.13) exprime essencialmente que, quando n é um número suficientemente grande (mas fixo), o valor da probabilidade de se obter uma sequência “típica” (3.10) tende a ser próxima do valor constante $\exp(-nH(X))$. Em linguagem matemática, um comportamento é típico quando as chances de ocorrer são muito grandes. Segundo o autor, pela lei dos grandes números, qualquer sequência longa o suficiente e aleatoriamente escolhida pode ser considerada “típica”, ou seja, com a probabilidade arbitrariamente próxima de 1 (RIOUL, 2018).

A partir dessa propriedade, o problema se resume em encontrar uma forma confiável de codificar uma fonte de informação com a menor taxa possível. Uma vez que as sequências não típicas possuem probabilidade arbitrariamente pequenas, escolhe-se codificar as típicas para uma compressão mais confiável. Assim, somando a equação 3.13 sobre todas as N sequências típicas ($p(\underline{x}) = \sum_i p(\underline{x}_i)$) e levando em consideração que n seja suficientemente grande, a probabilidade total de se escolher aleatoriamente uma sequência que seja típica é aproximadamente um:

$$1 \approx N \exp(-nH(X)). \quad (3.14)$$

Isso dá um número de aproximadamente $N \approx \exp(nH(X))$ sequências típicas. A taxa de codificação R de uma fonte é a razão entre o logaritmo de N e o número de elementos da sequência, n :

$$R = \frac{\log N}{n} \approx H(X). \quad (3.15)$$

A taxa mínima que uma fonte pode ser codificada confiavelmente é dada pelo valor de sua entropia $H(X)$. A partir dessa relação foi possível elaborar um teorema que ficou conhecido como *primeiro teorema da codificação de Shannon* (SHANNON, 1948). Esse teorema é reconhecido por garantir a melhor performance de qualquer algoritmo de compressão de dados e, por essa razão, recebe um significado operacional importante: “corresponde a taxa mínima de bits informativos em uma fonte de informação” (RIOUL, 2018). Esse resultado será uma das bases para nossa definição de entropia de um sistema de partículas.

3.5.2 Segundo Teorema da Codificação

O Segundo Teorema da Codificação surge para explicar o caso mais geral, em que o sistema de comunicação é afetado pela presença de um ruído. A partir dele foi possível solucionar o problema de transmissão de informação que consiste em transmitir uma informação de uma fonte para um canal de comunicação com a maior velocidade possível (RIOUL, 2018).

Se um ruído é adicionado ao sistema, a mensagem recebida irá conter certas distorções, erros e materiais estranhos. E por causa desse efeito sobre o canal, a incerteza sobre a mensagem cresce e, conseqüentemente, a quantidade de informação também aumenta, ou melhor, o sinal recebido é selecionado a partir de uma seqüência mais variada de elementos que o transmitido (SHANNON, 1948). A discussão que Weaver levanta é sobre o quanto de incerteza é produzida por uma fonte de informação se o sinal recebido é conhecido ou, de forma equivalente, qual a média de incerteza relativa ao sinal recebido, se a mensagem enviada é conhecida.

Shannon apresenta uma relação entre a capacidade de um canal de comunicação e a entropia correspondente à seleção de mensagens realizada pela fonte. Enquanto a entropia mede a quantidade de informação por símbolo, a capacidade de um canal é descrita em termos de sua habilidade de transmitir o que é produzido a partir da fonte de uma determinada informação, e é medida em bits por segundo. A relação entre a capacidade C e a entropia H mede, na verdade, a taxa de símbolos por segundo. Considerando que todo o processo de codificação do transmissor é bem sucedido, veremos que é possível transmitir símbolos a uma taxa média próxima de C/H , mas nunca excedendo essa taxa (SHANNON, 1948).

Considera-se, então, que exista uma outra variável aleatória X^* , referente ao sinal recebido, por exemplo. Enquanto a expressão 3.11 indica a probabilidade de uma seqüência \underline{x} com n elementos, utiliza-se

$$q(x) = \frac{n(X)}{n} \quad (3.16)$$

como a probabilidade empírica da seqüência \underline{x} . Assim, pela propriedade da equipartição assintótica e pela lei dos grandes números, pode-se reescrever a equação 3.13 como uma expressão exata

$$p(\underline{x}) \approx \left(\prod_x p(x)^{p(x)} \right)^n = \exp \left(-nH(X, X^*) \right), \quad (3.17)$$

onde $H(X, X^*)$ é conhecida como *entropia cruzada*

$$H(X, X^*) = \sum_x q(x) \log \frac{1}{p(x)}. \quad (3.18)$$

Como visto na seção anterior, a probabilidade de que a seqüência aleatoriamente escolhida seja típica é $N \exp(-nH(X, X^*))$, para N seqüências típicas possíveis. Nesse

caso, $N \exp(-nH(X, X^*)) \leq 1$ para qualquer valor de p ; para o caso particular em que $p = q$ tem-se $N \exp(-nH(X^*, X^*)) \leq 1$, ou melhor, $N \exp(-nH(X^*)) \leq 1$. Desse modo, chega-se a uma definição de *entropia relativa*, dada por:

$$H(X, X^*) - H(X^*) = \sum_x q(x) \log \frac{1}{p(x)} - \sum_x q(x) \log \frac{1}{q(x)} = D(q, p), \quad (3.19)$$

onde $D(q, p) = D(X^*, X) \geq 0$ é a entropia relativa ou divergência, conceito elucidado no anexo A.2. Observa-se, então, que a probabilidade de escolher aleatoriamente uma sequência que seja típica é limitada pela relação

$$N \exp(-nH(X, X^*)) \leq \exp(-nD(q, p)), \quad (3.20)$$

conhecida como *limite de grande desvio assintótico*. Ou seja, essa probabilidade é determinada pelas probabilidades das sequências que chegam e saem de um canal de comunicação, por exemplo.

Assume-se que uma sequência discreta $\underline{x} = (x_1, x_2, \dots, x_n)$ é o código de entrada em um canal sem memória e $\underline{y} = (y_1, y_2, \dots, y_n)$ o código de saída, que tem propriedade de ser típico junto com \underline{x} pela lei dos grandes números. Seria suficiente, na teoria, apenas decodificar a sequência \underline{y} depois de selecionar um único código \underline{x} que seja típico juntamente com \underline{y} . Na prática, “se outra sequência \underline{x}' de código for conjuntamente típica com \underline{y} , uma vez que as sequências de código são escolhidas para serem independentes, a distribuição de probabilidade real p da variável aleatória bivariada correspondente (X', Y) será $p(x)p(y)$, enquanto q é igual à distribuição conjunta $p(x, y)$ ” (RIOUL, 2018).

Se a probabilidade de se obter uma sequência típica obedece a relação acima 3.20, onde $D(q, p) = D(p(x, y), p(x)p(y)) = I(X; Y)$ pela definição A.7, então a probabilidade (média) de erro de decodificação de um canal que possui N sequências de códigos é limitada por

$$P_e \leq N \exp(-nI(X; Y)). \quad (3.21)$$

Para que esse erro seja exponencialmente pequeno à medida que $n \rightarrow +\infty$, “é suficiente que a taxa de transmissão por símbolo seja menor que a informação mútua” (RIOUL, 2018):

$$R_T = \frac{\log N}{n} < I(X; Y) \quad (3.22)$$

No intuito de maximizar a taxa de transmissão, a distribuição de probabilidade $p(x)$ deve ser escolhida de forma a maximizar a informação mútua $I(X; Y)$ entre as sequências. A *capacidade* de um canal,

$$C = \max_{p(x)} I(X; Y), \quad (3.23)$$

é a máxima quantidade de informação possível de ser transmitida por um canal de comunicação.

Com isso, chegamos ao resultado do *segundo teorema da codificação* de Shannon, o qual afirma existir um código que atinge uma probabilidade arbitrariamente pequena de erro de decodificação com a condição de que

$$R < C. \tag{3.24}$$

Ou seja, "uma comunicação arbitrariamente confiável pode ser alcançada desde que não exceda a capacidade de seu canal" (RIOUL, 2018).

A partir desse teorema, compreendeu-se que o ruído de transmissão não é um fator limitante para que se estabeleça uma comunicação confiável; ele apenas interfere na velocidade da transmissão. Portanto, o segundo teorema da codificação de Shannon foi o responsável por estabelecer a estrutura necessária para que a transmissão de dados sobre um canal ruidoso ocorra com a melhor performance possível.

4 Comparação das entropias

Pela história da termodinâmica, é possível observar as influências de Carnot nos trabalhos de Clausius e Kelvin, que postularam as duas leis da termodinâmica a partir das interpretações de seus trabalhos de 1824. O que pouco se comenta é que, embora Clausius leve o mérito de introduzir o conceito de entropia na termodinâmica, foi Carnot quem primeiro entendeu e possibilitou a compreensão dessa entidade, mas através do estudo do *calórico* (NASCIMENTO; BRAGA; FABRIS, 2004).

Na tentativa de solucionar o problema de rendimento das máquinas a vapor do século XIX, Sadi Carnot afirma que a potência máxima das máquinas é atingida através do restabelecimento do calórico, sem que haja perda deste durante o processo, ou seja, introduz a ideia de reversibilidade pelos processos adiabáticos. Carnot entende que a expansão de um gás no vácuo produz *vis viva*¹ e esta é convertida em calor por fricção, acarretando num aumento de calórico no gás (CALLENDAR, 1910). Carnot chega à conclusão de que a quantidade total de calórico permanece constante quando *a equalização da temperatura é efetuada por meio de motores reversíveis*. Além de estabelecer um sentido para o fluxo natural de calor, Carnot parece buscar uma relação entre o aumento e o restabelecimento desse calórico através de seu equilíbrio, indicando que a energia que está concentrada e organizada tende a se dispersar.

A ideia de calórico aparece posteriormente como “função termodinâmica” por Rankine, e também como “valor equivalente de transformação” por Clausius. Somente em 1865, após compreender a importância desse conceito, Clausius assimilará calórico ao conceito de entropia, e definirá como sendo a integral

$$\Delta S_C = \int \frac{dQ}{T}. \quad (4.1)$$

Mas foi Ludwig Boltzmann quem desenvolveu a teoria estatística que explica a assimetria temporal observada em processos irreversíveis de sistemas macroscópicos fora do equilíbrio, de maneira consistente com sua dinâmica microscópica reversível (LEBOWITZ et al., 1993). Enquanto a entropia proposta por Gibbs, S_G , é definida para um ensemble estatístico, Boltzmann faz uma separação entre a micro e macroescala associando a cada microestado X , de um sistema macroscópico no estado M , um valor S_B , a entropia de Boltzmann:

$$S_B(M(X)) = k \log |\Gamma_{M(X)}|. \quad (4.2)$$

¹ Do latim, ‘vis viva’ significa *força viva*, e é utilizada por Leibniz para representar o que entendemos hoje como energia cinética. Será utilizada mais tarde para demonstrar a conservação de energia proposta pela 1ª lei da termodinâmica.

Embora a derivação de Boltzmann seja feita em termos da mecânica clássica Newtoniana, acredita-se que a irreversibilidade quântica pode ser entendida também a partir de suas ideias, como é possível ver no artigo (LEBOWITZ et al., 1993).

Uma das principais contribuições de Boltzmann, segundo Lebowitz, foi estabelecer uma conexão direta entre sua função microscopicamente definida $S_B(M)$ e a entropia termodinâmica de Clausius, S_{eq} , para sistemas macroscópicos em equilíbrio. Assim, para um sistema em equilíbrio com uma dada energia E , volume V e um número de partículas N suficientemente grande, Boltzmann mostrou que

$$S_{eq}(E, V, N) \approx S_B(M_{eq}), \quad (4.3)$$

em que $M_{eq}(E, V, N)$ é o macroestado correspondente.

A segunda contribuição importante de Boltzmann foi quando ele utilizou sua equação (4.2) para definir entropia de sistemas fora do equilíbrio, e identificou um crescimento da entropia de Clausius a medida que o volume da região do espaço de fase $|\Gamma_{M(X)}|$ aumenta, comprovando a segunda lei da termodinâmica. O que acontece é que, ao remover a “parede restritiva” de um sistema macroscópico isolado, este evolui para um estado com maior entropia. Ou seja, o estado microscópico² X irá se mover com grande probabilidade para novas regiões disponíveis no espaço de fase até que seu estado $X(t)$ atinja a região $|\Gamma_{M_{eq}}|$ (LEBOWITZ et al., 1993).

Enquanto a segunda lei da termodinâmica busca expressar uma relação entre estados ordenados de acordo com a acessibilidade, a caracterização estatística assume que a dinâmica do sistema evolui de forma a manter o máximo de estados acessíveis. Ainda assim, ambas as teorias são compatíveis com relação a transições irreversíveis no que diz respeito a sistemas que partem de um estado menos provável para um estado mais provável. Ambas concordam com o fato de que a entropia corresponde a uma função monotonamente crescente no tempo. O processo inverso, envolvendo um decréscimo da entropia, é considerado como improvável na física estatística, ao passo que na termodinâmica clássica tal evento é postulado como sendo uma transição impossível (LANDAU; LIFSHITZ, 1958).

A análise de Boltzmann implica a existência da relação entre o volume do espaço de fase e a sua probabilidade (LEBOWITZ et al., 1993). Percebe-se que, embora apresentem diferenças conceituais, há uma equivalência entre a entropia de Boltzmann 4.2 e a entropia de Gibbs, definida por

$$S_G(\rho) = -k \int \rho(X) [\log \rho(X)] dX. \quad (4.4)$$

Nessa relação 4.4 $\rho(X)dX$ corresponde à densidade de ensemble, ou à probabilidade do estado microscópico do sistema ser encontrado no elemento de volume do espaço de fase

² A evolução de microestados é governada pela dinâmica Hamiltoniana - que conecta um microestado $X(t_0)$ a outro $X(t)$ em t .

dX , e cuja integral é sobre todo o espaço de fase $d\Gamma$.

A equivalência se dá quando $\rho(X)$ caracteriza um ensemble microcanônico generalizado, correspondente a um macroestado M , isto é, quando $\rho_{M(X)} = 1/|\Gamma_M|$ para X pertencente a Γ_M ou quando $\rho_{M(X)} = 0$ para qualquer outro caso:

$$S_G(\rho_M) = k \log |\Gamma_M| = S_B(M). \quad (4.5)$$

De maneira geral, as entropias de Gibbs (4.4), Boltzmann (4.2) e Clausius (4.1) coincidem para sistemas cuja densidade de partículas, de energia e de momento variam de forma quase imperceptível na escala microscópica e na condição de se ter um ponto de equilíbrio local. Nota-se, no entanto, que “a menos que o sistema esteja em completo equilíbrio e não haja mudanças sistemáticas de M ou ρ , a evolução temporal de S_B e S_G se comportam de maneira bem diferente” (LEBOWITZ et al., 1993). Enquanto S_G , que não se altera com o passar do tempo, mesmo para conjuntos que descrevem sistemas fora do equilíbrio, S_B possui um crescimento “típico” que explica a evolução em direção ao equilíbrio de tais sistemas. Portanto, segundo Lebowitz, a entropia relevante para se compreender a evolução temporal de um sistema macroscópico é S_B e não S_G .

Já a teoria de Claude Shannon surge em 1950 com um novo conceito de entropia que impulsiona a era digital, que identifica a entropia com uma medida da informação (ou desinformação) a respeito de um sistema. A entropia pode ser definida ainda como a quantidade de informação adicional sobre um sistema que ainda é desconhecido, ou seja, a informação adicional que seria necessária para identificar o arranjo específico que realmente ocorre (WOLFRAM, 2002).

Comparando a formula de Boltzmann 4.2 com a de Shannon 3.3, M equivale ao número de microestados de um sistema em equilíbrio termodinâmico Γ_M . A variante integral equipara-se com a primeira formula de entropia de Boltzmann (com sinal negativo) $S = - \int \rho \ln \rho dx$, onde a densidade de probabilidade ρ representa a fração de tempo gasto pelo sistema em torno de um ponto x dado no espaço de fase. Além disso, a formula de Shannon 3.2 assemelha-se matematicamente à formula de entropia de Von Neumann (1932) na mecânica quântica, $S = -Tr(\hat{D} \log \hat{D})$, onde $p(x)$ equipara-se aos autovalores do operador densidade \hat{D} (RIOUL, 2018).

A entropia de um sistema quântico também pode ser calculada utilizando a teoria de Shannon (NAZARENO; BRITO, 1999). A partir da equação de Schrödinger dependente do tempo

$$i\hbar \frac{d|\psi\rangle}{dt} = H|\psi\rangle, \quad (4.6)$$

o professor Nazareno demonstra que as *amplitudes de Wannier*, também dependentes do tempo $f_n(t) = \langle n|\psi\rangle$, satisfazem o conjunto de equações dado por

$$i\hbar \frac{df_n}{dt} = \varepsilon_n f_n + \sum_{r>0} \left(\frac{V}{r^\alpha} \right) (f_{n+r} + f_{n-r}). \quad (4.7)$$

O termo ε_n determina as energias locais, e o parâmetro α o alcance da interação entre diferentes pontos da rede, que é o objeto de estudo do artigo. O potencial V corresponde a um termo de perturbação. Sendo assim, a entropia nesse caso é obtida por

$$S(t) = - \sum |f_n(t)|^2 \ln |f_n(t)|^2. \quad (4.8)$$

Aqui, a função entropia pode assumir dois valores limite, $S = 0$ para um estado completamente localizado e $S = \ln N$ para um estado uniformemente distribuído.

A analogia formal que se estabelece entre as formulas é realmente intrigante e, embora a teoria de Shannon seja mais matemática do que física, os resultados obtidos a partir dela podem ser úteis quando aplicados à física com a devida interpretação (RIOUL, 2018). Os modelos propostos por Shannon enxergam a fonte de informação como dispositivo probabilístico que escolhe, dentre as mensagens possíveis, aquela que será emitida ao destinatário. No caso de já se ter a mensagem transmitida, a probabilidade expressa a expectativa, ou a hipótese, dessa mensagem ter sido realmente transmitida ou não.

Além da entropia calcular a quantidade média de informação que comunica uma mensagem, nota-se que o conceito de entropia derivado por Shannon está, na verdade, altamente relacionado com a entropia na mecânica estatística (ZUBAREV, 1974). De acordo com Zubarev, a entropia informacional é a medida de incerteza na informação correspondente à distribuição estatística. E de fato, analisando a formula de entropia de Shannon 3.2, H é um valor nulo quando a probabilidade de um evento p_k é 1 e o restante das probabilidades p_k forem zero, isto é, quando o resultado do experimento é previsível e não há incertezas. A entropia H obterá seu maior valor quando os eventos forem equiprováveis, ou seja, quando $p_k = 1/n$. Isso significa que quanto maior a incerteza, menos “informação” e maior a entropia H obtida.

A teoria ergódica, desenvolvida inicialmente como meio de quantificar a trajetória de sistemas físicos, foi utilizada por Shannon como ferramenta para caracterizar o desempenho ideal da fonte de informação ao comunicar uma mensagem através de códigos por um canal aleatório restrito (GRAY, 2011). No caso simples do processo de comunicação sem memória, onde a distribuição de probabilidade não se altera com o tempo e cujos códigos são definidos a partir de um conjunto finito, o teorema ergódico fornece os resultados necessários relativos ao comportamento assintótico da informação, que foram de grande importância para o desenvolvimento dos teoremas da codificação de Shannon. Ele também observa que as ideias básicas se estendem de maneira direta à fontes de comunicação mais complicadas como a fonte de Markov, ou cadeia de Markov, que é o caso mais geral de um processo estocástico em que as probabilidades dependem dos eventos anteriores. Ainda assim, o autor alega que essas generalizações estão longes de suprir a ideia geral das fontes estacionárias consideradas no teorema ergódico.

O próprio Weaver, na introdução do trabalho de Shannon, afirma que “a quan-

tidade que atende exclusivamente aos requisitos naturais que alguém estabelece para ‘informação’ acaba sendo exatamente aquela que é conhecida na termodinâmica como entropia” (SHANNON, 1948). Segundo ele, a ideia básica herdada da termodinâmica relaciona entropia ao grau de aleatoriedade, a tendência de sistemas físicos de se tornarem cada vez menos organizados e cada vez mais perfeitamente embaralhados; enquanto que na teoria da comunicação, a informação dada pela medida da entropia está associada à quantidade de liberdade de escolha que se tem ao construir uma mensagem.

É natural, e bastante curioso, que se estabeleça uma conexão direta entre os dois pontos de vista. Mas apesar da incrível compatibilidade, há quem acredite que esta não passe de uma mera coincidência. De fato, analogia entre as entropias se quebra pois o conceito de Shannon é derivado de uma propriedade lógica, e não dinâmica (DEACON, 2011). Segundo o autor, um exemplo que traduz essa contraposição é o fato de que a entropia de Shannon não aumenta espontaneamente na maioria dos sistemas de comunicação e, portanto, não há equivalente da segunda lei da termodinâmica quando se trata de entropia informacional.

Parte II

Medida da informação

5 Como medir informação

Ainda que o conceito de informação tenha diferentes significados em diferentes contextos, dentro da ciência da computação a informação é tudo aquilo que conseguimos medir através de dados. A partir da medida, análise e coleta de dados, seja de gráficos, tabelas ou imagens, é possível obter sequências de valores cujas variáveis quantitativas e qualitativas são geralmente dadas por representações numéricas que exprimem a quantidade de informação ali envolvida. Logo, se a quantidade de informação é adquirida a partir da interpretação de sinais disponíveis num fluxo de dados, é necessário primeiro entender como são armazenados os dados dentro da memória de um computador para, só então, compreender os algoritmos de compressão destes.

5.1 Variáveis inteiras

Se queremos medir a quantidade de informação existente em sequências de sinais gerados por uma fonte de codificação é necessário saber não só o tipo de dado com que estamos lidando, como a notação utilizada para armazená-los.

A unidade mais fundamental da memória de um computador é denominada por *bit*. Diferente da *unidade binária* utilizada para medir a quantidade de informação prevista pela entropia de Shannon, o *dígito binário* aqui se refere à menor parcela de informação armazenada ou transmitida por um computador, podendo ele ter apenas um de dois valores possíveis: “1” para quando o sistema está “ligado” e “0” para quando está “desligado”.

A coleção de 8 bits é chamado de *byte* ou *octeto* e, geralmente, armazena-se um dado qualquer utilizando um ou mais bytes de memória. A coleção de 4 bytes, ou 32 bits, é conhecida como *word* (ou *palavra*) e essa unidade já é capaz de representar uma informação útil na hora de transferir ou processar algum dado. Nesse sentido, os dados armazenados em notação binária são, em sua menor escala, uma série de conjuntos de “1” e “0” (Figura 4).



Figura 4 – Formato de armazenamento em bits, bytes e word.

Os números inteiros são comumente armazenados utilizando uma palavra de memória, que corresponde a 4 bytes, ou 32 bits de memória¹. Se cada bit possui dois valores possíveis, a combinação de k dígitos binários é capaz de representar até 2^k estados possíveis. Nesse caso, 2^{32} números inteiros podem ser representados utilizando uma palavra de memória na notação binária, de 0 até 4.294.967.295 ($2^{32} - 1$). Mas se o interesse é de incluir também os números negativos, utiliza-se um dos 32 bits para indicar o sinal (positivo ou negativo) do número inteiro, obtendo-se um novo intervalo de valores possíveis, entre $-2.147.483.647$ e $+2.147.483.647$ [$\pm 2^{31} - 1$]. Para números maiores é necessário mais do que 4 bytes de memória.

Além da forma binária existe também a forma octal e hexadecimal, que utilizam uma notação mais compacta de armazenamento. Enquanto o bit possui dois valores possíveis, o octal representa uma notação na qual cada dígito pode assumir o valor de 0 a 7. Como os dígitos octais são facilmente confundidos com os números decimais, os valores numéricos escritos nessa notação geralmente são precedidos por um dígito “0”. No armazenamento de caracteres, uma letra (ou qualquer outro caractere) pode ser armazenada em 1 byte de memória e sua representação pode ser convertida de 8 dígitos binários para 3 dígitos octais (ver em Tabela 1). A letra ‘A’, por exemplo, corresponde ao número decimal 65 na Tabela ASCII e, portanto, pode ser codificada na notação binária e separada em tripletos: 01 000 001. Sua equivalência na forma octal corresponde aos dígitos 101, ou então 0101 para ficar mais claro o fato de serem dígitos octais.

A notação hexadecimal é uma representação mais conveniente. Cada dígito hexadecimal pode ser representado por um valor entre 0 e 16, sendo os valores acima de “9” simbolizados pelas letras de “a” a “f”. Dessa forma, se um único dígito hexadecimal corresponde a 4 bits, um byte de memória pode ser reduzido em apenas 2 dígitos hexadecimais. E, assim como na notação octal, é comum encontrarmos os caracteres 0x ou # precedendo os dígitos hexadecimais para diferenciá-los. Tomando a letra ‘A’ como exemplo novamente, nota-se que sua notação binária em quadrupletos, 0100 0001, se converte nos dígitos hexadecimais 41, que também pode ser expresso por 0x41 ou #41.

5.2 Variáveis de ponto flutuante

O tipo de dado informa ao compilador como ele deve interpretar uma sequência de bits, define as operações que podem ser feitas e delimita os conjuntos de valores que uma expressão, seja variável ou função, pode atingir. Entre os tipos de dados numéricos mais comuns estão os tipos: inteiro, real, cadeia de caracteres e tipo lógico ou booleano.

¹ O tamanho de uma palavra é específico da CPU e, portanto, depende da arquitetura da máquina e do sistema operacional. Muitos tamanhos de palavras diferentes foram usados, incluindo 6-, 8-, 12-, 16-, 18-, 24-, 32-, 36-, 39-, 40-, 48-, 60- e 64-bit. Atualmente quase todos os processadores utilizam palavras de 64-bit. Disponível em <[https://en.wikipedia.org/wiki/Integer_\(computer_science\)](https://en.wikipedia.org/wiki/Integer_(computer_science))>.

ASCII Table

| Dec | Hex | Oct | Char | Dec | Hex | Oct | Char | Dec | Hex | Oct | Char | Dec | Hex | Oct | Char |
|-----|-----|-----|------|-----|-----|-----|---------|-----|-----|-----|------|-----|-----|-----|------|
| 0 | 0 | 0 | | 32 | 20 | 40 | [space] | 64 | 40 | 100 | @ | 96 | 60 | 140 | ` |
| 1 | 1 | 1 | | 33 | 21 | 41 | ! | 65 | 41 | 101 | A | 97 | 61 | 141 | a |
| 2 | 2 | 2 | | 34 | 22 | 42 | " | 66 | 42 | 102 | B | 98 | 62 | 142 | b |
| 3 | 3 | 3 | | 35 | 23 | 43 | # | 67 | 43 | 103 | C | 99 | 63 | 143 | c |
| 4 | 4 | 4 | | 36 | 24 | 44 | \$ | 68 | 44 | 104 | D | 100 | 64 | 144 | d |
| 5 | 5 | 5 | | 37 | 25 | 45 | % | 69 | 45 | 105 | E | 101 | 65 | 145 | e |
| 6 | 6 | 6 | | 38 | 26 | 46 | & | 70 | 46 | 106 | F | 102 | 66 | 146 | f |
| 7 | 7 | 7 | | 39 | 27 | 47 | ' | 71 | 47 | 107 | G | 103 | 67 | 147 | g |
| 8 | 8 | 10 | | 40 | 28 | 50 | (| 72 | 48 | 110 | H | 104 | 68 | 150 | h |
| 9 | 9 | 11 | | 41 | 29 | 51 |) | 73 | 49 | 111 | I | 105 | 69 | 151 | i |
| 10 | A | 12 | | 42 | 2A | 52 | * | 74 | 4A | 112 | J | 106 | 6A | 152 | j |
| 11 | B | 13 | | 43 | 2B | 53 | + | 75 | 4B | 113 | K | 107 | 6B | 153 | k |
| 12 | C | 14 | | 44 | 2C | 54 | , | 76 | 4C | 114 | L | 108 | 6C | 154 | l |
| 13 | D | 15 | | 45 | 2D | 55 | - | 77 | 4D | 115 | M | 109 | 6D | 155 | m |
| 14 | E | 16 | | 46 | 2E | 56 | . | 78 | 4E | 116 | N | 110 | 6E | 156 | n |
| 15 | F | 17 | | 47 | 2F | 57 | / | 79 | 4F | 117 | O | 111 | 6F | 157 | o |
| 16 | 10 | 20 | | 48 | 30 | 60 | 0 | 80 | 50 | 120 | P | 112 | 70 | 160 | p |
| 17 | 11 | 21 | | 49 | 31 | 61 | 1 | 81 | 51 | 121 | Q | 113 | 71 | 161 | q |
| 18 | 12 | 22 | | 50 | 32 | 62 | 2 | 82 | 52 | 122 | R | 114 | 72 | 162 | r |
| 19 | 13 | 23 | | 51 | 33 | 63 | 3 | 83 | 53 | 123 | S | 115 | 73 | 163 | s |
| 20 | 14 | 24 | | 52 | 34 | 64 | 4 | 84 | 54 | 124 | T | 116 | 74 | 164 | t |
| 21 | 15 | 25 | | 53 | 35 | 65 | 5 | 85 | 55 | 125 | U | 117 | 75 | 165 | u |
| 22 | 16 | 26 | | 54 | 36 | 66 | 6 | 86 | 56 | 126 | V | 118 | 76 | 166 | v |
| 23 | 17 | 27 | | 55 | 37 | 67 | 7 | 87 | 57 | 127 | W | 119 | 77 | 167 | w |
| 24 | 18 | 30 | | 56 | 38 | 70 | 8 | 88 | 58 | 130 | X | 120 | 78 | 170 | x |
| 25 | 19 | 31 | | 57 | 39 | 71 | 9 | 89 | 59 | 131 | Y | 121 | 79 | 171 | y |
| 26 | 1A | 32 | | 58 | 3A | 72 | : | 90 | 5A | 132 | Z | 122 | 7A | 172 | z |
| 27 | 1B | 33 | | 59 | 3B | 73 | ; | 91 | 5B | 133 | [| 123 | 7B | 173 | { |
| 28 | 1C | 34 | | 60 | 3C | 74 | < | 92 | 5C | 134 | \ | 124 | 7C | 174 | |
| 29 | 1D | 35 | | 61 | 3D | 75 | = | 93 | 5D | 135 |] | 125 | 7D | 175 | } |
| 30 | 1E | 36 | | 62 | 3E | 76 | > | 94 | 5E | 136 | ^ | 126 | 7E | 176 | ~ |
| 31 | 1F | 37 | | 63 | 3F | 77 | ? | 95 | 5F | 137 | _ | 127 | 7F | 177 | |

Tabela 1 – Tabela ASCII - Código Padrão Americano para o Intercâmbio de Informação. Disponível em <<https://www.asciitable.com/>>

A diferença entre eles é não só com relação à quantidade de bits utilizados, mas também no aspecto mais fundamental de como o número armazenado é representado: em formato de ponto fixo ou ponto flutuante (PRESS et al., 2007).

Os computadores armazenam os números reais com uma certa aproximação e não com um precisão infinita devido à parte fracionária. Diferente dos números inteiros, em que k dígitos conseguem representar 2^k diferentes estados - sendo '0' o primeiro, '1' o segundo, '2' o terceiro e assim por diante - os números reais não possuem uma ordem na qual podem ser expressos. Entre 0.1 e 0.2 existem os números 0.01, 0.001, 0.00001 e infinitos outros e, por essa razão, a precisão com que se consegue representar esses números é uma das dificuldades enfrentadas na hora de armazená-los. No entanto, com 8 bytes de memória é possível armazenar a distância exata da Terra até a Lua com precisão milimétrica², por exemplo. Dessa forma, a limitação na precisão numérica não é considerada um problema exatamente, já que não é comum encontrarmos medidas científicas com esse grau de acurácia.

Existem duas formas de representar os números fracionários, utilizando a notação de ponto fixo ou a de ponto flutuante. O formato de ponto fixo permite uma representação com precisão igual ao número de bits disponíveis para a parte fracionária, limitando a quantidade e precisão dos números reais que podem ser armazenados. Ou seja, a precisão

² Argumento retirado de <<https://statmath.wu.ac.at/courses/data-analysis/itdtHTML/node55.html>>.

é sempre fixa e não depende da magnitude da parte inteira do número representado. A notação de ponto flutuante possui uma faixa de representação bem maior. Esse formato é utilizado para representar números muito grandes ou muito pequenos, ou que requerem um alto grau de precisão e, por isso, exigem mais memória para armazenamento.

Os números representados utilizando a notação de ponto flutuante são divididos em mantissa (M), expoente (E) e base β , na forma $M \times \beta^e$. Similar ao padrão de notação científica, a representação dos números reais ($r \in \mathbb{R}$), na sua forma normalizada, é dada por

$$r = 0, d_1 d_2 d_3 \dots d_t \times \beta^e, \quad (5.1)$$

onde o primeiro algarismo após a vírgula é diferente de zero $d_1 \neq 0$ e $0 \leq d_j < \beta$; t é a precisão, que determina o número de dígitos na mantissa; β é a base de numeração³ e o expoente e determina a ordem de grandeza do número.

A estratégia de armazenamento mais usada pelos computadores modernos é a IEEE 754, criada em 1985. O Instituto de engenheiros elétricos e eletrônicos (IEEE) padronizou o formato para valores de ponto flutuante, uma maneira de representar números reais no hardware. Normalmente, os formatos usados pelo compilador são de *precisão simples* (4 bytes) e de *precisão dupla* (8 bytes), que são declaradas utilizando a palavra-chave **float** e **double**, respectivamente. Além destes, o padrão IEEE também especifica formatos de meia precisão (2 bytes), de precisão quádrupla (16 bytes) e um formato de precisão estendida dupla (10 bytes).

Embora o formato de ponto flutuante varie de máquina pra máquina no que diz respeito a ordem que os bits estão disponibilizados, o padrão IEEE-754 estabelece a estrutura de representação desses bits, como é possível ver na figura 5 para os casos de precisão simples e precisão dupla. Assim, para valores com precisão simples, a sequência de 32 bits é organizada da seguinte forma: o primeiro bit é reservado para o sinal do número armazenado, os 8 bits seguintes indicam o valor do expoente e os 23 bits restantes determinam a mantissa. Os valores com precisão dupla incluem um arranjo de 64 bits, em que 11 deles são destinados ao expoente, 53 à mantissa e 1 bit para o sinal.

Seguindo essa estrutura, o bit do sinal é o bit mais significativo, indicando “0” para números positivos e “1” para negativos. O expoente define o intervalo de representação, indicando a gama de valores possíveis. E a mantissa, ou significante, contém o binário equivalente aos algarismos significativos do número que vamos armazenar, isto é, a mantissa estabelece a precisão com que os valores podem ser representados. Com 23 bits é possível representar 2^{23} números reais diferentes; ainda assim existirá lacunas entre cada valor. Se lidamos com um intervalo de 0 a 10.000.000, por exemplo, só poderemos dar passos de $\frac{10.000.000}{2^{23}} \approx 1$, e não será possível distinguir valores que diferem um do outro por menos de 1.

³ Geralmente na base binária $\beta = 2$, mas pode ser na base decimal ou hexadecimal, $\beta = 10, 16$.

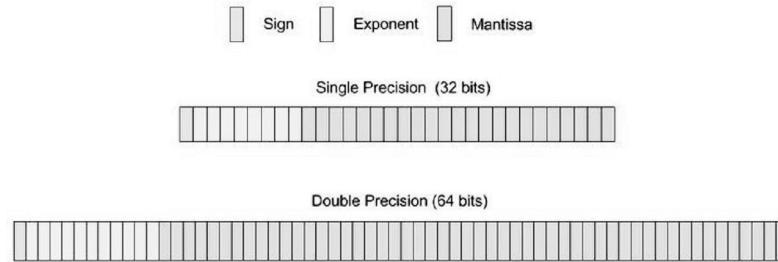


Figura 5 – Estrutura de ponto flutuante de acordo com a IEEE-754.

A representação numérica dos valores de ponto flutuante pode ser obtida a partir da expressão

$$S \times M \times b^{E-e}. \quad (5.2)$$

onde S é o bit referente ao sinal, M o conjunto de bits referente à mantissa e $b = 2$ a base da notação; E representa o expoente e e o “viés” do expoente, uma constante de número inteiro fixo para qualquer máquina (PRESS et al., 2007). Para valores flutuantes de precisão única, o viés vale $e = 127$, e para valores com precisão dupla $e = 1023$.

Os expoentes armazenados em 8 bits podem expressar um valor de 0 a 254, variando somente de -127 a 127. O valor 2^{127} equivale a aproximadamente 10^{38} , que é o limite real de precisão simples. Isso implica dizer que é possível armazenar números tão pequenos quanto 10^{-39} a números tão grandes quanto 10^{38} . Assim, depois de deduzir o valor do expoente, faz-se a subtração do viés para encontrar o valor real do expoente. Se o valor armazenado for menor que o viés, o expoente será, na verdade, um valor negativo e o ponto binário flutuará para a esquerda; se for maior, o expoente é positivo e o ponto binário flutuará para a direita.

A mantissa é armazenada como uma fração binária do formato 1.XXXXX, ou $1.F$, possuindo um valor maior ou igual a 1 e menor que 2. A fração F consiste no conjunto de bits da mantissa que são realmente armazenados (PRESS et al., 2007). Em outras palavras, a mantissa possui um valor implícito (não armazenado) igual a 1 e a ele somam-se as potências 2^{-j} , referente ao bit de posição j .

Um exemplo de como armazenar os números reais de 1.0 a 5.0 utilizando 4 bytes de memória é demonstrado na Tabela 2. Nesse formato, o primeiro bit é referente ao bit do sinal (S), os 8 bits seguintes são referentes ao expoente (E) e os 23 bits restantes são destinados à mantissa (M).

Como todos os números são positivos, o primeiro bit de cada número é 0. O expoente do número ‘1’ na primeira linha é dado por 0111111 1, que é 127 na notação decimal. Subtraindo o viés para obter o expoente real tem-se $E - e = 0$. A mantissa

| | | | | |
|----------|----------|----------|----------|---|
| 00111111 | 10000000 | 00000000 | 00000000 | 1 |
| 01000000 | 00000000 | 00000000 | 00000000 | 2 |
| 01000000 | 01000000 | 00000000 | 00000000 | 3 |
| 01000000 | 10000000 | 00000000 | 00000000 | 4 |
| 01000000 | 10100000 | 00000000 | 00000000 | 5 |

Tabela 2 – Números reais de 1.0 a 5.0 armazenados no formato de ponto flutuante com precisão simples.

é representada por 23 bits iguais a 0, restando apenas o seu valor implícito “1”. Assim, seguindo a relação 5.2, temos $2^0 \times 1 = 1$. Já o expoente do número ‘5’ é dado por 10000001, que equivale a 129, e $E - e = 2$. A mantissa é obtida a partir da soma do número implícito “1” com a parte fracionária dada pelos bits 01 (seguido de 21 bits iguais a 0), ou seja, $1 + 0 \times 2^{-1} + 1 \times 2^{-2} = 1.25$. Portanto, fazendo a multiplicação da base pela mantissa, temos $1.25 \times 2^2 = 5$.

A aritmética entre números na representação de ponto flutuante não é exata todavia. Mesmo aqueles que possuem uma representação exata de acordo com a expressão 5.2 terão sua precisão alterada quando somada ou subtraída de algum outro valor (PRESS et al., 2007). O menor valor que, somado ao número de ponto flutuante 1.0, resulta em um valor diferente de 1.0 determina a *acurácia da máquina* ϵ . Conhecido como *erro de arredondamento*, esse termo revela a precisão fracionária com a qual os números de ponto flutuante são representados, correspondendo a uma mudança de um bit menos significativo da mantissa. Esse tipo de erro deve ser adicionado a qualquer operação realizada entre números de ponto flutuante, podendo acumular com o aumento da quantidade de cálculos.

5.3 Algoritmos de compressão

Um dos objetivos deste trabalho envolve estabelecer uma relação entre a entropia de um sistema dinâmico molecular com a quantidade de informação que se tem sobre ele, ou melhor dizendo, com a quantidade de incerteza sobre o sistema.

A quantidade de informação medida pela expressão de entropia, na teoria de Shannon, está associada à quantidade de liberdade de escolha que se tem ao construir mensagens (SHANNON, 1948). A razão entre a entropia medida e a entropia máxima de uma fonte de informação é chamada de *entropia relativa* da fonte. De acordo com Weaver, se uma fonte possui entropia relativa de 0.8 significa que essa fonte tem aproximadamente 80% de liberdade de escolha ao formar uma mensagem, sendo os outros 20% a fração restante que é governada por leis estatísticas que determinam parte da estrutura da mensagem. Isso significa que parte dessa mensagem é desnecessária para fins de compreensão e é chamada de *redundância*.

Considerando que toda mensagem possui uma parcela redundante, que se repete, não há necessidade de contabilizá-la mais de uma vez. Partindo desse pressuposto, é possível reduzir a quantidade de memória destinada ao armazenamento dessa informação. No primeiro teorema da codificação, elucidado na seção 3.5.1, a entropia de Shannon recebe um outro significado operacional: é a taxa mínima de bits informativos na qual uma fonte de informação pode ser codificada de forma segura, isto é, a mensagem ainda pode ser compreendida (RIOUL, 2018). O teorema da codificação não só fornece o melhor desempenho possível de qualquer algoritmo de compressão de dados como também busca reduzir o tempo de transmissão destes.

Utilizaremos os resultados obtidos por Shannon para reduzir a quantidade de bits necessários para armazenar informação referente à posição de partículas indistinguíveis em um sistema molecular. A implementação algorítmica de Shannon utiliza o princípio de compressão de dados sem perda de ‘informação’, apenas restringindo as redundâncias. A modelagem de dados, ou seja, a caracterização prévia na fonte de dados, consiste em associar códigos maiores a símbolos menos prováveis e códigos menores a símbolos mais prováveis. Assim, a entropia informacional confere as probabilidades obtidas na hora de selecionar e compor uma mensagem. Portanto, o objetivo final desse trabalho consiste em verificar a hipótese de Shannon com relação à compressão de dados.

6 Informação de uma partícula

Depois de compreender o padrão de contagem e armazenamento de dados, seguiremos para a próxima etapa deste trabalho que consiste em medir a quantidade de informação contida primeiramente por uma partícula e, em seguida, por um sistema com mais partículas.

Designaremos Λ como a quantidade de informação referente ao número mínimo de bits necessário para armazenar a posição e a velocidade de todas as partículas do sistema. Para isso, assumiremos a existência de um algoritmo de compressão que se beneficie das simetrias, correlações, periodicidades e indistinguibilidade das partículas. A definição de Λ acima será aplicada depois que o algoritmo de um processo for aplicado às variáveis do sistema.

Conhecer as coordenadas das partículas com precisão infinita não é necessário nem possível para a simulação numérica da sua dinâmica. A cada medida estará associada uma margem de erro, e designaremos por δ^r e δ^v a incerteza admissível para a posição e a velocidade das partículas que compõem o sistema dinâmico ao longo de cada dimensão do espaço de configuração. O espaço de cada coordenada será discretizado em intervalos de comprimento igual à incerteza, sendo o intervalo ocupado por cada partícula representado por coeficientes que indicarão sua posição e a velocidade. Ao longo de cada dimensão j , representaremos esses coeficientes por θ^j e ω^j , respectivamente, para a posição e velocidade da partícula, ou seja, $r^j = \delta^r \theta^j$ e $v^j = \delta^v \omega^j$. Reversamente podemos escrever

$$\theta^j = \text{int} \left(\frac{r^j}{\delta^r} \right), \quad (6.1a)$$

$$\omega^j = \text{int} \left(\frac{v^j}{\delta^v} \right), \quad (6.1b)$$

onde a função $\text{int}(x)$ retorna o maior número inteiro menor que x .

As informações relativas às partículas serão armazenadas como números inteiros, ξ , compostos por λ bits, ou seja ξ pode representar θ^j ou ω^j . Utilizaremos um formato de armazenamento de precisão arbitrária, sendo o número de bits necessários para armazenar o número ξ dado pela função $\lambda(\xi)$. Para valores de ξ suficientemente grandes ($\lambda \gg 1$), assumiremos que

$$\lambda(\xi) \approx \log_2 |\xi|, \quad (6.2)$$

onde \log_2 é o logaritmo na base 2, que nos retorna a unidade binária utilizada para medir informação.

Como o objetivo é estender as medidas para sistemas com N partículas, distinguiremos a quantidade de informação referente à posição e velocidade de cada partícula pelo índice i , em que $i = 1, 2, \dots, N$. O número de bits necessários para armazenar as coordenadas j da posição e da velocidade da partícula i são, respectivamente

$$\lambda_i^{rj} = \lambda(\theta^j), \quad (6.3a)$$

$$\lambda_i^{vj} = \lambda(\omega^j). \quad (6.3b)$$

Portanto, o número de bits necessários para armazenar o estado da partícula i é dado pela soma das informações de posição e velocidade sobre cada dimensão j ,

$$\lambda_i = \sum_{j=1}^D \lambda_i^{rj} + \lambda_i^{vj}. \quad (6.4)$$

Vale ressaltar que, embora seja possível medir a quantidade de informação referente ao estado de uma partícula, a relação inversa não é. Ao somar as informações da posição e velocidade de uma partícula, obtemos uma quantidade de informação carregada por ela, mas o contrário não acontece. Medir a quantidade de informação de uma partícula não nos comunica em detalhes o estado em que ela se encontra. Isto porque existem inúmeras combinações possíveis para uma mesma quantidade de informação.

7 Informação de um sistema de partículas

Em sistemas computacionais de dinâmica molecular, as posições e velocidades das partículas são armazenadas usando um número fixo de bits (λ), e a quantidade de memória utilizada corresponde à quantidade de bits por partícula multiplicada pelo número de partículas, $n\lambda$ bits. Os valores que podem ser representados pelo modelo de armazenamento utilizado definem o espaço de configurações permitidas pela implementação numérica.

Algumas propriedades da dinâmica molecular limitam a região acessível pelo sistema a um subespaço do espaço permitido pela implementação computacional. Outras propriedades fazem com que o subespaço acessível não seja ocupado de forma uniforme. Esses dois processos podem ser utilizados na criação de algoritmos de compressão que reduzem a quantidade de bits necessários para armazenar a informação de um número grande de partículas.

Em boa parte dos casos, a economia de memória obtida na aplicação desses algoritmos não se justifica para sistemas de dinâmica molecular em vista do processamento envolvido na execução destes. A importância dos algoritmos reside na quantidade mínima de dados necessários para armazenar uma informação. O teorema da codificação de Shannon demonstra a implementação de um algoritmo que produza a maior compressão possível na memória utilizada, mesmo que pra isso seja necessário um alto custo operacional.

7.1 Partículas distinguíveis não correlacionadas

Caso as N partículas que compõem o sistema sejam distinguíveis e não exista correlação entre as variáveis que armazenam as posições e as velocidades, então a quantidade de bits necessários para armazenar a configuração do sistema é igual à soma do número de bits por partícula:

$$\Lambda = \sum_i^N \sum_j^D \lambda_i^{r_j} + \lambda_i^{v_j}. \quad (7.1)$$

A probabilidade de cada configuração possível dentro de um sistema pode ser escrita como o produto das densidades de probabilidades das componentes j da posição e da velocidade, $\rho^{r_j}(r)$ e $\rho^{v_j}(v)$, respectivamente. As funções distribuição de probabilidade satisfazem as seguintes propriedades:

$$\rho(x) \geq 0, \quad \forall x \in R_x, \quad (7.2a)$$

$$\int_{-\infty}^{+\infty} \rho(x) dx = 1. \quad (7.2b)$$

Embora o espaço de cada coordenada seja discretizado em pequenos intervalos (δ), assumiremos que ξ seja uma variável real, para simplificação dos cálculos. A função $\lambda(\xi)$ possui uma distribuição contínua de valores possíveis dentro de um dado intervalo. Para cada uma das coordenadas do subespaço, é possível escrever a quantidade de bits média por variável, correspondendo ao valor esperado de informação contida em cada uma delas. Assim, considerando que o espaço da posição tem comprimento L (em cada dimensão j) e que a velocidade de cada partícula pode variar entre $-\infty$ a $+\infty$, o valor esperado de informação para cada variável de cada partícula é dado por:

$$\langle \lambda_i^{rj} \rangle = \int_0^L \lambda \left(\frac{r^j}{\delta^r} \right) \rho^{rj}(r) dr, \quad (7.3a)$$

$$\langle \lambda_i^{vj} \rangle = \int_{-\infty}^{+\infty} \lambda \left(\frac{v^j}{\delta^v} \right) \rho^{vj}(v) dv. \quad (7.3b)$$

A função contínua λ depende de grandezas reais r/δ^r e v/δ^v . Essa substituição da variável inteira ξ por uma variável contínua é uma aproximação válida para θ^r e ω^r muito grandes.

7.2 Partículas distinguíveis em equilíbrio termodinâmico

No equilíbrio termodinâmico e na ausência de um campo gravitacional, a posição das partículas ao longo de cada uma das D dimensões se distribui uniformemente no intervalo $[0, L]$, ou seja,

$$\rho^{rj}(r) = \frac{1}{L}. \quad (7.4)$$

Associada à posição r^j , a variável inteira θ^j possui valores no intervalo $[0, L^\theta]$, onde $L^\theta = \text{int}(L/\delta^r)$; cada um dos valores inteiros ocorrendo com probabilidade $1/L^\theta$.

A velocidade das partículas, na condição de equilíbrio térmico, obedece à distribuição de Maxwell-Boltzmann,

$$\rho^{vj}(v) = \sqrt{\frac{m}{2\pi kT}} e^{-mv^2/2kT}. \quad (7.5)$$

Por conseguinte, ao aplicar as funções das equações 6.2 e 7.4 na equação 7.3a, adquirimos o número médio de bits necessários para armazenar cada coordenada j do vetor posição de cada partícula em equilíbrio. Tendo em vista que a relação logarítmica em 6.2 não é definida para $\xi = 0$, utilizaremos o valor 1 como limite inferior da integral. Assim, fazendo as devidas substituições obtemos

$$\langle \lambda_i^{rj} \rangle = \frac{\delta^r}{L} \int_1^{L/\delta^r} \log_2(\theta^j) d\theta^j \approx \log_2 \frac{L}{\delta^r} - \frac{1}{\ln 2}, \quad (7.6)$$

onde desprezamos o termo δ^r/L , que se anula no limite termodinâmico.

Logo, o número de bits necessário para armazenar todas as coordenadas da posição de uma partícula é igual a

$$\langle \lambda_i^r \rangle = \sum_{j=1}^D \langle \lambda_i^{rj} \rangle \approx D \left(\log_2 \frac{L}{\delta^r} - \frac{1}{\ln 2} \right) = \log_2 \frac{V}{V_0} - C_D \text{ bits.} \quad (7.7)$$

Portanto, o número médio de bits necessário para armazenar a posição de N partículas distinguíveis é dado pela soma dos bits referentes à posição de cada partícula individualmente:

$$\langle \Lambda^r \rangle = \sum_i^N \langle \lambda_i^r \rangle \approx N \log_2 \frac{V}{V_0} - NC_D, \quad (7.8)$$

onde $V = L^D$ representa o volume do espaço de configuração que contém o sistema e $V_0 = (\delta^r)^D$ o volume de um único cubículo da caixa, e $C_D = D \log_2 e$.

7.3 Partículas indistinguíveis

Nesta seção, apresentamos um algoritmo que permite reduzir o número de bits necessários para armazenar a posição de partículas indistinguíveis em equilíbrio termodinâmico, sem alterar a quantidade de bits que armazenam as velocidades das partículas, uma vez que elas estão associadas invariavelmente a cada partícula.

O algoritmo aqui apresentado consiste em encontrar um ordenamento entre as partículas de tal forma a maximizar a informação mútua (ver em anexo A.3) entre a posição de duas partículas subsequentes. Dessa forma, ao se subtrair duas posições de partículas subsequentes elimina-se a redundância e um número menor de bits é necessário para armazenar a informação.

O primeiro passo é dividir o espaço em volumes que contém, em média, uma partícula cada. Uma forma de fazer essa divisão é utilizar caixas de lado igual a

$$\Delta^r = \frac{L^r}{N^{1/D}} \quad (7.9)$$

A figura 6 ilustra a divisão do espaço em pequenos quadrados de dimensão δ^r usadas para definir a posição das partículas, e agrupadas em caixas de dimensão Δ^r , usados no algoritmo de compressão.

O próximo passo é impor uma “reorganização” para o sistema, onde ordena-se o mesmo de tal forma que as primeiras partículas são as que ocupam a primeira caixa no canto inferior da primeira linha, seguidas pelas partículas nos cubículos à sua direita, sucedendo-se às caixas à direita até o final dessa linha. As partículas seguintes na ordenação pertencem à próxima linha, de baixo pra cima, ocupando sempre da esquerda pra direita; o procedimento é repetido até que todas as linhas sejam percorridas. Quando

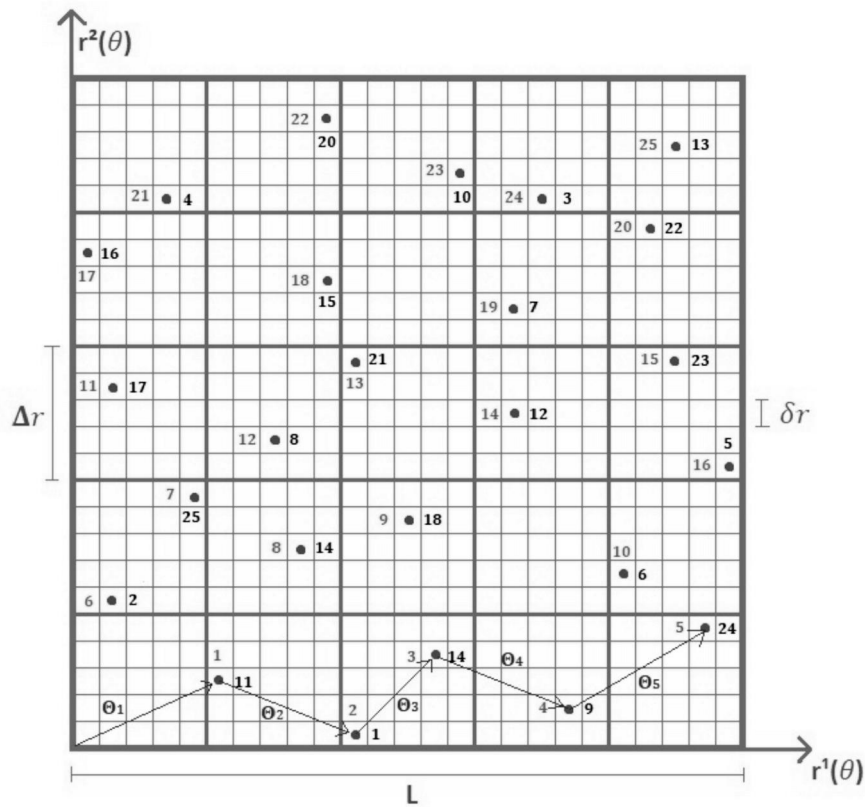


Figura 6 – Em um sistema bidimensional, o espaço quadrado de comprimento L é dividido em cubículos de comprimento δr , utilizadas para definir as posições das partículas com a precisão necessária para a evolução numérica da dinâmica. As caixas de dimensão Δr são utilizadas pelo algoritmo de compressão. Os índices originais i são mostrados em preto. Os índices α definidos pelo algoritmo de ordenação descritos em 7.10 são mostrados em vermelho.

| | | | | | | | | | | | | | | | | | | |
|----------|----|---|----|---|----|---|----|----|----|----|----|----|----|----|----|----|----|----|
| α | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 |
| i | 11 | 1 | 14 | 9 | 24 | 2 | 25 | 14 | 18 | 6 | 17 | 8 | 21 | 12 | 23 | 5 | 16 | 15 |

Tabela 3 – Correspondência entre os índices i e α das partículas contidas na caixa de dimensão L , mostradas na Fig.6.

uma caixa possui mais de uma partícula, elas são ordenadas da esquerda para direita, sorteando-se o ordenamento das partículas que possuam a mesma coordenada horizontal. Cada partícula recebe um índice α que segue essa ordenação e que estabelece uma relação biunívoca com o índice original, $i(\alpha)$. A Tabela 3 mostra a relação entre os índices i e α da figura 6. Aplicamos o mesmo reordenamento às velocidades das partículas, de forma que as velocidades continuem relacionadas às mesmas partículas.

Em seguida, ao definir a posição de cada partículas em relação à posição da partícula anterior, determinaremos o índice das posições relativas pelo índice α ,

$$\Theta_{\alpha}^j = \theta_{i(\alpha)}^j - \theta_{i(\alpha-1)}^j. \quad (7.10)$$

As coordenadas da primeira partícula de cada linha são definidas em relação ao canto inferior esquerdo da primeira caixa da linha. Ressaltamos que a transformação $\theta_i^j \rightarrow \Theta_\alpha^j$ definida pela equação 7.10 não é inversível no que se refere à identificação das partículas, ou seja, a partir do conjunto dos valores Θ_α^j é possível obter os valores de θ_α^j e ω_α^j , porém o ordenamento original é perdido. Observamos que, uma vez que as variáveis Θ e θ são inteiras, não há perda de informação por arredondamento na transformação definida pelo algoritmo 7.10.

Cada linha possui, em média, $N^{1/D}$ partículas, e suas coordenadas $\theta_{i(\alpha)}^1$ são uniformemente distribuídas dentro do intervalo $[0, L^\theta]$. Essa uniformidade resulta em uma distribuição para a coordenada Θ_α^1 equivalente a um processo de Poisson com frequência $\nu = N^{1/D}/L^\theta = \delta^r/\Delta^r = 1/\Delta^\theta$. Logo, os valores de Θ^1 obedecem à distribuição exponencial normalizada

$$\rho^{\Theta^1}(\Theta) = \nu e^{-\nu\Theta}, \quad (7.11)$$

com $0 \leq \Theta \leq \infty$. A densidade de probabilidade acima corresponde a aproximar a variável inteira Θ por uma grandeza real. Ao substituirmos essa expressão na integral 7.3a, com $r \rightarrow \delta^r\Theta$, precisamos lembrar que a aproximação 6.2 não se aplica em $\xi = 0$, e por isso, não podemos utilizar esse limite inferior na integral. Usando 1 como o limite da integral, obtemos

$$\begin{aligned} \langle \lambda^{r1} \rangle &= \nu \int_1^\infty \log_2 \Theta e^{-\nu\Theta} d\Theta \\ &= [-\log_2 \Theta e^{-\nu\Theta}]_1^\infty + \frac{1}{\ln 2} \int_1^\infty \frac{1}{\Theta} e^{-\nu\Theta} d\Theta \\ &= \frac{1}{\ln 2} \int_1^\infty \frac{1}{\Theta'} e^{-\Theta'} d\Theta' = \frac{1}{\ln 2} E_1(\nu), \end{aligned} \quad (7.12)$$

sendo E_1 a função exponencial integral. Para $\nu \ll 1$ tomamos apenas os dois primeiros termos da expansão $E_1(\nu) = -\gamma - \ln \nu - \sum_{k=1}^\infty (-\nu)^k/(kk!)$, onde $\gamma = 0,57721\dots$ é a constante de Euler-Mascheroni. Dessa forma podemos escrever

$$\langle \lambda^{r1} \rangle = \log_2 \frac{L}{\delta^r} - \frac{1}{D} \log_2 N - \gamma \log_2 e. \quad (7.13)$$

Para as outras dimensões $j > 1$, a coordenada Θ^j obedece à distribuição triangular

$$\rho^{\Theta^j}(\Theta) = \frac{1}{\Delta^{(\theta)^2}} \begin{cases} 0 & \text{se } \Theta \leq -\Delta^\theta, \\ \Theta + \Delta^\theta & \text{se } -\Delta^\theta < \Theta \leq 0, \\ \Delta^\theta - \Theta & \text{se } 0 < \Theta \leq \Delta^\theta, \\ 0 & \text{se } \Delta^\theta < \Theta. \end{cases} \quad (7.14)$$

Nesse caso, usando a integral 7.3a, podemos escrever

$$\begin{aligned}
 \langle \lambda^{rj} \rangle &= \frac{2}{\Delta^{\theta^2}} \int_1^{\Delta^\theta} (\Delta^\theta - \Theta) \log_2(\Theta) d\Theta & (7.15) \\
 &= \frac{2}{\Delta^{\theta^2} \ln 2} \left[\Theta \Delta^\theta (\ln \Theta - 1) - \frac{\Theta^2}{4} (2 \ln \Theta - 1) \right]_1^{\Delta^\theta} \\
 &= \log_2 \Delta^\theta - \frac{3}{2} \frac{1}{\ln 2} \\
 &= \log_2 \frac{L}{\delta^r} - \frac{1}{D} \log_2 N - \frac{3}{2} \log_2 e,
 \end{aligned}$$

tendo sido desprezados os termos de ordem igual ou superior a $(1/\Delta^\theta)$.

A partir dos resultados 7.13 e 7.15, a quantidade média de bits necessária para armazenar a posição de uma partícula é

$$\langle \lambda^r \rangle = \sum_{j=1}^D \langle \lambda^{rj} \rangle = \log_2 \frac{V}{V_0} - \log_2 N - C_I. \quad (7.16)$$

A constante $C_I = [\gamma + 3(D-1)/2] \log_2 e$ vale 0.8, 3.0 e 5.2, para D respectivamente igual a 1, 2 e 3. A quantidade de bits necessária para armazenar a posição de N partículas indistinguíveis é

$$\langle \Lambda^r \rangle \approx N \log_2 \frac{V}{V_0} - N \log_2 N - NC_I. \quad (7.17)$$

Comparando a quantidade média de bits das posições das partículas distinguíveis, 7.8, com a das partículas indistinguíveis, 7.17, conclui-se que a indistinguibilidade permite economizar $\log_2 N + [\gamma + (D-3)/2] \log_2 e$ bits por partícula. Ou seja, uma quantidade considerável de bits pode ser economizada quando lidamos com números grandes de partículas.

O próximo passo seria somar esse valor à quantidade de bits para armazenar as velocidades das partículas, onde espera-se obter uma expressão equivalente à entropia termodinâmica do gás ideal, mas não foi possível concluir neste trabalho, sendo este um ponto que pode ser abordado em trabalhos futuros.

Conclusão

Através deste trabalho foi possível investigar a evolução histórica do conceito de entropia e as várias definições existentes dentro de cada âmbito, um dos motivos pelos quais sua compreensão se torna tão confusa e complexa. O fato de haver uma mesma expressão matemática para quantificar certos conceitos, como o da informação ou o grau de desordem de um sistema não faz com que seus significados sejam exatamente os mesmos, e nem tampouco se excluam. As diferentes interpretações exploradas nesse trabalho, apesar das particularidades, se complementam e são essenciais para um entendimento mais profundo sobre as propriedades fundamentais da função entropia.

Embora a temperatura seja um conceito definido a partir da derivada parcial da energia interna de um sistema com relação à sua entropia:

$$T = \frac{\partial U_{(S,V,N)}}{\partial S}, \quad (7.18)$$

percebe-se que a definição de entropia é ainda mais fundamental, e independe de outros conceitos como temperatura e energia para sua existência. Pela discussão apresentada na seção 1.3 percebe-se que a entropia é uma grandeza termodinâmica essencial na definição de tempo, no sentido de que só é possível estabelecer uma sucessão temporal de eventos a partir da comparação dos valores de S . A Segunda Lei da termodinâmica afirma que *o aumento de entropia caracteriza os estados finais de um sistema macroscópico que podem ser alcançados a partir de um estado inicial, sem necessariamente deixar uma marca no resto do universo* (LIEB; YNGVASON, 2000). Entende-se por entropia uma função de “transformação” que prevê uma *codificação de possíveis mudanças de estados, $X \prec Y$* .

Ao mesmo tempo, na seção 2.2, é demonstrado a importância do tempo na definição da entropia estatística, no sentido de que *a entropia é uma quantidade que descreve as propriedades médias de um corpo dentro de um intervalo de tempo Δt diferente de zero* (LANDAU; LIFSHITZ, 1958). Ao tentar definir a entropia para sistemas fora do equilíbrio, divide-se o sistema em partes tão pequenas a ponto de existir um tempo de relaxamento $\tau < \Delta t$, onde Δt é um intervalo de tempo observável. Ou seja, não há mudança de estado quando se observa esses subsistemas pois estes já se encontrarão em “equilíbrio interno” e, por isso, Landau afirma não fazer sentido uma medida instantânea da entropia. Dito isso, compreende-se a razão pela qual a entropia é melhor definida em sistemas macroscópicos, onde convenientemente se observam as mudanças de estado.

Vale lembrar que a reversibilidade das leis da Mecânica, do Eletromagnetismo e de outras áreas da Física introduzem uma dificuldade na fundamentação da Segunda Lei da Termodinâmica, uma vez que esta estabelece uma irreversibilidade na dinâmica dos sistemas. De acordo com Lebowitz, a reversibilidade microscópica garante que as

flutuações presentes no estado de equilíbrio sejam tipicamente simétricas em relação aos momentos em que há um mínimo local na função de S_B . No entanto, Boltzmann associa o tempo ao volume do espaço de configuração. Nessa definição, afirma-se que *um sistema macroscopicamente isolado deve se encontrar em um estado macroscópico diferente M , para frações de tempo equivalente à razão de $|\Gamma_M|$ pelo volume total do espaço de fase $|\Gamma|$, possuindo a mesma energia* (LEBOWITZ et al., 1993). Boltzmann percebe ainda que, para sistemas macroscópicos, a fração de microestados que evoluem para macroestados com maior S_B é tão próximo de 1 que esse comportamento é considerado *típico*, atribuindo caráter estatístico à “seta do tempo”. Assim, utiliza-se essa lógica para a definir evolução posterior de um estado inicial (com baixa entropia). Através dos argumentos estatísticos envolvendo “expansão” do volume do espaço de fase, é possível *prever o comportamento futuro dos sistemas macroscópicos - mas não determinar seu passado*.

Contrapondo o argumento acima, relacionamos a entropia à medidas instantâneas de um sistema de partículas distinguíveis e indistinguíveis, em uma região do espaço de configuração de dimensão L^D . Pela equação 6.2, observa-se que a quantidade de bits utilizada para armazenar informação relativa a uma partícula é proporcional ao logaritmo do número de estados acessíveis por ela.

Ao impor um algoritmo de compressão que delimita, de certa forma, os estados acessíveis à partícula, menos bits serão necessários para armazenar esses dados. No contexto da teoria da informação, isso se reflete em uma transmissão mais rápida dos dados. Verifica-se, então, que quanto mais o estado macroscópico do sistema fornece ‘informações’ sobre o arranjo microscópico, menor a incerteza sobre os possíveis valores das variáveis que descrevem os microestados e, portanto, menor a sua entropia. Ou seja, de fato, quanto mais se sabe sobre o sistema, menor a incerteza que se tem sobre ele, e menor sua entropia A.13.

Embora essa afirmação possa ser feita no cenário ideal, acreditamos que essa hipótese possa ser estendida também no mundo real utilizando-se simulações computacionais. Em redes cristalinas como a Rede de Bravais, por exemplo, a disposição das moléculas é determinada de forma muito mais rígida pois sua estrutura possui um alto ordenamento microscópico. Em vista disso, a medida de informação a respeito dessas redes pode ser largamente reduzida se restringirmos as informações redundantes sobre o sistema e levarmos em conta as posições de cada molécula em relação aos seus vizinhos. Em conjunção com o a indistinguibilidade das partículas, esse fato permite reduzir a quantidade de informação necessária para descrever o estado microscópico. Ao abrir mão da identificação individual de cada partícula, descartamos informações irrelevantes para a dinâmica do sistema, diminuindo a quantidade de bits necessária para sua descrição.

Nota-se também uma certa similaridade entre a expressão encontrada em 7.17 e a entropia termodinâmica dos gases ideais 1.20 para um número grande e fixo de partículas.

Para uma melhor especificação do estado macroscópico desse sistema, é necessário ainda calcular a informação referente à velocidade das partículas para se obter a entropia em termos de sua energia total.

Por fim, conclui-se que há uma conexão entre a entropia termodinâmica e a entropia informacional de Shannon, representada aqui pela quantidade média de bits de armazenamento, que se verifica ser proporcional à quantidade de dados relativa às posições das partículas. Além disso, corrobora-se o caráter estatístico da entropia no sentido de que, dada uma distribuição de probabilidade, é possível inferir se o sistema está em um estado de menor ou maior entropia. Isto é, analisando as distribuições 7.4, 7.11 e 7.14 percebe-se que quanto mais acentuada for a curva, ou, quanto mais “concentrada” estiver essa informação, menor a entropia do sistema. E quanto mais uniforme for a distribuição de probabilidade referente às posições das partículas, maior a entropia desse sistema.

Referências

- CALLEN, H. B. *Thermodynamics and an Introduction to Thermostatistics*. [S.l.]: AAPT, 1998. Citado 6 vezes nas páginas 11, 26, 28, 30, 35 e 36.
- CALLENDAR, H. L. The caloric theory of heat and carnot's principle. *Proceedings of the Physical Society of London (1874-1925)*, IOP Publishing, v. 23, n. 1, p. 153, 1910. Citado 2 vezes nas páginas 35 e 65.
- DEACON, T. W. *Incomplete nature: How mind emerged from matter*. [S.l.]: WW Norton & Company, 2011. Citado na página 69.
- FERMI, E. *Thermodynamics*. Dover Publications, 1956. (Dover books in physics and mathematical physics). ISBN 9780486603612. Disponível em: <<https://books.google.com.br/books?id=VEZ1ljsT3IwC>>. Citado na página 25.
- GRAY, R. M. *Entropy and information theory*. [S.l.]: Springer Science & Business Media, 2011. Citado na página 68.
- LANDAU, L.; LIFSHITZ, E. *Statistical Physics (Course of Theoretical Physics vol 5)*. [S.l.]: Pergamon Oxford, 1958. Citado 12 vezes nas páginas 30, 35, 38, 39, 41, 42, 44, 48, 49, 50, 66 e 89.
- LARANJEIRAS, C. C.; PORTELA, S. I. The carnot cycle and the teaching of thermodynamics: a historical approach. *Physics Education*, IOP Publishing, v. 51, n. 5, p. 055013, 2016. Citado na página 35.
- LEBOWITZ, J. L. et al. Boltzmann's entropy and time's arrow. *Physics today*, American Institute of Physics, v. 46, p. 32–32, 1993. Citado 4 vezes nas páginas 65, 66, 67 e 90.
- LIEB, E.; YNGVASON, J. Physics today 53. *April issue*, v. 32, 2000. Citado 6 vezes nas páginas 19, 29, 31, 32, 38 e 89.
- NASCIMENTO, C. K.; BRAGA, J. P.; FABRIS, J. D. Reflexões sobre a contribuição da carnot à primeira lei da termodinâmica. *Química Nova*, SciELO Brasil, v. 27, n. 3, p. 513–515, 2004. Citado 4 vezes nas páginas 25, 26, 34 e 65.
- NAZARENO, H. N.; BRITO, P. E. de. Long-range interactions and nonextensivity in one-dimensional systems. *Physical Review B*, APS, v. 60, n. 7, p. 4629, 1999. Citado na página 67.
- PRESS, W. H. et al. *Numerical Recipes with Source Code CD-ROM 3rd Edition: The Art of Scientific Computing*. [S.l.]: Cambridge University Press, 2007. Citado 3 vezes nas páginas 75, 77 e 78.
- RIOUL, O. This is it: A primer on shannon's entropy and information. 2018. Citado 15 vezes nas páginas 39, 49, 53, 54, 55, 56, 60, 61, 62, 63, 67, 68, 79, 97 e 98.
- SHANNON, C. A mathematical theory of communication, bell systems technol. *J*, v. 27, n. 3, p. 379–423, 1948. Citado 13 vezes nas páginas 11, 53, 54, 55, 56, 57, 58, 59, 60, 61, 69, 78 e 97.

WOLFRAM, S. *A new kind of science*. [S.l.]: Wolfram media Champaign, IL, 2002. v. 5. Citado 6 vezes nas páginas 25, 34, 37, 45, 49 e 67.

ZUBAREV, D. N. *Nonequilibrium statistical thermodynamics*. [S.l.]: Consultants Bureau, 1974. Citado 3 vezes nas páginas 45, 47 e 68.

Anexos

ANEXO A – Conceitos usados na teoria de Shannon

A.1 Desigualdade de informação

Supõe-se duas sequências: uma sequência de símbolos correspondente à mensagem selecionada pela fonte e outra referente ao sinal recebido. As probabilidades desses dois conjuntos estão correlacionados, no sentido que a probabilidade de se receber um sinal com uma certa sequência de elementos depende da mensagem que foi enviada e, portanto, podemos calcular a entropia de uma relativa à outra (SHANNON, 1948).

Existe uma desigualdade fundamental, também conhecida como *desigualdade de informação*, derivada pela primeira vez por Gibbs no século XIX, na qual “toda teoria clássica envolvendo informação pode ser derivada a partir dela” (RIOUL, 2018). Essa relação diz que, para uma variável aleatória X , existem duas distribuições de probabilidade $p(X)$ e $q(X)$ no qual vale:

$$\mathbb{E} \log \frac{1}{p(X)} \leq \mathbb{E} \log \frac{1}{q(X)}, \quad (\text{A.1})$$

com o valor esperado definido em 3.5.

A desigualdade de Gibbs é uma consequência da propriedade de concavidade da função logaritmo (RIOUL, 2018) e, pela desigualdade de Jensen, a diferença entre os dois lados da relação A.1 é

$$\mathbb{E} \log \frac{q(X)}{p(X)} \leq \log \mathbb{E} \frac{q(X)}{p(X)} = \log 1 = 0. \quad (\text{A.2})$$

De fato, $\mathbb{E} \frac{q(X)}{p(X)} = \sum_x \frac{q(x)}{p(x)} p(x) = \sum_x q(x) = 1$ no caso discreto e $\mathbb{E} \frac{q(X)}{p(X)} = \int \frac{q(x)}{p(x)} p(x) = \int q(x) = 1$ no caso contínuo. Por essa razão, a igualdade só acontece no caso em que $p(x)$ e $q(x)$ coincidem.

A.2 Entropia Relativa ou Divergência

Se $p(x)$ representa a distribuição relacionada a X e $q(x)$ a X^* , a diferença entre os dois lados da desigualdade A.1 resulta na chamada *entropia relativa*, também conhecida como divergência de Kullback-Leibler

$$D(X, X^*) = \mathbb{E} \log \frac{p(X)}{q(X)} \geq 0 \quad (\text{A.3})$$

Diferente da entropia de Shannon, a entropia relativa é positiva para ambos os casos de distribuição discreta ou contínua, onde p e q se comportam como distribuições de probabilidade discretas e funções densidade de probabilidade, respectivamente:

$$D(X, X^*) = \sum_x p(x) \log \frac{p(x)}{q(x)} \geq 0, \quad (\text{A.4})$$

$$D(X, X^*) = \int p(x) \log \frac{p(x)}{q(x)} dx \geq 0, \quad (\text{A.5})$$

Nota-se, então, que a divergência $D(X, X^*) = D(p, q)$ mede a “distância informacional” relativa às duas distribuições p e q (RIOUL, 2018). Olhando pra desigualdade de Gibbs em A.1, pode-se escolher a distribuição $q(x)$ de forma que ela maximize a entropia do sinal recebido X^* , enquanto p representa a probabilidade referente à mensagem X escolhida. Nesse caso, a divergência nada mais é que a diferença entre a entropia e seu máximo valor:

$$D(X, X^*) = \begin{cases} H(X^*) - H(X) & \text{no caso discreto} \\ h(X^*) - h(X) & \text{no caso contínuo.} \end{cases} \quad (\text{A.6})$$

Quando a divergência se anula, isso equivale a dizer que o canal de comunicação não apresenta ruído e a mensagem que chega até ele é o mesmo que sai, por isso, $H(X^*) = H(X)$. Caso contrário, quanto maior a distância entre as medidas de informação, maior a incerteza, maior o ruído.

A.3 Informação Mútua e Entropia Condicional

Um outro conceito importante que surge na teoria da comunicação é o da *informação mútua*. Assim como a noção de entropia relativa, ele é capaz de medir a dependência mútua entre duas variáveis aleatórias X e Y . Se X e Y são independentes, então a distribuição conjunta desses eventos deve ser calculado como sendo o produto das probabilidades individuais $q(x, y) = p(x)p(y)$ (RIOUL, 2018). Dado que a mensagem enviada e o sinal recebido possuem algum tipo de interdependência, a distribuição conjunta $q(x, y) = p(x)p(y)$ e a probabilidade esperada $p(x, y)$ costumam não coincidir e, nesse caso, a divergência informacional $D(q, p) = \mathbb{E} \log \frac{p(x, y)}{q(x, y)}$ equivale a medir a *informação mútua*, dada pela relação:

$$I(X; Y) = \mathbb{E} \log \frac{p(X, Y)}{p(X)p(Y)}. \quad (\text{A.7})$$

A expressão A.7 é justamente a medida de dependência mútua entre as variáveis. E pela desigualdade A.1, $I(X; Y) \geq 0$, com igualdade $I(X; Y) = 0$ apenas para o caso em que $p = q$, ou seja, pra quando X e Y são independentes.

Embora a medida de divergência $D(p, q)$ seja não simétrica com relação a p e q , a medida de informação *mútua* é simétrica em (X, Y) e, deste modo, $I(X; Y) = I(Y; X)$. A informação mútua também pode ser interpretada como a medida de informação fornecida por Y sobre X . Assim, reescrevendo a equação A.7 tem-se

$$I(X; Y) = \mathbb{E} \log \frac{p(X|Y)}{p(X)}, \quad (\text{A.8})$$

onde $p(x|y) = p(x, y)/p(y)$ é a distribuição condicional de X sabendo $Y = y$, e $I(X; Y)$ é a medida da dependência mútua entre as variáveis.

A informação mútua é um conceito importante e central dentro da teoria de Shannon, e também está associada ao conceito de entropia. Reescrevendo a relação A.8 na forma:

$$I(X; Y) = \mathbb{E} \log \frac{1}{p(X)} - \mathbb{E} \log \frac{1}{p(X|Y)} \quad (\text{A.9})$$

A expressão revela que informação representada pela distribuição incondicional $p(x)$ de X (não sabendo Y) é afetada pelo conhecimento de Y . Dito de outra forma, a informação mútua é a diferença entre a entropia de Shannon de X e a *entropia condicional* de X sabendo Y :

$$I(X, Y) = \left\{ \begin{array}{ll} H(X) - H(X|Y) & \text{no caso discreto,} \\ h(X) - h(X|Y) & \text{no caso contínuo.} \end{array} \right\} \quad (\text{A.10})$$

Pela simetria, a relação A.10 também pode ser escrita como a medida de informação fornecida por X sobre Y : $I(X, Y) = H(Y) - H(Y|X)$. A entropia condicional, também chamada de *equivoca*, corresponde à incerteza média sobre a mensagem conhecendo-se o sinal recebido, ou ainda, à incerteza média sobre o sinal recebido quando a mensagem é conhecida. O caso discreto da entropia condicional é determinado por:

$$H(X|Y) = \sum_x \sum_y p(x, y) \log \frac{1}{p(x|y)} \quad (\text{A.11})$$

Além disso, a informação mútua pode ser estendido sobre mais de duas variáveis, de maneira que se for multivariada pode ter valores negativos. Assim, dada três variáveis aleatórias X, Y e Z , pode-se definir a *informação condicional mútua* como sendo $I(X; Y|Z) = H(X|Z) - H(X|Y, Z)$.

A.4 A relação entre as medidas de informação

Sabendo que a distribuição condicional é definida por $p(x|y) = p(x, y)/p(y)$, a entropia condicional A.11 pode ser escrita como uma média dos valores de entropia,

$$H(X|Y) = \sum_y p(y) \sum_x p(x|y) \log \frac{1}{p(x|y)} = \sum_y p(y) H(X|Y = y). \quad (\text{A.12})$$

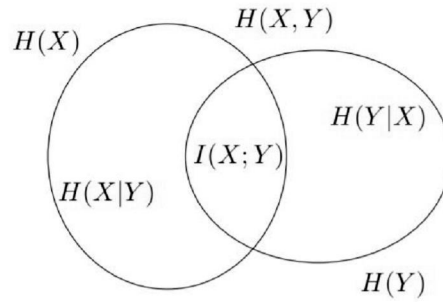


Figura 7 – Diagrama de Venn e o vínculo entre as medidas de informação de Shannon.

Portanto, para variáveis discretas, “enquanto $H(X)$ mede a incerteza sobre X , $H(X|Y)$ mede a incerteza média sobre X quando Y é conhecido”. Usando o fato de que a informação mútua é uma quantidade positiva, a partir da relação A.10, observa-se que $I(X;Y) = H(X) - H(X|Y) \geq 0$. Isso significa dizer que o ‘conhecimento’ reduz a incerteza sobre algo e, portanto,

$$H(X|Y) \leq H(X). \quad (\text{A.13})$$

Uma aplicação desse resultado é observado nos casos em que se observa uma mistura de sistemas. Considera-se dois sistemas X_1 e X_2 com distribuições de probabilidade p_1 e p_2 , cuja mistura é dada pela probabilidade linear $p_\lambda = \lambda_1 p_1 + \lambda_2 p_2$ com parâmetros λ_1 e λ_2 positivos tal que a soma $\lambda_1 + \lambda_2 = 1$. Nesse caso, se X_λ representa uma variável aleatória, pela relação A.13, nota-se que a entropia da mistura satisfaz

$$H(X_\lambda) \geq H(X_\lambda|\lambda) = \lambda_1 H(X_1) + \lambda_2 H(X_2). \quad (\text{A.14})$$

Com isso, conclui-se que a mistura aumenta a entropia resultante, isto é, “a entropia da mistura não é menor do que a mistura correspondente de entropias”.

O diagrama de Venn da Figura 7 consegue ilustrar de forma simples a relação entre conceitos tratados nessa seção. Nota-se do diagrama que a informação mútua, $I(X;Y)$, corresponde à intersecção dos dois “conjuntos de incerteza”, enquanto a entropia conjunta, $H(X, Y)$, equivale à união desses conjuntos.

Se existe uma bijeção entre X e Y , ou seja, se $X = Y$ suas incertezas se coincidem e suas respectivas entropias condicionais valem $H(X|Y) = H(Y|X) = 0$, logo:

$$H(X) = I(X; X) \quad (\text{A.15})$$

Esse resultado significa dizer que entropia também se refere à *auto-informação*: “ $H(X)$ pode ser visto como uma medida de informação provida por X sobre ela mesma”. Relembrando o fato de que entropia é uma quantidade positiva, $H(X) \geq 0$, a igualdade $H(X) = 0$ só é possível se X é independente dela mesma, isto é, para o caso em que $p(x) = 0$ ou 1. Ou seja, no caso em que a distribuição de probabilidade satisfaz a condição

determinística de certeza sobre aquele sistema. Por essa razão, a entropia $H(X)$ pode ser definida como uma medida de aleatoriedade ou incerteza sobre um sistema.

Se as variáveis X e Y são independentes, os dois conjuntos na figura 7 são separados, por conseguinte, a informação mútua entre eles vale $I(X;Y) = 0$. Nesse caso, a entropia conjunta equivale à soma das incertezas individuais $H(X,Y) = H(X) + H(Y)$. Mas no caso geral que as variáveis são dependentes, a entropia do conjunto é dado por $H(X,Y) = H(X) + H(Y|X) = H(Y) + H(X|Y)$, o que equivale a dizer que: “a incerteza conjunta é a soma da incerteza de uma variável e a incerteza da outra conhecendo-se a primeira”.

Por outro lado, se uma variável é totalmente dependente de outra, sua entropia será menor. Supondo que Y seja totalmente dependente de X de forma que $Y = f(X)$, observa-se pelo diagrama 7 que o conjunto de incerteza $H(Y)$ encontra-se dentro do conjunto $H(X)$ e, portanto, $H(Y|X) = 0$. E isso significa dizer que

$$H(f(X)) \leq H(X). \quad (\text{A.16})$$

De forma similar, observa-se que se os dois conjuntos no diagrama 7 forem escritos em função das variáveis aleatórias X e Y , para qualquer função f e g , suas incertezas $H(f(X))$ e $H(g(Y))$ se tornam menores e, conseqüentemente, a informação mútua entre elas também:

$$I(f(X); g(Y)) \leq I(X; Y). \quad (\text{A.17})$$

Esse resultado expressa que processamento de dados apenas reduzem a quantidade de informação. Esse exemplo é um importante resultado na teoria da informação conhecido como *desigualdade de processamento de dados*, utilizado para descrever as cadeias de Markov.