LEON BRILLOUIN

NAUKA A TEORIA INFORMACJI

WARSZAWA 1969
PAŃSTWOWE WYDAWNICTWO NAUKOWE

http://autonom.edu.pl

Dane oryginału:

SCIENCE AND INFORMATION THEORY

LEON BRILLOUIN, Member of the National Academy of Sciences
First ed. 1956, 2nd printing 1957, second ed. 1962
Academic Press Inc., New York — London

Library of Congress Catalog Card Number: 56—6601

Tłumaczyli z angielskiego
Stanisław Mazurek i Jan Szklanny

Pod redakcją i z przypisami
Jerzego Seidlera

Obwolutę i okładkę projektował
Stanisław Zamecznik

MIASTOPROJEKT - KATOWICE
BIBLIOTEKA TECHNICZNA
KATOWICACH
11768
007

Redaktor techniczny
Hanna Kęsicka

Korektor
T. Biernakiewicz

Printed in Poland

SPIS RZECZY

ERRATA

Str.	Wiersz od dołu	Jest	Powinno być
279	3	$\varphi = \frac{4\pi L}{\varphi}$	$\varphi = \frac{4\pi L}{\lambda}$
348	8	redukowaną	realizowaną
369	2	$P_0 = F/\varphi$	$P_0 = \left(\frac{F}{\varphi}\right)^N$

L. Brillouin, *Nauka a teoria informacji*

2. Zastosowanie definicji i rozważania ogólne

2.1. Definicje 35
2.2. Własność (A) 36
2.3. Własność (B) 37
2.4. Własność (C) 39
2.5. Zdarzenia łączne 42
2.6. Ilość informacji warunkowej 44

3. Rozwlekłość języka

3.1. Korelacja i zdarzenia łączne 46
3.2. Korelacja w języku 47
3.3. Rozwlekłość języka 49
3.4. Doświadczenia nad własnościami statystycznymi języka 51
3.5. Urządzenia kodujące 52

4. Zasady kodowania. Problemy związane z przepustowością kanału

4.1. Wstęp 54

Dane oryginału:

SCIENCE AND INFORMATION THEORY

LEON BRILLOUIN, Member of the National Academy of Sciences
First ed. 1956, 2nd printing 1957, second ed. 1962

II 7088

Redaktor techniczny
Hanna Kęsicka

Korektor
T. Biernakiewicz

Printed in Poland

SPIS RZECZY

Przedmowa 13

Przedmowa do drugiego wydania 15

Wstęp 17

1. Definicja ilości informacji

1.1. Definicja ilości informacji 22
1.2. Układy jednostek 23
1.3. Uogólnienia i przykłady 25
1.4. Informacja oparta o alfabet 26
1.5. Informacja zawarta w zespole symboli o różnych prawdopodobieństwach a priori 28
1.6. Uwagi ogólne 31

2. Zastosowanie definicji i rozważania ogólne

2.1. Definicje 35
2.2. Własność (A) 36
2.3. Własność (B) 37
2.4. Własność (C) 39
2.5. Zdarzenia łączne 42
2.6. Ilość informacji warunkowej 44

3. Rozwlekłość języka

3.1. Korelacja i zdarzenia łączne 46
3.2. Korelacja w języku 47
3.3. Rozwlekłość języka 49
3.4. Doświadczenia nad własnościami statystycznymi języka 51
3.5. Urządzenia kodujące 52

4. Zasady kodowania. Problemy związane z przepustowością kanału

4.1. Wstęp 54

4. 2. Definicja kanału i jego przepustowości 55
4. 3. Symbole, słowa i wiadomości przy kodowaniu sekwencyjnym 57
4. 4. Omówienie wyników 59
4. 5. Przykłady 62
4. 6. Obliczanie przepustowości kanału 66
4. 7. Dopasowanie kodu do kanału 66
4. 8. Zagadnienie ogólne — symbole o różnych długościach 69
4. 9. Zagadnienie dopasowania 73
4.10. Zagadnienia statystyki słów (Mandelbrot) 73
4.11. Rozwiązanie zagadnienia dopasowania 77
Dodatek do rozdziału 4 80

5. Zagadnienia kodowania

5.1. Kodowanie poszczególnych liter — system dwójkowy 82
5.2. Kodowanie poszczególnych liter — system trójkowy 84
5.3. Alfabet i liczby 86
5.4. Dwójkowe kodowanie słów 87
5.5. Alfabetyczne kodowanie słów 91
5.6. Kodowanie oparte na korelacji w obrębie bloków liter 91

6. Kody wykrywające i korygujące błędy

6.1. Kody wykrywające błędy 95
6.2. Kody korygujące pojedyncze błędy 96
6.3. Kody korygujące pojedyncze błędy i wykrywające błędy podwójne 100
6.4. Efektywność kodów korekcyjnych 101
6.5. Przepustowość binarnego kanału z szumem 104

7. Zastosowanie teorii informacji do rozwiązywania niektórych zagadnień specjalnych

7.1. Zagadnienia klasyfikacji z zastosowaniem komórki neutralnej 106
7.2. Metoda segregowania do wielu komórek 109
7.3. Najkorzystniejsza liczba sygnałów przypadających na komórkę elementarną 112

8. Analiza sygnałów: analiza Fouriera i metody próbkowania

8.1. Szeregi Fouriera 114
8.2. Zjawisko Gibbsa i zbieżność szeregów Fouriera 117
8.3. Całki Fouriera 120
8.4. Rola skończonej szerokości zajmowanego pasma częstotliwości 124
8.5. Zasada nieoznaczoności dla czasu i częstotliwości 127
8.6. Stopnie swobody wiadomości 132
8.7. Metoda próbkowania Shannona 136
8.8. Komórki informacyjne Gabora 139
8.9. Autokorelacja i widmo; twierdzenie Wienera-Chinczyna 141

8.10. Przekształcenia liniowe i filtry 144
8.11. Trójwymiarowa analiza Fouriera i metody próbkowania w trzech wymiarach 146
8.12. Analiza kryształów za pomocą promieni X 153
Dodatek do rozdziału 8 155

9. Zasady termodynamiki

9.1. Wstęp 157
9.2. Dwie zasady termodynamiki. Entropia i negentropia. 157
9.3. Niemożliwość wiecznego ruchu. Maszyny cieplne 161
9.4. Statystyczna interpretacja entropii 163
9.5. Przykłady rozważań statystycznych 165
9.6. Fluktuacje energii. Wzór Gibbsa 167
9.7. Oscylator kwantowy. 169
9.8. Fluktuacje 171

10. Pobudzenie cieplne i ruchy Browna

10.1. Ruch cieplny 174
10.2. Błądzenie przypadkowe 175
10.3. Efekt śrutowy. 179
10.4. Ruchy Browna 182
10.5. Pobudzenie cieplne w obwodzie elektrycznym 185
Dodatek do rozdziału 10 188

11. Szum cieplny w obwodach elektrycznych. Wzór Nyquista

11.1. Model z impulsami przypadkowymi. 190
11.2. Metoda Nyquista 192
11.3. Wnioski i przykłady zastosowania. 194
11.4. Uogólnienie wzoru Nyquista 196
11.5. Pobudzenie cieplne w obwodzie z prostownikiem 199

12. Zasada negentropii

12.1. Związek pomiędzy informacją a entropią 202
12.2. Negentropijna zasada informacji i uogólnienie zasady Carnota 204
12.3. Niektóre typowe przykłady fizyczne 207
12.4. Uwagi ogólne 211

13. Demon Maxwella a negentropijna zasada informacji

13.1. Demon Maxwella — podłoże historyczne 214
13.2. Wypędzanie demona 217
13.3. Omówienie 220
13.4. Postępowanie demona jako transformacja informacji w negentropię. . . 222

13.5. Negentropia potrzebna do obserwacji 226
13.6. Zagadnienie Szilarda: dobrze informowana maszyna cieplna 232
13.7. Rozważania Gabora 236
Dodatek I do rozdziału 13 240
Dodatek II do rozdziału 13 241

14. Negentropijna zasada informacji w fizyce ogólnej

14.1. Zagadnienie pomiarów w fizyce 242
14.2. Obserwacje dokonywane na oscylatorze 244
14.3. Rezonator wielkiej częstotliwości i cena obserwacji 247
14.4. Eksperymenty wymagające wielu jednoczesnych obserwacji przy niskich częstotliwościach 249
14.5. Zagadnienia wymagające dużej pewności obserwacji 254
14.6. Dokładniejsze omówienie doświadczeń, w których stosuje się wielkie . . częstotliwości 257
14.7. Przykład ilustrujący występowanie minimum negentropii potrzebnej do dokonania obserwacji 259

15. Obserwacja a informacja

15. 1. Błędy doświadczalne a informacja 264
15. 2. Pomiary odległości z małą dokładnością 266
15. 3. Pomiary odległości z dużą dokładnością 269
15. 4. Sprawność obserwacji 273
15. 5. Pomiar odległości za pomocą interferometru 274
15. 6. Inny schemat pomiaru odległości 278
15. 7. Pomiar okresów czasu 283
15. 8. Obserwacje pod mikroskopem 286
15. 9. Ognisko w falowodzie 291
15.10. Przykłady i omówienia 294
15.11. Podsumowanie 296

16. Teoria informacji, zasada nieoznaczoności i ograniczenia fizyczne obserwacji

16.1. Uwagi wstępne 298
16.2. Obserwacja jest procesem nieodwracalnym 300
16.3. Zasadnicze ograniczenia dokładności pomiarów fizycznych 302
16.4. Granice stosowalności geometrii euklidesowej 305
16.5. Możliwość zastąpienia fotonów cząstkami ciężkimi 307
16.6. Relacje nieoznaczoności w doświadczeniu z mikroskopem 309
16.7. Pomiar pędu 312
16.8. Nieoznaczoność w pomiarach pola 315

17. Negentropijna zasada informacji w telekomunikacji

17.1. Analiza sygnałów o skończonej szerokości pasma 317

17.2. Sygnały i szum termiczny — przedstawienie w hiperprzestrzeni 318
17.3. Przepustowość kanału z szumem 320
17.4. Omówienie wzoru Tullera–Shannona 321
17.5. Przykład praktyczny 325
17.6. Negentropijna zasada informacji w zastosowaniu do kanału z szumem . . 328
17.7. Zmodyfikowany wzór Gabora i rola zdudnień 331

18. Pisanie, drukowanie i czytanie

18.1. Przesyłanie informacji — informacja żywa 334
18.2. Zagadnienie czytania i pisania 335
18.3. Informacja martwa — jak ją ożywić 336
18.4. Pismo i druk 339
18.5. Omówienie przypadku specjalnego 340
18.6. Nowa informacja a rozwlekłość 341

19. Zagadnienie obliczeń

19.1. Maszyny matematyczne 343
19.2. Maszyna matematyczna jako element matematyczny 346
19.3. Maszyna matematyczna jako element obwodu. Próbkowanie i proces odwrotny do próbkowania (Salzer i Linvill) 349
19.4. Obliczenia na podstawie próbek danych w chwili t 352
19.5. Funkcja przenoszenia maszyny matematycznej 354
19.6. Układy zawierające maszynę matematyczną. Zagadnienie stabilności . . 356
19.7. Stabilność programu 358
19.8. Przykłady funkcji przenoszenia 361

20. Informacja, organizacja i inne zagadnienia

20.1. Informacja a organizacja 366
20.2. Informacja zawarta w prawie fizycznym 368
20.3. Informacja zawarta w tablicy liczb 371
20.4. Uwagi ogólne 373
20.5. Przykłady zagadnień nie mieszczących się w ramach współczesnej teorii 374
20.6. Zagadnienia informacji semantycznej 378

21. Nieuniknione błędy, determinizm i informacja

21.1. Informacja w nauce 384
21.2. Informacja jest skończona 385
21.3. Pogląd M. Borna 385
21.4. Obserwacje i błędy doświadczalne 386
21.5. Prosty przykład: wygnanie demona Laplace'a 388
21.6. Inne przykłady: oscylator anharmoniczny i prostownik 390
21.7. Anomalia oscylatora harmonicznego 393
21.8. Zagadnienie determinizmu 395

21. 9. Teoria informacji w świetle omówionych przykładów 397
21.10. Obserwacja i interpretacja 400
21.11. Wnioski 402

22. Zagadnienie skrajnie małych odległości

22.1. Trudności w mierzeniu skrajnie małych odległości 404
22.2. Możliwość wykorzystania poprzednich uwag do obliczania całek rozbieżnych występujących w zagadnieniach fizyki teoretycznej 406
22.3. Przykład: masa elektromagnetyczna elektronu 408
22.4. Obrona naszych założeń: ruch drżący Schrödingera 409
22.5. Omówienie i możliwość uogólnień 410

Przypisy 412

Skorowidz nazwisk 419

Skorowidz rzeczowy 421

Żonie, której miłość i pełne wiary poświęcenie były dla mnie ciągłym natchnieniem

PRZEDMOWA

W ciągu kilku ostatnich lat narodziła się nowa teoria naukowa, a mianowicie teoria informacji. Wzbudziła ona natychmiast ogromne zainteresowanie i rozwinęła się gwałtownie. Nowa ta teoria powstała w wyniku bardzo praktycznych i utylitarnych rozważań pewnych podstawowych zagadnień: w jaki sposób określić ilość informacji zawartej w wiadomości lub w telegramie, który ma być nadany? Jak mierzyć ilość informacji przenoszonej w obu tych przypadkach? Jak można porównać ilość informacji i jak mierzyć sprawność urządzeń kodujących? Wszystkie te i wiele podobnych zagadnień interesują inżyniera telekomunikacji i obecnie mogą być już traktowane ilościowo.

Z analizy tych zagadnień wyłoniła się nowa teoria o charakterze matematycznym i praktycznym, oparta na rozważaniach probabilistycznych. Raz w sposób ścisły sformułowana, może być ona wykorzystana przy rozpatrywaniu wielu podstawowych zagadnień naukowych. W szczególności wskazuje ona na możliwość rozwiązania problemu demona Maxwella i pokazania bezpośredniego związku pomiędzy informacją i entropią. Entropia w sensie termodynamicznym mierzy brak informacji o pewnym układzie fizycznym. Jeśli doświadczenie przeprowadzamy w laboratorium, to informację uzyskujemy kosztem wzrostu entropii, przy czym według uogólnionej zasady Carnota wzrost entropii musi być zawsze większy od uzyskanej ilości informacji. Ilość uzyskanej informacji odpowiada ujemnej entropii. Wielkość tę autor nazwał negentropią. Uogólnioną zasadę Carnota można więc także nazwać negentropową zasadą informacji. Zasada ta nakłada nowe ograniczenie na pomiary fizyczne niezależnie od zasad nieoznaczoności mechaniki kwantowej.

Niniejsza książka oparta jest na cyklu wykładów wygłoszonych dla inżynierów z International Business Machine Corporation (IBM) w Poughkeepsie

i Endicott (w stanie Nowy Jork) i później, na różnych uniwersytetach, w szczególności na Uniwersytecie Kalifornijskim w Berkeley.

Moim miłym obowiązkiem jest podziękować National Science Foundation za stypendium, które umożliwiło mi napisanie tej książki. Naukowa atmosfera Uniwersytetu Kolumbijskiego była mi również bardzo pomocna przy wykonywaniu tego zadania. Jestem także bardzo wdzięczny Pani Mary H. Payne za jej pomoc i cenne uwagi krytyczne.

LEON BRILLOUIN

Styczeń 1956

PRZEDMOWA DO DRUGIEGO WYDANIA

Ponieważ pierwsze wydanie zostało bardzo dobrze przyjęte, zdecydowano przygotować drugie wydanie, które zawiera wiele nowych spostrzeżeń. Nie zmieniając ogólnej struktury książki, w pierwszych rozdziałach wprowadzono pewną ilość poprawek i wyjaśnień. Ostatnie rozdziały: 21 i 22 — zupełnie nowe — są związane z kierunkiem badań przeprowadzonych przez autora w ciągu ostatnich kilku lat. Wprowadzono nowe zagadnienia. Ponownie rozważono rolę błędów w obserwacjach naukowych, co doprowadziło do krytycznego omówienia pojęcia determinizmu.

Uczeni wierzą w determinizm, lecz nie są w stanie tej wiary uzasadnić, ponieważ ich doświadczenia nigdy nie są dokładne. Brak dokładności jest bardzo poważnym mankamentem przy określaniu małych odległości. Eksperymentator nie zgadza się tutaj z matematykiem: nie chce rozważać rzeczy, których nie może zmierzyć. Takie podejście prowadzi do pewnych ciekawych konsekwencji, rozważanych w ostatnim rozdziale niniejszej książki.

LEON BRILLOUIN

Nowy Jork, 20 marca 1960

WSTĘP

Wraz z pojawieniem się i rozwojem teorii informacji nauka pozyskała nowe dziedziny badań. Teoria ta stała się obiektem zainteresowań wielu pionierów nauki i badaczy. Jest to zjawisko interesujące w historii nauki; tak gwałtowny rozwój badań jednej z gałęzi nauki zasługuje na głębszą analizę. Jak to się stało? Jak daleko sięga i jeszcze sięgnąć może? Czy rozwój ten oznacza inwazję wiedzy na obszary tradycyjnie należące do filozofii, czy może jest to odkrycie nowego obszaru, pewnego rodzaju „ziemi niczyjej", która wymknęła się dotychczasowym badaniom? Na wszystkie te pytania należałoby znaleźć możliwie dokładną odpowiedź.

Przede wszystkim: co to jest „informacja"? Zajrzyjmy do słownika Webstera: «przekazywanie lub percepcja (odbiór) wiedzy lub wiadomości; fakty gotowe do przekazania w odróżnieniu od zawartych w myślach lub wiedzy; dane, nowości, wiadomości, wiedza zdobyta na drodze studiów lub obserwacji...». Możemy powiedzieć, że informacja jest materiałem surowym. Jest jedynie zbiorem danych, wiedza natomiast zawiera pewną ilość rozmyślań i rozważań systematyzujących dane przez ich porównywanie i klasyfikację. Dalszy krok prowadzi już do wiedzy naukowej i formułowania praw naukowych.

Czy i w jaki sposób jest możliwe stworzenie podstaw naukowych teorii informacji? Należałoby przede wszystkim zacząć od ścisłej definicji. Nauka zaczyna się tam, gdzie pojawia się konieczność ścisłego rozgraniczenia znaczenia słów. Słowa mogą zostać wybrane z istniejącego już słownika. Można także stworzyć słowa zupełnie nowe; wszystkie muszą mieć jednak dokładnie określone znaczenia, dzięki czemu nie dochodzi do nieporozumień, nie ma dwuznaczności wewnątrz tej gałęzi nauki, która się nimi posługuje. Może się

zdarzyć, że to samo słowo w dwu różnych dziedzinach nauki ma różne znaczenia. Słowo „pierwiastek" ma ściśle określone znaczenie dla matematyka i inne, równie charakterystyczne znaczenie dla chemika. „Pierwiastek" z algebry nie tworzy związków chemicznych, a „pierwiastek" chemika nigdy nie jest niewymierny. Jednoznaczność słów jest charakterystyczna dla metod naukowych. Ponieważ naukowcy każdego kraju wprowadzają podobne określenia, łatwo jest dokonać przekładu przez wzajemnie jednoznaczne przyporządkowanie wyrazów w słownikach naukowych. Gdyby podobnie było ze słowami stosowanymi w życiu codziennym, porozumiewanie się między narodami byłoby o wiele łatwiejsze. Laik odczuwa zakłopotanie, jeżeli słowo potoczne używane jest w nowym, naukowym znaczeniu, i może być skłonny nazywać tego rodzaju stosowanie słów „żargonem naukowym". Jednakże przyjmowanie „żargonów" jest regułą w każdej wyspecjalizowanej dziedzinie — w teologii, filozofii, jak również w technice. Laik nie może zrozumieć języka specjalistów, ponieważ nie zna rozważanych przez nich problemów.

Precyzyjne definiowanie słów w języku naukowym opiera się zazwyczaj na dwóch różnych metodach. W matematyce definicje oparte są na pewnej liczbie starannie dobranych i określonych przez postulaty pojęć pierwotnych. Pojęcia bardziej złożone określa się za pomocą pojęć pierwotnych. Nowo wprowadzone definicje są równoznaczne tłumaczeniu (za pomocą słów) formuł zapisanych symbolicznie i opartych na postulatach. W naukach doświadczalnych wprowadzane są definicje innego typu, które mogą być nazwane „operacyjnymi". Na przykład siła, masa, prędkość itd. są definiowane przez krótki opis operacji, jakich należy dokonać w celu pomiaru tych wielkości. Operacyjny punkt widzenia przyjmuje wielu znanych uczonych, m. in. często przytaczane jest nazwisko P. W. Bridgmana. Stoją oni na stanowisku, że z zasady powinno się wprowadzać do języka naukowego tylko te wielkości, które mogą być zdefiniowane operacyjnie*. Słowa, dla których nie można znaleźć żadnej operacyjnej definicji, zazwyczaj okazują się w końcu niegodne zaufania i są eliminowane ze słownika naukowego. Przypomnijmy dla przykładu termin „eter", który teoria względności pozbawiła sensu.

Wracając do teorii informacji musimy zacząć od ścisłej definicji pojęcia „ilość informacji". Weźmy pod uwagę sytuacje, które mogą wystąpić w kilku

* Tzn. może być podany zespół działań (operacji), które należy wykonać, by zmierzyć tę wielkość. (*Przyp. tłum.*).

wariantach, przy czym ilość tych wariantów jest duża i nie mamy początkowo żadnej informacji o tym, jaki wariant w rzeczywistości ma miejsce. Jeżeli jednak zdarzy się, że mamy o tej sytuacji pewną informację, wówczas liczba możliwych wariantów ulegnie zmniejszeniu. Dokładna informacja o interesującej nas sytuacji może prowadzić do zawężenia zbioru wchodzących w grę wariantów do tylko jednego. Miernikiem ilości informacji może być funkcja stosunku liczby możliwych wariantów przed i po uzyskaniu informacji. Aby uzyskać addytywność ilości informacji zawartej w kilku niezależnych sytuacjach, wybierzemy zależność logarytmiczną. Problemy te i związane z nimi definicje rozważymy w rozdziale pierwszym. Stanowią one podstawę nowej teorii.

Teoria informacji może być z powodzeniem stosowana przy rozwiązywaniu wszystkich zagadnień technicznych, w których informacja odgrywa rolę, w szczególności zagadnień dotyczących: kodowania, telekomunikacji, maszyn matematycznych itd. W tych wszystkich przypadkach mamy do czynienia z przetwarzaniem informacji lub przekazywaniem jej z jednego miejsca na drugie. Teoria informacji jest bardzo użyteczna przy ustalaniu praw dotyczących przekazywania i przetwarzania informacji oraz przy określaniu dokładnych granic tego, co można robić, a czego nie. Jednakże nie jesteśmy na razie w stanie analizować procesów myślenia i nie możemy w tej chwili wprowadzić do naszej teorii żadnego elementu zawierającego pojęcie wartości informacji w sensie potocznym. Wyeliminowanie pojęcia wartości informacji jest bardzo poważnym ograniczeniem, lecz jest to cena, jaką musieliśmy zapłacić za możliwość stworzenia tej gałęzi wiedzy. Ograniczenie, o którym mowa, pozwala na podanie definicji ilości informacji; dzięki niej możemy traktować informację jako wielkość fizycznie mierzalną. Definicja ta nie rozróżnia jednak informacji o wielkim znaczeniu od niezbyt ważnej dla odbiorcy. Na pierwszy rzut oka definicja ta może się wydać sztuczna, lecz w rzeczywistości jest praktyczna i ścisła. Jest to definicja oparta na danych statystycznych o rozważanym systemie i każdy obserwator mógłby uzyskać takie same dane, jeśli w ogóle są one dostępne. Dlatego definicja informacji, którą wprowadzimy, jest obiektywna, niezależna od obserwatora. Z drugiej strony, „wartość informacji" jest oczywiście pojęciem subiektywnym, zależnym od obserwatora. Informacja zawarta w pewnym zdaniu może być dla mnie bardzo ważna, jednocześnie zaś zupełnie nieważna dla mojego kolegi. Wiadomość w gazecie może być czytana z zainteresowaniem przez wielu czytelników, natomiast

teoria Einsteina nie ma żadnej wartości dla laika, a przykuwa uwagę fizyka.

Tak więc w omawianej teorii pomijamy wszystkie te elementy wartości informacji. Nie znaczy to wcale, że należy je zawsze ignorować, lecz że w obecnej chwili możliwości uwzględnienia wartości informacji nie zostały dostatecznie zbadane i sklasyfikowane. Będą to prawdopodobnie następne zadania w programie naukowej analizy.

Teoria informacji leży na „ziemi niczyjej", jaką są zagadnienia związane z obiektywną miarą ilości informacji, których nigdy przedtem nie rozważali naukowcy ani filozofowie. Jeśli sięgniemy do zagadnień wartości dotyczących informacji, to wtargniemy na obszary zarezerwowane dla filozofii. Czy kiedykolwiek będziemy w stanie przekroczyć tę granicę i powiększyć zasięg wiedzy w tym kierunku? Jest to pytanie, na które odpowie przyszłość.

Definicja obiektywnej ilości informacji ma duże znaczenie praktyczne. Eliminacja czynnika wartości daje możliwość rozwiązania szeregu zagadnień. Inżyniera projektującego system telefoniczny nie obchodzi, czy linią będą przesyłane wiadomości dyplomatyczne, kursy akcji na giełdzie, czy nic nie znaczące rozmowy. Problem techniczny pozostanie zawsze taki sam: przekazać informację bezbłędnie i dokładnie bez względu na to, jaką może ona mieć dla kogoś wartość. Projektant maszyny matematycznej nie wie, czy będzie ona używana do obliczania tablic astronomicznych, czy też kont czekowych. Odrzucenie pojęcia wartości informacji stwarza możliwość naukowej analizy informacji bez uprzedzeń i wpływu czynników emocjonalnych.

Kiedy odkrywamy godne uwagi podobieństwo pomiędzy informacją i entropią, w grę wchodzi fizyka. Podobieństwo to zauważył dawno temu L. Szilard w pracy z 1929 r., która była jakby zwiastunem obecnej teorii informacji. Praca ta była rzeczywiście pionierską penetracją nowego terytorium, które obecnie szczegółowo badamy. Szilard rozważał w niej problem demona Maxwella, który w tej książce jest traktowany jako jeden z ważniejszych. Związek pomiędzy informacją i entropią na nowo odkrył Shannon rozpatrując różnego rodzaju zagadnienia telekomunikacyjne, my zaś poświęcimy wiele rozdziałów na porównanie tych pojęć. Udowodnimy, że ilość informacji należy traktować jako ujemny składnik ogólnej entropii układu fizycznego, czyli krótko mówiąc, że ilość informacji jest negentropią. Entropię układu fizycznego często definiuje się jako miarę przypadkowości struktury tego układu. Zagadnienie to możemy teraz sformułować nieco inaczej.

Żaden układ fizyczny nie jest w pełni określony (opisany). Znamy jedynie wartości pewnych zmiennych makroskopowych, natomiast nie jesteśmy w stanie podać dokładnych położeń i prędkości wszystkich cząstek tego układu. Dysponujemy jedynie częściową informacją na temat takiego układu, większość informacji o jego szczegółowej strukturze jest dla nas niedostępna. Entropia mierzy brak informacji, inaczej mówiąc, określa ona całkowitą ilość traconej przez nas informacji o mikrostrukturze układu.

Jest to tzw. negentropijna zasada informacji. Ponieważ entropii i ilości informacji nie możemy rozważać oddzielnie, przeto zasada ta prowadzi bezpośrednio do uogólnionej drugiej zasady termodynamiki. Ta negentropijna zasada informacji będzie poparta różnymi przykładami zaczerpniętymi z szerokiego zakresu dziedzin, zaczynając od fizyki teoretycznej, a kończąc na życiu codziennym. Zasadnicze zadanie polega na pokazaniu, że każda obserwacja lub eksperyment, dokonane na układzie fizycznym, automatycznie wywołują pewien wzrost entropii laboratorium. Jest więc możliwe porównanie straty negentropii (wzrostu entropii) z ilością uzyskanej informacji. Efektywność doświadczenia może być zdefiniowana jako stosunek ilości uzyskanej informacji do związanego z nią wzrostu entropii. Tak określona efektywność, zgodnie z uogólnioną zasadą informacji, jest zawsze mniejsza od jedności. Jak pokazują przykłady, sprawność może osiągać wartości bliskie jedności w pewnych, poszczególnych przypadkach, lecz może być także bardzo mała w innych.

Taki kierunek rozważań jest nader użyteczny przy porównywaniu podstawowych doświadczeń naukowych, a szczególnie fizycznych. Stwarza on nowe możliwości badań efektywności różnych metod obserwacji — ich dokładności i niezawodności.

Interesującym wynikiem tych rozważań jest wniosek, że pomiar skrajnie małych odległości jest fizycznie niemożliwy. Matematyk definiuje wielkości nieskończenie małe, lecz fizyk nie jest w stanie ich mierzyć, są więc one czystą abstrakcją bez żadnego znaczenia fizycznego. Jeśli przyjmiemy operacyjny punkt widzenia, powinniśmy wyeliminować wielkości nieskończenie małe z teorii fizycznych. Niestety, nie wiemy, w jaki sposób zrealizować taki program.

Ogólnie rzecz biorąc, mamy nadzieję, że teoria informacji stanowi początek nowego, ważnego działu badań naukowych, szczególnie w fizyce, a również i w biologii.

1

DEFINICJA ILOŚCI INFORMACJI

1.1. DEFINICJA ILOŚCI INFORMACJI

Teoria informacji, nauka rozwijająca się w ostatnich latach, znalazła szerokie zastosowanie w różnych dziedzinach, a w szczególności w telekomunikacji, technice obliczeniowej, fizyce teoretycznej oraz przy analizie podstawowego procesu, jakim jest obserwacja naukowa. Pokażemy, że można zdefiniować wielkość mierzącą ilość informacji w danej operacji i że jest ona ściśle związana z entropią — fizyczną wielkością stosowaną w termodynamice.

Definicja ilości informacji oparta jest na rozważaniach statystycznych. Zacznijmy od rozpatrzenia prostego przykładu.

Weźmy pod uwagę system, który może znajdować się w jednym spośród P_0 różnych stanów, przy czym stany te są a priori równoprawdopodobne. Tak jest w sytuacji początkowej, kiedy nie mamy żadnej informacji o stanie interesującego nas systemu. Jeżeli dokonamy obserwacji tego systemu, to może się nam udać ustalić, w jakim spośród P_0 możliwych stanów rzeczywiście się on znajduje. Im bardziej nieokreślona była początkowa sytuacja, tym większe było P_0 i tym większa będzie ilość informacji, jaką uzyskamy stwierdzając, w jakim stanie znajduje się system. Podsumowując, mamy:

sytuacja początkowa: $I_0 = 0$ przy P_0 równoprawdopodobnych stanach;
sytuacja końcowa: $I_1 \neq 0$ przy $P_1 = 1$ (tj. jednym stanie, który wchodzi w grę).

W związku z tym rzeczą naturalną jest ilość informacji, którą oznaczymy symbolem I, zdefiniować wzorem

$$I_1 = K \ln P_0, \tag{1.1}$$

gdzie K jest stałą, a ln oznacza logarytm naturalny.

Wprowadzenie logarytmu w równaniu (1.1) jest uwarunkowane wymaganiem, aby ilość informacji miała cechę addytywności. Sens tego wyjaśnimy na przykładzie. Rozpatrzymy dwa niezależne systemy. Niech w pierwszym z nich wchodzi w grę P_{01} równoprawdopodobnych stanów, w drugim zaś P_{02}. Każdy stan pierwszego systemu może być zestawiony ze stanem drugiego systemu. Stąd całkowita początkowa liczba stanów obu systemów rozpatrywanych łącznie wynosi

$$P_0 = P_{01} \cdot P_{02};$$

daje to

$$I_1 = K \ln (P_{01} \cdot P_{02}) = I_{11} + I_{12}, \tag{1.2}$$

gdzie

$$I_{11} = K \ln P_{01}, \quad I_{12} = K \ln P_{02}.$$

Całkowita ilość informacji wymagana dla określenia stanów obu systemów jest sumą dwóch odrębnych informacji I_{11} i I_{12}.

1.2. UKŁADY JEDNOSTEK

Zanim przejdziemy do dalszych rozważań, musimy zająć się sprawą jednostek. W teorii informacji przyjęto uważać ilość informacji I za wielkość bezwymiarową (liczbę oderwaną); stąd stała K jest także liczbą oderwaną. Najwygodniejszy okazał się przy tym dwójkowy układ jednostek [1.1]* (*binary digits* — w skrócie *bits*).

Weźmy pod uwagę system, którego stan jest zespołem n niezależnych stanów elementarnych, przy czym każdy ze stanów elementarnych może przyjmować jedną z dwu postaci, 0 lub 1. Całkowita liczba możliwych stanów tego systemu wynosi

$$P = 2^n.$$

Z równania (1.1) otrzymujemy ilość informacji

$$I = K \ln P = K n \ln 2.$$

Jeśli chcemy utożsamić I z n (liczbą jednostek dwójkowych), to trzeba wziąć

$$K = 1/\ln 2 = \lg_2 e. \tag{1.3a}$$

* W ten sposób znaczone są przypisy redaktora zamieszczone na końcu książki.

W ten sposób dochodzimy do wyrażenia na ilość informacji w jednostkach dwójkowych (bitach):

$$I = \lg_2 P. \tag{1.3b}$$

Dotychczas mówiliśmy o obserwacji pierwotnie nieznanego stanu systemu i ilości informacji, jaką uzyskujemy dzięki tej obserwacji. Zagadnienie takie możemy sformułować w nieco inny sposób: mamy do dyspozycji zbiór możliwości, przy czym każda z nich jest jednakowo prawdopodobna, i dokonujemy wyboru jednej z nich; interesuje nas ilość informacji, jaką uzyskujemy dzięki takiemu wyborowi.

Jako przykład weźmy talię 32 różnych kart, z których wybieramy jedną. Ponieważ ilość możliwości

$$P_0 = 32 = 2^5,$$

na podstawie wzoru (1.3b) powiemy, że ilość informacji, którą uzyskujemy po dokonaniu wyboru karty, wynosi

$$I = 5 \text{ bitów}.$$

Weźmy teraz dwie oddzielne talie, każdą zawierającą po 32 różne karty. Jeżeli wybieramy dwie karty, po jednej z każdej talii, to chcielibyśmy powiedzieć, że uzyskamy dwukrotnie więcej informacji niż poprzednio. Całkowita liczba możliwości wynosi

$$P = P_1 \cdot P_2,$$

gdzie

$$P_1 = P_2 = 32 = 2^5,$$

i stąd

$$P = 2^{10}.$$

Ze wzoru (1.1) mamy dalej

$$I = K \ln P = K \ln (P_1 \cdot P_2) = K \ln P_1 + K \ln P_2.$$

A więc

$$I = 10 \text{ bitów}.$$

Wobec tego przyjęcie logarytmu w definicji (1.1) wydaje się być słuszne.

Do innego układu jednostek dojdziemy, jeżeli porównamy „informację" z „entropią" stosowaną w termodynamice ([1.2]) i zdecydujemy się obie te

wielkości mierzyć w tych samych jednostkach. Entropia (rozdz. 9) ma wymiar energii dzielonej przez temperaturę, tzn. erg/°C (w układzie cgs). Dla entropii słuszny jest wzór Boltzmanna, podobny do wzoru (1.1), zawierający współczynnik

$$k = 1{,}38 \cdot 10^{-16}\ \text{erg/°C}. \tag{1.4}$$

Współczynnik ten nazywany jest stałą Boltzmanna (wzór (9.15)). Jeżeli do równania (1.1) wstawimy współczynnik k zamiast K, to informację będziemy mierzyć w jednostkach entropii.

Możemy pójść jeszcze dalej i tak dobrać jednostki, aby zarówno entropia, jak i informacja były wielkościami bezwymiarowymi* (liczbami oderwanymi). Można to osiągnąć mierząc temperaturę w jednostkach energii, przy czym, jeśli k jest liczbą bezwymiarową określoną równaniem (1.4), można stosować zwykłą skalę stustopniową. W takim przypadku stosunek między jednostkami naszych dwóch systemów jednostek staje się również liczbą bezwymiarową:

$$k/K = k \ln 2 \approx 10^{-16}. \tag{1.4a}$$

Liczba ta odgrywa ważną rolę we wszystkich zastosowaniach teorii.

1.3. UOGÓLNIENIA I PRZYKŁADY

Definicja ilości informacji może być uogólniona na przypadek, kiedy w początkowej sytuacji wchodzi w grę P_0 możliwości, a w końcowej nadal pozostaje ([1.3]) ich $P_1 \geqslant 1$:

na początku: $I_0 = 0$ przy P_0 równoprawdopodobnych możliwościach;
na końcu: $I_1 \neq 0$ przy P_1 równoprawdopodobnych możliwościach.

W takim przypadku przyjmujemy

$$I_1 = K \ln (P_0/P_1) = K \ln P_0 - K \ln P_1. \tag{1.5}$$

Definicja ta jest oczywistym uogólnieniem wzoru (1.1) i sprowadza się doń, jeżeli $P_1 = 1$.

Rozpatrzmy teraz przypadek, gdy mamy do czynienia z wyborem jednej cyfry z pewnego zbioru cyfr, przy czym wszystkie cyfry są równoprawdo-

*D. E. Bell, J. Appl. Phys. **23** (1952), str. 372.

podobne. Weźmy pod uwagę zbiór G cyfr systemu o podstawie [1.4] N. Mamy ogólnie

$$I = K \ln N^G = KG \ln N. \tag{1.6}$$

Jeżeli informację mierzyć w bitach, to

$$I = G \lg_2 N, \tag{1.7}$$

gdyż

$$(\lg_2 N)(\ln 2) = \ln N \text{ i } K = 1/(\ln 2).$$

Wobec tego, jeśli $N = 2$, mamy 2^G różnych możliwości i

$$I = G \text{ bitów}. \tag{1.8}$$

Łatwo pokazać, że możemy uzyskać tę samą ilość informacji, jeśli zmienimy podstawę N i odpowiednio dobierzemy liczbę cyfr. Weźmy dla przykładu system ósemkowy, $N' = 8$. Wówczas dla przedstawienia danej liczby w tym systemie będziemy potrzebowali tylko $G' = G/3$ cyfr i

$$I = G' \lg_2 N' = (G/3) \cdot 3 = G, \tag{1.9}$$

co jest zgodne ze wzorem (1.8).

Rozważmy jeszcze liczbę przedstawioną przez G_{10} cyfr systemu dziesiętnego:

$$N = 10, \quad I = G_{10} \lg_2 10 = G_{10} \cdot 3{,}32 \text{ bitów}. \tag{1.10}$$

Widzimy stąd, że cyfra systemu dziesiętnego daje 3,32 bitów informacji i w systemie dwójkowym dla przedstawienia tej samej liczby trzeba 3,32 razy więcej pozycji niż w systemie dziesiętnym.

1.4. INFORMACJA OPARTA O ALFABET

Większość informacji [1.5], z której korzystamy, jest przekazywana przy pomocy języka. W języku mówionym elementarnymi symbolami są dźwięki podstawowe (zwane często fonemami), język pisany zaś składa się z wyrazów złożonych z liter. Weźmy pod uwagę napisane słowo i spróbujmy obliczyć ilość zawartej w nim informacji [1.6]. To zagadnienie, o dużym znaczeniu

praktycznym, było rozpatrywane przez C. E. Shannona* i wielu innych autorów ([1.7]). Jak zobaczymy dalej, jego kompletne i ścisłe rozwiązanie nie zostało dotąd uzyskane z powodu braku pełnych danych statystycznych o języku.

Litery możemy uważać za symbole, które musimy wybierać, jeśli chcemy zbudować słowo ([1.8]). Pełny alfabet zawiera 27 symboli: 26 liter *plus* spacja (przerwa między literami). Gdyby te wszystkie symbole były a priori równoprawdopodobne, moglibyśmy powiedzieć, że ilość informacji zawartej w słowie złożonym z G liter jest równa

$$\left.\begin{aligned} I &= G \lg_2 27 \text{ bitów} \\ \text{lub} \qquad & \\ i &= \lg_2 27 = 4{,}76 \text{ bitów na literę,} \end{aligned}\right\} \tag{1.11}$$

co odpowiada bezpośredniemu zastosowaniu równania (1.7). Rozwiązanie powyższe nie jest jednakże zadowalające, ponieważ litery nie pojawiają się

Tablica 1.1

Prawdopodobieństwo występowania p i wartości $-\lg p$ dla liter języka angielskiego

Symbol	p	$-\lg p$	Symbol	p	$-\lg p$
przerwa między słowami	0,2	0,699	*L*	0,029	1,54
E	0,105	0,979	*C*	0,023	1,64
T	0,072	1,143	*FU*	0,0225	1,65
O	0,0654	1,184	*M*	0,021	1,68
A	0,063	1,2	*P*	0,0175	1,76
N	0,059	1,23	*YW*	0,012	1,92
I	0,055	1,26	*G*	0,011	1,96
R	0,054	1,27	*B*	0,0105	1,98
S	0,052	1,28	*V*	0,008	2,1
H	0,047	1,33	*K*	0,003	2,52
D	0,035	1,46	*X*	0,002	2,7
			JQZ	0,001	3,0

* C. E. Shannon, W. Weaver, *The Mathematical Theory of Communication*, Urbana, Ill. 1949, Univ. of Illinois Press.

w słowach danego języka z równymi prawdopodobieństwami a priori. Dla języka angielskiego prawdopodobieństwa te zostały wyznaczone ([1.9]) i są podane w tabl. 1.1.

Niech p_j będzie prawdopodobieństwem a priori pojawienia się j-tej litery ($j = 1, 2, \ldots, 27$). Shannon podał następującą definicję średniej ilości informacji przypadającej na jedną literę:

$$i = -K \sum_{j=1}^{27} p_j \ln p_j. \tag{1.12}$$

Rozpatrzmy teraz sensowność tego wzoru.

1.5. INFORMACJA ZAWARTA W ZESPOLE SYMBOLI O RÓŻNYCH PRAWDOPODOBIEŃSTWACH A PRIORI

Przypuśćmy, że używamy M różnych liter (w sensie ogólnym) ([1.10]): $1, 2, \ldots, j, \ldots, M$, których prawdopodobieństwa a priori wynoszą odpowiednio $p_1, p_2, \ldots, p_j, \ldots, p_M$, a poza tym nie wprowadzamy żadnych innych warunków i ograniczeń stosowania tych liter.

Chcemy wyprowadzić wzór Shannona, z którego wynika, że informacja zawarta w słowie (w sensie ogólnym) złożonym z G liter jest równa

$$I = G \cdot i = -GK \sum_{j=1}^{j=M} p_j \ln p_j, \tag{1.13}$$

gdzie i jest średnią ilością informacji na literę, a suma

$$\sum_{j=1}^{j=M} p_j = 1.$$

Zacznijmy od prostego przypadku, który łatwo później uogólnić.

Rozważmy alfabet złożony z dwu „liter", a mianowicie kreski i kropki, tj. taki, jakim posługujemy się w telegrafii. Literom tym możemy również przypisać symbole 0 i 1. Rozpatrzmy zespół pozycji, w które mogą być wpisywane litery obejmujące G pozycji, i załóżmy, że N_0 pozycji zajmuje 0, a N_1 pozycji zajmuje 1, przy czym $N_0 + N_1 = G$, a więc każda z wszystkich G pozycji jest zajęta przez jeden z symboli. Prawdopodobieństwo tego, że dana

pozycja jest zajęta przez 0, wynosi

$$p_0 = N_0/G. \tag{1.14}$$

To samo prawdopodobieństwo dla 1 wynosi

$$p_1 = N_1/G, \tag{1.15}$$

przy czym

$$p_0+p_1 = 1, \tag{1.16}$$

ponieważ prawdopodobieństwo, że w danej pozycji znajduje się zero lub jeden, wynosi 1.

Znajdźmy teraz liczbę sposobów, na które można zapełnić G pozycji (komórek) tak, aby każdą pozycję zajmowało 0 lub 1, lecz nigdy dwie litery jednocześnie. Jest to taki sam problem, jak w statystyce Fermiego. Ilość sposobów, na które możemy przy powyższych warunkach wypełnić G pozycji, jest równa ilości sposobów wypełnienia N_0 pozycji zerami, ponieważ w chwili, gdy pozycje te zajmują już zera, pozostałe N_1 pozycji będzie musiało zawierać jedynki. Liczba sposobów, na które możemy zapełnić N_0 pozycji zerami, jest równa ilości kombinacji N_0 elementów spośród G elementów, a więc jest równa

$$P = G!/N_0!\,N_1!. \tag{1.17}$$

Jest to liczba różnych „słów" zbudowanych z G „liter" alfabetu dwuliterowego, w którym jedna litera używana jest N_0 razy, druga zaś N_1 razy. Jeżeli wybierzemy jedno z tych słów, to zgodnie ze wzorem (1.1), uzyskamy ilość informacji

$$I = K \ln P = K\,(\ln G!-\ln N_0!-\ln N_1!). \tag{1.18}$$

Jeżeli słowo jest długie oraz G, N_0 i N_1 są dostatecznie duże, to logarytmy silni mogą być aproksymowane za pomocą wzoru Stirlinga

$$\ln Q! \approx Q(\ln Q-1). \tag{1.19}$$

Wiadomo, że wzór Stirlinga daje dobre przybliżenie dla $Q>100$. Tak więc jeśli $G \gg 1$, $N_0 \gg 1$ i $N_1 \gg 1$, mamy

$$I \approx K[G(\ln G-1)-N_0(\ln N_0-1)-N_1(\ln N_1-1)]$$

lub, pamiętając, że $G = N_0+N_1$,

$$I \approx K\,(G \ln G-N_0 \ln N_0-N_1 \ln N_1). \tag{1.20}$$

Korzystając jeszcze raz z równania $G = N_0 + N_1$ można przepisać równanie (1.20) w postaci

$$I \approx -KG\left(\frac{N_0}{G}\ln\frac{N_0}{G} + \frac{N_1}{G}\ln\frac{N_1}{G}\right). \tag{1.21}$$

Jeżeli ostatecznie podstawimy wzory (1.14) i (1.15) do wzoru (1.21) i tak uzyskane równanie podzielimy przez G, to otrzymamy

$$i = I/G \approx -K(p_0 \ln p_0 + p_1 \ln p_1), \tag{1.22}$$

gdzie i jest ilością informacji przypadającą na jedną literę. Jest to wzór Shannona (1.13) w sytuacji, gdy mamy tylko dwie litery. Łatwo zauważyć, że wyrażenie (1.22) jest zawsze dodatnie. Wynika to z tego, że prawdopodobieństwa p_0 i p_1 są mniejsze od jedności i stąd ich logarytmy są ujemne.

Uogólnienie rozumowania na sytuacje, kiedy liter jest więcej niż dwie, nie jest trudne. Oznaczmy przez $N_1, N_2, \ldots, N_j, \ldots, N_M$ ilość różnych liter, a ilość zajmowanych przez nie pozycji przez

$$G = \sum_{j=1}^{j=M} N_j. \tag{1.23}$$

Następnie określmy prawdopodobieństwo j-tej litery:

$$p_j = N_j/G. \tag{1.24}$$

Stąd

$$\sum_{j=1}^{j=M} p_j = 1. \tag{1.25}$$

Całkowita liczba P słów, które mogą być otrzymane przez losowy rozkład liter w G pozycjach (przy warunku, że żadna pozycja nie może być zajęta przez więcej niż jedną literę), jest równa

$$P = G!\Big/\prod_{j=1}^{j=M} N_j!, \tag{1.26}$$

co stanowi oczywiste uogólnienie wyrażenia (1.17). Stąd otrzymujemy wyrażenie na ilość informacji zawartej w jednym słowie:

$$I = K \ln P = K(\ln G! - \sum_{j=1}^{j=M} \ln N_j!) \approx K(G \ln G - \sum_{j=1}^{j=M} N_j \ln N_j). \tag{1.27}$$

Równościom w wyrażeniu (1.27) odpowiadają równości (1.18), (1.19) i (1.20). Zakładamy przy tym, że G i N_j $(j = 1, 2, \ldots, M)$ są tak duże, iż można stosować wzór Stirlinga. Postępując podobnie jak w przypadku (1.21) i (1.22), otrzymamy wyrażenie

$$I \approx -KG \sum_{j=1}^{j=M} (N_j/G) \ln (N_j/G) = -KG \sum_{j=1}^{j=M} p_j \ln p_j, \tag{1.28}$$

będące wzorem Shannona.

Jako przykład rozważmy słowo złożone z 10000 liter wybranych losowo, z równymi prawdopodobieństwami a priori, z 27-literowego alfabetu. W takim przypadku mamy

$$G = 10000 \text{ oraz } I = 10000 \lg_2 27 = 47600 \text{ bitów}; \tag{1.29}$$

stąd

$$i = I/10000 = 4{,}76 \text{ bitów na literę}. \tag{1.29a}$$

Jeśli budując słowo tej samej długości będziemy brali pod uwagę prawdopodobieństwa a priori poszczególnych liter zgodnie z tabl. 1.1, to do obliczenia ilości informacji musimy użyć zależności (1.28). Daje ona

$$i = I/G = I/10000 \approx -K \sum_{j=1}^{j=27} p_j \ln p_j = 4{,}03. \tag{1.30}$$

W następnych rozdziałach przeprowadzimy dokładniejszą analizę struktury języka naturalnego. Pokaże ona, że otrzymana tu wartość jest górną granicą oraz że rzeczywista ilość informacji przenoszonej przez literę (w sensie potocznym) jest znacznie mniejsza od 4 i zawiera się przypuszczalnie w przedziale od 1 do 2 bitów na literę.

1.6. UWAGI OGÓLNE

Już obecnie możemy ogólnie powiedzieć, że każdego rodzaju ograniczenie, każdy dodatkowy warunek nałożony na swobodę wyboru powoduje zmniejszenie ilości informacji. Weźmy pod uwagę system, w którym wchodzi w gr P różnych stanów, jeżeli nie ma żadnych ograniczeń nałożonych na jego zmi ne. Jeśli ograniczymy swobodę wyboru wartości tych zmiennych, to war te wyeliminują stany, które pierwotnie wchodziły w grę. Nowa liczba st

P' przy wprowadzonych ograniczeniach będzie oczywiście mniejsza od początkowej liczby stanów P i wobec tego dzięki obserwacji konkretnego stanu uzyskamy nową ilość informacji $I'<I$:

$$\left.\begin{array}{ll} \text{bez ograniczeń:} & P \text{ stanów; } I = K \ln P; \\ \text{z ograniczeniami:} & P' \text{ stanów; } P'<P;\ I' = K \ln P' < I. \end{array}\right\} \tag{1.31}$$

Przykład dotyczący informacji przenoszonej przez litery, podany w poprzednim paragrafie, ilustruje nasz wynik. Kiedy litery stosuje się z równymi prawdopodobieństwami a priori (dowolnie), ilość informacji wynosi 4,76 bitów na literę. Jeżeli zaś nałożymy ograniczenia, biorąc pod uwagę prawdopodobieństwa a priori poszczególnych liter zgodnie z tabl. 1.1, to ilość informacji na literę spada do 4,03 bita. Inne dodatkowe ograniczenia zmniejszyłyby tę wartość jeszcze bardziej. Drugim sposobem objaśnienia naszego ogólnego wyniku jest stwierdzenie, że ograniczenia przedstawiają pewną wstępną ilość informacji I_0 o wybieranym słowie, którą mamy przed wybraniem tego słowa; stąd

$$I' = I - I_c$$

jest informacją, jaka pozostaje do uzyskania, jeśli znamy I_c.

Obecnie omówimy nieco dokładniej kierunek naszych dalszych rozważań i naświetlimy związane z nim ograniczenia. W równaniu (1.1) wybraliśmy statystyczną definicję ilości informacji. Taka matematyczna definicja będzie korzystna przy analizie wielu naukowych i technicznych problemów. Pozwoli nam ona wyciągnąć pewne ogólne wnioski mające istotne znaczenie praktyczne i oznaczające się wielką ogólnością. Jednak ta niewątpliwie precyzyjna definicja stanowi także pewne ograniczenie. Aby uzyskać tę definicję, musimy odrzucić lub pominąć wiele potocznych sposobów rozumienia pojęcia „ilość informacji".

„Ilość informacji" definiujemy tu jako wynik wyboru. Nie traktujemy jednak uzyskanej informacji jako podstawy do przewidywania ani jako podstawy, na której może być oparty inny wybór. Definiując w ten sposób ilość informacji pomijamy inne sposoby potocznego rozumienia pojęcia ilości informacji. Na przykład ciągowi złożonemu ze stu liter przyporządkowujemy określoną ilość informacji nie patrząc na to, czy stanowi on sensowne zdanie angielskie, a jeżeli tak, to czy owo zdanie ma wagę z praktycznego punktu widzenia. Zgodnie z naszą definicją, zbiorowi stu liter wybranych losowo (z prawdopodo-

bieństwami jak w tabl. 1.1), stuliterowemu urywkowi z gazety, fragmentowi z dzieł Szekspira lub prac Einsteina przyporządkowujemy taką samą ilość informacji. Inaczej mówiąc, *ilość informacji* definiujemy tu jako coś odrębnego od *stopnia wiedzy*, dla którego oceny nie dysponujemy ilościową miarą.

Nie odróżniamy więc informacji pożytecznej od nieużytecznej i decydujemy się kompletnie ignorować pojęcie wartości informacji. Nasza statystyczna definicja ilości informacji oparta jest wyłącznie na tym, że pierwotnie mamy do czynienia z sytuacją niekompletnie określoną. Jeżeli sytuacja jest pierwotnie niekompletna, to niesie informację. Czy informacja ta jest wartościowa, czy nie — to nas nie interesuje. Pojęcie wartości informacji wiąże się z możliwością jej wykorzystania przez celowo działającego obserwatora (w szczególności człowieka). To jednak leży poza zasięgiem przedstawionej tu teorii ([1.11]). W szczególności nie będziemy zajmować się procesem myślenia ani użytkowaniem informacji przez żywe organizmy.

Mimo to wprowadzona tu definicja ilości informacji jest niewątpliwie ogromnie pożyteczna i wygodna. Odzwierciedla ona dokładnie sytuację inżyniera telekomunikacji, który musi być w stanie przesłać wiadomość zawartą na wypełnionym blankiecie telegraficznym, nie zwracając uwagi na wartość, jaką ta wiadomość przedstawia dla odbiorcy.

Wprowadzona tu ilość informacji jest wielkością absolutną, mającą tę samą wartość liczbową dla każdego obserwatora ([1.12]). Z drugiej strony, wartość, jaką przedstawia informacja dla ludzi, powinna być wielkością względną, mającą różne wartości dla różnych odbiorców, zgodnie z możliwością jej rozumienia i późniejszego użytkowania. Twierdzenie Einsteina dla matematyka będzie miało zapewne większą wartość niż artykuł z gazety. Dla laika artykuł w gazecie może mieć wartość lub nie, można natomiast z całą pewnością powiedzieć, że twierdzenie Einsteina nie będzie miało dla niego żadnej wartości.

Weźmy przykład bardziej trywialny. Przy wyborze jednej karty z 32 informacja wynosi 5 bitów; jest ona zawsze równa 5 bitów, niezależnie od tego, czy wyciągnięta karta okaże się asem, siódemką, czy królem. Wartość tych kart zależy jednak od reguł gry.

Zgodnie z naszą definicją, ilość informacji jest zawsze wielkością dodatnią. Tymczasem wartość informacji w sensie potocznym może w pewnych przypadkach nabierać znaczenia negatywnego. Profesor wygłaszając długi wykład

nagle odkrywa, że popełnił błąd, i mówi: „przepraszam, ale to wszystko, co powiedziałem, było błędne". To ostatnie zdanie ma wartość negatywną i niszczy całą informację zawartą w wykładzie.

Przykłady te dostatecznie ilustrują ograniczenia przedstawianej teorii i powinny być uwzględniane przy jakimkolwiek jej zastosowaniu. Do zagadnień tych wrócimy w rozdz. 20.

2

ZASTOSOWANIE DEFINICJI I ROZWAŻANIA OGÓLNE

2.1. DEFINICJE

W pierwszym rozdziale pokazaliśmy, że definicja miary ilości informacji (jako miernika stopnia nieoznaczoności, która istnieje przed dokonaniem wyboru) była ścisła, lecz z konieczności ograniczona. Tym samym „wartość" informacji nie mogła być w takiej definicji miary ilości informacji uwzględniona.

Pokazaliśmy tam także, że jeżeli dla możliwych wyborów istnieją niejednakowe prawdopodobieństwa a priori, to prawdopodobieństwa te mogą być interpretowane jako ograniczenia nałożone na nasz wybór, co w efekcie prowadzi do zmniejszenia ilości informacji.

Tak więc, jeśli prawdopodobieństwa a priori liter (1), (2), ..., (j), ... wynoszą odpowiednio: $p_1, p_2, \ldots, p_j, \ldots$, to (jak wykazano) ilość informacji na literę jest, zgodnie ze wzorem (1.12), równa

$$I/\text{literę} = -K \sum_j p_j \ln p_j. \tag{2.1}$$

Wzór ten otrzymano w gruncie rzeczy z wyrażenia określającego informację na literę w sytuacji, gdy na wybór nie nałożono żadnych ograniczeń, tzn. ze wzoru

$$I/\text{literę} = (K/G) \ln m^G = K \ln m, \tag{2.2}$$

gdzie m oznacza ilość różnych równoprawdopodobnych liter, G — całkowitą liczbę użytych liter (zamiast M z rozdz. 1.5 stosujemy tu m). Tak więc wy-

chodząc z równania (2.2) otrzymaliśmy (2.1). Wielu autorów, a przede wszystkim Shannon, za punkt wyjścia przyjmowało jednak równanie (2.1). Równanie (2.2) wynika z niego natychmiast. Załóżmy teraz, że mamy m różnych liter i że ich prawdopodobieństwa a priori są równe, tzn.:

$$p_1 = p_2 = \ldots = p_j = \ldots = p_m = 1/m. \tag{2.3}$$

Z równania (2.1) otrzymujemy

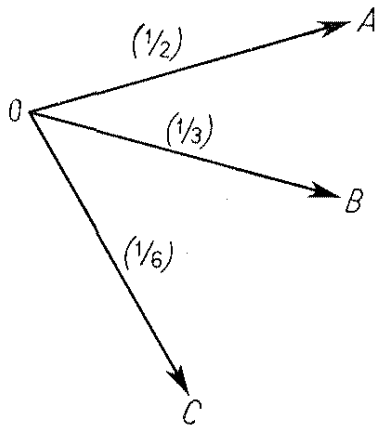
$$I/literę = -K \sum_{m \text{ wyrazów}} (1/m) \ln (1/m) = -K \ln (1/m) = K \ln m, \tag{2.4}$$

zgodnie z równaniem (2.2).

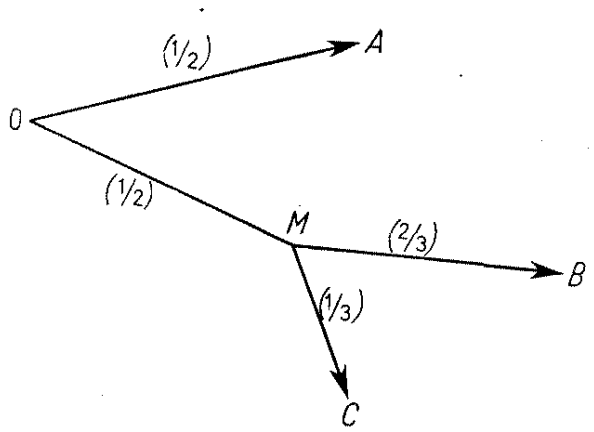
Rozpatrzmy teraz niektóre własności równania (2.1).

2.2. WŁASNOŚĆ (A)

Jeżeli wybór może być rozbity na dwa kolejne wybory, to początkowa ilość informacji powinna wyrażać się sumą ważoną ilości informacji częściowych. Rozpatrzmy przykład. Prawdopodobieństwo przejścia z punktu O do



Rys. 2.1. Prawdopodobieństwa przejścia z punktu O do punktów A, B i C



Rys. 2.2. Prawdopodobieństwa przejścia z punktu O do punktów A, B i C z pośrednim punktem M pomiędzy punktem O a punktami B i C

punktu A wynosi $^1/_2$, z O do B: $^1/_3$, a z O do C: $^1/_6$ (jak to pokazano schematycznie na rys. 2.1). Weźmy jednak pod uwagę rys. 2.2. Na rysunku tym pośredni punkt został wybrany w ten sposób, że prawdopodobieństwo przejścia z O do M jest także równe $^1/_2$. Oznacza to, że prawdopodobieństwo przejścia

z M do B musi być równe $^2/_3$, ponieważ początkowe prawdopodobieństwo przejścia z punktu O do B było równe $^1/_3$. Podobnie prawdopodobieństwo przejścia z M do C będzie równe $^1/_3$.

Dla sytuacji z rys. 2.1 ilość informacji oznaczymy przez $I(^1/_2, ^1/_3, ^1/_6)$, natomiast dla przypadku z rys. 2.2 ilość informacji dla odcinków OA i OM oznaczymy przez $I(^1/_2, ^1/_2)$, a dla MB i MC — przez $I(^2/_3, ^1/_3)$.

Całkowita ilość informacji dla rys. 2.1 i 2.2 musi być taka sama, tzn. równa $I(^1/_2, ^1/_3, ^1/_6)$. Z rysunku 2.1 mamy

$$I(^1/_2, ^1/_3, ^1/_6) = -K(^1/_2 \ln ^1/_2 + ^1/_3 \ln ^1/_3 + ^1/_6 \ln ^1/_6) \tag{2.5}$$

lub inaczej

$$I(^1/_2, ^1/_3, ^1/_6) = -K[^1/_2 \ln ^1/_2 + ^1/_2 \cdot ^2/_3 \ln(^1/_2 \cdot ^2/_3) + ^1/_2 \cdot ^1/_3 \ln(^1/_2 \cdot ^1/_3)]. \tag{2.6}$$

Rozwijając i grupując otrzymujemy

$$\begin{aligned} I(^1/_2, ^1/_3, ^1/_6) &= -K[(^1/_2 \ln ^1/_2 + ^1/_2 \ln ^1/_2) + ^1/_2(^2/_3 \ln ^2/_3 + ^1/_3 \ln ^1/_3)] = \\ &= -K(^1/_2 \ln ^1/_2 + ^1/_2 \ln ^1/_2) + ^1/_2[-K(^2/_3 \ln ^2/_3 + ^1/_3 \ln ^1/_3)] = \\ &= I(^1/_2, ^1/_2) + ^1/_2 I(^2/_3, ^1/_3). \end{aligned} \tag{2.7}$$

Wyrażenie ostateczne odpowiada sytuacji z rys. 2.2, dla której całkowita ilość informacji powinna być sumą ważoną ilości informacji odpowiadających wyborom OA, OM i MB, MC. Jest to wynik ogólny, będący konsekwencją faktu, że ilość informacji wyraża się za pomocą funkcji logarytmicznej.

2.3. WŁASNOŚĆ (B)

Własność (2.1) (B) jest nierównością, którą będziemy się posługiwać w dalszych rozważaniach. Aby otrzymać tę nierówność, rozważmy funkcję

$$\Psi = \sum_j p_j \ln q_j,$$

gdzie p_j jest prawdopodobieństwem j-tego wyboru, q_j mogą natomiast być traktowane jako inny zespół prawdopodobieństw, tak że

$$\sum_j p_j = \sum_j q_j = 1. \tag{2.8}$$

Przedstawmy q_j w postaci

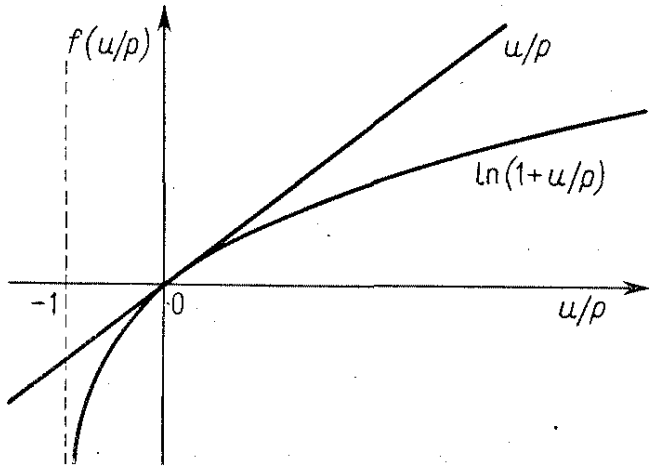
$$q_j = p_j + u_j = p_j(1+u_j/p_j).$$

Z powyższego wynika, że liczby u_j spełniają warunek

$$\sum_j u_j = 0. \tag{2.9}$$

Podstawiając q_j do Ψ otrzymujemy

$$\Psi = \sum_j p_j \ln[p_j(1+u_j/p_j)] = \sum_j p_j \ln p_j + \sum_j p_j \ln(1+u_j/p_j). \tag{2.10}$$



Rys. 2.3. Wykresy u/p i $\ln(1+u/p)$ w funkcji u/p

Jeżeli porównamy wykres funkcji $\ln(1+u/p)$ i u/p (rys. 2.3), stwierdzimy, że funkcja $\ln(1+u/p)$ jest określona wzorem

$$u/p \geqslant \ln(1+u/p). \tag{2.11}$$

Stosując wzór (2.11) do równania (2.10), otrzymujemy

$$\sum_j p_j \ln p_j + \sum_j p_j \ln(1+u_j/p_j) \leqslant \sum_j p_j \ln p_j + \sum_j p_j(u_j/p_j). \tag{2.12}$$

Korzystając z równania (2.9), otrzymujemy

$$\sum_j p_j \ln q_j \leqslant \sum_j p_j \ln p_j, \tag{2.13}$$

a po pomnożeniu tego wyrażenia przez $-K$, otrzymujemy własność (B):

$$-K \sum_j p_j \ln q_j \geqslant -K \sum_j p_j \ln p_j. \tag{2.14}$$

2.4. WŁASNOŚĆ (C)

Ilość informacji przyjmuje wartość największą, jeżeli wszystkie m prawdopodobieństw są równe:

$$p_1 = p_2 = \ldots = p_j = \ldots = p_m = 1/m. \tag{2.15}$$

Aby to udowodnić, zauważmy uprzednio, że ilość informacji jest funkcją jedynie $(m-1)$ zmiennych niezależnych, ponieważ

$$\sum_{j=1}^{j=m} p_j = 1 \quad \text{lub} \quad p_m = 1 - \sum_{j=1}^{j=m-1} p_j. \tag{2.16}$$

Wybierzmy zmienne $p_1, p_2, \ldots, p_{m-1}$ jako $(m-1)$ zmiennych niezależnych. Warunkiem* na to, by ilość informacji $I(p_1, p_2, \ldots, p_{m-1})$ osiągnęła maksimum, jest po pierwsze, by wszystkie pochodne cząstkowe pierwszego rzędu były równe zeru. W celu znalezienia drugiego warunku skontruujemy wyznacznik złożony z pochodnych drugiego rzędu, obliczonych w punkcie wyznaczonym przez pierwszy warunek:

$$\begin{vmatrix} I_{11} & I_{11} & \ldots & I_{1(m-1)} \\ I_{12} & I_{22} & \ldots & I_{2(m-1)} \\ \ldots & \ldots & \ldots & \ldots \\ I_{1(m-1)} & I_{2(m-1)} & \ldots & I_{(m-1)(m-1)} \end{vmatrix}, \tag{2.17}$$

gdzie $I_{ij} = \partial^2 I / \partial p_i \, \partial p_j = I_{ji}$, obliczone w punkcie, w którym I osiąga ekstremum.

Drugim warunkiem, aby ilość informacji I osiągnęła maksimum, jest to, aby I_{11} oraz wyznaczniki rzędu 2, 3, ..., $(m-1)$, otrzymane przez kolejne dołączanie następnego wiersza i kolumny w wyznaczniku (2.17), tworzyły ciąg znakozmienny.

Z drugiej postaci wyrażenia (2.16) mamy

$$\partial p_m / \partial p_j = -1$$

i wobec tego

$$\partial I / \partial p_j = -K(\ln p_j + 1 - \ln p_m - 1) = -K \ln (p_j / p_m).$$

---

* Por. np. W. F. Osgood, *Differential and Integral Calculus*, New York 1911, Macmillan, str. 342.

Stosując teraz pierwszy warunek, widzimy, że na to, aby wystąpiło maksimum I, musi zachodzić

$$p_j = p_m \quad \text{dla} \quad j = 1, 2, \ldots, (m-1);$$

stąd po uwzględnieniu wyrażenia (2.16) otrzymujemy

$$p_j = 1/m, \quad j = 1, 2, \ldots, m.$$

Drugie pochodne cząstkowe I, obliczone w tym punkcie, wynoszą

$$\partial^2 I/\partial p_j^2 = -K\left[\frac{1}{p_j} + \frac{1}{p_m}\right] = -2Km,$$

$$\partial^2 I/\partial p_i \partial p_j = -K(1/p_m) = -Km.$$

Aby spełnić drugi warunek, trzeba więc, aby wyznaczniki o postaci

$$\begin{vmatrix} -2Km & -Km & \ldots & -Km \\ -Km & -2Km & \ldots & -Km \\ \ldots & \ldots & \ldots & \ldots \\ -Km & -Km & \ldots & -2Km \end{vmatrix}$$

rzędu $1, 2, \ldots, (m-1)$ zmieniały znak przy wzroście rzędu. W danym przypadku jest to rzeczywiście spełnione, ponieważ łatwo wykazać, że wartość tego typu wyznacznika rzędu n jest równa

$$(-Km)^n (n+1).$$

W ten sposób udowodniliśmy własność (C).

W prostych przypadkach własność (C) może być zilustrowana graficznie. W przypadku dwóch możliwości, tzn. przy $p_1 + p_2 = 1$, mamy

$$I = -K[p_1 \ln p_1 + p_2 \ln p_2]. \tag{2.18}$$

Na rysunku 2.4 przedstawiliśmy I jako funkcję p_1 (lub p_2). I musi być równe zeru dla $p_1 = 0$ i $p_1 = 1$. Z uwagi na symetrię, I musi przyjmować maksimum dla $p_1 = {}^1/_2 = p_2$; wartość maksimum wynosi

$$I = -K\left[\frac{1}{2}\ln\frac{1}{2} + \frac{1}{2}\ln\frac{1}{2}\right] = -\left(\frac{1}{\ln 2}\right)\cdot(-\ln 2) = 1 \text{ bit}.$$

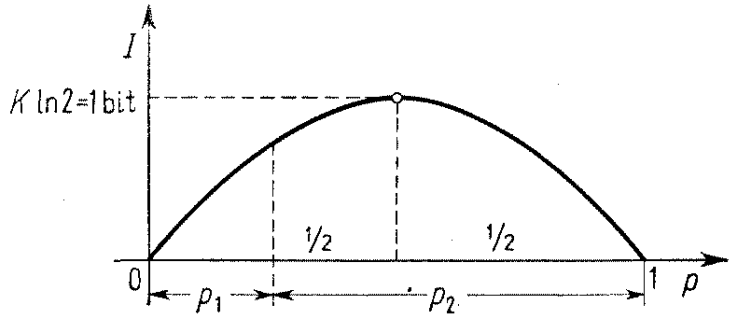
Dla trzech możliwości

$$p_1+p_2+p_3 = 1$$

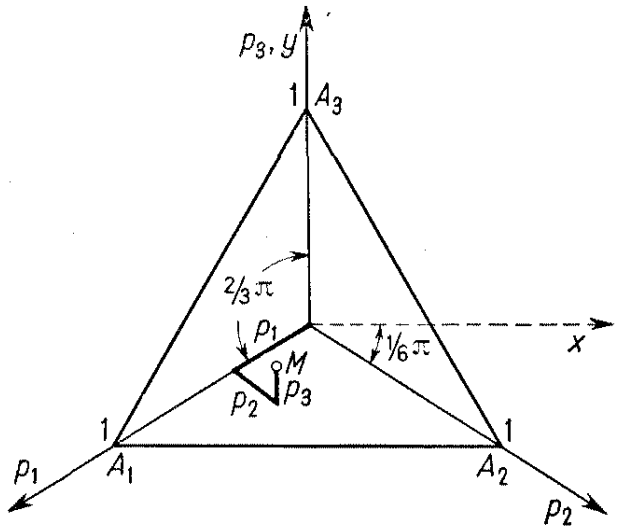
przy

$$I = -K[p_1 \ln p_1+p_2 \ln p_2+p_3 \ln p_3] \tag{2.19}$$

konstruujemy dwuwymiarowy wykres pokazany na rys. 2.5. Dwa wymiary są wystarczające, ponieważ tylko dwie, spośród trzech, wielkości p_1, p_2, p_3 są niezależne. Wybierzmy trzy kierunki na płaszczyźnie tworzące ze sobą



Rys. 2.4. Informacja jako funkcja $p_1 = 1-p_2$. Maksimum występuje dla $p_1 = p_2 = 1/2$ i wynosi $K \ln 2 = 1$ bit



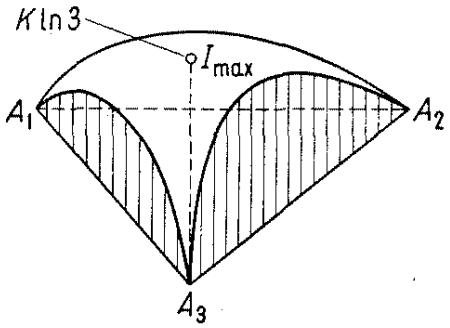
Rys. 2.5. Prawdopodobieństwom p_1, p_2, p_3 odpowiada punkt wewnątrz trójkąta $A_1A_2A_3$. Takie przyporządkowanie jest możliwe dzięki temu, że tylko dwa z trzech prawdopodobieństw są niezależne: $p_1+p_2+p_3 = 1$

kąty $2\pi/3$. Odkładając wzdłuż pierwszego kierunku odcinek o długości p_1, wzdłuż drugiego kierunku odcinek o długości p_2 i wzdłuż trzeciego — odcinek o długości p_3, otrzymamy punkt M. Z warunku (2.19) wynika, że punkt M musi leżeć wewnątrz równobocznego trójkąta A_1, A_2, A_3. Na przykład

punkt A_1 otrzymujemy, kiedy $p_1 = 1$, $p_2 = p_3 = 0$. Odcinek A_1A_2 odpowiada $p_1+p_2 = 1$, $p_3 = 0$. Można się o tym przekonać używając prostokątnych współrzędnych na płaszczyźnie i rzutując kontur $Op_1p_2p_3M$ najpierw na oś x, a następnie na oś y. W ten sposób otrzymamy współrzędne punktu M:

$$x = \sqrt{3}(-p_1+p_2)/2,$$
$$y = -(p_1+p_2)/2+p_3 = (3p_3-1)/2. \tag{2.20}$$

Równaniem odcinka A_1A_2 jest równanie $y = -1/2$, które po uwzględnieniu (2.20) daje $p_3 = 0$ i stąd $p_1+p_2 = 1$. Z uwagi na symetrię podobny wynik otrzymamy dla pozostałych boków trójkąta.



Rys. 2.6. Powierzchnia I (informacji) wykreślona jako funkcja p_1, p_2, p_3; $p_1+p_2+p_3 = 1$. Maksymalna wartość I występuje przy $p_1 = p_2 = p_3 = 1/3$ i wynosi $K \ln 3 = \ln 3/\ln 2$ bitów

W ten sposób możemy przyjąć za podstawę trójkąt $A_1A_2A_3$. Każdy jego wewnętrzny punkt odpowiada zbiorowi liczb p_1, p_2, p_3 spełniających warunek (2.19). Odkładając ilość informacji prostopadle do płaszczyzny trójkąta otrzymamy powierzchnię pokazaną w perspektywie na rys. 2.6. Wzdłuż każdego boku trójkąta otrzymamy krzywą, podobną do wykreślonej na rys. 2.4. Powierzchnia z rys. 2.6 osiąga maksimum dla $p_1 = p_2 = p_3 = 1/3$, tj. w środku trójkąta $A_1A_2A_3$.

2.5. ZDARZENIA ŁĄCZNE

Rozważmy teraz zagadnienie *zdarzeń łącznych*. Przypuśćmy, że mamy dwie zmienne losowe ([2.2]), x i y, przy czym

zmienna x może przyjmować wartość: 1, 2, 3,..., i,..., m,
zmienna y może przyjmować wartość: 1, 2, 3,..., j,..., n,

oraz że obserwujemy wartości, jakie każda z tych zmiennych faktycznie przyjmuje. Zdefiniujmy $p(i, j)$ jako prawdopodobieństwo tego, że zmienne x i y przyjmą w danej chwili wartości, odpowiednio, i oraz j. Wówczas

$$\sum_{i,j} p(i,j) = 1. \tag{2.21}$$

Jeżeli p_i jest prawdopodobieństwem zdarzenia, iż $x = i$ bez względu na wartość, jaką przyjmuje w tej obserwacji y, to

$$p_i = \sum_{j=1}^{j=n} p(i,j). \tag{2.22}$$

Podobnie prawdopodobieństwo tego, że $y = j$ (nie uwzględniając tego, jaką wartość zaobserwowano dla x), wynosi

$$p_j = \sum_{i=1}^{i=m} p(i,j). \tag{2.23}$$

Mamy zatem następujące związki:

$$\sum_{i=1}^{m} p_i = 1, \quad \sum_{j=1}^{n} p_j = 1 \quad \text{oraz} \quad \sum_{i,j} p_i \cdot p_j = 1. \tag{2.24}$$

Wprowadzimy obecnie pojęcie ilości informacji łącznej dla pary zmiennych losowych [2.3] (x, y) definiując ją wzorem

$$I(x,y) = -K \sum_{i,j} p(i,j) \ln p(i,j). \tag{2.25}$$

Ilość informacji dla samej zmiennej losowej x wynosi

$$I(x) = -K \sum_{i=1}^{m} p_i \ln p_i = -K \sum_{i,j} p(i,j) \ln p_i, \tag{2.26}$$

a dla samej zmiennej y

$$I(y) = -K \sum_{j=1}^{n} p_j \ln p_j = -K \sum_{i,j} p(i,j) \ln p_j. \tag{2.27}$$

Porównajmy sumę $I(x)+I(y)$ z $I(x, y)$. Ze wzorów (2.26) i (2.27) mamy

$$I(x)+I(y) = -K \sum_{i,j} p(i,j) \ln (p_i \cdot p_j). \tag{2.28}$$

Opierając się na własności (B) możemy łatwo udowodnić, że

$$I(x, y) \leqslant I(x)+I(y). \tag{2.29}$$

Rzeczywiście, jeżeli podstawimy zamiast q_j we wzorze (2.14) $p_i \cdot p_j$, otrzymamy

$$I(x, y) = -K \sum_{i,j} p(i, j) \ln p(i, j) \leqslant -K \sum_{i,j} p(i, j) \ln (p_i \cdot p_j) = I(x)+I(y).$$

Jest rzeczą jasną, że istnienie prawdopodobieństwa łącznego jest równoważne istnieniu ograniczenia, które, jak mówiliśmy w rozdziale pierwszym, prowadzi do zmniejszenia zawartości informacji. Równość (2.29) będzie miała miejsce tylko wtedy, gdy $p(i, j) = p_i \cdot p_j$, tzn. w przypadku, gdy zdarzenia są niezależne ([2.4]).

2.6. ILOŚĆ INFORMACJI WARUNKOWEJ

Aby zamiast nierówności (2.29) uzyskać równość, Shannon wprowadził pojęcie *ilości informacji warunkowej*, opierając się na pojęciu prawdopodobieństwa $p_i(j)$ tego, że zaobserwujemy $y = j$ przy warunku, że wiadomo już, iż zaszło $x = i$. Nazywać tu będziemy $p_i(j)$ prawdopodobieństwem warunkowym. Jest oczywiste, że

$$p_i \cdot p_i(j) = p(i, j) \quad \text{lub} \quad p_i(j) = p(i, j)/p_i. \tag{2.30}$$

Ilość informacji warunkowej jest ilością informacji związanej z prawdopodobieństwem $p_i(j)$. Oznaczymy ją symbolem $I_x(y)$ i zdefiniujemy wzorem

$$I_x(y) = -K \sum_{i,j} p(i, j) \ln p_i(j). \tag{2.31}$$

Łatwo wykazać, że $I(x, y) = I(x)+I_x(y)$.

Rozwijając prawą stronę równania (2.31) i korzystając z (2.30), mamy

$$I(x)+I_x(y) = -K[\sum_{i,j} p(i, j) \ln p_i + \sum_{i,j} p(i, j) \ln p_i(j)] =$$

$$= -K \sum_{i,j} p(i, j) \ln[p_i \cdot p_i(j)] = I(x, y); \tag{2.32}$$

następnie, łącząc wzór (2.29) i (2.32), otrzymujemy

$$I(x)+I(y) \geqslant I(x, y) = I(x)+I_x(y)$$

lub

$$I_x(y) \leqslant I(y). \tag{2.33}$$

Ilość informacji warunkowej $I_x(y)$ jest ilością informacji o zmiennej losowej ([2.5]) y przy warunku, że x jest już znane. Znajomość x nakłada ograniczenie, które obniża zawartość informacji związanej z wyborem y. Nierówność (2.32) przejdzie w równość tylko w przypadku, gdy x, y będą niezależne.

3

ROZWLEKŁOŚĆ JĘZYKA

3.1. KORELACJA I ZDARZENIA ŁĄCZNE

Aby wprowadzić pojęcie korelacji ([3.1]), zajmijmy się nieco bliżej zagadnieniami dotyczącymi zdarzeń łącznych.

Rozważaliśmy dwie zmienne, które mogą np. przybierać wartości

$$x = 1, 2, 3, \ldots, i, \ldots, m,$$
$$y = 6, 9, 8, \ldots, j, \ldots, n,$$

i zdefiniowaliśmy $p(i, j)$ jako prawdopodobieństwo zdarzenia, że zachodzi równocześnie $x = i$ oraz $y = j$. Związana z tym ilość informacji wynosi $I(x, y)$. Zdefiniowaliśmy także inne prawdopodobieństwa:

$$p_i = \sum_j p(i, j) = \text{prawdopodobieństwo tego, że } x = i, \text{ nie patrząc na to, jaką wartość przyjmuje zmienna } y$$

(z prawdopodobieństwem tym związana jest informacja $I(x)$) oraz

$$p_j = \sum_i p(i, j) = \text{prawdopodobieństwo tego, że } y = j, \text{ nie patrząc na to, jaką wartość przyjmuje zmienna } x$$

(związana z tym informacja wynosi więc $I(y)$).

W poprzednim rozdziale wykazaliśmy, że

$$I(x) + I(y) \geqslant I(x, y).$$

Nierówność ta wynikła z faktu, że zawartość informacji zmniejsza się w przypadku, gdy na system nakłada się jakieś ograniczenie.

Dla uzyskania równości wprowadziliśmy *prawdopodobieństwo warunkowe* $p_j(j)$, tzn. prawdopodobieństwo zdarzenia $y = j$ przy warunku, iż zaszło zdarzenie $x = i$. Z prawdopodobieństwem tym związana jest informacja $I_x(y)$.

Znaleziony związek pomiędzy tymi prawdopodobieństwami był następujący:

$$p_i \cdot p_i(j) = p(i, j),$$

a po zdefiniowaniu informacji warunkowej jako

$$I_x(y) = -K \sum_{i,j} p(i, j) \ln p_i(j),$$

uzyskaliśmy następującą równość:

$$I(x) + I_x(y) = I(x, y).$$

3.2. KORELACJA W JĘZYKU

Korelacja występuje we wszystkich językach*, ponieważ w przypadku wystąpienia określonej litery prawdopodobieństwo tego ([3.3]), że po niej nastąpi inna, określona litera, nie jest równe prawdopodobieństwu a priori pojawienia się tej litery ([3.4]). Jeżeli w języku angielskim wystąpi litera *t*, prawdopodobieństwo, że bezpośrednio po niej pojawi się *h*, jest znacznie większe od prawdopodobieństwa pojawienia się litery *n*. Podobnie, gdy po-

---

* Jako uzupełnienie pozycji literatury, na które będziemy się dalej powoływać, wymienimy tu następujące książki ([3.2]):

Fletcher Pratt, *Secret and Urgent*, „Blue Ribbon Books", New York 1939, 1942 (zawiera dane statystyczne o języku angielskim).

G. Dewey, *Relative Frequency of English Speech Sounds*, Cambridge, Mass. 1923, Harvard Univ. Press.

G. K. Zipf, *Human Behavior and the Principle of Least Effort*, Cambridge, Mass. 1949, Addison-Wesley. (W książce tej i w innych swoich publikacjach Zipf podaje dla różnych języków szereg krzywych w rodzaju krzywej przedstawionej na rys. 3.1).

jawi się zestaw *tio*, prawdopodobieństwo, że wystąpi n, jest bardzo duże. Tego typu korelację Shannon* określił jako *rozwlekłość* (*redundancję*).

Możemy teraz zastosować rozważania z poprzedniego rozdziału do oceny ilości informacji zawartej w słowie. Rozważania te można stosować jedynie w przypadku, kiedy korelacja występuje pomiędzy dwiema najbliższymi literami. Dla każdej pary liter związemy zmienną x z pierwszą literą, a zmienną y z drugą. Średnia ilość informacji na literę będzie więc ilością informacji warunkowej

$$I_x(y) = -K \sum_{i,j} p(i,j) \ln p_i(j) = -K \sum_{i,j} p_i \cdot p_i(j) \ln p_i(j). \tag{3.1}$$

Załóżmy, że mamy sytuację stacjonarną i prawdopodobieństwa nie zmieniają się w czasie.

Jeżeli pomiędzy literami nie występuje korelacja, będziemy mieli $p_i(j) = p_j$ oraz $\sum_i p_i = 1$. Stąd równanie (3.1) przejdzie w równanie

$$I(y) = -K \sum_j p_j \ln p_j. \tag{3.2}$$

Analiza korelacji pomiędzy pojedynczymi literami może być łatwo uogólniona na przypadek *bloków liter* w rodzaju wspomnianego *tio*. Oznaczmy blok $(N-1)$ liter symbolem $b_i(N-1)$, zaś prawdopodobieństwo wystąpienia tego zestawu — symbolem $p(b_i(N-1))$. Rozważmy następnie prawdopodobieństwo tego, że po tym zestawie pojawi się pewna litera j tworząc w ten sposób nowy blok N liter

$$b_{ij}(N) = (b_i(N-1), j).$$

Oznaczmy prawdopodobieństwo wystąpienia tego bloku, branego jako całość, przez

$$p(b_i(N-1), j).$$

Prawdopodobieństwo warunkowe $p_{b_i(N-1)}(j)$ jest prawdopodobieństwem, że j wystąpi po danym bloku $(N-1)$ liter $b_i(N-1)$. Definicja ta jest podobna

---

* C. E. Shannon, W. Weaver, *The Mathematical Theory of Communication*, Urbana, Ill. 1949, Univ. of Illinois Press, str. 3—28. Shannonowska „entropia informacji" odpowiada naszej „mierze informacji". Jak wykażemy później, miara informacji jest związana z „ujemną entropią".

C. E. Shannon, *Prediction and Entropy of Printed English*, Bell System Tech. J. **30** (1951), str. 50—64.

do definicji stosowanej w poprzednim rozdziale. Zachodzi tu również podobny związek pomiędzy prawdopodobieństwami:

$$p(b_i(N-1), j) = p(b_i(N-1)) \cdot p_{b_i(N-1)}(j). \tag{3.3}$$

Spróbujmy wyrazić teraz średnią ilość informacji na literę w ciągach, w których korelacja obejmuje N liter, wzorem

$$F_N = -K \sum_{i,j} p(b_i(N-1), j) \ln p_{b_i(N-1)}(j). \tag{3.4}$$

W równaniu tym, podobnym do równania (3.1), sumowanie po i rozciąga się na wszystkie możliwe zestawy $(N-1)$ liter.

3.3. ROZWLEKŁOŚĆ JĘZYKA

Stosując naszą definicję do typowego przykładu, jakim jest język angielski, powinniśmy otrzymać dokładną wartość średniej ilości informacji na literę, wyrażając ją granicą

$$I = \lim_{N \to \infty} F_N. \tag{3.5}$$

Wraz ze wzrostem N wzrasta ilość uwzględnianych *ograniczeń*, stąd ilość informacji maleje. Ciąg F_N musi zatem monotonicznie maleć do granicznej wartości stanowiącej rzeczywistą średnią ilość informacji na literę.

Prawdopodobieństwa, zarówno pojedynczych liter, jak i grup dwu- i trzyliterowych, zostały dla języka angielskiego określone, brak jest jednak wyczerpujących danych dla większych grup. Istniejące dane zostały zestawione poniżej; F oznacza tu informację w bitach na literę:

(1) wszystkie litery równoprawdopodobne (27 liter łącznie z przerwą) $F_0 = 4{,}76$ bitów na literę;

(2) uwzględniając prawdopodobieństwo poszczególnych liter $F_1 = 4{,}03$ bitów na literę;

(3) uwzględniając dane o grupach dwuliterowych $F_2 = 3{,}32$ bitów na literę;

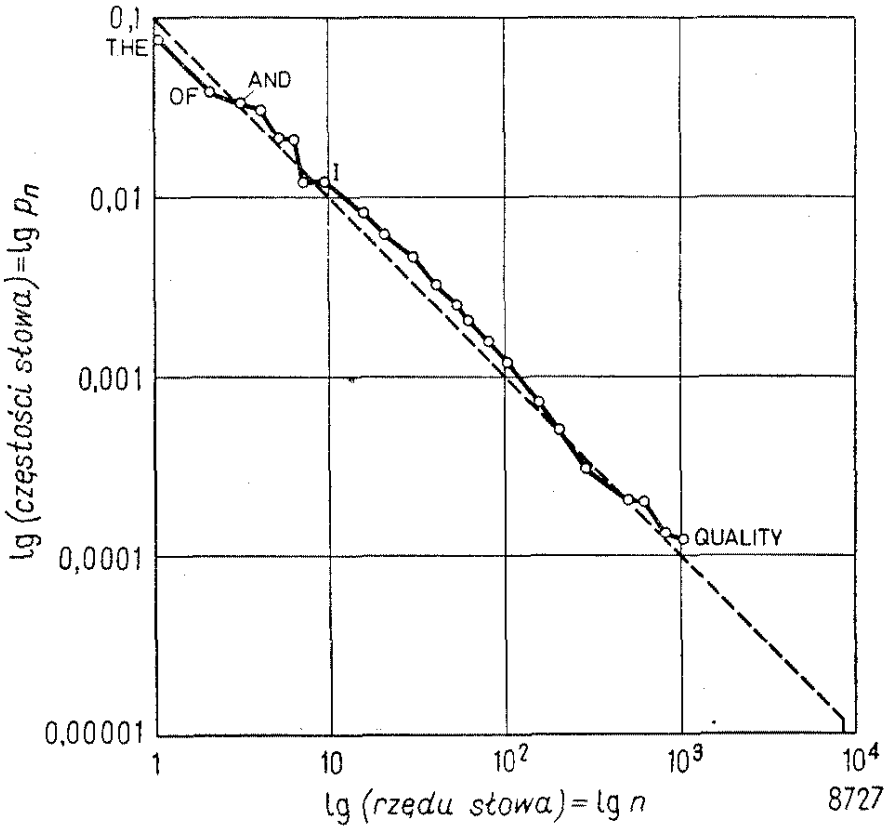
(4) uwzględniając dane o grupach trzyliterowych $F_3 = 3{,}1$ bitów na literę.

Rozwlekłość jest definiowana za pomocą F w następujący sposób:

$$R = 1 - F_{\text{lim}}/F_0. \tag{3.6}$$

Jednakże F_{lim} jest nieznane, ponieważ istnieje mało danych o zestawach więcej niż trzyliterowych. Do oceny wartości F_{lim} możemy próbować stosować

innego rodzaju statystykę, a mianowicie częstość występowania słów. Tego typu statystyka została zestawiona zarówno dla celów kodowania, jak i w związku z próbami zwiększenia sprawności nauki języka.



Rys. 3.1. Wykres (w skali podwójnie logarytmicznej) częstości słowa w zależności od rzędu słowa dla pierwszych 8727 słów

Słowa mogą być ustawione w ciąg zgodnie z malejącymi częstościami ich występowania. W ten sposób każdemu słowu można przypisać numer, jaki zajmuje ono w tym ciągu. Numer ten będziemy nazywali *rzędem słowa*. Na rysunku 3.1 pokazano zależność logarytmu częstości występowania słowa od logarytmu rzędu. Pierwszym słowem (najczęściej występującym w języku angielskim) jest „the", wykres zaś urywa się na numerze 8727, wybranym tak, by suma częstości równała się jeden. Jak można zauważyć, wykres na rys. 3.1 ma tę ciekawą własność, iż jest linią prostą w skali lg—lg. Linia ta może być opisana wzorem empirycznym

$$p_n \approx \frac{0{,}1}{n} \,. \tag{3.7}$$

Stosując wyrażenie (3.7) otrzymujemy

$$-K \sum_n p_n \ln p_n = 11{,}82 \text{ bitów na słowo.} \tag{3.8}$$

Następnie, jeżeli przyjmiemy, że średnie angielskie słowo (licząc także przerwę między słowami) składa się z 5,5 liter, to otrzymamy jako graniczną wartość $F_{\text{słowa}} = 2{,}14$ bitów na literę. Wartości tej nie można jednak utożsamiać z F_{lim}, ponieważ istnieją dodatkowe ograniczenia uwarunkowane związkami pomiędzy kolejnymi słowami.

3.4. DOŚWIADCZENIA NAD WŁASNOŚCIAMI STATYSTYCZNYMI JĘZYKA

W celu otrzymania bardziej szczegółowych danych proponuje się następującą grę. Wybiera się dowolne zdanie ([3.5]); następnie wybiera się osobę nie znającą tego zdania i żąda się od niej, aby, zaczynając od pierwszej, odgadywała kolejne litery (licząc przerwy). Po każdej próbie osobie tej daje się odpowiedź *tak* lub *nie* i grę prowadzi się dalej. Liczbę prób na każdą literę zapisuje się, a po zakończeniu gry, sumuje i dzieli przez całkowitą ilość liter w zdaniu. W ten sposób uzyskuje się liczbę bitów na literę, ponieważ każda próba, na którą odpowiadamy *tak* lub *nie*, daje jeden bit informacji.

Niektóre doświadczenia tego typu ([3.6]) były publikowane przez Shannona*. Dla pewnego zdania złożonego ze 102 liter całkowita liczba zgadywań potrzebna do odtworzenia zdania wyniosła 198 zgadywań. Wskazuje to, że na literę przypada 1,94 bitów.

W innym przykładzie grę prowadzono odmiennie. Po każdej próbie, gdy litera została odgadnięta prawidłowo, odpowiadano *tak*, a kiedy odpowiedź była błędna, podawano zgadywaną literę. W zdaniu złożonym ze 129 liter, 89 liter odgadnięto prawidłowo, natomiast w 40 próbach trzeba było podać prawidłową literę. Przyjmując, że ilość informacji przypadająca na jedną literę wynosi $F_1 = 4$ bity (co wykazano wcześniej), otrzymujemy całkowitą liczbę bitów

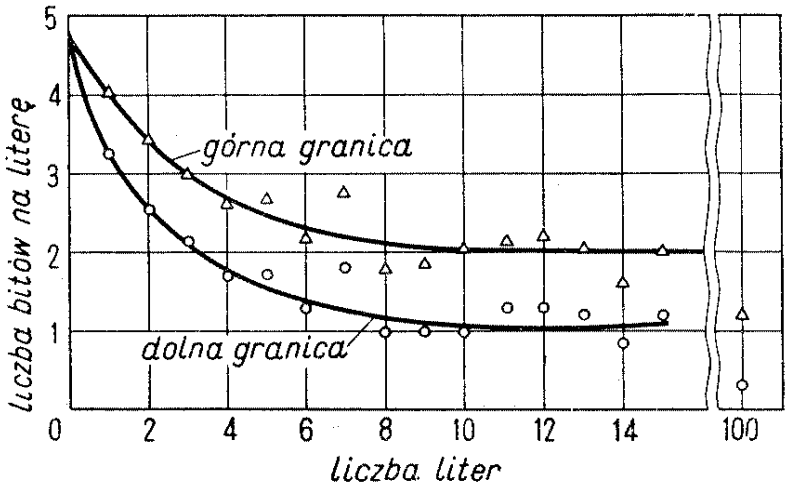
$$89 + 40 \cdot 4 = 249,$$

czyli

1,93 bitów na literę.

* C. E. Shannon, Bell System Tech. J. **30** (1951), str. 54—58.

W trzecim przykładzie otrzymano około 1,9 bitów na literę. Liczbę 1,9 powinno się jednak traktować jako górną granicę, ponieważ w wielu przypadkach litery, które trzeba było odgadnąć, były niemal oczywiste, a więc poprawne ich odgadnięcie (po którym następuje twierdząca odpowiedź) nie daje właściwie pełnego bitu informacji.



Rys. 3.2. Ilość bitów na literę jako funkcja ilości liter — do 15 liter z ekstrapolacją do 100 liter

Na rysunku 3.2, zaczerpniętym z publikacji Shannona*, przedstawiono zależność ilości informacji na literę od ilości liter, uzyskaną aż do 15 liter i dalej ekstrapolowaną do 100 liter. Jeśli przyjmiemy ekstrapolowaną wartość 1,4 bitów na literę, to porównując tę wartość z $F_0 = 4{,}76$ bitów na literę możemy stwierdzić, że język angielski cechuje znaczna rozwlekłość (redundancja). Wydaje się, że istotnie istnieje możliwość trzykrotnego skrócenia tego języka, co uczyniłoby go bardziej zwięzłym; jednakże mogą tu wejść w grę także i inne czynniki, jak zrozumiałość języka lub możliwość posługiwania się mową w obecności zakłóceń (szumów).

3.5. URZĄDZENIA KODUJĄCE ([3.7])

Rozważania przeprowadzone w poprzednich paragrafach prowadzą w sposób naturalny do zagadnień kodowania, a w szczególności do zagadnień związanych z budową urządzeń kodujących. Sugerują one od razu, iż dobre

* C. E. Shannon, Bell System Tech. J. **30** (1951), str. 54—58.

urządzenie kodujące ([3.8]), aby zmniejszyć rozwlekłość przychodzących liter, musi mieć pamięć. Inaczej mówiąc, koder musi pamiętać poprzednią literę, aby wytworzyć w sposób sprawny zespół symboli dla następnej przychodzącej litery ([3.9]). Jeżeli wprowadzimy oznaczenia:

α_n = stan pamięci w danej chwili,
x_n = następna litera,
$y_n = f(x_n, \alpha_n)$ = symbol lub zespół symboli odpowiadających literze x_n,

to pamięć musi zmienić swój stan na nowy, który będzie zależny zarówno od x_n, jak od α_n, a więc

$$\alpha_{n+1} = g(x_n, \alpha_n).$$

Wynik naszych probabilistycznych rozważań dotyczących zdarzeń łącznych został zwięźle sformułowany przez Shannona w twierdzeniu* 7:

«System kodujący może zmniejszać, a w najlepszym przypadku zachowywać ilość informacji.»

Metody kodowania rozpatrzymy w następnym rozdziale.

Urządzenia kodujące, które realizują odwracalny kod ([3.10]), mogą być użyte bądź do kodowania, bądź do dekodowania; zachowują one niezmienioną ilość informacji. Ilość ta może być obliczana dla całego słowa lub *na jednostkę czasu*, jeśli wiadomość jest dostarczana ze źródła w sposób ciągły. Rezultat Shannona stanowi pewną analogię do sytuacji w termodynamice, gdzie procesy odwracalne nie zmieniają entropii, podczas gdy w procesach nieodwracalnych entropia wzrasta.

---

* Por. C. E. SHANNON, W. WEAVER, *op. cit.*, str. 27.

4

ZASADY KODOWANIA. PROBLEMY ZWIĄZANE Z PRZEPUSTOWOŚCIĄ KANAŁU

4.1. WSTĘP

W poprzednich rozdziałach zdefiniowaliśmy ilość informacji zawartej w słowie (4.1). Przedyskutowaliśmy także rolę rozwlekłości, której skutkiem jest zmniejszenie ilości informacji przypadającej na literę w słowach danego języka. Jeżeli i jest średnią ilością informacji na literę i jeśli źródło podaje litery z szybkością m liter na sekundę, to szybkość podawania informacji wynosi

$$I' = mi \text{ bitów na sekundę.} \tag{4.1}$$

W niektórych sytuacjach może okazać się dogodne zastąpienie pierwotnego alfabetu stosowanego przez źródło innym zespołem liter, w celu zmniejszenia rozwlekłości i zredukowania liczby liter, które mają być przekazywane w jednej sekundzie. Ogólna procedura, jaką należy stosować w tego typu operacjach kodowania, została naszkicowana na końcu poprzedniego rozdziału. Jeśli kod realizowany przez urządzenie kodujące jest odwracalny, zachowuje ono stałą szybkość przepływu informacji I' i informacja nie jest tracona w procesie kodowania. Podobne urządzenie kodujące (pracujące w przeciwnym kierunku) może być użyte do dekodowania zakodowanych słów i zapisywania ich w alfabecie pierwotnym. Jeśli kod realizowany przez urządzenie kodujące nie jest odwracalny, może nastąpić strata części informacji i mogą zostać wprowadzone błędy i słowa niezrozumiałe. Stąd warunkiem, jaki przede wszystkim musimy spełnić przy kodowaniu, jest, aby kod był odwracalny.

Jak wybrać najbardziej efektywny kod? Brak jest ogólnej odpowiedzi

na to pytanie; jest ona uzależniona od przeznaczenia zakodowanych sygnałów. Przede wszystkim musimy rozważyć własności „kanału", którym sygnały mają być transmitowane (w przypadku telekomunikacji) lub magazynowane (jeśli rozpatrujemy np. zagadnienia pamięci maszyn cyfrowych).

Najlepszy będzie taki kod, który sprawnie dopasuje szybkość przepływu informacji do przepustowości kanału.

4.2. DEFINICJA KANAŁU I JEGO PRZEPUSTOWOŚCI

Jeżeli chcemy przekazać informację albo zmagazynować ją w urządzeniu pamiętającym, musimy przede wszystkim wybrać ośrodek fizyczny — kabel, łącze radiowe, rurę rtęciową, taśmę magnetyczną itp. Musimy także rozpatrzyć ograniczenia wnoszone przez fizyczne własności tego ośrodka. Na przykład, jeśli używamy kabla o paśmie przepuszczania szerokości $\Delta\nu$, najkrótszy sygnał, jaki może być przesyłany tym kablem, ma czas trwania Δt (rozdz. 8) taki, że

$$\Delta t \Delta \nu = 1. \tag{4.2}$$

Następnie wynika techniczny problem wybrania zespołu symboli dopasowanych do fizycznego ośrodka i takich, by było możliwe zastosowanie ich do komunikacji, a więc: kropek, kresek, spacji, sygnałów o różnej amplitudzie, polaryzacji itd. Jeżeli dokonamy takiego wyboru, to tym samym określimy kanał. W naszych statystycznych rozważaniach wystarczy zająć się tylko zbiorem n dyskretnych symboli $S_1, S_2, \ldots, S_j, \ldots, S_n$, które mogą być stosowane, i ich długościami (czasami trwania) $t_1, t_2, \ldots, t_j, \ldots, t_n$. Odtąd możemy pominąć fizyczną strukturę systemu przekazywania jak również techniczno-praktyczne przesłanki prowadzące do wyboru symboli S_j i ich czasu trwania t_j. Dla wszystkich teoretycznych i praktycznych rozważań *kanał* określimy po prostu przez zbiór symboli S_j o czasie trwania t_j. Możemy również nie brać pod uwagę fizycznej struktury sygnałów, bowiem tylko ich czas trwania będzie miał znaczenie dla naszych rozważań. Nie nakładamy żadnych ograniczeń na czasy trwania t_j i dopuszczamy możliwość występowania pewnej liczby sygnałów o tym samym czasie trwania.

Przyjmijmy, że symbole S_j są nawzajem całkowicie niezależne i zostały zdefiniowane w sposób, z którego nie wynika żadnego rodzaju korelacja.

Załóżmy, że dowolny ciąg symboli (4.2) może stanowić telegram. Całkowity czas trwania telegramu T jest sumą czasów trwania t_j poszczególnych symboli tworzących ten telegram.

Rozważmy wszystkie możliwe telegramy mające całkowity czas trwania T. Będą one odpowiadały wszystkim możliwym kombinacjom symboli S_j o całkowitym czasie trwania T. Oznaczmy przez $N(T)$ liczbę takich możliwych telegramów. Załóżmy, że stosowanie symboli S_j nie jest ograniczone przez żadne dodatkowe warunki czy korelacje. Wszystkie $N(T)$ telegramy o czasie trwania T uważa się za równoprawdopodobne a priori (4.3). Zatem, jeśli wybierzemy jeden z takich telegramów, to zgodnie z równaniem (1.1) otrzymamy informację

$$I = K \ln N,$$

przy czym wybór jednostek zależy tu od wyboru stałej K. Szybkość, z jaką informacja jest przekazywana przez kanał, jest równa I/T. Szybkość ta zmienia się wraz ze zmianą całkowitego czasu trwania T, można jednakże wykazać, że we wszystkich praktycznych przypadkach zbliża się ona do granicy, gdy T nieograniczenie rośnie. Shannon* przyjmuje tę granicę jako definicję przepustowości kanału

$$C = K \lim_{T \to \infty} \frac{\ln N(T)}{T}. \tag{4.3}$$

Jeżeli chcemy wyrazić informację w jednostkach entropii przyjmujemy $K = k$, gdzie k jest stałą Boltzmanna. W większości zagadnień będziemy mierzyli ilość informacji w bitach. Tak będzie, gdy przyjmiemy

$$K = \lg_2 \mathrm{e} \tag{4.4}$$

lub

$$C = \lim_{T=\infty} [\lg_2 N(T)]/T, \tag{4.5}$$

tzn. jeżeli T mierzymy w sekundach, przepustowość kanału wyrażamy w bitach na sekundę.

W następnych paragrafach tego rozdziału rozważymy sensowność powyższej definicji i warunek istnienia granicy C. Następnie wykażemy, że przepustowość kanału odpowiada maksymalnej średniej szybkości przepływu infor-

* C. E. Shannon, W. Weaver, *The Mathematical Theory of Communication*, Urbana, Ill. 1949, Univ. of Illinois Press, str. 7.

macji w kanale, jeśli wszystkie możliwe kombinacje symboli są dopuszczalne. W końcu rozważymy, jak kodować słowa, aby w pełni wykorzystać przepustowość kanału. Doprowadzi to do sformułowania zagadnienia kodowania z uwzględnieniem aspektu praktycznego. Dany jest mianowicie zbiór słów pisanych w pewnym języku, pewnym alfabetem. Za pośrednictwem kodera należy dopasować język do kanału i zbliżyć się, jak tylko jest to możliwe, do szybkości przekazywania informacji odpowiadającej przepustowości C kanału. Udowodnimy, że zagadnienie to można rozwiązać, i rozważymy niektóre ogólne warunki, jakie muszą być spełnione przez kod, aby osiągnięta była granica teoretyczna.

4.3. SYMBOLE, SŁOWA I WIADOMOŚCI PRZY KODOWANIU SEKWENCYJNYM

Przypuśćmy, że dokonaliśmy wyboru symboli i chcemy użyć kombinacji tych symboli do budowy słów lub telegramów ([4.4]). Nasze symbole mogą odpowiadać np. literom, z których mogą być złożone słowa, mogą też one reprezentować cyfry, których ciąg daje pewną liczbę. Bardziej ogólny problem obejmuje przypadek, kiedy cyfry lub litery odpowiadają całym słowom. W przypadku mowy każdemu słowu odpowiada kombinacja dźwięków podstawowych — fonemów, które grają tu rolę symboli.

Z wytworzeniem każdego symbolu związany jest pewien *koszt*. Sposób obliczenia tego kosztu zależy od fizycznego systemu przekazywania informacji. Na początku założymy, że koszt jest proporcjonalny do czasu trwania sygnału reprezentującego symbol. Tak jest w zagadnieniach telekomunikacji, gdzie wymagana jest mała moc, podczas gdy ilość przesyłanych wiadomości w czasie T ma być tak duża, jak tylko jest to możliwe. Każdy z symboli S_j ma pewien czas trwania t_j i taki sam czas trwania może mieć wiele symboli; zachodzi to wówczas, kiedy na przykład używamy impulsów o zmiennej amplitudzie lub o różnych częstotliwościach. Zachodzi to również w przypadku sygnałów telegraficznych, które mają postać kropek i przerw o tym samym czasie trwania.

Załóżmy, że na wybór możliwych ciągów symboli nie nałożono żadnych ograniczeń. Jeżeli używamy liter, pociąga to za sobą dopuszczenie jako słów czy telegramów bloków liter niemożliwych do wymówienia lub nie mających żadnego znaczenia w żadnym języku. Do oznaczania przerwy możemy wybrać

jeden spośród naszych symboli lub możemy przyjąć kombinację symboli. Uporządkujmy teraz nasze symbole zgodnie z ich kosztem (lub czasem trwania). Otrzymamy

$$\left.\begin{array}{l} a_1 \text{ symboli o czasie trwania } t_1, \\ a_2 \text{ symboli o czasie trwania } t_2, \\ a_3 \text{ symboli o czasie trwania } t_3, \quad a_j \geqslant 0. \\ a_j \text{ symboli o czasie trwania } t_j, \\ a_m \text{ symboli o czasie trwania } t_m = \sigma. \end{array}\right\} \tag{4.6}$$

Jeżeli σ oznacza czas trwania najdłuższego stosowanego przez nas symbolu, a n jest całkowitą liczbą symboli, to mamy

$$\sum_j a_j = n \quad \text{oraz} \quad 1 \leqslant t_1 < t_2 < \ldots < t_j < \ldots < t_m = \sigma. \tag{4.7}$$

Liczby a_j są nieujemnymi liczbami całkowitymi; możemy przyjąć, że czasy trwania t_j są także liczbami całkowitymi. Założenie to odpowiada sytuacji występującej w praktyce, ponieważ czas trwania może być określony tylko z pewnym błędem.

Zgodnie z programem, nakreślonym w rozdz. 4.2, chcemy obecnie obliczyć liczbę $N(t)$ różnych telegramów o całkowitym czasie trwania t. Liczba ta spełnia zależność

$$N(t) = a_1 N(t-t_1) + a_2 N(t-t_2) + \ldots + a_m N(t-t_m) = \sum_{j=1}^{j=m} a_j N(t-t_j). \tag{4.8}$$

Z zależności tej wynika, że aby utworzyć telegram o całkowitym czasie trwania t, trzeba uzupełnić telegram o czasie trwania $(t-t_1)$ którymkolwiek z a_1 symboli o czasie trwania t_1; podobnie można postępować dla $t_2, t_3, \ldots, t_m$.

Równanie (4.8) jest liniowym równaniem różnicowym*. Ma ono σ liniowo niezależnych rozwiązań szczególnych postaci

$$N_k(t) = X_k^t, \quad k = 1, 2, \ldots, \sigma, \tag{4.9}$$

rozwiązanie ogólne natomiast jest liniową kombinacją

$$N(t) = \sum_{k=1}^{k=\sigma} A_k X_k^t, \tag{4.10}$$

* Por. *Dodatek do rozdz.* 4.

ze stałymi współczynnikami A_k. Podstawiając rozwiązanie szczególne $N(t) = X^t$ do równania (4.8) widzimy, że X_k są pierwiastkami równania

$$X^t - \sum_{j=1}^{j=m} a_j X^{t-t_j} = 0$$

lub

$$1 - \sum_{j=1}^{j=m} a_j X^{-t_j} = 0. \tag{4.11}$$

Równanie to nazywa się *równaniem charakterystycznym* równania (4.8). W celu wyeliminowania ujemnych potęg X^σ możemy pomnożyć je przez X i w ten sposób otrzymamy równanie algebraiczne stopnia σ:

$$X^\sigma - \sum_{j=1}^{j=m} a_j X^{\sigma-t_j} = 0, \qquad a_j \geqslant 0. \tag{4.12}$$

W równaniu tym występuje tylko jedna zmiana znaku i stąd, zgodnie z regułą Kartezjusza, ma ono tylko jeden dodatni pierwiastek. Co więcej, wielomian staje się ujemny dla $X = 1$, ponieważ w naszym zagadnieniu występują co najmniej dwa symbole, tak że

$$\sum_{j=1}^{j=m} a_j \geqslant 2,$$

wielomian natomiast staje się dodatni dla bardzo dużych X. Dochodzimy więc do wniosku, że równanie (4.12) ma

$$\left.\begin{array}{l} \text{jeden rzeczywisty pierwiastek dodatni, większy niż 1 : } X_1 > 1, \\ \sigma - 1 \text{ ujemnych lub zespolonych pierwiastków: } X_2, X_3, \ldots, X_\sigma. \end{array}\right\} \tag{4.13}$$

4.4. OMÓWIENIE WYNIKÓW

Omówionym wyżej problemem zajmował się po raz pierwszy Shannon*, bardziej szczegółowe zaś rozważania na ten temat można znaleźć w pracy Mandelbrota**.

---

* Por. C. E. Shannon, W. Weaver, *op. cit.*, str. 5.

** B. Mandelbrot, *Contribution à la théorie mathématique des jeux de communication*, Ph. D. thesis, Paris, December 16, 1952; Publ. de l'Inst. de Statistique de l'Univ. de Paris, vol. 2, fasc. 1,2, 1953, str. 80—102; Trans. I.R.E. (P.G.I.T.) **3** (1954), str. 124.

Mamy więc rozwiązanie ogólne (4.10) zawierające σ stałych współczynników A_k. Współczynniki te są określone przez warunki początkowe, które obecnie rozważymy. Przeanalizujmy równanie (4.8) dla małych wartości t. W jednym z wyrazów w równaniu (4.8) wystąpi

$$t_j = t \quad \text{lub} \quad t - t_j = 0$$

i musimy wziąć $N(0) = 1$, gdyż jest oczywiste, że możemy użyć a_j symboli o czasie trwania t_j do reprezentowania niektórych wiadomości, jeżeli całkowity czas trwania telegramu wynosi $t = t_j$. Większe wartości $t_{j+1}, t_{j+2}, \ldots, t_m$ nie mogą być zastosowane. Stwierdzenie to jest słuszne dla wszystkich wartości j. Powyższe warunki spełnimy biorąc

$$N(0) = 1,$$

$$N(-1) = N(-2) = \ldots = N(-\sigma+1) = 0. \tag{4.14}$$

W ten sposób otrzymaliśmy σ warunków początkowych, które określają σ naszych współczynników A_k. Wartości funkcji $N(t)$ dla $t = -\sigma, -\sigma-1, \ldots$ nigdy nie występują w równaniu (4.8) i dlatego nie odgrywają one żadnej roli w naszym zagadnieniu.

Wydzielmy pierwiastek dodatni X_1 i napiszmy

$$N(t) = A_1 X_1^t + \sum_{k=2}^{k=\sigma} A_k X_k^t. \tag{4.15}$$

Pierwszy wyraz rośnie wykładniczo z t i jest dominującym składnikiem w równaniu dotyczącym $N(t)$. Pozostałe wyrazy mają charakter oscylacyjny. Jeżeli pierwiastek X_k jest ujemny, daje on na przemian dodatnie i ujemne wartości X_k^t, w miarę jak t wzrasta (t jest zawsze liczbą całkowitą). Jeżeli X_k jest pierwiastkiem zespolonym, to liczba zespolona X_k^* sprzężona z X_k jest także pierwiastkiem i ta para pierwiastków razem daje również oscylacje, gdy t wzrasta.

Rozwiązanie (4.15) jest stabilne, jeśli amplituda wszystkich członów oscylacyjnych maleje, i niestabilne, jeśli amplituda oscylacji wzrasta. Mówimy, że nasze rozwiązanie wykazuje

$$\left.\begin{array}{ll} \text{absolutną stabilność, jeśli} & |X_j| < 1 \text{ dla } j \geqslant 2, \\ \text{względną stabilność, jeśli} & |X_j| < X_1 \text{ dla } j \geqslant 2, \\ \text{niestabilność, jeśli} & |X_j| \geqslant X_1 \text{ dla co najmniej jednego } j \geqslant 2. \end{array}\right\} \tag{4.16}$$

W pierwszym przypadku amplituda oscylacji maleje; w drugim amplituda oscylacji rośnie wolniej niż wyraz dominujący.

Warunki początkowe (4.14) dają

$$N(0) = A_1+A_2+\ldots+A_\sigma = 1. \tag{4.17}$$

W dwóch pierwszych przypadkach (absolutnej i względnej stabilności), w wielu zagadnieniach możemy zaniedbać wyrazy oscylacyjne i zostawić tylko wyraz dominujący równania (4.15):

$$N'(t) = A_1X_1^t = A_1\,e^{\beta t}, \quad \text{gdzie} \quad \beta = \ln X_1. \tag{4.18}$$

$N(t)$ jest liczbą telegramów o całkowitym czasie trwania t. Nas jednakże może interesować liczba $N_1(t)$ telegramów o czasie trwania mniejszym lub równym t. Mając na uwadze, że t jest zawsze liczbą całkowitą, mamy

$$N_1(t) = \sum_{s=1}^{s=t} N(s) = \sum_{s=1}^{s=t}\sum_{k=1}^{k=\sigma} A_kX_k^s = \sum_{k=1}^{k=\sigma} A_k\left(\sum_{s=1}^{s=t} X_k^s\right). \tag{4.19}$$

Suma nie obejmuje przypadku $s = 0$, ponieważ telegram o długości zero nie ma żadnego sensu. Wykorzystując związek

$$\sum_{s=1}^{s=t} X_k^s = X_k(1+X_k+X_k^2+\ldots+X_k^{t-1}) = X_k\frac{1-X_k^t}{1-X_k} \tag{4.20}$$

mamy

$$N_1(t) = \sum_{k=1}^{k=\sigma} A_kX_k\frac{1-X_k^t}{1-X_k} = \sum_{k=1}^{k=\sigma} B_k(X_k^t-1), \tag{4.21}$$

gdzie

$$B_k = A_kX_k/(X_k-1) = A_k/(1-X_k^{-1}).$$

Wynik ten możemy napisać w postaci

$$N_1(t) = B_1X_1^t + \sum_{k=2}^{k=\sigma} B_kX_k^t - B_0, \tag{4.22}$$

gdzie

$$B_0 = \sum_{k=1}^{k=\sigma} B_k.$$

Liczba N_1 telegramów o czasie trwania krótszym lub równym t zawiera wyraz dominujący z X_1^t, wyrazy oscylujące ($k = 2, 3, \ldots, \sigma$) i stałą B_0. Pomijając wyrazy oscylacyjne otrzymujemy wyrażenie dla $N_1'(t)$:

$$N_1'(t) = B_1 X_1^t - B_0 = B_1 \, e^{\beta t} - B_0, \tag{4.23}$$

będzie to, w przypadkach stabilnych, dobre przybliżenie dla dużych t. Z równania (4.23) wynika dalej, że

$$t = \frac{1}{\beta} \ln \frac{N_1' + B_0}{B_1} = t_0 + \frac{1}{\beta} \ln (N_1' + B_0), \tag{4.24}$$

gdzie

$$t_0 = -\frac{1}{\beta} \ln B_1.$$

4.5. PRZYKŁADY

Dla zilustrowania metody obliczania $N(t)$ podamy teraz kilka przykładów.

4.5.1. Wszystkie symbole mają jednakową długość

Przyjmiemy dla tego przypadku $a_1 = n$, gdzie n jest całkowitą liczbą symboli o długości $t_1 = \sigma = 1$. Równanie (4.12) przyjmuje postać

$$X - n = 0. \tag{4.25}$$

Mamy tylko jeden rzeczywisty dodatni pierwiastek $X = n$ i stąd

$$N = n^t \quad \text{oraz} \quad N_1 = (n^t - 1)/(1 - n^{-1}).$$

Przypadek ten wykazuje absolutną stabilność — bez oscylacji.

4.5.2. Symbole o dwu różnych długościach: 1 i σ

(1) Stopień równania charakterystycznego jest równy σ. W przypadku ogólnym mamy

$$X^\sigma - a_1 X^{\sigma-1} - a_\sigma = 0. \tag{4.26}$$

(2) Najprostszy przypadek wystąpi dla $\sigma = 2$:

$$X^2 - a_1 X - a_2 = 0. \tag{4.27}$$

Równanie to ma pierwiastki

$$\left.\begin{matrix} X_1 \\ X_2 \end{matrix}\right\} = {}^1/_2(a_1 \pm \sqrt{a_1^2+4a_2}). \tag{4.28}$$

Ponieważ $|X_2| < X_1$, możemy być pewni, że względna stabilność jest zapewniona. Warunek absolutnej stabilności w tym przypadku ma postać

$$|X_2| = -X_2 < 1 \quad \text{lub} \quad a_1 > a_2 - 1. \tag{4.29}$$

Warunek ten jest spełniony np. przez kanał z dwoma tylko symbolami:

$$a_1 = a_2 = 1, \quad X_1 = \frac{1+\sqrt{5}}{2} = 1{,}62,$$

$$X_2 = \frac{1-\sqrt{5}}{2} = -0{,}62. \tag{4.30}$$

(3) Rozważmy przypadek, kiedy długie sygnały są trzykrotnie dłuższe niż krótkie:

$$t_1 = 1, \quad t_2 = 2, \quad t_3 = 3, \quad \sigma = 3,$$
$$a_1 < 0, \quad a_2 = 0, \quad a_3 > 0.$$

Równanie charakterystyczne ma postać

$$X^3 - a_1 X^2 - a_3 = 0. \tag{4.31}$$

Bierzemy

$$Y = X^{-1},$$

stąd mamy

$$Y^3 + pY + q = 0, \text{ gdzie } p = a_1/a_3, \text{ natomiast } q = -1/a_3. \tag{4.32}$$

Jest to dobrze znane równanie trzeciego stopnia*. Ponieważ wyrażenie $27q^2 + 4p^3$ jest dodatnie, istnieje jeden pierwiastek dodatni i dwa zespolone:

$$Y_1 = A + B, \quad Y_2 = \omega A + \tilde{\omega} B, \quad Y_3 = \tilde{\omega} A + \omega B, \tag{4.33}$$

---

* I. S. Sokolnikoff, E. S. Sokolnikoff, *Higher Mathematics for Engineers and Physicists*, New York 1941, Mc Graw-Hill, str. 86.

gdzie

$$\omega = (-1+\sqrt{3}i)/2, \quad \tilde{\omega} = (-1-\sqrt{3}i)/2, \quad \omega^2 = \tilde{\omega},$$

$$A^3 = \frac{1}{2}\left(-q + \sqrt{q^2 + \frac{4p^3}{27}}\right), \quad B^3 = \frac{1}{2}\left(-q - \sqrt{q^2 + \frac{4p^3}{27}}\right).$$

Wybieramy dla A i B rzeczywiste pierwiastki trzeciego stopnia, tak że Y_1 jest liczbą rzeczywistą, podczas gdy Y_2 i Y_3 są liczbami zespolonymi sprzężonymi.

Stabilność względna wystąpi, jeśli

$$|X_2| = |X_3| < X_1 \quad \text{lub} \quad |Y_2| = |Y_3| > Y_1. \tag{4.34}$$

Łatwo można obliczyć

$$|Y_2|^2 = Y_2\tilde{Y}_2 = Y_2Y_3 = A^2+B^2-AB,$$

$$Y_1^2 = A^2+B^2+2AB,$$

lecz

$$AB = -p/3 < 0.$$

Tak więc warunek względnej stabilności (4.34) jest spełniony. Aby wystąpiła absolutna stabilność trzeba, by

$$|Y_2|^2 = |Y_3|^2 > 1; \tag{4.35}$$

warunek ten jest spełniony, kiedy $|A|$ i $|B|$ nie są zbyt małe. Mamy

$$A^3+B^3 = (A+B)(A^2+B^2-AB) = -q = 1/a_3;$$

stąd

$$Y_1|Y_2|^2 = 1/a_3.$$

Ponieważ Y_1 jest mniejsze niż 1 ($X_1 > 1$), z warunku (4.35) wynika łatwo spełniony warunek

$$a_3 Y_1 < 1. \tag{4.36}$$

4.5.3. Każdy symbol ma inną długość

W tym przypadku mamy

$$a_1 = a_2 = \ldots = a_\sigma = 1$$

oraz odpowiednio

$$t_1 = 1, \quad t_2 = 2, \quad \ldots, \quad t_\sigma = \sigma.$$

Równanie charakterystyczne (4.12) przyjmuje postać

$$X^\sigma - X^{\sigma-1} - X^{\sigma-2} - \ldots - X - 1 = 0 \tag{4.37}$$

lub

$$X^\sigma - \frac{X^\sigma - 1}{X - 1} = 0.$$

Mnożąc przez $X-1$, otrzymujemy

$$X^{\sigma+1} - 2X^\sigma + 1 = 0. \tag{4.38}$$

Równanie powyższe ma pierwiastek obcy $X = 1$. Równanie (4.37) ma tylko jeden pierwiastek dodatni, leżący pomiędzy 1,62 (przypadek $\sigma = 2$, równanie (4.30)) a 2 (dla $\sigma \to \infty$). Wychodząc z równania (4.38), napiszemy

$$1{,}62 \leqslant X_1 \leqslant 2, \qquad X_1 = 2 - \frac{1}{X_1^\sigma}. \tag{4.39}$$

Jeśli σ jest parzyste, to istnieje jeden pierwiastek ujemny, kiedy zaś σ jest nieparzyste, pierwiastków ujemnych nie ma. Można to łatwo sprawdzić stosując do równania (4.38) regułę znaków Kartezjusza:

$$X' = -X \begin{cases} \sigma \text{ parzyste:} & -X'^{\sigma+1} - 2X'^\sigma + 1 = 0, \text{ jedna zmiana znaku,} \\ \sigma \text{ nieparzyste:} & X'^{\sigma+1} + 2X'^\sigma + 1 = 0, \text{ nie ma zmiany znaku.} \end{cases} \tag{4.40}$$

Wszystkie pozostałe pierwiastki są zespolone. Jeżeli σ jest parzyste, ujemny pierwiastek leży w przedziale pomiędzy $X' = 0$ oraz $X' = 1$ (co odpowiada $X = -1$) i dlatego nie wpływa on na stabilność rozwiązania. Dla dużych wartości σ pierwiastki zespolone są bliskie pierwiastkowi stopnia σ z jedności. Napiszmy

$$X = \mathrm{e}^{(2n\pi i/\sigma)+\varepsilon}, \tag{4.41}$$

gdzie n jest liczbą całkowitą, ε natomiast jest małe. Ta wartość X spełnia równanie (4.38), jeśli ε wyraża się w przybliżeniu wzorem

$$\mathrm{e}^{\sigma\varepsilon} \approx 1/(2 - \mathrm{e}^{2n\pi i/\sigma}), \qquad \sigma \gg 1.$$

Wyrażenie to daje małą wartość ε, jeśli σ jest duże. Pierwiastki zespolone będą więc miały wartość modułu bliską jedności i warunek względnej stabilności będzie spełniony.

4.6. OBLICZANIE PRZEPUSTOWOŚCI KANAŁU

Obecnie jesteśmy w stanie znaleźć przepustowość kanału zdefiniowaną wzorem (4.3). Przyjmijmy, że mamy do czynienia z przypadkiem stabilnym. Dla bardzo długich T możemy posługiwać się wyrazem dominującym (równanie (4.18)):

$$N(T) \longrightarrow A_1 X_1^T = A_1 \, e^{\beta T} \quad \text{przy} \quad T \longrightarrow \infty \tag{4.42}$$

i stąd

$$\ln N \longrightarrow \beta T + \ln A_1, \quad \text{gdzie} \quad \beta = \ln X_1.$$

Stosując teraz (4.3), otrzymujemy wzór na przepustowość kanału

$$C = \lim_{T \to \infty} \left(K\beta + \frac{K \ln A_1}{T} \right) = K\beta. \tag{4.43}$$

Shannon rozpatruje także zagadnienia, w których na dopuszczalne ciągi symboli S_i nakłada się pewne ograniczenia. My, jednakże, w obecnych rozważaniach zajmiemy się poprzednim zagadnieniem, bez żadnych ograniczeń.

Po zdefiniowaniu przepustowości kanału Shannon wprowadza swoją definicję miary ilości informacji, którą już podaliśmy w poprzednich rozdziałach. Następnie udowadnia on fundamentalne twierdzenie* dla kanału bezszumowego:

Jeśli kanał ma przepustowość C bitów na sekundę i otrzymuje wiadomości ze źródła informacji podawane z szybkością i bitów na symbol, to najlepszy z możliwych kodów pozwoli nam przesyłać przez kanał informację z szybkością $R = C/i$ symboli na sekundę.

Dowód tego twierdzenia wyniknie z rozważań, które przeprowadzimy poniżej.

4.7. DOPASOWANIE KODU DO KANAŁU

Zamierzamy obecnie zająć się dokładniej fizycznym znaczeniem fundamentalnego twierdzenia Shannona i określić własności kodu, który umożliwia pełne wykorzystanie przepustowości kanału. Wykażemy, że najlepszy jest taki kod, który prowadzi do najbardziej prawdopodobnego rozkładu symboli; podamy także reguły znajdowania takich optymalnych systemów kodowania.

* C. E. Shannon, W. Weaver, *op. cit.*, Sec. 9, str. 28.

Przypomnijmy najpierw niektóre wyniki z poprzednich rozdziałów. Założyliśmy, że w naszym systemie komunikacyjnym używamy n różnych symboli, i mogliśmy obliczyć całkowitą liczbę telegramów, to znaczy różnych kombinacji symboli, jeśli liczby $N_1, N_2, \ldots, N_j, \ldots, N_n$ powtórzeń tych różnych symboli były znane. Niech

$$G_0 = N_1+N_2+\ldots+N_n = \sum_{j=1}^{j=n} N_j \tag{4.44}$$

będzie całkowitą liczbą symboli, z których utworzone są telegramy, oraz

$$p_j = N_j/G_0, \quad \text{gdzie} \quad \sum_{j=1}^{j=n} p_j = 1, \tag{4.45}$$

oznacza względną częstość występowania j-ego symbolu. Całkowita liczba różnych telegramów tego typu jest określona przez wzór (1.26):

$$P = G_0!/\prod_j N_j!.$$

Stosując wzór Stirlinga otrzymujemy w przybliżeniu

$$\ln P = -G_0 \sum_{j=1}^{j=n} p_j \ln p_j. \tag{4.46}$$

Wzór ten był stosowany do określania ilości informacji, uzyskiwanej przy wyborze jednego telegramu z P możliwych (wzór (1.28)) i stąd mieliśmy informację na symbol

$$i = I/G_0 = -K \sum_{j=1}^{j=n} p_j \ln p_j. \tag{4.47}$$

4.7.1. Symbole o równej długości (jednakowym czasie trwania)

Na wstępie rozważmy prosty przykład, przyjmując, że wszystkie symbole mają jednakową długość t_1. Długi odcinek czasu T zawiera

$$G = T/t_1$$

pozycji (lub komórek) dla różnych symboli. W każdej pozycji możemy umieścić jeden z n różnych symboli i stąd całkowita liczba możliwych telegramów jest równa

$$N(T) = n^G = n^{T/t_1}. \tag{4.48}$$

Z definicji (4.3) otrzymamy teraz przepustowość takiego kanału:

$$C = (K \ln n)/t_1. \tag{4.48a}$$

W naszych obecnych rozważaniach dotyczących przepustowości kanału nie zakładamy znajomości częstości a priori występowania n różnych symboli. Kiedy definiowaliśmy informację, założyliśmy, że znamy częstości a priori symboli $p_1, p_2, \ldots, p_j, \ldots, p_n$, i otrzymaliśmy ilość informacji (na symbol) określoną przez równanie (4.47). Ilość informacji osiąga maksimum, jeżeli wszystkie p_j są sobie równe:

$$p_1 = p_2 = p_3 = \ldots = p_j = \ldots = p_n = 1/n$$

i to maksimum ma wartość

$$i_{\max} = -K \sum_{j=1}^{j=n} \frac{1}{n} \ln \frac{1}{n} = K \ln n,$$

ponieważ w sumie mamy n równych wyrazów. Jest to ilość informacji na symbol w przypadku, kiedy wszystkie symbole są równoprawdopodobne. Każdy symbol ma czas trwania t_1 i stąd szybkość, z jaką informacja jest przekazywana, wynosi

$$i/t_1 = (K \ln n)/t_1 \text{ bitów na sekundę.} \tag{4.49}$$

Jest to wzór identyczny ze wzorem (4.48a) dla przepustowości kanału. Podobieństwo pomiędzy wzorami (4.48) i (4.49) jest uderzające. Przy obliczaniu przepustowości kanału używaliśmy wszystkich możliwych rozkładów symboli. W przypadku wzoru (4.49) założyliśmy pewien dowolny zespół gęstości a priori, następnie zaś, w celu maksymalizacji informacji, przyjęliśmy $p_j = 1/n$ i w ten to sposób otrzymaliśmy identyczny wynik. Ze wzorów (4.48a) i (4.49) otrzymujemy szybkość przekazywania symboli

$$R = C/i = 1/t_1. \tag{4.50}$$

Istotnie nie możemy przekazywać więcej niż $1/t_1$ symboli o czasie trwania t_1 (sekund) na sekundę. Stąd w naszym prostym przykładzie najlepszym sposobem kodowania jest taki sposób, przy którym wszystkie symbole, zgodnie z twierdzeniem Shannona, występują w równych proporcjach*.

* Tzn. równie często. (*Przyp. tłum.*).

Zauważmy także, że warunek, by ilość informacji była maksymalna, daje najbardziej prawdopodobny rozkład symboli. Istotnie $N(t)$ jest całkowitą liczbą możliwych telegramów, zaś P, gdzie

$$\ln P = -\sum_{j=1}^{j=n} p_j \ln p_j = i/K$$

jest liczbą możliwych telegramów o rozkładzie symboli określonym przez prawdopodobieństwa p_j. Maksymalizacja $\ln P$ (lub ilości informacji i) jest równoważna maksymalizacji P. Jednakże P/N (gdzie N uważamy za stałą) jest prawdopodobieństwem wystąpienia rozkładu P (określonego przez p_j) i rozkład jest najbardziej prawdopodobny, jeżeli P osiąga maksimum.

W przypadku uproszczonym, jaki rozpatrujemy w tym paragrafie (wszystkie symbole mają taką samą długość), doszliśmy do warunku, że najbardziej prawdopodobny rozkład daje maksimum ilości informacji. Można także powiedzieć, że rozkład ten jest rozkładem asymptotycznym, co wynika z równania (4.48a), przy wyprowadzaniu którego nie uwzględnialiśmy rozkładu a priori. Ten asymptotyczny rozkład może być uważany za rozkład średni w tym znaczeniu, że przy uśrednianiu za długi okres czasu odgrywa rolę tylko rozkład asymptotyczny. W następnym rozdziale zobaczymy, że uwagi te są słuszne również w przypadku, kiedy symbole nie mają tego samego czasu trwania.

4.8. ZAGADNIENIE OGÓLNE — SYMBOLE O RÓŻNYCH DŁUGOŚCIACH

Chcemy obecnie rozważyć zagadnienie ogólne, z symbolami o różnych długościach; udowodnimy również, że nasz poprzedni wniosek, a mianowicie, że najbardziej efektywny jest kod, przy którym rozkład a priori częstości różnych symboli jest najbardziej prawdopodobny, pozostaje słuszny. Następnie będziemy chcieli znaleźć taki właśnie rozkład i rozpatrzyć związek zachodzący między nim a zależnościami (4.11) i (4.12).

Rozważania dla symboli o różnych długościach są nieco bardziej złożone niż dla symboli o równej długości. Załóżmy najpierw, że mamy do czynienia z telegramem o całkowitym czasie trwania T, oraz przyjmijmy, że będziemy stosować $N_1, N_2, \ldots, N_j, \ldots, N_n$ symboli, odpowiednio, typu 1, 2, ..., j, ..., n.

Całkowity czas trwania wiadomości wynosi

$$T = \sum_{j=1}^{j=n} N_j t_j. \tag{4.51}$$

Wprowadźmy całkowitą liczbę symboli G_0 i względne częstości występowania p_j:

$$G_0 = \sum_{j=1}^{j=n} N_j, \qquad p_j = N_j/G_0; \tag{4.52}$$

stąd

$$\sum_{j=1}^{j=n} p_j = 1. \tag{4.53}$$

Wzór (4.51) możemy zapisać w następującej postaci:

$$T = G_0 \sum_{j=1}^{j=n} p_j t_j. \tag{4.54}$$

Logarytm całkowitej liczby P różnych wiadomości, które możemy skonstruować przez zmianę pozycji różnych symboli, wyraża się w dalszym ciągu wzorem (4.46):

$$\ln P = -G_0 \sum_{j=1}^{j=n} p_j \ln p_j. \tag{4.55}$$

Informacja uzyskana przy wyborze jednego z rozpatrywanych telegramów (a jest ich P), zgodnie z naszą definicją (1.1), jest określona wzorem $K \ln P$. Wszystkie te telegramy mają ten sam czas trwania T, ponieważ składają się z takich samych liczb $N_1, N_2, \ldots, N_n$ różnych symboli; przyjmiemy również, że są one równoprawdopodobne a priori. Chcemy teraz znaleźć najbardziej prawdopodobny rozkład, co jest równoznaczne maksymalizacji ilości informacji przypadającej na telegram. Uzyskamy to, maksymalizując $\ln P$ oraz uwzględniając warunki (4.53) i (4.54). Można to uczynić stosując metodę mnożników Lagrange'a. Warunek wystąpienia ekstremum ma postać*

$$\mathrm{d}(\ln P) - \alpha\, \mathrm{d}\left(\sum_{j=1}^{j=n} p_j\right) - \beta\, \mathrm{d}T = 0, \tag{4.56}$$

* Warunek (4.56) jest warunkiem istnienia ekstremum $\ln P$. Dowód, czy jest to minimum, czy maksimum, może być przeprowadzony metodą stosowaną przy dowodzie własności (C) (rozdz. 2.3).

gdzie α i β są dowolnymi mnożnikami, które określimy później. Zmiennymi są tu $G_0, p_1, p_2, \ldots, p_j, \ldots, p_n$. Różniczkując równanie (4.55) otrzymujemy

$$\mathrm{d}(\ln P) = -(\mathrm{d}G_0) \sum_{j=1}^{j=n} p_j \ln p_j - G_0 \sum_{j=1}^{j=n} (1+\ln p_j)\, \mathrm{d}p_j.$$

Jeżeli podstawimy to wyrażenie do wzoru (4.56), to otrzymamy

$$-\mathrm{d}G_0 \left[\sum_{j=1}^{j=n} p_j \ln p_j + \beta \sum_{j=1}^{j=n} p_j t_j\right] + \sum_{j=1}^{j=n} \mathrm{d}p_j[-\alpha - \beta G_0 t_j - G_0(1+\ln p_j)] = 0. \tag{4.57}$$

Różniczki $\mathrm{d}G_0$ i $\mathrm{d}p_j$ musimy teraz traktować jako dowolne, niezależne wielkości. Różniczka zupełna $\mathrm{d}(\ln P)$ może być równa zeru tylko wtedy, gdy oba wyrażenia w nawiasach kwadratowych będą równe zeru, a więc jeżeli

$$\sum_{j=1}^{j=n} p_j (\ln p_j + \beta t_j) = 0, \tag{4.58}$$

$$\ln p_j = -1 - \beta t_j - \alpha/G_0. \tag{4.59}$$

Po podstawieniu równania (4.59) do (4.58) otrzymamy

$$\sum_{j=1}^{j=n} p_j(-1-\alpha/G_0) = 0.$$

Z uwagi na warunek (4.53) musimy przyjąć

$$1+\alpha/G_0 = 0 \quad \text{lub} \quad \alpha = -G_0. \tag{4.60}$$

Wobec tego równanie (4.59) sprowadza się do

$$\ln p_j = -\beta t_j \quad \text{lub} \quad p_j = \mathrm{e}^{-\beta t_j}. \tag{4.61}$$

Stała dowolna β jest określona przez warunek (4.53):

$$\sum_{j=1}^{j=n} p_j = \sum_{j=1}^{j=n} \mathrm{e}^{-\beta t_j} = 1. \tag{4.62}$$

Oznaczmy

$$X = \mathrm{e}^{\beta}.$$

Należy przy tym pamiętać, że w telegramie może występować więcej niż jeden symbol o tej samej długości.

Załóżmy, że występuje a_j symboli o czasie trwania t_j. Równanie (4.62) można wówczas zapisać w postaci

$$\sum_{j=1}^{j=n} a_j X^{-t_j} = 1. \tag{4.63}$$

Wyrażenie to jest identyczne z wyrażeniem (4.11) z rozdz. 4.3. Jeżeli będziemy rozpatrywać telegramy nieskończenie długie, dojdziemy ponownie do równoważności pomiędzy rozkładem asymptotycznym z rozdz. 4.3 a rozkładem odpowiadającym maksimum informacji, który jest zarazem (por. uzasadnienie w rozdz. 4.7) rozkładem najbardziej prawdopodobnym, jak również średnim. Rozkład (4.61) daje liczbę telegramów (4.55):

$$\ln P = -G_0 \sum_{j=1}^{j=n} p_j \ln p_j = G_0 \sum_{j=1}^{j=n} \beta t_j \, \mathrm{e}^{-\beta t_j} = G_0 \beta \sum_{j=1}^{j=n} t_j p_j \tag{4.64}$$

lub

$$\ln P = \beta T \quad \text{lub} \quad P = \mathrm{e}^{\beta T} = X^T$$

zgodnie z równaniem (4.54) oraz definicją X. Nasze P, najbardziej odpowiadające prawdopodobnemu rozkładowi, jest znowu identyczne z wielkością $N(T)$, odpowiadającą wszystkim możliwym rozkładom, przepustowość kanału zaś zgodnie ze wzorem (4.43) jest równa

$$C = K \ln X = K\beta. \tag{4.65}$$

Dla długich telegramów rozkład najbardziej prawdopodobny jest tak dalece bardziej prawdopodobny niż inne rozkłady, że te inne rozkłady mogą być po prostu pominięte. Najlepszym kodem jest kod, przy którym zostaje osiągnięty omawiany najbardziej efektywny rozkład symboli, mianowicie taki, przy którym zgodnie z równaniem (4.61) mamy:

$$(\ln p_j)/t_j = -\beta \quad \text{(stała)}. \tag{4.66}$$

Największy rzeczywisty pierwiastek β równania (4.62) odpowiada największemu rozwiązaniu X równania (4.12) i daje, zgodnie z (4.64), najbardziej prawdopodobny rozkład, z którym związana jest największa wartość P.

Jeżeli stosujemy najbardziej efektywny kod i rozkład symboli jest najbardziej prawdopodobny, a prawdopodobieństwa różnych symboli dane są wzorami (4.61), to ilość informacji na symbol wynosi

$$i = -K \sum_{j=1}^{j=n} p_j \ln p_j = -K\beta \sum_{j=1}^{j=n} t_j p_j = -K\beta \bar{t}, \tag{4.67}$$

gdzie $\bar{t}$ jest średnim czasem trwania symbolu. Szybkość przekazywania, zgodnie z równaniami (4.65) i (4.67), jest równa

$$R = C/i = 1/\bar{t}; \tag{4.68}$$

zauważmy, że wzór ten jest bardzo podobny do wzoru (4.50).

4.9. ZAGADNIENIE DOPASOWANIA

Kod, przy którym dopasowanie źródła słów do kanału jest najbardziej efektywne, jest określony przez warunek (4.61) lub (4.66). Warunek ten jest tak ważny w zagadnieniach kodowania jak warunek dopasowania impedancji charakterystycznych w zagadnieniach teorii obwodów. Shannon zajmując się pewnym szczególnym zagadnieniem posługuje się metodą prowadzącą do bardzo podobnej reguły*. Udowodniliśmy, że warunek dopasowania jest ogólny i słuszny dla wszystkich zagadnień kodowania. Specjalne przykłady i zastosowanie warunku dopasowania kodu do kanału rozważymy w następnym rozdziale.

Zaznaczmy na koniec, że pokrywanie się rozkładu średniego z rozkładem najbardziej prawdopodobnym jest dobrze znanym efektem występującym w przypadku systemów mogących przyjmować bardzo dużą ilość stanów. Pokrywanie się tych dwu rozkładów odgrywa niezmiernie ważną rolę w termodynamice statystycznej, gdzie tłumaczy zbieżność różnych definicji entropii, podanych przez J. W. Gibbsa, Maxwella, Boltzmanna i Plancka. Zagadnienie to zostało dokładnie rozpatrzone przez H.A. Lorentza w *Leçons au College de France* (*Les theories statistiques en thermodynamique*, Leipzig 1916, Teubner).

4.10. ZAGADNIENIA STATYSTYKI SŁÓW (MANDELBROT)

Wnioski, które podaliśmy w rozdz. 4.3, 4.4, 4.7, 4.8, Mandelbrot** zastosował przy badaniu statystyki słów języków naturalnych ([4.5]). Zagadnienie to było rozważane w poprzednim rozdziale z empirycznego punktu widzenia.

* C. E. Shannon, W. Weaver, *op. cit.*, str. 29—30.

** Por. notka ** na str. 59.

Jaka jest rola słowa w języku naturalnym? Znaczenie słowa można znaleźć w słowniku. Przedstawia ono rodzaj specjalnej wiadomości (inaczej informacji w sensie potocznym). W przypadku pisania — wiadomość ta jest kodowana przy pomocy liter, natomiast w przypadku mówienia — przy pomocy fonemów ([4.6]). W jaki sposób jest ona kodowana w mózgu? Jest to oczywiście problem otwarty, możemy jednak założyć pewien rodzaj kodowania oparty na elementarnych symbolach lub sygnałach, z których każdy charakteryzuje się określonym kosztem. Założenie takie pozwala nam stosować poprzednią analizę: musimy po prostu zamiast *czasem trwania* posługiwać się *kosztem*. Dalej, przyjmiemy, że kod jest, praktycznie biorąc, dopasowany do częstości występowania wiadomości w języku, tj. do względnych prawdopodobieństw ich używania. Założenia te przyjął Mandelbrot w swoich rozważaniach.([4.7])

Niech $S_1, S_2, \ldots, S_j, \ldots, S_n$ będą podstawowymi literami lub symbolami, $t_1, t_2, \ldots, t_j, \ldots, t_n$ zaś ich względnymi kosztami. Liczba słów ([4.8]) $N(t)$ o koszcie t jest dana przez równanie (4.15) i jeśli mamy do czynienia z przypadkiem stabilnym, możemy posłużyć się uproszczonym wyrażeniem (4.18). Liczba słów o koszcie mniejszym lub równym t jest określona przez równanie (4.22) i wynosi $N_1(t)$. Równanie (4.22) dla przypadku stabilnego przyjmuje uproszczoną postać (4.23):

$$N_1'(t) = B_1 e^{\beta t} - B_0.$$

Uporządkujmy teraz w słowniku wszystkie słowa zgodnie z rosnącym kosztem t. Niech N_1' będzie numerem porządkowym słowa w tej klasyfikacji. Równanie (4.24) wyraża średni w sensie statystycznym związek pomiędzy kosztem t i numerem N_1':

$$t = t_0 + \frac{1}{\beta} \ln (N_1' + B_0).$$

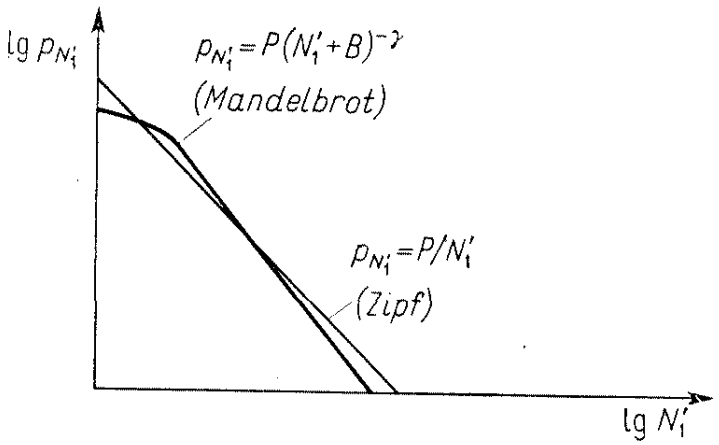
Rozważmy obecnie częstość występowania (lub prawdopodobieństwo) słowa. Pomiędzy kosztem i prawdopodobieństwem mógłby nie zachodzić żaden związek, nasze założenie jest jednakże inne. Zakładamy mianowicie, że prawdopodobieństwo słowa i jego koszt są do siebie dopasowane, tzn. długie, kosztowne ciągi kodowe odpowiadają rzadko występującym wiadomościom i odwrotnie, krótkie, tanie słowa odpowiadają wiadomościom używanym najczęściej.

Zagadnienie dopasowania rozważaliśmy w rozdz. 4.8 i 4.9; wyjaśniliśmy

tam, jak najlepiej wybrać prawdopodobieństwa symboli o różnych kosztach t_j. Wynik końcowy wyraziliśmy za pomocą równania (4.61):

$$\ln p_j = -\beta_1 t_j \qquad \text{lub} \qquad p_j = \mathrm{e}^{-\beta_1 t_j}.$$

Obecnie chcemy zastosować podobną analizę do słów, zamiast do liter czy symboli. Słowu o numerze N_1' odpowiada koszt $t_{N_1'}$; obie te wielkości są związane równaniami (4.23) i (4.24). Obecnie naszym warunkiem dobrego dopasowania jest warunek, by prawdopodobieństwo $p_{N_1'}$ słowa było związane



Rys. 4.1. Porównanie rozkładu Mandelbrota częstości słowa w funkcji numeru porządkowego słowa z rozkładem Zipfa

z kosztem tego słowa wzorem (4.61). Jest to ogólny warunek zawierający nową dowolną stałą β_1. Porównując wyrażenia (4.61) i (4.24) możemy wyeliminować koszt $t_{N_1'}$ i powiązać bezpośrednio numer porządkowy N_1' słowa z jego prawdopodobieństwem $p_{N_1'}$:

$$-\beta_1 t_{N_1'} = \ln p_{N_1'} = -\beta_1 t_0 - \frac{\beta_1}{\beta} \ln (N_1'+B) \tag{4.69}$$

i stąd

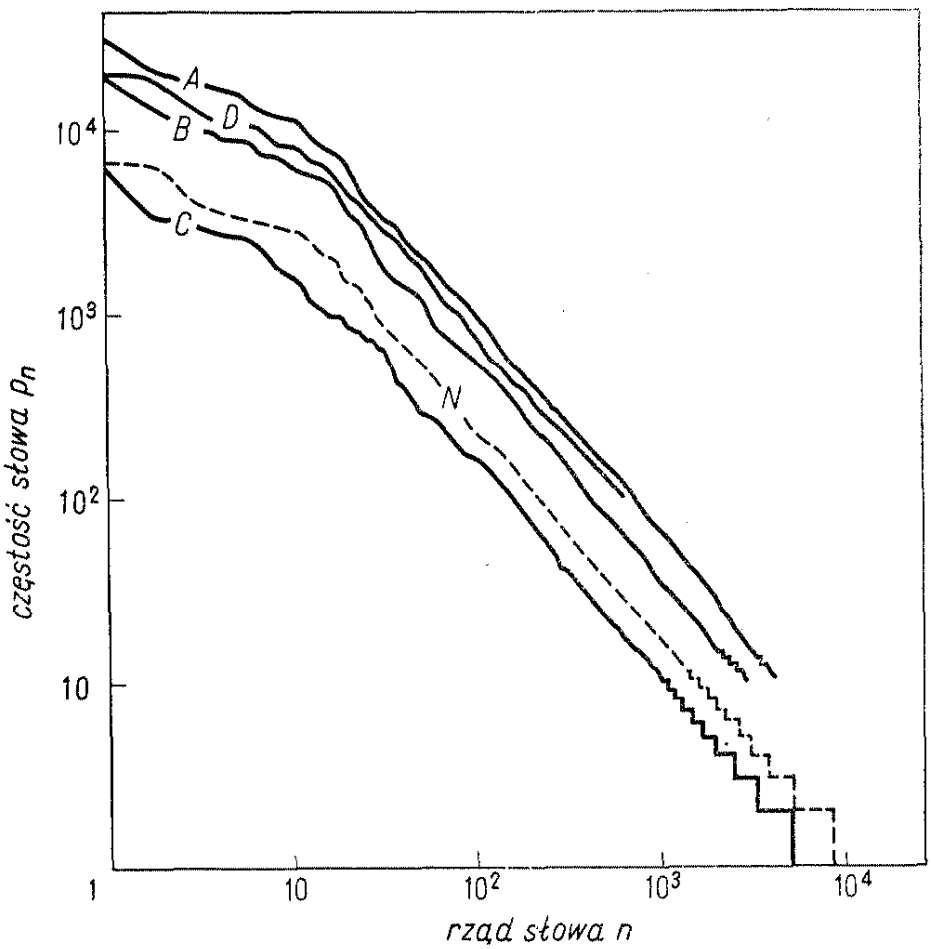
$$p_{N_1'} = P(N_1'+B)^{-\gamma}, \quad \text{gdzie} \quad P = \mathrm{e}^{-\beta_1 t_0}, \quad \gamma = \beta_1/\beta. \tag{4.70}$$

Wprowadzenie pojęcia pewnego rodzaju kodowania w mózgu wiadomości w słowa oraz założenie kosztu tych słów były niezbędne do uzyskania zależności (4.23) i (4.24). Nieznane koszty słów zdołaliśmy jednak wyeliminować, zakładając jedynie, że dopasowanie jest dobre. Wyrażenie (4.70) odpowiada prawu podanemu przez Mandelbrota; zawiera ono, jako specjalny przypadek, prawo zasugerowane przez Zipfa:

$$p_{N_1'} = P/N_1', \quad P \approx 1/10, \quad B = 0, \quad \gamma = 1, \quad (\beta_1 = \beta). \tag{4.71}$$

Inną ważną wielkością, którą należy rozważyć, jest całkowita liczba słów w słowniku. Jeżeli używamy słownika złożonego z R słów, mamy

$$1 \leqslant N_1' \leqslant R$$



Rys. 4.2. Krzywe doświadczalne częstości słowa w zależności od rzędu słowa. Krzywe A, B, C i D odnoszą się do języka norweskiego, krzywa N do niemieckiego

i suma prawdopodobieństw musi być równa jedności:

$$\sum_{N_1'=1}^{N_1'=R} p_{N_1'} = P \sum_{N_1'=1}^{N_1'=R} (N_1'+B)^{-\gamma} = 1. \tag{4.72}$$

Równanie to określa P lub, jeżeli P jest dane (Zipf przyjmuje $P = {}^1/_{10}$), określa całkowitą liczbę używanych słów. W taki właśnie sposób Shannon, przyjmując za Zipfem $P = {}^1/_{10}$, uzyskał liczbę słów w słowniku 8727. Dla większości języków można przyjąć $1 \leqslant \gamma \leqslant 1{,}2$, chociaż w szczególnych przypadkach γ może osiągać 1,6 (jak np. w mowie dzieci).

W dwóch cytowanych poprzednio pracach* Mandelbrot podaje wiele przykładów zastosowań teorii, są tam także rozważane niektóre inne własności języków. Na rysunku 4.1 pokazano dla porównania wykresy zależności wynikających z praw Mandelbrota i Zipfa, a na rys. 4.2 przedstawiono niektóre krzywe eksperymentalne, znalezione przez Zipfa.

4.11. ROZWIĄZANIE ZAGADNIENIA DOPASOWANIA

Podkreślaliśmy w rozdz. 4.9 potrzebę znalezienia sposobu kodowania, który dopasowywałby kanał do źródła podającego słowa. Wykazaliśmy, że równanie (4.61) stanowi warunek dobrego dopasowania. Obecnie chcemy pokazać, jak dopasowanie takie można praktycznie zrealizować.

W rozdziale 4.4 rozważyliśmy własności kanału. Przyjęliśmy mianowicie, że mamy n symboli $S_1, S_2, \ldots, S_j, \ldots, S_n$ o czasach trwania $t_1, t_2, \ldots, t_j, \ldots, t_n$, i wykazaliśmy, że średnia ilość informacji na symbol (rozdz. 4.8) ma wartość maksymalną, jeżeli prawdopodobieństwo symbolu p_j jest związane z jego czasem trwania wzorem (4.61):

$$p_j = \mathrm{e}^{-\beta t_j}, \qquad \beta = \text{stała}.$$

Rozważmy teraz wiadomość napisaną w pewnym języku przy użyciu pewnej liczby ν znaków Σ_λ. Znakiem ([4.9]) może być litera, blok liter, słowo itd. W języku chińskim np. każdy znak odpowiada słowu. W celu uproszczenia zagadnienia załóżmy, że każdy znak ma określone prawdopodobieństwo π_λ, niezależne od poprzedzających go znaków. Innymi słowy, aby uniknąć komplikacji, zaniedbamy korelację pomiędzy znakami.

Na początku chcemy rozważyć możliwie najprostszy sposób kodowania, polegający na przyjęciu ν symboli ([4.10]) i przyporządkowaniu w koderze najkrótszych symboli najbardziej prawdopodobnym znakom. W takim przypadku warunek (4.61) zazwyczaj nie jest spełniony i dopasowanie nie jest dobre. O wiele lepsze wyniki możemy otrzymać, jeżeli będziemy blokom znaków przyporządkowywać bloki symboli ([4.11]). Analiza takiego zagadnienia wymaga jednak uprzedniego zbadania własności prawdopodobieństw bloków znaków. Mamy

$$\left.\begin{array}{ll} \nu \text{ znaków:} & \Sigma_1, \Sigma_2, \ldots, \Sigma_\lambda, \ldots, \Sigma_\nu. \\ \text{prawdopodobieństwa:} & \pi_1, \pi_2, \ldots, \pi_\lambda \ \ldots, \pi_\nu. \end{array}\right\} \tag{4.73}$$

* Por. notka ** na str. 59.

Załóżmy dalej, że prawdopodobieństwa π_λ mogą być przedstawione w postaci

$$\pi_\lambda = \mathrm{e}^{-\theta_\lambda}. \tag{4.74}$$

Wzór ten będziemy traktować jako definicję wykładników θ_λ.

Prawdopodobieństwo bloku znaków $\Sigma_1\Sigma_2 \ldots \Sigma_\lambda$ można zapisać w postaci

$$\pi = \pi_1\pi_2 \ldots \pi_\lambda = \mathrm{e}^{(\theta_1+\theta_2+\ldots+\theta_\lambda)} = \mathrm{e}^{-\Theta}. \tag{4.75}$$

W ten sposób otrzymaliśmy prawo addytywności dla wykładników θ. Możemy teraz znaleźć liczbę $\mathcal{R}(\Theta)$ bloków symboli odpowiadających określonemu prawdopodobieństwu $\pi = \mathrm{e}^{-\Theta}$. Jest to zadanie zupełnie podobne do rozważanego w rozdz. 4.3. Musimy jedynie w miejsce czasów trwania t_j i T wstawić wykładniki θ_λ i Θ. Dla długich zespołów znaków wynik wyrażony przez równania (4.18) i (4.23) przepiszemy, jak następuje:

$$\mathcal{R}(\Theta) = a_1 \mathrm{e}^{\gamma\Theta}, \tag{4.76}$$

gdzie $\mathcal{R}(\Theta)$ jest średnią liczbą bloków znaków przy $\Sigma\theta = \Theta$, zaś współczynnik γ wprowadziliśmy na miejsce poprzedniego β. Mamy także

$$\mathcal{R}_1(\Theta) = b_1 \mathrm{e}^{\gamma\Theta} - b_0, \tag{4.77}$$

gdzie $\mathcal{R}_1(\Theta)$ jest średnią liczbą bloków znaków przy $\Sigma\theta \leqslant \Theta$.

Możemy obecnie porównać statystyczne własności bloków znaków i bloków symboli. Będziemy kodować $\mathcal{R}_1(\Theta)$ bloków znaków przy pomocy $N_1(t)$ najkrótszych bloków symboli; w ten sposób przypiszemy najkrótsze telegramy najczęściej występującym słowom. Porównując $\mathcal{R}_1(\Theta)$ i $N_1(t)$ i wykorzystując równania (4.77) i (4.23) dostajemy

$$\mathcal{R}_1(\Theta) = b_1 \mathrm{e}^{\gamma\Theta} - b_0 = B_1 \mathrm{e}^{\beta t} - B_0 = N_1(t). \tag{4.78}$$

Aby to uproszczone wyrażenie było słuszne, bloki znaków i symboli muszą być długie. W takim przypadku wyrazy z b_1 i B_1 w równaniu (4.78) będą o wiele większe niż stałe b_0 i B_0.

Wobec tego stałe te możemy pominąć. Dla celów praktycznych warunek (4.78) możemy więc zapisać w postaci

$$\left.\begin{aligned} b_1 \mathrm{e}^{\gamma\Theta} &= B_1 \mathrm{e}^{\beta t}, \\ \gamma\Theta &= \beta t + \ln (B_1/b_1). \end{aligned}\right\} \tag{4.79}$$

Ponieważ Θ i t są duże, drugi wyraz możemy odrzucić i napisać ostatecznie

$$\Theta = \beta t/\gamma. \tag{4.80}$$

Zgodnie z warunkiem (4.80) wykładnik Θ jest proporcjonalny do czasu trwania t.

Przy tym sposobie kodowania symbole S_j o czasie trwania t_j będą miały średnie prawdopodobieństwo

$$p_j = \mathrm{e}^{-\theta_j} = \mathrm{e}^{-\beta t_j/\gamma}. \tag{4.81}$$

Jest to dokładnie warunek (4.61), określający optymalny kod. Warunek (4.78) nie określa jednoznacznie kodu, tym niemniej udało się nam wykazać, że istnieje możliwość dopasowania źródła słów do kanału. Dobre dopasowanie wymaga kodowania przy użyciu długich bloków znaków i symboli, co jednak pociąga za sobą konieczność wprowadzenia opóźnień przy kodowaniu w nadajniku i przy dekodowaniu w odbiorniku.

Zagadnienia bardziej złożone, w których uwzględnia się korelacje pomiędzy kolejnymi znakami, a zatem i rozwlekłość, mogą być analizowane za pomocą metody podobnej do stosowanej w przypadku uproszczonym.

Wróćmy do warunku (4.68). Założyliśmy tam, że długie bloki symboli są stosowane dla rzadkich bloków znaków, a bloki krótkie dla często spotykanych bloków znaków. Ze stosowaniem zespołów znaków o zmiennej długości związana jest trudność polegająca na konieczności wprowadzenia specjalnego symbolu do oznaczania przerwy pomiędzy telegramami. Można tego uniknąć stosując bloki symboli o równej długości. Na pierwszy rzut oka takie rozwiązanie może wydawać się nieekonomiczne. Można jednakże wykazać, że nie jest ono gorsze od rozwiązania rozważanego powyżej. Zamieńmy warunek (4.78) warunkiem

$$\mathcal{R}_1(\Theta) = b_1 \mathrm{e}^{\gamma\Theta} - b_0 = A_1 \mathrm{e}^{\beta t} = N(t). \tag{4.82}$$

Ponieważ Θ i t są duże, możemy odrzucić stałą b_0 i otrzymamy

$$\gamma\Theta = \beta t + \ln (A_1/B_1). \tag{4.83}$$

Jeżeli w tym wyrażeniu odrzucimy ostatni wyraz, co jest możliwe, otrzymamy znów warunek (4.80). Postępowanie takie ilustruje dużą różnorodność możliwych metod kodowania, przy których uzyskuje się dobre dopasowanie.

DODATEK DO ROZDZIAŁU 4

Jednorodne, liniowe równanie różnicowe o stałych współczynnikach może być zapisane w postaci*

$$P(u(x)) = u(x+n)+p_{n-1}\,u(x+n-1)+\ldots+p_1 u(x+1)+p_0 u(x) = 0, \quad \text{(D.1)}$$

gdzie p są stałymi. Równanie to jest liniowe względem nieznanej funkcji $u(x)$, zaś przyrosty x są stałe i równe jedności. Ma ono n szczególnych rozwiązań $u_1(x), \ldots, u_n(x)$ niezależnych, jeżeli między rozwiązaniami tymi nie istnieją zależności liniowe. Rozwiązania te można znaleźć podstawiając

$$u(x) = \varrho^x. \quad \text{(D.2)}$$

Podstawiając (D.2) do (D.1) otrzymujemy

$$\varrho^x f(\varrho) = 0,$$

$$f(\varrho) = \varrho^n+p_{n-1}\varrho^{n-1}+\ldots+p_1\varrho+p_0 = 0. \quad \text{(D.3)}$$

Rozwiązanie (D.1) dostajemy, jeżeli ϱ jest pierwiastkiem równania charakterystycznego (D.3). Niech $\varrho_1, \varrho_2, \ldots, \varrho_n$ oznaczają pierwiastki tego równania. Załóżmy, że wszystkie one są różne. Można udowodnić, że odpowiadające tym pierwiastkom rozwiązania są niezależne i rozwiązanie ogólne ma postać

$$u(x) = \sum_{k=1}^{n} A_k \varrho_k^x, \quad \text{(D.4)}$$

gdzie A_k oznacza n dowolnych współczynników. W przypadku krotnego pierwiastka o krotności v rozwiązanie ogólne można otrzymać biorąc jako szczególne rozwiązania funkcje niezależne

$$\varrho_i^x, x\varrho_i^x, \ldots, x^{v-1}\varrho_i^x.$$

Rozważania dotyczące równań jednorodnych można łatwo uogólnić na przypadek równań niejednorodnych.

Ponieważ zagadnienie rozwiązania równania różnicowego sprowadza się do znalezienia pierwiastków równania algebraicznego (D.3), przypomnijmy

---

* L. M. Milne-Thomson, *The Calculus of Finite Differences*, London 1951, Macmillan, rozdz. 12, 13.

niektóre podstawowe własności* równań algebraicznych o współczynnikach rzeczywistych.

$$f(x) = a_0x^n+a_1x^{n-1}+\ldots+a_n = 0. \tag{D.5}$$

(1) Jeżeli $f(\alpha)$ i $f(\beta)$ mają różne znaki, istnieje co najmniej jeden pierwiastek pomiędzy α i β. Liczba takich pierwiastków jest nieparzysta.

(2) **Reguła Kartezjusza.** Liczba dodatnich rzeczywistych pierwiastków równania (D.5) jest albo równa liczbie zmian znaku $f(x)$, albo od niej mniejsza o liczbę parzystą.

(3) Całkowita liczba pierwiastków wynosi n. Jeżeli liczba zespolona $a+ib$ jest pierwiastkiem równania, to sprzężona z nią liczba zespolona $a-ib$ jest także pierwiastkiem tego samego równania.

---

* I. S. Sokolnikoff, E. S. Sokolnikoff, *op. cit.*, str. 27.

H. B. Fine, *College Algebra*, New York 1904, Ginn, str. 425.

L. E. Dickson, *First Course in the Theory of Equations*, New York 1939, Wiley, rozdz. 2.

5

ZAGADNIENIA KODOWANIA

5.1. KODOWANIE POSZCZEGÓLNYCH LITER — SYSTEM DWÓJKOWY

W poprzednich rozdziałach rozważyliśmy warunki najbardziej efektywnego kodowania. Nasze rozważania doprowadziły do warunku, że przy idealnym kodowaniu powinna być spełniona zależność

$$-\lg p_j = \beta t_j, \tag{5.1}$$

gdzie p_j jest prawdopodobieństwem a priori symbolu S_j, natomiast t_j czasem trwania albo długością tego symbolu, β zaś jest stałą.

Warunek (5.1) można uzasadnić w prosty sposób: symbole najbardziej prawdopodobne mają p_j stosunkowo bliskie jedności i stąd $-\lg p_j$ będzie liczbą dodatnią, stosunkowo małą. Zatem odpowiadające im sygnały zgodnie z warunkiem (5.1) powinny być krótkie, podczas gdy sygnały odpowiadające mało prawdopodobnym literom (znakom) mogą być długie. Wynik ten jest prawie oczywisty i równanie (5.1) określa warunek idealny, do którego, w sytuacjach praktycznych, w miarę możliwości należy dążyć. Wykorzystując dobrze znane prawdopodobieństwa występowania różnych liter w języku angielskim, wzór ten można też zastosować przy kodowaniu poszczególnych liter, z których zbudowane są słowa. W tablicy 5.1 (identycznej jak tabl. 1.1) podano prawdopodobieństwa p_j i ich ujemne logarytmy. Nie ma przy tym znaczenia, jakiego rodzaju logarytmy stosujemy, zmiana bowiem podstawy logarytmu prowadzi jedynie do przemnożenia stałej β przez stały mnożnik.

Jak widać z tabl. 5.1, logarytmy prawdopodobieństw liter ($^{5.1}$) zawarte są w przedziale od 0,7 do 3. Możemy dobrać kod binarny ($^{5.2}$), w którym

długość bloków sygnałów binarnych przyporządkowanych literom zmienia się w podobny sposób. Taki kod binarny można wykorzystać w kanale, w którym przekazuje się tylko jeden rodzaj impulsów: brak impulsu odpowiada 0, impuls odpowiada 1. Ponieważ nasze bloki sygnałów binarnych mają na ogół

Tablica 5.1

Prawdopodobieństwo występowania p i wartości $-\lg p$ dla liter języka angielskiego

Symbol	p	$-\lg p$	Symbol	p	$-\lg p$
przerwa między słowami	0,2	0,699	*L*	0,029	1,54
E	0,105	0,979	*C*	0,023	1,64
T	0,072	1,143	*FU*	0,0225	1,65
O	0,0654	1,184	*M*	0,021	1,68
A	0,063	1,2	*P*	0,0175	1,76
N	0,059	1,23	*YW*	0,012	1,92
I	0,055	1,26	*G*	0,011	1,96
R	0,054	1,27	*B*	0,0105	1,98
S	0,052	1,28	*V*	0,008	2,1
H	0,047	1,33	*K*	0,003	2,52
D	0,035	1,46	*X*	0,002	2,7
			JQZ	0,001	3,0

różną długość, musimy przewidzieć jakiś sposób rozpoznania końca bloku. Dwukrotne, trzykrotne lub czterokrotne zero może oznaczać koniec ciągu impulsów odpowiadających kodowanej literze. Oznacza to, że w blokach przyporządkowanych literom można stosować tylko pojedyncze zera. W tablicy 5.2 podano przykład tego rodzaju kodu.

Długość bloku zmienia się od 3 do 8 impulsów, co ogólnie biorąc jest wartością rozsądną. W długim telegramie, średni czas na literę można obliczyć mnożąc czasy t_j liter przez ich prawdopodobieństwa i sumując te iloczyny. Tak postępując otrzymujemy wartość

$$\bar{t} = \sum_{j=1}^{j=27} p_j t_j = 4{,}65, \tag{5.2}$$

gdzie p_j bierzemy z tabl. 5.1.

Wartość 4.65 jest także średnią ilością symboli na literę. Ponieważ używamy systemu dwójkowego, mamy

$$I_{\text{na literę}} = I_{\text{na symbol}} \cdot (\textit{ilość symboli na literę}) = \lg_2 2 \cdot 4{,}65 = 4{,}65.$$

Dla porównania rozważmy kilka innych sposobów kodowania ([5.3]). Jeżeli stosujemy bloki o równej długości, potrzebujemy pięciu cyfr dwójkowych do utworzenia 32 kombinacji, spośród których możemy wybrać 27 dla oznaczenia liter alfabetu. Kod, w którym stosujemy bloki o tej samej długości, równej 5, jest nieco gorszy niż nasz poprzedni kod z blokami o zmiennej długości. Otrzymana przez nas wielkość (informacja na literę) jest jednakże jeszcze zbyt duża, przekracza ona bowiem teoretyczną wartość 4,03 bitów na literę, wykazaną wzorem (1.30). Kod pokazany w tabl. 5.2 jest dużo lepszy niż zwykły kod Morse'a stosowany w telegrafii.

Tablica 5.2

Przykład kodu dwójkowego dla liter języka angielskiego

Litera	Kod	Litera	Kod	Litera	Kod
przerwa między słowami	000	*H*	101100	*W*	1101000
E	100	*D*	110000	*G*	1101100
T	1000	*L*	110100	*B*	1110000
O	1100	*C*	111000	*V*	1110100
A	10000	*F*	111100	*K*	1111000
N	10100	*U*	1010000	*X*	1111100
I	11000	*M*	1010100	*J*	10101000
R	11100	*P*	1011000	*Q*	10101100
S	101000	*Y*	1011100	*Z*	10110000

5.2. KODOWANIE POSZCZEGÓLNYCH LITER — SYSTEM TRÓJKOWY

W wielu przypadkach istnieje możliwość stosowania dodatnich i ujemnych impulsów oraz przerwy (braku impulsów). Ma to miejsce w przypadku taśmy magnetycznej lub systemu telegraficznego, w którym wykorzystuje się prąd

o zmiennym kierunku. Układy takie prowadzą do kodu trójkowego, w którym symbol może przyjmować jedną z postaci $-1, 0, +1$. Przykład (przypuszczalnie nie najlepszy) tego typu kodu podaliśmy w tabl. 5.3. Możemy stosować jeden z trzech symboli, np. -1, do oznaczania przerwy między blokami. Pozwala to nam kodować litery w kodzie dwójkowym.

Kod ten daje średnią liczbę 3,3 symboli na literę, co można łatwo obliczyć stosując prawdopodobieństwa z tabl. 5.1. Ilość bitów na literę wynosi

$$3{,}3 \lg_2 3 = 5{,}25. \tag{5.3}$$

Liczba ta jest znacznie większa od podanych liczb dla kodów z poprzedniego paragrafu. Wadą wszystkich kodów z blokami o zmiennej długości jest ko-

Tablica 5.3

Przykład kodu trójkowego dla liter języka angielskiego

Litera	Kod	Litera	Kod	Litera	Kod
przerwa między słowami	1,–1	*H*	011,–1	*W*	0100,–1
E	0,–1	*D*	100,–1	*G*	0101,–1
T	00,–1	*L*	101,–1	*B*	0110,–1
O	01,–1	*C*	110,–1	*V*	0111,–1
A	10,–1	*F*	111,–1	*K*	1000,–1
N	11,–1	*U*	0000,–1	*X*	1001,–1
I	000,–1	*M*	0001,–1	*J*	1010,–1
R	001,–1	*P*	0010,–1	*Q*	1011,–1
S	010,–1	*Y*	0011,–1	*Z*	1100,–1

nieczność wprowadzenia specjalnego sygnału wskazującego na zakończenie słowa kodowego. Kod z tabl. 5.3 nie jest dokładnie „wyważony", ponieważ symbol -1 jest mniej częsty niż 0 i 1. Bardziej dopracowany sposób kodowania mógłby wyrównać te różnice częstości i nieco zmniejszyć liczbę bitów na symbol.

Pod koniec tego rozdziału rozważymy problem kodowania całych słów zamiast kodowania poszczególnych liter i wykażemy, że taki sposób może być znacznie bardziej ekonomiczny.

5.3. ALFABET I LICZBY

Innym zagadnieniem jest zagadnienie kodowania oprócz liter alfabetu naturalnego także i liczb. Rozwiązanie tego zagadnienia będzie zależało od celu kodowania. Można założyć, że pisząc na maszynie częściej używamy liter niż liczb. W przypadku kalkulatora lub maszyny matematycznej sytuacja jest odwrotna. Podamy kilka przykładów:

(1) Stosujemy sześciopozycyjny, dwójkowy kod IBM (bitu korekcyjnego nie uwzględniamy), tak że bloki dla liter i dziesięciu dziesiętnych cyfr mają tę samą długość. Daje to średni stosunek prawdopodobieństw

$$\frac{\textit{prawdopodobieństwo litery}}{\textit{prawdopodobieństwo cyfry dziesiętnej}} = \frac{27}{10} = 2{,}7, \tag{5.4}$$

co jest wartością właściwą dla maszyny do pisania. Kod ten jest jednakże nieefektywny, ponieważ wykorzystuje się tylko 37 z możliwych 64 kombinacji, pozostawiając 27 nie wykorzystanych bloków symboli.

(2) Stosujemy 7 bloków sygnałów (symboli) dwójkowych dających 128 kombinacji i kodujemy wszystkie liczby od zera do 99 oraz 27 liter. Jedna kombinacja pozostaje nie wykorzystana. Jest to bardzo efektywny kod, dający średnio 3,5 bitów na liczbę (optimum wynosi 3,32). Traktujemy każdą z tych 100 liczb jako dwie cyfry dziesiętne (liczbę 7 można np. przedstawić jako 07); w ten sposób w liczbach występuje łącznie 200 cyfr dziesiętnych. Stąd mamy

$$\frac{\textit{prawdopodobieństwo litery}}{\textit{prawdopodobieństwo cyfry dziesiętnej}} = \frac{27}{200} = 0{,}135. \tag{5.5}$$

Wartość tego ilorazu wydaje się być rozsądna w przypadku kalkulatora.

(3) Przyjmujemy bloki 10 sygnałów dwójkowych dających 1024 kombinacji i kodujemy wszystkie liczby od 0 do 999 oraz 24 litery (dwie litery pozostaną nie zakodowane). Możemy zakodować tak samo literę *O* i zero a także *L* i 1. W ten sposób otrzymujemy 3,33 bitów na liczbę, co jest najlepszą praktycznie osiągalną wartością. Istnieje 1000 liczb trzycyfrowych, co daje łącznie 3000 cyfr dziesiętnych; stąd mamy stosunek

$$\frac{\textit{prawdopodobieństwo litery}}{\textit{prawdopodobieństwo cyfry dziesiętnej}} = \frac{26}{3000} = 0{,}00866. \tag{5.6}$$

Kod taki byłby odpowiedni dla elektronicznych maszyn cyfrowych, gdzie litery pojawiają się rzadko.

W kodach (2) i (3) dopełnienia do 99 i 999 będą odgrywały rolę zwykłego dopełnienia 9 i łatwo można je otrzymać stosując odpowiednie kody „nadmiar 14" lub „nadmiar 12". Ważne jest, aby przed wyborem kodu znać względne prawdopodobieństwo cyfr i liter.

5.4. DWÓJKOWE KODOWANIE SŁÓW

W rozdziale 3 rozważyliśmy wyniki Shannona dotyczące rozwlekłości języka angielskiego. W jednym ze swoich przykładów Shannon rozpatruje częstość występowania różnych słów w języku i podaje prawo

$$p_m = 1/10m \tag{5.7}$$

dla prawdopodobieństwa wystąpienia słowa o numerze porządkowym m w ciągu uporządkowanym zgodnie z malejącą częstością występowania. Prawa tego nie można stosować do nieskończonej liczby wyrazów, ponieważ w tym przypadku suma p_m będzie rozbieżna. Shannon założył, że używa się jedynie 8727 słów. Liczbę tę wybrał on dlatego, by suma p_m równała się jedności:

$$\sum_{m=1}^{m=8727} p_m = 1. \tag{5.8}$$

Przy takich założeniach można obliczyć średnią ilość informacji na słowo

$$\left.\begin{aligned} I &= 11{,}82 \text{ bitów na słowo,} \\ i &= \ 2{,}14 \text{ bitów na literę.} \end{aligned}\right\} \tag{5.9}$$

Przyjęto tu, że średnia długość słowa w 27-literowym alfabecie (26 liter *plus* przerwa między wyrazami) wynosi 5,5 liter. Liczba 2,14 bitów na literę stanowi duży postęp w porównaniu ze zwykłym kodowaniem poszczególnych liter, przy którym jest konieczne użycie około 5 bitów na literę.

---

* Obliczając graniczną liczbę słów: 8727 Shannon wykorzystywał, dla najczęściej używanych słów, prawdopodobieństwa znalezione empirycznie, natomiast dla słów mniej częstych ($m > 100$) stosował on wzór (5.7). Jeśliby stosować wzór (5.7) dla wszystkich słów, graniczna liczba słów w słowniku byłaby znacznie większa — wyniosłaby około 12000.

Rozpatrzmy teraz, jak można wykorzystać znajomość prawdopodobieństw słów. Na początek możemy rozważyć telegramy ([5.4]) o tej samej długości. Jeżeli stosujemy cyfry dwójkowe 0, 1 i przyjmujemy stałą długość telegramu 13 bitów, to otrzymamy $2^{13} = 8192$ różnych telegramów, których można użyć do kodowania 8192 słów, tworzących słownik wystarczający do wielu celów. Ten prosty sposób kodowania daje 13 bitów na słowo lub 2,36 bitów na literę, co jest już wartością bardzo bliską teoretycznej granicy (5.9).

Tablica 5.4

I	II	III	IV	V
1,00	10,00	100,00	1000,00	10000,00
	11,00	101,00	1010,00	10100,00
		110,00	1011,00	10101,00
		111,00	1100,00	10110,00
			1101,00	10111,00
			1110,00	11000,00
			1111,00	11010,00
				11011,00
				11100,00
				11101,00
				11110,00
				11111,0

Lepszy wynik można otrzymać stosując metodę podobną do opisanej w pierwszej części tego rozdziału (przykład z tabl. 5.2). W metodzie tej nie używamy wielokrotnych zer w telegramach; używamy ich tylko do oznaczania końców tych telegramów. Każdy telegram będzie zaczynał się od 1 i kończył zerami wielokrotnymi 00,000,0000,... Pierwsze takie kombinacje pokazane są w tabl. 5.4.

W czwartej kolumnie kombinacja 1001 została pominięta, ponieważ zawiera dwa zera obok siebie. Z tego powodu pomijamy 10001,00; 10010,00; 10011,00 i 11001,00 z piątej kolumny.

W dalszych rozważaniach opuścimy ostatnie dwa zera pojawiające się przy końcu każdego telegramu. Przypuśćmy, że znaleźliśmy wszystkie telegramy pewnej kolumny (np. siedem sygnałów IV kolumny, z czterema cyframi

w głównej części każdego sygnału). Wprowadzimy następujące oznaczenia:

n — numer kolumny;
$N_{n,1}$ — liczba telegramów zakończonych symbolem 1;
$N_{n,0}$ — liczba telegramów zakończonych pojedynczym zerem;
$N_{n,00}$ — liczba telegramów zakończonych grupą zer.

Rozważmy przykład w przypadku kolumny IV (nie licząc dwu ostatnich zer za kropką):

$$n = 4, \quad N_{4,1} = 3, \quad N_{4,0} = 2, \quad N_{4,00} = 2.$$

Chcemy teraz obliczyć liczby $N_{n+1,1}$, $N_{n+1,0}$ oraz $N_{n+1,00}$ dla telegramów z $(n+1)$ cyframi w głównej części. Telegramy $N_{n,0}$ oraz $N_{n,1}$ mogą być uzupełnione albo przez 0, albo przez 1. Telegram $N_{n,00}$ może być uzupełniony tylko przez 0 i stąd mamy następujące związki:

$$\left.\begin{aligned} N_{n+1,1} &= N_{n,1}+N_{n,0}, \\ N_{n+1,0} &= N_{n,1}, \\ N_{n+1,00} &= N_{n,0}+N_{n,00}. \end{aligned}\right\} \tag{5.10}$$

Stosując te związki krok za krokiem otrzymujemy wyniki podane w tabl. 5.5.

Jeśli zastosujemy do przesyłania słów telegramy z kolumn od 1 do 17, to liczba cyfr dwójkowych tworzących telegram będzie się zmieniała od 3 do 19 wliczając w to dwa zera dodawane na końcu każdego telegramu, co pozwoli zakodować słownik złożony z około 11000 słów. Obliczenie średniej liczby bitów na słowo nie jest zbyt trudne, jeśli wykorzysta się wzór Shannona dla prawdopodobieństw kolejnych słów. Prowadzi to do wyniku bliskiego 12 bitom na słowo, co jest wielkością zbliżoną do teoretycznej granicy 11,82. Ten typ kodowania wydaje się najbardziej efektywny spośród praktycznie realizowalnych.

W praktyce tego typu kod powinien być uzupełniony dodatkowym kodem pozwalającym kodować poszczególne litery. Potrzeba takiego kodu wynika stąd, że litery są potrzebne do tworzenia wszystkich słów, których nie ma w słowniku, szczególnie imion, nazwisk, nazw miast, krajów, rzek itp. Jeżeli takie słowa nie są zbyt częste, można zachować 27 ostatnich telegramów kodu do przesyłania liter, jeżeli natomiast pewne słowa występują bardzo często, ten sposób kodowania staje się niewygodny.

Związki (5.10) możemy wykorzystać do obliczenia względnych prawdopodobieństw $p_{n,1}$, $p_{n,0}$, $p_{n,00}$ telegramów, których część główna kończy się na 1, 0 lub 00:

$$N_{n,1} = N_n \cdot p_{n,1}, \quad N_{n,0} = N_n \cdot p_{n,0}, \quad N_{n,00} = N_n \cdot p_{n,00},$$
$$N_n = N_{n,1} + N_{n,0} + N_{n,00}. \tag{5.11}$$

Równania (5.10) i (5.11) określają razem kolejne wartości p. Kiedy n jest duże (np. $n > 10$), prawdopodobieństwa te stają się niezależne od n i łatwo można otrzymać

$$p_1 \approx p_{00} = {}^1/_2(3-\sqrt{5}) = 0{,}382, \quad p_0 = 0{,}236; \tag{5.12}$$

równania (5.10) przechodzą wówczas w

$$N_{n+1} = N_n(2p_1 + 2p_0 + p_{00}) = 1{,}618 N_n; \tag{5.13}$$

stąd

$$N_n = D_0 \cdot 1{,}618^n, \quad N_1 + N_2 + \ldots + N_n = D_1 \cdot 1{,}618^n, \tag{5.14}$$

gdzie

$$D_0 = 1{,}163, \quad D_1 \approx 3.$$

Porównanie otrzymanych wyników z dokładnymi wynikami z tabl. 5.5 wykazuje, że te przybliżone związki są dokładne dla $n > 10$. Przy pomocy tych zależności można dokładniej przeanalizować kod podany w niniejszym rozdziale.

Tablica 5.5

Binarne kodowanie słów

	Numer kolumny n																		
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
$N_{n,1}$	1	1	2	3	5	8	13	21	34	55	89	144	233	377	610	987	1597	2584	4181
$N_{n,0}$	0	1	1	2	3	5	8	13	21	34	55	89	144	233	377	610	987	1597	2584
$N_{n,00}$	0	0	1	2	4	7	12	20	33	54	88	143	232	376	609	986	1596	2583	4180
ΣN	1	2	4	7	12	20	33	54	88	143	232	376	609	986	1596	2583	4180	6764	10945
całkowita liczba zakodowanych słów	1	3	7	14	26	46	79	133	221	364	596	972	1581	2567	4163	6746	10926	17690	28635

5.5. ALFABETYCZNE KODOWANIE SŁÓW

Inny sposób kodowania słów może być oparty na wykorzystaniu jako symboli liter czy też ich bloków. Jeżeli jako symbole stosujemy pary liter: *aa*, *ab*, *ac*, ..., *zz*, możemy zakodować $26^2 = 676$ słów (^{5,5}) przy ilości informacji $2 \cdot 4{,}7 = 9{,}4$ bitów na słowo, ponieważ alfabet 26-literowy daje 4,7 bitów na literę. Dla uzyskania bardziej licznego zbioru telegramów możemy przyjąć telegramy trzyliterowe; daje to $26^3 = 17576$ telegramów, co stanowi już duży zbiór bez żadnej rozwlekłości, dający $3 \cdot 4{,}7 = 14{,}1$ bitów na telegram. Z obu tych przykładów widać, jak duża jest rozwlekłość zwykłego języka, gdzie na słowo wypada średnio 5,5 liter zamiast 2 lub 3, jak to ma miejsce w omawianych kodach. Oczywiście, telegramy tych kodów będą w większości niemożliwe do wymówienia.

Mniej efektywny kod, zawierający około 10000 telegramów, podał P. Luhn z IBM. Trzy- i czteroliterowe telegramy tego kodu są wymawialne i tak zakończone, że przerwa między tymi telegramami nie jest niezbędna. Jest to możliwy do zastosowania w praktyce kod, o średniej długości telegramu 3,8 litery zamiast 5,5 jak w przypadku języka angielskiego.

5.6. KODOWANIE OPARTE NA KORELACJI W OBRĘBIE BLOKÓW LITER

W zakończeniu rozdz. 3 omawialiśmy sposób kodowania ogólnego typu, podany przez Shannona. Sposób ten jest oparty na zastosowaniu układu z pamięcią oraz na wykorzystaniu przy kodowaniu danej litery powiązań statystycznych tej litery z literami sąsiednimi. Procedura taka stanowi bezpośrednie zastosowanie matematycznych rozważań z rozdz. 3, lecz może być użyta jedynie dla języka, dla którego przeprowadzono szczegółowe badania statystyczne i określono co najmniej prawdopodobieństwa digramów i trigramów (tzn. bloków dwu, trzy i więcej literowych).

Praktyczne metody kodowania tego typu rozważał B.M. Oliver*. Założył on, że do przekazywania liter w systemie wykorzystuje się 27 impulsów o amplitudach 0, 1, 2, ..., 27, i rozważał możliwości zmniejszenia do minimum

* B. M. Oliver, *Efficient Coding*, Bell System Tech. J. **31** (1952), str. 724.

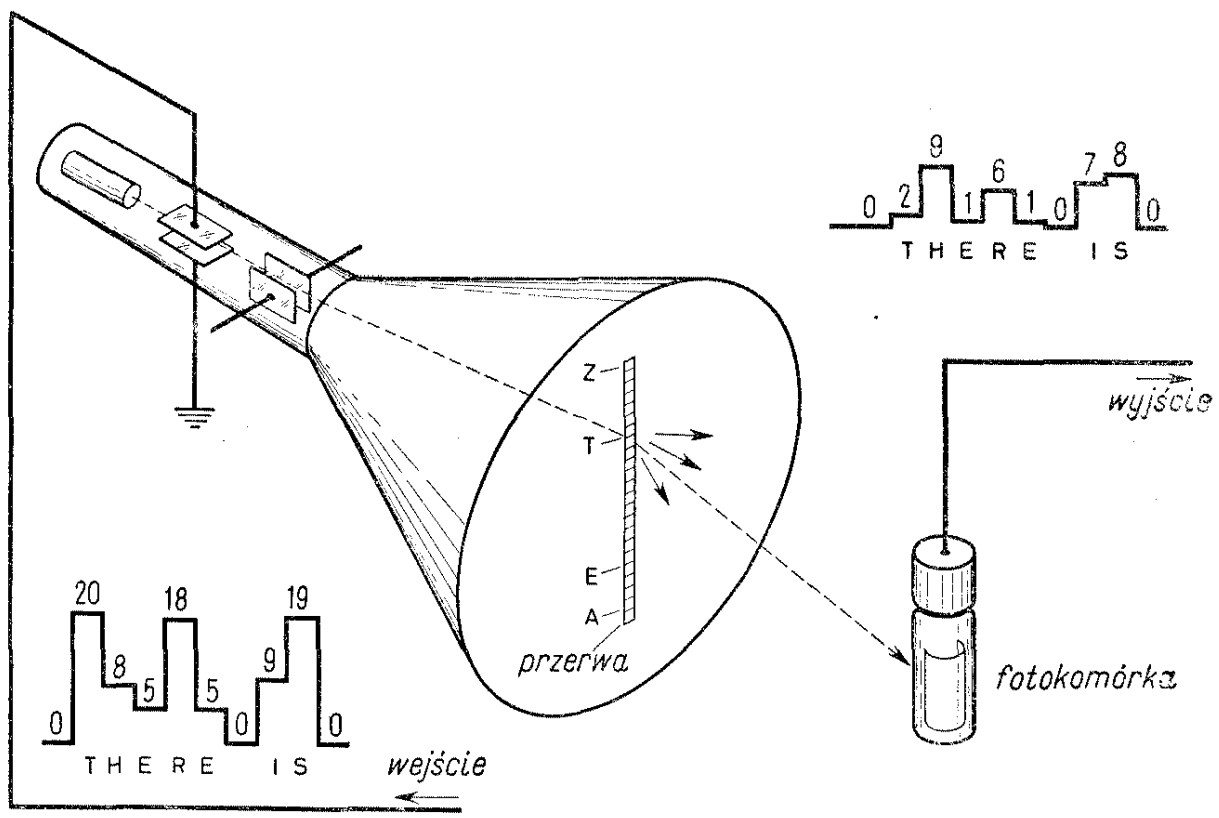
średniej mocy w kanale. Pierwszą możliwością jest użycie naszych skwantowanych impulsów zgodnie z porządkiem alfabetycznym:

symbol:	przerwa	A	B	C	D	E	...	
amplituda impulsu:	0	1	2	3	4	5	...	(5.15)

Rzecz oczywista, że nie jest to rozwiązanie dobre. Lepszy sposób kodowania może być oparty na prawdopodobieństwach z tabl. 5.1:

symbol:	przerwa	E	T	O	A	N	...	
amplituda impulsu:	0	1	2	3	4	5	...	(5.16)

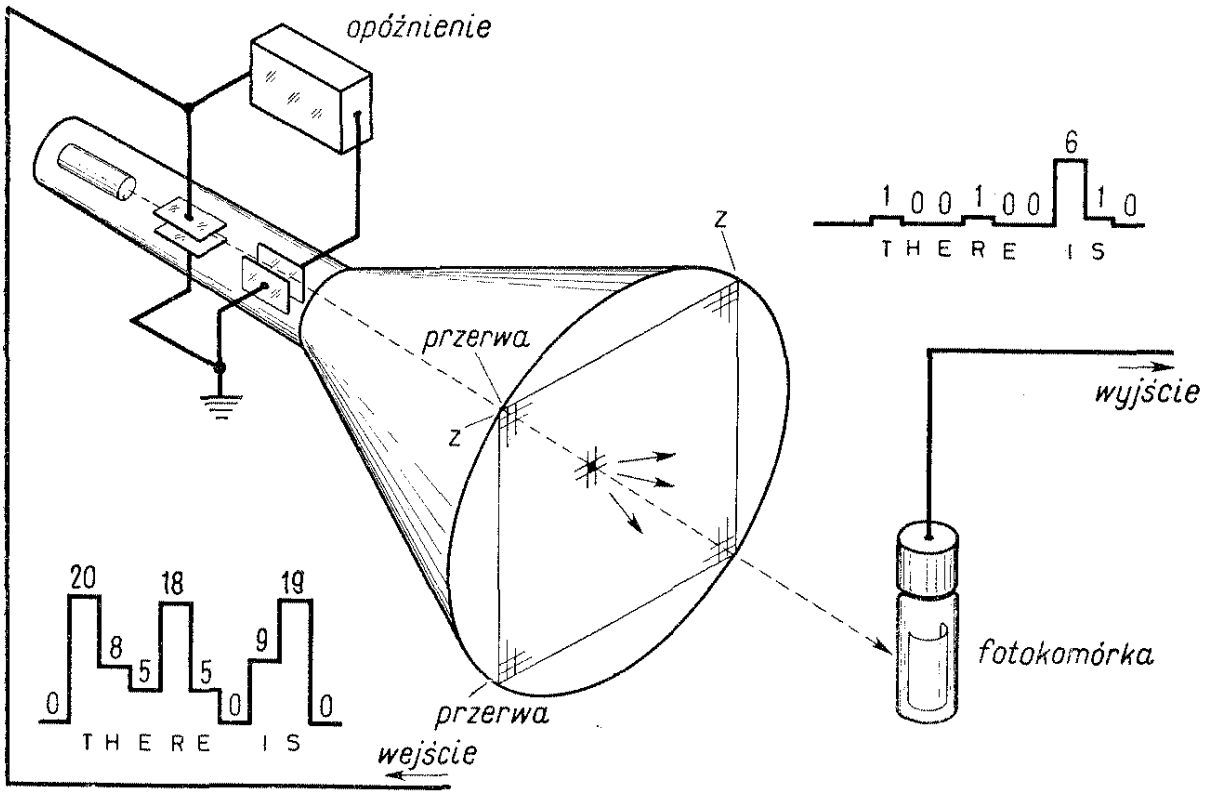
Rozwiązanie to jest bardziej efektywne, ponieważ impulsy o małej amplitudzie odpowiadają najbardziej prawdopodobnym literom.



Rys. 5.1. *Monogrammer* Olivera do kodowania pojedynczych liter

Zaproponowany przez Olivera układ realizujący taki sposób kodowania został przedstawiony na rys. 5.1. Impulsy odchylają promień lampy, na której ekran została nałożona maska o zmiennej przezroczystości (zgodnie z (5.16)). Natężenie światła padającego na fotokomórkę, a stąd i impuls z fotokomórki odpowiada przyporządkowaniu (5.16). Układ taki nazwano *monogrammerem*.

Digrammer pokazano na rys. 5.2 (także z pracy Olivera). W tym przypadku wykorzystuje się dwie pary płytek odchylających w lampie. Jedna para płytek jest sterowana poprzednią literą, druga zaś literą przychodzącą w danej chwili. Każdej parze liter (digramowi) (i, j) odpowiada punkt na ekranie. Przezro-



Rys. 5.2. *Digrammer* Olivera do kodowania par liter

czystość maski w punkcie wyznaczonym przez (i, j) odpowiada prawdopodobieństwu $p(i, j)$. Taki sposób kodowania prowadzi do dalszego zmniejszenia średniej mocy potrzebnej do transmisji. Rozważa się także układy oparte na trigramach i tetragramach, wydaje się jednak, że nie znajdą one praktycznego zastosowania.

Wielu różnych autorów* podaje inną jeszcze metodę kodowania. Najlepiej można ją wyjaśnić na przykładzie telefonii. Mamy przesłać ciągłą krzywą i dysponujemy danymi statystycznymi co do typu występujących krzywych. Wykorzystując tę informację oraz niektóre już przesłane wartości możemy przewidzieć kolejną najbardziej prawdopodobną wartość i przesyłać

* P. Elias, *Harvard University Thesis*, Proc. I. R. E. **39** (1951), str. 839; E. R. Kretzmer, Bell System Tech. J. **31** (1952), str. 751; C. W. Harrison, *ibid.*, str. 764.

tylko poprawkę, tzn. odchyłkę od wartości przewidzianej ([5.6]). Zaletą tej metody jest możliwość jej zastosowania do dwóch wymiarów (telewizja), przy czym warunkiem zastosowania jest znajomość odpowiednich danych statystycznych. W przypadku telewizji rozpatrzymy system, w którym wykorzystuje się $2^6 = 64$ poziomów kwantyzacji. Jeśli wszystkie poziomy są równoprawdopodobne, potrzebujemy 6 bitów na punkt obrazu. Jednakże różne poziomy mają różne prawdopodobieństwa i dzięki temu informacja na punkt maleje do 5 bitów. Wykorzystując korelację pomiędzy sąsiednimi punktami można dalej zmniejszyć tę wartość do 2,4—2,6 bitów na punkt. Odpowiada temu możliwość dwukrotnego zmniejszenia przepustowości kanału telewizyjnego. Podobnie jak w innych tego typu zagadnieniach, trzeba jednak brać pod uwagę fakt, że transmisja z małą rozwlekłością jest znacznie mniej odporna na zakłócenia. Spowodowane to jest tym, że zmniejszenie rozwlekłości zmniejsza także możliwość przewidywania i korekcji błędu w odbiorniku.

6

KODY WYKRYWAJĄCE I KORYGUJĄCE BŁĘDY

6.1. KODY WYKRYWAJĄCE BŁĘDY

W poprzednich rozdziałach rozważaliśmy metody pozwalające na eliminowanie rozwlekłości i konstruowanie kodów, w których liczba bitów na literę jest najmniejsza. Metody te można wykorzystać nie tylko w przypadku liter, lecz także w przypadku dowolnego typu symboli przy rozpatrywaniu bardziej ogólnych problemów teorii informacji.

Obecnie musimy jednakże postawić pytanie, czy w niektórych zagadnieniach praktycznych rozwlekłość nie jest czynnikiem korzystnym. Jeżeli bowiem przesyłamy lub magazynujemy telegramy w systemie, który charakteryzuje się dużym poziomem szumów, pewne błędy mogą być spowodowane szumami i rozwlekłość może być bardzo pomocna w procesie wykrywania, a nawet korygowania błędów. Na przykład, korekcji takiej dokonuje człowiek czytając książkę zawierającą pewną ilość błędów drukarskich lub telegram (w sensie potocznym), w którym pewne litery zostały zniekształcone. Zamiast wykorzystywać bardzo dużą naturalną rozwlekłość języka angielskiego do wykrywania błędów, możemy wprowadzić specjalny mechanizm kontroli, stosowany w połączeniu z kodem o małej rozwlekłości. Jest to rozwiązanie o wiele bardziej skuteczne, ponieważ nie można korzystać z naturalnej rozwlekłości języka przy wykrywaniu błędów w nazwiskach, nazwach miast, rzek itp.

Jeżeli mamy do czynienia z kodem binarnym, najprostszą formą kontroli jest test parzystości. Możemy na przykład na podstawie wyników obserwacji wnioskować, że w pewnej liczbie sygnałów binarnych (powiedzmy, $n = 6$) nie pojawi się więcej niż jeden błąd. Aby móc stosować test parzystości,

łączymy sygnały w bloki pierwotne po sześć i dodajemy do każdego bloku tworzącego telegram siódmą cyfrę binarną tak dobraną, aby całkowita liczba jedynek była liczbą parzystą. Na przykład:

blok pierwotny 101100, cyfra binarna 1,
blok pierwotny 100 010, cyfra binarna 0.

Pozwala to na wykrycie po stronie odbiorczej jednego błędu w ciągu, lecz bez możliwości określenia miejsca, w którym pojawił się błąd. Nie ma także możliwości wykrycia błędów podwójnych.

6.2. KODY KORYGUJĄCE POJEDYNCZE BŁĘDY

Możemy rozważyć następnie możliwość wprowadzenia metod kontroli umożliwiających określenie po stronie odbiorczej pozycji błędu, co pozwoli go skorygować.

W swej interesującej pracy R. W. Hamming* rozważał zagadnienie kodów pozwalających określić pozycję błędu i korygować go. Hamming konstruuje takie kody wprowadzając kilka testów parzystości; w swej pracy podaje on także niektóre zastosowania tej metody.

Rozważmy, dla wyjaśnienia metody Hamminga, telegram będący ciągiem cyfr binarnych (6.1) i załóżmy, że w tym ciągu nigdy nie wystąpi więcej niż jeden błąd. Spośród n pozycji ciągu wybieramy m pozycji, które rezerwujemy do przenoszenia informacji, natomiast resztę: k pozycji rezerwujemy dla kontroli. Cyfry binarne na k pozycjach kontrolnych dają 2^k różnych liczb binarnych, które powinny wskazywać:

(1) że nie ma błędu lub
(2) jeśli jest, to na jakiej pozycji.

Ponieważ mamy n pozycji, wszystkich wyżej wymienionych sytuacji jest $n+1$ i stąd otrzymujemy warunek

$$2^k \geqslant n+1. \tag{6.1}$$

* R. W. Hamming, Bell System Tech. J. **29** (1950), str. 147.

Kontrola jest najbardziej efektywna, jeżeli zachodzi równość. W tablicy 6.1 podaliśmy maksymalną liczbę symboli, które mogą być skontrolowane przy użyciu k pozycji kontrolnych.

Drugi wiersz w tabl. 6.1 odpowiada założeniu, że w ciągu trzech cyfr binarnych nie może wystąpić więcej niż jeden błąd. Jednej z pozycji w ciągu

Tablica 6.1

Maksymalna liczba symboli, które mogą być skontrolowane za pomocą jednego do pięciu symboli kontrolnych

Liczba symboli kontrolnych k	Całkowita liczba symboli n	Informacja $m = k-n$
1	1	0
2	3	1
3	7	4
4	15	11
5	31	26

użyjemy do przenoszenia informacji, dwie pozostałe będą służyły do kontroli (tabl. 6.2).

Cyfry binarne* na pozycjach B i C wybiera się w nadajniku tak, że

$$A+B \text{ jest parzyste oraz } A+C \text{ jest parzyste.}$$

Odbiornik oblicza te dwie sumy: jeśli tylko jedna z nich jest nieparzysta, to sygnał z pozycji kontrolnej objętej tą sumą jest błędny. Jeśli, na przykład, $A+B$ jest nieparzyste, a $A+C$ jest parzyste, sygnał B jest błędny. W przypadku, gdy błąd tkwi w A, obie sumy będą nieparzyste. Stąd zawsze wiadomo, który sygnał musi być skorygowany.

Trzeci wiersz w tabl. 6.1 odpowiada kodowi, który będziemy tu nazywać kodem *cztery z siedmiu*. W kodzie tym cztery cyfry binarne przenoszą informację, trzy służą do kontroli. Tu również błędy mogą być wykrywane i korygo-

* Autor oznacza tak samo pozycję ciągu i cyfrę na tej pozycji. (*Przyp. tłum.*).

Tablica 6.2

Kod o jednej pozycji informacyjnej i dwu pozycjach kontrolnych*

Informacja	Symbole kontrolne
A	B C
x	x
x	x

* Symbole kontrolne zostały wybrane tak, by suma x-ów w każdej ścieżce była parzysta.

Tablica 6.3

Kod o czterech pozycjach informacyjnych i trzech pozycjach kontrolnych*

Informacja	Symbole kontrolne
A B C D	E F G
x—x—x	x
x—x x	x
x x—x	x

* Symbole kontrolne zostały wybrane tak, by suma x-ów w każdej ścieżce była parzysta.

wane przy założeniu, że nigdy nie może wystąpić więcej niż jeden błąd w siedmiu cyfrach binarnych. Przebieg procesu kontroli przedstawiono w tabl. 6.3.

Sygnałami A, B, C i D mogą być zero lub jeden. Cyfry w pozycjach kontrolnych powinny być tak wybrane w nadajniku, aby następujące sumy kontrolne były parzyste:

$$\left.\begin{aligned} A+B+C+E \\ A+B+D+F \\ A+C+D+G \end{aligned}\right\} \text{parzyste.} \tag{6.2}$$

Te trzy sumy oblicza odbiornik dla ciągu odebranego.

Jeżeli wszystkie sumy są parzyste, nie ma żadnego błędu, ponieważ założyliśmy, że w siedmiu pozycjach nie może wystąpić więcej niż jeden błąd. Jeżeli jedna suma jest nieparzysta, błąd tkwi w pozycji kontrolnej objętej tą sumą. Na przykład, gdy tylko pierwsza suma jest nieparzysta, znaczy to,

że cyfra na pozycji E jest błędna. Jeżeli dwie sumy dają liczby nieparzyste, błąd tkwi w pozycji, która występuje w obu tych sumach jednocześnie i nie występuje w trzeciej sumie. Jeśli na przykład, dwie pierwsze sumy są nieparzyste, błąd tkwi w pozycji B. Gdy wszystkie sumy są nieparzyste, błąd jest w pozycji A.

Tablica 6.4

Kod o jedenastu pozycjach informacyjnych i czterech pozycjach kontrolnych*

Informacja											Symbole kontrolne			
Pozycje	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
	x	x	x	x		x	x	x				x			
	x	x	x		x	x			x	x			x		
	x	x		x	x		x		x		x			x	
	x		x	x				x		x	x				x

* Symbole kodowe zostały wybrane tak, by suma x-ów w każdej ścieżce była parzysta.

Powyższa metoda różni się nieznacznie od metody Hamminga, prowadzi ona jednakże do podobnych rezultatów.

Rozważmy teraz czwarty przypadek z tabl. 6.1 i ponownie załóżmy, że nie może wystąpić więcej niż jeden błąd w ciągu 15 cyfr binarnych, tzn., że mamy

$$m = 11 \text{ pozycji informacyjnych,}$$
$$k = 4 \text{ pozycje kontrolne.}$$

System kontroli dla tego przypadku jest pokazany w tabl. 6.4. Wynika stąd, że mają być parzyste następujące sumy kontrolne:

$$\left.\begin{aligned} &A + B + C + D + F + G + H + L \\ &A + B + C + E + F + I + J + M \\ &A + B + D + E + G + I + K + N \\ &A + C + D + E + H + J + K + O \end{aligned}\right\} \text{ parzyste.} \qquad (6.3)$$

Nieparzystość pojedynczej sumy oznacza, że błąd występuje w którejkolwiek z pozycji kontrolnych. Błędy na pozycjach informacyjnych są przyczyną nieparzystości dwu, trzech lub czterech sum kontrolnych.

Ogólna metoda staje się oczywista na tle powyższych przykładów. Jeżeli mamy k pozycji kontrolnych i pewną liczbę m pozycji informacyjnych $A, B, C, D, \ldots$, włączamy cyfrę binarną z pozycji A do wszystkich k sum kontrolnych. Każda cyfra binarna z następnych k pozycji występuje w $(k-1)$ spośród k sum kontrolnych. Następnie wykorzystujemy $(k-2)$ spośród k sum kontrolnych dla następnych $\frac{1}{2}k(k-1)$ pozycji informacyjnych itd. Każda suma kontrolna zawiera cyfrę binarną tylko z jednej pozycji kontrolnej i cyfra binarna z danej pozycji kontrolnej występuje tylko raz. Stąd w ogólnym przypadku k pozycji kontrolnych daje

$$m = 1+k+\frac{k(k-1)}{2}+\frac{k(k-1)(k-2)}{2\cdot 3}+\ldots+ \\ +\frac{k(k-1)}{2} = 2^k-k-1 \tag{6.4}$$

możliwości kontroli, co stanowi maksymalną liczbę pozycji informacyjnych, które mogą być skontrolowane. Pojedyncze błędy mogą także występować na k pozycjach kontrolnych, przypadek zaś braku błędów daje jeszcze jedną możliwość. Dlatego mamy

$$2^k = m+k+1 = n+1,$$

co jest zgodne z równaniem (6.1), i stąd możemy wnioskować, że podana tu metoda jest ogólna.

Jeżeli mamy 6 pozycji informacyjnych, jak w kodzie stosowanym przez IBM, to potrzebne nam są cztery pozycje kontrolne (tabl. 6.1). W takim przypadku możemy dowolnie wybrać 6 z 11 pozycji (A do K) z tabl. 6.4.

6.3. KODY KORYGUJĄCE POJEDYNCZE BŁĘDY I WYKRYWAJĄCE BŁĘDY PODWÓJNE

Nasze poprzednie postępowanie możemy rozszerzyć tak, aby wykrywać błędy podwójne. Zacznijmy od kodu korygującego pojedyncze błędy. Kod ten możemy uzupełnić jeszcze jedną pozycją kontrolną, aby móc sprawdzić wszystkie poprzednie pozycje posługując się testem na parzystość. Mamy więc następujące przypadki:

(1) Nie ma błędu — wszystkie testy na parzystość dają wynik dodatni.

(2) Pojedynczy błąd — pewne testy dają wynik ujemny i wskazują pozycję błędu. Ostatni dodatkowy test daje także wynik ujemny.

(3) Dwa błędy. Ostatni test daje wynik dodatni, niektóre zaś z poprzednich testów — wyniki ujemne. Świadczy to o tym, że na pewnych pozycjach wystąpiły co najmniej dwa błędy. Metoda ta jest ogólna i może być zastosowana do któregokolwiek z przykładów rozważanych w poprzednich paragrafach. Stosując reprezentację ciągów jako punktów w wielowymiarowej przestrzeni, Hamming podał interpretację geometryczną wprowadzonych pojęć. Na tej interpretacji można oprzeć ogólny dowód słuszności poprzednich metod.

Nowsze osiągnięcia teorii kodów korekcyjnych można znaleźć w pracy E. N. Gilberta*. Obszerny materiał dotyczący kodów korygujących wielokrotne błędy jest zawarty w „Transactions of the I.R.E. Professional Group on Information Theory"**.

6.4. EFEKTYWNOŚĆ KODÓW KOREKCYJNYCH

Kody korekcyjne, które rozważaliśmy w poprzednich rozdziałach, można wykorzystać następująco. Wybierzmy jako szczególny przykład kod „cztery z siedmiu" (tabl. 6.3). Przyjmijmy, że pewne słowo zakodowane jest następująco:

$$0000100010100100. \tag{6.5}$$

Telegram ten podzielmy na bloki czterocyfrowe:

$$0000,\quad 1000,\quad 1010,\quad 0100, \tag{6.6}$$

które uzupełnimy trzema cyframi odpowiadającymi testom parzystości — równanie (6.2):

$$0000000,\quad 1000111,\quad 1010010,\quad 0100110. \tag{6.7}$$

W tej postaci telegram przekazywany jest przez kanał i w odbiorniku wykorzystuje się testy parzystości do korekcji, a następnie odrzuca zbędne już trzy ostatnie cyfry i przywraca jego początkową ((6.6) i (6.5)) postać.

Postępowanie takie jest słuszne w przypadku, gdy na n cyfr przypada

---

* E. N. Gilbert, Bell System Tech. J. **31** (1952), str. 504.

** Trans. I.R.E. (P.G.I.T.) **1—4**; przede wszystkim zob. Golay, Reed, Silverman, Balser i in. Zob. także *Convention Record of the IRE Nat. Conv. 1954.*

nie więcej niż jeden błąd (w naszym przykładzie $n = 7$). Warunek ten nigdy nie jest dokładnie spełniony. Zazwyczaj mamy do czynienia z sytuacją, w której błędy pojawiają się przypadkowo ze średnim prawdopodobieństwem p. Dowiedziemy, że kod korekcyjny pozwala znacznie zmniejszyć liczbę błędów i daje małe prawdopodobieństwo P wystąpienia błędów nie skorygowanych. Oznaczmy przez p prawdopodobieństwo błędu jednej cyfry w telegramie. Błąd polega na tym, że 0 może przejść w 1 lub odwrotnie. Weźmy pod uwagę kod (k, n, m) tego typu co w tabl. 6.1, o k pozycjach kontrolnych i całkowitej liczbie pozycji n. Dla danego ciągu n cyfr binarnych mamy następujące prawdopodobieństwa:

$$\left.\begin{aligned}
&\pi_0 = (1-p)^n && \text{— prawdopodobieństwo braku błędu,}\\
&\pi_1 = np(1-p)^{n-1} && \text{— prawdopodobieństwo wystąpienia jednego błędu,}\\
&\pi_2 = \frac{n(n-1)}{2}\,p^2(1-p)^{n-2} && \text{— prawdopodobieństwo dwu błędów na dwu różnych pozycjach,}\\
&\pi_l = \frac{n}{l!(n-l)!}\,p^l(1-p)^{n-l} && \text{— prawdopodobieństwo } l \text{ błędów na } l \text{ różnych pozycjach.}
\end{aligned}\right\} \quad (6.8)$$

Aby pokazać, jak uzyskuje się te wzory, obliczmy prawdopodobieństwo wystąpienia jednego błędu. W takim przypadku mamy $(n-1)$ prawidłowych cyfr z prawdopodobieństwem $(1-p)^{n-1}$ i jedną fałszywą (z prawdopodobieństwem p), która może się znajdować na dowolnej spośród n pozycji. Prowadzi to do wzoru na π_1 z zestawienia (6.8).

Inne prawdopodobieństwa obliczane są przy założeniu, że błędy występują zawsze na różnych pozycjach. Takie założenie jest uzasadnione tym, że dwa błędy nałożone na siebie w jednej pozycji nawzajem się znoszą. Jeśli na przykład na pewnej pozycji mamy cyfrę 1, to jeden błąd da nam 0, zaś po następnym znowu otrzymamy 1. W ten sposób podwójne błędy zawsze prowadzą do zmniejszenia ogólnej liczby błędów. Współczynniki π_l są rzeczywiście prawdopodobieństwami, ponieważ ich suma (zgodnie z twierdzeniem o dwumianie) równa się jeden:

$$\sum_{l=0}^{l=n} \pi_l = \sum_{l=0}^{l=n} \frac{n!}{l!(n-l)!}\,p^l(1-p)^{n-l} = [(1-p)+p]^n = 1^n = 1. \quad (6.9)$$

Jeżeli w odebranym ciągu n cyfr nie ma żadnego błędu lub jest tylko jeden, można prawidłowo odtworzyć m cyfr przenoszących informację. Jeśli zaś w ciągu wystąpi 2, 3, ..., l, ... błędów, cyfr tych nie można prawidłowo odtworzyć. Prawdopodobieństwo błędnego odebrania m cyfr przenoszących informację będzie więc równe

$$P = \pi_2+\pi_3+\ldots+\pi_n = 1-\pi_0-\pi_1. \tag{6.10}$$

Załóżmy, że p jest małe, i pomińmy wyrazy z p^3 i p^4, ... Wyrazem dominującym w P jest π_2 i stąd

$$P \approx \pi_2 \approx {}^1/_2 n(n-1)p^2. \tag{6.11}$$

Wartość P można porównać z oczekiwaną liczbą błędów w przypadku, kiedy nie stosujemy kodu korekcyjnego. Dla m pozycji przenoszących informację liczba ta jest równa

$$1-(1-p)^m = mp-{}^1/_2 m(m-1)p^2 \ldots \approx mp.$$

Stosunek mp/P może służyć jako miara efektywności kodu. Możemy także

Tablica 6.5

Własności różnych kodów korekcyjnych

Kod			mp	P	mp/P	p'
k	n	m				
2	3	1	p	$3p^2$	$\frac{1}{3p}$	$3p^2$
3	7	4	$4p$	$21p^2$	$\frac{4}{21p}$	$\frac{21}{4}p^2$
4	15	11	$11p$	$105p^2$	$\frac{11}{105p}$	$\frac{105}{11}p^2$
5	31	26	$26p$	$465p^2$	$\frac{26}{465p}$	$\frac{465}{26}p^2$

zdefiniować wielkość $p' = P/m$, którą można uważać za nowe prawdopodobieństwo błędu na cyfrę po korekcji. W tablicy 6.5 podano wartości mp, P, mp/P i p' dla kilku różnych kodów.

Rozpatrzmy sytuację, w której występuje średnio jeden błąd na sto cyfr kodu bez korekcji, tzn. $p = 10^{-2}$. Jeżeli przyjmiemy kod korekcyjny „cztery z siedmiu", będziemy mieli $P = 21 \cdot 10^{-4}$, co po korekcji daje średnio jeden błąd na 475 cyfr. Efektywność tego kodu wyraża się liczbą $4 \cdot 10^2/21 = 19$. Prawdopodobieństwo p' błędu na cyfrę po korekcji wynosi $p' = 5{,}25 \cdot 10^{-4}$.

6.5. PRZEPUSTOWOŚĆ BINARNEGO KANAŁU Z SZUMEM

Porównajmy otrzymane wyniki z teoretycznym wzorem Shannona*. Rozpatrując ogólne zagadnienie dyskretnego kanału z szumem, Shannon otrzymał dla kanału binarnego następujący wzór:

$$C = 1 + p \lg_2 p + (1-p) \lg_2 (1-p), \tag{6.12}$$

gdzie C jest przepustowością kanału z szumem w bitach na symbol, p zaś prawdopodobieństwem błędu jednego znaku binarnego. Wzór ten można uzasadnić w sposób elementarny. Przesyłamy przez kanał telegram złożony z N cyfr binarnych; z tych N cyfr Np będzie przesłanych nieprawidłowo. Do skorygowania tych błędów trzeba znać jedynie ich pozycję w ciągu cyfr tworzących telegram. Jeżeli fałszywą cyfrą jest 1, błąd korygujemy zastępując tę cyfrę zerem i odwrotnie. Informację o położeniu błędów możemy przesłać dołączając do telegramu dodatkowe sygnały (1 dla błędu, 0 dla braku błędu). Do przesłania tych dodatkowych sygnałow wymagany jest kanał o przepustowości

$$-p \lg_2 p - (1-p) \lg_2 (1-p)$$

bitów na symbol. Ta dodatkowa informacja musi być przesłana przez nasz kanał, co obniża jego przepustowość. Kanał bez szumu pierwotnie ma przepustowość $C = 1$, zaś kanał z szumem przepustowość określoną przez równanie (6.12). Teoretycznie powinniśmy znaleźć sposób kodowania umożliwiający osiągnięcie takiej przepustowości.

W naszym poprzednim przykładzie $p = 10^{-2}$, wyrażenie (6.12) daje wartość przepustowości rzędu 0,92, co oznacza, że do korekcji błędów po-

* C. E. Shannon, W. Weaver, *The Mathematical Theory of Communication*, Urbana, Ill. 1949, Univ. of Illinois Press, str. 38.

winna wystarczyć tylko jedna dodatkowa cyfra binarna na dziewięć wyjściowych. Nasz kod *cztery z siedmiu* nie zapewnia pełnej korekcji i wydłuża telegram o $^3/_4$ zamiast o $^1/_9$. Jesteśmy więc jeszcze daleko od granicy teoretycznej.

E. N. Gilbert* dochodzi do tego samego wniosku nie podając żadnego praktycznego rozwiązania. Praca jego zawiera jednak wiele interesujących rozważań dotyczących własności wielu rodzajów kodów i ich efektywności.

Problem korekcji błędów jest jednym z najważniejszych w telekomunikacji i obecnie ukazało się wiele prac dotyczących tej tematyki. Metoda Hamminga, opisana w rozdz. 6.2, została stworzona w celu automatycznej korekcji błędnych symboli. Bardziej ekonomiczną metodę podał C. A. Wagner z Massachusetts Institute of Technology. Polega ona na określaniu cyfry binarnej, dla której prawdopodobieństwo błędu jest największe, i korygowaniu go. Metodę tę rozważali również Silverman i Balser**, wydaje się być ona obiecująca. Dla określenia, któremu z symboli odpowiada największe prawdopodobieństwo błędu, autorzy ci wykorzystywali pewne wstępne dane, dotyczące korelacji w przesyłanym telegramie. Jednakże przyjęcie korelacji jest równoważne przyjęciu rozwlekłości. Wynika stąd, że w takim przypadku liczba symboli wykorzystywanych przy transmisji jest zbyt wysoka. Wydaje się, że w obliczeniach, przeprowadzonych przez Silvermana i Balsera, dotyczących porównania liczby wymaganych symboli w ich metodzie oraz w metodzie Hamminga — nie uwzględniono tego faktu. Ogólna idea tego typu korekcji jest jednakże obiecująca. Dalszy rozwój teorii kodów korekcyjnych, z uwagi na jej duże znaczenie w telekomunikacji, należałoby z uwagą śledzić ([6.2]).

* E. N. Gilbert, Bell. System J. **31** (1952), str. 504.

** Silverman, Balser, Proc. I.R.E. **42** (1954), str. 1428; Trans. I.R.F. (P.G.I.T.) **4** (1954), str. 50.

7

ZASTOSOWANIE TEORII INFORMACJI DO ROZWIĄZYWANIA NIEKTÓRYCH ZAGADNIEŃ SPECJALNYCH

7.1. ZAGADNIENIE KLASYFIKACJI Z ZASTOSOWANIEM KOMÓRKI NEUTRALNEJ

Jeżeli chcemy klasyfikować dane (teksty, rośliny, błędy itd.) przez umieszczanie ich w komórkach klasyfikacyjnych („szufladkach") wchodzących w skład pewnego zespołu, to wyłania się zagadnienie optymalnego wyboru rozmiaru tych komórek („szufladek"). Jeśli bowiem „rozmiar" komórek jest zbyt „mały" w porównaniu z „rozmiarem" klasyfikowanych danych, wystąpi niejednoznaczność, jeśli zaś „rozmiar" jest zbyt „duży", klasyfikacja jest małą efektywna.

Obecnie omówimy dwa różne sposoby klasyfikacji. Pierwszy z nich polega na tym, że klasyfikowane dane, których przynależność do żadnej z komórek nie jest oczywista, umieszczamy w specjalnej komórce, którą będziemy nazywali *komórką neutralną**. Drugi sposób, to sposób *segregowania do wielu komórek* — tak nazywamy system, gdy klasyfikowane dane, których nie możemy umieścić w jednej z komórek, przypisujemy wielu komórkom naraz**. Interesujące jest porównanie efektywności*** obu tych sposobów

---

* Odpowiada to w bibliotekarstwie szufladce „różne".

** Ten sposób również jest często stosowany w bibliotekarstwie.

*** Zagadnienie to po raz pierwszy sformułował i rozważył D.K.C. MacDonald, J. Appl. Phys. 23 (1952), str. 529.

klasyfikacji. Rozważania nasze oprzemy na fundamentalnym wzorze (1.13):

$$I = -K \sum_j p_j \ln p_j. \tag{7.1}$$

Wyraża on ilość informacji w przypadku, gdy dokonujemy wyboru spośród pewnej liczby przypadków $a, b, \ldots, j, \ldots$, mających prawdopodobieństwa a priori $p_a, p_b, \ldots, p_j, \ldots$ Jeżeli chcemy wyrażać informację w jednostkach dwójkowych, to

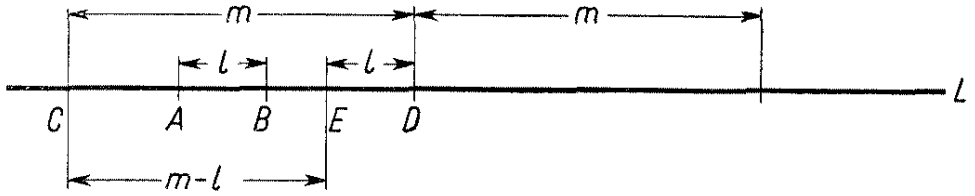
$$\left.\begin{aligned} K &= \lg_2 \mathrm{e}, \\ I &= -\sum_j p_j \lg_2 p_j \text{ bitów.} \end{aligned}\right\} \tag{7.1a}$$

Zajmiemy się uproszczonym modelem klasyfikacji, a mianowicie rozważymy klasyfikację położenia odcinka o zadanej długości l, na znacznie dłuższym odcinku o całkowitej długości L.

Przyjmijmy, że $L \geqslant l$ tak, aby można było pominąć sytuacje powstające na końcach odcinka. Odcinek o długości L dzielimy na N przylegających do siebie odcinków o długości m tak, że

$$Nm = L. \tag{7.2}$$

Odcinki te odpowiadają komórkom klasyfikacyjnym, a długość odcinka odpowiada rozmiarowi komórki. Chcemy znaleźć wartość optymalną m, w tym znaczeniu, iż daje ona największą zawartość informacji. Model ten został przedstawiony graficznie na rys. 7.1. Umieśćmy w sposób przypadkowy



Rys. 7.1. Model klasyfikacji z komórkami klasyfikacyjnymi o długości $m = CD$ i klasyfikowanymi elementami o długości $l = AB$

segment AB o długości l na odcinku L. Odcinek CD będzie jedną z komórek o długości m. Jeśli punkt A pada pomiędzy C i E, to cały odcinek AB leży wewnątrz tej komórki. Jeśli A znajduje się pomiędzy E i D, to odcinek AB przykrywa punkt styku D pomiędzy komórkami. Mamy wówczas

$$CD = m, \quad ED = l, \quad CE = m-l. \tag{7.3}$$

Załóżmy, że prawdopodobieństwa trafienia odcinka AB do różnych komórek są sobie równe. Prawdopodobieństwo zdarzenia, że AB w całości leży w i-tej komórce, wynosi

$$p_i = (m-l)/L. \tag{7.4}$$

Gdyby klasyfikowane odcinki były nieskończenie krótkie, to mielibyśmy $p_i = m/L$. Całkowite prawdopodobieństwo tego, że odcinek AB w całości leży w pewnej komórce, wynosi

$$P = \sum_{i=1}^{i=N} (m-l)/L = N(m-l)/L = 1-l/m. \tag{7.5}$$

Przypadek $m = 0$ jest trywialny i nie będziemy go rozważać, a więc obecność m w mianowniku nie komplikuje sytuacji.

Prawdopodobieństwo tego, że odcinek AB leży na granicy między komórkami i w związku z tym musi być umieszczony w pewnej dodatkowej neutralnej komórce, jest oczywiście równe

$$1-P \equiv q = l/m. \tag{7.6}$$

Zawartość informacji na symbol danych jest określona wzorem (7.1) i wynosi

$$\begin{aligned} I/K &= -\left\{\sum_{i=1}^{i=N} p_i \ln p_i + q \ln q\right\} = \\ &= -\left\{N\left(\frac{m-l}{L}\right)\ln\left(\frac{m-l}{L}\right) + \frac{l}{m}\ln\frac{l}{m}\right\} = \\ &= -\left\{(1-l/m)\ln\left(\frac{m-l}{L}\right) + \frac{l}{m}\ln\frac{l}{m}\right\}. \end{aligned} \tag{7.7}$$

Ponieważ chcemy obliczyć maksimum informacji, przyrównujemy pochodną I względem m do zera:

$$\frac{I}{K}\frac{\mathrm{d}I}{\mathrm{d}m} = -\frac{l}{m^2}\ln\left(\frac{m-l}{L}\right) - \left(1-\frac{l}{m}\right)\left(\frac{1}{m-l}\right) + \frac{l}{m^2}\cdot\ln\frac{l}{m} + \frac{l}{m^2} = 0$$

lub

$$-\frac{l}{m}\ln\left(\frac{m-l}{L}\right) - 1 + \frac{l}{m}\ln\frac{l}{m} + \frac{l}{m} = 0.$$

Stąd maksimum I wystąpi przy

$$\frac{l}{m}\left(1+\ln\left\{\frac{l}{m}\left(\frac{L}{m-l}\right)\right\}\right)=1$$

lub

$$m=l\left(1+\ln\frac{lL}{m(m-l)}\right). \tag{7.8}$$

Jest więc rzeczą oczywistą, że optymalna wartość m jest tego samego rzędu co l. Dokładne rozwiązanie zależy jednak (raczej nie krytycznie, z uwagi na zależność logarytmiczną) od wyboru L, a ściślej mówiąc od stosunku „długości" całego systemu klasyfikacji do „długości" klasyfikowanych elementów l. Weźmy np. $L=e^5 l$, tj. $L\approx 150\,l$. Oznaczając m/l przez x, z równania (7.8) otrzymujemy

$$x=6-\ln\{x(x-1)\}. \tag{7.9a}$$

Rozwiązaniem tego równania jest $x\approx 3{,}7$, tzn. że „rozmiar" komórki powinien być około cztery razy większy od „rozmiaru" klasyfikowanych elementów. Jeżeli jednak wziąć $L=e^{10}l$ (tzn. $L\approx 22\,000\,l$), otrzymamy

$$x=11-\ln\{x(x-1)\}, \tag{7.9b}$$

co daje $x\approx 8{,}0$.

7.2. METODA SEGREGOWANIA DO WIELU KOMÓREK

Rozważmy teraz model odpowiadający metodzie segregowania do wielu komórek, którą stosujemy w przypadku danych nie leżących w całości w jednej z komórek. Przede wszystkim załóżmy, że jeden klasyfikowany element może pokrywać tylko dwie sąsiednie komórki (ma to miejsce np. w przypadku systemu, w którym występują lokalizowane błędy). Dalsze nasze rozważania będą uzależnione od sposobu znajdowania miejsca pokrycia i jego interpretacji. Jeśli na przykład zajmiemy się systemem segregacji tekstów, to przydzielenie tekstu równocześnie do dwu komórek segregatora oznacza po prostu, że tekstowi przypisano szersze miejsce w przestrzeni klasyfikacji. Wyrażając to w terminologii, którą posługiwaliśmy się dla modelu z rozdz. 7.1, możemy powiedzieć, że długość odcinka, w którym może leżeć lewy koniec A klasyfiko-

wanego odcinka, wynosi $(2m-l)$. W ten sposób, pomijając znów sytuacje powstające na końcach, mielibyśmy teraz

$$I/K = \left\{\left[\sum_{i=1}^{i=N} \frac{m-l}{L} \ln\left(\frac{m-l}{L}\right)\right] + \frac{l}{m} \ln\left(\frac{2m-l}{L}\right)\right\}. \tag{7.10}$$

Zauważmy, że mnożnik w drugim wyrazie pozostał niezmieniony przy założeniu, że nie zmieniło się prawdopodobieństwo a priori l/m zachodzenia klasyfikowanego odcinka naraz na dwie komórki. Ilość informacji wynikająca z zaklasyfikowania do dwu komórek równocześnie: $\ln (2m-l)/L$ jest jednakże różna. Maksymalizując I względem m i oznaczając jak poprzednio $m/l = x$, otrzymujemy wyrażenie na wartość optymalną

$$x = \frac{2x-1}{2x+1} \ln \frac{2x-1}{x-1}. \tag{7.11}$$

Daje to $x \approx 1{,}06$. Jest rzeczą ciekawą, że rozwiązanie to nie zależy od L.

Jednakże w przypadku, kiedy w naszym systemie segregacji interesuje nas bezpośrednia akumulacja informacji statystycznej, znajomość miejsca zachodzenia klasyfikowanego odcinka na dwie komórki naraz pozwala na zlokalizowanie tego odcinka. Prowadzi to do zależności

$$I/K = -\left\{\left(1 - \frac{l}{m}\right) \ln \frac{m-l}{L} + \frac{l}{m} \ln \frac{l}{L}\right\}, \tag{7.12}$$

a stąd mamy warunek

$$x = -\ln (x-1), \tag{7.13}$$

którego rozwiązaniem jest $x \approx 1{,}28$. Wartość ta także nie zależy od L.

Porównamy obecnie ilość informacji przypadającej na klasyfikowany element dla różnych metod klasyfikacji, stosowanych w sposób optymalny. W pierwszym przypadku klasyfikacji do dwu komórek, przyjmując optymalną wartość m/l otrzymujemy:

$$\left(\frac{I}{K}\right)_{\text{opt}} = \ln \frac{L}{l} + \ln\left(\frac{l}{m-l}\right) - \frac{2m+l}{2m-l} \tag{7.14}$$

lub

$$(I/K)_{\text{opt}} = \ln (L/1) + 0{,}03, \tag{7.14a}$$

podczas gdy w drugim przypadku (wzór (7.12))

$$(I/K)_{\text{opt}} = \ln (L/l)+0{,}28. \tag{7.15}$$

Z porównania tej ostatniej wartości z przypadkiem, gdy stosujemy system z komórką neutralną, rozpatrywanym w rozdz. 7.1, widać, że system segregacji do wielu komórek jest lepszy przy warunku

$$\frac{m-l}{l} + \ln \frac{m-l}{l} \geqslant -0{,}28,$$

tj. gdy

$$L/l \geqslant 1{,}11. \tag{7.16}$$

Wartość L spełniająca ten warunek jest jednak na tyle mała, że segregowanie traci sens.

Analiza zagadnienia w przypadku klasyfikacji, kiedy klasyfikowane elementy mają pewien statystyczny rozkład długości, nie jest trudna. Jeżeli $p(l)$ jest gęstością rozkładu prawdopodobieństwa l, to dla systemu z komórką neutralną (równanie (7.7)) otrzymujemy

$$I/K = -\int_0^{\infty} \left\{ \left(1 - \frac{l}{m}\right) \ln \left[\left(1 - \frac{l}{m}\right) \frac{m}{L} \right] + \frac{l}{m} \ln \frac{l}{m} \right\} p(l)\, \mathrm{d}l. \tag{7.17}$$

Jeżeli przyjmiemy

$$p(l) = \begin{cases} l/l_0, & \text{gdy} \quad 0 \leqslant l \leqslant l_0, \\ 0, & \text{gdy} \quad l > l_0, \end{cases}$$

to po przekształceniach otrzymamy warunek uzyskania maksimum zawartości informacji:

$$m/l_0 = 1+\ln \left(\frac{Ll_0}{m(m-l_0)} \right) - \left(\frac{m}{l_0} \right)^2 \ln \left(\frac{m}{m-l_0} \right),$$

z którego dla rozpatrywanego poprzednio przypadku $L/l_0 = \mathrm{e}^5$ otrzymujemy

$$m/l_0 = 2{,}2. \tag{7.18}$$

We wszystkich tych przypadkach zajmowaliśmy się elementami, które mogą być klasyfikowane przez rozmieszczanie na linii, a więc w jednym wy-

miarze. Podane zależności możemy uogólnić na elementy zależne od 2, 3 lub więcej zmiennych, które powinny być klasyfikowane przez rozmieszczanie na płaszczyźnie, w przestrzeni lub w przestrzeni wielowymiarowej. Jednakże przy takich warunkach jeden element może pokrywać kilka komórek, co znacznie komplikuje sprawę.

7.3. NAJKORZYSTNIEJSZA LICZBA SYGNAŁÓW PRZYPADAJĄCYCH NA KOMÓRKĘ ELEMENTARNĄ

Zagadnienie to, zależnie od sposobu podejścia, ma różne rozwiązania. Przyjmijmy tu punkt widzenia inżyniera telekomunikacji, który mając do dyspozycji kanał charakteryzujący się określoną stałą czasu dąży do uzyskania jego maksymalnej efektywności. Poniższe rozważania przytaczamy zgodnie z R. M. Walkerem (IBM, Watson Laboratory). Ograniczymy się do kanałów bezszumowych. Zagadnieniami z szumem zajmiemy się w następnych rozdziałach. Przyjmijmy, że sieci elektryczne składające się na kanał charakteryzują się pewną stałą czasu τ. Wiadomo, że sieć inercyjna, w której zmagazynowano energię E_n, traci tę energię zgodnie z prawem

$$E = E_n \, \mathrm{e}^{-t/\tau} \, . \tag{7.19}$$

Przyjmijmy teraz, że sygnał ma postać impulsu o energii, która może przyjmować jedną z wartości

$$0, E_0, 2E_0, \ldots, E_n = (n-1)E_0. \tag{7.20}$$

Jeśli energia sygnału jest najwyższa i wynosi E_n oraz jest tracona zgodnie z równaniem (7.19), to aby odczytać poziom 0 lub E_0, musimy czekać, aż E opadnie poniżej $E_0/2$. Prowadzi to do warunku

$$\left.\begin{aligned} E_0/2 &= (n-1)E_0 \, \mathrm{e}^{-t/\tau}, \\ t/\tau &= -\ln \frac{1}{2(n-1)} = \ln 2(n-1). \end{aligned}\right\} \tag{7.21}$$

Jeżeli chcemy rozróżniać impulsy o energiach danych przez równanie (7.20), to czas t określony przez równanie (7.21) odpowiada najkrótszemu odcinkowi czasu pomiędzy kolejnymi rozróżnialnymi impulsami.

Dla długiego odcinka czasu całkowitą liczbę impulsów obliczamy, jak następuje:

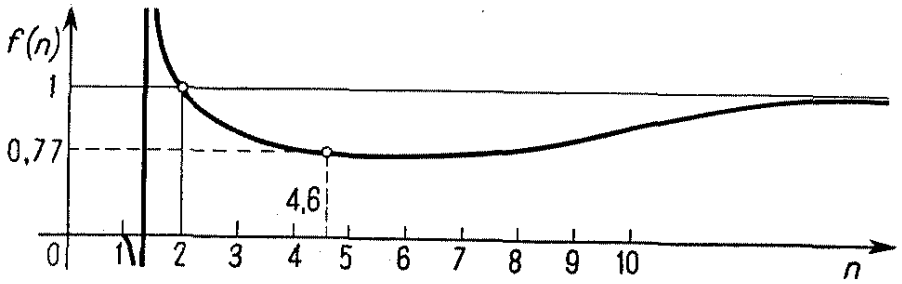
$$G = T/t = T/(\tau \ln 2(n-1)). \tag{7.22}$$

Przyjmując, że wszystkie impulsy mają równe prawdopodobieństwa a priori, otrzymujemy maksymalną możliwą zawartość informacji (wzór (1.6))

$$I = KG \ln n = (KT/\tau) f(n), \tag{7.23}$$

gdzie

$$f(n) = (\ln n)/(\ln 2(n-1)). \tag{7.24}$$



Rys. 7.2. Wykres $f(n) = (\ln n)/(\ln 2(n-1))$ w funkcji n

Możemy wykreślić krzywą $f(n)$ zmieniając n od $n = 2$ (kod binarny) do ∞ (zmienna ciągła). Krzywa ta została przedstawiona na rys. 7.2. Funkcja $f(n)$ przyjmuje wartość 1 dla $n = 2$, dla $n = 4{,}6$ osiąga minimum 0,77 i rośnie znów do 1 przy $n \to \infty$.

Jak więc widzimy, ilość informacji, jaka może być przesłana w danym czasie, jest maksymalna w przypadku kodu binarnego (telegrafia) lub w telefonii przy przesyłaniu zmiennej ciągłej.

8

ANALIZA SYGNAŁÓW: ANALIZA FOURIERA I METODY PRÓBKOWANIA

8.1. SZEREGI FOURIERA

W następnych rozdziałach będziemy korzystali z niektórych wyników analizy sygnałów. Obecny rozdział będzie stanowił systematyczny przegląd metod matematycznych stosowanych do analizy sygnałów i wzajemnych powiązań tych metod ([8.1]).

Zacznijmy od analizy Fourierowskiej funkcji okresowych*. Niech $f(t)$ będzie funkcją okresową o okresie τ. Funkcja ta może być przedstawiona w postaci sumy składowych harmonicznych. Będziemy stosowali postać zespoloną szeregu Fouriera:

$$f(t) = \sum_{n=-\infty}^{+\infty} C_n \, e^{in\omega_0 t}, \qquad \omega_0 = \frac{2\pi}{\tau}, \tag{8.1}$$

gdzie

$$C_n = \frac{1}{\tau} \int_{\cdot}^{\tau} f(t) \, e^{-in\omega_0 t} \, dt. \tag{8.2}$$

Jeżeli $f(t)$ jest funkcją rzeczywistą, to spełniony jest warunek

$$C_{-n} = C_n^*, \tag{8.3}$$

---

* Przejrzysty i ścisły wykład podstaw teorii szeregów Fouriera można znaleźć w pracy H. Bohra, *Almost Periodic Functions*, New York 1947, Chelsea, rozdz. 1.

gdzie gwiazdka oznacza liczbę zespoloną sprzężoną. Podano wiele praktycznych metod obliczania całki (8.2). Ze względu na ścisły związek z zagadnieniami, którymi będziemy się zajmować dalej, rozważymy tu metodę przedziałów skończonych. Wybierzmy N równoodległych punktów w przedziale $[0, \tau]$, który odpowiada jednemu okresowi, podstawiając

$$t_m = m\theta, \qquad \theta = \frac{\tau}{N}, \tag{8.4}$$

gdzie m jest liczbą całkowitą. Jako przybliżenie całki (8.2) weźmy sumę

$$\gamma_{n'} = \frac{1}{N} \sum_{m=1}^{N} f(t_m)\, \mathrm{e}^{-in'\omega_0 t_m}. \tag{8.5}$$

Podstawiając zamiast $f(t_m)$ szereg Fouriera (8.1), otrzymujemy

$$\gamma_{n'} = \frac{1}{N} \sum_{m=1}^{N} \sum_{n=-\infty}^{+\infty} C_n\, \mathrm{e}^{i(n-n')m\omega_0\theta}, \qquad \omega_0\theta = \frac{2\pi}{N}. \tag{8.6}$$

Możemy skorzystać z następującej zależności:

$$\sum_{m=1}^{N} \mathrm{e}^{ipm\frac{2\pi}{N}} = \begin{cases} N, & \text{gdy } p = qN, \\ 0, & \text{gdy } p \neq qN, \end{cases} \tag{8.7}$$

gdzie q jest liczbą całkowitą.
Dolny wzór wynika stąd, że kąty $pm\,2\pi/N$ są równorozłożone pomiędzy 0 i $2p\pi$. Wyrazy niezerowe sumy (8.6) odpowiadają wskaźnikom

$$p = n - n' = qN,$$

gdzie q jest dowolną liczbą całkowitą; stąd

$$\gamma_{n'} = \sum C_n = \sum_q C_{n'+qN}. \tag{8.8}$$

Współczynnik $\gamma_{n'}$ obliczony ze wzoru (8.5) jest sumą współczynników C_n ze wskaźnikami $n = n' + qN$. Poza tym współczynniki $\gamma_{n'}$ są okresowe względem n' z okresem równym N. Wnioski te są szczególnym przypadkiem wyniku bardziej ogólnego*.

* L. Brillouin, *Wave Propagation in Periodic Structures*, 1st ed. New York 1946, McGraw-Hill, 2nd ed., New York 1952, Dover Publications; zob. np. rozdz. 1, str. 7.

Interesujący jest przypadek, gdy w rozwinięciu Fouriera funkcji rzeczywistej składowe harmoniczne o numerach większych od pewnego granicznego n_M stają się znikomo małe, tak że można je odrzucić:

$$C_n = 0, \quad \text{jeżeli} \quad |n| > n_M, \quad C_{-n} = C_n^*. \tag{8.9}$$

Wówczas możemy wziąć

$$N = 2n_M + 1. \tag{8.10}$$

Stąd mamy

$$\gamma_n = C_n, \quad \text{jeżeli} \quad |n| \leqslant n_M. \tag{8.11}$$

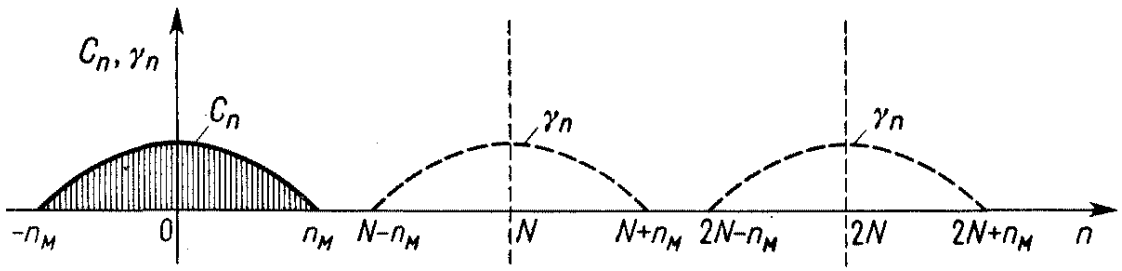
Pierwsze N wyrazów γ_n daje dokładnie N pierwszych współczynników Fouriera C_n, wyrazy γ_n o większych numerach są powtórzeniem współczynników C_n — okres powtarzania wynosi N (rys. 8.1).

Bardzo ważnym wynikiem analizy Fouriera jest *wzór Parsevala*. Jest on słuszny w przypadku, kiedy funkcja $f(t)$ jest rzeczywista, ograniczona i całkowalna w przedziale $[0, \tau]$ odpowiadającym jednemu okresowi. Wzór Parsevala ma postać:

$$\frac{1}{\tau}\int_0^\tau f^2\,\mathrm{d}t = \sum_{n=-\infty}^{+\infty} |C_n|^2. \tag{8.12}$$

Wzór ten łatwo można udowodnić podstawiając w miejsce jednej z funkcji $f(t)$, w wyrażeniu podcałkowym, po lewej stronie wzoru (8.12), jej rozwinięcie Fouriera (8.1):

$$\frac{1}{\tau}\int_0^\tau f^2\,\mathrm{d}t = \frac{1}{\tau}\int_0^\tau f \sum_n C_n\,\mathrm{e}^{\mathrm{i}n\omega_0 t}\,\mathrm{d}t = \sum_n C_n C_n^* = \sum_n |C_n|^2.$$



Rys. 8.1. Wykres ilustrujący równość współczynników Fouriera C_n i współczynników γ_n dla $-n_M \leqslant n \leqslant +n_M$ oraz okresowość γ_n dla większych wartości n

W wielu zastosowaniach wielkości występujące we wzorze (8.12) przedstawiają energię; w takim przypadku twierdzenie Parsevala mówi, że energię przypadającą na okres τ można obliczyć bądź to całkując f^2, bądź sumując energie składowych harmonicznych.

8.2. ZJAWISKO GIBBSA I ZBIEŻNOŚĆ SZEREGÓW FOURIERA

Warunki zbieżności szeregów Fouriera omawiane są w wielu podręcznikach*. Rozwinięcie Fouriera można stosować dla funkcji posiadającej skończoną liczbę punktów nieciągłości w obrębie okresu τ. Niech t_0 oznacza punkt nieciągłości z lewostronną i prawostronną wartością $f(t_0+0)$ i $f(t_0-0)$ oraz załóżmy, że w punkcie t_0 istnieje lewostronna i prawostronna pochodna. Można udowodnić, że szereg Fouriera dla $t = t_0$ jest zbieżny do wartości

$$\tfrac{1}{2}[f(t_0-0)+f(t_0+0)]. \tag{8.13}$$

Wynik ten ilustruje szczególny, lecz typowy przykład, który teraz rozpatrzymy. Rozważmy funkcję o okresie 2π mającą nieciągłości dla $t = \pm\pi, \pm 3\pi, \ldots$ (rys. 8.2):

$$f(t) = t, \qquad -\pi < t < +\pi, \tag{8.14}$$

przy czym wielkość uskoku dla $t = \pi$ wynosi -2π. Szereg Fouriera tej funkcji ma postać

$$f(t) = 2\left(\frac{\sin t}{1} - \frac{\sin 2t}{2} + \frac{\sin 3t}{3} \ldots\right). \tag{8.15}$$

Oznaczmy przez S_p sumę pierwszych p wyrazów i porównajmy ją z rozwijaną funkcją. Różnica Δ wynosi

$$\Delta = f(t) - S_p = 2\sum_{n=p+1}^{\infty} (-1)^{n+1}\frac{\sin nt}{n}. \tag{8.16}$$

Chcemy zbadać, jak będzie się zachowywać różnica Δ w okolicy nieciągłości

* R. V. Churchill, *Fourier Series and Boundary Value Problems*, New York 1941, Mc Graw-Hill; J. S. Sokolnikoff, E. S. Sokolnikoff, *Higher Mathematics for Engineers and Physicists*, New York 1941, Mc Graw-Hill, str. 68.

przy $t = \pi$. Weźmy

$$t = \pi - \theta,$$

wówczas

$$\sin nt = (-1)^{n+1} \sin n\theta,$$

skąd

$$\Delta = f - S_p = 2 \sum_{n=p+1}^{\infty} \frac{\sin n\theta}{n}.$$

Kolejne wartości n różnią się o $\Delta n = 1$, zaś o θ zakładamy, że jest bardzo małe. Wprowadźmy nową zmienną

$$\eta = n\theta, \qquad \Delta\eta = \theta, \qquad \theta \ll 1.$$

Zgodnie z podstawowym twierdzeniem rachunku całkowego możemy sumę we wzorze (8.16) zamienić na całkę

$$\Delta = 2 \int_{\eta=(p+1)\theta}^{\infty} \frac{\sin \eta}{\eta} \, d\eta = -2 \operatorname{Si}(p+1)\theta, \tag{8.17}$$

gdzie przez $\operatorname{Si}(x)$ oznaczamy, jak zazwyczaj, sinus całkowy*. Funkcja Si przyjmuje w punkcie $x = 0$ wartość $-\pi/2$ i przy x dążącym do nieskończoności wykazuje tłumione oscylacje. Kolejne maksima i minima funkcji Si występują dla

$$\frac{(p+1)\theta}{\pi} = \frac{x}{\pi} = \begin{cases} 1 \\ 2 \\ 3 \\ 4 \end{cases}; \qquad \text{wówczas} \qquad \operatorname{Si}(x) = \begin{cases} +0{,}281 \\ -0{,}153 \\ +0{,}104 \\ -0{,}079 \end{cases}. \tag{8.18}$$

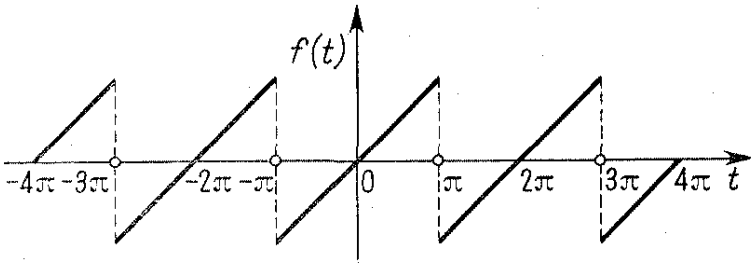
Wykresy kolejnych przybliżeń funkcji $f(t)$ pokazano na rys. 8.2. Zanotujmy następujące fakty:

(1) Wszystkie krzywe odpowiadające sumom częściowym S_p przechodzą

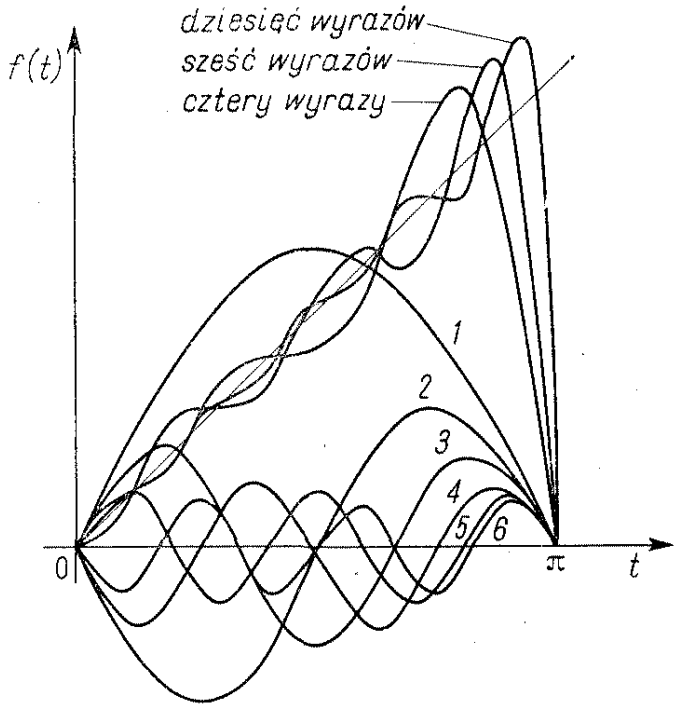
* Jahnke — Emde, *Tables of Functions*, New York 1945, Dover Publications, str. 9.

przez jeden punkt $t = \pi$, $S_p = 0$. Stąd punkt w środku skoku jest punktem, do którego szereg jest zbieżny, co jest zgodne ze wzorem (8.13).

(2) Sumy częściowe wykazują duże oscylacje w otoczeniu punktu nieciągłości, podczas gdy nie wykazuje ich funkcja rozwijana. Kiedy liczba p



Rys. 8.2a. Wykres nieciągłej funkcji okresowej o okresie 2π
$f(t) = t, \quad -\pi < t < +\pi$



Rys. 8.2b. Sześć pierwszych wyrazów szeregu Fouriera funkcji $f(t)$. Pokazano także sumy pierwszych czterech, sześciu i dziesięciu wyrazów tego szeregu

wyrazów w sumie częściowej jest dostatecznie duża, oscylacje te można przedstawić standardową krzywą Δ (równanie (8.17)). Oscylacje mają zawsze taką samą amplitudę, natomiast położenia maksimów i minimów wyznaczają zależności (8.18). Pierwszy przerzut występuje dla $\theta = \pi/p$; w tym przypadku

suma przekracza wartość funkcji o 0,281, co daje błąd względny

$$\frac{\Delta}{f} = \frac{\Delta}{\pi} = \frac{2}{\pi}(0{,}281) = 18\%. \tag{8.19}$$

Jeśli liczba p wyrazów sumy częściowej wzrasta, długość przedziału, w którym występuje błąd, maleje, jednakże maksima i minima zachowują tę samą amplitudę. Efekt ten jest znany jako *zjawisko Gibbsa*. Na rysunkach 8.2 łatwo jest zauważyć, że wysokość przerzutu wynosi zawsze 18%, gdy suma obejmuje 4, 6 lub 10 wyrazów.

Wyniki, do jakich doszliśmy w tym szczególnym przypadku, mają charakter ogólny i pozostają słuszne dla dowolnej funkcji z uskokami. Rozważmy funkcję okresową $F(t)$ o okresie 2π mającą uskoki $D_1, D_2, D_3, \ldots$ w punktach $t_1, t_2, t_3, \ldots$. Możemy skonstruować funkcję ciągłą

$$G = F + \frac{D_1}{2\pi} f(t - t_1 + \pi) + \frac{D_2}{2\pi} f(t - t_2 + \pi) \ldots$$

dodając do funkcji F odpowiednią liczbę funkcji typu (8.14) pomnożonych przez odpowiednie współczynniki. Szereg Fouriera funkcji ciągłej jest szybkozbieżny; jego współczynniki maleją jak $1/n^2$ lub jeszcze szybciej. Nieciągłości funkcji f dają natomiast współczynniki malejące powoli, a mianowicie tylko jak $1/n$ (wzór (8.15)). Zachowanie się szeregu Fouriera funkcji F w pobliżu nieciągłości jest w głównej mierze określone tymi, wolno malejącymi wyrazami, wykazującymi własności takie same, jak w omówionym wyżej przypadku szczególnym.

Na zakończenie podkreślimy jeszcze inną ważną własność szeregu Fouriera. Suma częściowa S_p jest mianowicie najlepszym możliwym, w sensie najmniejszego błędu średniokwadratowego, przybliżeniem rozwijanej funkcji*.

8.3. CAŁKI FOURIERA

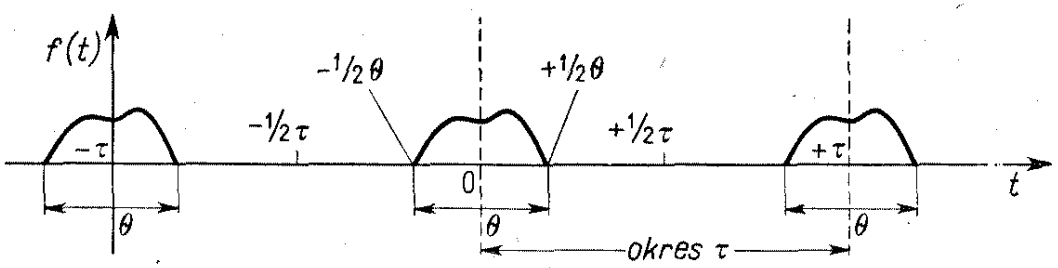
Całki Fouriera zilustrujemy najpierw na prostym przykładzie związanym z praktycznymi zagadnieniami techniki impulsowej. Rozważmy impuls o czasie trwania θ, powtarzany okresowo (okres τ). Rozwińmy ciąg takich

* Por. notka * do str. 114.

impulsów w szereg Fouriera zgodnie z równaniami (8.1) i (8.2) (zob. rys. 8.3):

$$\left.\begin{aligned} f(t) &= \sum_{n=-\infty}^{+\infty} C_n \, \mathrm{e}^{\mathrm{i}n\omega_0 t}, \quad \omega_0 = 2\pi\nu_0 = \frac{2\pi}{\tau}, \\ C_n &= \frac{1}{\tau}\int_{-\tau/2}^{\tau/2} f(t)\, \mathrm{e}^{-\mathrm{i}n\omega_0 t}\, \mathrm{d}t, \quad C_{-n} = C_n^*, \end{aligned}\right\} \tag{8.20}$$

przy czym zakładamy, że $f(t)$ jest funkcją rzeczywistą. Ponieważ na zewnątrz przedziału θ impuls jest równy zero, można przyjąć dla C_n granice całko-



Rys. 8.3. Impulsy o czasie trwania θ i okresie powtarzania τ

wania $\pm\theta/2$. Przypuśćmy teraz, że zwiększamy stopniowo odległość τ pomiędzy kolejnymi impulsami i przechodzimy do granicy $\tau\to\infty$. Funkcja $f(t)$ może być zapisana w postaci

$$f(t) = \sum_{-\infty}^{\infty} C_n \, \mathrm{e}^{\mathrm{i}2\pi\nu_n t}\, \Delta n = \sum_{-\infty}^{\infty} C_n\tau \, \mathrm{e}^{\mathrm{i}2\pi\nu t}\, \Delta\nu, \quad \nu_n = n\nu_0, \quad \Delta n = 1,$$

$$\Delta\nu = \nu_0 = 1/\tau.$$

Dla bardzo dużych τ sumę możemy zastąpić całką:

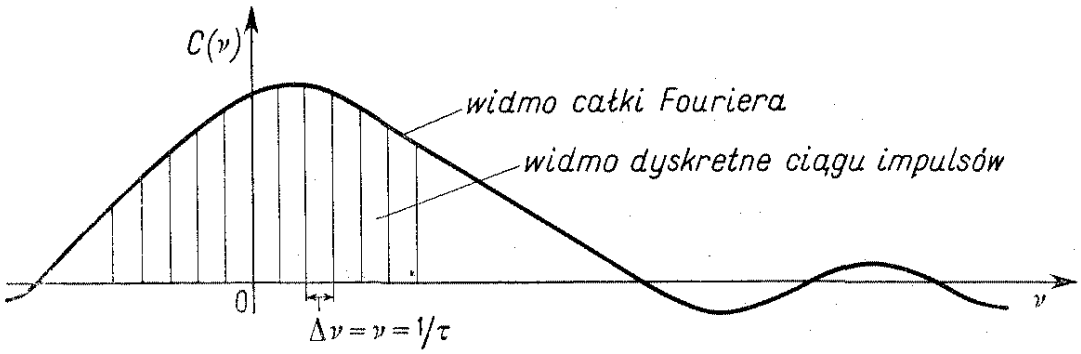
$$\tau\to\infty: \quad f(t) = \int_{\nu=-\infty}^{\infty} C(\nu)\, \mathrm{e}^{\mathrm{i}2\pi\nu t}\, \mathrm{d}\nu \tag{8.21}$$

oraz

$$\lim_{\tau\to\infty}(C_n\tau) = C(\nu) = \int_{t=-\infty}^{\infty} f(t)\, \mathrm{e}^{-\mathrm{i}2\pi\nu t}\, \mathrm{d}t = C^*(-\nu). \tag{8.22}$$

W całce ze wzoru (8.22) całkowanie obejmuje faktycznie tylko przedział θ, odpowiadający pojedynczemu impulsowi, ponieważ $f(t)\neq 0$ tylko dla $-\theta/2 < t < \theta/2$.

Dla dużego τ, współczynniki $C_n\tau$ tworzą widmo prążkowe, obwiednią którego jest ciągłe widmo odpowiadające całce Fouriera. Im większe jest τ, tym mniejszy jest przedział $\Delta\nu = 1/\tau$ pomiędzy prążkami widma prążko-



Rys. 8.4. Ciągłe widmo $C(\nu)$ całki Fouriera jako obwiednia widma dyskretnego $C_n\tau$

wego (rys. 8.4). Z równań (8.21) i (8.22) wynika, że zmienne t i ν odpowiadają sobie nawzajem w tym sensie, że jeśli impuls $f(t)$ ma widmo $C(\nu)$, to impuls $C(t)$ ma widmo $f(\nu)$.

Jeżeli impuls jest symetryczny, to jego widmo również jest symetryczne:

$$f(t) = f(-t) = \int_{\nu=-\infty}^{\infty} C(\nu) \cos 2\pi\nu t \, d\nu, \tag{8.23}$$

$$C(\nu) = C(-\nu) = \int_{t=-\infty}^{\infty} f(t) \cos 2\pi\nu t \, dt. \tag{8.24}$$

Są to wyrażenia podstawowe w naszych rozważaniach. Podamy teraz kilka przykładów:

impuls prostokątny

$$f(t) = \begin{cases} 1 & |t| < (1/2)\theta_1 \\ 0 & |t| > (1/2)\theta_1 \end{cases}$$

widmo

$$C(\nu) = \theta_1 \frac{\sin \pi\nu\theta_1}{\pi\nu\theta_1}; \tag{8.25}$$

impuls typu $\frac{\sin t}{t}$

$$f(t) = 2\nu_M \frac{\sin 2\pi\nu_M t}{2\pi\nu_M t}$$

widmo prostokątne

$$C(\nu) = \begin{cases} 1 & |\nu| < \nu_M \\ 0 & |\nu| > \nu_M \end{cases}; \tag{8.26}$$

impuls trójkątny

$$f(t) = \begin{cases} \left(1 - \frac{|t|}{\theta_1}\right) & |t| < \theta_1 \\ 0 & |t| > \theta_1 \end{cases}$$

$$C(\nu) = \theta_1 \left(\frac{\sin \pi\nu\theta_1}{\pi\nu\theta_1}\right)^2; \tag{8.27}$$

impuls gaussowski

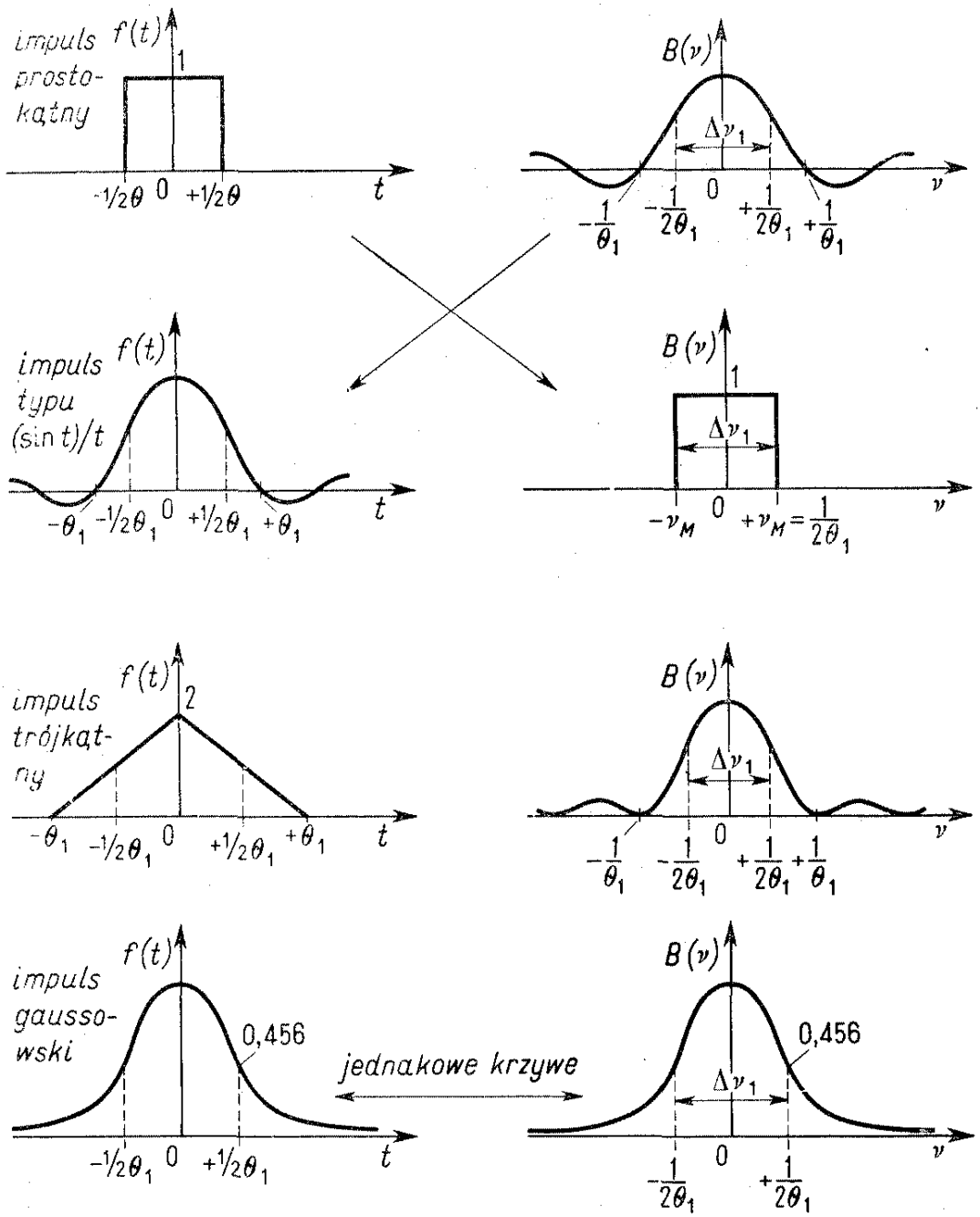
$$f(t) = e^{-\pi(t/\theta_1)^2}$$

widmo gaussowskie

$$C(\nu) = \theta_1 e^{-\pi(\nu\theta_1)^2}. \tag{8.28}$$

Odpowiednie wykresy pokazano na rys. 8.5. Pierwsze dwa przypadki ilustrują omawianą poprzednio odpowiedniość zmiennych t, ν oraz ich symetrię w równaniach (8.23) i (8.24).

Dla całki Fouriera słuszny jest wzór Parsevala podobny do wzoru (8.12), słusznego dla szeregu Fouriera. Wzór ten wyprowadzimy w następujący



Rys. 8.5. Przykłady funkcji i ich widm. Krzywe te odpowiadają wyrażeniom (8.25) do (8.28). Warto zwrócić uwagę na symetrię t—ν (zaznaczoną strzałkami) dla pierwszych dwóch przypadków oraz dla impulsu i jego widma w ostatnim przypadku

prosty sposób:

$$\int |f|^2 \, dt = \int f^* f \, dt = \int f^* \, dt \int C(\nu) e^{i2\pi\nu t} \, d\nu,$$

w wyrażeniu tym w miejsce funkcji $f(t)$ podstawiliśmy jej całkę Fouriera (8.21). Całkując ze względu na t mamy

$$\int f^*(t) \, e^{i2\pi\nu t} \, d\nu = C^*(\nu),$$

stąd

$$\int_{-\infty}^{+\infty} |f(t)|^2 \, dt = \int_{-\infty}^{+\infty} |C(\nu)|^2 \, d\nu = E. \tag{8.29}$$

W wielu zagadnieniach fizycznych całka ta odpowiada całkowitej energii fali o amplitudzie $f(t)$ i wielkość E może być obliczana albo bezpośrednio na podstawie funkcji $f(t)$, bądź na podstawie jej fourierowskiego widma.

8.4. ROLA SKOŃCZONEJ SZEROKOŚCI ZAJMOWANEGO PASMA CZĘSTOTLIWOŚCI

Jedną z charakterystycznych cech każdego systemu telekomunikacyjnego jest to, że zawsze wykorzystuje się w nim tylko skończone pasmo częstotliwości. System może być typu *dolnoprzepustowego*, jeśli wykorzystuje się częstotliwości od 0 do pewnej maksymalnej wartości ν_M, lub może być *środkowoprzepustowy*, jeśli przyjęto w nim częstotliwości pomiędzy dwoma wartościami ν_1 i ν_2. Systemy, które można nazwać *górnoprzepustowymi*, są czysto hipotetyczne, ponieważ nie istnieje żaden układ fizyczny bez górnej granicy częstotliwości. Analizę systemu *pasmowoprzepustowego* można sprowadzić do analizy systemu dolnoprzepustowego, modulującego pewną częstotliwość nośną. Dlatego możemy ograniczyć nasze rozważania do systemów dolnoprzepustowych.

Rozważmy zagadnienie, podczas gdy dany jest system o pewnej górnej granicy częstotliwości ν_M. Jakie sygnały możemy przekazywać w takim systemie? Innymi słowy, jakie są cechy charakterystyczne funkcji $f(t)$, których widmo Fouriera nie przekracza wartości ν_M?

Warunek ten oznacza, że wykorzystujemy pasmo częstotliwości $\Delta\nu_1$:

$$|\nu| \leqslant \nu_M, \quad -\nu_M \leqslant \nu \leqslant \nu_M, \quad \Delta\nu_1 = 2\nu_M. \tag{8.30}$$

Na wstępie chcemy wykazać, że dla odtworzenia sygnału o czasie trwania θ_1 konieczne jest pasmo częstotliwości o szerokości

$$\Delta\nu_1 \geqslant \frac{1}{\theta_1}, \quad \Delta\nu_1 \cdot \theta_1 \geqslant 1. \tag{8.30a}$$

W praktyce pasmo częstotliwości nie jest dokładnie określone. Dla częstotliwości leżących wewnątrz pasma, wartości $C(\nu)$ są duże w porównaniu z wartościami dla częstotliwości leżących na zewnątrz pasma. Podobnie funkcja $f(t)$ przedstawia sygnał o stosunkowo dużej amplitudzie w przedziale czasu θ_1 i stosunkowo małej amplitudzie dla pozostałych chwil na osi czasu. Iloczyn $\Delta\nu_1 \cdot \theta_1$ ma kres dolny. Możemy to zauważyć na tle przykładów (8.25) do (8.28), gdzie sygnały o krótszym czasie trwania mają szersze pasmo częstotliwości.

Zagadnienie ilościowego określenia szerokości zajmowanego pasma częstotliwości i czasu trwania sygnału jest zagadnieniem o charakterze empirycznym. Wyniki obserwacji z dziedziny telewizji, które obecnie omówimy, dają dobre podstawy do wyboru definicji ilościowych. Ustalono mianowicie, że do przesłania sygnału telewizyjnego przy 525 liniach na obraz, 30 obrazach na sekundę i 360 impulsach na linię potrzebne jest pasmo o szerokości około 6 megacykli*. Daje to długość impulsu

$$\theta_1 = \frac{1}{525 \cdot 30 \cdot 360} = \frac{1}{5{,}7} \cdot 10^{-6};$$

stąd

$$\Delta\nu_1 \cdot \theta_1 = 6 \cdot 10^6 \cdot \frac{1}{5{,}7} \cdot 10^{-6} = 1{,}05.$$

Przykład ten sugeruje, że szerokość zajmowanego pasma i czas trwania sygnału muszą spełniać nierówność (8.30a):

$$\Delta\nu_1 \cdot \theta_1 \geqslant 1.$$

Rozważmy teraz przykłady (8.25) do (8.28) z poprzedniego rozdziału. Są to krótkie, symetryczne impulsy o maksymalnej amplitudzie dla $t = 0$ i rzeczy-

* Jest to standard amerykański. (*Przyp. tłum.*).

wistej fourierowskiej gęstości widmowej $C(\nu)$ mającej maksimum dla $\nu = 0$. Pokażemy, że dla następujących definicji $\Delta\nu_1$ i θ_1 spełniona jest nierówność* (8.30a):

$$f(0) \cdot \theta_1 = \int_{-\infty}^{+\infty} f(t)\, dt, \tag{8.31}$$

gdzie θ_1 — nominalny czas trwania impulsu,

$$C(0) \cdot \Delta\nu_1 = \int_{-\infty}^{+\infty} C(\nu)\, d\nu, \tag{8.32}$$

gdzie $\Delta\nu_1$ — nominalna szerokość pasma.

Zauważmy teraz, że całka w równaniu (8.31) jest równa wielkości $C(0)$ występującej w równaniu (8.24). Podobnie całka w równaniu (8.32) jest równa wielkości $f(0)$ występującej w równaniu (8.23). Zatem

$$\theta_1 = \frac{C(0)}{f(0)}, \qquad \Delta\nu_1 = \frac{f(0)}{C(0)}, \qquad \theta_1 \cdot \Delta\nu_1 = 1. \tag{8.33}$$

Definicje te dają równość (8.33) zamiast nierówności (8.30). Łatwo można to sprawdzić na naszych poprzednich przykładach. Jednakże są to przypadki szczególne, nie odzwierciedlające ogólnej sytuacji, a więc wyrażenia (8.31) i (8.32) nie mogą być zastosowane do ogólnej, mającej praktyczne znaczenie, definicji czasu trwania sygnału lub szerokości pasma. Tak więc na przykład przy obliczaniu nominalnego czasu trwania sygnału θ_1, ujemne wartości f powinny dawać dodatni przyrost, podczas gdy w równaniu (8.31) występują one ze znakiem ujemnym. Możemy wprowadzić następujące zmodyfikowane definicje oparte na wartościach bezwzględnych:

$$|f(0)|\theta_2 = \int_{-\infty}^{+\infty} |f(t)|\, dt \geqslant \left| \int_{-\infty}^{+\infty} f(t)\, dt \right| = |C(0)| \tag{8.34}$$

i podobnie

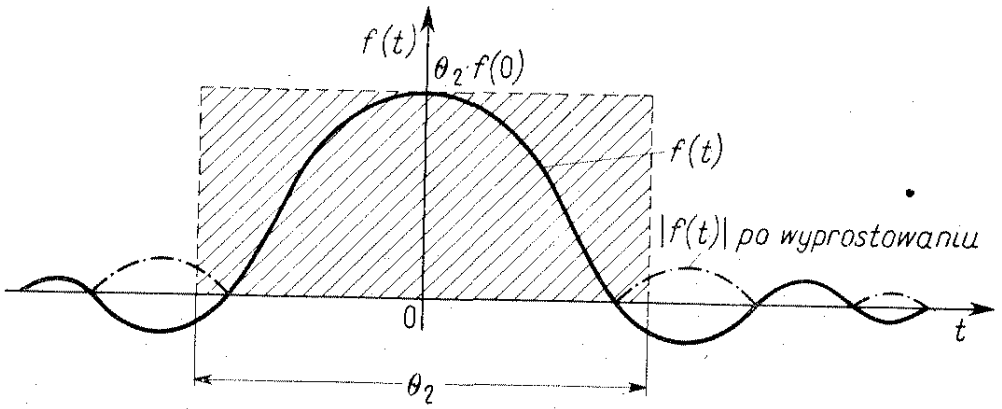
$$|C(0)|\Delta\nu_2 = \int_{-\infty}^{+\infty} |C(\nu)|\, d\nu \geqslant \left| \int_{-\infty}^{+\infty} C(\nu)\, d\nu \right| = |f(0)|. \tag{8.35}$$

Z definicji otrzymujemy

$$\theta_2 \geqslant \theta_1, \qquad \Delta\nu_2 \geqslant \Delta\nu_1, \qquad \theta_2\Delta\nu_2 \geqslant \theta_1\Delta\nu_1 = 1, \tag{8.36}$$

* H. A. Wheeler, Proc. I.R.E. **27** (1939), str. 359; P. Le Corbeiller, *Electronic Circuits and Tubes*, New York 1947, McGraw-Hill, str. 586—590.

co stanowi warunek (8.30a), który chcieliśmy spełnić. Rysunek 8.6 ilustruje sposób korzystania z definicji (8.34). Długość θ_2 jest wybrana tak, aby pole prostokąta o podstawie θ_2 i wysokości $f(0)$ było równe polu pod krzywą $|f(t)|$. Jeśli chodzi o szerokość pasma sytuacja przedstawia się podobnie. Czas trwania θ_2 może być zmniejszony, jeżeli początek osi czasu przesuniemy



Rys. 8.6. Czas trwania θ_2 impulsu otrzymano przez porównanie iloczynu $f(0)\cdot\theta_2$ z obszarem pomiędzy wyprostowaną krzywą $|f(t)|$ i osią t

do punktu, gdzie $|f|$ osiąga maksimum. W podobny sposób, przez odpowiedni wybór początku osi częstotliwości możemy zmniejszyć szerokość pasma $\Delta\nu_2$. Przesuwanie osi częstotliwości możemy zrealizować mnożąc widmo przez współczynnik $e^{2\pi i\nu t}$; operacja taka odpowiada demodulacji sygnału. Jednakże nawet tak zdefiniowane θ_2' i $\Delta\nu_2'$ zawsze spełniają warunek (8.36).

8.5. ZASADA NIEOZNACZONOŚCI DLA CZASU I CZĘSTOTLIWOŚCI

Nasz ogólny warunek (8.30a) jest szczególnym przypadkiem zasady nieoznaczoności znanej w mechanice kwantowej. Zasada ta mówi, że zmienna q oraz sprzężony z nią moment p nie mogą być jednocześnie zmierzone z dowolną dokładnością, lecz że istnieje związek pomiędzy błędem Δq pomiaru q i błędem Δp pomiaru p. Związek ten jest nierównością

$$\Delta q\cdot\Delta p\geqslant\alpha, \tag{8.37}$$

przy czym α w pewnych przypadkach jest równe stałej Plancka h, a w innych — równe stałej Diraca $\hbar = h/2\pi$. Obie te stałe są tak małe, że z praktycznego punktu widzenia nie ma większego znaczenia, którą z nich przyjmiemy.

Jak wiadomo z mechaniki, energia jest „momentem sprzężonym" z czasem t, tak że zgodnie z równaniem (8.37)

$$\Delta E \cdot \Delta t \geqslant \alpha. \tag{8.38}$$

W przypadku drgań o częstotliwości ν energia ($^{8.2}$) $E = h\nu$ i dlatego

$$\Delta E = h\Delta\nu;$$

stąd, jeżeli wykorzystamy empiryczny związek (8.30a), otrzymamy

$$\Delta E \cdot \Delta t = h\Delta\nu \cdot \Delta t \geqslant h. \tag{8.39}$$

Wynika z tego, że powinniśmy przyjąć raczej $\alpha = h$ niż $\alpha = \hbar$. Zagadnienie to rozważymy dalej, przy końcu tego rozdziału.

Obecnie przedstawimy rozumowanie, stosowane w mechanice kwantowej, które prowadzi do zasady nieoznaczoności*. Oblicza się tam funkcję falową $f(t)$, nie przypisując jej jednak żadnego bezpośredniego znaczenia fizycznego, zakłada się natomiast, że kwadrat $|f|^2$ daje gęstość prawdopodobieństwa zmiennej t. Ściślej mówiąc prawdopodobieństwo tego, że zmienna t leży pomiędzy t oraz $t+\mathrm{d}t$, jest definiowane, jak następuje:

$$p(t')\,\mathrm{d}t = \frac{1}{E}\,|f(t')|^2\,\mathrm{d}t, \tag{8.40}$$

gdzie E jest całkowitą energią fali, określoną przez (8.29). Warunek (8.29) daje także

$$\int_{-\infty}^{+\infty} p(t)\,\mathrm{d}t = 1. \tag{8.40a}$$

Podobnie prawdopodobieństwo tego, że częstotliwość ν leży w przedziale ν' oraz $\nu'+\mathrm{d}\nu$, jest określone przez wzory

$$P(\nu')\,\mathrm{d}\nu = \frac{1}{E}\,|C(\nu')|^2\,\mathrm{d}\nu \tag{8.41}$$

oraz

$$\int_{-\infty}^{+\infty} P(\nu)\,\mathrm{d}\nu = 1. \tag{8.41a}$$

* Zob. np. D. Gabor, *M.I.T. Lectures*, 1951.

Zdefiniujmy teraz średni czas trwania sygnału wzorem

$$t_0 = \int_{-\infty}^{\infty} tp(t)\,\mathrm{d}t, \tag{8.42}$$

średnią zaś częstotliwość — wzorem

$$\nu_0 = \int_{-\infty}^{+\infty} \nu\, P(\nu)\,\mathrm{d}\nu. \tag{8.43}$$

Odchylenia t_1 i ν_1 od tych średnich są określone wzorami

$$t = t_0 + t_1, \qquad \nu = \nu_0 + \nu_1.$$

Obliczmy wartości średnie kwadratów tych odchyłek

$$\overline{t_1^2} = \int_{-\infty}^{+\infty} t_1^2\, p(t)\,\mathrm{d}t, \qquad \overline{\nu_1^2} = \int_{-\infty}^{+\infty} \nu_1^2 P(\nu)\,\mathrm{d}\nu. \tag{8.44}$$

Średnie błędy pomiaru czasu Δt i pomiaru częstotliwości $\Delta\nu$ określimy, jak następuje:

$$\Delta t = \sqrt{a\overline{t_1^2}}, \qquad \Delta\nu = \sqrt{a\overline{\nu_1^2}}. \tag{8.45}$$

Stałą a wybierzemy tak, by warunek (8.30a) był spełniony, tzn. chcemy, aby

$$\Delta t \cdot \Delta\nu = \sqrt{a^2 \overline{t_1^2}\,\overline{\nu_1^2}} \geqslant 1 \quad \text{lub} \quad a^2 \overline{t_1^2}\,\overline{\nu_1^2} \geqslant 1. \tag{8.45a}$$

Jeżeli wykorzystamy równania (8.40), (8.41) i (8.44), to zamiast równania (8.45a) będziemy mieli [8.3]

$$a^2 \int_{t=-\infty}^{\infty} t_1^2 f^*(t) f(t)\,\mathrm{d}t \int_{\nu=-\infty}^{\infty} \nu_1^2 C^*(\nu)\, C(\nu)\,\mathrm{d}\nu \geqslant E^2. \tag{8.46}$$

Nie zawężając zagadnienia możemy przyjąć, że $t_0 = 0$ oraz $\nu_0 = 0$; zatem $t_1 = t$ oraz $\nu_1 = \nu$. Przystępując do doboru wartości stałej a, udowodnimy tożsamość

$$\int_{\nu=-\infty}^{\infty} \nu^2\, C^*(\nu)\, C(\nu)\,\mathrm{d}\nu = -\frac{1}{4\pi^2} \int_{t=-\infty}^{\infty} f^*(t)\,\frac{\mathrm{d}^2 f}{\mathrm{d}t^2}\,\mathrm{d}t. \tag{8.47}$$

Dowód tej tożsamości łatwo można przeprowadzić, przy założeniu, że $f(t)$ oraz $C(\nu)$ są różniczkowalne i dostatecznie szybko maleją dla dużych wartości

argumentów. Korzystając z całki Fouriera (8.21) otrzymujemy

$$\frac{\mathrm{d}^2 f}{\mathrm{d}t^2} = \int_{\nu=-\infty}^{\infty} C(\nu)(-4\pi^2\nu^2)\mathrm{e}^{2\pi \mathrm{i}\nu t}\,\mathrm{d}\nu,$$

stąd

$$\int_{t=-\infty}^{\infty} f^* \frac{\mathrm{d}^2 f}{\mathrm{d}t^2}\,\mathrm{d}t = -4\pi^2 \int_{\nu=-\infty}^{\infty} \nu^2\, C(\nu)\,\mathrm{d}\nu \int_{t=-\infty}^{\infty} f^*\, \mathrm{e}^{2\pi \mathrm{i}\nu t}\,\mathrm{d}t.$$

Występująca po prawej stronie druga całka zgodnie z równaniem (8.22) jest niczym innym jak $C^*(\nu)$. W ten sposób dowiedliśmy tożsamości (8.47).

Korzystając z tożsamości (8.47) równanie (8.46) możemy napisać w postaci

$$-\frac{a^2}{4\pi^2} \int_{t=-\infty}^{\infty} t^2 f^* f\,\mathrm{d}t \int_{t=-\infty}^{\infty} f^* \frac{\mathrm{d}^2 f}{\mathrm{d}t^2}\,\mathrm{d}t \geqslant \left| \int_{t=-\infty}^{\infty} f^* f\,\mathrm{d}t \right|^2. \tag{8.48}$$

W dalszym ciągu będziemy korzystać z nierówności Schwarza:*

$$4 \int F^* F\,dt \int G^* G\,\mathrm{d}t \geqslant \left| \int (F^* G + F G^*)\,\mathrm{d}t \right|^2, \tag{8.49}$$

gdzie

$$F = tf \quad \text{oraz} \quad G = \mathrm{d}f/\mathrm{d}t.$$

Drugą całkę obliczamy całkując przez części i korzystając z tego, że funkcja f znika w nieskończoności:

$$\int_{-\infty}^{+\infty} G^* G\,\mathrm{d}t = \int_{-\infty}^{\infty} \frac{\mathrm{d}f^*}{\mathrm{d}f} \frac{\mathrm{d}f}{\mathrm{d}f}\,\mathrm{d}t = f^* \frac{\mathrm{d}f}{\mathrm{d}t}\Bigg|_{-\infty}^{+\infty} - \int_{-\infty}^{\infty} f^* \frac{\mathrm{d}^2 f}{\mathrm{d}t^2}\,\mathrm{d}t =$$

$$= -\int_{-\infty}^{\infty} f^* \frac{\mathrm{d}^2 f}{\mathrm{d}t^2}\,\mathrm{d}t. \tag{8.50}$$

Również prawą stronę nierówności Schwarza całkujemy przez części:

$$\int_{-\infty}^{+\infty} t\left(f^* \frac{\mathrm{d}f}{\mathrm{d}t} + f \frac{\mathrm{d}f^*}{\mathrm{d}t}\right)\mathrm{d}t = (t|f|^2)\Big|_{-\infty}^{+\infty} - \int_{-\infty}^{\infty} |f|^2\,\mathrm{d}t - E.$$

---

* Dowód nierówności Schwarza podano w *Dodatku* przy końcu tego rozdziału.

W ten sposób nierówność Schwarza przyjmuje postać

$$-4\int_{-\infty}^{+\infty} t^2 ff^*\,\mathrm{d}t \int_{-\infty}^{\infty} f^*\frac{\mathrm{d}^2 f}{\mathrm{d}t^2}\,\mathrm{d}t \geqslant E^2.$$

Z porównania tego wyrażenia z (8.48) wynika (8.4), że należy przyjąć

$$a^2 = 16\pi^2 \quad \text{lub} \quad a = 4\pi,$$

tzn. że mamy

$$\Delta t = \sqrt{4\pi\, \overline{t_1^2}}, \quad \Delta\nu = \sqrt{4\,\pi\overline{\nu_1^2}}.$$

Interesujące jest porównanie tych definicji z równaniami (8.31) i (8.32). W poprzednim rozdziale, w przykładach (8.25), (8.27) i (8.28) mieliśmy czas trwania θ_1 i szerokość pasma $1/\theta_1$. Nowe nasze definicje dają wartości następujące:

$$(8.25):\quad E = \theta_1, \qquad \Delta t = \sqrt{\frac{\pi}{3}}\,\theta_1, \quad \Delta\nu = \infty, \qquad \Delta t\,\Delta\nu = \infty,$$

$$(8.27):\quad E = \frac{2}{3}\theta_1, \quad \Delta t = \sqrt{\frac{2\pi}{5}}\,\theta_1, \quad \Delta\nu = \frac{1}{\theta_1}\sqrt{\frac{3}{\pi}}, \quad \Delta t\,\Delta\nu = \sqrt{\frac{6}{5}},$$

$$(8.28):\quad E = \frac{\theta_1}{\sqrt{2}}, \quad \Delta t = \sqrt{2}\,\theta_1, \qquad \Delta\nu = \frac{1}{\theta_1}\sqrt{2}, \qquad \Delta t\,\Delta\nu = 2.$$

Obliczenie E, $\Delta\nu$ oraz Δt w tych przykładach pozostawiamy czytelnikowi. Podamy tylko niektóre z całek oznaczonych, które występują w obliczeniach:

$$\int_{-\infty}^{\infty}\left(\frac{\sin x}{x}\right)^2 \mathrm{d}x = \pi, \qquad \int_{-\infty}^{\infty}\frac{\sin^4 x}{x^2}\,\mathrm{d}x = \frac{\pi}{2}, \qquad \int_{-\infty}^{\infty} x^2\,\mathrm{e}^{-x^2}\,\mathrm{d}x = \sqrt{\pi}.$$

W wielu pracach z mechaniki kwantowej przyjmuje się $a = 2$. W takim przypadku

$$\Delta_1 t = \sqrt{2\overline{t_1^2}} = \frac{\Delta t}{\sqrt{2\pi}}, \qquad \Delta_1\nu = \frac{\Delta\nu}{\sqrt{2\pi}}.$$

Wynika stąd

$$\Delta_1 t \cdot \Delta_1\nu \geqslant \frac{1}{2\pi}.$$

Jeśli podobnie określimy $\Delta_1 q$ i $\Delta_1 p$, relacja nieoznaczoności przyjmie postać

$$\Delta_1 q \cdot \Delta_1 p \geqslant \hbar.$$

Definicje te z punktu widzenia naszych poprzednich rozważań wydają się nierealne. Zgodnie z nimi kres dolny $\Delta\nu \cdot \Delta t$ ma być 2π razy mniejszy, co we wspomnianym w poprzednim rozdziale przypadku kanału telewizyjnego oznaczałoby, że wystarczające jest pasmo o szerokości około 1 megaherca. Rzecz prosta ([8.5]), że gdyby to było rzeczywiście możliwe, wprowadzono by w telewizji, zamiast jednego kanału o szerokości 6 megaherców, sześć kanałów o szerokości jednego megaherca. Znaczenie fizyczne wartości Δt i $\Delta\nu$ łatwo można odczytać z rys. 8.5.

Czytelnik na pierwszy rzut oka oceni więc, jak nierealistyczne byłoby przyjęcie wartości mniejszych od wprowadzonych przez nas.

8.6. STOPNIE SWOBODY WIADOMOŚCI

Rozważmy obecnie następujące zagadnienie. Mamy funkcję czasu* $f(t)$, określoną w przedziale czasu τ, której widmo nie zawiera częstotliwości wyższych od wartości ν_M. Ile parametrów (lub stopni swobody) potrzeba do określenia takiej funkcji? Wykażemy, że wystarcza tylko

$$N = 2\nu_M \tau \tag{8.51}$$

(przy założeniu, że $\nu_M \tau \gg 1$) niezależnych parametrów dla określenia takiej funkcji. Rozpatrzymy także niektóre ogólne własności funkcji $f(t)$ oraz różne możliwości wyboru jej parametrów.

Przede wszystkim musimy zaznaczyć, że funkcja, której wartości znamy jedynie dla skończonego przedziału, np. przedziału $0 < t < \tau$, nie jest w pełni określona. Jeżeli chcemy poprawnie ustawić zagadnienie, musimy uwzględnić przebieg funkcji przed zerem i po τ. Należy to zrobić nie dodając niczego do informacji zawartej w wiadomości ([8.6]) $f(t)$. Są tu możliwe dwa założenia:

(1) *Funkcja $f(t)$ jest okresowa* — przebieg f z przedziału od 0 do τ po-

* Funkcję tę autor nazywa wiadomością. (*Przyp. tłum.*).

wtarza się nieskończoną ilość razy, a więc

$$f(t+q\tau) = f(t), \quad q \text{ jest liczbą całkowitą.} \tag{8.52}$$

(2) *Przebieg funkcji powtarza się jednorazowo* w przedziale $0 < t < \tau$:

$$f(t) = 0 \quad \text{dla} \quad t < 0 \quad \text{lub} \quad t > \tau. \tag{8.53}$$

Ten ostatni przypadek był rozpatrywany przez Shannona.

Zacznijmy od założenia (1) i rozważmy funkcję okresową o okresie τ. Rozwińmy tę funkcję w szereg Fouriera:

$$f(t) = \frac{1}{2} a_0 + \sum_{n=1}^{n_M} (a_n \cos n\varphi + b_n \sin n\varphi), \tag{8.54}$$

gdzie

$$\varphi = \frac{2\pi t}{\tau} = 2\pi \nu_0 t, \qquad \nu_0 = \frac{1}{\tau}. \tag{8.55}$$

Przyjmijmy, że częstotliwość maksymalna ν_M równa się dokładnie jednej z harmonicznych częstotliwości ν_0:

$$n_M = \nu_M \tau \quad \text{lub} \quad n_M \nu_0 = \nu_M. \tag{8.56}$$

Szereg Fouriera zawiera w takim przypadku skończoną liczbę wyrazów — do wyrazu z numerem n_M. Każdej harmonicznej odpowiadają dwa parametry a_n oraz b_n, stąd całkowita liczba parametrów licząc stałą a_0 wynosi

$$N = 2n_M + 1 = 2\nu_M \tau + 1. \tag{8.57}$$

Jeśli czas trwania sygnału τ jest dostatecznie duży, wyrażenie (8.57) przechodzi w (8.51). Współczynniki a_n i b_n mogą stanowić jeden z możliwych zespołów parametrów określających funkcję. Zamiast szeregu Fouriera o wyrazach rzeczywistych możemy przyjąć zespolony szereg Fouriera — wyrażenia (8.1) i (8.2):

$$\left.\begin{aligned} f(t) &= \sum_{n=-n_M}^{+n_M} c_n \, \mathrm{e}^{\mathrm{i}n\omega_0 t}, \quad \omega_0 = 2\pi\nu_0 = \frac{2\pi}{\tau}, \\ c_n &= \frac{1}{2}(a_n - \mathrm{i}\, b_n) = \frac{1}{\tau}\int_0^{\tau} f(t')\mathrm{e}^{-\mathrm{i}n\omega_0 t'}\, \mathrm{d}t' = c_{-n}^{*}, \end{aligned}\right\} \tag{8.58}$$

gdzie gwiazdka oznacza liczbę zespoloną sprzężoną.

Zamiast szeregu Fouriera możemy, do wyboru zespołu parametrów określających funkcję okresową $f(t)$, zastosować metodę próbkowania. Polega ona na tym, że w przedziale τ wybieramy N równoodległych chwil próbkowania, np.

$$t_m = m\theta, \qquad \theta = \frac{\tau}{N} = \frac{1}{2\nu_M + \nu_0}, \tag{8.59}$$

gdzie

$$N = 2n_M + 1 = \tau(2\nu_M + \nu_0) = 2\nu_M\tau + 1;$$

f_m nazwiemy wartością próbki:

$$f_m = f(m\theta), \tag{8.59a}$$

$$f_{m+qN} = f_m, \qquad q \text{ jest liczbą całkowitą,}$$

zgodnie z warunkiem (8.52).

Naszą wyjściową funkcję $f(t)$ możemy odtworzyć ([8.7]), jeśli znamy jej $2n_M+1$ wartości próbek w przedziale τ. Zapiszmy

$$f(t) = \sum_{m=1}^{N} f_m g(t - m\theta), \tag{8.60}$$

gdzie $g(t)$ jest pewną funkcją impulsową powtarzającą się okresowo w chwilach czasu $t = q\tau$. Przyjmijmy, że funkcja $g(t)$ ma następujące własności:

$g(t) = 1$ dla $t = 0$ lub $t = q\tau$, gdzie q — liczba całkowita,
$g(t) = 0$ dla $t = m\theta$ lub $m \neq qN$.

Nasza funkcja impulsowa przyjmuje wartość zero dla wszystkich innych chwil próbkowania w przedziale τ. Nietrudno jest skonstruować funkcję o takich własnościach przy warunku, że jej widmo jest skończone i nie przekracza wartości ν_M. Wykorzystamy tutaj tożsamość Lagrange'a

$$\frac{1}{2} + \sum_{n=1}^{n_M} \cos \varphi n = \frac{1}{2} \sum_{n=-n_M}^{+n_M} e^{in\varphi} = \frac{\sin\left(n_M + \frac{1}{2}\right)\varphi}{2 \sin(\varphi/2)}, \tag{8.61}$$

gdzie $\varphi = \omega_0 t$.

Funkcja występująca po prawej stronie (8.61) dla $t = 0$ lub $t = q\tau$ jest

równa $n_M + {}^1/_2$, w tych bowiem punktach mianownik przyjmuje wartość zero. Taka funkcja oscyluje i przyjmuje wartości zerowe w punktach

$$\left(n_M + \frac{1}{2}\right)\omega_0 t = \pi, 2\pi, \ldots, m\pi,$$

$$t = \frac{\tau}{N}, \quad \frac{2\tau}{N}, \quad \ldots, \quad \frac{m\tau}{N}, \quad m \neq qN,$$

gdzie m nie może być całkowitą wielokrotnością N.

Możemy założyć, że nasza funkcja impulsowa $g(t)$ ma postać

$$g(t) = \frac{\sin\left(N\frac{\omega_0 t}{2}\right)}{N\sin\left(\frac{\omega_0 t}{2}\right)} = \frac{1}{N}\sum_{-n_M}^{+n_M} e^{in\omega_0 t}. \tag{8.62}$$

Porównując równania (8.60), (8.62) i (8.58) otrzymujemy

$$f(t) = \frac{1}{N}\sum_{-n_M}^{+n_M}\sum_{m=1}^{N} f_m\, e^{in\omega_0(t-m\theta)}.$$

Stąd mamy

$$c_n = \frac{1}{N}\sum_{m=1}^{N} f_m\, e^{-inm\omega_0\theta}. \tag{8.63}$$

Wyrażenie to wiąże bezpośrednio współczynniki Fouriera* c_n z wartościami próbek f_m. Relację odwrotną otrzymujemy ze wzoru (8.58):

$$f_m = f(m\theta) = \sum_{n'=-n_M}^{+n_M} c_{n'} e^{-in'm\omega_0\theta}. \tag{8.63a}$$

W zespolonym szeregu Fouriera mamy $N = 2n_M + 1$ zespolonych amplitud, które dla $\pm n$ są liczbami zespolonymi sprzężonymi. Daje to N niezależnych zmiennych rzeczywistych; wynika stąd, że naszych N chwil próbkowania

* Równanie (8.63) jest identyczne z (8.5) i (8.11), ponieważ zajmujemy się tutaj funkcjami mającymi skończoną szerokość pasma, jak w równaniu (8.9).

zapewnia taką samą ilość stopni swobody. Wzory (8.63) i (8.63a) można łatwo sprawdzić:

$$\left.\begin{aligned} c_n &= \frac{1}{N}\sum_{m=1}^{N}\sum_{n'=-n_M}^{n_M} c_{n'}\, e^{i(n'-n)m\omega_0\theta}; \\ &\text{jednakże} \\ \frac{1}{N}\sum_{m=1}^{N} e^{ipm\omega_0\theta} &= \begin{cases} 1 \text{ dla } p = 0, \\ 0 \text{ dla } p \neq 0 \text{ (całkowitego)}, \end{cases} \end{aligned}\right\} \quad (8.63b)$$

ponieważ dla $p = 0$ mamy w sumie N wyrazów równych jedności, podczas gdy dla $p \neq 0$ fazy N składowych harmonicznych są równomiernie rozłożone w przedziale od 0 do 2π, co w rezultacie daje sumę równą zeru. Mamy więc

$$c_n = \sum c_{n'}\,\delta(n-n'), \qquad \text{c.b.d.o.}$$

Zamiast próbkowania w chwilach czasu $m\theta$ można przyjąć chwile próbkowania $m\theta+\theta'$, gdzie θ' jest ustalonym odcinkiem czasu i $\theta' < \theta$. Da to inny zbiór wartości próbek $f'_m = f(m\theta+\theta')$, którym możemy zastąpić zbiór wartości f_m. W takim przypadku wzory (8.63) i (8.63a) należy zastąpić wzorami

$$c_n = \frac{1}{N}\sum_{m} f'_m\, e^{-in\omega_0(m\theta-\theta')}, \qquad f'_m = \sum_{n'} c_{n'}\, e^{in'\omega_0(m\theta+\theta')}.$$

Wartości f_m i f'_m są powiązane ze sobą zespołem N związków liniowych.

8.7. METODA PRÓBKOWANIA SHANNONA

Shannon wykorzystuje metodę próbkowania w związku z zagadnieniem (wzór (8.53)) funkcji występującej jednorazowo, o której zakładamy, że przyjmuje wartość zero dla $t < 0$ lub $t > \tau$. Zagadnienie to możemy rozwiązać w sposób następujący. Weźmy funkcję okresową o długim okresie $\tau_1 > \tau$, załóżmy, że przyjmuje ona wartość naszej funkcji $f(t)$ w przedziale od 0 do τ, natomiast w przedziale od τ do τ_1 równa się zero. Aby nie wprowadzać częstotliwości wyższych od ν_M, załóżmy przy tym, że przejście $f(t)$ w zero na obu końcach 0 i τ jest płynne ([8.8]). Tę nową funkcję bę-

dziemy próbkować na odcinku τ_1 w N_1 równooddalonych chwilach próbkowania, zgodnie z równaniami (8.59) i (8.62). Nowy okres próbkowania θ_1 zgodnie z równaniem (8.59) wynosi

$$\theta_1 = \frac{1}{2\nu_M + \nu_1}, \quad \nu_1 = \frac{1}{\tau_1}, \quad \omega_1 = \frac{2\pi}{\tau_1}. \tag{8.64}$$

Jeśli τ i τ_1 są duże, to ν_0 i ν_1 są do pominięcia w porównaniu z częstotliwością maksymalną i mamy

$$\theta_0 \approx \theta_1 \approx \frac{1}{2\nu_M}.$$

Całkowita liczba chwil próbkowania wynosi teraz

$$N_1 = 2\nu_M \tau_1 + 1. \tag{8.65}$$

Spośród tych N_1 chwil N leży w przedziale $0 < t < \tau$, a $N_1 - N$ w przedziale $\tau < t < \tau_1$. Pierwszy zbiór chwil próbkowania daje niezerowe wartości próbek:

$$f_m = f(m\theta_1) \neq 0, \quad 0 < m < N, \tag{8.66}$$

podczas gdy dla drugiego zbioru wszystkie próbki mają wartość zero:

$$f_m = f(m\theta_1) = 0, \quad N < m < N_1.$$

Nasza funkcja impulsowa $g(t)$ (równanie (8.62)) przyjmuje postać

$$g(t) = \frac{\sin\left(N_1 \frac{\omega_1 t}{2}\right)}{N_1 \sin\left(\frac{\omega_1 t}{2}\right)}. \tag{8.67}$$

Jeżeli założymy teraz, że τ_1 rośnie nieograniczenie, ν_1 zaś maleje do zera, to w granicy otrzymamy okres próbkowania

$$\theta_{\text{lim}} = \frac{1}{2\nu_M}$$

i liczba próbek f_m równych zero będzie nieskończona. Próbki o wartościach niezerowych odpowiadają chwilom próbkowania z przedziału

$$0 < m\,\theta_{\text{lim}} < \tau.$$

Liczba ich jest równa

$$N_{\text{lim}} = \tau/\theta_{\text{lim}} = 2\nu_M\tau. \tag{8.68}$$

Otrzymany przez nas wynik jest zgodny z równaniem (8.51), natomiast funkcja impulsowa $g(t)$ upraszcza się do postaci

$$g(t) \approx \frac{\sin\left(N_1 \frac{\omega_1 t}{2}\right)}{N_1 \frac{\omega_1 t}{2}} \approx \frac{\sin 2\pi\nu_M t}{2\pi\nu_M t}, \tag{8.69}$$

ponieważ

$$N_1\omega_1 = (2\nu_M\tau_1+1)\frac{2\pi}{\tau_1} = 4\pi\nu_M + \frac{2\pi}{\tau_1} \approx 4\pi\nu_M.$$

W granicy funkcję $f(t)$ można przedstawić równaniem (8.60), które przyjmuje postać

$$f(t) = \sum_m f_m \frac{\sin \pi(2\nu_M t - m)}{\pi(2\nu_M t - m)}. \tag{8.70}$$

Można łatwo wykazać, że rozwinięcie równania (8.70) przyjmuje wartość f_m dla naszych wszystkich chwil próbkowania. Rozważmy np. chwilę q:

$$t_q = \frac{q}{2\nu_M}.$$

Wyraz $m = q$ w sumie (8.70) daje składową f_q, natomiast wszystkie inne wyrazy są równe zeru:

$$m \neq q: \quad \sin \pi (q-m) = 0.$$

Suma (8.70) nie daje dokładnie wartości $f(t) = 0$ dla $t < 0$ oraz $t > \tau$, lecz stanowi funkcję obustronnie szybko malejącą z małymi oscylacjami o częstotliwości ν_M. Podany wyżej sposób przedstawiania funkcji $f(t)$ stosował Shannon.

Jak wykazaliśmy wyżej, występująca jednorazowo funkcja $f(t)$, określona zgodnie z równaniem (8.53), ma tylko N stopni swobody. Jeżeli do analizy takiej funkcji zastosujemy rozwinięcie Fouriera, to zamiast szeregu Fouriera otrzymamy całkę Fouriera. Liczba wyrazów w rozwinięciu Fouriera jest

nieskończona, lecz jest wśród nich tylko N niezależnych zmiennych f_m. Wniosek ten można łatwo sprawdzić, ponieważ nasza funkcja $g(t)$ jest identyczna z impulsem z przykładu (8.26) i ma następujące widmo:

$$B(\nu) = \left\{ \begin{array}{ll} \dfrac{1}{2\nu_M} & \text{dla} \quad |\nu| < \nu_M, \\ 0 & \text{dla} \quad |\nu| > \nu_M; \end{array} \right\} \tag{8.70a}$$

stąd

$$g(t) = \frac{1}{2\nu_M} \int_{\nu=-\nu_M}^{+\nu_M} \cos 2\pi\nu t \, \mathrm{d}\nu = \frac{1}{2\nu_M} \int_{\nu=-\nu_M}^{+\nu_M} \mathrm{e}^{\mathrm{i}2\pi\nu t} \, \mathrm{d}\nu \tag{8.71}$$

oraz

$$f(t) = \sum_m \frac{f_m}{2\nu_M} \int_{-\nu_M}^{+\nu_M} \mathrm{e}^{\mathrm{i}2\pi\nu(t-m\theta)} \, \mathrm{d}\nu, \tag{8.72}$$

gdzie

$$\theta = \frac{1}{2\nu_M}.$$

Ogólna zasada rozważanego tu zagadnienia została sformułowana niezależnie przez wielu uczonych. Wydaje się jednak, że pierwsze interesujące wyniki otrzymał Whittaker* w r. 1915.

Dalsze szczegóły dotyczące omówionej wyżej metody oraz możliwości jej zastosowania znajdzie czytelnik w rozdz. 17.6 i 17.7.

8.8. KOMÓRKI INFORMACYJNE GABORA

Oba rozważane sposoby przedstawiania funkcji są przypadkami szczególnymi podanej przez D. Gabora** bardziej ogólnej metody. Obszar zajmowany przez funkcję na płaszczyźnie $(^{8.9})$ (ν, t) jest prostokątem $\nu_m \cdot \tau$ (rys. 8.7).

---

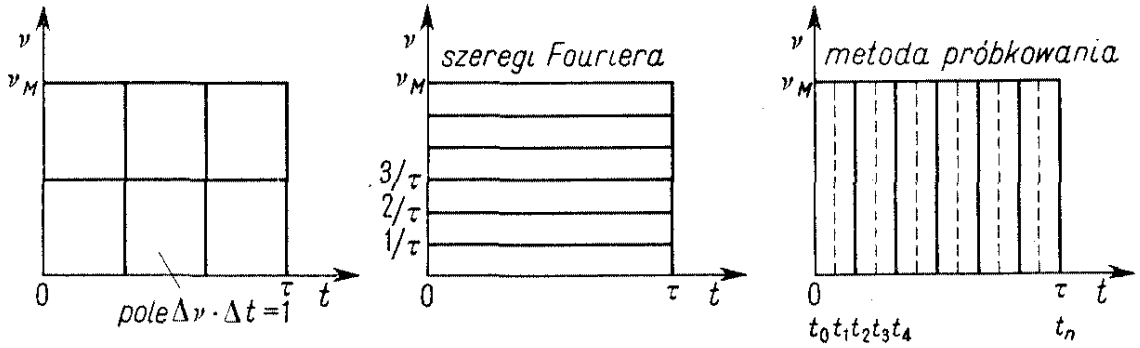
* E. T. Whittaker, Proc. Roy. Soc. Edinburgh **35** (1915), str. 181.

** D. Gabor, *Theory of Communication*, J. Inst. Elec. Engrs. **93**, III (1946), str. 429; **94**, III (1947), str. 369; także Phil. Mag. [7] **41** (1950), str. 1161; *La Cybernetique*, Paris 1951, Editions Revue d'optique, str. 115.

Prostokąt ten można podzielić na elementarne komórki o polu jednostkowym

$$\Delta\nu \cdot \Delta t = 1. \tag{8.73}$$

Każdej komórce odpowiadają dwie składowe funkcji — symetryczna i asymetryczna, a mianowicie dwie elementarne funkcje o czasie trwania Δt, zajmujące pasmo szerokości $\Delta\nu$. Amplitudy obu tych funkcji muszą być



Rys. 8.7. Różne sposoby wyboru komórek o jednostkowym polu na płaszczyźnie (ν,t)

znane; określają one jednoznacznie naszą funkcję $f(t)$. Liczba komórek wynosi

$$n = \nu_M \tau,$$

natomiast liczba elementarnych funkcji

$$N = 2n = 2\nu_M \tau, \tag{8.74}$$

podobnie jak w równaniu (8.51). Kształt komórek jest bez znaczenia, ważne jest tylko ich pole. Przy rozwijaniu na szereg Fouriera (równanie (8.58)) wykorzystuje się rozkład na komórki poziome, w metodzie próbkowania zaś (równanie (8.61)) — na komórki pionowe. Jako funkcje elementarne Gabor zaleca przyjmować funkcje sinus i cosinus z obwiednią Gaussa, jak to pokazano na rys. 8.8.

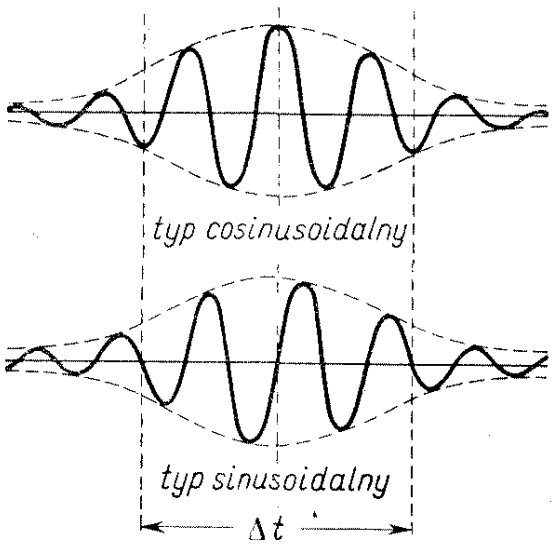
Komórki (8.73) odpowiadają komórkom występującym w mechanice kwantowej, gdzie energia E i czas t są dwiema zmiennymi sprzężonymi, sama komórka zaś jest określona przez warunek

$$\Delta E \cdot \Delta t = h, \tag{8.75}$$

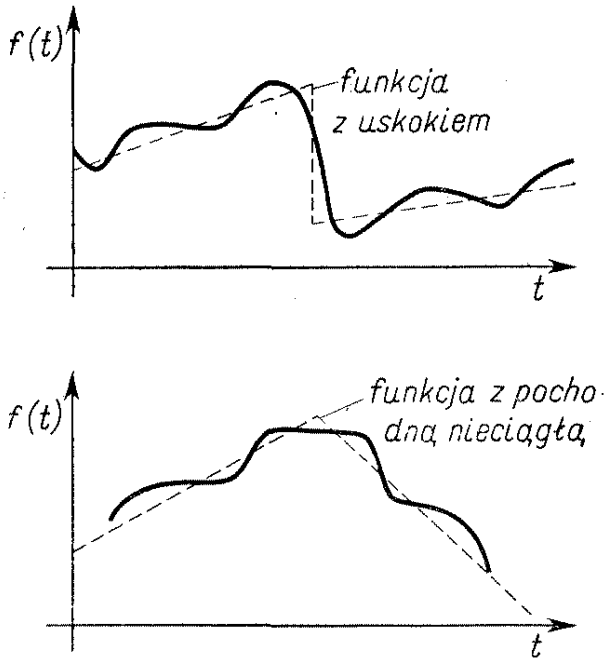
gdzie

$$E = h\nu, \qquad \Delta\nu\, \Delta t = 1.$$

Funkcje o skończonej szerokości pasma są funkcjami szczególnego rodzaju; nie mają one żadnych nieciągłości, żadnych ostrych kątów; przebieg ich jest łagodny, jak to pokazano na rys. 8.9.



Rys. 8.8. Funkcje wiadomości Gabora: sinusoidalna i cosinusoidalna



Rys. 8.9. Funkcje z ograniczonym pasmem częstotliwości

8.9. AUTOKORELACJA I WIDMO; TWIERDZENIE WIENERA—CHINCZYNA

Rozważmy ponownie zagadnienie (1) z rozdz. 8.6, gdzie rozpatrywaliśmy funkcję okresową $f(t)$ o okresie τ (równanie (8.52)), której widmo nie zawiera częstotliwości wyższych od ν_M. Rozwińmy funkcję $f(t)$ na zespolony szereg Fouriera (8.1) i (8.58):

$$\left.\begin{aligned} f(t) &= \sum_{-m_M}^{+n_M} C_n \, e^{in\omega_0 t}, \qquad \omega_0 = \frac{2\pi}{\tau} = 2\pi\nu_0, \\ C_{-n} &= C_n^*, \qquad\qquad n_M = \frac{\nu_M}{\nu_0} = \nu_M \tau. \end{aligned}\right\} \tag{8.76}$$

Zdefiniujmy dla tej funkcji funkcję autokorelacji ([8.10]), $R(t_0)$, wzorem

$$R(t_0) = \overline{f(t)f(t+t_0)} = \frac{1}{\tau}\int_0^{\tau} f(t)\,f(t+t_0)\,dt. \tag{8.77}$$

Podstawmy zamiast funkcji $f(t+t_0)$ jej rozwinięcie Fouriera (8.76). Mamy w ten sposób rozwinięcie Fouriera $R(t_0)$:

$$R(t_0) = \frac{1}{\tau} \sum_{-n_M}^{n_M} C_n \, e^{in\omega_0 t_0} \int_0^{\tau} f(t) \, e^{in\omega_0 t} \, dt = \sum_{-n_M}^{+n_M} C_n C_n^* \, e^{in\omega_0 t_0}, \qquad (8.78)$$

przy czym w ostatnim przekształceniu korzystamy ze wzorów (8.2) i (8.3).

Zauważmy dalej, że suma ze wzoru (8.78) zawiera jedynie kwadraty modułów współczynników Fouriera funkcji $f(t)$:

$$I_n = C_n C_n^* = |C_n|^2 = |C_{-n}|^2 = I_{-n}. \qquad (8.79)$$

Zależność od faz znikła w procesie uśredniania (8.77) i (8.78). Wykorzystując symetrię (8.79), wyrażenie (8.78) możemy przepisać, jak następuje:

$$R(t_0) = I_0 + 2 \sum_{n=1}^{n_M} I_n \cos n\omega_0 t_0 = R(-t_0). \qquad (8.80)$$

Funkcja autokorelacji jest więc parzystą funkcją opóźnienia t_0. Jest ona także funkcją okresową, przy czym okres jej jest równy okresowi τ funkcji wyjściowej. Taka sama jest także górna częstotliwość $|\nu| \leqslant \nu_M$. Równanie (8.80) stanowi szczególny przypadek twierdzenia Wienera–Chinczyna*, wiążącego funkcję autokorelacji funkcji $f(t)$ z jej widmem Fouriera.

Jak pokazaliśmy w rozdz. 8.6, nasza funkcja wyjściowa $f(t)$ ma skończoną liczbę N stopni swobody (zob. równanie (8.57)):

$$N = 2n_M + 1,$$

wśród których możemy rozróżnić

1	składową stałą	$C_0 = \gamma_0$,
n_M	amplitud składowych harmonicznych	γ_n,
n_M	faz składowych harmonicznych	φ_n;

$$C_n = \gamma_n e^{i\varphi n}, \qquad C_{-n} = \gamma_n e^{-i\varphi n}, \qquad (8.81)$$

stąd

$$I_n = \gamma_n^2, \qquad n = 0,1, \ldots, n_M. \qquad (8.82)$$

* N. Wiener, Acta Math. **55** (1930), str. 117; A. Chinczyn, Math. Ann. **109** (1934), str. 604; Y. W. Lee, P. Cheatham, J. B. Wiesner, M.I.T. Report No. 141, 1949.

Rozwinięcie (8.80) ma tylko n_M+1 stopni swobody, ponieważ pozbyliśmy się zależności od n_M faz.

Zamiast rozwinięć Fouriera (8.78) i (8.80) przy analizie funkcji autokorelacji możemy wykorzystać metodę próbkowania (wzory (8.59) i (8.63)). Wybierzmy w tym celu N równoodległych chwil próbkowania w przedziale τ rozciągającym się od $-\tau/2$ do $+\tau/2$:

$$t_{0m} = m\theta, \quad \theta = \frac{\tau}{N}, \quad N = 2n_M+1, \tag{8.83}$$

$$m = -n_M, -n_M+1, \ldots, -1, 0, 1, \ldots, n_M-1, n_M$$

i rozważmy wartości próbek

$$R_m = R(m\theta) = R_{-m}. \tag{8.84}$$

Z warunku symetrii (8.80) wynika, że istnieje tylko n_M+1 niezależnych wartości próbek odpowiadających

$$m = 0, 1, \ldots, n_M. \tag{8.85}$$

Wynik ten pokrywa się z naszymi poprzednimi rozważaniami, gdzie wykazaliśmy, że liczba stopni swobody jest równa n_M+1. Funkcję $R(t_0)$ możemy odtworzyć wykorzystując wartości próbek podobnie jak w równaniu (8.60):

$$R(t_0) = \sum_{m=1}^{n_M} R_m[g(t_0-m\theta)+g(t_0+m\theta)]+R_0\, g(t_0), \tag{8.86}$$

gdzie funkcja g jest zdefiniowana przez równanie (8.62).

Przy takim sposobie postępowania wykorzystujemy tylko n_M+1 stopni swobody funkcji wyjściowej (z ogólnej liczby $2n_M+1$). Sposób ten można z powodzeniem stosować we wszystkich przypadkach, gdzie fazy φ_n nie są znane albo z uwagi na ich losowy charakter (zagadnienia szumu), albo z uwagi na to, że metoda pomiaru nie pozwala zmierzyć kątów fazowych (np. badania budowy kryształów za pomocą promieni X).

Zamiast zespolonego szeregu Fouriera (8.76) funkcji $f(t)$ możemy przyjąć postać rzeczywistą (8.54):

$$\left.\begin{aligned} f(t) &= \frac{1}{2}A_0 + \sum_{n=1}^{n_M} A_n \cos(n\omega_0 t+\varphi_n), \\ A_0 &= 2\gamma_0, \quad A_n = 2\gamma_n, \quad I_0 = \frac{A_0^2}{4}, \quad I_n = \frac{A_n^2}{4} \end{aligned}\right\} \tag{8.87}$$

i zgodnie z (8.80) otrzymamy

$$R(t_0) = {}^1/_4 A_0^2 + {}^1/_2 \sum_{n=1}^{n_M} A_n^2 \cos(n\omega_0 t_0). \tag{8.88}$$

W przypadku szumu białego (8.13) współczynniki A_n^2 są stałe, niezależne (8.13) od n.

8.10. PRZEKSZTAŁCENIA LINIOWE I FILTRY

Bardzo skuteczna metoda badania funkcji jest oparta na przekształceniach liniowych*. Rozważmy funkcję $f(t)$ i jej całkę Fouriera

$$f(t) = \int c(\nu)\, e^{2\pi i \nu t}\, d\nu \tag{8.89}$$

oraz funkcję $\varphi(\theta)$ malejącą dostatecznie szybko i jej całkę Fouriera

$$\varphi(\theta) = \int \gamma(\nu)\, e^{2\pi i \nu \theta}\, d\nu. \tag{8.90}$$

Funkcję $\varphi(\theta)$ określa następujące przekształcenie:

$$F(t) = \int f(t+\theta)\, \varphi(\theta)\, d\theta. \tag{8.91}$$

Funkcję $F(t)$ możemy także przedstawić w postaci całki Fouriera:

$$F(t) = \int C(\nu)\, e^{i2\pi t \nu}\, d\nu. \tag{8.92}$$

Wykorzystując (8.91) i podstawiając $t' = t+\theta$ mamy

$$C(\nu) = \int F\, e^{-i2\pi\nu t}\, dt = \int f(t')\, e^{-i2\pi\nu t'}\, dt' \int \varphi(\theta)\, e^{+i2\pi\nu\theta}\, d\theta.$$

---

* H. Labrouste, Y. Labrouste, Compt. rend. **184** (1927), str. 259; Ann. Inst. Phys. Globe Univ. Paris **7** (1929), str. 190; **9** (1931), str. 99; **11** (1933), str. 93; Terrestrial Magnetism and Atm. Elec. **41**, No. 15 (1936), str. 105;

H. Deve, *Tables numeriques*, Paris 1930, Evreux;

M. Levy, Compt. rend. **198** (1934), str. 2222; **199** (1934), str. 1031; **200** (1935), str. 646; *Ph. D. thesis*, Paris 1951.

H. Labrouste rozważał dokładniej zagadnienie obliczeń numerycznych, gdzie korzysta się z dyskretnych wartości funkcji. Przyjmował on ogólne przekształcenie typu

$$F(t) = \sum_n \varphi_n f(t+n\tau),$$

gdzie n jest całkowite. M. Levy uogólnił wyniki na przypadek przekształceń całkowych typu (8.91).

Ponieważ całki po prawej stronie są równe $C(\nu)$ oraz $\gamma^*(\nu)$, otrzymujemy wyrażenie

$$C(\nu) = c(\nu)\,\gamma^*(\nu). \tag{8.93}$$

Wynika stąd, że przekształcenie (8.91) jest równoważne filtracji ([8.14]) przy pomocy filtru o funkcji przenoszenia $\gamma(\nu)$. Labrouste i Levy rozważali różne funkcje określające przekształcenie (8.91) i pozwalające na wyizolowanie składowych sinusoidalnych ze złożonych krzywych eksperymentalnych. Na przykład impulsy (8.25) do (8.26) dają funkcje przenoszenia odpowiadające ich widmom i mogą być użyte jako funkcje określające przekształcenia.

Funkcja $F(t)$ jest funkcją korelacji skrośnej funkcji f i φ. Jeżeli $\varphi(t)$ pokrywa się z $f(t)$, to otrzymujemy funkcję autokorelacji funkcji $f(t)$ i równania (8.91) i (8.93) stają się podobne do (8.77) i (8.79).

Aby pokazać związek pomiędzy wynikami otrzymanymi w tym rozdziale, rozpatrzmy przykład. Przypuśćmy, że chcemy zbadać eksperymentalną krzywą $f(t)$, o której wiemy, że częstotliwości powyżej ν_M nie są zapisane dokładnie. Przede wszystkim chcemy więc odrzucić te częstotliwości. Możemy to zrobić stosując przekształcenie (8.91) o funkcji przekształcenia (8.26); w tym celu należy obliczyć

$$F(t) = 2\nu_M \int f(t+\theta)\,\frac{\sin(2\pi\nu_M\theta)}{2\pi\nu_M\theta}\,\mathrm{d}\theta. \tag{8.94}$$

Funkcja $\varphi(\theta)$ ma częstotliwość odcięcia ν_M, widmo jej ma postać

$$\gamma(\nu) = \begin{cases} 1, & \text{gdy} \quad |\nu| \leqslant \nu_M, \\ 0, & \text{gdy} \quad |\nu| > \nu_M. \end{cases}$$

Przekształcenie to odpowiada idealnemu filtrowi dolnoprzepustowemu — tłumi ono wszystkie wysokie częstotliwości, tzn. wygładza krzywą i eliminuje ostre załamania i nieciągłości, o których przypuszczamy, że nie występują w rzeczywistości. Ponieważ nowa funkcja $F(t)$ posiada wyraźnie określoną częstotliwość graniczną ν_M, przeto jeżeli obserwujemy ją w odcinku czasu τ, to zawiera ona tylko

$$N = 2\nu_M\tau$$

stopni swobody (por. (8.51)). Funkcję tę będziemy traktować jako funkcję występującą jednorazowo (rozdz. 8.6, punkt (2)) i zastosujemy metodę próbko-

wania Shannona (rozdz. 8.7) z okresem próbkowania (8.68):

$$\theta_{\text{lim}} = \frac{1}{2\nu_M}.$$

W wyniku próbkowania funkcję ciągłą zastępujemy ciągiem impulsów o wysokościach $F_M = F(m\theta_{\text{lim}})$. Jak pokazaliśmy w rozdz. 8.1 (równania (8.9.), (8.11) i rys. 8.1), taki ciąg impulsów ma nieskończenie szerokie widmo. Składa się ono z widma pierwotnego ($|\nu| \leqslant \nu_M$), powtarzającego się okresowo wzdłuż osi ν z okresem $2\nu_M$. Impulsy te mogą być przekazywane w systemie telekomunikacyjnym za pomocą jednej z metod współczesnej techniki impulsowej. Wartości próbek są odtwarzane w stacji odbiorczej i zagadnienie sprowadza się do rekonstrukcji krzywej ciągłej $F(t)$ na podstawie próbek. Można zrealizować to za pomocą układu elektrycznego, a mianowicie za pomocą filtru pasmowego wycinającego jedno z identycznych pasm częstotliwości z rys. 8.1. Operację tę można także zrealizować w sposób rachunkowy obliczając sumę (8.70):

$$F(t) = \sum_m F_m \frac{\sin \pi(2\nu_M t - m)}{\pi(2\nu_M t - m)}, \tag{8.94a}$$

co znów odpowiada matematycznemu opisowi idealnej filtracji dolnoprzepustowej. Ściśle rzecz biorąc, aby obliczyć wartości funkcji dla ustalonej chwili t, musimy znać wartość wszystkich próbek przed i po t. Jeśli chcemy odtworzyć $F(t)$ z dokładnością do 10^{-4}, musimy uwzględnić co najmniej 10^4 próbek branych przed i po chwili t. Jest to konieczne z uwagi na fakt, że wyrazy sumy maleją bardzo powoli — tylko jak m^{-1} lub $(t'-t)^{-1}$. Wydaje się jednak, że w przypadkach, kiedy wymagana jest duża dokładność, obliczenie sumy (8.94a) jest najlepszą metodą odtwarzania $F(t)$.

8.11. TRÓJWYMIAROWA ANALIZA FOURIERA I METODY PRÓBKOWANIA W TRZECH WYMIARACH

Metody rozważane w poprzednich rozdziałach mogą być zastosowane do analizy zagadnień, w których występują dwa lub trzy wymiary. Rozpatrzymy teraz pokrótce tego typu zagadnienia i wskażemy na ich związek

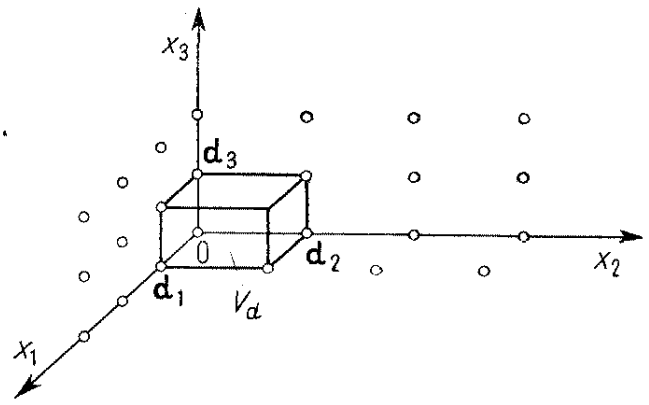
z badaniami budowy kryształów przy pomocy promieni X lub z wykorzystaniem dyfrakcji neutronów.

Okresowość w przestrzeni znajduje swój wyraz w sieci krystalicznej i może być scharakteryzowana trzema translacjami $\mathbf{d}_1, \mathbf{d}_2, \mathbf{d}_3$. Funkcję $F(\mathbf{r})$, gdzie $\mathbf{r}$ jest wektorem o składowych (x, y, z), nazywamy funkcją okresową o okresach d_1, d_2, d_3, jeśli ma ona następującą własność:

$$F(\mathbf{r}+\varrho_1\mathbf{d}_1+\varrho_2\mathbf{d}_2+\varrho_3\mathbf{d}_3) = F(\mathbf{r}), \tag{8.95}$$

gdzie $\varrho_1, \varrho_2, \varrho_3$ są dowolnymi, dodatnimi lub ujemnymi liczbami całkowitymi.

W przypadku ogólnym, kiedy wektory $\mathbf{d}_1, \mathbf{d}_2, \mathbf{d}_3$ nie są ortogonalne, występują pewne komplikacje, których można uniknąć w przypadku wektorów ortogonalnych. Nasze obecne rozważania ograniczymy do najprostszego



Rys. 8.10. Sieć prostokątna, określona przez ortogonalne wektory $\mathbf{d}_1$, $\mathbf{d}_2$, $\mathbf{d}_3$

przypadku i przyjmiemy współrzędne prostokątne x_1, x_2, x_3 o kierunkach pokrywających się z kierunkami $\mathbf{d}_1, \mathbf{d}_2, \mathbf{d}_3$. Wektory, które będziemy rozpatrywać, mają następujące składowe (rys. 8.10):

$$\left.\begin{aligned} \mathbf{r}: &\quad (x_1, x_2, x_3),\\ \mathbf{d}_1: &\quad (d_1, 0, 0),\\ \mathbf{d}_2: &\quad (0, d_2, 0),\\ \mathbf{d}_3: &\quad (0, 0, d_3), \end{aligned}\right\} \tag{8.96}$$

i własność okresowości (8.95) można wyrazić, jak następuje:

$$F(x_1+\varrho_1 d_1,\ x_2+\varrho_2 d_2,\ x_3+\varrho_3 d_3) = F(x_1, x_2, x_3). \tag{8.96a}$$

Wektory $\mathbf{d}_1, \mathbf{d}_2, \mathbf{d}_3$ określają podstawową komórkę sieci o objętości

$$V_d = d_1 d_2 d_3.$$

Funkcję okresową $F(i)$ można rozwinąć w potrójny szereg Fouriera

$$F(x_1, x_2, x_3) = \sum_{h_1 h_2 h_3} C_{h_1 h_2 h_3} \, e^{2\pi i (h_1 b_1 x_1 + h_2 b_2 x_2 + h_3 b_3 x_3)}, \tag{8.97}$$

gdzie

$$b_1 = \frac{1}{d_1}, \quad b_2 = \frac{1}{d_2}, \quad b_3 = \frac{1}{d_3}, \quad h_1, h_2, h_3 \text{ — liczby całkowite.}$$

Trzy wektory odwrotne ([8.15]) $\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3$ są także ortogonalne i określają komórkę sieci odwrotnej ([8.16]) (rys. 8.11):

$$V_b = b_1 b_2 b_3 = \frac{1}{V_d}.$$

Sieć odwrotną można określić także dla struktur skośnokątnych i rozważania tego przypadku prowadzą do podobnych wyrażeń. Współczynniki Fouriera $C_{h_1 h_2 h_3}$ otrzymujemy za pomocą następujących całek:

$$C_{h_1 h_2 h_3} = \frac{1}{V_d} \int_0^{d_1} \int_0^{d_2} \int_0^{d_3} F(x_1, x_2, x_3) \, e^{-2\pi i (h_1 b_1 x_1 + h_2 b_2 x_2 + h_3 b_3 x_3)} \times \\ \times \, dx_1 \, dx_2 \, dx_3. \tag{8.98}$$

Jeżeli funkcja $F(i)$ jest rzeczywista, to mamy

$$C_{-h_1, -h_2, -h_3} = C^*_{h_1, h_2 h_3}.$$

Wyrażenia te są prostym uogólnieniem równań (8.1), (8.2) i (8.3). Oznaczenia odpowiadają sobie według następującego schematu:

okres τ:	d_1, d_2, d_3,
częstotliwość podstawowa ν_0:	b_1, b_2, b_3,
numer harmonicznej n:	h_1, h_2, h_3.

W widmie trójwymiarowym mogą występować różnego rodzaju ograniczenia częstotliwości:

(1) *Trzy niezależne granice w trzech kierunkach*:

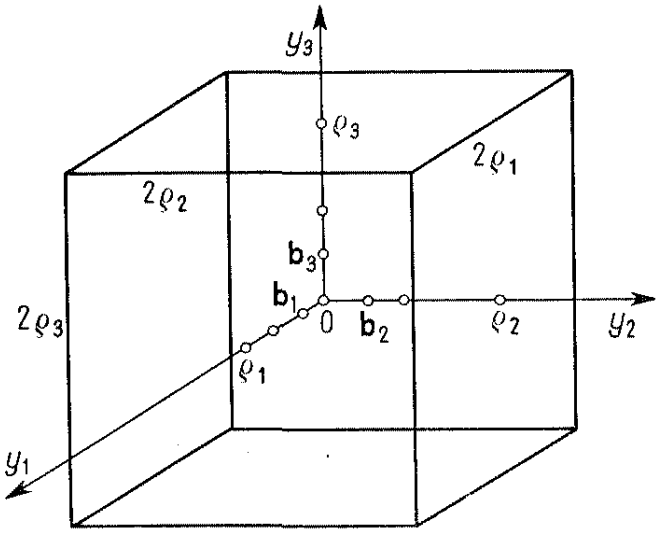
$$|h_1 b_1| \leqslant \varrho_1, \quad |h_2 b_2| \leqslant \varrho_2, \quad h_3 b_3 \leqslant \varrho_3; \tag{8.99}$$

stąd (rys. 8.11):

$$|h_1| \leqslant \frac{\varrho_1}{b_1} = \varrho_1 d_1, \quad |h_2| \leqslant \varrho_2 d_2, \quad |h_3| \leqslant \varrho_3 d_3.$$

Wielkości $\varrho_1, \varrho_2, \varrho_3$ reprezentują maksymalne częstotliwości w kierunkach *1*, *2*, *3*. Sytuacja taka jest po prostu złożeniem trzech oddzielnych 1-wymiarowych sytuacji typu (1), rozważanych w rozdz. 8.6. W tym przypadku zachodzi następująca odpowiedniość oznaczeń:

$$\text{maksymalna częstotliwość } \nu_M \quad \dots \quad \varrho_1, \varrho_2, \varrho_3.$$



Rys. 8.11. Sieć odwrotna jest określona przez wektory $\mathbf{b}_1$, $\mathbf{b}_2$ i $\mathbf{b}_3$. Punkty siatki leżą wewnątrz pokazanego prostopadłościanu, jeśli ϱ_1, ϱ_2 i ϱ_3 są maksymalnymi częstotliwościami w kierunkach *1*, *2*, *3*

(2) Mamy *pojedynczą granicę częstotliwości* ϱ z warunkiem (rys. 8.11):

$$h_2^1 b_1^2 + h_2^2 b_2^2 + h_3^2 b_3^2 \leqslant \varrho^2. \tag{8.100}$$

Warunek ten odpowiada najmniejszej długości fali $1/\varrho$. Rozważmy dokładniej jego znaczenie fizyczne. Każda składowa h_1, h_2, h_3 w rozwinięciu (8.97) przedstawia pewien sinusoidalny rozkład w przestrzeni. Wszystkie punkty leżące na płaszczyźnie

$$h_1 b_1 x_1 + h_2 b_2 x_2 + h_3 b_3 x_3 = 0$$

mają taką samą fazę jak punkt w początku układu współrzędnych. Inne płaszczyzny równoległe o tej samej fazie można opisać równaniem

$$h_1 b_1 x_1 + h_2 b_2 x_2 + h_3 b_3 x_3 = m,$$

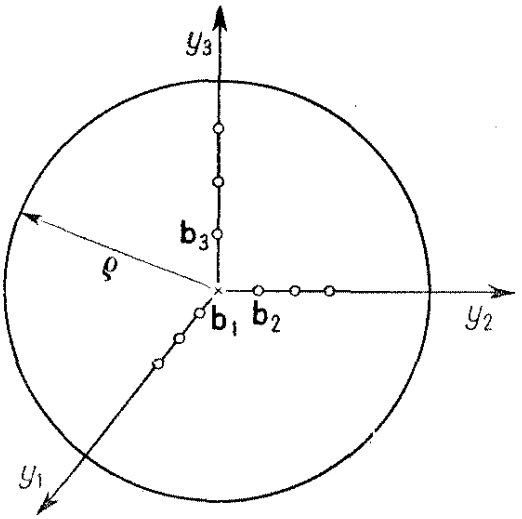
gdzie m jest liczbą całkowitą. Długość fali Λ jest odległością pomiędzy dwiema

sąsiednimi płaszczyznami $\Delta m = \pm 1$ i stąd bez trudu otrzymujemy

$$\frac{1}{A^2} = (h_1 b_1)^2 + (h_2 b_2)^2 + (h_3 b_3)^2 \leqslant \varrho^2, \tag{8.100a}$$

gdzie ϱ określa (8.100). Duże wartości h dają małe długości fali. Wektor $(h_1 b_1, h_2 b_2, h_3 b_3)$ jest ortogonalny do płaszczyzny fal i jego długość jest równa $1/A$.

Funkcje z ograniczonym widmem (warunek (1) lub (2)) mają jedynie skończoną liczbę stopni swobody. Warunek (1) jest łatwiejszy do analizy,



Rys. 8.12. Sieć odwrotna jest określona przez wektory $\mathbf{b}_1$, $\mathbf{b}_2$ i $\mathbf{b}_3$. Maksymalna częstotliwość ϱ nie zależy od kierunku i punkty sieci leżące na zewnątrz kuli o promieniu ϱ są wyłączone

ponieważ stanowi on bezpośrednie uogólnienie zagadnienia (1) z rozdz. 8.6. Analogia z równaniem (8.57) jest następująca:

$$\left.\begin{aligned} N_1 &= 2\varrho_1 d_1 + 1 \quad \text{stopni swobody dla } x_1, \\ N_2 &= 2\varrho_2 d_2 + 1 \quad \text{stopni swobody dla } x_2, \\ N_3 &= 2\varrho_3 d_3 + 1 \quad \text{stopni swobody dla } x_3. \end{aligned}\right\} \tag{8.101}$$

Stąd całkowita liczba stopni swobody jest równa

$$N = N_1 N_2 N_3 \approx 8 V_d \varrho_1 \varrho_2 \varrho_3. \tag{8.102}$$

Powyższe przybliżenie jest słuszne w przypadku, kiedy parametry $\varrho_1, \varrho_2, \varrho_3$ są dużo większe niż b_1, b_2, b_3:

$$\frac{\varrho_1}{b_1} = \varrho_1 d_1 \gg 1, \quad \varrho_2 d_2 \gg 1, \quad \varrho_3 d_3 \gg 1.$$

Metoda próbkowania podana w rozdz. 8.6 (równania (8.59) do (8.64)) może być zastosowana bezpośrednio, jeśli wziąć prostokątną siatkę N punktów próbkowania* wewnątrz komórki $d_1 d_2 d_3$:

$$\left.\begin{aligned} x_{1s} = m_1\xi_1, \quad x_{2s} = m_2\xi_2, \quad x_{3s} = m_3\xi_3, \\ \xi_i = \frac{d_i}{N_i} = \frac{1}{2\varrho_i + b_i} \approx \frac{1}{2\varrho_i}, \\ 0 \leqslant m_i < N_i, \quad i = 1, 2, 3. \end{aligned}\right\} \tag{8.103}$$

Okresowi próbkowania θ odpowiadają ξ_1, ξ_2, ξ_3.

Funkcja wyjściowa może być łatwo odtworzona z wartości próbek przy pomocy potrójnej sumy

$$F = \sum_{m_1 m_2 m_3} F_{m_1 m_2 m_3}\, g_1(x_1 - m_1\xi_1)\, g_2(x_2 - m_2\xi_2)\, g_3(x_3 - m_3\xi_3), \tag{8.104}$$

gdzie

$$F_{m_1 m_2 m_3} = F(m_1\xi_1, m_2\xi_2, m_3\xi_3).$$

Funkcja g została zdefiniowana przez równanie (8.62) i jeżeli uwzględnimy odpowiedniość oznaczeń, otrzymamy

$$g_i(x_i) = \frac{\sin(N_i \pi b_i x_i)}{N_i \sin(\pi b_i x_i)}, \qquad i = 1, 2, 3. \tag{8.105}$$

W ten sposób rozwiązaliśmy postawione zagadnienie; metoda próbkowania może być więc stosowana również w zagadnieniach trójwymiarowych.

Zagadnienie (2) z pojedynczą granicą ϱ (równanie (8.100)) jest bardziej skomplikowane. Pokażemy, że również w tym przypadku funkcja posiada skończoną liczbę stopni swobody, jednakże sposób próbkowania można tylko naszkicować i pozostaje on zagadnieniem otwartym.

W rozwinięciu wzoru (8.97) składowe Fouriera h_1, h_2, h_3 mogą być interpretowane jako punkty siatki prostokątnej:

$$y_1 = h_1 b_1, \quad y_2 = h_2 b_2, \quad y_3 = h_3 b_3, \tag{8.106}$$

gdzie h_1, h_2, h_3 są to dodatnie lub ujemne liczby całkowite. Siatka ta jest znana jako *siatka odwrotna*. Każdy punkt (h_1, h_2, h_3) tej siatki odwrotnej

---

* Punkt próbkowania w przypadku funkcji trzech zmiennych x, y, z odpowiada chwili próbkowania w przypadku funkcji czasu $f(t)$. (*Przyp. tłum.*).

odpowiada jednej składowej Fouriera. Podstawowa komórka V_b tej siatki ma objętość

$$V_b = b_1 b_2 b_3 ;$$

skąd mamy

$$\frac{1}{V_b} = V_d \tag{8.107}$$

punktów siatki odwrotnej, przypadających na jednostkę objętości. Granica (8.100) wyznacza obszar sferyczny w siatce odwrotnej. Ma on objętość $^4/_3\pi\varrho^3$ i zawiera, zgodnie z (8.107),

$$N \approx V_d \, {}^4/_3\pi\varrho^3 \tag{8.108}$$

punktów. Wyrażenie to daje liczbę składowych Fouriera, a zatem i liczbę niezależnych stopni swobody naszej funkcji. Podobieństwo pomiędzy równaniami (8.102) i (8.108) jest oczywiste. Równanie (8.102) odpowiada obszarowi prostokątnemu w siatce odwrotnej, równanie zaś (8.108) — obszarowi sferycznemu.

W zasadzie byłoby możliwe skonstruowanie metody próbkowania przez wykorzystanie N punktów w komórce podstawowej $d_1 d_2 d_3$ siatki, jednakże zagadnienie to nie zostało dotąd w pełni opracowane, nie można bowiem przyjąć prostokątnej lub sześciennej siatki punktów próbkowania, ponieważ nie wykazuje ona wymaganej symetrii sferycznej i odpowiada odmiennemu zagadnieniu (8.99). Można sobie wyobrazić zbiór punktów siatki rozmieszczonych w środkach ciasno ułożonych kul. Prowadzi to do sieci płasko centrowanych, ponieważ inne sieci typu heksagonalnego wykazują mniejszą symetrię*.

Rozważmy zbiór punktów próbkowania rozmieszczonych w płasko centrowanej siatce sześciennej o krawędzi δ. Sześcian δ^3 zawiera cztery punkty próbkowania** i stąd objętość u przypadająca na punkt próbkowania jest równa

$$u = {}^1/_4\delta^3. \tag{8.109}$$

W elementarnej komórce V_d sieci krystalicznej mamy

$$N' = \frac{V_d}{u} = \frac{4V_d}{\delta^3} \tag{8.110}$$

---

* L. Brillouin, *op. cit.*, str. 149.

** L. Brillouin, *op. cit.*, notka 8 na str. 135.

punktów próbkowania. Liczba tych punktów powinna być równa liczbie N stopni swobody (por. równanie (8.108)), stąd związek

$$N' = N, \qquad \frac{4}{3}\pi\varrho^3 = \frac{4}{\delta^3}, \qquad \varrho\delta = \left(\frac{3}{\pi}\right)^{1/3}. \tag{8.111}$$

Odległość pomiędzy płaszczyznami punktów próbkowych równoległymi do osi x_1, x_2 lub x_3 wynosi $\delta_1 = \delta/2$; odległość pomiędzy płaszczyznami prostopadłymi do głównych przekątnych jest równa $\delta_2 = \delta/\sqrt{3}$. Musimy porównać te odległości z najmniejszą długością fali $\lambda = \varrho^{-1}$. Daje to

$$2\varrho\delta_1' = \left(\frac{3}{\pi}\right)^{1/3} \approx 0{,}986, \qquad 2\varrho\delta_2 = \left(\frac{3}{\pi}\right)^{1/3} \frac{2}{\sqrt{3}} \approx 1{,}14 \tag{8.112}$$

zamiast warunku

$$2\varrho\xi = 1$$

w przypadku trójwymiarowym (por. (8.103)).

Aby konsekwentnie rozwinąć metodę próbkowania, trzeba rozważyć funkcję impulsową $g(\mathbf{r})$ odpowiadającą danemu zagadnieniu. Funkcja ta powinna zawierać tylko fale o długości większej niż $1/\delta$. Pozostaje do wyjaśnienia, czy w trzech (lub dwóch) wymiarach warunek ten będzie zgodny z innymi założeniami dotyczącymi równania (8.60), a mianowicie założeniami, że funkcja impulsowa $g(\mathbf{r})$ powinna być równa jedności dla $\mathbf{r} = 0$ i zeru dla innych punktów próbkowania. Jeśli ten drugi warunek nie może być spełniony, to stosowanie metody próbkowania poważnie się komplikuje. Wydaje się, że jedynym praktycznym rozwiązaniem jest postępowanie podobne do rozważonego przy końcu rozdz. 8.10, a mianowicie wprowadzenie ograniczenia typu (1) (równanie (8.99)) przez trzykrotne całkowanie podobnie do (8.94), a następnie wprowadzenie prostokątnej sieci punktów próbkowania (8.103) i (8.104). W każdym razie jest to metoda konsekwentna, o dużym znaczeniu praktycznym.

8.12. ANALIZA KRYSZTAŁÓW ZA POMOCĄ PROMIENI X

Zagadnienia rozważane w poprzednich rozdziałach są ściśle związane z zagadnieniem analizy kryształów za pomocą promieni X. Kryształ jest trójwymiarową okresową strukturą o podstawowej komórce $\mathbf{d}_1, \mathbf{d}_2, \mathbf{d}_3$, nato-

miast funkcja okresowa $F(\mathbf{r})$ w tej sieci odpowiada gęstości elektronów. Natężenie promieni X, odbitych w różnych kierunkach, określa kwadraty modułów $|C_{h_1h_2h_3}|^2$ współczynników Fouriera w rozwinięciu (8.97), jednakże przy pomocy danych eksperymentalnych nie można określić faz tych współczynników*. Dalej, istnieje graniczna liczba współczynników, których moduły można wyznaczyć doświadczalnie przy zastosowaniu promieni X o długości fali λ.

Sinusoidalny rozkład gęstości elektronów o długości fali Λ odbija promienie X padające pod kątem φ, jeżeli jest spełniony warunek Bragga:

$$\lambda = 2\Lambda \cos \varphi. \tag{8.113}$$

Jeżeli badany kryształ obracamy w strumieniu promieni X we wszystkich kierunkach, wówczas jest możliwe określenie wszystkich wartości Λ większych od $^1/_2\lambda$:

$$\Lambda \geqslant \frac{1}{2}\lambda = \frac{1}{\varrho}. \tag{8.114}$$

Z podobnymi ograniczeniami spotykamy się, jeśli posługujemy się metodą Debye-Scherrera. W ten sposób powróciliśmy do zagadnienia (2) z rozdz. 8.11. Przy takich ograniczeniach funkcja F ma nie więcej niż N stopni swobody (równanie (8.108)) i z wyników badań nad rozpraszaniem promieni X przez kryształy otrzymujemy $N/2$ danych.

A. L. Patterson** rozważał funkcję

$$P(u_1, u_2, u_3) = \frac{1}{V_d} \int_0^{d_1} \int_0^{d_2} \int_0^{d_3} F(x_1+u_1, x_2+u_2, x_3+u_3) \times$$
$$\times F(x_1x_2x_3)\mathrm{d}x_1\,\mathrm{d}x_2\,\mathrm{d}x_3. \tag{8.115}$$

Jest to po prostu funkcja autokorelacji (8.77) dla trzech wymiarów. Biorąc zamiast funkcji $F(x_1+u_1 \ldots)$ w równaniu (8.115) jej rozwinięcie Fouriera

---

* Zob. np. L. Brillouin, *op. cit.*, rozdz. 6 i 7; R. W. James, *The Optical Principles of the Diffraction of X-Rays*, London 1950, G. Bell, rozdz. 7.

** A. L. Patterson, Phys. Rev. **46** (1934), str. 372; Z. Krist, Phys. Rew. **90** (1935), str. 517; D. Harker, J. Chem. Phys. **4** (1936), str. 381; R. W. James, *op. cit.*

(8.97) otrzymujemy

$$P(u_1, u_2, u_3,) = \sum_{h_1, h_2, h_3} C_{h_1 h_2 h_3} \, e^{2\pi i \sum h_i b_i u_i} \times$$

$$\times \frac{1}{V_d} \int_0^{d_1} \int_0^{d_2} \int_0^{d_3} F(x_1 x_2 x_3) \, e^{2\pi i \sum h_i b_i x_i} \cdot dx_1 \, dx_2 \, dx_3 =$$

$$= \sum_{h_1, h_2, h_3} C_{h_1 h_2 h_3} C^*_{h_1 h_2 h_3} \, e^{2\pi i (h_1 b_1 u_1 + h_2 b_2 u_2 + h_3 b_3 u_3)}. \tag{8.116}$$

Wyrażenie to jest bezpośrednim uogólnieniem równania (8.78) i zawiera tylko kwadraty modułów CC^*, nie zawiera zaś faz. Metoda Pattersona, podana niezależnie od Wienera, stanowi bezpośrednie rozwinięcie metody funkcji autokorelacji Wienera. Relacje symetrii rozważane w rozdz. 8.9 są słuszne również tutaj i funkcja Pattersona zawiera tylko $^1/_2N$ niezależnych parametrów. Do przedstawienia funkcji Pattersona przy pomocy próbek potrzeba, z uwagi na symetrię, tylko $^1/_2N$ próbek na obszar połowy podstawowej komórki kryształu.

Inną skuteczną metodę analizy o nazwie *splot* podał R. Pepinsky*. Praktycznie odpowiada ona przekształceniom rozważanym w rozdz. 8.10, uogólnionym na trzy wymiary. Metoda ta okazała się nadzwyczaj pożyteczna przy badaniu skomplikowanych struktur. Przy końcu rozdziału 8.10 zalecaliśmy stosowanie impulsów (8.25) ÷ (8.28) jako funkcji określających przekształcenia liniowe. Funkcję (8.26) Pepinsky nazywa *jądrem Dirichleta*, a funkcję (8.27) *jądrem Fejera*, przy czym $C(t)$ jest funkcją, a $f(v)$ widmem (f i C zamienione). Papinsky podał inne typy funkcji określających przekształcenia i rozważył przykłady licznych ważnych zastosowań swojej metody.

W ten sposób wykazaliśmy tu, że badanie budowy kryształów przy użyciu promieni X stanowi trójwymiarowe uogólnienie analizy sygnałów w przypadku jednowymiarowym.

DODATEK DO ROZDZIAŁU 8

Nierówność Schwarza

Aby udowodnić nierówność (8.49), rozważmy całkę

$$I = \int |F + cG|^2 \, dt = \int (F^* + cG^*)(F + cG) \, dt > 0, \tag{8D.1}$$

* R. Pepinsky, *Report to the Office of Naval Research*, Feb. 1, 1952.

gdzie c jest stałą rzeczywistą. Rozwijając wyrażenie podcałkowe otrzymujemy

$$I = c^2 \int G^* G \, \mathrm{d}t + c \int (F^* G + F G^*) \, \mathrm{d}t + \int F^* F \, \mathrm{d}t.$$

Daje to związek:

$$Ac^2 + Bc + C > 0. \tag{8D.2}$$

Ponieważ nie może istnieć pierwiastek rzeczywisty, mamy

$$B^2 - 4AC \leqslant 0$$

lub

$$4AC \geqslant B^2, \tag{8D.3}$$

co jest nierównością Schwarza.

9

ZASADY TERMODYNAMIKI

9.1 WSTĘP

W poprzednich rozdziałach zajmowaliśmy się tym, w jaki sposób można zdefiniować ilość informacji. Ograniczenia wprowadzonej przez nas definicji rozważyliśmy w pierwszym rozdziale, następne zaś rozdziały poświęciliśmy na rozwinięcie naszych ogólnych rozważań oraz przedyskutowanie możliwości ich zastosowania.

Chcemy obecnie omówić nowy aspekt teorii informacji, wynikający stąd, że istnieje ścisły związek pomiędzy teorią informacji i termodynamiką. Stwierdzono mianowicie, że informacja jest związana z entropią. Zaczniemy od podania zarysu termodynamiki i jej zasad*. ([9.1])

9.2. DWIE ZASADY TERMODYNAMIKI. ENTROPIA I NEGENTROPIA

1. Zasada zachowania energii.

2. Zasada Carnota. Według Kelvina, zasada ta jest zasadą degradacji energii:

---

* M. W. Zemansky, *Heat and Thermodynamics*, New York 1951, 3rd ed, McGraw — Hill.

K. K. Darrow, *The Concept of Entropy*, Am. J. Phys. **12** (1944), str. 183; *Entropy and Statictics*, Bell System Tech. J. **21** (1942), str. 51; **22** (1943), str. 108 i str. 362.

L. Brillouin, *Quantenstatistik*, Berlin 1931, Springer.

J. E. Mayer, M. G. Mayer, *Statistical Mechanics*, New York 1940, Wiley.

R. C. Tolman, P. C. Fine, Revs. Modern Phys. **20** (1948), str. 51.

energią wyższego stopnia jakości jest energia mechaniczna lub elektryczna;
energią średniego stopnia jakości jest energia chemiczna;
energią niższego stopnia jakości jest ciepło (energia cieplna).

W układzie izolowanym, zamkniętym całkowita ilość energii jest stała. Przemiany lub reakcje chemiczne zachodzące wewnątrz układu, jeżeli są to przemiany odwracalne, nie zmieniają stopnia jakości energii układu. Jeżeli są to przemiany nieodwracalne, to powodują one zmniejszenie stopnia jakości energii w tym układzie.

W podanej klasyfikacji energia chemiczna zajmuje miejsce pośrednie, ponieważ w procesach przemiany przechodzi ona zarówno w energię cieplną, jak i mechaniczną (elektryczną) i odwrotnie. Wiadomo np. że przy ładowaniu i rozładowywaniu baterii następuje wymiana ciepła i energii elektrycznej.

Stopień jakości energii można dokładnie zdefiniować przy pomocy ujemnej entropii $(-S)$. Druga zasada termodynamiki mówi, że entropia S musi zawsze wzrastać lub co najmniej pozostawać stała, stąd ujemna entropia zawsze maleje.

Definicja entropii. Jeżeli układ otrzymuje ilość ciepła Δq, to jego entropia wzrasta o wartość

$$\Delta S = \Delta q/T, \qquad \Delta q = T\,\Delta S, \tag{9.1}$$

gdzie T jest temperaturą bezwzględną. W stopniach Celsjusza:

$$T^\circ\,[\mathrm{K}] = t^\circ[\mathrm{C}] + 273{,}15.$$

Podamy dla przykładu kilka typowych przypadków.

Rozważmy dwa układy połączone A i B. Przyjmijmy, że mogą one wzajemnie wymieniać ciepło i pracę. Przyjmijmy także, że układ połączony, który oznaczymy AB, jest układem izolowanym. Niech pewna przemiana zachodząca w takim układzie prowadzi do następujących warunków:

Układ	Temperatura	Wykonywana praca	Ciepło otrzymane	Wzrost entropii
A	T_A	W_A	q_A	$\Delta S_A = q_A/T_A$
B	T_B	W_B	q_B	$\Delta S_B = q_B/T_B$

$$\left.\begin{aligned} &\text{Pierwsza zasada: } W_A - q_A + W_B - q_B = 0. \\ &\text{Druga zasada: } \quad \Delta S_A + \Delta S_B \geqslant 0. \end{aligned}\right\} \tag{9.2}$$

Całkowita energia układu AB nie ulega zmianie.

Przykłady przemian nieodwracalnych. Przepływ ciepła z ciała o wyższej temperaturze do ciała o temperaturze niższej:

$$T_A > T_B.$$

Nie zostaje wykonana żadna praca:

$$W_A = 0, \quad W_B = 0.$$

Przekazane ciepło:

$$q_A = -q_B, \quad q_B > 0,$$

$$\Delta S_A + \Delta S_B = q_B\left(\frac{1}{T_B} - \frac{1}{T_A}\right) > 0. \tag{9.3}$$

Całkowita entropia układu wzrosła.

Tarcie, tłumienie lepkie: $T_A = T_B = T$ — temperatura jest jednakowa. Układ A wykonuje pracę, w układzie B wydziela się ciepło:

$$W_A > 0, \quad q_A = 0, \quad W_A = 0, \quad q_B > 0.$$

$$W_A - q_B = 0,$$

$$\Delta S_A + \Delta S_B = \frac{q_B}{T} = \frac{W_A}{T} > 0. \tag{9.4}$$

Całkowita entropia układu wzrosła.

Energię całkowitą układu możemy określić, jeżeli znamy sposób połączenia elementów składowych układu, znamy energię tych elementów i jeśli uwzględnimy wykonywaną pracę i wydzielane ciepło w procesie powstawania tego układu. Całkowita energia U jest funkcją wszystkich fizycznych i chemicznych parametrów określających układ. Jeżeli układ powstał w wyniku łańcucha *przemian odwracalnych*, to jego *całkowitą entropię* można określić jako:

$$S = S_1 + S_2 + \ldots + S_n \tag{9.5}$$

czyli sumę entropii odpowiadających każdemu odwracalnemu krokowi.

Negentropia ($N = -S$) określa stopień jakości energii i musi zawsze maleć. W stwierdzeniu tym tkwi sens zasady Kelvina *degradacji jakości*

energii i dlatego zrozumiały jest fakt, że znaczenie negentropii pierwszy docenił bliski przyjaciel Kelvina, który pisał*:

«Byłoby bardzo pożądane znaleźć słowo wyrażające *zdolność* ciepła zmagazynowanego w danym zbiorniku do wykonywania pracy, tj. termin określający tę wielkość, stratę której nazywamy *rozpraszaniem*. Doskonałe słowo *entropia*, które w związku z tym wprowadził Clausius, zostało przez niego zastosowane niestety do oznaczenia *negacji* wielkości, którą chcemy wyrazić w sposób możliwie najbardziej naturalny. Wprowadzenie drugiego, bardziej odpowiadającego naszym celom, terminu mogłoby prowadzić do pewnego zamieszania. Konieczność wprowadzenia mimo wszystko takiego terminu stanie się bardziej zrozumiała po rozpatrzeniu poniższych, dobrze ilustrujących zagadnienie, przykładów.»

Układ izolowany dysponuje *negentropią***, jeżeli ma on możliwość wykonania pracy mechanicznej lub elektrycznej. Jeżeli temperatura T układu nie jest jednorodna, lecz układ składa się z kilku części o różnych temperaturach, zawiera on pewną ilość negentropii. Ta negentropia może być wykorzystana do wytworzenia pracy mechanicznej, może też zostać po prostu rozproszona i stracona wskutek przewodności cieplnej. Różnica ciśnienia pomiędzy poszczególnymi częściami układu stanowi inny przykład negentropii. Podobnym przykładem może być również różnica potencjałów. Zbiornik sprężonego gazu w pomieszczeniu o ciśnieniu atmosferycznym, naczynie, w którym znajduje się gaz pod zmniejszonym ciśnieniem w podobnym pomieszczeniu, naładowana bateria czy inny układ mogący wytwarzać energię o wysokim stopniu jakości (np. pracę mechaniczną) lub ulegać degradacji na skutek pewnych nieodwracalnych procesów (przewodności cieplnej, oporności elektrycznej, tarcia, tłumienia) — są źródłem negentropii.

E. Schrödinger w swej książce *What is Life?**** podkreślił znaczenie negentropii. Jeżeli żywy organizm szuka pożywienia, to tylko w celu uzyskania z niego negentropii. Negentropia ta jest niezbędna dla pokrycia strat spowodowanych przez pracę mechaniczną lub proste procesy degradacji energii w żywym organizmie. Energia zawarta w pożywieniu nie odgrywa w istocie większej roli, ponieważ jej wartość jest zachowywana i nigdy nie ulega zmniejszeniu; istotnym czynnikiem jest tu negentropia.

---

* P. G. Tait, *Sketch of Thermodynamics*, Edinburgh 1868, Edmonston and Douglas, str. 100.

** Skrót od *negative entropy*. (*Przyp. tłum.*).

*** E. Schrödinger, *What is Life?*, New York 1945, Cambridge Univ. Press.

9.3. NIEMOŻLIWOŚĆ WIECZNEGO RUCHU. MASZYNY CIEPLNE

Maszyna korzystająca tylko z jednego zbiornika ciepła o temperaturze T nie może zamieniać ciepła na pracę, ponieważ byłoby to sprzeczne z warunkiem (9.4) i oznaczałoby zmniejszenie entropii. Maszyna może wykonywać pracę, jeżeli korzysta ona z dwóch zbiorników ciepła A i B o dwu różnych temperaturach:

$$T_A > T_B.$$

Praca maszyny składa się z kolejnych etapów:

(1) Przekazanie ciepła q_A ze zbiornika A do maszyny (ciepło otrzymane).

(2) Wykonanie pracy W przez maszynę (praca wykonana).

(3) Przekazanie ciepła q_B z maszyny do zbiornika B o niższej temperaturze.

Ponieważ *ciepło otrzymane* traktujemy jako dodatnie, mamy tu do czynienia z ujemnym ciepłem otrzymanym $q_B < 0$.

Po zakończeniu pełnego cyklu pracy maszyna powraca do stanu początkowego. Oznacza to, że jej całkowita energia przyjęła wartość początkową:

$$W - q_A - q_B = 0. \tag{9.6}$$

Jeżeli proces jest odwracalny, całkowita entropia maszyny powinna również pozostać bez zmiany:

$$\Delta S_A + \Delta S_B = 0; \tag{9.7}$$

stąd

$$\frac{q_A}{T_A} + \frac{q_B}{T_A} = 0$$

lub

$$q_B = -\frac{T_B}{T_A} q_A. \tag{9.8}$$

Jak już wyjaśniono, q_B jest ujemne.

Innymi słowy: pewna ilość entropii, a mianowicie

$$\Delta S = +\Delta S_A = -\Delta S_B \tag{9.9}$$

przechodzi od temperatury wysokiej T_A do niskiej T_B. Ilości ciepła otrzymywane przy tych dwu temperaturach były równe odpowiednio:

$$q_A = +T_A \Delta S, \quad q_B = -T_B \Delta S. \tag{9.10}$$

Suma $q_A+q_B = (T_A-T_B)\,\Delta S > 0$ jest dodatnia. Maszyna otrzymała ze źródła o wyższej temperaturze więcej ciepła niż go oddała do źródła o temperaturze niższej. Brakująca energia cieplna odpowiada pracy wykonanej przez maszynę (wzór (9.6)). *Współczynnik sprawności cieplnej* definiujemy tu jako stosunek pracy wykonanej do otrzymanego ciepła.

$$\frac{W}{q_A} = \frac{T_A-T_B}{T_A} < 1. \tag{9.11}$$

Przykładem jest maszyna cieplna, w której ciepło przepływa od ciała o wysokiej temperaturze do ciała o temperaturze niższej: zostaje wykonana praca.

Chłodziarka jest odwrotnym typem maszyny cieplnej. *Praca dostarczana* wywołuje przepływ ciepła od źródła o niższej temperaturze do źródła o temperaturze wyższej.

W przykładach tych łatwo można rozpoznać różne stopnie jakości energii. Cała praca może zostać zamieniona na ciepło (wzór (9.4)), ale ciepło w pracę można zamienić tylko częściowo (wzór (9.11)).

Jeżeli ciepło mierzymy w jednostkach energii, to w układzie CGS q wyraża się w ergach, a nie w kaloriach. Mierząc temperaturę w stopniach Kelvina, entropię otrzymujemy w ergach na stopień. Jeśli przyjmiemy, że T ma wymiar energii (rozdz. 1.1), entropia będzie wielkością bezwymiarową. W takim przypadku będziemy uważali, że stała k jest liczbą bezwymiarową:

$$k = 1{,}38 \cdot 10^{-16}.$$

Wielkość kT będzie miała nadal wymiar energii.

Termodynamika dostarcza ważnych informacji o zagadnieniach związanych z ciepłem i temperaturą. Jeśli rozważa się tylko procesy odwracalne, możliwe jest określenie całkowitej energii U i całkowitej entropii S dla danego układu fizycznego. Obie wielkości U i S są funkcjami zmiennych fizycznych określających stan układu.

Jako przykład zastosowania metod termodynamiki rozważmy dielektryk o przenikalności dielektrycznej $\varepsilon(T)$ zależnej od temperatury. Niech na przykład:

$$D = \varepsilon E, \qquad \varepsilon(T) = \varepsilon_0 + \frac{C}{T-\theta}, \tag{9.12}$$

przy czym

D — przesunięcie elektryczne,

C — stała Curie,

E — pole elektryczne,

θ — temperatura Curie.

Z rozważań teoretycznych wynika, że pojemność cieplna substancji zależy od polaryzacji, a w związku z tym i od przesunięcia D. Dokładna analiza prowadzi do ustalenia zależności pomiędzy tymi różnymi wielkościami: całkowita energia U i entropia S substancji dane są przez

$$U(T, D) = U_0(T) + \frac{1}{2} D^2 \left(\frac{1}{\varepsilon} - T \frac{\partial}{\partial T} \left(\frac{1}{\varepsilon} \right) \right), \tag{9.13}$$

$$S(T, D) = S_0(T) - \frac{1}{2} D^2 \frac{\partial}{\partial T} \left(\frac{1}{\varepsilon} \right), \tag{9.14}$$

gdzie U_0 i S_0 odnoszą się do materiału niespolaryzowanego ($D = 0$).

Zakłada się, że materiał ten jest twardy i nieściśliwy, i zaniedbuje się elektrostrykcję. Podobne wyrażenia można wyprowadzić dla pola magnetycznego*, które ma duże znaczenie jako podstawa znanej metody oziębiania przez rozmagnesowywanie (Debye–Giauque), pozwalającej na osiąganie temperatur rzędu ułamków stopnia w skali absolutnej.

9.4. STATYSTYCZNA INTERPRETACJA ENTROPII

Wielkość zwana *entropią* formalnie jest zdefiniowana wzorem (9.1) i wykazaliśmy na kilku przykładach jej związek ze *stopniem jakości energii*. Druga zasada termodynamiki mówi, że entropia w układzie zamkniętym zawsze wzrasta, podczas gdy stopień jakości energii zawsze maleje (degradacja energii Kelvina).

Jasną interpretację tej ogólnej zależności podali J. W. Gibbs, J. C. Maxwell, L. Boltzmann; uzupełnił ją M. Planck. Podstawę tych wyjaśnień znaleziono w atomowej strukturze materii i kwantowych prawach stabilności atomów, cząsteczek i kryształów. Atomy są tak małe, że w każdej ilości materii, na której możemy przeprowadzać doświadczenia, znajduje się ich ogromna liczba. Gramoatom materii zawiera około $6 \cdot 10^{23}$ atomów. Z tej przyczyny zupełnie

* L. Brillouin, H. P. Iskenderian, Elec. Commun. **25** (1948), str. 300.

nie jesteśmy w stanie śledzić ruchów pojedynczych atomów. Wielkości, które jedynie możemy zaobserwować, są w rzeczywistości wielkościami średnimi. Statystyczna teoria termodynamiki oparta jest na tych ogólnych uwagach i stanowi obecnie jeden z najbardziej zaawansowanych działów fizyki. Wykazano, że entropia jest związana z prawdopodobieństwem. Sztucznie można stworzyć układ o bardzo mało prawdopodobnej strukturze. Jednakże jeśli układ ten pozostawić samemu sobie, to jego struktura będzie stopniowo dążyła ku strukturze bardziej prawdopodobnej. Prawdopodobieństwo wykazuje więc naturalną tendencję wzrostu, podobnie jak entropia. Dokładny związek pomiędzy prawdopodobieństwem a entropią podaje znana zależność Boltzmanna–Plancka:

$$S = k \ln P, \tag{9.15}$$

gdzie $k = 1{,}38 \cdot 10^{-16}$ erg/stopień (stała Boltzmanna), S jest entropią rozważanego układu, P zaś liczbą *elementarnych mikrostanów*, które prowadzą do takiego samego stanu makroskopowego. Aby pokazać związek entropii z rachunkiem prawdopodobieństwa, rozpatrzmy przede wszystkim dokładne znaczenie liczby P.

Zgodnie z teorią kwantową stany atomów i cząsteczek mogą przyjmować tylko skończoną liczbę różnych postaci. Na poziomie atomowym materii nie ma ciągłości, lecz tylko dyskretne, stabilne lub niestabilne stany i atomy przeskakują z jednego stanu do drugiego, emitując przy tym lub pochłaniając energię. Każdą z tych dyskretnych konfiguracji stanów skwantowanego układu fizycznego nazwał Planck *kompleksją* i nazwa ta pozostała w literaturze naukowej*.

Rozważmy obecnie układ fizyczny: część kryształu lub gaz w zbiorniku w określonych warunkach, tzn. o danych: objętości, ciśnieniu, energii całkowitej, temperaturze i składzie chemicznym. *Zmienne makroskopowe* są wielkościami, które możemy mierzyć w laboratorium. Nie wystarczą one jednak do jednoznacznego opisania stanu układu. Istnieje bowiem olbrzymia liczba *zmiennych mikroskopowych*, których nie jesteśmy w stanie dokładnie zmierzyć (położenia i prędkości wszystkich atomów, stany kwantowe tych atomów lub struktur molekularnych itd.). Wszystkie te nieznane wielkości umożliwiają przyjmowanie przez układ wielkiej liczby stanów kwantowych — mikrosta-

* W oryginale *complexion*. W literaturze polskiej przyjął się termin mikrostan. (*Przyp. tłum.*).

nów. Całkowitą liczbę* P mikrostanów wykorzystaliśmy we wzorze (9.15), wiążąc ją z entropią układu. Liczba ta jest bardzo duża, jednakże skończona i wzór (9.15) ma określony sens.

W naszych rozważaniach będziemy korzystać także z innego ważnego związku. Jest to znana zależność pomiędzy energią i częstotliwością

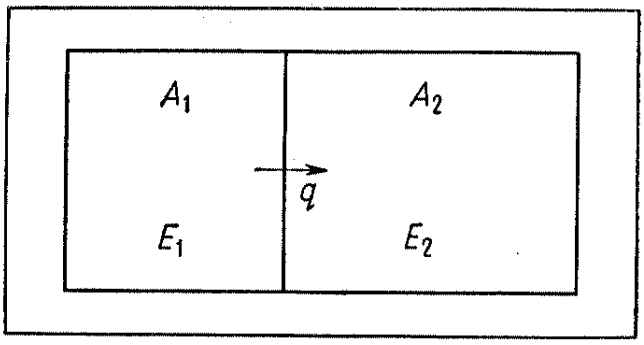
$$E_1 - E_2 = h\nu, \tag{9.16}$$

gdzie E_1 oznacza energię początkową, E_2 energię końcową układu atomowego lub cząsteczkowego emitującego drgania o częstotliwości ν oraz $h = 6{,}6 \cdot 10^{-27}$ erg · s (stała Plancka). Kiedy układ fizyczny traci energię, następuje emisja promieniowania, a kiedy ją zyskuje, następuje absorpcja. Ponieważ poziomy energetyczne są dyskretne, częstotliwości emisji (lub absorpcji) tworzą widmo dyskretne charakterystyczne dla rozważanego układu atomowego lub cząsteczkowego.

9.5. PRZYKŁADY ROZWAŻAŃ STATYSTYCZNYCH

Na kilku przykładach wykażemy obecnie, w jakim stopniu nasze rozważania statystyczne odpowiadają danym eksperymentalnym. Rozpatrzmy dwa ciała A_1 i A_2, będące w kontakcie umożliwiającym jedynie wymianę ciepła. Układ A_1A_2 umieszczamy w izolującym zbiorniku V, tak jak to pokazano na rys. (9.1). W takich warunkach całkowita energia

$$E_1 + E_2 = E \tag{9.17}$$



Rys. 9.1. Dwa ciała A_1 i A_2 mające, odpowiednio, energie E_1 i E_2 mogą wymieniać tylko ciepło. Układ A_1A_2 umieszczony jest w izolującym zbiorniku V

* Liczbę P określa się w literaturze polskiej terminem prawdopodobieństwo termodynamiczne. (*Przyp. tłum.*).

układu pozostaje stała. Musimy jednak rozpatrzyć możliwość przekazania q kalorii od A_1 do A_2:

$$\Delta E_1 = -q, \quad \Delta E_2 = +q. \tag{9.18}$$

Chcemy znaleźć najbardziej prawdopodobny rozkład energii pomiędzy A_1 i A_2. Dla A_1 liczba mikrostanów P_1 jest funkcją E_1, podobnie dla A_2 liczba mikrostanów P_2 jest funkcją E_2. Całkowita liczba mikrostanów dla układu złożonego wynosi

$$P = P_1(E_1) \cdot P_2(E_2), \tag{9.19}$$

ponieważ każdy mikrostan ciała A_1 może być ustawiony w parze z dowolnym spośród $P_2(E_2)$ mikrostanów ciała A_2. Najbardziej prawdopodobnym rozkładem energii jest rozkład, przy którym P osiąga maksimum. Jeżeli P jest maksymalne, to nie ulega ono zmianie przy przekazywaniu bardzo małej ilości ciepła q od A_1 do A_2 (wzór (9.18)). Załóżmy, że P_1 i P_2 są ciągłymi funkcjami E_1 i E_2 oraz że są one różniczkowalne. Jeżeli więc uwzględnimy znaki we wzorze (9.18), otrzymamy

$$\Delta P = \frac{\mathrm{d}P}{\mathrm{d}q} q = \left(-\frac{\mathrm{d}P_1}{\mathrm{d}E_1} P_2 + P_1 \frac{dP_2}{\mathrm{d}E_2} \right) q = 0. \tag{9.20}$$

Z poprzedniego wyrażenia wynika

$$\frac{1}{P_1} \frac{\mathrm{d}P_1}{\mathrm{d}E_1} = \frac{1}{P_2} \frac{\mathrm{d}P_2}{\mathrm{d}E_2} \tag{9.21}$$

lub

$$\frac{\mathrm{d} \ln P_1}{\mathrm{d}E_1} = \frac{\mathrm{d} \ln P_2}{\mathrm{d}E_2}. \tag{9.22}$$

Jeśli teraz skorzystamy ze wzoru Boltzmanna–Plancka (9.15), określającego entropię S_1 i S_2 ciał A_1 i A_2, otrzymamy

$$\frac{1}{k} \frac{\mathrm{d}S_1}{\mathrm{d}E_1} = \frac{1}{k} \frac{\mathrm{d}S_2}{\mathrm{d}E_2}. \tag{9.23}$$

W naszym przykładzie jedynym rodzajem energii wymienianej pomiędzy A_1 i A_2 było ciepło, stąd zgodnie ze wzorem (9.1)

$$\frac{\mathrm{d}S_1}{\mathrm{d}E_1} = \frac{\mathrm{d}S_1}{\mathrm{d}Q_1} = \frac{1}{T_1}. \tag{9.24}$$

Lewa strona równania (9.23) odpowiada odwrotności temperatury bezwzględnej $\frac{1}{T_1}$ ciała A_1, strona prawa zaś odpowiada $\frac{1}{T_2}$. W ten sposób ostatecznie mamy:

$$\frac{1}{T_1} = \frac{1}{T_2} \quad \text{lub} \quad T_1 = T_2. \tag{9.25}$$

Najbardziej prawdopodobny rozkład energii pomiędzy A_1 i A_2 odpowiada więc jednakowym temperaturom obu ciał.

W ten sposób pokazaliśmy, że otrzymane przez nas zależności prowadzą do wyników zgodnych z dobrze znanymi danymi eksperymentalnymi.

Ważnym punktem w naszym rozumowaniu było założenie, że $P_1(E_1)$ i $P_2(E_2)$ są funkcjami ciągłymi i różniczkowalnymi. Jest to słuszne jedynie dla dużych układów fizycznych, zawierających ogromną liczbę atomów, w przypadku kiedy E_1 i E_2 są bardzo duże w porównaniu z kwantowymi skokami energii.

9.6. FLUKTUACJE ENERGII. WZÓR GIBBSA

Rozważmy bardzo duże ciało. Przyjmijmy np., że A_1 jest bardzo dużym zbiornikiem wody starannie odizolowanym od otoczenia. Odpowiada to *termostatowi* — urządzeniu utrzymującemu stałą temperaturę. Temperatura T_1 układu jest istotnie tak długo stała, jak długo ciepło wymieniane q pozostaje bardzo małe w porównaniu z całkowitą energią E_1 termostatu A_1.

Dla takiego termostatu, na podstawie wzorów (9.15) i (9.24) otrzymujemy

$$\frac{\mathrm{d} \ln P_1}{\mathrm{d}E_1} = \frac{1}{kT_1} \tag{9.26}$$

lub po scałkowaniu

$$\ln P_1 = \frac{E_1}{kT_1} + C_1, \tag{9.27}$$

gdzie C_1 jest stałą całkowania; stąd

$$P_1 = P_0 \, \mathrm{e}^{E_1/kT_1}, \tag{9.28}$$

gdzie

$$P_0 = \mathrm{e}^{C_1}.$$

Taka zależność pomiędzy liczbą mikrostanów P_1 i energią E_1 jest typowa dla termostatu.

Wprowadźmy do termostatu A_1 element A_2 (drobną ilość substancji) i rozważmy jego zachowanie w stałej temperaturze T_1, utrzymywanej przez termostat. Zależności (9.21), (9.24) i (9.25) określają najbardziej prawdopodobną energię E_2. Mogą również wystąpić inne wartości energii, odpowiadają im jednak mniejsze prawdopodobieństwa. Ciało A_2 nie zawsze będzie miało energię najbardziej prawdopodobną, lecz będzie wykazywać *fluktuacje energii.*

Z warunku (9.17) mamy

$$E_1 = E - E_2. \tag{9.29}$$

Po podstawieniu tej wartości do wzoru (9.28) otrzymamy

$$P_1 = B_1 \, e^{-E_2/kT_1}, \tag{9.30}$$

gdzie

$$B_1 = P_0 \, e^{E/kT_1}.$$

Liczba mikrostanów całego układu w dalszym ciągu wyraża się wzorem (9.19), stąd

$$P = P_2(E_2) B_1 \, e^{-E_2/kT_1}. \tag{9.31}$$

Wynika stąd, że w termostacie o temperaturze T_1 każdy z P_2 mikrostanów ciała A_2 odpowiadających pewnej energii E_2 ma prawdopodobieństwo

$$p(T) = B_1 \, e^{-E_2/kT_1}. \tag{9.32}$$

To ostatnie wyrażenie po raz pierwszy wprowadził Gibbs. Przy określaniu stanu ciała A_2 utrzymywanego w temperaturze T_1 wykorzystał on wyrażenie (9.32), zamiast jako punkt wyjścia potraktować równanie (9.15), jak to zrobił Boltzmann. Można wykazać, że oba te sposoby rozumowania są sobie równoważne. Wzór Gibbsa (9.32) wykorzystamy dalej przy analizie zagadnień, w których mamy do czynienia z fluktuacjami.

Uzyskane wyniki ilustrują możliwości wykorzystania rozważań statystycznych przy wyjaśnianiu zagadnień termodynamiki.

Ciało o temperaturze T znajduje się w stanie ciągłego ruchu. W przypadku gazu cząsteczki wewnątrz zbiornika poruszają się w różnych kierunkach

i bardzo często zderzają się ze sobą. Średnia energia kinetyczna E_k cząsteczek wynosi

$$\overline{E}_k = {}^1/_2 kT \text{ na stopień swobody} \tag{9.33}$$

i energia tego *ruchu cieplnego* jest właśnie energią zawartą w gazie o temperaturze T.

W krysztale atomy zajmują określone miejsce w przestrzeni, tworząc regularną siatkę krystaliczną. Ruch cieplny w kryształach objawia się w drganiach atomów wokół ich punktów równowagi. Jak długo częstotliwość tych drgań jest mała, ich energia kinetyczna wyraża się wzorem (9.33) i łatwo można obliczyć średnią całkowitą energię każdego oscylatora jako sumę energii kinetycznej i potencjalnej. A więc dla niskoczęstotliwościowego oscylatora mamy:

$$\overline{E}_c = \overline{E}_k + \overline{E}_p, \qquad \overline{E}_p = \overline{E}_k = {}^1/_2 kT, \tag{9.34}$$

$$\overline{E}_c = kT. \tag{9.35}$$

9.7. OSCYLATOR KWANTOWY

Proste te wyrażenia przestają być słuszne dla oscylatorów o wyższych częstotliwościach, kiedy to kwant $h\nu$, w porównaniu z kT, nie jest już wielkością małą. W takim przypadku musimy uwzględnić fakt, że oscylator kwantowy może przyjmować tylko dyskretne wartości energii:

$$E = 0,\ h\nu,\ 2h\nu\ \ 3h\nu,\ \ldots,\ nh\nu,\ \ldots \tag{9.36}$$

Jeżeli oscylator jest umieszczony w termostacie o temperaturze T, to poziomy energii mają prawdopodobieństwa określone przez wzór (9.32):

$$p_0 = B, \qquad p_1 = B\,\mathrm{e}^{-h\nu/kT}, \qquad p_2 = B\,\mathrm{e}^{-2h\nu/kT}, \ \ldots$$

$$\ldots, p_n = B\,\mathrm{e}^{-nh\nu/kT}, \ \ldots \tag{9.37}$$

Współczynnik B można obliczyć korzystając z warunku, że suma prawdopodobieństw (9.37) musi być równa jedności:

$$1 = \sum_{i=0}^{\infty} p_i = B(1+\mathrm{e}^{-x}+\mathrm{e}^{-2x}+\ldots+\mathrm{e}^{-nx}+\ldots) = \frac{B}{1-\mathrm{e}^{-x}}, \tag{9.38}$$

gdzie

$$x = \frac{h\nu}{kT}. \tag{9.39}$$

Stąd

$$B = 1 - e^{-x} \tag{9.40}$$

oraz

$$p_0 = 1 - e^{-x}, \quad p_1 = e^{-x} - e^{-2x}, \quad \ldots$$
$$\ldots, p_n = e^{-nx} - e^{-(n+1)x}, \ldots \tag{9.41}$$

Korzystając ze wzoru (9.41) można otrzymać proste wyrażenie na sumę prawdopodobieństw poziomów energetycznych leżących powyżej pewnej granicy:

$$E \geqslant mh\nu, \quad n \geqslant m,$$

$$\sum_{n=m}^{\infty} p_n = p_m + p_{m+1} + p_{m+2} + \ldots = e^{-mx}. \tag{9.42}$$

Mając obliczone prawdopodobieństwa p_n dyskretnych poziomów energii, możemy obliczyć średnią energię $\bar{E}$ oscylatora kwantowego

$$\bar{E} = \bar{n}h\nu,$$

gdzie $\bar{n}$ — średnia liczba kwantów $h\nu$:

$$\bar{n} = \sum_{n=0}^{\infty} np_n = B(0 + e^{-x} + 2e^{-2x} + \ldots + n\,e^{-nx} + \ldots) = B\frac{e^{-x}}{(1-e^{-x})^2}. \tag{9.43}$$

Sumę w tym wyrażeniu obliczyliśmy korzystając z tego, że jest ona pochodną względem x następującej sumy:

$$-\sum_{n=0}^{\infty} e^{-nx} = -1/(1-e^{-x}).$$

Podstawiając zamiast B prawą stronę wzoru (9.40) mamy

$$\bar{n} = 1/(e^x - 1), \tag{9.44}$$

gdzie

$$x = h\nu/kT,$$

tak że

$$\overline{E} = h\nu/(e^{h\nu/kT}-1). \tag{9.45}$$

Jest to znany wzór Plancka, będący podstawą teorii promieniowania *ciała doskonale czarnego.*

W zagadnieniu oscylatora kwantowego (wzory (9.38) ÷ (9.44)) należy podkreślić różnicę pomiędzy:

stanem najbardziej prawdopodobnym, odpowiadającym największej wartości p, tzn. $p_0 = B$, $n = 0$;

średnią liczbą kwantów $\overline{n}$ (wzór (9.44)).

Dla bardzo małych wartości x:

$$x \ll 1, \quad h\nu \ll kT, \quad \frac{1}{e^x-1} \approx \frac{1}{x},$$

mamy

$$\overline{E} = h\nu/x = kT. \tag{9.46}$$

Jest to przypadek niskich częstotliwości i wynik ten jest identyczny ze wzorem (9.35).

9.8. FLUKTUACJE

Obecnie, przy analizie dokładności pomiarów fizycznych posłużymy się otrzymanymi w poprzednim rozdziale wynikami. W danej temperaturze T pobudzenie cieplne wszystkich części aparatu pomiarowego wywołuje nieuporządkowane drgania, uniemożliwiające pomiar mierzonej wielkości z nieograniczenie dużą dokładnością.

Pobudzenie cieplne wywołuje ruchy Browna, powoduje powstawanie szumów w obwodach elektrycznych, w urządzeniach akustycznych itd. Wszystkie te zagadnienia rozpatrzymy pod kątem zastosowania teorii informacji do zagadnień fizycznych. Pozwoli to dostrzec interesujący związek pomiędzy informacją i entropią oraz przeprowadzić wnikliwą analizę fundamentalnych ograniczeń dokładności pomiarów fizycznych.

Energia cieplna nie zmniejszałaby dokładności pomiarów, jeśli jej wartość byłaby stała. Istotnym czynnikiem utrudniającym obserwację jest ciągły stan fluktuacji, przy którym nie można przewidzieć rzeczywistej energii układu

w danym czasie. Energia ta może różnić się znacznie od wartości średniej.

W poprzednim paragrafie podkreśliliśmy różnicę pomiędzy średnią i najbardziej prawdopodobną wartością energii. Rozważmy teraz zjawisko fluktuacji energii oscylatora kwantowego. Przyjmijmy, że w pewnej chwili czasu układ posiada

$$n = \bar{n}+m \text{ kwantów}, \tag{9.47}$$

gdzie m jest fluktuacją ponad wartość średnią $\bar{n}$. Obliczmy wartość średniokwadratową $\overline{n^2}$:

$$n^2 = (\bar{n}+m)^2 = (\bar{n})^2+2\bar{n}m+m^2. \tag{9.48}$$

Obliczając średnią $\overline{n^2}$ musimy pamiętać, że $\bar{m} = 0$, ponieważ $\bar{n}$ jest wartością średnią n; dlatego

$$\overline{n^2} = (\bar{n})^2+\overline{m^2}. \tag{9.49}$$

Jeżeli znamy $\overline{n^2}$, łatwo możemy ze wzoru (9.49) obliczyć wartość średnio kwadratową fluktuacji. Obliczenie $\overline{n^2}$ również nie jest trudne, ponieważ znamy prawdopodobieństwa stanów kwantowych oscylatora (wzór (9.37)). Mianowicie, korzystając ze wzoru (9.43) możemy napisać

$$\overline{n^2} = \sum_{n=0}^{\infty} n^2 p_n = B \sum_{n=0}^{\infty} n^2 \mathrm{e}^{-nx} = -B \frac{\mathrm{d}}{\mathrm{d}x} \sum_{n=0}^{\infty} n \mathrm{e}^{-nx} =$$

$$= -B \frac{\mathrm{d}}{\mathrm{d}x} [\mathrm{e}^{-x}/(1-\mathrm{e}^{-x})^2]. \tag{9.50}$$

Po zróżniczkowaniu i podstawieniu wartości B ze wzoru (9.40) otrzymamy

$$\overline{n^2} = B \frac{\mathrm{e}^{-x}+\mathrm{e}^{-2x}}{(1-\mathrm{e}^{-x})^3} = \frac{\mathrm{e}^{-x}+\mathrm{e}^{-2x}}{(1-\mathrm{e}^{-x})^2} = \frac{\mathrm{e}^{x}+1}{(\mathrm{e}^{x}-1)^2}. \tag{9.51}$$

Porównując wzory (9.49) i (9.51) oraz podstawiając wzór (9.44) otrzymujemy

$$\overline{m^2} = \overline{n^2}-(\bar{n})^2 = \frac{\mathrm{e}^{x}+1-1}{(\mathrm{e}^{x}-1)^2} = (\bar{n})^2+\bar{n}. \tag{9.52}$$

Jest to znana zależność wiążąca fluktuacje z wartościami średnimi, otrzymana po raz pierwszy przez Plancka i Einsteina.

Dla *oscylatora niskoczęstotliwościowego* ($h\nu \ll kT$) średnia liczba

kwantów $\bar{n}$ jest bardzo duża i wyrażenie (9.52) możemy zastąpić przez

$$\overline{m^2} = (\bar{n})^2. \tag{9.53}$$

Korzystając z ostatniego równania możemy obliczyć fluktuacje energii ε:

$$E = \bar{E} + \varepsilon, \qquad \overline{\varepsilon^2} = \overline{m^2}(h\nu)^2 = (\bar{E})^2, \tag{9.54}$$

gdzie $\bar{E} = \bar{n}h\nu$ jest średnią energią oscylatora.

Dla *niskich częstotliwości* pierwiastek kwadratowy ze średniokwadratowej fluktuacji jest równy średniej energii, fluktuacje są więc w przybliżeniu tego samego rzędu co wartość średnia energii:

$$h\nu \ll kT, \qquad (\overline{\varepsilon^2})^{1/2} = \bar{E} = kT. \tag{9.55}$$

Dla *wysokich częstotliwości* ($h\nu \gg kT$) wartość średnia $\bar{n}$ jest mała, a $(\bar{n})^2$ jeszcze mniejsza; dlatego

$$\overline{m^2} \approx \bar{n} \ll 1 \tag{9.56}$$

lub

$$\overline{\varepsilon^2} \approx \bar{n}(h\nu)^2 = \bar{E}h\nu. \tag{9.57}$$

Średnia energia fluktuacji dla wysokich częstotliwości jest znacznie większa, ponieważ $\bar{E}$ jest mniejsze niż $h\nu$, i stąd

$$\bar{E}h\nu \gg \overline{E^2}. \tag{9.58}$$

Fluktuacje przy wysokich częstotliwościach (wzór (9.57)) są większe niż fluktuacje przy częstotliwościach niskich (wzór (9.54)). Wynik ten jest typowy dla zjawisk kwantowych.

Rozważmy teraz układ N niskoczęstotliwościowych oscylatorów harmonicznych. Całkowita energia średnia układu $\bar{E}_t$ jest równa

$$\bar{E}_t = N\bar{E} = NkT. \tag{9.59}$$

W tym przypadku występują fluktuacje ε_t:

$$E_t = \bar{E}_t + \varepsilon_t. \tag{9.60}$$

Fluktuacje N oscylatorów są niezależne i stąd, zgodnie ze wzorami (9.54) i (9.55), mamy

$$\overline{\varepsilon_t^2} = N\overline{\varepsilon^2} = N(kT)^2. \tag{9.61}$$

10

POBUDZENIE CIEPLNE I RUCHY BROWNA

10.1. RUCH CIEPLNY

Zanim przejdziemy do praktycznych zastosowań, przypomnijmy sobie jeden z najważniejszych wniosków z poprzedniego rozdziału, a mianowicie, że przy niskiej częstotliwości na każdy stopień swobody przypada* $E_k = kT/2$.

Ruchy Browna

Obserwując pod mikroskopem małe cząstki zanurzone w cieczy, można zauważyć, że poruszają się one nieustannie i że ruch ten ma charakter przypadkowy. Pierwsi obserwatorzy tego zjawiska sądzili, że są to cząstki żywe. Szybko jednak ustalono, że ich ruch jest wywołany wyłącznie przypadkowym bombardowaniem przez cząsteczki cieczy. W sposób zadowalający wyjaśnił to po raz pierwszy Einstein, a następnie Perrin uzyskał potwierdzenie na drodze eksperymentalnej. Dalsze doświadczenia Perrina pozwoliły mu określić rozmiary atomów, ponieważ mógł on bezpośrednio znaleźć liczbę Avogadra A, tj. liczbę atomów lub cząsteczek w jednym molu substancji.

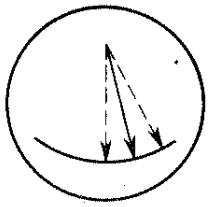
Graniczna dokładność pomiarów

Rozważmy zagadnienie mierzenia prądu przy użyciu amperomierza. Strzałka przyrządu umocowana jest na osi i w związku z tym ma jeden stopień swobody, jak pokazano na rys. 10.1. Energia kinetyczna przypadkowych

---

* Wniosek ten w fizyce jest znany jako twierdzenie o ekwipartycji energii kinetycznej na stopnie swobody. (*Przyp. tłum.*).

drgań strzałki jest równa $E_k = kT/2$; wielkość tych drgań można łatwo obliczyć. Ich amplitudę możemy zmniejszyć ochładzając instrument. Daje to jednak wynik tylko częściowy i, jeśli chcemy uzyskać stabilny stan niskiego pobudzenia cieplnego, musimy również ochłodzić resztę obwodu.

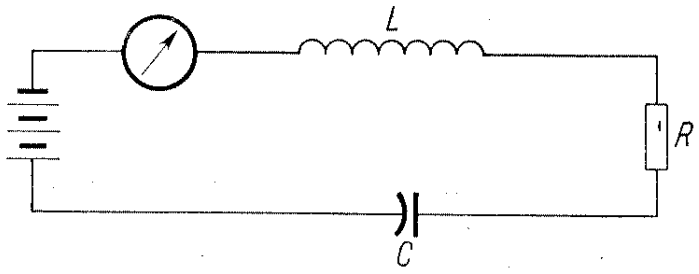


Rys. 10.1. Strzałka amperomierza umocowana jest na osi i dlatego ma jeden stopień swobody

Problem szumu (np. w opornikach) był starannie badany. Jednym z pierwszych, którzy zajmowali się tym zagadnieniem, był Nyquist. Analizując obwód szeregowy L, C, R (pokazany na rys. 10.2) przy jednakowej temperaturze wszystkich elementów, łatwo zauważyć, że ma on jeden stopień swobody (prąd I); stąd mamy

$$^1/_2 kT = E_k = {}^1/_2 LI^2. \tag{10.1}$$

Wynika stąd, że przez obwód płyną przypadkowe prądy, których wielkość określa powyższe równanie. Jedynym sposobem zmniejszenia tych prądów jest, jak poprzednio, ochłodzenie całego obwodu. Można określić rozkład częstotliwości tych prądów; zagadnieniem tym zajmiemy się dalej.



Rys. 10.2. Obwód szeregowy złożony ze źródła siły elektromotorycznej, amperomierza, indukcyjności L, opornika R i pojemności C. W oporniku jest generowany szum

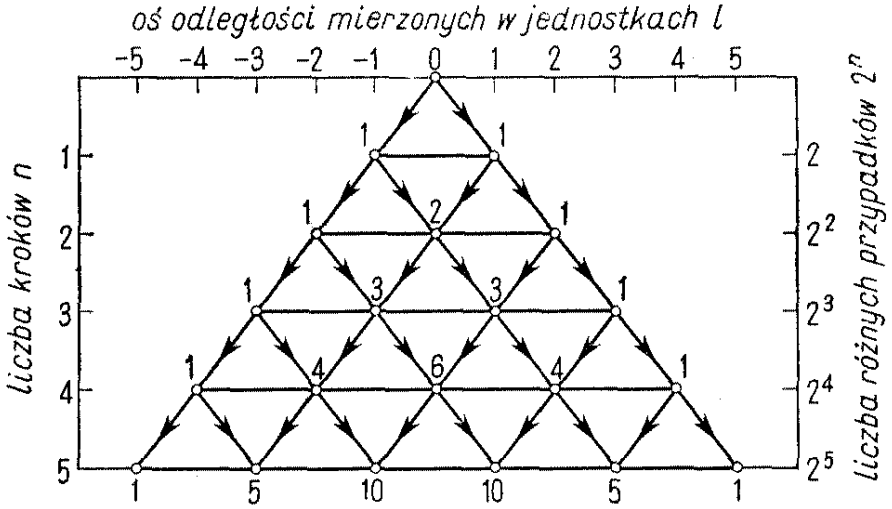
10.2. BŁĄDZENIE PRZYPADKOWE

Rozważania nasze rozpoczniemy od zagadnienia statystycznego, mającego duże znaczenie we wszystkich przypadkach ruchu przypadkowego. Cząstka porusza się wzdłuż linii prostej, pokonując w równych odcinkach czasu równe odległości w kierunku na lewo lub na prawo. Oznaczmy jednostkę przesunięcia cząstki przez l, natomiast momenty czasu, w jakich odbywają

się te przesunięcia, przez 1, 2, 3 itd. Na rysunku 10.3 pokazaliśmy ogólny schemat. Widzimy, że jest to dobrze znany trójkąt Pascala; liczba różnych dróg złożonych z n kroków, którymi można dojść do wybranego punktu, jest równa $\frac{n!}{q!(n-q)!}$, jeśli odległość tego punktu od początku* wynosi

$$\left.\begin{aligned} m &= -n+2q, \\ q &= \frac{n+m}{2}, \qquad n-q = \frac{n-m}{2}. \end{aligned}\right\} \tag{10.2}$$

Prawdopodobieństwo $\pi_{n,m}$ osiągnięcia punktu m jest równe stosunkowi ilości dróg, którymi można dojść do tego punktu, do całkowitej ilości możli-



Rys. 10.3. Przypadkowe błądzenie o krokach równej długości odtwarza trójkąt Pascala dla współczynników dwumianu. Liczby przy punktach oznaczają ilość możliwych dróg prowadzących do danego punktu. Całkowita liczba przypadków jest równa 2^n

wych dróg, która jest oczywiście równa 2^n, ponieważ przy każdym kroku każda z dróg dzieli się na dwie (± 1):

$$\pi_{n,m} = \frac{1}{2^n}\frac{n!}{q!(n-q)!} = \frac{n!}{2^n\left(\frac{n+m}{2}\right)!\left(\frac{n-m}{2}\right)!}. \tag{10.3}$$

Z warunku (10.2) wynika, że parzystość lub nieparzystość m jest określona przez n, tzn. jeśli n jest parzyste, to m również jest parzyste i odwrotnie.

* Początkiem jest punkt O na osi odległości (rys. 10.3), natomiast punktem końcowym jest punkt leżący m jednostek na prawo lub na lewo od punktu O. (*Przyp. tłum.*).

Średnie przesunięcie $\overline{R_n}$ po n krokach jest oczywiście równe zeru, ponieważ szanse ruchu w prawo lub w lewo w każdym kroku są sobie równe, natomiast wartość średnia kwadratu przesunięcia wzrasta ze wzrostem liczby kroków. Z rysunku 10.3 wynika

$$\overline{R_n^2} = \overline{m^2} l^2 = l^2 \sum_{m=-n}^{+n} m^2 \pi_{n,m}; \tag{10.4}$$

$$\left.\begin{aligned} \overline{R_0^2} &= 0, \\ \overline{R_1^2} &= {}^1/_2(l^2+l^2) = l^2, \\ \overline{R_2^2} &= {}^1/_4[(2l)^2+2\cdot 0+(2l)^2] = 2l^2, \\ \overline{R_3^2} &= {}^1/_8[(3l)^2+3l^2+3l^2+(3l)^2] = {}^{24}/_8 l^2 = 3l^2. \end{aligned}\right\} \tag{10.5}$$

Wyniki te sugerują wzór

$$\overline{R_n^2} = nl^2, \tag{10.6}$$

który można udowodnić przez indukcję. W tym celu wystarczy wykazać, że jeśli

$$\overline{m_n^2} = \sum_{m=-n}^{m=n} \pi_{n,m} m^2 = n, \tag{10.7}$$

to wynika stąd

$$\overline{m_{n+1}^2} = \sum_{m=-n-1}^{m=n+1} \pi_{n+1,m} m^2 = n+1. \tag{10.7a}$$

Zauważmy przede wszystkim, że jeśli po $n+1$ krokach osiągamy odległość $m < n+1$, to odległość osiągnięta w poprzednim kroku musiała być równa albo $m+1$, albo $m-1$ i prawdopodobieństwa obu tych możliwości są sobie równe. Stąd

$$\pi_{n+1,m} = {}^1/_2(\pi_{n,m+1}+\pi_{n,m-1}), \qquad m < n+1 \tag{10.8}$$

oraz

$$\pi_{n+1,n+1} = \pi_{n+1,-n-1} = \frac{1}{2^{n+1}}. \tag{10.8a}$$

Oba wskaźniki w $\pi_{n,m}$ są jednocześnie parzyste lub nieparzyste i przy sumowaniu po jednym z nich otrzymujemy tylko co drugi wyraz różny od zera. Korzystając z tego faktu oraz z równań (10.8) i (10.8a) — sumę w równaniu

(10.7a) możemy napisać:

$$\overline{m_{n+1}^2} = 2\frac{(n+1)^2}{2^{n+1}} + \frac{1}{2}\sum_{m=-n+1}^{m=n-1}(\pi_{n,m+1}+\pi_{n,m-1})m^2 =$$

$$= \frac{(n+1)^2}{2^n} + \frac{1}{2}\sum_{m=-n+1}^{m=n-1}\pi_{n,m+1}m^2 + \frac{1}{2}\sum_{m=-n+1}^{m=n-1}\pi_{n,m-1}m^2.$$

W pierwszej sumie podstawiamy $k = m+1$, a w drugiej $l = m-1$. Mamy zatem

$$\overline{m_{n+1}^2} = \frac{(n+1)^2}{2^n} + \frac{1}{2}\sum_{k=-n+2}^{k=n}\pi_{n,k}(k-1)^2 + \frac{1}{2}\sum_{l=-n}^{k=n-2}\pi_{n,l}(l+1)^2 =$$

$$= \frac{(n+1)^2}{2^n} + \frac{1}{2}\sum_{k=-n}^{k=n}\pi_{n,k}(k-1)^2 - \frac{1}{2}\pi_{n,-n}(-n-1)^2 +$$

$$+ \frac{1}{2}\sum_{l=-n}^{l=n}\pi_{n,l}(l+1)^2 - \frac{1}{2}\pi_{n,n}(n+1)^2.$$

Ponieważ $\pi_{n,n} = \pi_{n,-n} = \frac{1}{2^n}$, pozostają tylko dwie sumy, które można połączyć:

$$\overline{m_{n+1}^2} = \frac{1}{2}\sum_{k=-n}^{k=n}\pi_{n,k}[(k-1)^2+(k+1)^2] = \frac{1}{2}\sum_{k=-n}^{k=n}\pi_{n,k}(2k^2+2) =$$

$$= \sum_{k=-n}^{k=n}\pi_{n,k}k^2 + \sum_{k=-n}^{k=n}\pi_{n,k} = n+1, \qquad \text{c.b.d.o. (10.9)}$$

uwzględniając wzór (10.7) oraz fakt, że suma $\pi_{n,k}$ po k jest równa jedności.

Rozkład punktów rozszerza się wraz z upływem czasu i wartość pierwiastka średniego kwadratu przesunięcia dla równych kroków o długości l jest równa

$$(\overline{R^2})^{1/2} = n^{1/2}l. \qquad (10.10)$$

Można wykazać, że jeśli liczba kroków n nieograniczenie rośnie, rozkład (10.3) dąży do rozkładu Gaussa (zob. *Dodatek* na końcu tego rozdziału). Pokażemy dalej, jak można wykorzystać otrzymany wynik przy rozpatrywaniu zagadnienia ruchów Browna oraz przy ocenie błędów popełnianych w trakcie wieloetapowych obliczeń.

W tym celu przyjmijmy graniczny przypadek $|l| = {}^1/_2$; błąd średniokwadratowy wynosi wówczas $\sqrt{\overline{R_n^2}} \approx {}^1/_2 \sqrt{n}$, tzn. przyjmujemy, że ostatnia cyfra znajduje się w przedziale $\pm {}^1/_2$. Z drugiej strony błąd maksymalny będzie proporcjonalny do ${}^1/_2 n$, co odpowiada mało prawdopodobnemu przypadkowi nałożenia się wszystkich błędów.

Możemy również rozważyć przypadek, kiedy poszczególne kroki mają różną długość. Mamy wówczas do czynienia z rozkładem losowym o średniej

$$\overline{R_n^2} = \overline{l^2} n \quad \text{lub} \quad \sqrt{\overline{R_n^2}} = \sqrt{\overline{l^2}}\,\sqrt{n}. \tag{10.11}$$

Wyrażenie to może być bardziej odpowiednie do oceny błędów.

10.3. EFEKT ŚRUTOWY

Pojęciem *błądzenia przypadkowego* możemy posłużyć się przy rozpatrywaniu problemów szumu w *lampach elektronowych* (*efekt śrutowy*) i przemian radioaktywnych.

Podczas przemiany radioaktywnej pierwiastek emituje cząstki ze średnią szybkością N cząstek na sekundę. Jest to zjawisko bez powiązań statystycznych pomiędzy emitowanymi kolejno cząstkami.

Ilość cząstek emitowanych w czasie t można przedstawić wzorem

$$P = Nt + \nu, \tag{10.12}$$

gdzie ν jest poprawką odpowiadającą szumowi.

Opis ten dokładnie odpowiada zjawisku emisji elektronów z gorącej katody i dlatego szum występujący w lampach elektronowych ma ten sam charakter. Na rysunku 10.4 pokazano wykres P w funkcji czasu.

Przede wszystkim zagadnienie emisji musimy sprowadzić do problemu błądzenia przypadkowego. Można tego dokonać wieloma różnymi sposobami, zależnie od aktualnych warunków emisji. Zacznijmy od prostego przypadku odpowiadającego następującym założeniom. Przyjmijmy, że można znaleźć na tyle mały okres czasu θ, że może w nim zajść jedno z dwu równoprawdopodobnych zdarzeń: brak emisji i emisja jednej cząstki. W ten sposób w czasie $t = \theta$ zmiana P wyniesie:

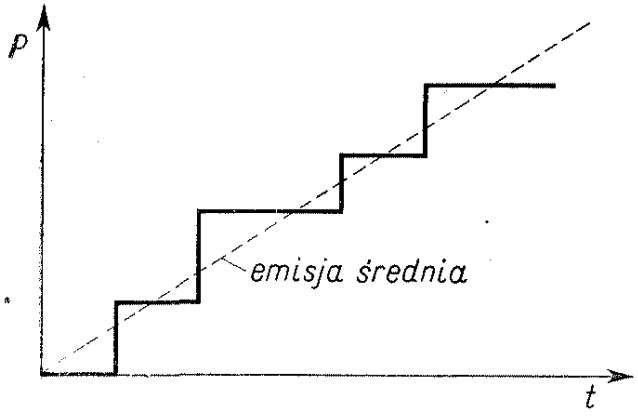
$$\Delta t = \theta, \quad \Delta P = \begin{cases} 0 \\ 1 \end{cases} = {}^1/_2 + \begin{cases} -{}^1/_2 \\ +{}^1/_2 \end{cases}, \tag{10.13}$$

stąd

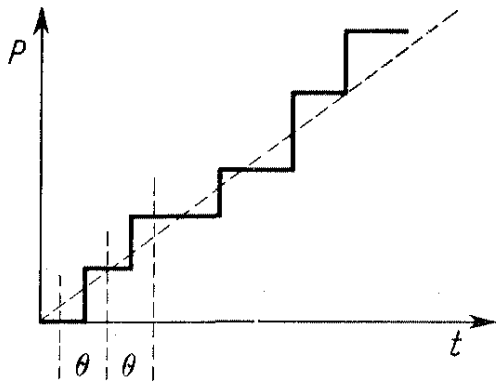
$$N = 1/(2\theta)$$

oraz

$$\Delta\nu = \begin{cases} -1/2 \\ +1/2 \end{cases}.$$



Rys. 10.4. Wykres P w funkcji czasu jest funkcją schodkową. Wykres wartości średniej P jest linią prostą. „Schodki" odpowiadają wyrazowi ν we wzorze (10.12)



Rys. 10.5. Funkcja schodkowa z rys. 10.4 z okresem θ tak wybranym, aby prawdopodobieństwo emisji jednej cząstki w tym okresie wynosiło 1/2

Jeśli więc przyjmiemy $|l| = 1/2$, sprowadzimy zagadnienie emisji do zagadnienia jednostajnego wzrostu $N(t)$, z jednoczesnym błądzeniem przypadkowym. Średnia kwadratu fluktuacji jest nadal równa

$$\overline{\nu_n^2} = nl^2 = (1/4)n, \tag{10.14}$$

gdzie n odpowiada całkowitemu czasowi $t = n\theta$. W czasie t średnia wartość P wyniesie

$$\overline{P} = Nt = n\theta\frac{1}{2\theta} = \left(\frac{1}{2}\right)n,$$

skąd

$$\overline{\nu_n^2} = (1/2)\overline{P}. \tag{10.15}$$

W tym przypadku średnia kwadratu fluktuacji jest równa połowie średniej liczby wyemitowanych cząstek. Przy innych założeniach, odpowiadających *w pełni przypadkowej emisji*, dojdziemy do ważniejszego, bardziej zbliżonego do rzeczywistości zagadnienia.

Wybierzmy bardzo mały odcinek czasu θ i załóżmy, że w tym czasie nastąpił wzrost ilości wyemitowanych cząstek o ΔP

$$\Delta P = \begin{cases} 0 \text{ z prawdopodobieństwem } 1-N\theta, \\ 1 \text{ z prawdopodobieństwem } N\theta \ll 1. \end{cases}$$

Odcinek czasu θ jest tak mały, że emisja dwóch lub więcej cząstek jest nieprawdopodobna. Średnia szybkość emisji wynosi N, poprzedni warunek możemy więc napisać

$$\Delta P = N\theta + \begin{cases} -N\theta & \text{prawdopodobieństwo } (1-N\theta), \\ 1-N\theta & \text{prawdopodobieństwo } N\theta. \end{cases} \tag{10.16}$$

Rozdzieliliśmy tutaj średni wzrost $N\theta$ i fluktuacje, dochodząc w ten sposób do zagadnienia błądzenia przypadkowego

$$\Delta v = \begin{cases} -N\theta & \text{prawdopodobieństwo } (1-N\theta), \\ 1-N\theta & \text{prawdopodobieństwo } N\theta. \end{cases} \tag{10.16a}$$

Ponieważ jest to przypadek nierównych kroków, zastosujemy wyrażenie (10.11).

Średnia kwadratu długości kroku jest

$$\overline{l^2} = \overline{\Delta v^2} = (N\theta)^2(1-N\theta) + (1-N\theta)^2 N\theta = N\theta(1-N\theta),$$

stąd, ponieważ $N\theta$ jest bardzo małe,

$$\overline{l^2} = N\theta.$$

Korzystając z tego wyrażenia i biorąc pod uwagę wyrażenie (10.11) otrzymujemy wyrażenie na średni kwadrat fluktuacji i średnią wartość P za okres t:

$$t = n\theta, \qquad n \text{ duże},$$

$$\overline{P} = Nt = nN\theta,$$

$$\overline{v^2} = n\overline{l^2} = nN\theta = \overline{P}.$$

Średni kwadrat fluktuacji dla emisji w pełni przypadkowej jest dokładnie równy średniej liczbie wyemitowanych cząstek.

10.4. RUCHY BROWNA

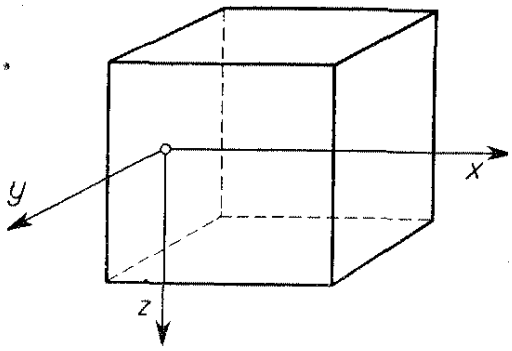
Rozważając ruchy Browna będziemy się wzorować na ujęciu H. A. Lorentza*. Na wstępie podamy kilka stałych:

$$k = 1{,}38 \cdot 10^{-16} \text{ ergów/cząstkę.°K},$$

$$kT = 4{,}14 \cdot 10^{-14} \text{ ergów/cząstkę} \approx 1/40 \text{ eV},$$

$$\text{przy } T = 300°K \approx 27°\text{C}.$$

Rozważmy cząstkę umieszczoną w ośrodku o współrzędnych pokazanych na rys. 10.6. Będziemy rozpatrywać jej ruch wyłącznie w kierunku osi x.



Rys. 10.6. Współrzędne x, y i z cząstki umieszczonej w lepkim ośrodku

Ponieważ ruch cząstki jest przypadkowy, więc jej ruch w kierunku x jest niezależny od ruchu w kierunku y i z. Jeżeli wprowadzimy współczynnik tarcia lepkiego w, to równanie ruchu cząstki o masie m można będzie zapisać

$$F_x = m\frac{dv}{dt} + wv, \qquad (10.17)$$

gdzie F_x oznacza przypadkową siłę działającą na cząstkę. F_x musi być siłą przypadkową, ponieważ w każdym rozsądnie małym przedziale czasu ma miejsce bardzo duża liczba zderzeń pomiędzy cząstką a cząsteczkami cieczy.

Weźmy teraz przedział czasu τ na tyle mały, aby zmiana prędkości cząstki Δv była znikoma, i obliczmy całkowity popęd, jaki działa na naszą cząstkę w tym czasie:

$$X = \int_0^\tau F_x \, dt. \qquad (10.18)$$

Prędkość cząstki dla $t = 0$ jest równa v_0, dla $t = \tau$ zaś wynosi v_1; po scałkowaniu równania (10.17) przy granicach całkowania $0, \tau$ mamy

$$m(v_1 - v_0) + wv_0\tau = X. \qquad (10.19)$$

* H. A. Lorentz, *Les théories statistiques en thermodynamique*, Leipzig 1916, Teubner, str. 49; G. L. de Haas-Lorentz, *Die Brownsche Bewegung*, Braunschweig 1913, Vieweg.

Grupując wyrazy otrzymujemy

$$mv_1 = v_0(m-w\tau)+X. \tag{10.20}$$

Ponieważ siła F_x jest całkowicie przypadkowa, możemy przypuszczać, że

$$\overline{X} = 0, \tag{10.21}$$

musimy jednakże uwzględnić fakt, że $\overline{X^2}$ jest różne od zera. Podnieśmy wzór (10.20) do kwadratu:

$$v_1^2 = v_0^2\left(1 - \frac{2w\tau}{m} + \frac{w^2\tau^2}{m^2}\right) + \frac{X^2}{m^2} + \frac{2Xv_0}{m}\left(1 - \frac{w\tau}{m}\right), \tag{10.22}$$

ponieważ τ jest bardzo małe, mamy

$$\frac{w^2\tau^2}{m^2} \ll \frac{2w\tau}{m}.$$

Jeśli teraz uśrednimy wzór (10.22) i uwzględnimy, że $\overline{X}$ jest równe zeru, wówczas będziemy mogli znaleźć wartość średnią v_1^2:

$$\overline{v_1^2} = \overline{v_0^2}\left(1 - \frac{2w\tau}{m}\right) + \frac{\overline{X^2}}{m^2}. \tag{10.23}$$

Z uwagi na to, że nasz układ (cząstka+ciecz) ma pewną temperaturę T, jest dla niego słuszna zależność wynikająca z teorii kinetycznej

$$^1/_2 m\overline{v^2} = {}^1/_2 kT.$$

Dlatego mamy

$$\overline{v_1^2} = \overline{v_0^2} = \frac{kT}{m}. \tag{10.24}$$

Po podstawieniu wzoru (10.24) do (10.23) otrzymujemy wyrażenie na $\overline{X^2}$:

$$\overline{X^2} = 2w\tau kT. \tag{10.25}$$

Widzimy więc, że średnia kwadratowa impulsu siły jest różna od zera. Sprowadzając wzór (10.23) do postaci

$$\frac{1}{2}m\overline{v_1^2} = \frac{1}{2}m\overline{v_0^2}\left(1 - \frac{2w\tau}{m}\right) + \frac{\overline{X^2}}{2m},$$

możemy zauważyć, że $(^1/_2)m\overline{v_0^2}(2w\tau/m)$ przedstawia stratę energii kinetycznej

cząstki, wywołaną ruchem w lepkim ośrodku, a $\overline{X^2}/2m$ odpowiada energii kompensującej, jaką cząstka uzyskuje na skutek przypadkowych zderzeń z cząsteczkami cieczy. W ten sposób obserwujemy podwójną wymianę energii:

$$\begin{cases} \text{energia kinetyczna cząstki} \rightarrow \text{ruch cząsteczek cieczy.} \\ \text{ruch cząsteczek cieczy} \rightarrow \text{energia kinetyczna cząstki.} \end{cases}$$

Dalej możemy zauważyć, że jeżeli tarcie lepkie jest duże, to energia kompensująca jest także duża. Duże tarcie lepkie oznacza silne sprzężenia między cząstką a otaczającym ją lepkim ośrodkiem, a więc również i silne zderzenia, którym odpowiadają duże przypadkowe impulsy X. Praca wykonywana przez siły przypadkowe ma źródło w energii ruchu cieplnego i na skutek tarcia znowu zamienia się na ciepło.

Aby otrzymać średniokwadratowe przesunięcie cząstki, wprowadzimy pewne uproszczenia. Założymy mianowicie, że dla $t = 0$ jest $v_0 = 0$, dla $t = \tau$ zaś $v = v_1$ oraz że po czasie τ na cząstkę nie działają żadne siły. To ostatnie założenie oznacza, że wydzielamy jedno ze zderzeń i traktujemy je jako zdarzenie odosobnione. Zatem po czasie τ prędkość cząstki maleje zgodnie ze wzorem

$$v = v_1 \, e^{wt/m}. \tag{10.26}$$

Po scałkowaniu obliczymy x:

$$x = \int_\tau^\infty v \, dt = v_1 \int_\tau^\infty e^{-wt/m} \, dt = \frac{v_1 m}{w} \left[e^{-wt/m}\right]_\tau^\infty.$$

Ponieważ τ jest małe, możemy zamiast przedziału τ, ∞ przyjąć przedział $0, \infty$, skąd przy $v_0 = 0$ ze wzoru (10.20) otrzymujemy

$$x = \frac{mv_1}{w} = \frac{X}{w}. \tag{10.27}$$

Wielkość x oznacza przesunięcie cząstki wynikające z impulsu X działającego w czasie τ. Jest to pewien rodzaj błądzenia przypadkowego, ponieważ $\overline{X} = 0$ i dlatego również $\bar{x} = 0$. Jednakże $\overline{X^2} \neq 0$, podnosząc do kwadratu i uśredniając równanie (10.27) mamy

$$\overline{x^2} = \overline{X^2}/w^2. \tag{10.28}$$

Z podstawienia wzoru (10.25) do ostatniego mamy

$$\overline{x^2} = \frac{2kT}{w}\tau. \tag{10.29}$$

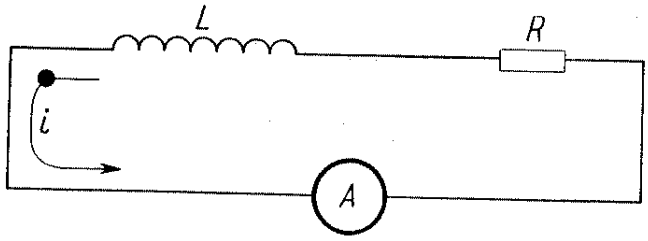
Dla dużego przedziału czasu t, zawierającego wiele małych przedziałów τ, kwadraty przesunięć sumują się i

$$\overline{x^2} = \frac{2kTt}{w}. \tag{10.30}$$

Wynik ten możemy sprawdzić na drodze doświadczalnej, obserwując położenie cząstek koloidalnych w zawiesinie jako funkcję czasu.

10.5. POBUDZENIE CIEPLNE W OBWODZIE ELEKTRYCZNYM

Zajmiemy się obecnie zagadnieniem podobnym do ruchów Browna — przypadkowym przepływem ładunków w obwodzie elektrycznym. Rozpatrzmy obwód elektryczny pokazany na rys. 10.7. Obwód ten składa się z induk-



Rys. 10.7. Obwód rozpatrywany przy analizie szumów cieplnych. Indukcyjność, oporność i amperomierz są połączone szeregowo i utrzymywane w temperaturze T

cyjności, oporu i amperomierza służącego do mierzenia prądu. Wszystkie te elementy mają pewną temperaturę T. Ponieważ obwód ma jeden stopień swobody, więc

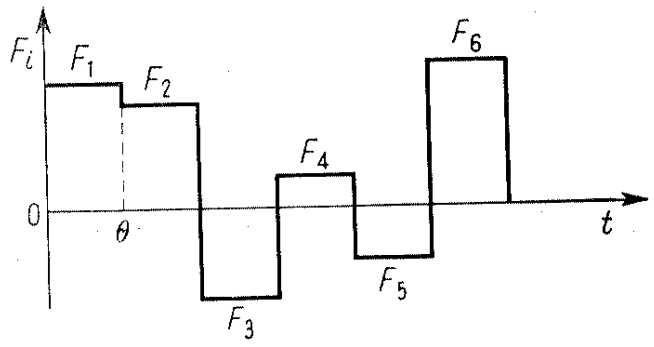
$$\overline{E}_k = {}^1/_2 kT = {}^1/_2 L\,\overline{i^2}. \tag{10.31}$$

Prąd fluktuuje przypadkowo, $\overline{i} = 0$. Jeśli napiszemy równanie przepływu prądu

$$L\frac{\mathrm{d}i}{\mathrm{d}t} + Ri = F, \tag{10.32}$$

gdzie F oznacza przypadkową siłę elektromotoryczną (sem.) w obwodzie, łatwo możemy zauważyć, że równanie to jest identyczne z równaniem ruchu (10.17) cząstki wykonującej ruchy Browna. Ponadto mechanizm wymiany energii jest także podobny.

Elektron oddaje swoją energię kinetyczną cząsteczkom opornika, a *zderze-*



Rys. 10.8. Siła przypadkowa jest funkcją schodkową, stałą w każdym przedziale czasu $(n-1)\theta, n\theta$

nia cząsteczek dostarczają energii elektronowi. Kompensacja tej straty jest tu określona przez średni kwadrat popędu elektromotorycznego* i otrzymujemy związek podobny do (10.25)

$$\overline{X^2} = 2R\tau kT. \tag{10.33}$$

Ładunek q zajmuje miejsce przesunięcia x, wobec czego słuszny jest wzór podobny do (10.30):

$$\overline{q^2} = \frac{2kT}{R}\, t. \tag{10.34}$$

Mamy więc dwa źródła szumu w układach elektronicznych: szum cieplny wywołany pobudzeniem termicznym i fluktuacjami ładunku w takich elementach obwodów jak oporniki i szum wywołany efektem śrutowym w lampach. W następnym rozdziale zajmiemy się widmem częstotliwości szumu cieplnego.

---

* Popędem elektromotorycznym analogicznie do wzoru (10.18) nazywamy tu całkę

$$X = \int_0^{\tau} F(t)\, \mathrm{d}t,$$

gdzie $F(t)$ — siła elektromotoryczna. (*Przyp. tłum.*).

Możemy również skonstruować model bardziej dokładny, uwzględniający wzrost w czasie wartości średniokwadratowego popędu w przypadku ruchów Browna lub popędu elektromotorycznego w obwodzie elektrycznym. Przyjmijmy, że siła zmienia się w czasie w sposób przypadkowy, tak jak to pokazano na rys. 10.8, i że jest ona stała w każdym odcinku czasu θ (θ — bardzo małe, $\tau = n\theta$). Ponieważ siła ma charakter przypadkowy, nie występuje żadna korelacja i

$$\left.\begin{aligned} \overline{F_i F_k} &= 0, \quad i \neq k, \\ \overline{F_i^2} &= \overline{F^2} \neq 0. \end{aligned}\right\} \tag{10.35}$$

Obliczamy

$$X = \int_{t=0}^{t=\tau=n\theta} F\,\mathrm{d}t = \theta(F_1+F_2+\ldots+F_n)$$

oraz

$$X^2 = \theta^2(F_1^2+F_2^2+\ldots+F_n^2+F_n^2+2F_1F_2+2F_1F_3+\ldots). \tag{10.36}$$

Po uśrednieniu otrzymujemy wyrażenie

$$\overline{X^2} = \theta^2(\overline{F_1^2+F_2^2+\ldots}).$$

Ponieważ wartości średnie iloczynów są równe zero (wzór (10.35)), otrzymujemy

$$\overline{X^2} = n\theta^2\overline{F^2}, \tag{10.37}$$

natomiast po podstawieniu $\tau = n\theta$

$$\overline{X^2} = \tau(\overline{F^2\theta}). \tag{10.38}$$

Jak wykazaliśmy, wartość średniokwadratowa popędu lub popędu elektromotorycznego jest proporcjonalna do czasu. Jest to zgodne ze wzorami (10.25) i (10.33) i wyjaśnia mechanizm wzrostu tych wielkości.

DODATEK DO ROZDZIAŁU 10

W rozdziale 10.2 rozważaliśmy zagadnienie błądzenia przypadkowego i obliczyliśmy prawdopodobieństwo $\pi_{n,m}$ osiągnięcia punktu m po n krokach (wzór (10.3)):

$$\pi_{n,m} = \frac{n!}{2^n \frac{n+m}{2}! \frac{n-m}{2}!}.$$

Zastanówmy się obecnie, co się stanie, jeśli liczba kroków będzie bardzo duża, podczas gdy m pozostanie ograniczone. Korzystając ze wzoru Stirlinga po prostych przekształceniach otrzymamy

$$n! \to \left(\frac{n}{e}\right)^n \sqrt{2\pi n}, \qquad n \gg 1,$$

$$\pi_{n,m} \to \frac{n^n}{(n+m)^{(n+m)/2}(n-m)^{(n-m)/2}} \frac{\sqrt{2\pi n}}{2\pi \sqrt{\frac{n+m}{2}\frac{n-m}{2}}} =$$

$$= \frac{2\sqrt{n}}{\sqrt{2\pi(n^2-m^2)}} \frac{1}{\left(1+\frac{m}{n}\right)^{(n+m)/2}\left(1-\frac{m}{n}\right)^{(n-m)/2}}.$$

Wyrazy w mianowniku zachowują się jak $(1+\varepsilon)^{1/\varepsilon}$, tzn. dla bardzo małych ε dążą do e. Spróbujemy jednak znaleźć lepsze przybliżenie. Skorzystamy z rozwinięcia

$$\ln(1+\varepsilon)^{1/\varepsilon} = \frac{1}{\varepsilon}\ln(1+\varepsilon) = \frac{1}{\varepsilon}\left(\varepsilon - \frac{\varepsilon^2}{2} + \ldots\right) = 1 - \frac{\varepsilon}{2} + \ldots,$$

skąd

$$(1+\varepsilon)^{1/\varepsilon} \to e^{1-\varepsilon/2};$$

daje to przy $\varepsilon = m/n$,

$$\left[\left(1+\frac{m}{n}\right)^{\frac{n}{m}}\right]^{\frac{m}{2}+\frac{m^2}{2n}} \to \left[e^{1-\frac{m}{2n}}\right]^{\frac{m}{2}+\frac{m^2}{2n}} = e^{\frac{m}{2}+\frac{m^2}{4n}+\ldots};$$

podobnie

$$\left(1-\frac{m}{n}\right)^{\frac{n-m}{2}} \to e^{-\frac{m}{2}+\frac{m^2}{4n}}$$

i ostatecznie mamy

$$\pi_{n,m} \to \frac{2\sqrt{n}}{\sqrt{2\pi(n^2-m^2)}}\, e^{-m^2/2n} \to \frac{2}{\sqrt{2\pi n}}\, e^{-m^2/2n},$$

co odpowiada rozkładowi prawdopodobieństwa typu gaussowskiego. Musimy pamiętać, że m i n są równocześnie parzyste lub nieparzyste. Dla danego n odległość pomiędzy kolejnymi punktami wynosi $\Delta m = 2$ (rys. 10.3). Dla przykładu weźmy przypadek, kiedy n i m są parzyste

$$n = 2h, \quad m = 2y, \quad \Delta y = 1, \quad h \gg y.$$

Możemy pominąć wyrazy z m^2 w mianowniku $\pi_{n,m}$ i napisać

$$\pi_{h,y} = \frac{1}{\sqrt{\pi h}}\, e^{-y^2/h}.$$

Porównując ten wynik ze standardowym zapisem rozkładu normalnego

$$P = \frac{1}{\sqrt{2\pi}\,\sigma}\, e^{-y^2/2\sigma^2}, \quad \overline{y^2} = \sigma^2,$$

gdzie σ oznacza odchylenie standardowe, otrzymujemy

$$\sigma^2 = \frac{h}{2}, \quad \overline{m^2} = 4\overline{y^2} = 2h = n.$$

Odpowiada to ogólnemu związkowi (10.7).

11

SZUM CIEPLNY W OBWODACH ELEKTRYCZNYCH. WZÓR NYQUISTA

11.1. MODEL Z IMPULSAMI PRZYPADKOWYMI

Niektóre wyniki, uzyskane w poprzednim rozdziale, wykorzystamy teraz do wyprowadzenia wzoru Nyquista, określającego rozkład częstotliwości widma szumu wynikającego z przypadkowej, termicznej sem*. (11.1) Przypomnijmy (wynika to ze wzorów (10.33) i (10.38), że wartość średniokwadratowa popędu elektromotorycznego $X = \int_0^\tau F\,dt$, obliczonego za okres czasu τ, jest równa

$$\overline{X^2} = 2RkT\tau, \tag{11.1}$$

$$\overline{X^2} = \tau(\theta\overline{F^2}). \tag{11.2}$$

Wyrażenie (11.2) otrzymaliśmy przyjmując model przypadkowych impulsów sem., stałych w małym przedziale czasu θ takim, że

$$\tau = n\theta.$$

Łącząc wzory (11.1) i (11.2) otrzymujemy

$$\overline{F^2}\theta = 2RkT. \tag{11.3}$$

Mimo że w naszym modelu mamy do czynienia z przypadkowym ciągiem impulsów prostokątnych, widmo mocy tego ciągu będzie się pokrywać z wid-

* J. L. Lawson, G. E. Uhlenbeck, *Threshold Signals*, New York 1950, Radiation Laboratory, Series 24, Mc Graw-Hill.

mem pojedynczego impulsu prostokątnego, ponieważ przy impulsach pojawiających się przypadkowo (bez korelacji) sumuje się kwadraty amplitud składowych harmonicznych.

Widmo pojedynczego impulsu sem. o wysokości F oraz czasie trwania θ jest określone przez wzory (8.23) i (8.25):

$$f(t) = \int_{\nu=-\infty}^{+\infty} C(\nu) \cos 2\pi\nu t \, d\nu = \int_{\nu=0}^{+\infty} 2C(\nu) \cos 2\pi\nu t \, d\nu, \tag{11.4}$$

gdzie

$$C(\nu) = F\theta\left(\frac{\sin \pi\nu\theta}{\pi\nu\theta}\right), \tag{11.5}$$

θ jest czasem trwania impulsu, natomiast nominalna szerokość pasma wynosi $1/\theta$ (rozdz. 8.4):

$$\Delta\nu = 2\nu_M = \frac{1}{\theta}, \qquad -\nu_M \leqslant \nu \leqslant +\nu_M, \qquad \nu_M = \frac{1}{2\theta}. \tag{11.6}$$

Z dużym przybliżeniem za widmo częstotliwości impulsu prostokątnego o wysokości F i czasie trwania θ można uważać prostokąt o podstawie $-\nu_M$, $+\nu_M$ i wysokości

$$C(\nu) = F\theta. \tag{11.7}$$

Jeśli nie chcemy wprowadzać częstotliwości ujemnych, możemy wziąć:

$$0 \leqslant \nu \leqslant \nu_M, \qquad \nu_M = \frac{1}{2\theta}, \tag{11.8}$$

i podwoić amplitudę widma (jak we wzorze (11.4)):

$$C'(\nu) = 2C(\nu) = 2F\theta. \tag{11.9}$$

Rozważymy obecnie ciąg impulsów sem. o czasie trwania θ i wartości średniokwadratowej $\overline{F^2}$. Wartość średniokwadratowa sem. $\overline{E}$, zgodnie ze wzorem (11.3) jest równa

$$\overline{E^2} = \overline{F^2} = 2RkT\frac{1}{\theta}. \tag{11.10}$$

Pokazaliśmy również, że wartość ta jest w przybliżeniu stała w pasmie $0 \leqslant \nu \leqslant \nu_M$. Mamy więc

$$\overline{E_\nu^2} = 4RkT\nu_M \tag{11.11}$$

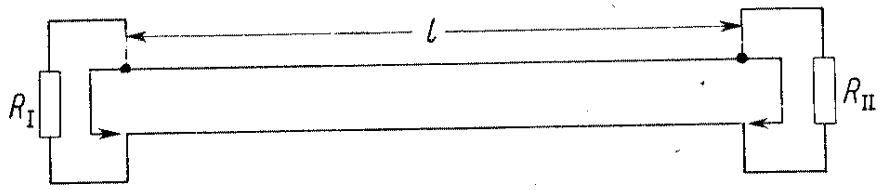
przy prawie stałej amplitudzie w pasmie od 0 do ν_M. Dla mniejszego przedziału częstotliwości $\Delta\nu$, tzn. od ν do $\nu+\Delta\nu$, wartość tę musimy zredukować do wartości:

$$\overline{e^2} = 4RkT\,\Delta\,\nu. \tag{11.12}$$

Jest to właśnie wzór Nyquista*.

11.2. METODA NYQUISTA

Zajmiemy się obecnie metodą wyprowadzenia wzoru (11.12) podaną przez Nyquista. Na wstępie skonstruujmy obwód umożliwiający przesyłanie sygnałów o szerokim pasmie częstotliwości, taki jak np. linia długa (rys. 11.1). Niech nasza linia ma długość l i impedancję charakterystyczną Z. Zakończmy ją opornikami R_{I} i R_{II}, spełniającymi warunek $R_{\mathrm{I}} = R_{\mathrm{II}} = Z$. Cały obwód będziemy utrzymywać w temperaturze T. Każdy z oporników jest źródłem



Rys. 11.1. Linia długa o długości l zakończona opornikami R_{I} i R_{II} oraz przełącznikami do zwierania oporników

termicznej sem., która jest przekazywana od R_{I} i absorbowana przez R_{II} oraz odwrotnie. Przesyłane sygnały zajmują szerokie pasmo częstotliwości.

Jeżeli zostanie osiągnięta równowaga termiczna, zwieramy oba końce linii i ponieważ równowaga cieplna musi się utrzymywać, sygnały biegnące

---

* Niektóre prace źródłowe:

P. W. Bridgman, *Note on the Principle of Detailed Balancing*, Phys. Rev. **31** (1928), str. 101.

J. B. Johnson, *ibidem*, **29** (1927), str. 367; **32** (1928), str. 97.

H. Nyquist, *ibidem*, **32** (1928), str. 110.

J. Bernamont, *Fluctuations de potential aux bornes d'un conducteur metallique*, Ann. phys. [**11**], **7** (1937).

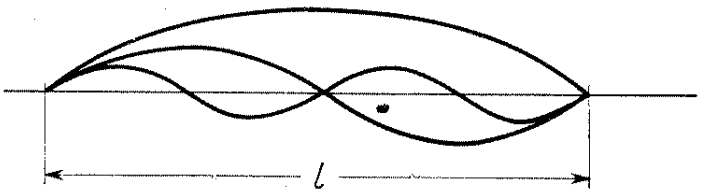
M. Courtines, *Les fluctuations dans les appareils de mesures*, en: *Congr. intern, d'électricité*, vol. **2**, Paris 1932.

w linii nie będą dłużej absorbowane (ani też nie będą emitowane nowe), lecz po prostu będą ulegać odbiciu od końców linii. W warunkach równowagi sygnały tworzą w zwartej na końcu linii fale stojące z węzłami na końcach linii (rys. 11.2). Są to fale o częstotliwościach własnych określonych wzorem

$$\frac{n\lambda}{2} = l, \tag{11.13}$$

gdzie każdemu n odpowiada jedna częstotliwość własna, tzn. jeden stopień swobody linii.

Ponieważ dla każdej częstotliwości własnej linię długą możemy traktować jako oscylator harmoniczny, energia średnia dla każdej częstotliwości własnej



Rys. 11.2. Fale stojące w linii przy zwartych opornikach

(stopnia swobody) musi być sumą energii potencjalnej i kinetycznej. Ze względu zaś na fakt, że każda z nich jest równa $(1/2)kT$, na stopień swobody przypada średnia energia kT.

Aby obliczyć energię w linii dla przedziału częstotliwości $\Delta\nu$, musimy znać liczbę stopni swobody Δn odpowiadającą temu przedziałowi. Ze wzoru (11.13) mamy

$$n = \frac{2l}{\lambda} = 2l\frac{\nu}{c}, \tag{11.14}$$

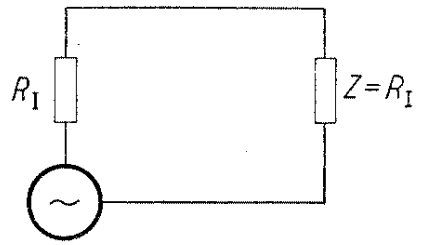
ponieważ $\lambda = c/\nu$, gdzie c jest prędkością fali. Stąd

$$\Delta n = \frac{2l}{c}\Delta\nu \tag{11.15}$$

jest liczbą stopni swobody w przedziale częstotliwości $\Delta\nu$. Całkowita energia średnia w zwartej na końcach linii w przedziale $\Delta\nu$ wynosi

$$\overline{E}_c = \Delta nkT = \frac{2l}{c}\Delta\nu kT\ . \tag{11.16}$$

Energia ta została dostarczona na początku przez dwa jednakowe oporniki R_{I} i R_{II}. Stąd całkowita energia w pasmie częstotliwości $\Delta\nu$, dostarczona przez każdy z oporników, wynosi $(l/c)kT\,\Delta\nu$.



Rys. 11.3. Opornik R_{I} i impedancja charakterystyczna linii $Z = R_{\mathrm{I}}$ zastępująca tę linię

Aby otrzymać wartość sem. generowanej w każdym oporniku, obliczmy średnią moc P, tzn. energię na jednostkę czasu

$$\overline{P} = \frac{(l/c)kT\,\Delta\nu}{(l/c)},$$

ponieważ l/c jest czasem przejścia sygnału od R_{I} do R_{II}. Średnia moc generowana w każdym oporniku i wysyłana w linię jest więc równa:

$$\overline{P} = kT\,\Delta\nu. \tag{11.17}$$

Jednakże każdy opornik jako źródło termicznej siły elektromotorycznej pracuje na obciążenie $2R$ (rys. 11.3) — oporność R *plus* impedancja charakterystyczna $Z = R$ linii. Mamy więc

$$e_\nu = 2Ri_\nu$$

i średnia moc wydzielana w Z wynosi

$$\overline{P} = R\overline{i_\nu^2} = \frac{\overline{e_\nu^2}}{4R}; \tag{11.18}$$

skąd, zgodnie ze wzorem (11.17)

$$\frac{\overline{e_\nu^2}}{4R} = kT\,\Delta\nu \tag{11.19}$$

lub

$$\overline{e^2} = 4kTR\,\Delta\nu. \tag{11.20}$$

Jest to wyrażenie otrzymane już poprzednio.

11.3. WNIOSKI I PRZYKŁADY ZASTOSOWANIA

Dla skontrolowania wyprowadzonego wzoru zastosujemy go do szeregowego obwodu L,R,C i obliczymy wartość średniokwadratową prądu wywołanego termiczną sem. Obwód ten jest pokazany na rys. 11.4. Ponieważ ter-

miczna sem. jest określona wzorem (11.20) lub (11.22), mamy

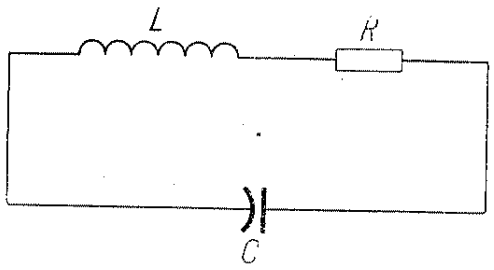
$$\overline{i_\nu^2} = \frac{\overline{e_\nu^2}}{|Z|^2} = \frac{\overline{e_\nu^2}}{R^2 + \left(L\omega - \frac{1}{\omega C}\right)}, \qquad \omega = 2\pi\nu. \tag{11.21}$$

Jeżeli podstawimy

$$\omega_0 = 2\pi\nu_0 = 1/\sqrt{LC}, \tag{11.22}$$

i skorzystamy ze wzoru (11.20), wyrażenie (11.21) przyjmie postać

$$\overline{i_\nu^2} = \frac{4RkT\,\Delta\nu}{R^2 + L^2\left(\omega - \frac{\omega_0}{\omega}\right)^2}. \tag{11.23}$$



Rys. 11.4. Oporność R, indukcyjność L oraz pojemność C połączone szeregowo tworzą oscylator harmoniczny. Szum jest generowany w oporności

Wprowadzając pojęcie dobroci obwodu

$$Q_0 = \frac{\omega_0 L}{R} \tag{11.24}$$

otrzymamy

$$\overline{i_\nu^2} = \frac{2}{\pi}\frac{kT}{L}\frac{Q_0}{\omega_0}\frac{\Delta\omega}{1 + \frac{Q_0^2}{\omega_0^2}\left(\omega - \frac{\omega_0^2}{\omega}\right)^2}. \tag{11.25}$$

Energia kinetyczna dla częstotliwości ν jest równa

$$\overline{E}_{k\nu} = {}^1/_2 L\overline{i_\nu^2}. \tag{11.26}$$

Całkowitą energię kinetyczną otrzymamy sumując $\overline{E}_k$ po wszystkich częstotliwościach od $\nu = 0$ do $\nu = \infty$; sumę tę możemy zastąpić całką

$$\overline{E}_k = \sum \overline{E_{k\nu}} = \sum \frac{1}{2} L\overline{i_\nu^2} = \frac{kT}{\pi}\frac{Q_0}{\omega_0}\int_0^\infty \frac{d\omega}{1 + \frac{Q_0^2}{\omega_0^2}\left(\omega - \frac{\omega_0^2}{\omega}\right)^2}. \tag{11.27}$$

Nietrudno jest obliczyć tę całkę i pokazać, że*

$$\frac{Q_0}{\omega_0}\int_0^\infty \frac{d\omega}{1+\frac{Q_0^2}{\omega_0^2}\left(\omega-\frac{\omega_0^2}{\omega}\right)^2}=\frac{\pi}{2}, \tag{11.28}$$

skąd całkowita (średnia) energia kinetyczna jest równa

$$\overline{E_k}={}^1/_2 kT. \tag{11.29}$$

Istnieje również energia potencjalna, odpowiadająca ładunkom kondensatora, której wartość średnia jest równa ${}^1/_2 kT$, a więc ostatecznie całkowita energia średnia wynosi

$$\overline{E}=\overline{E_k}+\overline{E_p}=kT, \tag{11.30}$$

co jest dobrze znanym wynikiem dla niskoczęstotliwościowego oscylatora harmonicznego.

11.4. UOGÓLNIENIE WZORU NYQUISTA

Wzór Nyquista wyprowadziliśmy dla przypadku obwodu elektrycznego z opornością R. Szum jest generowany tylko w oporności, gdzie istnieje sprzężenie pomiędzy prądem elektrycznym i ruchem cieplnym cząsteczek w materiale przewodzącym. Nie ma takiego sprzężenia w indukcyjności ani w pojemności i w związku z tym nie wprowadzają one żadnych dodatkowych szumów.

---

* Podstawiając $\omega=\omega_0 s$ i rozbijając przedział całkowania na dwa: od 0 do 1 i od 1 do ∞ mamy

$$Q_0\int_0^1 \frac{ds}{1+Q_0^2(s-1/s)^2}+Q_0\int_1^\infty \frac{ds}{1+Q_0^2(s-1/s)^2}=Q_0\int_1^\infty \frac{ds/s^2+ds}{1+Q_0^2(s-1/s)^2}=$$

$$=Q_0\int_0^\infty \frac{dx}{1+Q^2x^2}=\frac{\pi}{2},$$

gdzie $x=s-1/s$; $x=0$ dla $s=1$ i $x=\infty$ dla $s=\infty$.

W impedancji

$$Z = R + iX \tag{11.31}$$

tylko składowa R wprowadza szumy.

Zamiast obwodu elektrycznego możemy rozważać układ mechaniczny z tarciem lepkim. Korzystaliśmy już z analogii drgań mechanicznych i elektrycznych w rozdz. 10.5; możemy z niego skorzystać również w przypadku wzoru Nyquista. Podsumowując wyniki rozdz. 10 (równania (10.18), (10.25), (10.30), (10.33) oraz (10.34)) i rozdziału omawianego obecnie, możemy podać ogólne wyrażenie na szum w dowolnym układzie z tłumieniem R:

$$\overline{X^2} = 2RkTt, \tag{11.32}$$

gdzie X jest popędem w czasie t;

$$\overline{x^2} = \frac{2kT}{R}\, t, \tag{11.33}$$

gdzie x jest przesunięciem za okres t w przypadku swobodnych ruchów Browna;

$$\overline{E_k} = {}^1/_2 kT \tag{11.34}$$

jest energią kinetyczną przypadającą na stopień swobody; wreszcie

$$\overline{f^2} = 4RkT\nu \tag{11.35}$$

jest wartością średniokwadratową sił z częstotliwościami z przedziału pomiędzy ν oraz $\nu + \Delta\nu$.

Wszystkie te wyrażenia mają szerokie zastosowanie przy niskich częstotliwościach. De Haas-Lorentz rozpatrzyła szereg przykładów zastosowania wzorów (11.32), (11.33) i (11.34)*. Uogólnienie wzoru (11.35) na szereg zagadnień nieelektrycznych dokładnie rozważał Callen.** Wykazał on całkowitą uniwersalność tego wyrażenia. Rozważał on także zagadnienia z różnego

---

* H. A. Lorentz, *Les théories statistiques en thermodynamique*, Leipzig 1916, Teubner.

G. L. de Haas-Lorentz, *Die Brownsche Bewegung*, Braunschweig 1931, Vieweg.

** H. B. Callen, Ph. D. thesis, M.I.T., 1948.

R. F. Greene i H. B. Callen, Phys. Rev. **83** (1951), str. 1231.

H. B. Callen i T. A. Welton, *ibidem* **83** (1951), str. 34.

H. B. Callen i R. F. Greene, *ibidem* **86** (1952), str. 702.

typu warunkami ograniczającymi. Uzyskane wyniki są ważnym wynikiem zastosowania termodynamiki do analizy procesów nieodwracalnych.*

Sytuacja zmienia się w sposób zasadniczy, kiedy mamy do czynienia z wysokimi częstotliwościami. Przede wszystkim musimy zdać sobie sprawę z wpływu pojawiających się w takim przypadku wielu nowych stopni swobody. Na przykład w obwodzie elektrycznym wystąpią drgania o wyższych częstotliwościach, określonych przez parametry rozłożone (indukcyjność i pojemność), wzbudzone fluktuacjami cieplnymi, w układzie mechanicznym będą natomiast wzbudzane różne rodzaje drgań poprzecznych. Istnieją prócz tego ograniczenia kwantowe. Nasza poprzednia teoria może więc być stosowana tylko dla niskich częstotliwości, kiedy spełniony jest warunek

$$h\nu \ll kT, \tag{11.36}$$

gdzie h jest stałą Plancka. Jak wiadomo, dla wyższych częstotliwości wyrażenie (11.34) nie jest spełnione i średnia energia oscylatora ν jest wówczas równa

$$\overline{E}(\nu) = \frac{h\nu}{e^{h/kT} - 1} \tag{11.37}$$

zamiast kT. Inne wyrażenia powinny być również odpowiednio skorygowane.

Podłoże fizyczne omawianych zjawisk jest jasne: opór lepki oznacza sprzężenie pomiędzy pewnym ruchem a ruchami termicznymi. Sprzężenie to powoduje tłumienie ruchu i zamianę energii o wyższym stopniu jakości (jaką jest energia mechaniczna) na niskojakościowe ciepło. Ta wymiana energii nie może jednak przebiegać wiecznie, ponieważ doprowadziłoby to układ do stanu zupełnego spoczynku, wiemy zaś, że w temperaturze T musi istnieć w układzie pewien rodzaj ruchów Browna. Dlatego ten sam mechanizm sprzężenia, który wywołuje tłumienie, musi także tworzyć przypadkowe siły, działające w układzie i utrzymujące go w ruchu cieplnym. Rozumowanie to było podstawą przy wyprowadzaniu wyrażenia (11.32) lub (11.35).

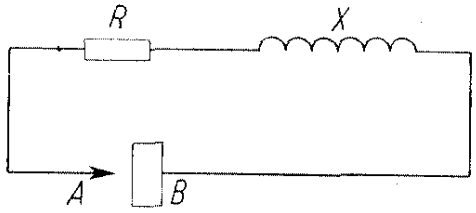
Wyjaśnienie ruchów Browna cząstki w wodzie jest bardzo proste: zderzenia cząstki z cząsteczkami wody są przyczyną tarcia lepkiego. Przypadkowe zde-

* L. Onsager, Phys. Rev. **37** (1931), str. 405; **38** (1931), str. 2265.
H. B. G. Casimir, Revs. Modern Phys. **17** (1945), str. 343.
S. R. De Groot, *Thermodynamics of Irreversible Processes*, Amsterdam 1951, North Holland.

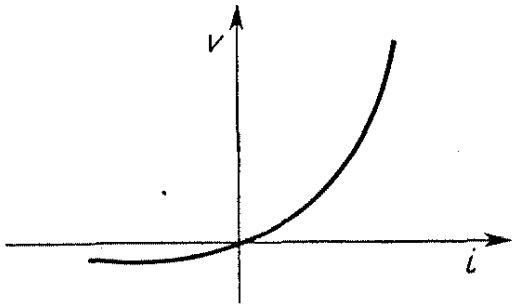
rzenia cząsteczek wody z cząstką wywołują przypadkowe siły utrzymujące ją w ruchu cieplnym. We wszystkich innych możliwych przypadkach sytuacja jest analogiczna.

11.5. POBUDZENIE CIEPLNE W OBWODZIE Z PROSTOWNIKIEM*

Przyjmijmy, że prostownik połączono szeregowo z opornością R i reaktancją X, jak na rys. 11.5. Oporność R (a także oporność prostownika) będzie wytwarzać w obwodzie przypadkową sem. wywołaną pobudzeniem cieplnym. Prostownik nie może prostować prądu wywołanego przypadkową sem.,



Rys. 11.5. Oporność R, reaktancja X oraz prostownik AB połączone szeregowo. Szum jest generowany w oporności. Pojawia się wyprostowane napięcie, lecz brak jest wyprostowanego prądu



Rys. 11.6. Charakterystyka napięciowo-prądowa prostownika

ponieważ wyprostowany prąd mógłby wykonywać pracę, co jest niezgodne z drugą zasadą termodynamiki.

Na rysunku 11.6 pokazano charakterystykę prostownika. Ponieważ krzywa ta rośnie szybciej po prawej stronie osi i niż po lewej, możemy wnioskować, że fluktuacje *w lewo* muszą być większe niż *w prawo*. W przeciwnym przypadku średnia wartość prądu nie równałaby się zeru i popłynąłby wyprostowany prąd. Może jednak powstać wyprostowane napięcie i jego wartość chcemy właśnie obliczyć. W tym celu rozwińmy przypadkową sem. (generowaną w danej chwili w R oraz w oporności prostownika r) w zespolony szereg

* L. Brillouin, Phys. Rev. **78** (1950), str. 627.

Fouriera:

$$\varphi = \varphi_0 + \sum_\nu \varphi_\nu. \tag{11.38}$$

Ponieważ φ_ν są wykładniczymi funkcjami zespolonymi, mamy

$$\overline{\varphi_\nu} = 0 \tag{11.39}$$

dla wszystkich ν, oraz

$$\overline{\varphi} = \varphi_0. \tag{11.40}$$

Jeżeli usunąć prostownik z obwodu, wówczas $\varphi_0 = 0$. Chcemy obliczyć wartość φ_0, tj. średnią stałą wartość sem. wywołaną obecnością prostownika.

Termiczna sem. wywołuje przypadkowy prąd, który także można rozwinąć w szereg Fouriera:

$$i = \sum_\nu i_\nu. \tag{11.41}$$

W rozwinięciu brakuje wyrazu i_0, ponieważ w obwodzie nie płynie wyprostowany prąd. Jak poprzednio, wyrazy i_ν są funkcjami wykładniczymi zespolonymi i ich wartości średnie są równe zeru:

$$\bar{i}_\nu = 0 \text{ dla wszystkich } \nu, \quad \text{skąd} \quad \bar{i} = 0. \tag{11.42}$$

Spadek potencjału na prostowniku wynosi

$$\begin{aligned} V_A - V_B = f &= ri + bii^* = \\ &= r\sum_\nu i_\nu + b\Big[\sum_\nu i_\nu i_\nu^* + \sum_\nu \sum_{\mu \neq \nu} i_\nu i_\mu^*\Big], \end{aligned} \tag{11.43}$$

ponieważ przypadkowe prądy są tak małe, że prostownik pracuje na początkowej części charakterystyki, którą możemy aproksymować parabolą. Uśredniając wzór (11.43) mamy

$$\bar{f} = b\sum_\nu i_\nu i_\nu^*, \tag{11.44}$$

ponieważ iloczyny ii^* są zespolonymi funkcjami wykładniczymi o wykładnikach różnych od zera.

Wykorzystując teraz równanie znane z teorii obwodów, mamy

$$\varphi = \varphi_0 + \sum_\nu \varphi_\nu = R\sum_\nu i_\nu + j\sum_\nu X_\nu i_\nu + f, \tag{11.45}$$

gdzie $j = \sqrt{-1}$. Po uśrednieniu otrzymamy

$$\overline{\varphi} = \varphi_0 = \overline{f} = b \sum_{\nu} i_\nu i_\nu^*. \tag{11.46}$$

Aby znaleźć wyrażenie na φ_0, ponownie korzystamy ze wzoru (11.45) i porównujemy wyrazy o tych samych wskaźnikach:

$$\varphi_\nu = (R+r+jX_\nu)i_\nu. \tag{11.47}$$

Wyraz ri_ν zależy od f (wzór (11.43)). Iloczyny $i_\nu i_\mu^*$ $(\mu \neq \nu)$, także zależne od f, odrzucimy, ponieważ są one proporcjonalne do kwadratów małych prądów i przez to są bardzo małe w porównaniu z i_ν. Ze wzoru (11.47) otrzymujemy

$$\overline{i_\nu i_\nu^*} = \frac{\overline{\varphi_\nu \varphi_\nu^*}}{(R+r)^2+X_\nu^2} \tag{11.48}$$

Aby obliczyć $\overline{\varphi_\nu \varphi_\nu^*}$, możemy skorzystać ze wzoru (11.20), ponieważ tylko oporności w obwodzie generują przypadkową sem.:

$$\overline{\varphi_\nu \varphi_\nu^*} = 4(R+r)kT\Delta\nu. \tag{11.49}$$

Ostatecznie wyprostowana przez prostownik sem. wynosi

$$\varphi_0 = 4kT(R+r) \sum_{\nu} \frac{\Delta\nu}{(R+r)^2+X_\nu^2}. \tag{11.50}$$

Jeśli sumę w tym wyrażeniu zastąpimy przez całkę i przyjmiemy indukcyjny charakter reaktancji, tzn. $X = 2\pi L\nu$, łatwo otrzymamy

$$\varphi_0 = bkT/L. \tag{11.51}$$

Przypomnijmy na zakończenie, że w obwodzie nie ma wyprostowanego prądu, natomiast wyprostowane napięcie jest równe spadkowi potencjału na prostowniku, tak że praca nie może być wykonywana. Zastosowanie pokazanej metody w szczególnych przypadkach zawsze jest związane z taką całkowitą kompensacją stałego napięcia średniego. Jest ona przykładem bardziej ogólnej metody składnikowego równoważenia, rozważanej przez Bridgmana*.

* P. W. Bridgman, Phys. Rev. **31** (1928), str. 101.

12

ZASADA NEGENTROPII

12.1. ZWIĄZEK POMIĘDZY INFORMACJĄ A ENTROPIĄ

Definicję ilości informacji opartą na rozważaniach statystycznych podaliśmy w rozdziale pierwszym. Rozpatrywaliśmy tam sytuację, w której mamy pierwotnie P_0 stanów lub możliwości o równych prawdopodobieństwach a priori. W celu zmniejszenia liczby wchodzących w grę stanów czy możliwości do P_1 wymagana jest ilość informacji I_1, którą określa logarytm stosunku P_0/P_1 (wzór (1.6)):

$$\left.\begin{array}{lll} \text{sytuacja początkowa:} & I_0 = 0, & P_0 \text{ możliwości,} \\ \text{sytuacja końcowa:} & I_1 = 0, & P_1 \text{ możliwości,} \end{array}\right\} \tag{12.1}$$

gdzie

$$I_1 = K \ln(P_0/P_1).$$

W jednostkach termodynamiki $K = k =$ stała Boltzmanna.

Podzielimy obecnie informację na dwie klasy:

(1) Informacja swobodna, która pojawia się w sytuacji, kiedy możliwości uważamy za abstrakcyjne i nie przypisujemy im żadnego znaczenia fizycznego. Ilość informacji swobodnej będziemy oznaczali symbolem I_f.

(2) Informacja związana wystąpi w sytuacji, kiedy możliwości mogą być interpretowane jako mikrostany pewnego układu fizycznego. Ilość informacji związanej będziemy oznaczali symbolem I_b.

Podział ten wprowadziliśmy z następujących względów: Chcemy obecnie rozważyć związek pomiędzy ilością informacji i entropią (a także jej ujemną wartością — negentropią), przy czym terminu „entropia" będziemy używać

tylko w zwykłym, termodynamicznym znaczeniu. Dlatego przyjmiemy dalej, że tylko ilość informacji dotyczącej zjawisk fizycznych, tzn. ilość informacji związanej może być powiązana pewnymi zależnościami z entropią.

Aby wyprowadzić zależność pomiędzy ilością informacji związanej i entropią, przyjmiemy, że równoprawdopodobne możliwości odpowiadają mikrostanom Plancka pewnego układu fizycznego (np. zbiorowiska cząstek). Mamy zatem

Sytuacja	Ilość informacji związanej	Liczba mikrostanów	Entropia
początkowa	$I_{b0}=0$	P_0	$S_0=k\ln P_0$
końcowa	$I_{b1}\neq 0$	$P_1<P_0$	$S_1=k\ln P_1$

(12.2)

Jest sprawą oczywistą, że układ fizyczny w przyjętym schemacie nie jest izolowany. Dostarczana informacja zmniejsza liczbę mikrostanów i entropia układu maleje. Informacja ta musi być dostarczona przez jakiś czynnik zewnętrzny, którego entropia rośnie. Związek pomiędzy ubytkiem entropii naszego układu i wymaganą do tego celu ilością informacji, po uwzględnieniu wzoru (12.1), jest oczywisty:

$$I_{b1}=k(\ln P_0-\ln P_1)=S_0-S_1 \tag{12.3}$$

lub

$$S_1=S_0-I_{b1}. \tag{12.3a}$$

Ilość informacji związanej pojawia się jako ujemny wyraz w całkowitej entropii układu fizycznego; stąd możemy powiedzieć, że

$$\begin{aligned} &\textit{ilość informacji związanej} = \textit{ubytek entropii } S = \\ &= \textit{wzrost negentropii } N, \end{aligned} \tag{12.4}$$

gdzie negentropię N definiujemy jako entropię ze znakiem ujemnym. Stwierdzenie to nazwiemy *negentropijną zasadą informacji*. Dalej rozważymy szereg przykładów ilustrujących tę zależność i wykażemy, w jaki sposób informacja może przechodzić w negentropię i odwrotnie.

W przypadku informacji swobodnej nie należy mówić o zależności pomiędzy ilością informacji i entropią, ponieważ związek pomiędzy entropią i liczbą możliwości określony jest jedynie w sytuacji, kiedy możliwości są wchodzącymi w grę mikrostanami układu fizycznego.

12.2. NEGENTROPIJNA ZASADA INFORMACJI I UOGÓLNIENIE ZASADY CARNOTA

W sytuacji opisanej przez wzór (12.2) entropia zmalała na skutek oddziaływania czynnika zewnętrznego. Dla układu izolowanego zasada Carnota mówi, że przy naturalnej ewolucji układu zachodzi

$$\Delta S_1 \geqslant 0 \quad \text{lub} \quad \Delta(S_0 - I_{b1}) \geqslant 0. \tag{12.5}$$

Wzrost entropii może być więc spowodowany wzrostem S_0 albo $-I_{b1}$ lub też obu tych wielkości jednocześnie. Jeśli układ izolowany pozostawimy samemu sobie, będzie on w sposób naturalny dążył ku najbardziej prawdopodobnej strukturze, odpowiadającej określonym przez ustalone wartości pewnych makroskopowych parametrów (objętości, energii, struktury chemicznej itd.), warunkom fizycznym układu. Układ może być zdeterminowany nawet w swoich *początkowych* warunkach, odpowiadających pewnej entropii S_0. Dodatkowa informacja I_{b1} w wyrażeniach (12.2) i (12.3) odpowiada większemu stopniowi zdeterminowania. Możemy określić wiele mikroskopowych parametrów, których zmian nie jesteśmy w stanie dokładnie śledzić. Możemy na przykład zbudować układ z pewnych określonych ilości związków chemicznych, lecz później związki te mogą ze sobą reagować i na skutek tego skład chemiczny układu może się zmieniać w czasie, nie jesteśmy zatem w stanie obserwować dokładnie tego składu. Tego typu wstępne zdeterminowanie struktury może zawierać się albo w S_0, albo w I_{b1} i podział całkowitego wzrostu entropii pomiędzy te dwie składowe może nastręczać wiele trudności.

Interesujący jest przypadek, kiedy S_0 odpowiada ogólnej strukturze po swobodnej ewolucji układu izolowanego tak, że wielkość S_0 nie jest zdeterminowana. W takim przypadku S_0 pozostaje stałe i wzór (12.5) upraszcza się do

$$\Delta I_{b1} \leqslant 0 \tag{12.5a}$$

oraz

$$N = -S, \quad \Delta N_1 \leqslant 0, \quad \Delta(N_0 + I_{b1}) \leqslant 0, \tag{12.6}$$

a więc $\Delta I_{b1} \leqslant 0$, jeśli układu nie można zdeterminować w stanie początkowym. Jeśli wykorzystamy negentropię jako miarę jakości energii (rozdz. 9.2), otrzymamy zasadę Kelvina degradacji energii wyrażoną symbolicznie przez wzór (12.6). Dla przemian odwracalnych otrzymamy równość, podczas gdy w przypadku przemian nieodwracalnych mamy do czynienia z nierównością.

Fakt, że $\Delta I_{b1} \leqslant 0$ (wzór (12.5a)), jest zgodny z uzyskanym już wcześniej wynikiem dotyczącym informacji swobodnej, mianowicie, że

$$\Delta I_f \leqslant 0. \tag{12.7}$$

Warunek ten otrzymaliśmy przy przeprowadzaniu dowodu, że ilość informacji (swobodnej, ponieważ nie identyfikowaliśmy abstrakcyjnych symboli z mikrostanami układu fizycznego) osiąga maksimum dla równoprawdopodobnych przypadków.

Rozdział 12.3 poświęcimy na rozważenie różnych przykładów ilustrujących fakt, że oba rodzaje informacji można zamienić w negentropię i że informację, zarówno swobodną jak i związaną, można otrzymać jedynie kosztem negentropii jakiegoś układu fizycznego. Równanie (12.4) możemy przepisać w formie reakcji odwracalnej:

$$I \rightleftarrows N. \tag{12.8}$$

Takie sformułowanie (jeśli włączymy obie klasy informacji) prowadzi do uogólnienia zasady Carnota*. Jeśli

$$\Delta I_f \leqslant 0, \qquad \Delta N_1 \leqslant 0,$$

to

$$\Delta N_1 + \Delta I_f \leqslant 0,$$

lub

$$\Delta(S_1 - I_f) = \Delta(S_0 - I_{b1} - I_f) = \Delta(S_0 - I) \geqslant 0 \tag{12.9}$$

w przypadkach, kiedy odgrywa rolę zarówno informacja związana jak i swobodna. Korzystając z pojęcia negentropii zasadę degradacji energii możemy wyrazić za pomocą nierówności

$$\Delta(N_0 + I) \leqslant 0. \tag{12.10}$$

Wielkość N_0 jest teraz negentropią pewnego układu fizycznego. Składowa informacyjna zawiera informację związaną, która może powodować wzrost N_0 (w przykładzie opisanym przez wzór (12.2) I_{b1} powoduje wzrost N_0 do N_1), oraz informację swobodną. Suma negentropii i ilości informacji może pozostawać stała w przypadku przemian odwracalnych i maleje dla przemian nieod-

* L. Brillouin, J. Appl. Phys. **24** (1953), str. 1152.

wracalnych. Podane nierówności są słuszne jedynie dla wartości średnich. We wszystkich przykładach stosowania zasady Carnota istnieje możliwość wystąpienia przypadkowych fluktuacji.

Podamy teraz przykład ilustrujący rolę informacji swobodnej i związanej. Zacznijmy od tego, że pewien człowiek dysponuje informacją, która tak długo jest swobodna, jak długo pozostaje w umyśle tego człowieka, w takiej bowiem sytuacji nie jest ona bezpośrednio związana z żadnym układem fizycznym. Śledząc straty informacji w czasie jej przekazywania drugiemu człowiekowi, na każdym etapie będziemy ustalać, jakiego rodzaju informacja jest tracona, swobodna czy związana:

(1) Człowiek dysponuje informacją swobodną.

(2) Informację tę przekazuje przyjacielowi, powiedzmy w języku angielskim; będzie to już teraz informacja związana, ponieważ została przetransformowana w fale dźwiękowe, impulsy elektryczne lub innego rodzaju zaburzenia fizyczne, które mogą być stosowane do transmisji. Przy kodowaniu, przed transmisją, mogą wystąpić błędy. Będzie to strata informacji swobodnej, ponieważ występuje ona przed transformacją w zaburzenie fizyczne.

(3) Na skutek zniekształceń i szumu cieplnego w systemie telekomunikacyjnym mogą wystąpić dalsze straty informacji. Będą to straty informacji związanej.

(4) Przyjaciel ma przytępiony słuch i części słów nie odbiera. Na skutek tego występują straty informacji związanej, lecz informacja odebrana przez przyjaciela jest obecnie swobodna, ponieważ znajduje się w jego umyśle.

(5) Po jakimś czasie przyjaciel zapomina pewną część informacji i są to już straty informacji swobodnej.

Straty informacji występują na każdym etapie; w najlepszym przypadku istnieje możliwość nie wystąpienia strat. Taka sytuacja może powstać np., kiedy agent giełdowy podaje przez telefon informacje o kursie akcji.

Należy zwrócić uwagę, że nie można jednoznacznie określić rodzaju traconej informacji w krokach (2), (4) i (5). Czasami, szczególnie w sytuacjach, kiedy występuje człowiek-obserwator, trudno jest odróżnić informację swobodną od związanej. We wszystkich przypadkach musimy oczywiście pominąć każdy element wartości informacji i dlatego nie zajmujemy się np. sytuacją, kiedy uczony gromadzący wyniki pomiarów po ich przestudiowaniu odkrywa prawo fizyczne. Do tego typu zagadnień naszych definicji nie można zastosować. Co więcej, tylko przez wyeliminowanie procesów myślenia mogliśmy

rozwinąć teorię informacji w sposób zadowalający. Możemy mieć oczywiście nadzieję, że w przyszłości uda się nam rozwinąć teorię informacji tak, aby objęła analizę zagadnień związanych z myśleniem i wartością informacji.

Wracając do równań (12.6) i (12.10) możemy podsumować wszystko, co powiedzieliśmy na temat związku pomiędzy teorią informacji i termodynamiką w sposób następujący:

$$\left.\begin{array}{l} \text{negentropia } N \text{ odpowiada ilości informacji } I; \\ \text{temperatura } T \text{ oznacza, że występuje szum cieplny zakłócający} \\ \quad \text{transmisję informacji;} \\ \text{energia zachowuje swoje zwykłe znaczenie;} \\ \Delta Q = T\Delta S \text{ jest ilością ciepła występującego w pewnym procesie;} \\ \Delta W = T\Delta N = T\Delta I \text{ odpowiada pracy mechanicznej, jaka może} \\ \quad \text{być wykonana.} \end{array}\right\} \quad (12.11)$$

Przy wyższej temperaturze T szum jest większy i ilość pracy wymaganej dla transmisji określonej informacji musi być większa, aby przezwyciężyć zakłócenia.

12.3. NIEKTÓRE TYPOWE PRZYKŁADY FIZYCZNE

Omówimy obecnie niektóre klasyczne zagadnienia fizyki; wskażemy, gdzie powstaje informacja związana i jaką odgrywa rolę. Przy okazji sprawdzimy niektóre wcześniejsze rozważania dotyczące entropii i wyjaśnimy kilka trudnych punktów.

Rozważmy przypadek, kiedy mamy idealny jednoatomowy gaz w zbiorniku o objętości V, cieplnie odizolowanym od otoczenia. W warunkach równowagi entropia w tym przypadku jest równa*

$$S = kn\left\{\frac{5}{2} + \ln\left[g\frac{V}{n}\left(\frac{4\pi mE}{3h^2n}\right)^{3/2}\right]\right\}, \quad (12.12)$$

gdzie n — liczba atomów gazu, m — masa atomu, k, h — odpowiednio stałe Boltzmanna i Plancka, E — całkowita energia i g — liczba nierozróżnialnych stanów na poziomie podstawowym atomu.

* J. E. Mayer, M. G. Mayer, *Statistical Mechanics*, New York 1940, Wiley, str. 115, wzór (5.19); również str. 132—140 i 209.

Jeśli stan podstawowy nie jest zdegenerowany, to $g = 1$. Zachodzi to dla przypadku atomu bez momentu pędu na poziomie podstawowym ($j = 0$). Jeżeli atom ma całkowity moment pędu j (spin *plus* moment orbitalny), wówczas występuje $g_j = 2j+1$ stanów, o których zakłada się, że są one równoprawdopodobne. Jest to słuszne przy założeniu, że stan podstawowy jest dostatecznie niski i temperatura nie jest zbyt wysoka tak, że nie może zachodzić pobudzanie do wyższych stanów energetycznych.

Dane otrzymane na drodze doświadczalnej potwierdzają wzór znaleziony teoretycznie, lecz jest bardzo dużo wielkości, które trzeba wprowadzić w teorii, a które są w istocie zbyt małe, aby można je było obserwować w większości eksperymentów fizycznych*. Przepiszmy więc wzór (12.12) w następujący sposób:

$$\left.\begin{array}{ll} \text{dla } g = 1\text{: brak degeneracji, } j = 0, & S_1, \\ \text{dla } g = 2\text{: } j = {}^1/_2 & S_2 = S_1 + kn \ln 2, \\ \text{dla } g = 3\text{: } j = 1 & S_3 = S_1 + kn \ln 3. \end{array}\right\} \quad (12.13)$$

Przypadek, kiedy $g = 2$, odpowiada atomom o spinie ${}^1/_2$ i zwrotach $(\pm {}^1/_2)$. Dla $G = 3$ mamy $j = 1$ i trzy zwroty $(-1, 0, +1)$. Wszystkie zwroty są równoprawdopodobne w przypadku, kiedy pole zewnętrzne jest równe zeru (małe pole zewnętrzne powoduje rozszczepienie poziomów energetycznych i usuwa degenerację).

Różnice między S_1, S_2 i S_3 są bardzo małe, a poza tym nie możemy ich bezpośrednio zaobserwować, ponieważ nie znamy sposobu zmiany wartości j danego atomu. Jednakże w przemianach radioaktywnych możemy obserwować przejście atomu o masie m oraz spinie j w inny atom, o masie m' oraz spinie j'; w takich przypadkach wielkości te miałyby sens fizyczny.

Przeprowadzona dyskusja wykazuje, że o różnicy entropii ΔS możemy mówić tylko w przypadku, jeśli przemiana jest fizycznie możliwa, w przeciwnym razie wielkość ΔS nie ma żadnego sensu fizycznego. Przyjmijmy, że posiadamy dodatkową informację o stanie gazu, np. wiemy, że chwilę wcześniej gaz ten zajmował mniejszą objętość. Ma to miejsce w przypadku, kiedy gaz znajduje się w zbiorniku o objętości V' i nagle otwieramy zawór łączący ten

---

* O. Stern, Revs. Modern Phys. **21** (1949), str. 534.
E. Schrödinger, Proc. Roy. Irisch Acad. **53 A** (1950), str. 189.
K. Darrow, Bell System Tech. J. **21** (1942), str. 51; **22** (1943), str. 108.

zbiornik z innym, o objętości V'':

$$V = V' + V'', \quad V' < V.$$

Entropia początkowa S' jest mniejsza niż entropia S po rozprężeniu o wartość daną przez wzór (12.12):

$$S - S' = kn(\ln V - \ln V') = I, \tag{12.14}$$

gdzie I jest ilością informacji. Po otwarciu zbiornika V'' gaz napłynie do niego i wystąpią oscylacyjne zmiany gęstości gazu w obu zbiornikach. Stan równowagi, o równomiernej gęstości w całej objętości V, ustala się stopniowo. Wzrost entropii i strata informacji występują jednocześnie. Możemy więc powiedzieć, że gaz stopniowo „zapomina" informację*.

Jako inny przykład może posłużyć przypadek dyfuzji gazu. Aby skoncentrować się na istotnych punktach tego zagadnienia, przyjmiemy, że mamy dwa gazy o atomach o tej samej masie m oraz $g = 1$, przy czym n_1 atomów typu 1 początkowo zajmuje objętość V_1, natomiast n_2 atomów typu 2 — objętość V_2:

$$\left.\begin{array}{l} n_1 + n_2 = n, \quad V_1 + V_2 = V, \\ n_1 = p_1 n, \quad n_2 = p_2 n, \quad V_1 = p_1 V, \quad V_2 = p_2 V, \quad p_1 + p_2 = 1. \end{array}\right\} \tag{12.15}$$

Związki (12.5) implikują równość początkowych koncentracji obu składników i koncentracji końcowej:

$$V_1/n_1 = V_2/n_2 = V/n.$$

Zakładamy prócz tego ekwipartycję energii:

$$E_1/n_1 = E_2/n_2 = E/n. \tag{12.16}$$

Z przyjętych założeń wynika, że ciśnienia początkowe i temperatury obu gazów nie różnią się.

---

* J. von Neumann rozważa również przypadek, kiedy obserwator zapomina informację, i stwierdza, że proces ten także oznacza wzrost entropii. Widzimy tu znowu, że niemożliwe jest rozróżnianie w sposób ścisły informacji swobodnej od związanej we wszystkich przypadkach. Potwierdza to potrzebę wprowadzenia uogólnionej zasady Carnota (rozdz. 12.2). Dalsze rozważania na ten temat można znaleźć u J. Von Neumanna, *Mathematische Grundlagen der Quanten-Mechanik* (Berlin 1932, Springer) lub u P. Jordana, Philosophy of Sci. **16** (1949), str. 269.

Oznaczmy przez Q wyrażenie w nawiasie wzoru (12.12):

$$Q = [(4\pi mE)/(3h^2n)]^{3/2}.$$

Z naszych założeń wynika także, że Q ma tę samą wartość dla obu gazów i zachowuje ją w czasie procesu mieszania się gazów. Mamy więc przed zmieszaniem

$$S_{12} = k(n_1+n_2)\left(\frac{5}{2} + \ln\frac{V}{n} + \ln Q\right) = S. \tag{12.17}$$

Natomiast po zmieszaniu

$$S'_{12} = kn_1\left(\frac{5}{2} + \ln\frac{V}{n_1} + \ln Q\right) + kn_2\left(\frac{5}{2} + \ln\frac{V}{n_2} + \ln Q\right) =$$

$$= S_{12} + kn_1(\ln n - \ln n_1) + kn_2(\ln n - \ln n_2), \tag{12.18}$$

$$S'_{12} = S_{12} - kn(p_1 \ln p_1 + p_2 \ln p_2).$$

Nieodwracalny wzrost entropii jest równy

$$\Delta S' = S'_{12} - S_{12} = -kn(p_1 \ln p_1 + p_2 \ln p_2) > 0. \tag{12.19}$$

Wielkość ta jest dodatnia, ponieważ $p_1<1$ oraz $p_2<1$, a więc logarytmy są ujemne. W sytuacji początkowej mieliśmy o układzie pewną informację I (związaną).

$$I = -kn(p_1 \ln p_1 + p_2 \ln p_2), \tag{12.20}$$

którą straciliśmy w procesie dyfuzji. Możemy porównać wzór (12.20) z wyrażeniem Shannona na informację (2.1). W stanie końcowym gaz w postaci mieszaniny odpowiada sytuacji z prawdopodobieństwami a priori p_1 oraz p_2 dla dwóch symboli. Każdy symbol jest tu przedstawiony przez jeden z typów atomów. Wybór jednej szczególnej kombinacji symboli daje (Shannon) informację na symbol $-k(p_1 \ln p_1 + p_2 \ln p_2)$. Stan początkowy gazu odpowiada selekcji atomów *1* i umieszczeniu ich w V_1, podczas gdy atomy *2* pozostają w V_2. Wymaga to informacji wyrażonej równaniem (12.20).

Możemy teraz rozważyć cokolwiek inne zagadnienie. Przyjmijmy mianowicie, że atomy mają spin $^1/_2$: typ *1* odpowiada orientacji $+^1/_2$, podczas gdy typ *2* odpowiada orientacji $-^1/_2$. Zderzenia będą powodowały przejścia pomiędzy dwoma typami atomów i stan końcowy będzie odpowiadać

$$p_1 = {}^1/_2, \quad p_2 = {}^1/_2, \tag{12.21}$$

oraz entropii

$$S''_{12} = S_{12} + kn \ln 2. \tag{12.22}$$

Mamy tu znowu do czynienia ze wzrostem entropii i stratą informacji

$$\Delta S'' = S''_{12} - S'_{12} = kn[\ln 2 + p_1 \ln p_1 + p_2 \ln p_2] > 0. \tag{12.23}$$

Entropia końcowa S''_{12} jest równa entropii S_2 ze wzoru (12.13) dla gazu o spinie $^1/_2$ i dwu równoprawdopodobnych zwrotach. Przeprowadzone rozważania w sposób dostatecznie jasny wskazują na konieczność wprowadzenia wyrazu g do wyrażenia na entropię i wyjaśniają jego znaczenie fizyczne. Wzrost entropii odpowiada stracie informacji, podobnie jak w naszych poprzednich przykładach.

Porównajmy teraz otrzymane wyrażenie z wynikami rozważań przeprowadzonych w rozdz. 2. Pokazaliśmy, że wybór jednej kombinacji dwu symboli, mających prawdopodobieństwa a priori p_1, p_2 daje ilość informacji na symbol

$$I_p = -K(p_1 \ln p_1 + p_2 \ln p_2). \tag{12.24}$$

Maksimum tej wielkości występuje dla

$$p_1 = p_2 = {}^1/_2, \qquad I_{\max} = k \ln 2. \tag{12.25}$$

Jeśli teraz przechodzimy od przypadku p_1, p_2 do przypadku $^1/_2, ^1/_2$, zmiana ilości informacji na symbol wyniesie

$$\Delta I = I_{\max} - I_p = k[\ln 2 + p_1 \ln p_1 + p_2 \ln p_2]. \tag{12.26}$$

Jest to wyrażenie dokładnie takie samo jak (12.23).

12.4. UWAGI OGÓLNE

Obecnie możemy już wyciągnąć pewne interesujące wnioski z rozważonych zagadnień.

Jeśli ilość informacji, jaką posiadamy o układzie fizycznym, zwiększy się, to entropia tego układu ulegnie zmniejszeniu. Małej wartości entropii odpowiada stan niestabilny, który wcześniej lub później, poprzez naturalną ewolucję, przejdzie w stan stabilny o dużej entropii.

Druga zasada termodynamiki nie mówi nic na temat czasu trwania tego

procesu i w związku z tym nie wiemy, jak długo układ będzie „pamiętał" informację. Ponieważ klasyczna termodynamika nie daje odpowiedzi na to ważne pytanie, musimy odwołać się do modelu molekularnego lub atomowego i pomocy teorii kinetycznej. Przy użyciu odpowiednich modeli, możemy obliczać szybkości tłumienia różnego rodzaju fal, szybkość dyfuzji, szybkość reakcji chemicznych itd. Obliczone czasy mogą się zmieniać w szerokim zakresie od ułamków sekundy do lat i wieków.

Tego typu opóźnienia czasowe można w praktyce wykorzystać w wielu przypadkach. Nie potrzeba wiele czasu, aby ciąg impulsów (odpowiadających np. kropkom i kreskom) przesyłanych przez linię zanikł i został zapomniany, jednakże ten krótki odcinek czasu jest wystarczający dla przesłania tego ciągu nawet na bardzo dużą odległość. Fakt ten umożliwia telekomunikację.

Układ zdolny do przechowywania informacji przez pewien czas można wykorzystać jako pamięć w maszynie matematycznej.

Przykłady rozważane w rozdz. 12.3 są interesujące nie tylko z teoretycznego punktu widzenia. Pokazują one również, w jaki sposób można rozwiązać szereg zagadnień praktycznych. Rozważmy dla przykładu problemy dyfuzji i rozkładu spinów (por. wzory (12.15)÷(12.26). Są one ściśle związane z magnetycznymi urządzeniami pamięciowymi; spin zawsze jest związany z momentem magnetycznym, stąd sytuacja w rodzaju opisanej przez wzór (12.15) oznacza, że objętość V' jest namagnesowana do nasycenia w pewnym kierunku, podczas gdy objętość V'' jest namagnesowana w kierunku przeciwnym. Informacji zmagazynowanej w układzie odpowiada ubytek entropii. Rozważania nasze pokazują, w jaki sposób, na skutek dyfuzji i zderzeń powodujących wzrost entropii i „wymazywanie" informacji, sytuacja stopniowo ulega zmianie.

Entropię definiujemy zazwyczaj jako miarę nieuporządkowania w układzie fizycznym. Bardziej precyzyjnie możemy powiedzieć, że *entropia mierzy brak informacji o rzeczywistej strukturze układu*. Taki brak informacji prowadzi do tego, że bardzo duża ilość różnych mikrostruktur, których w praktyce nie jesteśmy w stanie odróżnić od siebie, jest możliwa. Ponieważ w danej chwili może wystąpić jedna z tych mikrostruktur, brak informacji odpowiada rzeczywistemu nieuporządkowaniu na poziomie ukrytych stopni swobody. Obraz ten dobrze ilustruje przykład gazu idealnego. Możemy określić całkowitą ilość n atomów, ich masę m, współczynnik g oraz całkowitą energię E (wzór (12.11)), nie znamy jednak położenia i prędkości poszczególnych atomów.

Jest to brak informacji, któremu odpowiada entropia określona wzorem (12.12). Ponieważ nie znamy położeń i prędkości atomów, nie jesteśmy w stanie odróżnić dwóch próbek gazu różniących się jedynie położeniami i prędkościami atomów. Sytuację możemy jedynie określić jako nieuporządkowany ruch atomów.

Początki nowoczesnych rozważań dotyczących pojęcia entropii można znaleźć w jednym ze starych artykułów Szilarda*, którego pionierska praca nie została wtedy dobrze zrozumiana. Związek pomiędzy entropią i informacją odkrył ponownie Shannon**. Zdefiniował on entropię ze znakiem przeciwnym niż znak zwykłej, termodynamicznej entropii. Stąd to, co Shannon nazywa entropią informacji, jest w rzeczywistości negentropią. Można to łatwo zauważyć na podstawie dwóch przykładów (str. 27 i str. 61 książki Shannona), w których udowadnia on, że w pewnych nieodwracalnych procesach (nieodwracalny przekształtnik lub filtr) entropia maleje. Aby być w zgodzie z naszymi określeniami należy zmienić znak i czytać negentropia.***

Związkiem między entropią i informacją zajmował się w swoich ostatnich pracach J. Rothstein.**** Jego sposób podejścia do rozważanego zagadnienia, cechujący się dużą jasnością, w pełni pokrywa się z punktem widzenia autora przedstawionym w niniejszym rozdziale.

---

* L. Szilard, Z. Physik **53** (1929), str. 840.

** C. E. Shannon, W. Weaver, *The Mathematical Theory of Information*, Urbana Ill. 1949, Univ. of Illinois Press.

*** Wystarczy czytać: *negentropia*. (Przyp. tłum.).

**** J. Rothstein, Science **114** (1951), str. 171; Phys. Rev. **85** (1952), str. 135.

13

DEMON MAXWELLA A NEGENTROPIJNA ZASADA INFORMACJI

13.1. DEMON MAXWELLA — PODŁOŻE HISTORYCZNE

Zagadnienie demona Maxwella dostarcza dobrego przykładu zastosowania teorii informacji i wyraźnie pokazuje powiązanie między informacją i entropią.

Demon sortujący pojawił się po raz pierwszy w 1871 roku w *Teorii ciepła* Maxwella* jako «istota, której zmysły są tak wyostrzone, że mogłaby śledzić każdą cząsteczkę w czasie jej ruchu i byłaby w stanie robić to, co obecnie jest dla nas niemożliwe... Przypuśćmy, że naczynie podzielone jest na dwie części A i B przegrodą, w której znajduje się mały otworek, i że istota, która *może dostrzegać pojedyncze cząsteczki*, zatyka i odtyka ten otworek, pozwalając cząsteczkom szybszym przechodzić z A do B, a powolniejszym tylko z B do A. Istota ta, bez wykonywania pracy, będzie więc podnosić temperaturę w części B, a obniżać w A — wbrew II zasadzie termodynamiki».**

Paradoks ten rozpatrywało kilka pokoleń fizyków. M. Smoluchowski*** pierwszy zwrócił uwagę na możliwy wpływ ruchu Browna na drzwiczki zatrzaskowe, co prowadziłoby do przypadkowego otwierania lub zamykania tych drzwiczek i poważnie zakłóciłoby działanie całego układu. Miałoby to szczególne znaczenie w jakimś przyrządzie automatycznym, takim jak zawór sprężynowy, i całkowicie uniemożliwiłoby długotrwałe działanie takiego układu. Smoluchowski doszedł do wniosku, że ruchy Browna stanowią tylko

---

* Tytuł oryginału: *Theory of Heat*, str. 328.

** Pełny cytat przytacza J. H. Jeans w *Dynamical Theory of Gases*, 3rd ed., Cambridge, Mass. 1921, Cambridge Univ. Press, str. 183.

*** M. Smoluchowski, Physik Z. **13** (1912), str. 1069.

pozorne naruszenie II zasady termodynamiki ze względu na przypadkową, nie dającą się przewidzieć naturę i ze względu na krótki czas trwania ruchu. Stałe działanie układu z jakimś zatrzaskiem jest niemożliwe i niemożliwe jest z pewnością istnienie perpetuum mobile drugiego rodzaju. Układ z zatrzaskiem może działać, lecz tylko nieregularnie i w sposób niesystematyczny.

Interesująco przedstawia się problem działania zaworu sprężynowego: gdyby nie jego własne ruchy Browna, to zawór mógłby wytwarzać różnicę ciśnień między dwoma połączonymi naczyniami. Jest to przypadek podobny do prostownika, który omówiono pod koniec rozdz. 11. Idealny prostownik działający na pojedyncze elektrony prostowałby ruch termiczny elektronów wbrew II zasadzie. Nie ma jednak idealnego prostownika, a jedynym wynikiem, jakiego możemy się spodziewać, jest naruszenie symetrii ruchu termicznego — jak to wyjaśniono w rozdz. 11.

Szilard* w swojej znakomitej pracy jako pierwszy wyjaśnił, że demon wpływa na informację o mikrostrukturze ruchu gazu i w rzeczywistości zmienia informację w negentropię. Omówimy później niektóre ciekawsze zagadnienia poruszane przez Szilarda.

Lewis** przedyskutował zagadnienia podobne do rozważonych w poprzednim rozdziale, a mianowicie rozdzielanie i dyfuzję gazu. Rozważał on składową entropii związaną z mieszaniem lub rozdzielaniem gazów i wyciągnął wnioski podobne do naszych. Slater*** postawił pytanie, czy zasada nieoznaczoności może odgrywać jakąś rolę w tego rodzaju zagadnieniu.

Demon Maxwella powinien mierzyć jednocześnie położenie i prędkość danego atomu. Obu tych wielkości nie można mierzyć z nieskończenie dużą dokładnością ze względu na dobrze znane ograniczenie:

$$\Delta p \, \Delta q \geqslant h, \qquad \text{gdzie} \qquad p = mv. \tag{13.1}$$

Nieoznaczoność ta może odgrywać rolę w przypadku atomów lekkich (mała masa m) i przy dużych ciśnieniach, kiedy to położenia atomów powinny być określone z dużą dokładnością. Ograniczenie związane z nieoznaczonością, jak wykazał Demers****, nie ma jednak znaczenia w przypadku atomów ciężkich i niskich ciśnień. W istocie rzeczy mamy do czynienia z bardziej funda-

* L. Szilard, Z. Physik **53** (1920), str. 840.

** G. N. Lewis, Science **71** (1930), str. 569.

*** J. C. Slater, *Introduction to Chemical Physics*, New York 1939, Mc Graw-Hill.

**** P. Demers, Can. J. Research **22** (1944), str. 27; **23** (1945), str. 47.

mentalnym zagadnieniem postawionym po raz pierwszy przez Demersa* i Brillouina**: Czy możliwe jest, aby demon mógł rzeczywiście dostrzegać pojedyncze atomy?

Przyjmijmy, że cały układ*** jest odizolowany i znajduje się początkowo w danej temperaturze T. Demon znajduje się w zamkniętym naczyniu będącym w stanie równowagi przy stałej temperaturze, gdzie promieniowanie musi być promieniowaniem ciała czarnego, a przecież wewnątrz ciała czarnego niczego zobaczyć nie można. Podnoszenie temperatury nie pomogłoby nic.

Przy „czerwonej" temperaturze maksimum promieniowania przypada na czerwień i ma to samo natężenie, bez względu na to, czy nie ma cząsteczek, czy są ich miliony wewnątrz zamkniętego naczynia. Nie tylko zostaje zachowane to samo natężenie, lecz i fluktuacje są te same. Demon odczuwałby promieniowanie cieplne wraz z jego fluktuacjami, ale nigdy nie dostrzegałby cząsteczek.

Nie powinno dziwić to, że Maxwell nie pomyślał o włączeniu promieniowania do układu będącego w stanie równowagi przy temperaturze T. Promieniowanie ciała czarnego poznano dopiero w 1871 roku. Upłynęło jednak jeszcze 30 lat, zanim właściwie zostało zrozumiane zagadnienie termodynamiki promieniowania i pojawiła się teoria Plancka.

Demon nie może dostrzegać cząsteczek, stąd też nię może wprawiać w ruch drzwiczek i naruszyć tym samym drugiej zasady termodynamiki. Demon nie jest w stanie dostrzegać cząsteczek, lecz może próbować je wykryć jedną z następujących metod: może mierzyć siły Van der Waalsa lub pola pochodzące od dipoli elektrycznych, czy też momenty magnetyczne.

Tutaj sytuacja jest inna niż w przypadku obserwacji światła: wszystkie te pola mają mały zasięg i natężenie ich maleje co najmniej z r^{-2}, gdzie r jest odległością. Demon mógłby wykrywać cząsteczki tylko wtedy, gdy znalazłyby się one blisko ścianki i drzwiczek. Jest to już zbyt późno, aby był on w stanie wprawiać w ruch drzwiczki nie wykonując żadnej pracy. Te same siły, na których polegałby demon przy wykrywaniu cząsteczek, działałyby również na drzwiczki lub zawór. Przeto poruszanie drzwiczkami wymagałoby wykonania pewnej pracy, a to już komplikuje bardzo analizę układu.

---

* P. Demers, *loc. cit.*

** L. Brillouin, Am. Scientist **37** (1949), str. 554 — zob. str. 565; **38** (1950), str. 594; również J. Appl. Phys. **22**, (1951), str. 334.

*** Należy zwrócić uwagę na różnicę znaczenia słowa „układ". W tym przypadku mamy na myśli układ fizyczny, jakim jest np. gaz zawarty w jakimś naczyniu. (*Przyp. tłum.*)

W założeniach Maxwella istotne jest, aby cząsteczka została wykryta na długo przed osiągnięciem ścianki, i to w takiej odległości, żeby wszystkie siły o małym zasięgu można było pominąć bez uszczerbku dla eksperymentu. Potrzebne jest tu jakieś promieniowanie. Najłatwiej zagadnienie to omówić zakładając, że jest to promieniowanie świetlne.

13.2. WYPĘDZENIE DEMONA

Zbadajmy możliwości demona. Możemy wyposażyć go w latarkę elektryczną, tak aby mógł dostrzegać cząsteczki. Latarka jest źródłem promieniowania, które nie znajduje się w stanie równowagi. Wprowadza ona ujemną entropię do układu. Z tej ujemnej entropii demon czerpie informację. Wykorzystując tę informację wprawia w ruch drzwiczki i odtwarza entropię ujemną, zamykając w ten sposób cykl:

$$\text{negentropia} \rightarrow \text{informacja} \rightarrow \text{negentropia}. \tag{13.2}$$

W całości procedurę taką rozpatrywał Demers*, nie zwracając uwagi na etap pośredni, w którym demon uzyskuje informację wykorzystywaną w ostatniej operacji. Demers na różnych przykładach wykazał, że strata negentropii latarki jest większa niż zysk końcowy. Innymi słowy początkowy przyrost entropii jest większy niż ubytek końcowy, a całkowity bilans czyni zadość zasadzie Carnota.

Szilard wcześniej zwrócił uwagę na etap pośredni z informacją. Przeprowadził on bardzo dokładną analizę transformacji informacji w negentropię, lecz swoje badania dotyczące pierwszego etapu oparł na całkowicie sztucznym modelu.

Zacznijmy od ogólnych rozważań, które zilustrują najbardziej istotne punkty. Badania dokładniejsze oparte o wyniki Jacobsona** przedstawimy później.

W celu przedyskutowania bilansu entropii musimy zdefiniować układ odizolowany, do którego bezpiecznie można stosować II zasadę termodynamiki.

W układzie naszym wyróżnić można następujące części składowe:

---

* P. Demers, *op. cit.*

** H. Jacobson, Trans. N. Y. Acad. Sci. **14** (1951), str. 6.

(1) Naładowaną baterię i żarówkę przedstawiające latarkę elektryczną.

(2) Gaz znajdujący się w stałej temperaturze T_0 w zamkniętym naczyniu Maxwella*. Naczynie posiada przegrodę z otworkiem, która dzieli je na dwie części.

(3) Demon wprawia w ruch drzwiczki znajdujące się przy otworku. Cały układ jest odizolowany i zamknięty.**

Bateria rozgrzewa włókno latarki do wysokiej temperatury T_1:

$$T_1 > T_0. \tag{13.3}$$

Warunek ten jest konieczny dla uzyskania światła widzialnego o częstotliwości ν_1, tak aby

$$h\nu_1 > kT_0 \tag{13.4}$$

i żeby było ono rozróżnialne na tle promieniowania ciała czarnego w naczyniu o temperaturze T_0.***

W czasie doświadczenia bateria dostarcza całkowitej energii E, ale nie wprowadza zmian entropii. Rozżarzone włókno wypromieniowuje energię E i traci entropię. Zmiana entropii włókna wyniesie

$$S_f = -\frac{E}{T_1} = \text{negentropii wprowadzonej do gazu.} \tag{13.5}$$

Jeżeli demon nie interweniuje, energia E absorbowana jest przez gaz o temperaturze T_0 i obserwujemy całkowity przyrost entropii

$$S = \frac{E}{T_0} + S_f > 0. \tag{13.6}$$

Prześledźmy teraz pracę demona. Może on wykryć cząsteczkę wtedy, gdy co najmniej jeden kwant energii $h\nu_1$ zostanie rozproszony przez nią i zaabsorbowany przez jego oko (lub fotokomórkę, jeżeli taką się posługuje). Końcowy przyrost

---

* Naczynie, o którym mowa, przyjęto nazywać naczyniem Maxwella. Jest to naczynie podzielone na dwie części; połączone są one małym otworkiem, który można w jakiś sposób zatykać i odtykać. (*Przyp. tłum.*).

** Demona można zastąpić urządzeniem automatycznym z *magicznym okiem*, które to urządzenie otwiera drzwiczki w odpowiednim momencie czasu. Jest to wyłącznie kwestia zaprojektowania takiego urządzenia i nie zmienia ogólnych wniosków wynikających z całego zagadnienia.

*** Zob. wzór (9.45). Warunek (13.4) wymagany jest po to, aby średnia energia E rezonatora kwantowego była znacznie mniejsza od kT_0.

entropii wyniesie zgodnie z warunkiem (13.4)

$$S_d = \frac{h\nu_1}{T_0} = kb, \tag{13.7}$$

gdzie

$$\frac{h\nu_1}{kT_0} = b > 1.$$

Raz otrzymana informacja może być wykorzystana do zmniejszenia entropii układu. Entropia układu zgodnie ze wzorem Boltzmanna wyniesie

$$S_0 = k \ln P_0, \tag{13.8}$$

gdzie P_0 jest pewną wagą statystyczną układu i oznacza całkowitą ilość mikrostanów układu (kompleksjony) Plancka. Po uzyskaniu informacji układ zostaje dokładniej określony. Ilość mikrostanów P_0 maleje o wielkość p, a zatem mamy

$$P_1 = P_0 - p,$$

$$\Delta S_i = S - S_0 = k\,\Delta(\ln P) = -k\left(\frac{p}{P_0}\right). \tag{13.9}$$

Oczywiste jest, że we wszystkich praktycznych przypadkach $p \ll P_0$. Równowagę entropii wyrazić można wzorem:

$$\Delta S_d + \Delta S_i = k\left[b - \left(\frac{p}{P_0}\right)\right] > 0, \tag{13.10}$$

skoro tylko $b > 1$ oraz $p/P_0 < 1$.

Końcowym efektem jest wzrost entropii układu odizolowanego, czego wymaga II zasada termodynamiki. Demon może jedynie odzyskać małą część negentropii i wykorzystać informację do zmniejszenia degradacji energii. W pierwszym etapie procesu (wzór (13.7)) otrzymujemy przyrost entropii ΔS_d, a tym samym zmianę negentropii

$$\Delta N_d = -kb < 0,$$

tzn. ubytek negentropii. Z tej utraconej negentropii pewna jej ilość ulega zamianie na informację i w ostatnim etapie procesu (równanie (13.9)) informacja ta zamienia się znowu w negentropię

$$\Delta N_i = k\left(\frac{p}{P_0}\right) > 0,$$

tj. przyrost negentropii. Usprawiedliwia to przyjęcie podstawowego schematu (13.2).

13.3. OMÓWIENIE

Rozpatrzmy bardziej szczegółowo pierwotne zagadnienie Maxwella. Można przyjąć, że po pewnym czasie demon był w stanie uzyskać różnicę temperatur ΔT:

$$\left.\begin{aligned} T_B > T_A, \quad & T_B = T + {}^1/_2\Delta T, \\ T_B - T_A = \Delta T, \quad & T_A = T - {}^1/_2\Delta T. \end{aligned}\right\} \tag{13.11}$$

W następnym kroku demon wybiera szybko poruszającą się cząsteczkę z naczynia A o energii kinetycznej ${}^3/_2kT\,(1+\varepsilon_1)$ i kieruje ją prosto do naczynia B. Później wybiera cząsteczkę poruszającą się powoli, z naczynia B, o energii kinetycznej ${}^3/_2kT(1-\varepsilon_2)$ i pozwala jej przejść do naczynia A. Aby zaobserwować te dwie cząsteczki, demon potrzebuje dwóch kwantów światła i stąd pojawia się przyrost entropii, podobny do obliczonego w wyrażeniu (13.7):

$$\Delta S_d = 2kb, \quad \text{gdzie } b = h\nu_1/kT > 1. \tag{13.12}$$

Wymiana cząsteczek prowadzi do przeniesienia z A do B energii

$$\Delta Q = {}^2/_3kT(\varepsilon_1+\varepsilon_2), \tag{13.13}$$

która odpowiada ubytkowi całkowitej entropii zgodnie ze wzorem (13.11):

$$\Delta S_i = \Delta Q\left(\frac{1}{T_B} - \frac{1}{T_A}\right) = -\Delta Q\,\frac{\Delta T}{T^2} = \frac{3}{2}\,k(\varepsilon_1+\varepsilon_2)\,\frac{\Delta T}{T}. \tag{13.14}$$

Wielkości ε_1 oraz ε_2 są na ogół małe, chociaż czasami mogą przyjmować wartości rzędu kilku jedności.

Przyrost ΔT jest dużo mniejszy od T i stąd

$$\Delta S_i = -{}^3/_2k\eta, \quad \eta \ll 1,$$

a zatem

$$\Delta S_d + \Delta S_i = k(2b - {}^3/_2\eta) > 0, \tag{13.15}$$

co spełnia zasadę Carnota.

Demers* rozpatrywał inny przykład, zakładając, że demon znajduje się w naczyniu zamkniętym o temperaturze niższej $T_2 \ll T_0$. Można by wyobrazić sobie doświadczenie, w którym demon mógłby rozróżniać kwanty $h\nu$ emito-

* P. Demers, *op. cit.*

wane przez cząsteczki przy temperaturze T_0. Zamiast wzoru (13.4) mamy wówczas

$$h\nu > kT_2, \quad T_2 < T_0 \tag{13.16}$$

i rozumować można podobnie.

Bez względu na to, czy posługujemy się temperaturą wyższą T_1 (wzór (13.3)), czy niższą T_2, niezbędna jest zawsze różnica temperatur, gdyż w przeciwnym razie demon nie będzie mógł spełniać swojej roli. Skoro zaś mamy różnicę temperatur, to zbędna jest pomoc demona. W tych warunkach wystarczy posłużyć się dowolną maszyną cieplną. Jeżeli chcemy np. wytworzyć różnicę temperatur i posługujemy się żarówką o temperaturze T_1, to najprościej podgrzewać promieniowaniem jedną połowę gazu, a drugą utrzymywać przy temperaturze T_0. Postępowanie takie będzie bardziej efektywne niż praca demona.

Nasze pierwsze rozważania były powierzchowne, gdyż przyjęliśmy założenia odpowiadające *zwykłym* warunkom, mianowicie: $p \ll P_0$ we wzorze (13.9) lub $\Delta T \ll T$ we wzorze (13.11). Zagadnienie pracy demona prześledzimy bardziej szczegółowo w rozdz. 13.4 i wykażemy, że zasada Carnota jest zawsze spełniona, nawet w wyjątkowych sytuacjach.

Wykryliśmy jednak ważne prawo fizyczne wyrażone równaniem (13.10), które mówi, że każdy pomiar wymaga odpowiedniego przyrostu entropii oraz że istnieje dolna granica zmian entropii, poniżej której pomiar staje się niemożliwy. Granica ta odpowiada zmianie entropii rzędu k, tj. stałej Boltzmanna. Bardziej szczegółowe rozważania wykażą, że dokładna granica wynosi $k \ln 2$ lub w przybliżeniu $0{,}7k$ na jeden bit uzyskanej informacji. Jak mówi Gabor*: «Nie możemy dostać za darmo nic, nawet obserwacji». To bardzo ważne prawo jest prostym następstwem ogólnej negentropijnej zasady informacji i powrócimy do niego później. Wydaje się niezrozumiałe, dlaczego jeszcze do niedawna nie dostrzegaliśmy tak ogólnego prawa.

Znamienną cechą naszych rozważań było to, że uwzględniały one warunki kwantowe, natomiast stała Plancka h w wynikach końcowych nie występuje; zależą one tylko od stałej Boltzmanna k. Dowodzi to, że nasze wyniki są niezależne od kwantów i zasady nieoznaczoności i w rzeczywistości analizę przeprowadzić można w ramach klasycznej teorii, bez wprowadzania warunków kwantowych.

---

* D. Gabor, *M.I.T. Lectures*, 1951.

13.4. POSTĘPOWANIE DEMONA JAKO TRANSFORMACJA INFORMACJI W NEGENTROPIĘ

Zasadnicze zagadnienie jest tak ważne, że wymaga ściślejszego i bardziej szczegółowego prześledzenia. W rozdziale tym rozwiniemy zagadnienie poruszone przez Jacobsona*. Wybieramy najpierw przykład bardziej uproszczony od pierwotnie wymyślonego przez Maxwella. Zamiast demona temperaturowego, jak można określić omawianego poprzednio demona sortującego, będziemy rozważać demona ciśnieniowego. Demon taki wprawiając odpowiednio w ruch drzwiczki pozwala cząsteczkom przechodzić z B do C, a nie pozwala powracać z C do B. Po pewnym czasie wzrośnie ciśnienie w C, a zmaleje w B. Operację taką łatwiej przedyskutować, ponieważ demon ciśnieniowy nie wymaga dokładnego pomiaru prędkości, zbędne jest zatem uwzględnianie zasady nieoznaczoności. W celu wyeliminowania nieistotnych parametrów, określimy dokładnie warunki doświadczenia, którego schemat pokazano na rys. 13.1.

Naczynia B i C połączone są rurką o przekroju A równym powierzchni drzwiczek. Wiązka światła ze źródła zewnętrznego przechodzi przez okienko i światło rozproszone przez cząsteczki służy do wykrywania tych cząsteczek. Wykrycie atomów musi zajść w jakiejś odległości x od drzwiczek, a prędkość atomów musi być znana co najmniej w przybliżeniu, aby można było określić czas otwarcia drzwiczek. Możemy np. posłużyć się przesłoną obrotową pozwalającą przechodzić cząsteczkom o prędkościach zbliżonych do prędkości średniej, a odbijającą wszystkie cząsteczki o innych prędkościach.

W przyrządzie takim mielibyśmy dobrze określoną powierzchnię A, natomiast prędkości uderzających cząsteczek byłyby określone tylko w przybliżeniu. Naszym zadaniem jest określenie ilości informacji wymaganej przez demona, aby mógł on efektywnie działać. Rozważmy długi przedział czasu t, podzielmy go na małe podprzedziały τ. Demon musi wiedzieć, czy ma otworzyć drzwiczki na okres τ, czy trzymać je zamknięte. W każdym przedziale τ wymagana jest informacja, czy istnieje cząsteczka (lub cząsteczki), która powinna uderzyć w drzwiczki o powierzchni A z lewej strony. Sprowadza się to do typowego pytania TAK czy NIE. Istnieje prawdopodobieństwo a priori p_1 TAK oraz p_2 NIE. W ciągu długiego przedziału czasu t średnio przez N_1 małych odcinków czasu drzwiczki będą otwarte i przez N_2 — zamknięte. Możemy

* H. JACOBSON, *op. cit.*

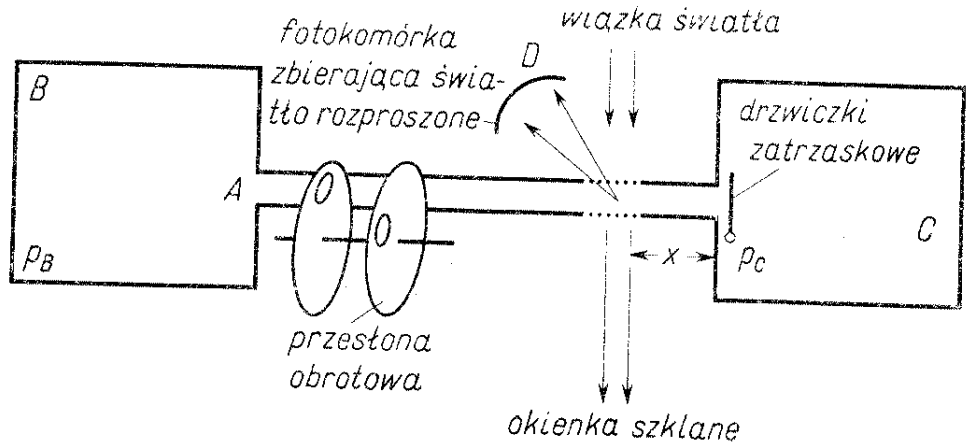
to zapisać, jak następuje:

$$t = G\tau, \quad N_1 = Gp_1, \quad N_2 = Gp_2, \quad p_1+p_2 = 1. \tag{13.17}$$

Jeżeli podaną przez nas w rozdz. 1 definicję ilości informacji zastosujemy bezpośrednio do tego zagadnienia, to za okres τ otrzymujemy ilość informacji

$$I_\tau = -k(p_1 \ln p_1 + p_2 \ln p_2) \text{ jednostek termodynamicznych.} \tag{13.18}$$

Zastanówmy się nad teorią kinetyczną naszego zagadnienia. Średnia liczba cząsteczek uderzających w drzwiczki o powierzchni A w czasie τ (przy tempera-



Rys. 13.1. W naczyniach A i B znajdują się cząsteczki odpowiednio pod ciśnieniami P_B i P_C. Cząsteczki te mogą przechodzić z B do C przez rurkę o przekroju A. Dwie obrotowe przesłony z otworkami również o powierzchni A są ustawione tak, że cząsteczka przechodząca przez otworek w lewej przesłonie przejdzie również przez otworek w przesłonie prawej, jeżeli ma prędkość zbliżoną do pewnej danej prędkości. Do wykrywania cząsteczek przechodzących przez rurkę służy wiązka światła oświetlająca szklaną część rurki. Światło rozproszone przez cząsteczkę trafia do detektora, który daje sygnał otwierający drzwiczki, aby cząsteczka mogła przejść do C

turze T_0) jest proporcjonalna do gęstości gazu. Mamy średnio:

$$\left.\begin{aligned} u &= bA\tau \text{ cząsteczek uderzających z lewej strony,} \\ v &= ru = cA\tau \text{ cząsteczek uderzających z prawej strony,} \end{aligned}\right\} \tag{13.19}$$

natomiast

$$r = \frac{c}{b} = \frac{P_C}{P_B}, \tag{13.20}$$

gdzie P_C i P_B oznacza odpowiednio ciśnienie w naczyniach B i C.

Przesłona wybierająca cząsteczki o prędkości średniej z lewej strony będzie eliminować niektóre cząsteczki uderzające z tej strony (dzięki odbiciom od dru-

giej przesłony), dlatego równanie (13.20) trzeba poprawić na nierówność:

$$r \geqslant \frac{P_C}{P_B} = r_p. \tag{13.21}$$

Między p_1 oraz u zachodzi zależność

$$p_1 = 1 - e^{-u}; \quad p_2 = e^{-u} \tag{13.22}$$

i jeżeli $u \ll 1$, to wówczas $p_1 \approx u$, a $p_2 \approx 1 - u$. W *Dodatku I* przy końcu tego rozdziału przytoczono wyprowadzenie tej zależności. Ilość informacji za okres τ otrzymujemy ze wzoru (13.21):

$$I_\tau = k[-(1-e^{-u})\ln(1-e^{-u}) + ue^{-u}], \tag{13.23}$$

który dla małych wartości u sprowadza się do postaci

$$I_\tau \approx k(-u \ln u + u) = ku(1 - \ln u), \quad u \ll 1. \tag{13.24}$$

Możemy wyliczyć wielkość, o jaką zmalała entropia ($\Delta S < 0$) wskutek wprawiania w ruch drzwiczek.

Liczba cząsteczek w naczyniu C wzrosła w czasie τ o wielkość u (tzn. o liczbę cząsteczek przepuszczonych z lewej strony na prawą), zmniejszoną o vp_1 (tj. liczbę cząsteczek przechodzących z prawej strony na lewą), ponieważ drzwiczki otwarte są tylko przez część p_1 całkowitego czasu. Zatem

$$\Delta n_\tau = u - up_1 \approx u(1 - ru), \quad u \ll 1, \tag{13.25}$$

a zmiana entropii wynosi

$$\Delta S_\tau = -k \ln\left(\frac{P_C}{P_B}\right) \Delta n_\tau = -ku(1 - ru) \ln r_p. \tag{13.26}$$

Zależność (13.26) można otrzymać z równania (12.12), jak następuje. Równanie (12.12) na entropię jednoatomowego gazu idealnego możemy zapisać w następującej postaci:

$$S = k\left[\alpha n + n \ln\left(\frac{V}{n}\right)\right],$$

gdzie α zależy od masy cząsteczek i ich energii średniej. W naszym zagadnieniu masa cząsteczek w obu naczyniach B i C jest taka sama. Jeżeli ich średnia energia początkowa jest taka sama, to ze względu na sposób, w jaki demon przenosi

cząsteczki z B do C, pozostanie nie zmieniona przez czas trwania doświadczenia. Entropia układu wynosi więc

$$S = S_B + S_C = k\left\{\alpha(n_B+n_C)+n_B \ln \frac{V_B}{n_B} + n_C \ln \frac{V_C}{n_C}\right\}.$$

Przy przechodzeniu cząsteczek z B do C objętości pozostają stałe i wówczas

$$\Delta n_C = \Delta n_\tau = -\Delta n_B,$$

tak że

$$\Delta S = k\left\{\alpha(\Delta n_B+\Delta n_C)+\Delta n_B \ln \frac{V_B}{n_B} + \Delta n_C \ln \frac{V_C}{n_C} + \right.$$

$$\left. + n_B \frac{n_B}{V_B}\left(-\frac{V_B}{n_B^2}\right)\Delta n_B + n_C \frac{n_C}{V_C}\left(-\frac{V_C}{n_C^2}\right)\Delta n_C\right\} = -k\,\Delta n_\tau \ln\left(\frac{V_B}{n_B}\frac{n_C}{V_C}\right).$$

Z dobrze znanych zależności z kinetycznej teorii gazów mamy

$$P_B = \frac{n_B \overline{v_B^2} m}{3V_B}, \qquad P_C = \frac{n_C \overline{v_C^2} m}{3V_C},$$

a ponieważ założyliśmy już, że energie średnie są te same, otrzymujemy bezpośrednio wzór (13.26).

Ze wzoru (13.26) wynika, że ubytek entropii można otrzymać tylko wtedy, gdy $ru<1$, a najbardziej korzystne warunki przyjmują postać

$$ru \ll 1; \quad \text{jeżeli} \quad u \ll 1, \quad \text{to} \quad r_p \approx r. \tag{13.27}$$

Oznacza to, że liczba cząsteczek uciekających z powrotem jest mała. Obliczmy teraz sprawność postępowania demona na II etapie przemiany (13.2):

$$\text{informacja} \rightarrow \text{negentropię}.$$

Definiujemy tę sprawność jako stosunek:

$$\varepsilon_{II} = -\frac{\Delta S_\tau}{I_\tau} = \frac{(1-ru)\ln r}{1-\ln u}. \tag{13.28}$$

Prawą stronę tego równania otrzymamy po uwzględnieniu wzorów (13.24), (13.26) i (13.27). Weźmy $ru = e^{-\beta} \ll 1$; wówczas $\beta > 1$ oraz

$$u = \frac{e^{-\beta}}{r}, \qquad \ln u = -\beta - \ln r; \tag{13.29}$$

wobec tego

$$\varepsilon_{II} = \frac{(1-e^{-\beta}) \ln r}{1+\beta+\ln r}. \tag{13.30}$$

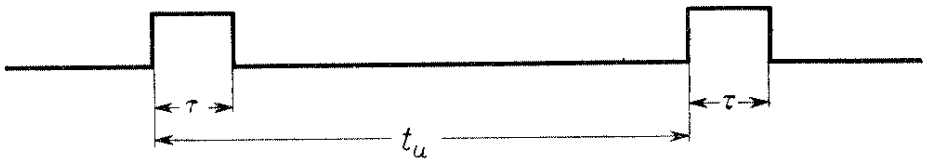
Sprawność jest zawsze mniejsza od 1 i może przyjąć wartość 1, gdy β jest duże oraz r bardzo duże. W tym przypadku u jest z pewnością bardzo małe. Okoliczność ta usprawiedliwia nasze założenia (13.24) i (13.27). Dużą sprawność na II etapie przemiany (13.2) uzyskuje się przy dużej różnicy ciśnień pomiędzy C i B.

13.5. NEGENTROPIA POTRZEBNA DO OBSERWACJI

Rozważymy teraz pierwszy etap przemiany i obliczymy ilość negentropii wymaganej do uzyskania informacji.

Aby dowiedzieć się, czy cząsteczka uderzy w drzwiczki w określonym przedziale czasu τ, powinniśmy się posłużyć ciągiem impulsów promieniowania o czasie trwania τ. Jeżeli cząsteczki nie ma, promieniowanie przechodzi przez urządzenie pomiarowe (rys. 13.1) nie będąc absorbowanym i wówczas entropia nie wzrasta. Jeżeli jest cząsteczka (lub więcej cząsteczek), promieniowanie ulega rozproszeniu i jest absorbowane przez fotoelement. Entropia wzrośnie o E/T_0, gdzie T_0 jest temperaturą układu, a E zaabsorbowaną energią.

Przyjęliśmy, że średnia liczba cząsteczek uderzających w drzwiczki (wzór (13.19)) w pewnym czasie τ wynosi u, oraz znaleźliśmy prawdopodobieństwo



Rys. 13.2. Ciąg impulsów prostokątnych o czasie trwania τ powtarzanych z okresem t_u

p_1 uderzeń (jednego lub więcej) w tym czasie (wzór (13.22)). Wynika stąd, że średni przedział czasu t_u między uderzeniami cząsteczek (lub grup uderzających cząsteczek) wynosi

$$t_u = \frac{\tau}{p_1}. \tag{13.31}$$

Jest to zarazem średni czas między kolejnymi chwilami otwierania drzwi-

czek. Trzeba więc zarejestrować za pomocą fotoelementu ciąg impulsów o czasie trwania τ i średnim odstępie t_u między impulsami (jak pokazano na rys. 13.2).

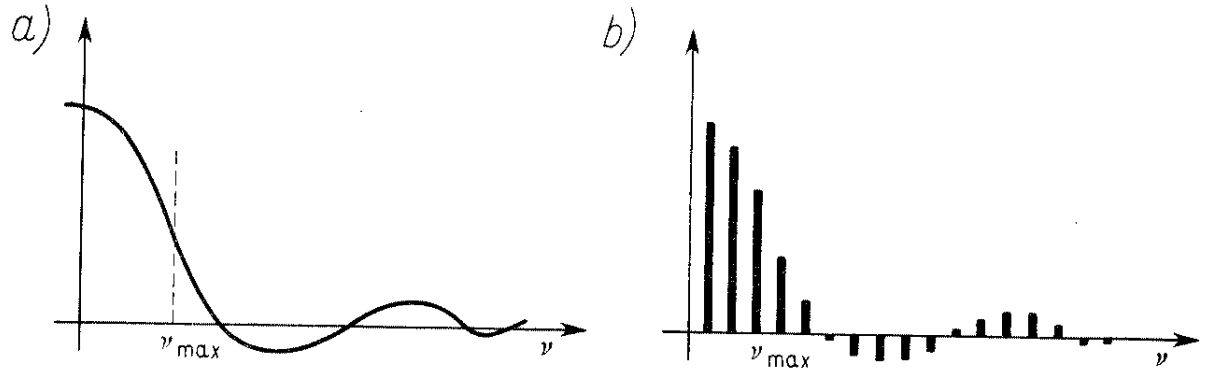
Nasuwa się pytanie: ile potrzebujemy kwantów i jak wielkie powinny one być, aby dokonywać dokładnej rejestracji cząsteczek?

Pojedynczy impuls można przedstawić całką Fouriera* (por. równania (8.25) ÷ (8.28)); odpowiada mu widmo ciągłe o efektywnej szerokości pasma

$$\nu_{\max} = \frac{1}{2\tau}. \tag{13.32}$$

Impuls zawiera zatem nieskończoną liczbę składowych harmonicznych. W klasycznej teorii każda składowa harmoniczna ma nieskończenie małą energię, ale całkowita energia impulsu pozostaje skończona. Teoria kwantowa dałaby skończoną energię $h\nu$ dla każdej składowej i całkowita energia impulsu mogłaby przyjąć wartość nieskończoną.

Inaczej przedstawia się sytuacja w przypadku ciągu impulsów o średnim okresie powtarzania t_u. Zagadnienie to rozpatrzymy przy założeniu uprasz-



Rys. 13.3. a) Widmo ciągłe pojedynczego impulsu; b) Widmo dyskretne ciągu impulsów. Obwiednia ma taką samą postać jak widmo ciągłe

czającym, przyjmując, że czas powtarzania t_u jest niezmienny, co odpowiada założeniu modelu periodycznego. Ciąg impulsów można rozłożyć w szereg Fouriera o widmie dyskretnym zawierającym tylko pewne częstotliwości, jak

* Bardziej ogólne rozważania przytoczono w *Dodatku II* na końcu tego rozdziału. Zastąpienie pojedynczego impulsu (całka Fouriera) ciągiem impulsów (szereg Fouriera) jest odwróceniem rozważań z rozdz. 8.3 przy wprowadzaniu całek Fouriera.

to pokazano na rys. 13.3. Częstotliwościami tymi są:

$$0, \nu_u, 2\nu_u, \ldots \text{ aż do } n\nu_u, \qquad \nu_u = \frac{1}{t_u} = \frac{p_1}{\tau}. \tag{13.33}$$

Maksymalna częstotliwość musi leżeć dostatecznie wysoko, aby można było wytworzyć wąskie impulsy o czasie trwania τ. Wypływa stąd warunek podobny do (13.22), który w przybliżeniu można zapisać:

$$\nu_{\max} = n\nu_u = \frac{1}{2\tau}. \tag{13.34}$$

Należy podkreślić, że uproszczonym modelem periodycznym posługiwać się możemy tylko tak długo, dopóki t_u jest dłuższe od 2τ, tak iż przerwa $(t_u - \tau)$ nie jest krótsza od czasu trwania impulsu τ. W ogólności otrzymujemy $n+1$ składowych harmonicznych łącznie z częstotliwością zerową:

$$n \approx \frac{1}{2\tau\nu_u} = \frac{1}{2p_1} = \nu_{\max} t_u. \tag{13.34a}$$

Rozważmy przykład zakładając, że amplituda wyrazu zerowego wynosi $A/2$ i A dla innych częstotliwości. Załóżmy dalej, że harmoniczne zawierają tylko składowe cosinusoidalne (wyrazy z $\sin k\varphi = 0$). Wypadkowy sygnał można zapisać w postaci

$$f(t) = A\frac{1}{2} + \cos\varphi + \cos 2\varphi + \ldots + \cos n\varphi = A\frac{\sin\left(n+\frac{1}{2}\right)\varphi}{2\sin\frac{1}{2}\varphi}. \tag{13.35}$$

Prawa strona równości wynika z tożsamości Lagrange'a, którą wykorzystaliśmy już we wzorze (8.61), gdzie

$$\varphi = 2\pi\nu_u \qquad t = 2\pi t/t_u.$$

Otrzymujemy przeto ciąg impulsów rozmytych (już nieprostokątnych), jak pokazano na rys. 13.4. Kształt ich opisany jest przez wzór (13.35). Maksymalna amplituda wynosi $(n+{}^1/_2)A$ dla $t = 0$, a pierwsze zera po obu stronach występują przy

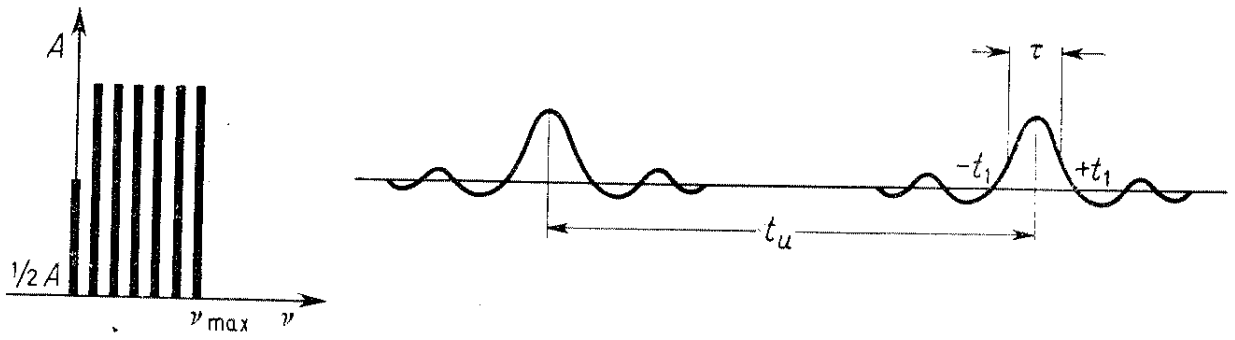
$$\pm n + \frac{1}{2}\varphi_1 = \pi, \qquad \pm t_1 = \frac{t_u}{2n+1}.$$

Całkowity czas, który upływa od jednego zerowania się impulsu do następ-

nego, wynosi $2t_1$, lecz efektywny lub nominalny czas trwania impulsu (rozpatrywany w rozdz. 8.4 i 8.5) odpowiada w przybliżeniu połowie tego czasu. Jest to więc czas, w ciągu którego amplituda jest rzeczywiście duża. Definiujemy więc

$$\tau = t_1 = \frac{t_u}{2n+1} \qquad \text{lub} \qquad n + \frac{1}{2} = \frac{t_u}{2}. \tag{13.36}$$

Jest to relacja, która zastępuje przybliżony warunek (13.34a); t_u jest okresem powtarzania impulsów.



Ryc. 13.4. Ciąg impulsów rozmytych opisanych równaniem (13.35) o efektywnym czasie trwania τ powtarzanych z okresem t_u oraz ich widmo Fouriera

Średnią energię impulsu obliczymy ze wzoru Parsevala (wzór (8.12)):

$$\overline{f^2} = \frac{1}{t_u}\int_0^{t_u} f^2\,\mathrm{d}t = \frac{1}{4}A^2 + \sum_1^n \frac{1}{2}A^2 = \frac{1}{2}A^2\left(n+\frac{1}{2}\right).$$

Energia impulsu składa się z wyrazu zerowego w kwadracie oraz wyrazów $^1/_2A^2$, które pochodzą od $A^2\cos^2 m\varphi$.

Moc średnia (energia na sekundę) ciągu impulsów przedstawiona jest przez $\overline{bf^2}$, gdzie współczynnik liczbowy b zależy od wyboru jednostek.

$$\overline{P} = \overline{bf^2} = \frac{1}{2}\left(n+\frac{1}{2}\right)bA^2. \tag{13.37}$$

Prawie cała energia skoncentrowana jest w krótkich impulsach o szerokości τ. Można to łatwo sprawdzić. Największa amplituda $(n+^1/_2)A$ (przy $t=0$) odpowiada mocy szczytowej

$$P_M = b(n+^1/_2)^2A^2 = 2(n+^1/_2)\overline{P},$$

zgodnie zaś ze wzorami (13.37) i (13.36) mamy

$$P_M\tau = \overline{P}t_u, \tag{13.37a}$$

co dowodzi, że energia w impulsie jest równa całkowitej energii za okres t_u.

Musimy teraz prześledzić pobudzenie termiczne różnych stopni swobody, ale wpierw założymy, że wszystkie częstotliwości są stosunkowo niskie, tj.

$$h\nu < kT,$$

gdzie T jest temperaturą układu. Ze wzoru (13.34) otrzymaliśmy $n+1$ stopni swobody za okres czasu t_u i wynik ten zgadza się z rozważaniami przytoczonymi w rozdz. 8 (wzór (8.57)). Wyraz zerowy ma średnią energię $^1/_2kT$, a każda z n składowych harmonicznych posiada średnią energię kT. W sumie, zgodnie ze wzorem (13.36) średnia energia termiczna za okres t_u wynosi

$$\overline{E}_T = \left(n + \frac{1}{2}\right)kT = kT\frac{t_u}{2\tau}. \tag{13.38}$$

Energia ta jest równomiernie rozłożona w całym okresie t_u.

Prawdopodobieństwo zaabsorbowania światła przez fotokomórkę jest proporcjonalne do energii dostępnej w każdym podprzedziale czasu. W czasie trwania impulsu τ, zgodnie ze wzorami (13.37a) i (13.38), jej wielkość wynosi

$$E_1 = P_M\tau + \overline{E}_T\frac{\tau}{t_u} = \overline{P}t_u + \frac{1}{2}kT.$$

W pozostałym czasie $(t_u - \tau)$ mamy

$$E_2 = \overline{E}_T(t_u - \tau)/t_u = \frac{1}{2}kT[(t_u/\tau) - 1].$$

Składowa E_2 będzie powodować fałszywe absorpcje, które wpływają na niepożądane wprawianie w ruch drzwiczek. Aby w pracy układu zmniejszyć błędy tego rodzaju, przyjmujemy

$$E_1 = \alpha E_2, \quad \alpha \gg 1; \tag{13.39}$$

stąd

$$\overline{P}t_u + \frac{1}{2}kT = \frac{1}{2}\alpha kT[(t_u/\tau) - 1].$$

Ponieważ α jest duże, zaś $\frac{\tau}{t_u}$ małe, to warunek ten sprowadza się do postaci

$$\overline{P} \approx \alpha k T / 2\tau \tag{13.40}$$

i wtedy $\overline{P} t_u$ oznacza energię promieniowania zaabsorbowanego w czasie t_u w fotokomórce. Średnia energia zaabsorbowana w czasie τ wynosi $\overline{P}\tau$ i odpowiednio średnia entropia wzrasta o wartość

$$\Delta S_\tau = \overline{P} \frac{\tau}{T} = \frac{1}{2} \alpha k. \tag{13.41}$$

Wszystkie przybliżenia opierają się na założeniu, że

$$t_u \gg \tau \quad \text{oraz} \quad u \ll 1.$$

Możemy teraz omówić sprawność postępowania demona w pierwszym etapie, kiedy negentropia zmieniana jest w informację. Ze wzorów (13.18) i (13.24) otrzymujemy sprawność za okres τ

$$\varepsilon_{\mathrm{I}} = \frac{I_\tau}{\Delta S_\tau} \leqslant 2(-p_1 \ln p_1 - p_2 \ln p_2)/\alpha \approx 2u(1 - \ln u)/\alpha. \tag{13.42}$$

Ostatnie wyrażenie jest słuszne dla bardzo małych wartości u. Jeżeli u zbliża się do jedności, wówczas rozważania są bardzo trudne. Sprawność ε_{I} może w tym przypadku być wyższa. Dla małych wartości u sprawność staje się coraz mniejsza.

Podsumowując otrzymane wyniki widzimy, że:

ε_{I}, tj. sprawność w I etapie osiąga maksimum, gdy $u \approx 1$,
$\varepsilon_{\mathrm{II}}$, tj. sprawność w II etapie osiąga maksimum, gdy u jest bardzo małe.

Sprawność całkowitą obliczymy posługując się wzorami (13.28) i (13.42). Daje to

$$\varepsilon_0 = \varepsilon_{\mathrm{I}} \cdot \varepsilon_{\mathrm{II}} = \frac{2u(1 - ru) \ln r}{\alpha}. \tag{13.43}$$

Tak jak we wzorze (13.29) bierzemy

$$ru = \mathrm{e}^{-\beta} = \eta < 1$$

231

$$\varepsilon_0 = \frac{2\eta(1-\eta)\ln r}{\alpha r} \leqslant \frac{\ln r}{2r\alpha}, \tag{13.44}$$

tością optymalną jest $\eta = {}^1/_2$. Ostatnie wyrażenie przyjmuje maksimum dla $r = e$, dając $u = \frac{\eta}{r} = \frac{e}{2}$. Jest to niewielka wartość, dla której wszystkie nasze przybliżenia są słuszne:

$$\varepsilon_{0\max} = \frac{1}{2e\alpha} \approx \frac{1}{5{,}44\alpha}, \qquad \alpha \gg 1.$$

Całkowita sprawność spada do zera, gdy $r = 1$ (równe ciśnienia w B i C), i również wtedy, gdy r jest bardzo duże (duże ciśnienie w C). Wykazaliśmy w ten sposób, że całkowita sprawność zawsze jest mniejsza od jedności, jak również, że sprawności ε_{I} oraz $\varepsilon_{\mathrm{II}}$ są mniejsze od jedności, nawet w anormalnych okolicznościach, gdy stosunek ciśnień r jest bardzo duży.

13.6. ZAGADNIENIE SZILARDA: DOBRZE INFORMOWANA MASZYNA CIEPLNA

W roku 1929 Szilard* opublikował ważny artykuł dotyczący zagadnienia demona Maxwella i jako pierwszy wykrył powiązanie między informacją a entropią. Była to rzeczywiście pionierska praca. Znaczenia jej nie doceniano przez długi okres czasu, dopiero obecny rozwój teorii wysunął ją na pierwszy plan.

Rozpatrzymy teraz niektóre ciekawsze zagadnienia poruszone przez Szilarda i omówimy nową pracę Gabora* poruszającą też te zagadnienia.

Szilard rozważa następujący przypadek: zamknięty cylinder o objętości V może być podzielony na dwie części V_1 i V_2 przez umieszczenie przegrody w pewnej pozycji. W cylindrze znajduje się tylko jedna cząsteczka. W chwili gdy obserwator wsuwa przegrodę, może on w jakiś sposób dowiedzieć się, gdzie znajduje się cząsteczka: w V_1 czy V_2. Przypuśćmy, iż jest ona w V_1. Przegroda przesuwa się tak, jak tłoczek wewnątrz cylindra i objętość V_1 powiększa się powoli do wartości początkowej V, podczas gdy cały układ pozostaje w stałej

* L. Szilard, *op. cit.*

* D. Gabor, *op.* cit.

temperaturze T utrzymywanej przez termostat. W czasie tego powolnego ruchu cząsteczka wielokrotnie zderzać się będzie z tłoczkiem i zderzenia te wywołają średnie ciśnienie, podobne do ciśnienia gazu idealnego. Zostaje przy tym wykonana pewna praca. Następnie wysuwamy przegrodę umieszczając ją znowu w położeniu wyjściowym, a cała operacja może być powtórzona od nowa. Układ ten wykonuje pracę mechaniczną. Układ pracuje przy jednej temperaturze, lecz wymagana jest informacja o położeniu cząsteczki. Naszym zadaniem jest omówić powiązania informacji ze zmianą entropii. Wprowadzona przez nas definicja ilości informacji i wyniki uzyskane na tle poprzednich przykładów umożliwią uproszczenie i dokładniejsze sprecyzowanie rozważania.

Kolejność etapów w schemacie Szilarda jest następująca:

(1) Ustawiamy przegrodę w pewnej pozycji.

(2) Wykrywamy, czy cząsteczka znajduje się w V_1, czy w V_2.

(3) Przesuwamy tłoczek w górę lub w dół — zgodnie z punktem B.

Pierwszy etap nie wymaga omówienia. W punkcie B musimy wymyśleć przyrząd umożliwiający zlokalizowanie cząsteczki w czasie doświadczenia. Możemy posłużyć się wiązką światła B_1 przechodzącą przez V_1 i drugą wiązką B_2 przechodzącą przez V_2. Fotokomórki C_1 i C_2 mogą reagować na światło rozproszone przechodzące z V_1 lub V_2. Wykorzystajmy najpierw wiązkę promieniowania B_1. Jeżeli pojawi się światło rozproszone w C_1, to dowodzi to, iż cząsteczka znajduje się w V_1. Jeżeli fotoelement C_1 nie wykryje światła rozproszonego, to jeszcze niczego to nie dowodzi:

cząsteczki może nie być w V_1;

cząsteczka może nie rozpraszać światła;

kwant światła rozproszonego omija fotokomórkę C_1.

Powtarzamy następne doświadczenie z wiązką B_2 i V_2. Ostatecznie kwant $h\nu$ zostaje rozproszony w obszarze V_1, bądź V_2 i w ten sposób lokalizujemy cząsteczkę.

Absorpcja $h\nu$ w jednej z fotokomórek odpowiada wzrostowi entropii zgodnie ze wzorem (13.7):

$$\Delta S = \frac{h\nu}{T} \geqslant k. \tag{13.45}$$

Dokładniejsze rozważania (wzór (14.17)) nieznacznie obniżą granicę, dając

$$\Delta S \geqslant k \ln 2 = 0{,}7\, k. \tag{13.45a}$$

Obliczamy teraz ilość uzyskanej informacji. Prawdopodobieństwo p_1 znalezienia cząsteczki w V_1 jest oczywiście równe $\frac{V_1}{V}$; p_2 jest prawdopodobieństwem tego, że cząsteczka znajduje się w V_2, przy czym $p_2 = \frac{V_2}{V}$:

$$p_1 = \frac{V_1}{V}; \quad p_2 = \frac{V_2}{V}; \quad p_1 + p_2 = 1.$$

Rozważmy oba przypadki oddzielnie. Możemy wykorzystać wzory (1.5) i (1.6) oraz obliczyć ilość informacji dla przypadków:

$$\left.\begin{aligned} &\text{cząsteczka w } V_1\text{: ilość informacji } i_1 = k \ln \frac{V}{V_1} = -k \ln p_1; \\ &\text{cząsteczka w } V_2\text{: ilość informacji } i_2 = k \ln \frac{V}{V_2} = -k \ln p_2. \end{aligned}\right\} \tag{13.46}$$

Średnia ilość informacji przypadająca na jedną operację

$$\bar{I} = p_1 i_1 + p_2 i_2 = -k(p_1 \ln p_1 + p_2 \ln p_2) > 0. \tag{13.46a}$$

Wzór ten odpowiada wzorowi Shannona (2.1). Każda pojedyncza operacja może dostarczyć ilości informacji mniejszej lub większej od przyrostu entropii ΔS wymaganego do uzyskania tej ilości informacji, lecz średnia ilość informacji wyrażona przez wzór (13.46a) jest mniejsza od średniego przyrostu entropii za operację:

$$\overline{\Delta S} \geqslant \bar{I}. \tag{13.47}$$

Znak równości zachodzi wtedy, gdy objętości V_1 i V_2 są równe:

$$V_1 = V_2 = {}^1/_2 V; \quad p_1 = p_2 = {}^1/_2; \quad \bar{I} = k \ln 2,$$

co odpowiada maksimum średniej ilości informacji. Mogą wystąpić fluktuacje ilości informacji uzyskiwanej w poszczególnych operacjach. Jednak w wielu próbach średnia ilość informacji jest mniejsza niż zapłacona za nią cena wyrażona negentropią ΔN.

Uogólniona zasada Carnota z rozdz. 12.2 jest zachowana dla wartości średnich, przy czym występują dodatnie lub ujemne fluktuacje

$$\Delta N = -\Delta S, \quad \Delta(N+I) \leqslant 0. \tag{13.47a}$$

Omówimy teraz zmiany entropii gazu i ostatnią operację naszej maszyny. Entropia gazu idealnego (wzór (12.12)) wynosi

$$S = S_0 + k \ln V.$$

W chwili wsuwania przegrody otrzymujemy pewien ubytek entropii:

$$\left.\begin{aligned} \Delta\sigma_i = k \ln\left(\frac{V_1}{V}\right) = k \ln p_1 < 0, \text{ jeżeli cząsteczka jest w } V_1, \\ \Delta\sigma_2 = k \ln p_2 < 0, \text{ jeżeli cząsteczka znajduje się w } V_2. \end{aligned}\right\} \quad (13.48)$$

Prawdopodobieństwa obu możliwości wynoszą odpowiednio p_1 i p_2. Widzimy, że ubytek entropii ($-\Delta\sigma_1$ lub $-\Delta\sigma_2$) jest zawsze równy poszczególnym ilościom informacji i_1 lub i_2, ponieważ wyrażenia (13.46) i (13.48) są identyczne z wyjątkiem znaku. Co więcej, średnia entropia gazu przypadająca na operację maleje o

$$\overline{\Delta\sigma} = p_1\Delta\sigma_1 + p_2\Delta\sigma_2 = -I. \quad (13.49)$$

Gdy przesuwamy tłoczek i objętość wzrasta do początkowej wartości V, to odzyskujemy początkową wartość entropii.

Wykonana praca W równa jest ciepłu Q pobranemu z zasobnika ciepła (przyjmując powolny odwracalny proces):

$$W = Q = \int_{V_1}^{V} p \, dv = kT \int_{V_1}^{V} \frac{dV}{V} = kT \ln\left(\frac{V}{V_1}\right); \quad pV = kT,$$

co otrzymuje się wykorzystując prawo Boyle'a dla jednej cząsteczki i biorąc pod uwagę przypadek (1), tj. cząsteczkę w objętości V_1. Przyrost entropii za okres powiększania objętości wynosi

$$\frac{Q}{T} = k \ln\left(\frac{V}{V_1}\right) = -\Delta\sigma_1 > 0.$$

Analogicznie przedstawia się sprawa, jeżeli cząsteczka znajduje się w V_2. Ostatnie obliczenie potwierdza słuszność wzoru (13.48) i pokazuje, jak otrzymuje się początkowy ubytek entropii. W ogólności przyrost entropii netto wynosi $\Delta S - I$ (wzór (13.47)) za cały cykl przemiany.

Udowodniliśmy przeto przykładowo, że ilość informacji odpowiada ujemnej entropii (wzory (13.47) i (13.48)), jednakże musieliśmy uważać na fluktuacje, ponieważ założyliśmy, że wpierw ustawiliśmy przegrodę nie znając poprzedniego położenia cząsteczki.

13.7. ROZWAŻANIA GABORA

Gabor* badał podobne zagadnienie, lecz przyjął inny sposób funkcjonowania maszyny. Odwraca on kolejność operacji (1) i (2) i postępuje w sposób następujący:

(a) Upewnia się, że cząsteczka znajduje się w V_1 (u Szilarda etap (2)).

(b) Umieszcza przegrodę (u Szilarda etap (1)).

(c) Przesuwa tłoczek.

Gabor zakłada, że wiązka światła oświetlająca V_1 wykorzystywana jest do uzyskania informacji. Jego aparaturę pokazano na rys. 13.5. Pojedyncza cząsteczka porusza się w cylindrze V utrzymywanym w stałej temperaturze T. Dolna część cylindra V_1 ma przezroczyste ścianki (rys. 13.5) i oświetlona jest przez światło o pewnej częstotliwości ν pochodzące z rozżarzonego włókna. Układ zwierciadeł i soczewek utrzymuje obieg światła bez pochłaniania, dopóki pewna ilość światła nie zostanie rozproszona przez cząsteczkę i zaabsorbowana w fotoelemencie. Jeżeli to zajdzie, przekaźnik wprawia w ruch mechanizm: tłoczek rozdzielający wsuwa się bez tarcia w cylinder, natomiast dwa zwierciadła opadają w dół zamykając drogę promieniom świetlnym. Cząsteczka stopniowo podnosi tłoczek i wykonuje pracę mechaniczną.

Po wprowadzeniu tłoczka entropia zmaleje o wartość

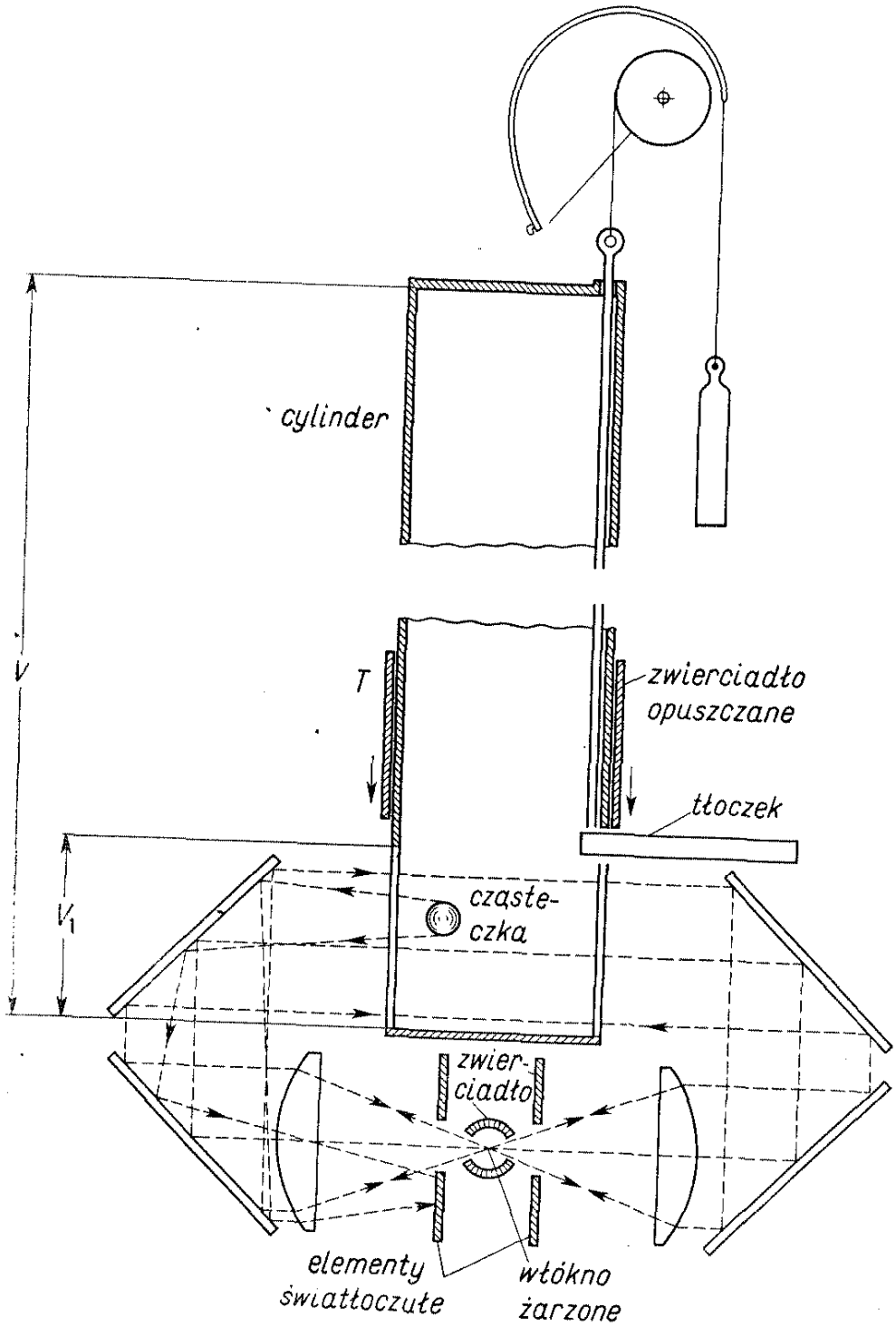
$$\Delta S_1 = k \ln\left(\frac{V_1}{V}\right) = k \ln p_1 < 0. \tag{13.50}$$

W cyklu rozprężania następuje wzrost entropii, który pozwala przywrócić jej pierwotną wartość.

Po zakończeniu rozprężania się V_1 tłoczek wysuwa się w bok i zajmuje pozycję wyjściową. Rysunek tego nie pokazuje. Entropię (wzór (13.50)) można uczynić tak małą, jak chcemy, zwiększając objętość V. Trudność, jaką tu napotykamy, polega na tym, by udowodnić, że ilość światła wymagana podczas obserwacji wzrasta również wraz z objętością V w taki sposób, że początkowy przyrost entropii musi być zawsze większy od obliczonego przyrostu $|\Delta S_1|$.

Rozważmy pierwszą część operacji, gdy oczekujemy, aby cząsteczka weszła do V_1 i rozproszyła światło. Średni czas oczekiwania t_u jest proporcjo-

* D. Gabor, *op. cit.*



Rys. 13.5. Maszyna Gabora. Światło przechodzi przez V_1 tak długo, aż pojawi się cząsteczka, która rozproszy foton wykrywalny przez jeden z elementów światłoczułych. Jeśli to nastąpi, przekaźnik wsuwa tłoczek do cylindra, opadające zwierciadła przesłaniają promienie świetlne, cząsteczka zaś zaczyna podnosić tłoczek

nalny do V, natomiast czas trwania błysku światła rozproszonego wynosi τ i proporcjonalny jest do V_1. Wobec powyższego nasza obserwacja musi dawać dwie składowe informacji:

(1) Cząsteczka znajduje się w V_1, zamiast pierwotnej informacji, że znajduje się w V.

(2) Zachodzi to w przedziale czasu τ, będącego częścią całkowitego czasu t_u.

Jest to odmienna sytuacja od tej, która występuje na rys. 13.1, gdzie położenie cząsteczki określone było przez geometrię urządzenia, natomiast niezbędne obserwacje dotyczyły czasu.

Możemy obliczyć całkowitą ilość wymaganej (w naszym przykładzie) informacji korzystając ze wzoru (12.1). Mamy

$$I = k \ln\left(\frac{P_0}{P_1}\right),$$

gdzie P_0 równe jest liczbie równoprawdopodobnych możliwości w przypadku ogólnym i P_1 równe jest liczbie równoprawdopodobnych możliwości, gdy dana jest ilość informacji I.

Oczywiście, mamy

$$\frac{P_0}{P_1} = \frac{t_u}{\tau}\frac{V}{V_1} = \left(\frac{V}{V_1}\right)^2,$$

ponieważ

$$\frac{t_u}{\tau} = \frac{V}{V_1}. \tag{13.51}$$

Ilość informacji w tego rodzaju eksperymencie wyniesie

$$I = k \ln\left(\frac{V}{V_1}\right)^2 = 2k \ln\left(\frac{V}{V_1}\right) = -2k \ln p_1. \tag{13.52}$$

Ilość informacji I jest dwukrotnie większa od przyrostu entropii ΔS_1 (wzór (13.50)), który mierzy negentropię otrzymywaną przy pomocy informacji I. Sprawność przemiany informacji w negentropię wynosi $^1/_2$. Obliczamy następnie przyrost entropii potrzebnej do otrzymania tej informacji. Na całym odcinku L (proporcjonalnym do V) musimy wytworzyć wiązkę promieniowania świetlnego pokrywającego tylko odcinek L_1 (proporcjonalną do V_1).

Można tego dokonać na drodze superpozycji pewnej ilości drgań własnych o długości L.

Właściwej formie drgań odpowiadają długości fali

$$2L,\ L,\ \frac{2}{3}L,\ \ldots,\ \frac{2L}{M}$$

i aby uzyskać dobry pomiar długości L_1, najkrótsza fala musi być rzędu $2L_1$. W takim razie mamy

$$M = \frac{L}{L_1} = \frac{V}{V_1}. \tag{13.53}$$

Jest to liczba drgań własnych przypadająca na objętość V.

Każdej formie tych drgań odpowiadają błyski o czasie trwania τ powtarzane w odstępach t_u. Sytuacja jest zbliżona do poruszonej w rozdz. 13.5 (wzory (13.39) oraz (13.41)) i wymaga przyrostu entropii na drganie

$$\frac{\alpha k t_u}{2\tau} \text{ przy } t_u \gg \tau \text{ oraz } \alpha \gg 1.$$

Ostatecznie znajdujemy

$$\Delta S = \frac{M\alpha k t_u}{2\tau} = \frac{\alpha k}{2}\left(\frac{V}{V_1}\right)^2. \tag{13.54}$$

Jest to całkowita entropijna* cena obserwacji. Porównajmy ten wynik ze wzorem (13.52) Otrzymamy

$$\Delta S - I = k\left[\left(\alpha\frac{F}{2}\right) - \ln F\right] > 0, \qquad \alpha \gg 1, \tag{13.55}$$

gdzie

$$F = \left(\frac{V}{V_1}\right)^2 = \left(\frac{1}{p_1}\right)^2.$$

Wyrażenie to jest zawsze dodatnie. Gdy p_1 maleje do zera, F wzrasta do nieskończoności, wobec tego wyrażenie (13.55) rośnie do nieskończoności. Jest to wyjaśnienie zagadnienia Szilarda. Dalsze szczegóły można znaleźć w rozdz. 15.

---

* Do oznaczenia kosztu obserwacji wyrażanego zmianą entropii używać będziemy terminu: entropijna cena obserwacji. (*Przyp. tłum.*)

DODATEK I DO ROZDZIAŁU 13

Wykorzystując rozważania H. Jacobsona chcemy dowieść słuszności wzoru (13.22). Bierzemy bardzo krótki przedział czasu θ, w ciągu którego nie może zajść więcej niż jedno zderzenie cząsteczki z powierzchnią A.

Prawdopodobieństwo jednego zderzenia: $\alpha\theta \ll 1$;

Prawdopodobieństwo tego, że zderzenia nie ma: $1-\alpha\theta$.

Rozważmy teraz przedział czasu $\tau = m\theta$. Otrzymujemy

$\pi_0 = (1-\alpha\theta)^m$ — prawdopodobieństwo tego, że nie nastąpi zderzenie cząsteczki z powierzchnią A w czasie τ,

$\pi_1 = m\alpha\theta(1-\alpha\theta)^{m-1}$ — prawdopodobieństwo tego, że nastąpi jedno zderzenie cząsteczki z powierzchnią A w jakimś przedziale m o długości θ,

$\pi_2 = \frac{m(m-1)}{2}(\alpha\theta)^2(1-\alpha\theta)^{m-2}$ — prawdopodobieństwo tego, że nastąpią dwa zderzenia w czasie τ,

$\pi_n = \frac{m!}{n!(m-n)!}(\alpha\theta)^n(1-\alpha\theta)^{m-n}$ — prawdopodobieństwo tego, że nastąpi n zderzeń w czasie τ.

Oczywiście

$$\pi_0+\pi_1+\ldots+\pi_n+\ldots+\pi_m = [\alpha\theta+(1-\alpha\theta)]^m = 1.$$

Jeżeli θ staje się bardzo małe, natomiast m bardzo duże, to

$$\pi_0 = (1-\alpha\theta)^m = (1-\alpha\theta)^{\frac{\alpha\tau}{\alpha\theta}} \to e^{-\alpha\tau}.$$

Zauważmy, że π_0 pokrywa się z wyrażeniem na p_2 podanym we wzorze (13.22), natomiast $\alpha\tau = u$. Dowiedliśmy więc w ten sposób słuszności wzoru (13.22). Następnie

$$p_1 = \pi_1+\pi_2+\ldots+\pi_m = 1-e^{-\alpha\tau} = 1-e^{-u}.$$

Jest to prawdopodobieństwo jednego lub więcej zderzeń cząsteczki z powierzchnią A.

DODATEK II DO ROZDZIAŁU 13

Możemy uogólnić rozumowanie, z którego skorzystaliśmy we wzorze (13.33). Załóżmy, że w miejsce impulsów odległych od siebie o równą wartość czasu, m impulsów jest przypadkowo rozłożonych na bardzo długim odcinku czasu t:

$$t = \frac{m\tau}{p_1},$$

gdzie $m \gg 1$ oraz $p_1 =$ prawdopodobieństwu wystąpienia aktywnego przedziału τ. [Przez przedział aktywny rozumiemy ten odcinek czasu τ, w którym nastąpi zderzenie cząsteczki z powierzchnią A.]

Zakładamy również, że ciąg tych impulsów powtarzany jest w następujących po sobie przedziałach o długości t, co periodyzuje proces z długim okresem t. Ten okresowy rozkład ma widmo o częstotliwościach 0, ν_0, $2\nu_0$, ..., $n\nu_0$ (wzory (13.34) ÷ (13.36)), gdzie

$$\nu_0 = \frac{1}{t} = \frac{p_1}{m\tau}, \qquad n\nu_0 = \nu_{\max} \approx \frac{1}{2\tau}.$$

Maksymalna wartość n wynosi

$$n = \frac{1}{2\tau\nu_0} = \frac{m}{2p_1}.$$

Dyskusja przytoczona przy wzorach (13.37) ÷ (13.41) odnosi się bezpośrednio do tego przypadku, dając średni przyrost entropii w czasie τ

$$\Delta S_\tau = \alpha k/2,$$

jak we wzorze (13.41).

14

NEGENTROPIJNA ZASADA INFORMACJI W FIZYCE OGÓLNEJ

14.1. ZAGADNIENIE POMIARÓW W FIZYCE

Do tej pory mieliśmy możność wyciągnąć kilka ściśle sprecyzowanych wniosków, które można podsumować, jak następuje:

(1) Informacja może być zmieniona w negentropię i odwrotnie. Jeżeli przemiana jest odwracalna, to straty nie występują, jeżeli zaś przemiana jest nieodwracalna, to zawsze występują straty.

(2) Dowolne doświadczenie dające informację o układzie fizycznym, średnio rzecz biorąc, prowadzi do wzrostu entropii układu lub jego otoczenia. Średni przyrost entropii jest zawsze większy (lub równy) ilości uzyskanej informacji. Innymi słowy, za informację płaci się negentropią, ale cena płacona będzie zawsze nie mniejsza od ilości uzyskanej informacji. Obserwacja jest zawsze powiązana z przyrostem entropii, a tym samym pociąga za sobą proces nieodwracalny.

(3) Najmniejsza możliwa ilość negentropii wymaganej podczas obserwacji jest rzędu k, a szczegółowe rozważania dają

$$\textit{minimum negentropii} = k \ln 2 = 0{,}7k \approx$$
$$\approx 10^{-16}\ ^{\circ}\mathrm{K} \text{ (w układzie CGS)}. \tag{14.1}$$

Jest to dokładna wartość graniczna, która zgadza się z wcześniej przytoczonym rozważaniem Szilarda. W jednostkach binarnych minimum to oznacza wartość jednego bita.

(4) Wnioski (1)–(3) prowadzą do wytłumaczenia zagadnienia demona Maxwella, który jest niczym innym, jak tylko urządzeniem zamieniającym negentropię w informację i odwrotnie. Chcemy teraz bardziej szczegółowo zbadać punkty (2) i (3). Prowadzi to do rozważań na temat pewności i dokładności doświadczenia fizycznego.

Czytelnik może zauważyć, że musieliśmy wprowadzić dowolny współczynnik α (por. wzory (13.39) i (13.54)), o którym założyliśmy, że jest dostatecznie duży, aby zredukować błędy obserwacji oraz zmniejszyć ilość niepożądanych zadziałań układu.

Współczynnik α zabezpiecza *pewność** urządzenia doświadczalnego. Zagadnienie pewności doświadczenia jest bardzo ważne i omówimy je szczegółowo. Zajmiemy się również zagadnieniem ustalenia minimum negentropii wymaganej w czasie obserwacji oraz dowiedziemy słuszności wyrażenia (14.1).

Wnioski z punktów (2) i (3) podano już przy omawianiu demona Maxwella. Obserwator, fizyk, bądź demon wymagają zawsze źródeł negentropii, gdyż każda obserwacja opłacana jest negentropią z otoczenia. Potrzebne są więc baterie, zasilacze, sprężone gazy itp. — wszystko to, co przedstawia źródła negentropii. Potrzebne jest również w laboratorium światło, aby można było odczytać wskazania amperomierzy lub innych przyrządów.

Jaka jest najmniejsza możliwa ilość negentropii potrzebnej do dokonania obserwacji? Przypuśćmy, że chcemy odczytać wskazanie amperomierza utrzymywanego w temperaturze T. Wskazówka i sprężynka przedstawiają układ drgający o bardzo małej częstotliwości własnej. Wskazówka porusza się ruchem drgającym Browna o średniej energii kinetycznej $\frac{1}{2}kT$ i ze średnią energią całkowitą

$$\overline{E}_t = kT. \tag{14.2}$$

G. Ising** przyjął, że do poprawnego odczytu wymagana jest energia

$$E_I \geqslant 4\overline{E}_t = 4kT. \tag{14.3}$$

Energia ta po dokonaniu odczytu rozproszona będzie na tarcie, efekt Joule'a,

---

* W literaturze polskiej przyjęto termin *niezawodność*. Ponieważ w tym przypadku ma to inne znaczenie, gdyż niezawodność *(reliability)* odnosi się do urządzeń, które zapewniają pewność obserwacji — tutaj używać będziemy terminu: *pewność*. (*Przyp. tłum.*).

** G. Ising, Phil. Mag. **[6] 51** (1926), str. 827.

M. Courtines, *Les Fluctuations dans les appareils de mesures*, Congr. intern. d'electricite, vol. 2, Paris 1932.

tarcie lepkie przyrządu itp. Oznacza to przyrost entropii całego układu o

$$\Delta S = \frac{E_I}{T} \geqslant 4k. \tag{14.4}$$

Według kryterium Isinga entropia rzędu $4k$ wymagana jest dla dokonania obserwacji. Wykażemy teraz, że $k \ln 2 = 0{,}7k$ jest wartością graniczną, przy której mamy 50% szans na poprawną obserwację i 50% na to, że obserwowane wskazanie przyrządu spowodowane jest pobudzeniem cieplnym.

14.2. OBSERWACJE DOKONYWANE NA OSCYLATORZE

Rozważmy teraz oscylator harmoniczny o częstotliwości ν i poziomach energetycznych $E_n = nh\nu$. Poruszone wyżej zagadnienie amperomierza odpowiada przypadkowi granicznemu — bardzo małej częstotliwości. Niektóre ogólne wnioski dotyczące statystycznych własności takiego rezonatora, utrzymanego w temperaturze T, uzyskaliśmy już poprzednio. Prawdopodobieństwo poziomu energetycznego $nh\nu$ wynosi

$$\left.\begin{aligned} p_n &= B\,\mathrm{e}^{-nx} = \mathrm{e}^{-nx} - \mathrm{e}^{-(n+1)x}, \qquad x = \frac{h\nu}{kT}; \\ \sum_{n=0}^{\infty} p_n &= \sum_{n=0}^{\infty} B\mathrm{e}^{-nx} = 1, \qquad B = 1 - \mathrm{e}^{-x}; \end{aligned}\right\} \tag{14.5}$$

stąd na prawdopodobieństwo stanów leżących powyżej $q-1$ i poniżej q otrzymujemy, odpowiednio, wyrażenia

$$\left.\begin{aligned} P_{n \geqslant q} &= p_q + p_{q+1} + \ldots = \mathrm{e}^{-qx}, \\ P_{0<n<q} &= p_0 + p_1 + \ldots + p_{q-1} = 1 - \mathrm{e}^{-qx}. \end{aligned}\right\} \tag{14.5a}$$

Jako „środkową" liczbę kwantową m bierzemy medianę* określoną przez warunek

$$P_{n<m} = P_{n \geqslant m},$$

* Zob. np. M. Fisz, *Rachunek prawdopodobieństwa i statystyka matematyczna*, Warszawa 1958. (*Przyp. tłum.*).

co daje

$$e^{-mx} = 1/2 \quad \text{lub} \quad mx = \ln 2 \approx 0{,}7,$$

tak że

$$E_m = mh\nu = kT \ln 2. \tag{14.6}$$

Medianę można porównać z wartością średnią

$$\bar{n} = \sum_{n=0}^{\infty} np_n = \frac{1}{e^x - 1}. \tag{14.7}$$

Dla małych częstotliwości

$$h\nu \ll kT \quad \text{oraz} \quad x \ll 1,$$

tak że wzór (14.7) daje

$$\bar{n}x = 1 \quad \text{lub} \quad \bar{E} = \bar{n}h\nu = kT. \tag{14.8}$$

Pozwala to wnioskować, że mediana jest mniejsza od wartości średniej. W przypadku wielkich częstotliwości sytuacja przedstawia się inaczej. Omówimy to w rozdz. 14.3.

Przystąpimy następnie do obliczenia średniej liczby kwantowej dla stanów $n \geqslant q$:

$$\bar{n}_q = \frac{\sum_{n=q}^{\infty} np_n}{P_{n \geqslant q}} = \frac{B e^{-qx} \sum_{r=0}^{\infty} (q+r) e^{-rx}}{e^{-qx}}, \quad n = q+r$$

i wykorzystując warunek (14.5) i (14.7) otrzymamy

$$\bar{n}_q = q + \bar{n}, \quad \text{ponieważ} \quad \bar{r} = \bar{n}. \tag{14.9}$$

Średnia liczba kwantowa $\overline{n_{q'}}$ zbioru stanów $n < q$ otrzymywana jest bezpośrednio z relacji

$$P_{n<q}\overline{n_{q'}} + P_{n \geqslant q}\bar{n}_q = \bar{n},$$

co daje

$$\overline{n_{q'}} = \bar{n} - \frac{q}{e^{+qx} - 1} \quad \text{dla} \quad n < q. \tag{14.10}$$

Porównajmy te wyniki z medianą $q = m$:

$$\left.\begin{aligned} \bar{n}_m &= m+\bar{n}, \quad n \geqslant m, \\ \bar{n}_{m'} &= \bar{n}-m, \quad n < m. \end{aligned}\right\} \tag{14.11}$$

Dla małych częstotliwości

$$\left.\begin{aligned} \bar{n}_m &= \bar{n}(1+\ln 2) \approx 1{,}7\bar{n}, \\ \bar{n}_{m'} &= \bar{n}(1-\ln 2) \approx 0{,}3\bar{n}. \end{aligned}\right\} \tag{14.12}$$

Możemy teraz rozpatrzyć zagadnienie dokonywania obserwacji oraz przewidywania jej pewności. Przypuśćmy, że obserwujemy w pewnej ustalonej chwili czasu t energię odpowiadającą $n \geqslant q$ kwantom. Prawdopodobieństwo tego, że ta wysoka liczba kwantowa może być uwarunkowana fluktuacjami termicznymi i wynosi $P_{n\geqslant q}$. Prawdopodobieństwo zajęcia przez rezonator stanu $n < q$ na drodze normalnej, kiedy to absorbuje on pewną ilość dodatkowych kwantów (ze źródła zewnętrznego), aby osiągnąć poziom powyżej poziomu q, wynosi $P_{n<q}$. Jeżeli przyjmiemy równe prawdopodobieństwa w obu przypadkach, to za q musimy brać medianę, tzn.

$$q = m = \frac{\ln 2}{x}.$$

Z drugiej strony, gdybyśmy przyjęli wartość energii podaną przez Isinga (wzór (14.3)) na obserwację pewną, to otrzymalibyśmy

$$\left.\begin{aligned} q_1 &= 4\frac{1}{x}, \\ (P_{n\geqslant q})/(P_{n<q}) &= \mathrm{e}^{-4}/(1-\mathrm{e}^{-4}) \approx \frac{1}{54}, \end{aligned}\right\} \tag{14.13}$$

co daje około 2-procentowe prawdopodobieństwo błędu spowodowanego pobudzeniem termicznym rezonatorów.

W chwili t obserwujemy więcej niż q kwantów w rezonatorze. Jeżeli uwarunkowane są one fluktuacjami cieplnymi, to średni nadmiar energii pochodzi z otaczającego rezonator termostatu T i zgodnie ze wzorem (14.9) średnia energia w tym przypadku wyniesie

$$\bar{E} = \bar{n}qh\nu = qh\nu + \bar{n}h\nu. \tag{14.14}$$

Jeżeli q kwantów absorbowanych pochodzi rzeczywiście ze źródła zewnętrznego, to nadmiar energii $qh\nu$ przychodzi z tego źródła i dodaje się do normal-

nej energii średniej $\bar{n}hv$ dając znów wzór (14.14). Nadmiar energii qhv rozprasza się później na skutek tarcia, lepkości czy też strat omowych i jest absorbowany przez termostat o temperaturze T.

Jeżeli równanie (14.14) odzwierciedla efekt fluktuacji, to mówi ono o tym, że termostat odzyskał część energii, którą utracił poprzednio. Jeżeli równanie (14.14) odzwierciedla efekt absorpcji energii ze źródła zewnętrznego, to mówi ono, że wówczas termostat absorbuje q kwantów i zatem przyrost entropii wyniesie

$$\Delta S = qhv/T = qkx = \begin{cases} k \ln 2, & \text{jeżeli } q = m, \\ 4k, & \text{według Isinga.} \end{cases} \tag{14.15}$$

Ten przyrost entropii otaczającego układ termostatu jest ceną, jaką płaci się za obserwację. Najmniejsza jej wartość wynosi $k \ln 2$ lub $0{,}7k$ i otrzymujemy ją wtedy, gdy dopuszczamy prawdopodobieństwo błędu 50%.

Jest bardzo ciekawe to, że nawet gdy wybierzemy granicę $m < \bar{n}$ (por. wzory (14.6) i (14.8)), musi być absorbowana pewna ilość energii, aby rezonator powrócił do stanu normalnego. Zachodzi to dlatego, że graniczna wartość m może być mniejsza niż $\bar{n}$, lecz średnia wartość $\bar{n}_m$ kwantów (o stanach wyższych od m) jest większa niż $\bar{n}$ (wzór (14.12)).

Wynik powyższy jest jak najbardziej ogólny i stosuje się tak do małych, jak i wielkich częstotliwości, chociaż przykład liczbowy dotyczył tylko częstotliwości małych.

14.3. REZONATOR WIELKIEJ CZĘSTOTLIWOŚCI I CENA OBSERWACJI

Rozważmy rezonator wielkiej częstotliwości, dla którego hv może być rzędu kT lub większego. Wielkość x nie jest już mała i warunek (14.6) na medianę nie będzie dawać zwykle wartości całkowitej. Musimy więc wybrać najbliższą liczbę całkowitą:

$$q \geqslant m, \qquad q,\ m \text{ całkowite.} \tag{14.16}$$

Pewność obserwacji jest wówczas większa niż 50%, a ΔS jest większe od $k \ln 2$ (wzór (14.15)), jeżeli nadmiar energii ulega rozproszeniu i oddawany jest termostatowi o temperaturze T.

Wybieramy nieco odmienny przypadek, odpowiadający zagadnieniu

axwella. Cząsteczki znajdują się w naczyniu zamkniętym o tempe-
By obserwować te cząsteczki, posługujemy się dodatkowym źród-
łem tła, emitującym kwanty $h\nu$. Cząsteczka jest obserwowana wtedy, gdy rozproszy ona co najmniej jeden kwant $h\nu$, który później trafia do fotoelementu lub oka obserwatora. Kwant ten musi leżeć powyżej poziomu promieniowania ciała czarnego, aby można go było rozróżnić od tła.

Założyliśmy poprzednio (wzory (13.4) i (13.38)), że wymagania powyższe prowadzą do warunku

$$h\nu \geqslant kT.$$

Wykażemy teraz, że granicę tę można obniżyć do

$$h\nu \geqslant kT \ln 2. \tag{14.17}$$

W tym celu zauważamy wpierw, że najmniejszą absorbowaną energią jest kwant. Dalej, jeśli żądamy, aby pewność obserwacji jednego kwantu (lub większej ilości) rezonatora była większa niż 50%, musimy przyjąć $m = 1$. To wraz z warunkiem (14.6) i (14.7) prowadzi do zależności

$$m = \bar{n} = 1 \quad \text{oraz} \quad x = \frac{h\nu}{kT} = \ln 2. \tag{14.18}$$

Wartość średnia i mediana pokrywają się i to właśnie odpowiada absorbcji jednego kwantu.

Stan podstawowy bez zaobserwowanych kwantów ma prawdopodobieństwo $^1/_2$ równe prawdopodobieństwu wszystkich stanów z 1, 2, 3 lub większą ilością kwantów. Jeżeli obserwujemy w rezonatorze jeden kwant (lub więcej), to prawdopodobieństwa tego, że pochodzi on od rzeczywistej absorpcji, oraz tego, że może być wywołany fluktuacjami, równe są $^1/_2$. Gdy rozproszony kwant zostanie zaabsorbowany przez ośrodek otaczający o temperaturze T, otrzymujemy przyrost entropii otoczenia

$$\Delta S = k \ln 2. \tag{14.19}$$

Graniczną wartość entropii k, przytoczoną w rozdz. 13.3, należy zatem poprawić na $k \ln 2$, co nieznacznie podnosi sprawność postępowania demona.

Jeżeli w naczyniu zamkniętym o temperaturze T rozpatrzymy rezonator o większej częstotliwości ν, takiej że

$$h\nu > kT \ln 2 \quad \text{lub} \quad x > \ln 2,$$

to otrzymamy $m < 1$ i powinniśmy przyjąć $q = 1$ — zgodnie ze wzorem

(14.16). Daje to pewność większą od 50%, natomiast ΔS większe od $k \ln 2$. We wszystkich tych przykładach obserwacja dostarcza jednego bitu informacji TAK lub NIE, co jest odpowiedzią na pytania:

Czy dany rezonator zaabsorbował energię?

Czy w danym miejscu znajduje się cząsteczka gazu?

Jedna binarna jednostka informacji odpowiada entropii $k \ln 2$ i opłacona musi być negentropią większą niż $k \ln 2$.

Należy zauważyć, że informacja ta odnosi się do krótkiego okresu czasu: tłumienie rezonatora spowoduje szybkie rozpraszanie nadmiaru energii lub też cząsteczka wyjdzie poza pole obserwacji. Jest to prawo zaniku, które charakteryzuje II zasadę termodynamiki.

Przedstawione rozumowanie jasno wykazuje, że II zasada spełniona jest tylko dla wartości średnich, co wykazano w punkcie (2), rozdz. 14.1. Możliwość wystąpienia nieprzewidzianych fluktuacji zawsze ogranicza zakres zastosowań II zasady termodynamiki. Może się zdarzyć, że jedna szczegółowa obserwacja może być dokonana za wyjątkowo niską cenę, nie ma jednak sposobu na to, by przewidzieć, kiedy i jak jej dokonać. Tylko wartości średnie można przewidzieć w sposób pewny.

Ilość informacji odpowiadająca większej liczbie bitów wpływa na bardzo mały przyrost entropii układu ze względu na mnożnik 10^{-16} występujący we wzorze (14.1).

Zagadnienia, które obecnie omawiamy, nie mają wielkiego znaczenia dla termodynamiki, ale związek między entropią i informacją jest fundamentalny dla teorii informacji, której nie można by konsekwentnie zbudować, nie uwzględniając tego powiązania. Dalej rozpatrzymy specjalne warunki, przy których entropijna cena obserwacji może stać się dużo wyższa niż wartość graniczna (14.1).

14.4. EKSPERYMENTY WYMAGAJĄCE WIELU JEDNOCZESNYCH OBSERWACJI PRZY NISKICH CZĘSTOTLIWOŚCIACH

W rozdziałach 14.2 i 14.3 omówiliśmy prosty przykład, w którym doświadczenie polegało na obserwacji jednego oscylatora, i znaleźliśmy granicę przyrostu entropii

$$\Delta S = k \ln 2$$

przy prawdopodobieństwie poprawnej obserwacji równym $^1/_2$. Granicę tę uzyskano przy założeniu małych częstotliwości. Jeżeli przyjmiemy przypadek, gdzie mogą być pożądane wielkie częstotliwości (rozdz. 14.3), wyniki można podsumować, jak następuje:

$$\Delta S \geqslant A_1 k, \text{ gdzie } A_1 = \begin{cases} \ln 2 & \text{— częstotliwości małe z } h\nu \leqslant kT, \\ \dfrac{h\nu}{kT} & \text{— częstotliwości wielkie z } h\nu > kT. \end{cases} \tag{14.20}$$

Jest to przyrost entropii dla pojedynczej obserwacji. W praktyce promieniowanie ciała czarnego nie zawiera częstotliwości wielkich, a jeden kwant energii $h\nu$ wystarcza, aby zapewnić pewność obserwacji.

Główny powód ograniczenia, polegający na tym, że informacja uzyskiwana jest tylko z odczytów pozytywnych (zaobserwowanie wysokiej energii rezonatora), tkwi w tym, że nie można ściśle zdefiniować pewności odczytu negatywnego.

Rozważymy teraz bardziej złożone zagadnienie, w którym doświadczenie wymaga obserwacji n oscylatorów. W niektórych przypadkach obserwacje muszą być dokonywane jednocześnie na n rezonatorach, w innych zaś możemy obserwować je kolejno. Przy obserwacji można żądać, aby tylko jeden z tych rezonatorów w ustalonej, danej chwili wykazywał energię większą od pewnej granicznej wartości E_l, chociaż w innych urządzeniach eksperymentalnych obserwację można oprzeć na obserwowaniu wysokiej energii pewnej liczby rezonatorów m:

$$1 \leqslant m \leqslant n.$$

Przyjmując, że wszystkie rezonatory utrzymywane są w termostacie przy stałej temperaturze T, rozważamy najpierw przypadek małych częstotliwości ($h\nu \ll kT$). Problem sprowadza się do określenia granicy E_l, dla której prawdopodobieństwo poprawnej obserwacji wynosi $^1/_2$, oraz prawdopodobieństwo obserwacji fałszywej, spowodowanej fluktuacjami wynosi również $^1/_2$.

Rozpatrzmy dowolną wartość graniczną E_l i omówmy przypadek jednego z n oscylatorów. Ze wzoru (14.5a) mamy

$$P_1 = \mathrm{e}^{-E_l/kT}. \tag{14.21a}$$

Jest to prawdopodobieństwo tego, że energia rezonatora jest większa od E_l i pochodzi od fluktuacji cieplnych, natomiast

$$P_2 = 1 - \mathrm{e}^{-E_l/kT} \tag{14.21b}$$

oznacza prawdopodobieństwo tego, że fluktuacje dostarczają rezonatorowi energii mniejszej niż E_l.

Obserwując n rezonatorów, łatwo obliczyć prawdopodobieństwa zależne od fluktuacji termicznych (tabl. 14.1).

Tablica 14.1

Wzór	Określenie
$p_{n,n} = p_1^n$	prawdopodobieństwo tego, że wszystkie n rezonatorów ma energię $\geqslant E_l$
$p_{n,n-1} = n\,p_1^{n-1}\,p_2$	prawdopodobieństwo tego, że jeden z rezonatorów ma energię $< E_l$, podczas gdy pozostałe $(n-1)$ mają energię $\geqslant E_l$
$p_{n,n-2} = \frac{n(n-1)}{2}\,p_1^{n-2}\,p_2^2$	prawdopodobieństwo tego, że jakieś dwa rezonatory mają energię $< E_l$, natomiast $(n-2)$ mają energię $\geqslant E_l$
$p_{n,n-m} = \frac{n!}{m!(n-m)!}\,p_1^{n-m}\,p_2^m$	prawdopodobieństwo tego, że m rezonatorów ma energię $< E_l$, a pozostałe $(n-m)$ mają energię $\geqslant E_l$
$p_{n,0} = p_2^n$	prawdopodobieństwo tego, że wszystkie rezonatory mają energię $<$ od E_l

Należy wspomnieć, iż w rozdz. 14.3 założono, że obserwacja polega na odczytaniu energii rezonatora większej od wartości E_l. Odczyt negatywny, tj. obserwowanie niskiej energii rezonatora, nie daje żadnej informacji ze względów przytoczonych w rozdz. 13.6 podczas omawiania przez nas maszyny cieplnej Szilarda. Suma prawdopodobieństw rozciągnięta na wszystkie przypadki równa się, oczywiście, jedności:

$$\sum_{m=0}^{n} P_{n,\,n-m} = P_1^n + \ldots + \frac{n!}{m!(n-m)!}\,P_1^{n-m}\,P_2^m + \ldots + P_2^n =$$
$$= (P_1+P_2)^n = 1^n = 1.$$

Określmy teraz szczegółowo warunki doświadczenia polegającego na obserwacji. Zakładamy najpierw, że obserwacja daje wynik pozytywny, gdy któryś z n oscylatorów posiada energię większą od E_l. Prawdopodobieństwo P_2^n oznacza, że wynik ten nie jest spowodowany fluktuacjami, a $1-P_2^n$ jest prawdopodobieństwem tego, że jeden lub więcej oscylatorów dzięki fluktu-

acjom termicznym podniosło swoją energię do wartości wyższej od E_l. Prawdopodobieństwo poprawnej obserwacji wynosi zatem P_2^n, ponieważ obserwowana energia jest większa od E_l. W tym przypadku musi być ona wynikiem absorpcji energii ze źródła zewnętrznego. Prawdopodobieństwo obserwacji fałszywej spowodowanej fluktuacjami wynosi $1-P_2^n$. Żądamy teraz, aby poprawna obserwacja zachodziła z prawdopodobieństwem $^1/_2$, i otrzymujemy warunek

$$P_{\dot{z}}^n = (1-e^{-A_n})^n = \frac{1}{2} \quad \text{przy} \quad E_l = A_n kT \tag{14.22}$$

lub

$$e^{-A_n} = 1-\left(\frac{1}{2}\right)^{1/n}. \tag{14.22a}$$

Dla $n = 1$ otrzymujemy przytoczony wcześniej wzór (14.15):

$$E_{l1} = kT \ln 2, \qquad A_1 = \ln 2.$$

Gdy n wzrasta, E_l musi także wzrastać, ponieważ prawa strona wzoru (14.22a) maleje wraz ze wzrostem n. Funkcja e^{-A_n} również maleje. Możemy zlogarytmować obie strony wzoru (14.22) i otrzymać wyrażenie przybliżone dla A_n:

$$n \ln(1-e^{-A_n}) \approx -n\, e^{-A_n} \approx -\ln 2$$

lub

$$A_n \approx \ln n - \ln(\ln 2) \approx \ln n + 0{,}3667. \tag{14.23}$$

Poniżej przytoczono niektóre wartości współczynnika A_n:

$n =$	1	2	4	100	10000
$A_n \approx$	0,69	1,23	1,84	4,95	9,58

(14.24)

W naszym doświadczeniu wzbudzony rezonator przyjął nadmiar energii, ze źródła zewnętrznego, przy obserwacji poprawnej. Jeżeli obserwacja jest fałszywa, nadmiar ten pochodzi z termostatu. W obu przypadkach ten nadmiar energii zostanie zaabsorbowany później przez termostat, co w przypadku poprawnej obserwacji prowadzi do przyrostu entropii o

$$\Delta S = \frac{E_l}{T} = A_n k.$$

Rozważania przytoczone tutaj są podobne do rozważań przytoczonych przy omawianiu wzorów (14.14) i (14.15).

Należy podkreślić, że wyliczony tu przyrost ΔS oznacza tylko granicę dolną. Rzeczywisty przyrost może być większy od kA_n. Absorpcję energii ze źródła zewnętrznego (zasobnika) obserwujemy tylko wtedy, jeżeli zaabsorbowana energia podnosi rezonator do poziomu wyższego od E_l. Inne rezonatory mogą absorbować energię ze źródła zewnętrznego w mniejszych ilościach, co jednak nie podniesie energii rezonatorów powyżej poziomu E_l i ujdzie naszej uwadze. Ta zaabsorbowana energia w późniejszej fazie przejdzie również do termostatu i wniesie dodatkowy nieznany przyrost entropii. Dochodzimy ostatecznie do warunku

$$\Delta S \geqslant A_n k. \tag{14.25}$$

Ponieważ A_n jest funkcją rosnącą n, dolna granica przyrostu entropii staje się większa, jeżeli liczba rezonatorów niezbędnych do obserwacji wzrasta.

Równania (14.22), (14.22a), (14.23) i (14.24) na A_n otrzymano w oparciu o założenie, że tylko jeden z rezonatorów dostarczył odczytu pozytywnego. Równania te są jednak ważne w przypadku, kiedy otrzymujemy m $(1 \leqslant m \leqslant n)$ odczytów pozytywnych, jeżeli żądamy, aby wszystkie były prawdziwe z prawdopodobieństwem $^1/_2$. Oznacza to, tak jak poprzednio, iż prawdopodobieństwo tego, że wszystkie n rezonatorów mają energię większą od E_l, powinno wynosić $^1/_2$, jeżeli uwzględniamy tylko fluktuacje. W takim razie wszystkie przytoczone poprzednio wzory na A_n są słuszne. Musimy jednak zmodyfikować wzór (14.25) wyrażający przyrost entropii, ponieważ termostat absorbuje teraz energię mE_l. W takim razie

$$\Delta S = \frac{mE_l}{kT} = mA_n^1 k$$

lub, jeżeli uwzględnimy nieoznaczoność w odczytach negatywnych,

$$\Delta S \geqslant mA_n k. \tag{14.26}$$

W rozważaniach tych wykorzystaliśmy tylko niedużą część tabl. 14.1. Inne prawdopodobieństwa mogłyby być wykorzystane w innych zagadnieniach, takich np., w których tylko pewna część odczytów pozytywnych uważana jest za prawdziwą.

Inaczej wygląda sprawa w przypadku wielkich częstotliwości. Jeden pojedynczy kwant $h\nu$ przypadający na rezonator wystarcza do przeprowadzenia obserwacji pewnej i oczywiście reprezentuje on minimum wymaganej energii. Częstotliwości wielkie, w promieniowaniu ciała czarnego praktycznie nie

ują. Wobec tego weźmy

$$E_l = h\nu = AkT, \qquad A = \frac{h\nu}{kT}.$$

Podsumowując otrzymane wyniki możemy stwierdzić, że doświadczenie wymagające n równoczesnych obserwacji n rezonatorów zamkniętych w termostacie prowadzi do przyrostu entropii

$$\Delta S = Ak \tag{14.27}$$

na każdy rezonator, którego energia jest większa od E_1, gdzie

$$A = \begin{cases} A_n & \text{— zgodnie ze wzorami (14.23) i (14.24) dla częstotliwości niskich, o } h\nu \ll kT, \\ \dfrac{h\nu}{kT} & \text{— dla częstotliwości wysokich, o } h\nu \gg kT. \end{cases} \tag{14.28}$$

Przypadek pośredni $h\nu \approx kT$ powinien być przeanalizowany oddzielnie z uwzględnieniem specyfiki konkretnego zagadnienia.

14.5. ZAGADNIENIA WYMAGAJĄCE DUŻEJ PEWNOŚCI OBSERWACJI

Nasze poprzednie rozważania opierały się na założeniu, że możemy być zadowoleni, jeśli w naszym doświadczeniu mamy 50% błędów. To bardzo słabe ograniczenie wybrano w celu uzyskania najmniejszej wartości przyrostu entropii związanej z doświadczeniem. Interesująco przedstawia się sprawa, jeżeli nasze badania rozciągniemy na zagadnienia wymagające większej pewności i obserwować będziemy, jak wzrasta entropijna cena pomiarów.

Podkreślmy tu różnicę między pewnością i dokładnością. Pojęcie dokładności zostanie szczegółowo omówione w rozdz. 15, a zdefiniować ją możemy, jak następuje: Mamy wielkość x mierzoną z błędem Δx, przy czym cały zakres możliwych obserwacji wynosi l. Dokładność definiujemy wzorem

$$\mathcal{A} = \frac{l}{\Delta x}. \tag{14.29a}$$

Na przykład przy posługiwaniu się miarką metrową z działkami co milimetr $\mathcal{A} = 1000$, natomiast posługując się miarką jardową z działkami co $^1/_2$ cala otrzymamy $\mathcal{A} = 144$.

Pewność stanowi zagadnienie odrębne, które teraz omówimy. Przypuśćmy, że *dane jest postępowanie doświadczalne z wymaganiami pewnej dokładności.* Zawsze istnieje możliwość, że fluktuacje termiczne spowodują fałszywe odczyty i mogą wpłynąć na niepoprawne wskazania. Będzie to zachodzić z pewnym prawdopodobieństwem. Zamiast rozważanego poprzednio prawdopodobieństwa $^1/_2$, chcemy teraz zająć się przypadkiem, w którym prawdopodobieństwo błędu może być zmniejszone do wartości

$$P = \frac{1}{r}, \quad r > 2. \tag{14.29b}$$

Warunek Isinga (wzory (14.3), (14.4) (14.13) i (14.15)) odpowiada $r = 54$.

Zacznijmy od doświadczenia z rezonatorem małej częstotliwości, w którym dokonuje się jedynego odczytu. Można tu przytoczyć bezpośrednio nasze rozważania z rozdz. 14.2 z tym, że musimy wybrać wartość graniczną q, która daje

$$P_{n \geqslant q} = \mathrm{e}^{-qx} = \frac{1}{r}.$$

Zgodnie ze wzorem (14.5a) otrzymamy

$$\ln r = qx = \frac{qhv}{kT}. \tag{14.30}$$

Energia graniczna rezonatora wynosi qhv. Po dokonaniu obserwacji ta właśnie dodatkowa energia zostaje rozproszona na skutek tłumienia rezonatora (wzór (14.14)), w wyniku czego wzrasta entropia termostatu T (wzór (14.15)) o

$$\Delta S = \frac{qhv}{T} = k \ln r. \tag{14.31}$$

Entropijna cena pomiaru wzrasta więc logarytmicznie ze wzrostem pewności r.

Rozpatrzmy teraz eksperyment, w którym dokonuje się wielu obserwacji w celu rozszerzenia wyników z rozdz. 14.4. Nasz wcześniejszy wzór (14.22) zastąpmy równaniem

$$P = (1 - \mathrm{e}^{-A_n})^n = 1 - \frac{1}{r}, \quad E_l = A_n kT, \tag{14.32}$$

gdzie P jest prawdopodobieństwem tego, że dla n rezonatorów dokonujemy m pozytywnych odczytów i wszystkie te odczyty są prawdziwe, natomiast E_l jest energią graniczną każdego z n rezonatorów.

Poprzedni wynik (14.31) odpowiada $n = 1$ oraz

$$A_1 = \ln r. \tag{14.33}$$

Ponieważ dla dużych wartości r oraz n wielkość A_n staje się bardzo mała, można wzory uprościć korzystając z wyrażeń asymptotycznych. Logarytmując obie strony wzoru (14.32) mamy

$$\left.\begin{aligned} & n \ln (1 - e^{-A_n}) \approx -n e^{-A_n}, \qquad \ln\left(1 - \frac{1}{r}\right) \approx -\frac{1}{r} \\ & \text{i stąd} \\ & n e^{-A_n} \approx \frac{1}{r} \\ & \text{lub} \\ & A_n \approx \ln (rn), \qquad rn \gg 1. \end{aligned}\right\} \tag{14.34}$$

Współczynnik A_n dla częstotliwości małych może w praktyce zmieniać się od 0,7 do kilkuset. Wzór (14.34) może dawać dobre przybliżenie nawet przy małych wartościach n, ponieważ jest on słuszny dla $n = 1$.

Ostatecznie energia rozproszona w m rezonatorach wynosi mA_nkT, a całkowita cena entropijna obserwacji wynosi

$$\Delta S \geqslant kmA_n = km \ln (rn). \tag{14.35}$$

Przypadek wielkich częstotliwości jest znowu inny. Jeden pojedynczy kwant $h\nu$ przypadający na rezonator wystarcza do obserwacji, ponieważ drgania o wielkich częstotliwościach w promieniowaniu ciała czarnego praktycznie nie występują:

$$\left.\begin{aligned} & E_l = h\nu = AkT, \\ & A = \frac{h\nu}{kT}, \; h\nu \gg kT, \\ & \Delta S \geqslant mkA. \end{aligned}\right\} \tag{14.36}$$

Dochodzimy zatem do współczynnika

$$A = \begin{cases} A_n \text{ dane przez wzór (14.34) dla częstotliwości bardzo małych, o } h\nu \ll kT, \\ \dfrac{h\nu}{kT} \text{ dla częstotliwości wielkich, o } h\nu \gg kT. \end{cases} \tag{14.37}$$

Przejście od małych do wielkich częstotliwości byłoby trudniejsze do zanalizowania; uzależnione jest również od specyfiki każdego rozważanego zagadnienia.

We wszystkich poruszonych do tej pory przykładach zakładaliśmy, że obserwowano n oscylatorów utrzymywanych w temperaturze termostatu T. W pewnej chwili obserwowaliśmy energie tych oscylatorów, a później ustalała się kolejno równowaga temperatury między oscylatorami i termostatem. Pewna ilość energii zostaje przy tym rozproszona i entropia termostatu wzrasta. Sytuacja taka jest najprostsza do analizy, niemniej jednak przypadek wielkich częstotliwości wymaga bardziej szczegółowego omówienia.

14.6. DOKŁADNIEJSZE OMÓWIENIE DOŚWIADCZEŃ, W KTÓRYCH STOSUJE SIĘ WIELKIE CZĘSTOTLIWOŚCI

Warunki, jakie założyliśmy w niniejszym rozdziale, stanowią zadowalająco dokładne potwierdzenie teorii w przypadku małych częstotliwości. Czasami jednak doświadczenie wymaga posługiwania się wielkimi częstotliwościami i może być wówczas interesujące, czy całkowita energia wymagana podczas doświadczenia musi ostatecznie zawsze ulec rozproszeniu. Równanie (14.36) określa z pewnością całkowitą ilość energii wymaganej do obserwacji, lecz zamiast posługiwać się odbiornikiem w postaci rezonatora utrzymywanego w temperaturze T, możemy pomyśleć o innym przyrządzie, takim jak fotokomórka utrzymywana w temperaturze T. Promieniowanie wykorzystywane w eksperymencie może uderzyć w metal, z którego wybije elektron o energii kinetycznej

$$E_{k_0} = h\nu - W_0, \tag{14.38}$$

gdzie W_0 jest pracą wyjścia elektronu. Elektron może być potem hamowany przez pole elektryczne, tracąc energię W_1, równą pracy wykonanej przez pole. W końcu elektron osiągnie ekran fluorescencyjny o temperaturze T i zostanie pochłonięty. W chwili absorpcji elektron posiada małą energię kinetyczną

$$E_{k_1} = h\nu - W_0 - W_1 \tag{14.39}$$

i energia ta zmienia się na ciepło dając przyrost entropii ekranu

$$\Delta S = \frac{h\nu - W_0 - W_1}{T}. \tag{14.40}$$

Jeżeli pole hamujące jest słabe, ΔS będzie nieznacznie mniejsze niż $h\nu/T$. Jeżeli pole jest silne, elektron nie może osiągnąć ekranu i obserwacja jest niemożliwa. Doświadczenie dawać może wyniki pozytywne wtedy, gdy elektron osiągnie ekran z energią większą niż fluktuacje termiczne ekranu, to zaś prowadzi do zagadnienia omówionego w rozdz. 14.5.

Należy zwrócić uwagę na różnicę między

$$\Delta E = nh\nu \tag{14.41}$$

będącą najmniejszą, całkowitą energią wymaganą w doświadczeniu z wykorzystaniem n komórek odbiorczych, a najmniejszym przyrostem entropii n odbiorników przy obserwacji o pewności r,

$$\Delta S = knA_n. \tag{14.42}$$

Występujący tu współczynnik A_n został zdefiniowany przez wzory (14.23), (14.24) i (14.34).

W przypadku wielkich częstotliwości mamy

$$\Delta Q = T\Delta S < \Delta E. \tag{14.43}$$

Tylko część całkowitej energii ΔE zamienia się w ciepło ΔQ, a pozostała różnica energii może być przekształcona na pracę elektryczną lub mechaniczną ΔW:

$$\Delta W = \Delta E - \Delta Q. \tag{14.44}$$

W większości urządzeń doświadczalnych nie dbamy o to, czy energia ΔE ulega rzeczywiście przemianie w ciepło lub czy zostaje odzyskana, mimo to poprzednie nasze rozważania teoretyczne pozostają słuszne.

Posługiwanie się wielkimi częstotliwościami do uzyskania pewnej ilości informacji okazuje się zatem bardziej kosztowne niż posługiwanie się częstotliwościami małymi, chociaż można by wprowadzić jakieś specjalne urządzenia pozwalające odzyskać (w postaci pracy) część energii związanej z wielkimi częstotliwościami. Jednakże zawsze niezbędne jest rezerwowanie dostatecznej ilości energii do uaktywnienia detektora małej częstotliwości, tak że poprzednie wyniki odnoszące się do rezonatorów małej częstotliwości dają minimalny koszt obserwacji. Wspomniane urządzenie minimalizujące straty energii w przypadku wielkich częstotliwości można po prostu wyobrazić

sobie jako przyrząd przekształcający początkowe sytuacje, wymagające wielkich częstotliwości, w sytuacje, dla których mamy już do czynienia z małymi częstotliwościami.

Warto w tym miejscu wspomnieć o metodzie fotograficznej. Podczas uderzenia kwantu $h\nu$ w kliszę fotograficzną następuje jonizacja atomu i emitowany zostaje elektron o energii kinetycznej E_{k_0}, wyrażonej przez wzór (14.38), gdzie W_0 jest teraz progiem jonizacji. Elektron przebywa pewną odległość i ostatecznie zostaje zatrzymany w emulsji tracąc swoją energię E_{k_0} (częściowo na wykonywanie pracy, częściowo na ciepło wytwarzane podczas licznych zderzeń). W_0 jest tu dużo większe od kT (ponieważ jonizacja nie zachodzi pod wpływem fluktuacji termicznych) i ciepło wytworzone w emulsji będzie większe od kT (w przeciwnym razie elektron nie przebyłby dostatecznie długiej drogi i mógłby być odrzucony na skutek pobudzenia termicznego). Warunki te nie odbiegają od dopiero co omówionych i prowadzą do podobnych wniosków. Procesu wywołania kliszy nie musimy rozpatrywać. Jest on podobny do wzmacniania (z punktu widzenia entropii — bardzo kosztowny) efektów mikroskopowych w efekty makroskopowe, które można już bezpośrednio obserwować.

14.7. PRZYKŁAD ILUSTRUJĄCY WYSTĘPOWANIE MINIMUM NEGENTROPII POTRZEBNEJ DO DOKONANIA OBSERWACJI

Rozważymy teraz uproszczony przykład, na tle którego rzeczywiście możliwe jest wykazanie występowania minimum negentropii wymaganej do obserwacji. Mianowicie weźmy znów pod uwagę zagadnienie poruszone w rozdz. 13.4 dotyczące określania położenia cząsteczki przy pomocy promieniowania świetlnego. Weźmy ciąg impulsów pokazanych na rys. 14.1. Czas trwania impulsu świetlnego wynosi τ_1, przerwy zaś między impulsami τ_2. Całkowity okres powtarzania $\tau = \tau_1 + \tau_2$. Funkcja

$$f(t) = \begin{cases} 1 \text{ dla } |t| < \dfrac{\tau_1}{2} \\ 0 \text{ dla } \dfrac{\tau_1}{2} < |t| < \dfrac{\tau}{2} \end{cases} \tag{14.45}$$

opisuje ciąg, który zajmuje pasmo częstotliwości określone w głównej mierze

przez czas trwania najwęższego impulsu. Jeżeli założymy, że $\tau_1 < \tau_2$, to zgodnie ze wzorem (13.32)

$$0 < \nu < \nu_M, \qquad \nu_M \approx \frac{1}{2\tau_1}. \tag{14.46}$$

Zamiast ciągu impulsów $f(t)$ możemy rozpatrzyć impulsy odwrócone $F(t)$ pokazane na rys. 14.2:

$$F(t) = \begin{cases} 0 \text{ dla } |t| < \dfrac{\tau_1}{2} \\ \\ 1 \text{ dla } \dfrac{\tau_1}{2} < |t| < \dfrac{\tau}{2} \end{cases} \quad \text{z okresem } \tau, \tag{14.47}$$

$$F(t) = 1 - f(t). \tag{14.48}$$

Obie funkcje*, f i F mają widmo Fouriera tej samej postaci, chociaż jedna z nich składa się z impulsów wąskich i dużych przerw, a druga z szerokich impulsów i krótkich przerw. Maksymalna częstotliwość ν_M związana jest w obu przypadkach z minimalnym czasem τ_1 lub τ_2. Ta maksymalna częstotliwość jest najmniejsza wtedy, gdy f oraz F są takie same, tj. przy

$$\tau_1 = \tau_2 = \frac{\tau}{2}, \quad \nu_M = \frac{1}{\tau}. \tag{14.49}$$

Odpowiadać to będzie najmniejszej ilości energii wymaganej do obserwacji, a stąd i najmniejszemu przyrostowi entropii. Rozważmy tę sytuację bardziej szczegółowo.

Łatwo możemy obliczyć ilość informacji. Zgodnie z poprzednim założeniem posługujemy się ciągiem impulsów $f(t)$ (rys. 14.1) i obserwujemy światło rozpraszane przez cząsteczkę. Oznacza to, że cząsteczka może się znaleźć w wiązce światła podczas jednego z impulsów o czasie trwania τ_1. Oznaczmy przez P_0 całkowitą liczbę możliwości a priori. Jest ona proporcjonalna do τ. Ilość możliwości wchodzących w grę po uzyskaniu informacji oznaczmy przez P_1.

---

* Funkcje $f(t)$ oraz $F(t)$ brane są w pierwszym okresie. Dla n-tego okresu należy oczywiście przyjmować

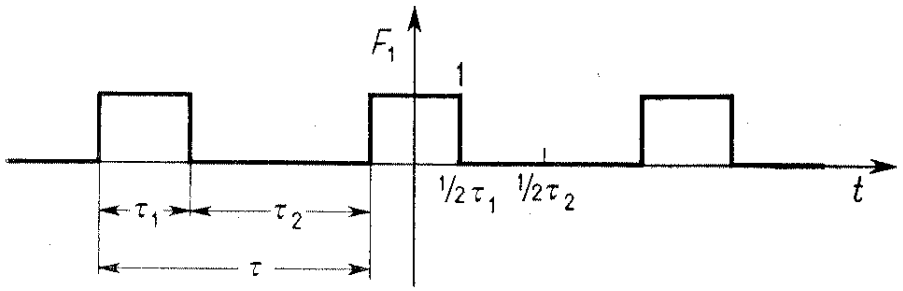
$$|t| < \frac{n\tau_1}{2}, \qquad n\frac{\tau_1}{2} < |t| < \frac{n\tau}{2}.$$

(*Przyp. tłum.*).

Jest ona proporcjonalna do $\tau_1 < \tau$. Stosując wzór (1.6) znajdujemy ilość informacji

$$I = k \ln\left(\frac{P_0}{P_1}\right) = k \ln\left(\frac{\tau}{\tau_1}\right). \tag{14.50}$$

Krótszy czas τ_1 zwiększa dokładność a tym samym większa będzie ilość informacji.

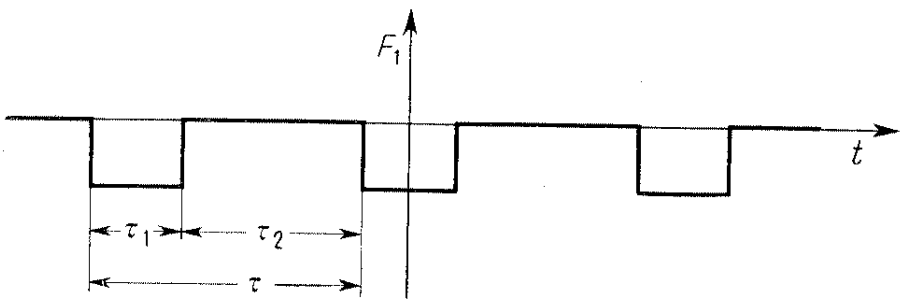


Rys. 14.1. Impulsy świetlne powtarzane z okresem τ. Czas trwania impulsu wynosi τ_1, natomiast przerwa między impulsami $\tau_2 = \tau - \tau_1$

W celu obliczenia entropii, koniecznej dla dokonania obserwacji, przeanalizujemy najpierw widmo Fouriera funkcji (14.45)

$$\left.\begin{aligned} f(t) &= \frac{\tau_1}{\tau} + \sum_{n=1}^{\infty} a_n \cos n\Phi, \\ \Phi = \frac{2\pi t}{\tau}, \qquad a_n &= \frac{2}{\pi n} \sin\frac{n\pi\tau_1}{\tau} = \frac{2}{\pi n}(-1)^{n+1} \sin\frac{n\pi\tau_2}{\tau}. \end{aligned}\right\} \tag{14.51}$$

Symetria względem τ_1 oraz τ_2 jest oczywista i odpowiada relacji (14.48). Amplitudy a_n są maksymalne dla małych wartości n i stopniowo maleją do wartości bardzo małych.

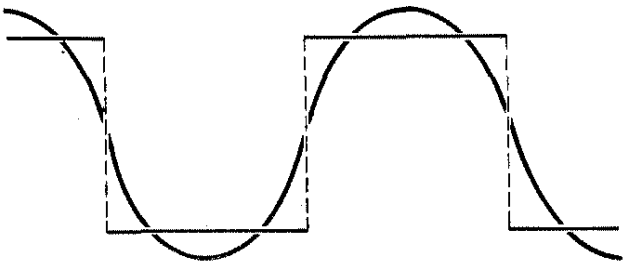


Rys. 14.2. Impulsy świetlne o czasie trwania τ_2 powtarzane z okresem τ. Jest to odwrócenie fali z rys. 14.1

Przyjmijmy $\tau_1 < \tau_2$; wówczas pierwsze zerowanie się amplitudy wystąpi dla

$$n_0 = \frac{\tau}{\tau_1}, \qquad \nu_0 = \frac{n_0}{\tau} = \frac{1}{\tau_1}, \qquad \tau_1 < \tau_2.$$

Musimy pamiętać o obecności szumu. Bardzo małe amplitudy w widmie (14.51) zostaną zatarte przez szumy. Załóżmy, że możemy pozbyć się wielkich



Rys. 14.3. Aproksymacja ciągu impulsów $\tau_1 = \tau_2 = {}^1/_2\tau$ dwoma wyrazami szeregu Fouriera: $f = \frac{1}{2} + \frac{2}{\pi}\cos\frac{2\pi t}{\tau}$

częstotliwości i otrzymać tylko skończone widmo rozciągające się do częstotliwości ${}^1/_2\nu_0$ (równe połowie częstotliwości odpowiadającej pierwszemu zerowaniu się amplitudy). Mamy wówczas

$$\left.\begin{aligned} \nu_{\max} &= \frac{1}{2\tau_1}, \quad n_{\max} = \frac{\tau}{2\tau_1}, \quad \tau_1 < \tau_2, \\ \nu_{\max} &= \frac{1}{2\tau_2}, \quad n_{\max} = \frac{\tau}{2\tau_2}, \quad \tau_2 < \tau_1. \end{aligned}\right\} \tag{14.52}$$

Przypadek $\tau_1 = \tau_2$ godny jest szczególnej uwagi. W tym przypadku nasz warunek (14.52) daje $n_{\max} = 1$, co oznacza stosowanie aproksymacji jednowyrazowej, jak pokazano na rys. 14.3:

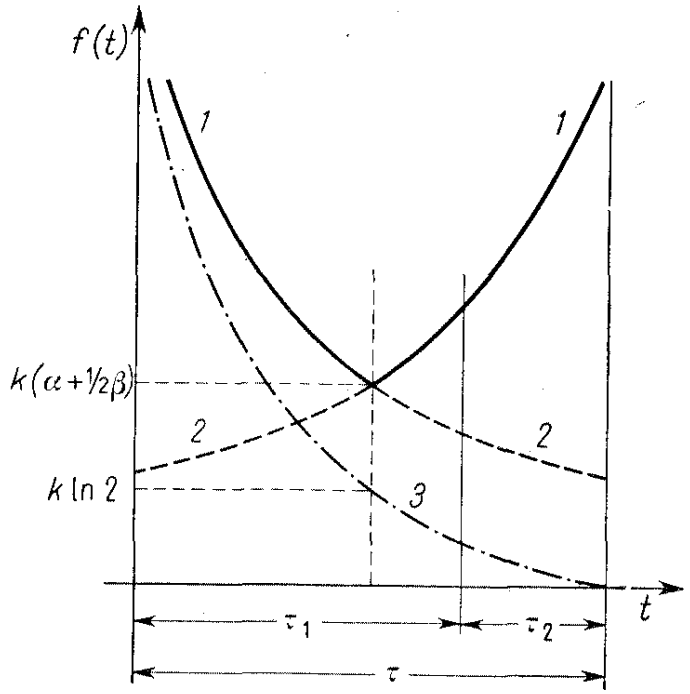
$$f = \frac{1}{2} + \frac{2}{\pi}\cos\frac{2\pi t}{\tau}, \qquad \tau_1 = \tau_2 = \frac{\tau}{2}.$$

W miejsce fali prostokątnej mamy falę cosinusoidalną.

Aby przezwyciężyć wpływ szumu cieplnego, każdy z otrzymanych składników musi mieć energię rzędu kT. Ostatecznie całkowita energia sygnału z ograniczonym widmem wynosi

$$E = n_{\max}\alpha kT + \beta\frac{kT}{2}, \tag{14.53}$$

gdzie α i β są współczynnikami liczbowymi rzędu jedności. Ostatni wyraz odpowiada energii, jaką powinna mieć składowa stała, aby przewyższyć



Rys. 14.4. Informacja I i związany z nią przyrost entropii jako funkcja $\tau_1 = \tau - \tau_2$ dla

$$0 < \tau_1 < \tau$$

1 — przyrost entropii; *2* — niewykorzystana część funkcji $\alpha n_{\max} + {}^1/_2\beta$; *3* — informacja.

energię ${}^1/_2\, kT$. Energia ta jest rozpraszana i absorbowana (np. przez fotokomórkę) podczas obserwacji i pociąga za sobą przyrost entropii

$$\Delta S = k\left(n_{\max} + \frac{1}{2}\beta\right). \tag{14.54}$$

Odpowiadająca wzorowi (14.53) krzywa ma minimum przy

$$\tau_1 = \tau_2 = \frac{\tau}{2}.$$

Jej przebieg pokazano na rys. 14.4, na którym także pokazano wykres ilości informacji I (wzór (14.50)). Rozpatrzony tu przykład wskazuje, że przyrost entropii jest zawsze większy niż uzyskana ilość informacji, zaś ilość entropii wymaganej do obserwacji nie może być mniejsza od pewnej granicznej wartości — w naszym przypadku równej $k(\alpha + {}^1/_2\beta)$.

15

OBSERWACJA A INFORMACJA

15.1. BŁĘDY DOŚWIADCZALNE A INFORMACJA

Zwróciliśmy już uwagę na fakt, że ilość informacji uzyskiwanej z doświadczenia jest ściśle związana z dokładnością działania aparatury doświadczalnej. Zaczęliśmy od zdefiniowania ilości informacji (rozdz. 12):

$$I_1 = k \ln \frac{P_0}{P_1}, \tag{15.1}$$

gdzie

P_0 — liczba równoprawdopodobnych możliwości przed obserwacją,
P_1 — liczba równoprawdopodobnych możliwości po obserwacji,
I_1 — informacja uzyskana z obserwacji.

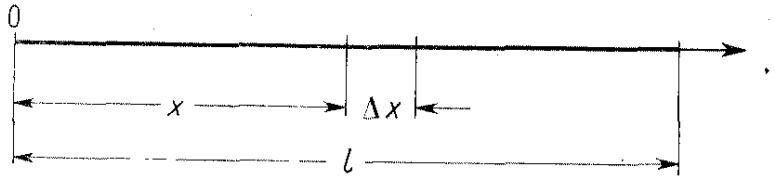
Definicję tę zastosowaliśmy w rozdz. 13 i 14 w różnych warunkach doświadczeń. Omówimy teraz zasadnicze zagadnienia, jakimi są: dokładność, błędy eksperymentalne i informacja. Zacznijmy od prostego i dającego się dokładnie opisać przykładu. Na odcinku o długości L obserwujemy położenie punktu x z błędem Δx, jak pokazano na rys. 15.1. Wprowadźmy określenia:

$$\left.\begin{aligned} &\text{błąd absolutny: } \varepsilon_a = \Delta x, \\ &\text{błąd względny: } \varepsilon_r = \frac{\Delta x}{x}, \\ &\text{błąd znormalizowany: } \varepsilon_c = \frac{\Delta x}{L}. \end{aligned}\right\} \tag{15.2}$$

Dokładność eksperymentu $\mathcal{A}$ określona jest przez wzór (14.29) jako odwrotność błędu znormalizowanego ε_c:

$$\mathcal{A} = \frac{1}{\varepsilon_c} = \frac{L}{\Delta x}. \tag{15.2a}$$

Warunek, że wzrost dokładności powinien odpowiadać coraz to mniejszym błędom znormalizowanym, jest spełniony przez tę definicję. W fizyce znane



Rys. 15.1. Miejsce położenia cząstki określone z błędem Δx w punkcie x w zakresie możliwych położeń L

jest pojęcie błędu bezwzględnego (absolutnego) i błędu względnego. O błędzie znormalizowanym niewiele się wspomina. Niemniej jest to ważna wielkość w powiązaniu z teorią informacji. Ilość informacji uzyskiwanej z obserwacji, według wzoru (15.1), wyniesie

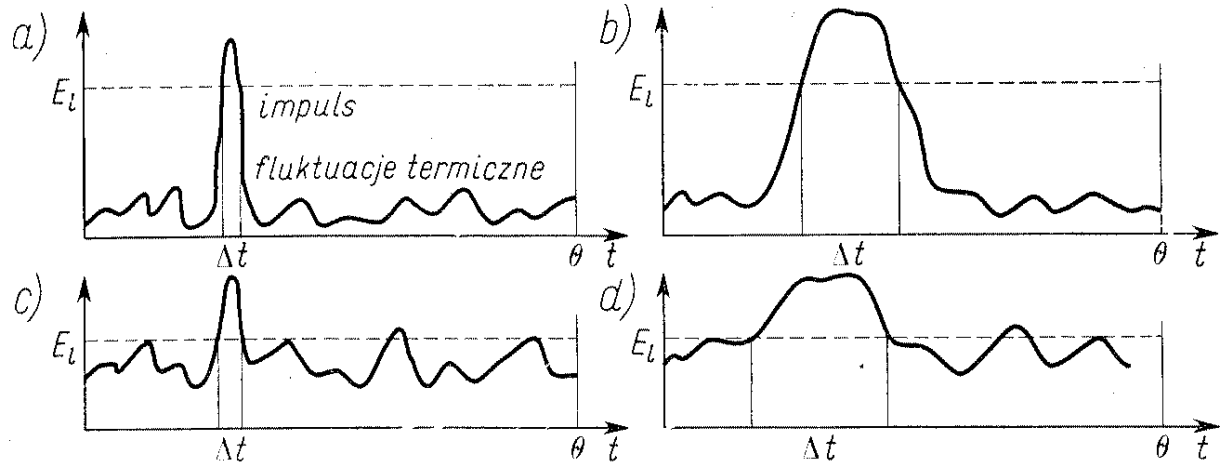
$$\Delta I = k \ln \left(\frac{L}{\Delta x}\right) = -k \ln \varepsilon_c = k \ln \mathcal{A}. \tag{15.3}$$

Równość ta zawiera stosunek nieoznaczoności Δx, pozostającej po dokonaniu obserwacji, do nieoznaczoności początkowej L, odpowiadającej całemu zakresowi możliwych obserwacji. W przyrządzie optycznym byłby to stosunek zdolności rozdzielczej do całkowitej apertury.

Posłużyliśmy się już podobną definicją odnoszącą się do pomiaru czasu w rozdz. 14.7. Omówiliśmy przypadek wąskich impulsów o czasie trwania τ powtarzalnych z okresem T i wykazaliśmy, że ważną wielkością w rozważaniach był stosunek τ/T, gdyż obie te wielkości w takiej postaci występują w wyrażeniach na informację i ceny entropijne. Omówimy podobne zagadnienia dotyczące przestrzeni i czasu oraz prześledzimy tok mierzenia położenia cząstki o współrzędnych x, y, z w czasie t z błędami Δx, Δy, Δz oraz Δt. Definicja wymaga wprowadzenia wielkości a, b, c oraz θ będących miarami zakresów możliwych wartości x, y, z oraz t. Jeżeli zakresy obserwacji nie są ściśle określone, wówczas informacja uzyskana z pomiaru nie może być określona i cena entropijna staje się nieskończenie duża.

Przypomnijmy różnicę między pojęciami „dokładność" i „pewność". W rozdziale 14 zdefiniowaliśmy pewność w oparciu o rozważania możliwości wystąpienia niepoprawnych odczytów spowodowanych fluktuacjami termicznymi w przyrządzie pomiarowym. Prosty rysunek wyjaśni różnicę między tymi pojęciami. Na rysunku 15.2 zakładamy, że używa się impulsów rozchodzących się wzdłuż linii prostej i impulsy te obserwowane są przez odpowiednie urządzenie odbiorcze. Całkowity czas trwania ciągu impulsów wynosi θ, a urządzenie odbiorcze może mierzyć najwęższy impuls o czasie trwania Δt. Dokładność zatem będzie wyrażona stosunkiem $\theta/\Delta t$.

Urządzenie odbiorcze nastawione jest na pewną graniczną wartość energii impulsów E_l, którą odbiornik ten może zarejestrować. Impulsy o energii



Rys. 15.2. Przykłady impulsów z szumem termicznym wyczerpujące cztery kombinacje dużej i małej dokładności z dużą i małą pewnością

mniejszej od E_l nie będą rejestrowane, podczas gdy impulsy o energii większej od E_l zostaną zarejestrowane. Jeżeli E_l jest dużo większa od średniej wartości fluktuacji, pewność jest duża, gdyż istnieje wówczas mała szansa przekroczenia progu przez fluktuacje. Niski próg E_l bliski jest natężenia fluktuacji termicznych, a to powoduje małą pewność. Omówioną sytuację ilustruje rys. 15.2.

15.2. POMIARY ODLEGŁOŚCI Z MAŁĄ DOKŁADNOŚCIĄ

Zacznijmy od bardzo prostego zagadnienia, mianowicie pomiaru długości pręta za pomocą miarki. Miarka podzielona jest na mniejsze części Δx (np. 1 mm), i tę miarkę przykładamy do mierzonego pręta, tak aby jeden

koniec pręta i początek miarki pokrywały się. Niezbędne jest do tego celu światło, by można było obserwować miarkę i pręt oraz patrzeć na każdą działkę miarki, tak aby wykryć tę, która pokrywa się z drugim końcem pręta. W celu uściślenia rozważania wyobraźmy sobie następującą sytuację: zakładamy, że cząstka znajduje się gdzieś w przedziale od 0 do L wzdłuż kierunku osi x, i chcemy zmierzyć położenie tej cząstki z dokładnością Δx. Aby to zrobić, odcinek $0L$ dzielimy na przedziały o długości Δx:

$$n = \frac{L}{\Delta x} = \frac{1}{\varepsilon_c} = \mathcal{A}. \tag{15.4}$$

Zagadnienie sprowadza się do wykrycia przedziału, w którym leży cząstka. Przesuwamy wiązkę światła, np. przy pomocy układu zwierciadeł, przez każdy z n przedziałów. Każdy przedział o długości Δx sprzężony jest z rezonatorem, który będzie przyjmował część światła odbitego przez cząstkę, jeżeli znajdzie się ona w danym przedziale. Każdy z rezonatorów utrzymywany jest w temperaturze T i chwilowo zakładamy, że posługujemy się światłem o częstotliwości niskiej ($h\nu \ll kT$).

Postępować można według jednego z dwu następujących sposobów:

Przypadek I. Zaczynamy od badania pierwszego rezonatora. Obserwujemy następnie każdy rezonator po kolei, aż zauważymy, że pewien rezonator został pobudzony. Pobudzenie może być spowodowane odbiciem światła od cząstki w i-tym przedziale ($i \leqslant n$) bądź może być spowodowane fluktuacjami. W pierwszym przypadku pobudzenie nazwiemy poprawnym, a w drugim fałszywym.

Przypadek II. Badamy jednocześnie wszystkie rezonatory. Załóżmy, że obserwujemy j pobudzonych rezonatorów ($1 \leqslant j \leqslant n$), przy czym tylko jeden pobudzony rezonator odpowiada obecności cząstki, pozostałe zaś pobudzenia są fałszywe.

Obliczamy uzyskaną ilość informacji i związany z nią przyrost entropii, aby wykazać, że

$$\Delta(S-I) \geqslant 0,$$

w każdym z tych dwóch przypadków. Aby to zrobić, wykorzystujemy wyniki z rozdz. 14 dotyczące obserwacji wymagających wielu rezonatorów. Podobnie jak w rozdz. 14, nie przywiązujemy wagi do odczytów negatywnych. Tylko odczyty pozytywne (obserwacje dużej energii rezonatora) dostarczają informa-

cji. I znów podobnie jak w rozdz. 14, przyrost entropii będzie granicą dolną.

Przypadek I. W przypadku tym obserwujemy cząstkę w przedziale i, przy czym $i \leqslant n$. W ten sposób cząstka lokalizowana była w jednym z i a priori możliwych położeń i stąd otrzymana ilość informacji wyniesie

$$\Delta I = k \ln i. \tag{15.5}$$

Załóżmy, że dokonano obserwacji i rezonatorów, z których jeden został pobudzony wskutek zaabsorbowania energii, która później ulegnie rozproszeniu w termostacie, tak że związany z nią przyrost entropii wyniesie zgodnie ze wzorem (14.26)

$$\Delta S \geqslant k A_i. \tag{15.5a}$$

Jeżeli i jest duże, to wykorzystując wzory (15.5), (15.5a) i (14.23) na A_i mamy

$$\Delta(S-I) \geqslant k[\ln i - \ln(\ln 2)] - k \ln i = -k \ln(\ln 2) > 0. \tag{15.6}$$

Oczywiście, i może nie być duże, a więc możemy zaobserwować cząstkę w pierwszym lub drugim przedziale. W tym przypadku nie możemy stosować

Tablica 15.1

	i			
	1	2	3	4
A_i	0,69	1,23	1,58	1,84
$\Delta I/k$	0,00	0,69	1,10	1,39
$\Delta(S-I)/k$	0,69	0,54	0,48	0,45

wyrażenia przybliżonego (14.23) na A_i, choć nierówność (15.6) jest spełniona. W tablicy 15.1 podano wyniki dla niektórych małych wartości i.

Przypadek II. W tym przypadku mamy j pozytywnych odczytów, tak że liczba możliwych pozycji cząstki została zredukowana z n do j. Uzyskana informacja wynosi

$$\Delta I = k \ln (n/j). \tag{15.7}$$

W tym przypadku j rezonatorów zaabsorbowało energię, która musi ulec później rozproszeniu, a związany z nią przyrost entropii, w związku z dokonaniem obserwacji wszystkich n rezonatorów, wyniesie

$$\Delta S \geqslant kjA_n. \tag{15.7a}$$

Posłużyliśmy się znów równaniem (14.26). Przy założeniu dość dużego n, pozwalającego wykorzystać wyrażenia przybliżone na A_n (wzór (14.23)), z uwzględnieniem wzorów (15.7) i (15.7a), mamy

$$\Delta(S-I) \geqslant k\left\{j\,[\ln n - \ln(\ln 2)] - \ln\frac{n}{j}\right\} =$$

$$= k\,[\ln j + (j-1)\ln n - j\ln(\ln 2)] > 0. \tag{15.8}$$

Wyrażenie (15.8) pokrywa się z (15.6), jeżeli $j = 1$, tzn. że tylko jeden odczyt jest pozytywny. Dla $j>1$ przypadek II dostarcza mniej informacji i daje większy przyrost entropii.

W dalszym ciągu tego rozdziału wykorzystamy wyniki uzyskane dla przypadku II biorąc $j = 1$. W rozpatrywanych przykładach przypadek ten daje najmniejszą różnicę pomiędzy przyrostem entropii a ilością informacji, a jednocześnie można założyć, że n jest duże, i stosować wyrażenie przybliżone na A_n.

Należy zaznaczyć, że inne metody pomiarowe będą prowadzić do jeszcze większego przyrostu entropii przy tej samej ilości uzyskanej informacji. Na przykład ktoś może ustawić rezonatory tak, aby przyjmowały one światło bezpośrednio; wówczas odczyt pozytywny oznacza brak cząstki. W tym przypadku za odczyt negatywny uważa się przesłonięcie wiązki świetlnej na danym odcinku. Łatwo sprawdzić, że taka metoda pomiaru daje dużo większy przyrost entropii, ponieważ większa część rezonatorów absorbuje energię.

15.3. POMIARY ODLEGŁOŚCI Z DUŻĄ DOKŁADNOŚCIĄ

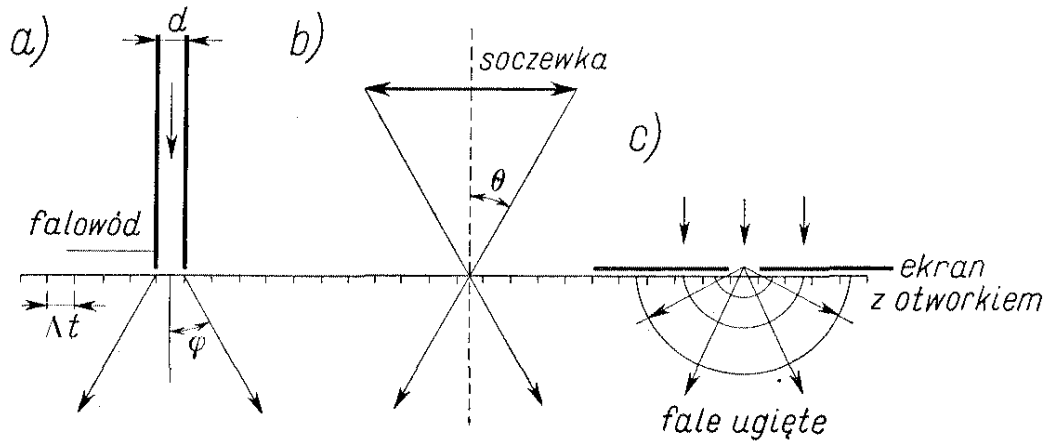
Poprzednie rozważanie oparte było na założeniu małej dokładności, tj. gdy dopuszczalny błąd Δx nie jest bardzo mały, tak że posługiwać się można światłem o małej częstotliwości. Gdy żądamy, aby błąd Δx był bardzo mały, niezbędne jest posługiwanie się światłem o krótkich falach, a zatem

o wielkiej częstotliwości promieniowania, tak by oświetlać w danej chwili tylko jeden z przedziałów.

Nie można praktycznie skupiać światła na pojedynczym przedziale, jeżeli długość fali światła λ nie jest mniejsza niż $2\Delta x$:

$$\lambda \leqslant 2\Delta x. \tag{15.9}$$

Na rysunku 15.3 przedstawiono trzy typy urządzeń oświetlających. W pierwszym (rys. 15.3a) wykorzystuje się falowód o średnicy $d = \Delta x$. Falowód ten ma dolną częstotliwość odcięcia, tzw. *częstotliwość krytyczną* odpowiadającą



Rys. 15.3. Urządzenia służące do skupiania światła na odcinku Δx: a) falowód o grubości $d = \Delta x$; b) soczewka o aperturze θ; c) przesłona ze szczeliną o szerokości Δx

$^1/_2\lambda = d$ i może przesyłać częstotliwości powyżej tej granicy. Odpowiada to dokładnie warunkowi (15.9). Światło wychodzi z falowodu pod kątem φ, który staje się coraz mniejszy ze wzrostem częstotliwości. Wiązkę można skupić na pojedynczym odcinku Δx.

Na rysunku (15.3b) wykorzystujemy soczewkę o aperturze θ. Ognisko, zgodnie z dobrze znanym wzorem na zdolność rozdzielczą soczewki, ma szerokość

$$\Delta x = \frac{\lambda}{2 \sin \theta}.$$

Daje to znów warunek (15.9). Na rysunku 15.3c próbujemy posłużyć się długimi falami i przesłoną z otworkiem z szerokości Δx, lecz dyfrakcja powo-

duje rozchodzenie się światła na sąsiednie przedziały i na pojedynczym odcinku Δx nie można skupić dokładnie światła.

Z powyższego wynika, że aby można było otrzymać dobre skupienie światła, częstotliwość światła używanego do obserwacji musi spełniać warunek

$$\nu = \frac{c}{\lambda} \geqslant \frac{c}{2\Delta x}. \tag{15.10}$$

Zdefiniujemy teraz charakterystyczną długość związaną z temperaturą T, w której znajdują się rezonatory w czasie obserwacji:

$$\frac{h\nu}{kT} = \frac{hc}{2kT\Delta x} = \Delta x_T \frac{1}{2\Delta x}, \tag{15.11}$$

gdzie

$$\Delta x_T = \frac{hc}{kT} \approx \frac{1{,}44}{T}. \tag{15.12}$$

Przyjęto tu minimalną wartość ν spełniającą warunek (15.10).

Możemy powołać się teraz na warunki z rozdz. 15.2, gdzie wymagano małej częstotliwości, tak że

$$\frac{h\nu}{kT} \ll 1,$$

co prowadzi do zależności

$$\Delta x \gg \Delta x_T/2. \tag{15.13}$$

Oznacza to, że dokładność może być mała, gdyż długość przedziału Δx nie może być mniejsza od pewnej małej wartości.

Z drugiej strony, dla uzyskania dużej dokładności żądamy, aby

$$\Delta x \ll \frac{\Delta x_T}{2}, \tag{15.13a}$$

co prowadzi (wzór (15.11)) do nierówności

$$\frac{h\nu}{kT} \gg 1. \tag{15.14}$$

Przy omawianiu wielkich częstotliwości będziemy chcieli jednak bardziej precyzyjnie określić wartość $h\nu$. Z poprzednich rozważań możemy przyjąć,

że wielkim częstotliwościom odpowiada

$$h\nu > E_l = kTA_n, \tag{15.14a}$$

gdzie E_l jest graniczną energią oscylatorów małych częstotliwości.

Jeżeli przyjąć, że n jest duże, tj. posługujemy się dużą ilością rezonatorów, wówczas warunek ten przyjmuje postać

$$h\nu > kT[\ln n - \ln(\ln 2)]. \tag{15.14b}$$

W tym przypadku, gdy zakładamy wielkie częstotliwości, możemy obliczyć $\Delta(S-I)$ w sytuacjach opisanych przy końcu rozdz. 15.2 (obserwacje n rezonatorów dają tylko jeden odczyt pozytywny). Ponieważ absorpcja pojedynczego kwantu wystarcza na jeden odczyt, mamy wówczas

$$\left.\begin{aligned} \Delta S &= \frac{h\nu}{T} > k[\ln n - \ln(\ln 2)], \\ \Delta I &= k \ln n, \\ \Delta(S-I) &> k \ln(\ln 2) > 0, \end{aligned}\right\} \tag{15.15}$$

tak jak w przypadku, gdy posługujemy się małymi częstotliwościami. W rzeczywistości przyrost entropii może być tu dużo większy, gdyż ν jest bardzo duże. Nie możemy jednak zapomnieć o rozważaniach z rozdz. 14.6, gdzie podano możliwość zmniejszenia rozpraszanej części energii $h\nu$. Nie możemy jednak uzyskać wyników lepszych niż w przypadku, gdy rozważamy małe częstotliwości, gdyż prędzej czy później obserwacja musi być dokonana za pomocą przyrządu pracującego na małych częstotliwościach. Uogólniona zasada Carnota jest więc spełniona i rozważymy później rolę, jaką odgrywa duża ilość energii wymaganej w doświadczeniu. Podobne sytuacje wystąpią w innych przykładach.

Pozostaje przypadek $h\nu \approx kT$; będzie on wymagać specjalnego omówienia w każdej poszczególnej sytuacji. Jeżeli na przykład A_n jest tak małe, że spełniona jest nierówność (15.14a), ale nie jest spełniona (15.14), to sytuacja wymaga szczegółowego rozważenia, opartego na dokładnych wyrażeniach różnych wartości średnich, niezbędnych przy wyprowadzaniu wartości A_n.

Należy podkreślić, że wartości granicznej Δx_T, podanej przez wzór (15.12), nie nadajemy specjalnego sensu fizycznego. Przy temperaturze T wartość ta określa po prostu granicę między różnymi metodami obliczeń; obie metody prowadzą jednak do tych samych wyników.

15.4. SPRAWNOŚĆ OBSERWACJI

Sprawność doświadczalnej metody obserwacji można zdefiniować jako stosunek uzyskanej ilości informacji ΔI do straty negentropii $|\Delta N|$, którą jest przyrost entropii ΔS układu związany z obserwacją:

$$\mathcal{E} = \frac{\Delta I}{\Delta N} = \frac{\Delta I}{\Delta S}, \qquad \Delta N = -\Delta S. \tag{15.16}$$

Gdy dokładność jest mała, to dochodzimy do warunków z rozdz. 15.2 i uwzględniając przypadek II, dla jednego pozytywnego odczytu mamy

$$\mathcal{E} = \frac{\Delta I}{\Delta N} = \frac{\ln n}{A_n} \approx \frac{1}{1 - \dfrac{\ln(\ln 2)}{\ln n}} \approx 1 + \frac{\ln(\ln 2)}{\ln n}; \tag{15.17}$$

przy czym n jest duże, a pewność równa się 2.

Jeżeli dokładność jest większa ($\mathcal{A} = n$ jest duże), to większa jest sprawność obserwacji $\mathcal{E}$ i współczynnik ten zmierza do jedności dla bardzo dużych wartości n.

Wzór (15.17) odnosi się do przypadku małej pewności z prawdopodobieństwem $^1/_2$ tego, iż obserwacja spowodowana jest fluktuacjami termicznymi. Możemy posłużyć się wzorami z rozdz. 14.5 odpowiadającymi zagadnieniom wymagającym dużych pewności. Dowiedliśmy (wzór (14.34)), że jeżeli pewności r oraz n są duże, to

$$A_n = \ln rn, \qquad r \gg 1, \qquad n \gg 1$$

daje przyrost entropii

$$\Delta S = k \ln rn \tag{15.18}$$

i sprawność

$$\mathcal{E}_r = \frac{\Delta I}{\Delta S} \approx \frac{\ln n}{\ln rn} = \frac{1}{1 + \dfrac{\ln r}{\ln n}}. \tag{15.19}$$

Sprawność będzie zależała od wartości n i r.

Zwróciliśmy uwagę na to, że duża dokładność wymaga wielkich częstotliwości. Jeżeli jednak istnieją dodatkowe urządzenia zmniejszające ilość rozpraszanej energii, to powyższe relacje, słuszne w przypadku małych częstotli-

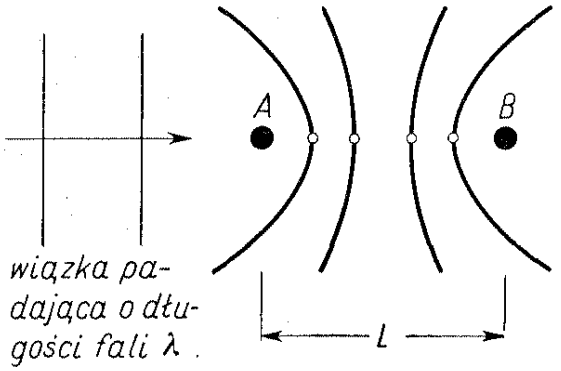
wości, zachowają również swoją ważność w przypadku wielkich częstotliwości. Jeżeli urządzeń tych nie ma, to wzór (15.19) ulega modyfikacji i otrzymujemy

$$\mathcal{E}_r = \frac{\Delta I}{\Delta S} \approx \frac{\ln n}{hv} \tag{15.19a}$$

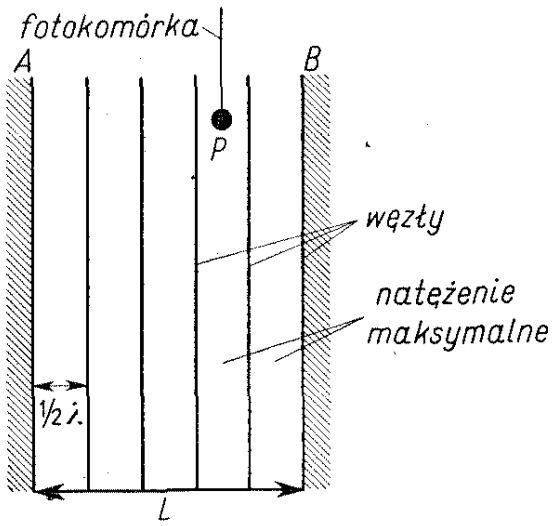
z $hv > E_l \approx \ln rn$, tak że sprawność wychodzi mniejsza.

15.5. POMIAR ODLEGŁOŚCI ZA POMOCĄ INTERFEROMETRU

Metoda pomiaru odległości lub położenia rozpatrzona w tym rozdziale daje względnie mały przyrost entropii; innymi słowy, sprawność pomiaru, przynajmniej przy małych pewnościach, zbliża się do jedności i wzrasta wraz ze wzrostem dokładności. Metoda ta jest również bardzo efektywna przy



Rys. 15.4. Pasma interferencyjne wytwarzane przez światło rozpraszane przez dwie cząstki A i B odległe od siebie o L



Rys. 15.5. Fale stojące między dwiema równoległymi płytkami A i B odbijającymi światło i ustawionymi od siebie w odległości L. Fotokomórka P służy do określania powierzchni o maksimum natężenia

większej pewności, przy czym dokładna wartość sprawności zależy (w myśl wzoru (15.19)) od pewności i dokładności.

Pomiary interferometryczne są bardzo kosztowne z punktu widzenia przyrostu entropii i wydaje się, że sprawność jest mała nawet przy małych pewnoś-

ciach. Jak zobaczymy, głównym powodem tego jest konieczność rejestracji wielu pozytywnych odczytów.

Wyobraźmy sobie, że w celu pomierzenia odległości L między punktami A i B umiejscawiamy w punktach A i B cząstki materialne. Punkty te oświetlamy, a następnie obserwujemy obraz interferencyjny światła rozproszonego przez te cząstki. Liczba n pasm interferencyjnych między punktami A i B daje liczbę połówek fal na odcinku L (rys. 15.4):

$$n\frac{\lambda}{2} = L. \tag{15.20}$$

Dla większej dokładności, musimy wziąć pod uwagę rzeczywisty interferometr, np. Fabry-Perota, który stosuje się do pomiaru długości wzorca metra. Ograniczając się tylko do głównych cech metody, wyobraźmy sobie dwie płytki równoległe A i B oraz system fal stojących między tymi płytkami, tak jak pokazano na rys. 15.5. Aby znaleźć odległość między płytkami, wystarczy policzyć ilość powierzchni o maksimum natężenia między płytkami A i B. Posłużyć się można w tym celu fotoelementem P, który może się przesuwać od A do B.

Wzór (15.20) daje częstotliwość światła

$$\nu = \frac{nc}{2L}. \tag{15.21}$$

Jeżeli wykrycie powierzchni o maksimum natężenia wymaga absorpcji q kwantów $h\nu$ przez fotoelement P, wówczas całkowita wymagana energia promieniowania wyrazi się wzorem:

$$\Delta E = nqh\nu = n^2 q\frac{hc}{2L}. \tag{15.22}$$

Dla małych częstotliwości liczba q będzie duża, ale przy wykorzystaniu częstotliwości wielkich jeden kwant ($q = 1$) na pasmo może okazać się wystarczający. Rozważmy to bardziej szczegółowo.

Jeżeli całkowita energia ΔE zostanie zaabsorbowana i zamieniona w termostacie na ciepło, wówczas odpowiadający jej przyrost entropii podczas obserwacji wyniesie

$$\Delta S = \frac{\Delta E}{T} = n^2 q\frac{hc}{2L} = kn^2 q\frac{\Delta x_T}{2L}. \tag{15.23}$$

Wprowadziliśmy tu znowu długość charakterystyczną Δx_T dla temperatury T, zdefiniowaną przez wzór (15.12). Załóżmy, że korzystamy z najmniejszej liczby q kwantów, dopuszczalnej w istniejących warunkach doświadczenia. (Wrócimy jeszcze do tego później). Mierzymy odległość L z błędem nie przekraczającym $\frac{\lambda}{2}$:

$$\Delta L \approx \frac{\lambda}{2} = \frac{L}{n}. \tag{15.24}$$

Ponieważ spełnione są warunki (15.23), więc

$$\Delta S = knq \frac{\Delta x_T}{2\Delta L}. \tag{15.25}$$

Zastanówmy się teraz nad tym, ile kwantów q potrzeba do obserwacji położenia pasma.

15.5.1. Duża odległość, mała dokładność

Warunki te odpowiadają małym kwantom i małym częstotliwościom obserwowanym w rezonatorach umieszczonych w termostacie o temperaturze T:

$$h\nu \ll kT, \tag{15.26}$$

$$\frac{h\nu}{kT} = n \frac{hc}{2LkT} = \frac{\Delta x_T}{2\Delta L} \ll 1.$$

Mamy więc teraz sytuację rozważaną w rozdz. 14.4, gdzie omówiliśmy ogólną zasadę działania urządzenia doświadczalnego, w którym wykorzystuje się dużą liczbę n równoczesnych obserwacji. Uzyskaliśmy teraz następujący wynik dla przypadku pewności równej 2 (prawdopodobieństwo błędu spowodowanego fluktuacjami wynosi $^1/_2$):

$$E_l = qh\nu = A_n kT \quad \text{przy} \quad A_n > 0{,}7.$$

Daje to

$$q = A_n \frac{kT}{h\nu} = 2A_n \frac{\Delta L}{\Delta x_T} \gg 1.$$

Współczynnik A_n osiąga wartość kilku jednostek. Większa pewność prowadzi do większej wartości A_n.

Ze wzoru (15.25) możemy obliczyć przyrost entropii dla rozważanego przypadku, ponieważ w tym przypadku nie ma wątpliwości, że cała energia ΔE rozpraszana zostaje w czasie obserwacji. Daje to nam

$$\Delta S = knA_n. \tag{15.27}$$

Konkretyzując załóżmy, że obserwacja ma na celu pomiar położenia punktu B względem punktu A. Tak jak w poprzednich rozważaniach uzyskujemy ilość informacji:

$$\Delta I = k \ln \left(\frac{L}{\Delta L}\right) = k \ln n, \tag{15.28}$$

natomiast sprawność obserwacji dla dużych n wyniesie

$$\mathcal{E} = \frac{\Delta I}{\Delta S} = \frac{\ln n}{nA_n} \approx \frac{\ln n}{n[\ln n - \ln(\ln 2)]} \approx \frac{1}{n}\left[1 + \frac{\ln(\ln 2)}{\ln n}\right] \leqslant 1. \tag{15.29}$$

Jest ona dużo mniejsza niż w poprzednich przypadkach, gdzie żądano, aby tylko jedna obserwacja była pozytywna.

Gdybyśmy wymagali większej pewności, z małym prawdopodobieństwem fałszywych obserwacji powodowanych fluktuacjami termicznymi, to trzeba by A_n obliczać ze wzoru z rozdz. 14.5. Przyrost entropii będzie n razy większy niż w odpowiednim przypadku z rozdz. 14.3, a sprawność zmaleje w stosunku $\frac{1}{n}$.

Jakakolwiek byłaby pewność, A_n jest zawsze większe od $\ln n$ i negentropijna zasada informacji prowadzi do uogólnionej zasady Carnota

$$\Delta(S-I) = k(nA_n - \ln n) \geqslant 0. \tag{15.30}$$

15.5.2. Małe odległości lub duża dokładność i niska temperatura

W tych warunkach niezbędne są wielkie częstotliwości:

$$h\nu \gg kT, \qquad \frac{h\nu}{kT} = \frac{\Delta x_T}{2\Delta L} \gg 1. \tag{15.31}$$

Warunki te powodują to, że trzeba daleko wybiegać poza granice widma promieniowania ciała czarnego o temperaturze T. Można wykorzystać kwanty o małym natężeniu, przypadające na każde pasmo, mimo to i tak całkowita

ilość energii potrzebnej do pomiaru wyniesie

$$q = 1; \quad \Delta E = nhv = \frac{nhc}{2\Delta L}; \quad L = \frac{L}{n}. \tag{15.32}$$

Znaleźliśmy się w sytuacji opisanej w rozdz. 14.6 i musimy rozróżnić całkowitą, wymaganą do obserwacji energię, od tej jej części, która w końcu musi ulec rozproszeniu w postaci ciepła ΔQ. Jak wyjaśniono, musimy posługiwać się fotokomórką z polem hamującym, chociaż ze względów teoretycznych w praktyce byłby to bardzo niesprawny przyrząd. Pozwoli to nam zmniejszyć energię rozpraszaną do takiej samej wartości jak przy małych częstotliwościach (rozdz. 15.5.1) i zastosować wzory (15.27)÷(15.30).

Należy mieć jednak na uwadze fakt, że do doświadczenia wymagana jest duża energia ΔE, na co wskazuje równość (15.32), którą przepisać można w postaci

$$\Delta E\,\Delta L = \frac{1}{2}\,nhc; \quad n = \frac{L}{\Delta L} = \mathcal{A} = \text{dokładność}. \tag{15.33}$$

Przytoczymy później interpretację fizyczną tego wzoru, który wskazuje na to, że energia ΔE może wzrastać bez ograniczenia, jeżeli ΔL staje się coraz mniejsze.

15.6. INNY SCHEMAT POMIARU ODLEGŁOŚCI

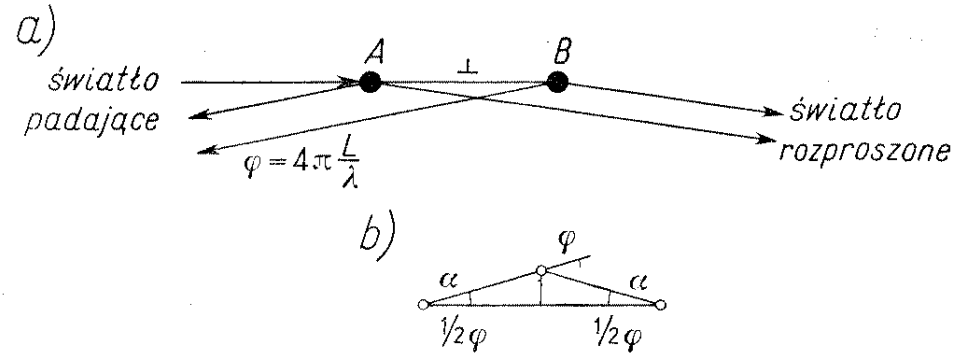
Poprzednie rozważania uwydatniają duże trudności mierzenia ekstremalnie małych odległości przy pomocy interferometrów oraz wysoką cenę takich pomiarów. Z punktu widzenia ważności tego wniosku jest słuszne, by wykazać, że inne metody są niemniej kosztowne. Ktoś np. pomyśleć może, że trudności, wymienione poprzednio w tym rozdziale mogą wynikać z faktu posługiwania się falami dużo krótszymi od mierzonych odległości. Doprowadziło to do skrajnie krótkich fal w zagadnieniu pomiaru bardzo małych odległości. Możemy postępować według innej procedury i opierać się na małych częstotliwościach, a więc wykorzystywać promieniowanie o falach dużo dłuższych niż odległość L. Stosownie do wzoru (15.24) i definicji Δx_T (15.12) rozważmy następujące warunki:

$$hv \ll kT, \quad \lambda \gg \Delta x_T. \tag{15.34}$$

Mamy również

$$\lambda \gg L. \tag{15.35}$$

Sytuację tę zilustrowano na rys. 15.6. Będziemy obserwowali w tym przypadku natężenie światła odbitego przez cząstki A i B w kierunku zgodnym z kierunkiem promieniowania padającego i w kierunku przeciwnym. Fale rozpraszane



Rys. 15.6. Światło rozpraszane przez cząstki A i B w kierunku światła padającego jest w fazie (rys. a). Światło rozpraszane wstecz ma fazę $\varphi = 4\pi L/\lambda$ w stosunku do światła padającego. Podany wykres wektorowy (rys. b) odnosi się do rozpraszania wstecznego. Częstotliwość światła jest mała, a amplituda wynosi α

przez cząstki A i B w kierunku zgodnym z promieniowaniem padającym są w fazie. Niech α będzie amplitudą fali rozpraszanej przez każdą z cząstek. Całkowite natężenie fal rozpraszanych w kierunku zgodnym z promieniowaniem padającym wyniesie

$$i_f = 4\alpha^2, \tag{15.36}$$

podczas gdy fale odbite w kierunku przeciwnym do kierunku światła padającego mają różnicę fazy φ i dają wypadkowe natężenie

$$i_b = 4\alpha^2 \cos^2\left(\frac{\varphi}{2}\right) \approx 4\alpha^2\left(1 - \frac{\varphi^2}{4}\right); \tag{15.37}$$

natomiast

$$\varphi = \frac{4\pi L}{\varphi} \tag{15.38}$$

i faza ta jest mała ze względu na wzór (15.35), tak że rozłożenie $\cos^2\left(\frac{\varphi}{2}\right)$ we wzorze (15.37) jest słuszne.

Różnica $i_f - i_b$ jest miarą L:

$$i = i_f - i_b = \frac{i_f \varphi^2}{4} = i_f 4\pi^2 \left(\frac{L}{\lambda}\right)^2. \tag{15.39}$$

W celu zmierzenia i_f oraz i_b fotokomórka musi absorbować światło rozproszone, względnie musi je odebrać jakiś odbiornik. W takim razie energia rozproszona równa jest $i_f + i_b$ i suma ta jest w przybliżeniu równa $2i_f$.

Związany z tym przyrost entropii wynosi zgodnie ze wzorem (15.39)

$$\Delta S = 2\frac{i_f}{T} = i\frac{\left(\frac{\lambda}{L}\right)^2}{2\pi^2 T}. \tag{15.40}$$

Pobudzenie termiczne w odbiornikach jest rzędu kT (wzór (15.34)) i stąd mamy błędy rzędu kT przy pomiarze i_f oraz i_b, tak że

$$\Delta i = 2kT \tag{15.41}$$

est błędem pomiaru różnicy natężenia i. Musimy więc stosować duże natężenia w celu uzyskania dostatecznej dokładności.

Jeżeli założymy, że

$$i = \beta kT, \quad \beta \gg 1, \quad i_f = fkT, \tag{15.41a}$$

to ze wzorów (15.39) i (15.41a) otrzymujemy

$$i = \beta kT = fkT\, 4\pi^2 \left(\frac{L}{\lambda}\right)^2,$$

co po uwzględnieniu wzoru (15.38) daje

$$\frac{\beta}{f} = 4\pi^2 \left(\frac{L}{\lambda}\right)^2 \ll 1 \quad \text{lub} \quad f \gg \beta \gg 1.$$

Następnie ze wzoru (15.41) otrzymujemy

$$\frac{\Delta i}{i} = \frac{2}{\beta} \tag{15.42}$$

i dalej ze wzoru (15.39):

$$\Delta i = \Delta i_f\, 4\pi^2 \left(\frac{L}{\lambda}\right)^2 + i_f 4\pi^2 \frac{L}{\lambda^2} 2\Delta L$$

lub ze wzoru (15.42)

$$\frac{\Delta i}{i} = \frac{\Delta i_f}{i_f} + \frac{2\Delta l}{L} = \frac{1}{f} + \frac{2\Delta L}{L} = \frac{2}{\beta}.$$

Ostatnia równość i warunek $f \gg \beta$ dają

$$\frac{\Delta L}{L} \approx \frac{1}{\beta}. \tag{15.42a}$$

Współczynnik β spełnia tu rolę dokładności $\mathcal{A}$ w pomiarze L, tak że wyrażenie (15.40) oznaczające przyrost entropii ΔS_l staje się w myśl wzoru (15.41a) równe

$$\Delta S_l = k \frac{\mathcal{A}}{2\pi^2} \left(\frac{\lambda}{L}\right)^2, \tag{15.43}$$

przy założeniu, że wykorzystujemy fale długie o $\lambda \gg L$.

Otrzymany wynik porównamy teraz ze wzorami z rozdz. 15.5. Rozważymy ponownie dwa przypadki.

15.6.1. Fale krótkie, duże odległości (wzory (15.26) i (15.27))

Mamy tu do czynienia z warunkiem

$$L \gg \Delta x_T. \tag{15.44}$$

Cena entropijna przy stosowaniu fal krótkich wynosi

$$\Delta S_{S,A} = knA_n, \quad \text{dla} \quad n = \mathcal{A} \quad \text{oraz} \quad A_n \approx \ln(rn), \tag{15.45}$$

gdzie r jest pewnością eksperymentu. Ostatecznie metoda krótkofalowa daje

$$\Delta S_{S,A} = k\mathcal{A} \ln(r\mathcal{A}), \tag{15.46}$$

co oznacza przemnożenie $k\mathcal{A}$ przez współczynnik rzędu kilku lub najwyżej kilkudziesięciu.

Jeżeli zastosujemy fale długie, to ΔS_l wyraża się wzorem (15.43). Aby był spełniony wyrunek (15.35), bierzemy

$$\lambda = \gamma L, \quad \gamma \gg 1.$$

Wartość γ rzędu kilkudziesięciu wystarcza i jeżeli uwzględni się warunek (15.44), to warunek (15.34) zostaje automatycznie spełniony. Ostatecznie

$$\Delta S_I = k\mathcal{A}\frac{\gamma^2}{2\pi^2}, \tag{15.47}$$

gdzie mnożnik $\gamma^2/2\pi^2$ dochodzi do kilkudziesięciu.

Także i w tym przypadku obie metody prowadzą do zbliżonych rzędów wielkości negentropijnej ceny pomiaru.

15.6.2. Fale krótkie, małe odległości (wzory (15.31), (15.32) i (15.34))

Przyjmijmy

$$L \ll \Delta x_T. \tag{15.48}$$

Zgodnie ze wzorem (15.32) wymagana energia fal krótkich wynosi

$$\Delta E = n^2\frac{hc}{2L} = kTn^2\frac{\Delta x_T}{2L},$$

przy czym tylko część jej ulega niezbędnemu rozproszeniu. Przyjmując najbardziej niekorzystny przypadek (cała energia ulega rozproszeniu) otrzymamy przyrost entropii

$$\Delta S_{S,B} \geqslant kn^2\frac{\Delta x_T}{2L} = k\mathcal{A}^2\frac{\Delta x_T}{2L}. \tag{15.49}$$

Stosując wzór (15.43) dla przypadku fal długich, wracamy znów do warunków (15.34) i (15.35). Bardziej ścisły okazuje się teraz warunek (15.34). Przyjmujemy

$$\gamma' h\nu = kT \quad \text{dla} \quad \gamma' \gg 1,$$

co prowadzi do

$$L = \gamma'\Delta x_T \quad \text{oraz} \quad \frac{\lambda}{L} = \frac{\Delta x_T}{L}.$$

Współczynnik γ' rzędu kilku jest dostatecznie duży, aby uczynić energię cieplną równą kT, jak tego wymaga równość (15.41). Mamy przeto

$$\Delta S_I = k\mathcal{A}\frac{\gamma'^2}{2\pi^2}\left(\frac{\Delta x_T}{L}\right)^2. \tag{15.50}$$

Mnożnik $\left(\frac{\gamma'}{\pi}\right)^2$ jest rzędu kilku. Dla fal długich przy małych odległościach L, cena entropijna wzrasta jak L^{-2} (wzór 15.50)), podczas gdy dla fal krótkich jak L^{-1} (wzór (15.49)). Przy dostatecznie małych odległościach metoda posługiwania się falami krótkimi daje niższą cenę entropijną od metody posługiwania się falami długimi; rzeczywiście, przy pomiarach bardzo małych odległości fale długie są bardzo kosztowne. Procedura podana w rozdz. 15.5 jest wygodniejsza niż przytoczona tutaj.

Rozważania nie są jeszcze zupełne. Przyrost entropii zależy od dokładności $\mathscr{A}$ i wzrasta jak $\mathscr{A}^2$ dla fal krótkich i jak $\mathscr{A}$ dla fal długich. Jednak przy danej dokładności główne wnioski, wyciągnięte wyżej, są ważne — przynajmniej dla dostatecznie małych odległości. Przypadek *umiarkowanie* małych odległości wymagałby ściślejszego przebadania, lecz nie ma to większego znaczenia praktycznego.

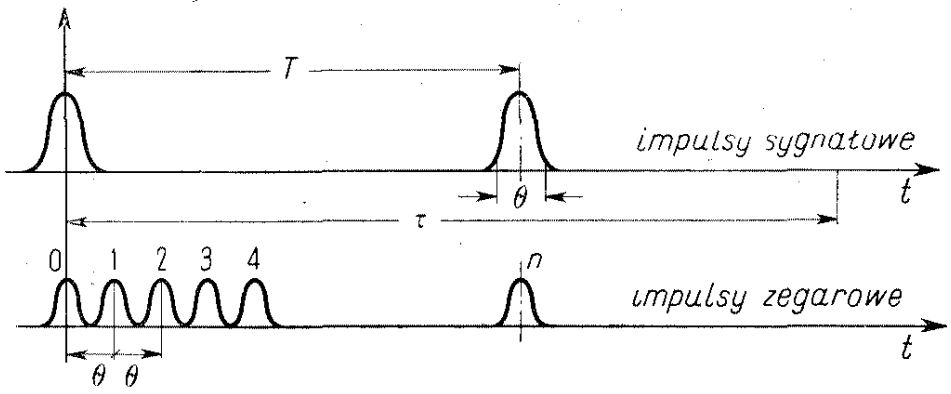
15.7. POMIAR OKRESÓW CZASU

Po szczegółowym omówieniu pomiaru odległości wrócimy teraz do innego fundamentalnego zagadnienia: mierzenia czasu. W większości zegarów mechanizm przesuwa wskazówki po tarczy. Sprowadza to po prostu pomiar czasu do pomiaru kąta lub odległości. Obserwacji w tym przypadku dokonuje się tak samo, jak opisano poprzednio. Procedura taka jednak nie jest bardzo dokładna i będziemy rozważali inny sposób pomiaru, który odpowiada zasadniczo metodzie stosowanej w chronometrach elektronicznych, gdzie można uzyskiwać bardzo dużą dokładność. Okres czasu t określamy jako odległość w czasie między dwoma impulsami, które mogą być impulsami elektrycznymi przesyłanymi po linii do odbiornika. Sytuację tę ilustruje rys. 15.7. Pierwszy impuls pokrywa się z impulsem przychodzącym z zegara elektronicznego. Zegar elektroniczny wysyła ciąg wąskich impulsów o czasie trwania θ i przerwie między impulsami równej również θ. Każdy impuls zegara służy do włączenia pewnego odbiornika. Ponumerowane od 1 do n odbiorniki podłączane są kolejno do linii elektrycznej, z której spodziewamy się nadejścia impulsu. Jeżeli nie ma sygnału w okresie włączenia odbiornika, to nic się nie dzieje. Jeżeli sygnał się pojawi, to odbiornik zareaguje. Odnotujemy np., że odbiornik O i odbiornik n zareagowały; wówczas dochodzimy do wniosku,

że przedział czasu miał czas trwania

$$t = n\theta. \tag{15.51}$$

Mechanizm przełączający nie wymaga z zasady żadnej energii. Może on działać w oparciu o zmianę napięcia na diodzie krystalicznej, która jest zatkana do chwili pojawienia się impulsu. W prostszy sposób dokonać tego można zmianą napięcia na siatce lampy. Kiedy pojawi się sygnał, to musi on być zarejestro-



Rys. 15.7. Dwa sygnały o czasie trwania θ pojawiające się z odstępem czasu T w możliwym przedziale τ. Zegar elektroniczny generuje krótkie impulsy z okresem θ uaktywniające odbiorniki dla sygnałów, między którymi chcemy mierzyć odstęp czasu

wany, a to oznacza już rozpraszanie energii. Można pomyśleć o użyciu nastrojonego odbiornika, ale wzmacniacz jest nieodzowny. Jest to właśnie pomocniczy przyrząd do wzmocnienia mocy i wobec tego nie będziemy rozważać energii wymaganej przez sam system wzmacniający.

Aby w zagadnieniach tego typu obliczyć ilość informacji i negentropijną cenę eksperymentu, musimy określić całkowity czas trwania doświadczenia. Wykorzystamy definicje podobne do wprowadzonych w rozdz. 15.1.

Rozpatrujemy błąd proporcjonalny ε_c i dokładność $\mathcal{A}$:

$$\varepsilon_c = \frac{\Delta t}{\tau} = \frac{\theta}{\tau} = \frac{1}{\mathcal{A}}. \tag{15.52}$$

W ogólności sytuacja jest podobna do omówionej poprzednio przy pomiarach odległości.

Ilość otrzymanej informacji zgodnie z wzorem (15.3) wynosi

$$\Delta I = k \ln \mathcal{A}, \tag{15.53}$$

podczas gdy cena entropijna obserwacji wyniesie

$$\Delta S \geqslant 2kA_{n+1}. \tag{15.54}$$

Ponieważ obserwowaliśmy dwa pozytywne odczyty na $(n+1)$ rezonatorach, za A_{n+1} można przyjąć wartość podaną przez wzór (14.23) dla przypadku o pewności 2 (50% prawdopodobieństwa błędu) lub

$$A_{n+1} = \ln[r(n+1)]$$

dla zagadnień wymagających dużej pewności (rozdz. 14.6). Uogólniona zasada Carnota jest spełniona, tzn.

$$\Delta(S-I) \geqslant 0. \tag{15.55}$$

W zagadnieniach pomiaru czasu, na mocy relacji nieoznaczoności, pozostajemy zawsze przy warunkach odpowiadających przypadkowi z rozdz. 15.5.1. Impulsy o czasie trwania θ zajmują pasmo

$$\Delta\nu = 2\nu_M = \frac{1}{\theta}.$$

Wykażemy, że warunek ten prowadzi do małych częstotliwości:

$$h\nu_M < kT, \qquad \text{skoro } \theta = \frac{1}{2\nu_M} > \frac{h}{2kT}, \tag{15.56}$$

co można przepisać w postaci

$$\theta > \frac{\Delta x_T}{2c} = \frac{1{,}44 \cdot 10^{-10}}{6T} \text{ sekund}, \tag{15.57}$$

gdzie

T — temperatura w stopniach Kelvina,

Δx_T — odległość charakterystyczna odpowiadająca temperaturze T, określona wzorem (15.12).

Warunek (15.56) praktycznie zawsze jest spełniony. Można go badać również z punktu widzenia relacji nieoznaczoności.

Jak wiadomo, czas t i energia E stanowią parę zmiennych sprzężonych. Dlatego, gdy mierzymy t z dokładnością Δt, to na pewno wprowadzamy nieznaną perturbację ΔE w pomiarze energii E, taką że

$$\Delta E\, \Delta t \approx h. \tag{15.58}$$

Nieoznaczoność ΔE trzeba porównać z normalnymi fluktuacjami energii ε, wynikającymi z ruchu termicznego w temperaturze T:

$$\Delta E \approx \sqrt{\overline{\varepsilon^2}} = \sqrt{n}\, kT, \tag{15.59}$$

gdzie n oznacza liczbę aktywnych stopni swobody układu lub ilość drgań małej częstotliwości ($hv \leqslant kT$).

Drgania wielkiej częstotliwości w praktyce nie są wzbudzane i wywierają mały wpływ na fluktuacje energii. Tak więc trzeba porównać ΔE oraz $\Delta_1 E$. Wchodzą tu w grę dwa przypadki.

$$\text{Przypadek 1. Jeżeli } \Delta E \ll \Delta_1 E, \text{ to } \Delta t \gg \frac{h}{\sqrt{n}\, kT}. \tag{15.60}$$

Jest to przypadek teraz rozważany. Perturbacje spowodowane pomiarem czasu można więc w tym przypadku pominąć. Warunki (15.57) i (15.60) różnią się tylko współczynnikiem $\sqrt{n}$.

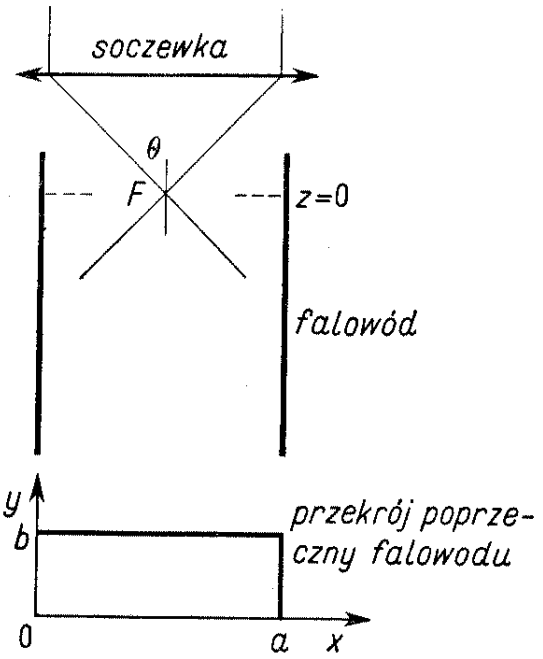
$$\text{Przypadek 2. Jeżeli } \Delta E \geqslant \Delta_1 E, \text{ to } \Delta t \leqslant \frac{h}{\sqrt{n}\, kT}. \tag{15.61}$$

Warunki te mogą wywołać znaczne perturbacje układu. Duża dokładność w określaniu czasu mogłaby nawet uszkodzić obserwowany układ, gdyż potrzebne byłyby wielkie częstotliwości związane z dużymi kwantami hv rzędu ΔE lub większymi.

15.8. OBSERWACJE POD MIKROSKOPEM

Omówione w niniejszym rozdziale zagadnienia pomiaru odległości i czasu prawie całkowicie pokrywają swoimi zastosowaniami zakres zagadnień, z jakimi mamy do czynienia w miernictwie. Na przykład pomiar prądu elektrycznego polega na odczycie wskazania amperomierza, a to sprowadza się w zasadzie do pomiaru położenia na skali, co z kolei można uważać za pomiar odległości. Pomiar odległości, pomiar czasu oraz liczenie stanowią główne metody obserwacji eksperymentalnych. Nasze rozważania pokazują, jak do tych zagadnień zastosować negentropijną zasadę informacji. Zagadnienia pomiarów są jednak tak ważne, że prześledzimy jeszcze inne, dodatkowe przykłady. Zastanówmy się najpierw nad określaniem położenia cząstki za pomocą mikroskopu koncentrującego światło w ognisku F, jak pokazano na rys. 15.8. W tym celu niezbędny jest pewien zakres częstotliwości i bardzo duża ilość fotonów padających na soczewkę. Zagadnienie zostaje całkowicie zdefiniowane, jeżeli ograniczymy pole obserwacji, tak jak zilustrowano to na rys. 15.8.

Światło przechodzi przez falowód, którego przekrój poprzeczny może być prostokątem o bokach a i b leżących na osiach układu współrzędnych x i y. Falowód określa dyskretny zbiór fal własnych*, których superpozycja prowadzi do powstania ogniska na płaszczyźnie ogniskowania** o współrzędnej $z = 0$.



Rys. 15.8. Falowód ogranicza pole obserwacji, jakie może zapewnić soczewka o aperturze θ. Ognisko soczewki F leży na płaszczyźnie $z = 0$ wewnątrz falowodu o przekroju prostokątnym

Wybierzmy falę H z pionowo skierowanym polem magnetycznym (poprzecznym polem elektrycznym), którego pochodna normalna musi być równa zeru przy ściankach falowodu:

$$H_z = \sum_{l,m} A_{lm} \cos \frac{\pi l x}{a} \cos \frac{\pi m y}{b} \cos \omega \times \times \left(t - \frac{z}{V_{lm}} \right), \qquad (15.62)$$

gdzie l oraz m są liczbami węzłów w kierunku osi x oraz y, natomiast $V_{l,m}$ jest prędkością fazową składowej fali l, m w kierunku osi z. Liczby całkowite l oraz m mogą przyjmować wartości z przedziału $0 \div \infty$, ale kombinacja 0, 0 jest wykluczona. Równanie propagacji fal

$$\nabla^2 H_z - \frac{1}{c^2} \frac{\partial^2 H_z}{\partial t^2} = 0 \qquad (15.63)$$

daje

$$\omega^2 \left(\frac{1}{c^2} - \frac{1}{V_{l,m}^2} \right) = \left(\frac{\pi l}{a} \right)^2 + \left(\frac{\pi m}{b} \right)^2. \qquad (15.64)$$

Prędkość fazowa $V_{l,m}$ jest większa niż prędkość światła c, podczas gdy pręd-

* Zob. np. C. B. Montgomery, R. H. Dicke, F. M. Purcell, *Principles of Microwave Circuits*, M.I.T. Radiaton Lab. Series, vol. 8, New York 1948, McGraw-Hill, str. 34, 37, 55.

** Płaszczyzną ogniskowania nazywa się tu płaszczyznę, która przecina ognisko i jest prostopadła do ogniskowej soczewki. (*Przyp. tłum.*).

kość grupowa $U_{l,m}$ jest mniejsza niż c. Obie te prędkości osiągają wartość graniczną c dla częstotliwości bardzo dużych. Prędkość fazowa staje się nieskończenie wielka (a $U_{l,m}$ zbliża się do zera) przy częstotliwości granicznej określonej ze wzorów

$$\frac{1}{c^2}\omega_{l,m}^2 = \left(\frac{\pi l}{a}\right)^2 + \left(\frac{\pi m}{b}\right)^2 \tag{15.65}$$

stąd

$$\frac{\omega_{l,m}}{c} = \frac{2\pi}{\lambda_{l,m}},$$

natomiast odpowiednia długość fali dana jest przez

$$\frac{4}{\lambda^2_{l,m}} = \left(\frac{l}{a}\right)^2 + \left(\frac{m}{b}\right)^2.$$

Daną l, m-falę można wytworzyć za pomocą fali płaskiej przechodzącej przez falowód nachylony pod pewnym kątem do osi i odbijającej się od ścianek falowodu. Oznaczmy przez φ_x, φ_y oraz θ kąty utworzone przez padającą falę płaską i odpowiednie osie x, y, u. Otrzymamy

$$\left.\begin{aligned} &\cos\varphi_x = \pm\frac{l\lambda}{2a}, \qquad \cos\varphi_y = \frac{m\lambda}{2b} \\ &\text{i stąd} \\ &V_{l,m} = \frac{c}{\cos\theta} > c, \\ &\frac{\omega^2}{c^2}\sin^2\theta = \left(\frac{\pi l}{a}\right)^2 + \left(\frac{\pi m}{b}\right)^2 \\ &\text{lub} \\ &\frac{4}{\lambda^2}\sin^2\theta = \left(\frac{l}{a}\right)^2 + \left(\frac{m}{b}\right)^2. \end{aligned}\right\} \tag{15.66}$$

Warunki graniczne odpowiadają $\theta = \pi/2$. W przypadku sytuacji pokazanej na rys. 15.8 wzbudzane wyższe fale składowe l, m powstają na skutek nachylenia promieni w wiązce zbieżnej. Poprzez superpozycję odpowiednich składników (wzór (15.62)) można wytworzyć ognisko F. Jeżeli dana jest częstotliwość wiązki

ω, to wzór (15.65) ogranicza wartości wskaźników l, m tak, aby zawsze zachodziło

$$\omega_{l,m} < \omega.$$

Gdy dany jest kąt apertury θ_0, to wzór (15.66) daje ostrzejsze ograniczenie wskaźników l, m. Duża dokładność w wyznaczaniu wymiarów ogniska będzie wymagała wysokich składowych l, m w sumie (15.62), a to w konsekwencji pociąga za sobą konieczność stosowania wysokiej częstotliwości i dużej apertury θ_0.

Zagadnienie nasze jest jednak odmienne. Zadowolimy się ograniczoną dokładnością w wyznaczaniu wymiarów ogniska, ale chcemy za to dokładność tę powiązać z informacją otrzymywaną w obserwacji. Bierzemy skończoną ilość składowych sumy (15.62)

$$0 \leqslant l \leqslant l_M \quad \text{oraz} \quad 0 \leqslant m \leqslant m_M, \tag{15.67}$$

jak również częstotliwości większe od częstotliwości granicznych określonych przez wzór (15.66). Warunki (15.67) określają ognisko optyczne o skończonych wymiarach:

$$\Delta x \geqslant \frac{a}{l_M}, \qquad \Delta y \geqslant \frac{b}{m_M}. \tag{15.68}$$

Zmiennych z oraz t nie możemy rozdzielić, gdyż fale nasze rozchodzą się swobodnie w kierunku osi z, zakładamy więc, że falowód jest bądź nieskończenie długi, bądź tak obciążony na końcu, aby uniknąć odbić*. Możemy stosować wąskie impulsy o czasie trwania τ, tak jak uczyniono to w rozdz. 13. W celu uzyskania skończonej liczby stopni swobody, impulsy musimy powtarzać z okresem t. Daje to nam liczbę stopni swobody dla każdej fali (wzór (13.36)):

$$n \geqslant \frac{t}{2\tau}. \tag{15.69}$$

W ogólności wiązka nasza ma $l_M m_M n$ stopni swobody, jeżeli chodzi o wybór częstotliwości ω. Załóżmy niską częstotliwość przy temperaturze T: $h\nu < kT$. Na każdy stopień swobody będzie przypadać energia kT, więc całkowita energia cieplna wyniesie

$$E_T = l_M m_M n k T. \tag{15.70}$$

* Jest to przypadek dopasowania obciążenia do impedancji charakterystycznej falowodu. (*Przyp. tłum.*).

Jeżeli chcemy obserwować ognisko, które można odróżnić na tle promieniowania cieplnego, to użyć musimy energii E co najmniej równej AE_T, gdzie A jest współczynnikiem zdefiniowanym w rozdz. 14.4 i 14.5. W czasie obserwacji energia promieniowania jest rozpraszana przez cząstkę i absorbowana przez fotokomórkę. Energia ta daje przyrost entropii

$$\Delta S = \frac{E}{T} \geqslant l_M m_M n k A \tag{15.71}$$

lub ze wzorów (15.68) oraz (15.69)

$$\Delta S \geqslant Ak\frac{F}{2}, \quad A > 1 \quad \text{oraz} \quad F = \frac{a}{\Delta x}\frac{b}{\Delta y}\frac{t}{\tau}. \tag{15.72}$$

Ilość informacji uzyskanej ze wzorów (15.1) i (15.3) wynosi

$$\Delta I = k \ln F \tag{15.73}$$

i jest zawsze mniejsza od ΔS. Negentropijna zasada informacji zostaje spełniona i stąd

$$\Delta(S-I) \geqslant k\left(\frac{AF}{2} - \ln F\right) > 0. \tag{15.74}$$

Nierówność ta przypomina wyrażenie (13.55) podane przez Gabora, dotyczące zagadnienia Szilarda. Wyrażenie (15.74) jest zawsze dodatnie. Osiąga ono minimum przy $F = \frac{2}{A}$, a wartość tego minimum wynosi

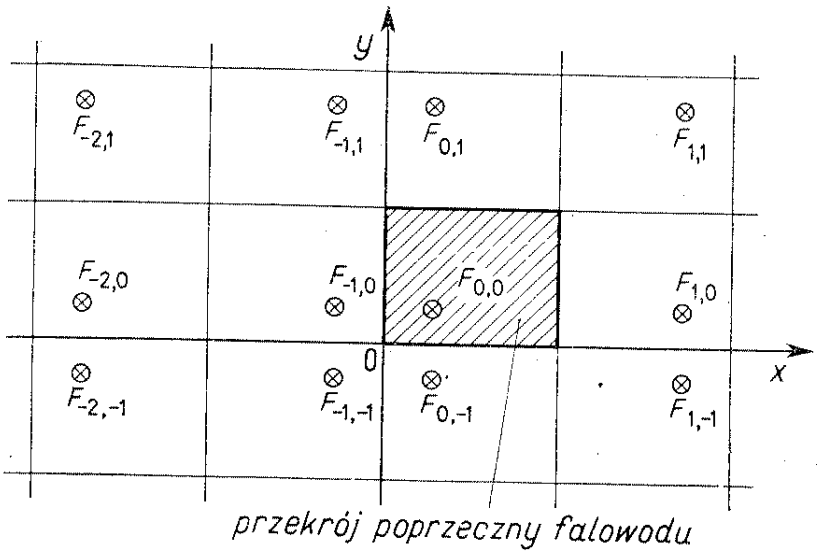
$$k\left[1 - \ln\left(\frac{2}{A}\right)\right] > 0,$$

gdyż A jest większe od jedności. Jeżeli pole obserwacji jest bardzo duże i dokładność duża, wówczas wielkość F staje się również bardzo duża, natomiast cena entropijna obserwacji ΔS bardzo znacznie przewyższa ilość uzyskanej informacji ΔI. Uwaga ta może mieć duże znaczenie w zastosowaniu do wielu zagadnień fizycznych. Obserwacje ogniska bez ograniczenia czasu i pola prowadziłyby do nieskończenie wielkiej ceny entropijnej, ale nie odpowiada to żadnemu zagadnieniu fizycznemu*.

* Zagadnienie to poruszył G. Tonaldo di Francia, J. Opt. Soc. Amer. **45** (1955), str. 497—501.

15.9. OGNISKO W FALOWODZIE

Powróćmy do warunków (15.67) i dokładniej prześledźmy warunki panujące w pobliżu ogniska w falowodzie. Powinniśmy mieć na uwadze fakt, że ścianki rury zachowują się jak doskonałe zwierciadła, odbijające nieskończoną ilość razy przekrój poprzeczny rury w górę i w dół, na lewo i prawo, tak jak ilustruje to rys. 15.9.



Rys. 15.9. Zakreskowany prostokącik przedstawia przekrój poprzeczny falowodu z ogniskiem soczewki $F_{0,0}$. Ponieważ ścianki falowodu doskonale odbijają, obrazy ogniska będą pojawiać się tak, jak pokazano na rysunku

Ognisko $F_{0,0}$ wewnątrz falowodu ma swoje obrazy (rys. 15.9): $F_{-1,0}$, $F_{0,-1}$, $F_{-1,-1}$ w sąsiednich prostokącikach i tak aż do nieskończoności. Jak można przedstawić sytuację taką za pomocą układu fal typu (15.62)? Zauważmy napierw, iż ognisko optyczne $z = 0$ charakteryzuje się tym, że wszystkie składowe w wyrażeniu (15.62) są w fazie. Dla innej wartości z te same składowe nie będą już w fazie, gdyż ich prędkości są różne i wiązka rozbiega się.

Rozpatrując tylko płaszczyznę ogniskowania $z = 0$ mamy

$$H_{z=0} = \sum_{l,m} A_{lm} \cos l\varphi \cos m\psi \cos \omega t, \tag{15.75}$$

gdzie

$$\varphi = \frac{\pi x}{a}, \qquad \psi = \frac{\pi y}{b}.$$

Podwójny szereg Fouriera możemy rozłożyć na dwa szeregi: pierwszy względem x, drugi zaś względem y. Otrzymujemy

$$\left.\begin{aligned} H_{z=0} &= X(x)\,Y(y)\cos\omega t,\\ X(x) &= \sum_{l} X_l \cos l\varphi, \quad 1 \leqslant l \leqslant l_M,\\ Y(y) &= \sum_{m} Y_M \cos m\psi, \quad 1 \leqslant m \leqslant m_M,\\ A_{lm} &= X_l Y_m. \end{aligned}\right\} \tag{15.76}$$

Wyłączamy $l = 0$, $m = 0$ w celu wyeliminowania zabronionego składnika 0,0.

Zagadnienie jest symetryczne względem x i y, dlatego rozpatrzymy tylko współrzędną x. Chcemy otrzymać ognisko $F_{0,0}$ ze wszystkimi jego obrazami. Oznacza to, że musimy mieć ostre maksima $X(x)$ w punktach

$$x = x_0 + 2pa \quad \text{oraz} \quad x = -x_0 + 2pa, \tag{15.77}$$

gdzie p jest dodatnią lub ujemną liczbą całkowitą. W dalszym ciągu skonstruujemy szereg ze skończoną liczbą wyrazów, przedstawiający tego typu sytuację z daną dokładnością. Potrzebne nam jednak będą niektóre zależności matematyczne, które teraz podamy. Zacznijmy od tożsamości Langrange'a i zapiszmy ją w innej formie niż wzór (13.35):

$$\cos\varphi + \cos 2\varphi + \ldots + \cos l_M\varphi = \sum_{1}^{l_M} \cos l\varphi = \frac{1}{2}\left(\frac{\sin\left(l_M + \frac{1}{2}\right)\varphi}{\sin\frac{1}{2}\varphi} - 1\right) =$$

$$= \frac{\sin\frac{l_M}{2}\varphi\,\cos\frac{l_M+1}{2}\varphi}{\sin\frac{\varphi}{2}}. \tag{15.78}$$

Wynik ten otrzymujemy na drodze prostych przekształceń trygonometrycznych opartych o zależność

$$\left(l_M + \frac{1}{2}\right)\varphi = \frac{l_M}{2}\varphi + \frac{l_M+1}{2}\varphi.$$

Bierzemy teraz

$$1 \leqslant l \leqslant l_M, \qquad x_l = \cos l\varphi_0, \qquad \varphi_0 = \frac{\pi x_0}{a}, \tag{15.79}$$

co daje

$$X(x) = \sum_{0}^{l_M} \cos l\varphi \cos l\varphi_0 = \frac{1}{2} \sum_{0}^{l_M} [\cos l(\varphi-\varphi_0) + \cos l(\varphi+\varphi_0)]$$

lub wykorzystując (15.78)

$$X(x) = \frac{\sin \frac{l_M}{2}(\varphi-\varphi_0) \cos \frac{l_M+1}{2}(\varphi-\varphi_0)}{2 \sin \frac{\varphi-\varphi_0}{2}} + \frac{\sin \frac{l_M}{2}(\varphi+\varphi_0) \cos \frac{l_M+1}{2}(\varphi+\varphi_0)}{2 \sin \frac{\varphi+\varphi_0}{2}}. \tag{15.80}$$

Pierwszy człon daje ostre maksima przy

$$\varphi = \varphi_0 + 2p\pi$$

tj. wtedy, gdy mianownik staje się zerem. Człon drugi daje maksima przy

$$\varphi = -\varphi_0 + 2p\pi$$

i daje to ognisko i jego obrazy określone przez wzór (15.77). Cały ten układ ulega degeneracji, gdy $\varphi_0 = 0$, tj. wtedy, gdy ognisko znajduje się w jednym z rogów falowodu.

Ciąg współczynników we wzorze (15.79) daje dokładnie wymagane warunki. Rozważmy samo pierwotne ognisko i jego otoczenie:

$$x = x_0 + \delta x, \quad \varphi = \varphi_0 + \delta\varphi, \quad \delta x = a\frac{\delta\varphi}{\pi}. \tag{15.81}$$

Drugi składnik (wzór (15.80)) jest do pominięcia.

Pierwszy składnik ma maksimum $= l_M/2$ w ognsiku ($\delta\varphi = 0$) i pierwsze zera przy

$$\delta\varphi = \pm\frac{\pi}{l_M+1},$$

tj. wtedy, gdy wyrażenie cosinusoidalne licznika pierwszego składnika jest zerem. Szerokość maksimum można praktycznie zdefiniować jako połowę

odległości między pierwszym zerem z prawej strony i pierwszym z lewej:

$$\left.\begin{aligned} \Delta\varphi &= \frac{2\delta\varphi}{2} = \frac{\pi}{l_M+1}, \\ \Delta x &= \frac{a}{l_M+1} = \frac{a}{l_M}. \end{aligned}\right\} \tag{15.82}$$

W przypadku kiedy l_M jest duże, warunek ten upraszcza się do postaci danej poprzednim wzorem (15.67).

Podobne rozważanie stosować można dla współrzędnej y.

Rozważmy aperturę kwadratową i takie same ograniczenia na x oraz y:

$$a = b = r, \qquad \Delta x = \Delta y, \qquad l_M = m_M = M.$$

Promień ogniska wyniesie

$$\Delta r = \sqrt{(\Delta x)^2+(\Delta y)^2} = r\sqrt{\frac{1}{l_M^2}+\frac{1}{m_M^2}} = \sqrt{2}\,\frac{r}{M}.$$

Ograniczenie kąta θ_0 otrzymamy ze wzoru (15.66):

$$\frac{4\nu^2}{c^2}\sin^2\theta_0 = \left(\frac{l_M}{a}\right)^2 + \left(\frac{m_M}{r}\right)^2 = 2\left(\frac{M}{r}\right)^2, \tag{15.83}$$

tak że

$$\Delta r = \frac{c}{\nu \sin\theta_0} = \frac{\lambda}{\sin\theta_0} \tag{15.84}$$

jest po prostu wzorem na zdolność rozdzielczą, będącą funkcją długości fali i kąta aperturowego θ_0.

15.10. PRZYKŁADY I OMÓWIENIA

Czytelnik może się zdziwić, dlaczego otrzymane zależności (15.67) i (15.69) są tak podobne z wyjątkiem współczynnika $^1/_2$, który występuje we wzorze (15.69). Powodem jest to, że wzdłuż osi x i y wystąpić muszą układy fal stojących. Fale stojące powstają z dwu jednakowych, ale rozchodzących się w przeciwnych kierunkach fal. Fale takie powstają na skutek odbić od ścianek falowodu. W kierunku osi z mamy natomiast tylko jedną falę rozchodzącą się bez odbić. Tym tłumaczy się współczynnik $^1/_2$ we wzorze (15.69) odnoszący się do układu o n stopniach swobody. Wyjaśniliśmy na początku rozdz. 15.9,

że wszystkie fale składowe w płaszczyźnie ogniskowania $z = 0$ są zgodne w fazie i fazy ich zaczynają się szybko różnić dla innych wartości z, co spowodowane jest różnymi prędkościami V_{lm}. Moglibyśmy równie dobrze rozważyć ognisko jako wynik zbieżności wiązki światła o maksymalnym kącie θ_0 określonym przez (15.83). Wiązka światła przechodzi i poza ognisko, a wówczas θ_0 jest kątem rozchodzenia się wiązki.

Zamiast ostrego ogniska możemy potrzebować długiego promienia *szpilkowego*. Można go otrzymać stosując wielkie częstotliwości oraz małe kąty θ. Wzory (15.66) lub (15.83) wskazują na możliwość dobrania tych parametrów w taki sposób, że liczby całkowite l oraz m pozostają duże, a wiązka świetlna posiada małą powierzchnię przekroju.

Mamy również możliwość wytworzenia, co najmniej w przybliżeniu, wiązki płaskiej o zadanej grubości. Można to zrobić biorąc soczewkę cylindryczną, małe kąty θ i długie szczeliny ograniczające wiązkę w kierunku osi x. W kierunku y wiązki nie ogranicza się. Sytuację tę zilustrowano na rys. 15.10. Rozłożenie fali (wzór (15.76)) daje przy płaszczyźnie ogniskowania

$$H_{z=0} = X(x) \cos \omega t,$$

co jest niezależne od y.

Wypadkową wiązkę opisuje równanie

$$H = \sum_{l=1}^{l_M} X_l \cos l\varphi \cos \omega \left(t - \frac{z}{V_1}\right) \tag{15.85}$$

przy $m = 0$. Wielka częstotliwość ω, dużo większa od częstotliwości granicznej

$$\omega \gg \omega_{l_M} = \pi c \frac{l_M}{a}, \tag{15.86}$$

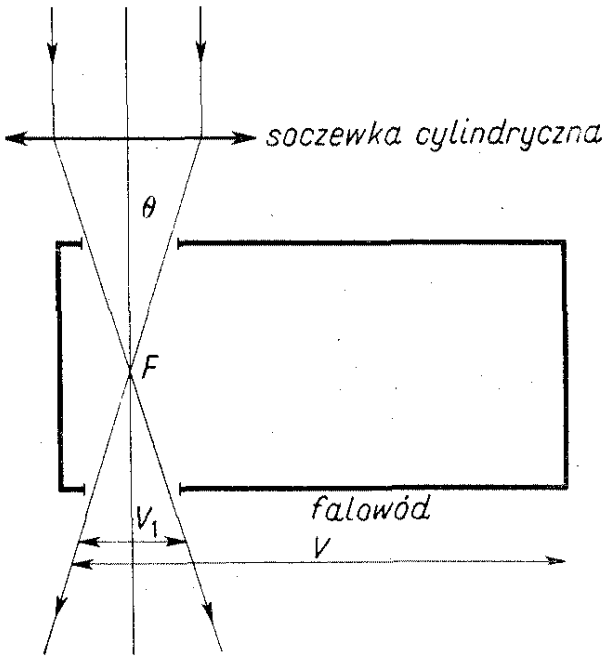
daje wiązkę o bardzo małej rozbieżności. Warunki te odpowiadają zagadnieniu maszyny Szilarda opisanej w rozdz. 13.6.

Rysunek 15.10 obrócony o 90° odpowiada sytuacji z rys. 13.5. W tym przypadku mamy $\Delta y = b$ i wyrażenie na F z równania (15.72) sprowadza się do postaci

$$F = \frac{a}{\Delta x} \frac{t}{\tau} = \left(\frac{V}{V_1}\right)^2. \tag{15.87}$$

Jest to dokładnie wartość F ze wzoru (13.55), a połączenie wzorów (15.74) i (15.87) daje wynik identyczny z poprzednim wynikiem z rozdz. 13, który otrzymano na innej drodze.

Nasze wyrażenia (15.72) i (15.74) zawierają współczynnik liczbowy, którego wartości omówiono w rozdz. 14.5. Dla małych częstotliwości współczynnik ten przybiera wartość A_n, która może być rzędu kilku jednostek, ale niektóre



Rys. 15.10. Wiązka promieni świetlnych o małej rozbieżności wytwarzana przez soczewkę cylindryczną. Wiązka taka nadaje się do zastosowania w maszynie Szilarda

warunki eksperymentalne mogą wymagać wielkich częstotliwości; prowadzi to do dużo większych wartości A, tak że

$$A = \begin{cases} A_n, & \text{gdy } h\nu \ll kT, \\ \dfrac{h\nu}{kT}, & \text{gdy } h\nu \gg kT. \end{cases} \tag{15.88}$$

Górna równość odnosi się do obserwacji klasycznych, natomiast dolna odpowiada obserwacji, która wprowadza zakłócenie do obserwowanego układu i w której uwzględnić trzeba efekty kwantowe.

15.11. PODSUMOWANIE

We wszystkich przykładach omówionych w tym i poprzednim rozdziale doszliśmy do wniosku, że informacja otrzymywana z doświadczenia zależy logarytmicznie od dokładności:

$$\Delta I = k \ln \mathscr{A} \tag{15.89}$$

i że związany z nią przyrost entropii spełnia nierówność

$$\Delta S \geqslant \Delta I. \tag{15.90}$$

Przykłady te posłużyły do sprawdzenia uogólnionej zasady Carnota, którą nazwaliśmy *negentropijną zasadą informacji*:

$$\Delta(S-I) \geqslant 0. \tag{15.91}$$

Zbadaliśmy sprawność obserwacji, która w myśl negentropijnej zasady spełnia nierówność

$$\varepsilon = \frac{\Delta I}{\Delta S} \leqslant 1. \tag{15.92}$$

W przypadku prostego pomiaru odległości (rozdz. 15.2 i 15.3) sprawność jest bliska jedności, a negentropijna cena obserwacji jest mała. We wszystkich innych rozważanych przypadkach sprawność jest dużo mniejsza od jedności, a przyrost entropii zależy co najmniej liniowo od dokładności:

$$\Delta S = a\mathcal{A}+b. \tag{15.93}$$

Stała b jest zwykle mała. Współczynnik a w niektórych przypadkach jest stały lub niekiedy jest logarytmiczną funkcją $\mathcal{A}$. Metoda interferometryczna pomiaru odległości i pomiar położenia przy pomocy mikroskopu prowadzą do przyrostu entropii danego przez wzór (15.93). Oczywiście, w tych przypadkach ze wzrostem dokładności sprawność gwałtownie maleje.

Ogólnie mówiąc, zobaczyliśmy, że sprawność jest większa przy posługiwaniu się małymi częstotliwościami. Inaczej jest jedynie w przypadkach, w których występują warunki kwantowe, a wymagana dokładność jest bardzo duża. W takich okolicznościach użycie wielkich częstotliwości jest bardziej opłacalne, chociaż sprawność jest bardzo mała. W tym przypadku można zwiększyć sprawność przez wprowadzenie odpowiednich przyrządów zmniejszających rozpraszanie energii. W wyniku takiego postępowania zagadnienia związane z wielkimi częstotliwościami dają się sprowadzić do zagadnień związanych z małymi częstotliwościami i wobec tego we wszystkich tych przypadkach można wykorzystać wyniki dotyczące małych częstotliwości.

Można zauważyć, że podana poprzednio prosta metoda pomiaru odległości, która daje bardzo dużą sprawność, nie należy do tych metod pomiarowych, które łatwo wykorzystać można w laboratorium do pomiaru małych odległości, gdyż trudno jest skonstruować odpowiedni aparat. Bardzo małe odległości mierzy się przeważnie metodą interferometryczną, która, jak pokazaliśmy, odznacza się małą sprawnością.

16

TEORIA INFORMACJI, ZASADA NIEOZNACZONOŚCI I OGRANICZENIA FIZYCZNE OBSERWACJI

16.1. UWAGI WSTĘPNE

Omówiliśmy różnorodne doświadczenia i doszliśmy do ogólnego wniosku, jakim jest *negentropijna zasada informacji*, która mówi, że każda informacja otrzymana z obserwacji fizycznej musi być opłacona przyrostem entropii w laboratorium. Średnio przyrost tej entropii jest większy od otrzymanej informacji, jeśli obie wielkości mierzone są w tym samym układzie jednostek. Warunki te narzucają nowe ograniczenia na możliwości obserwacji; chcemy porównać je z dobrze znanymi relacjami nieoznaczoności.

Pół wieku toczyła się dyskusja na temat związku między fizyką klasyczną i kwantową. Oryginalny punkt widzenia podaje N. Bohr w swojej *zasadzie dopełnienia*, a jego interpretacja relacji nieoznaczoności wydaje się właściwym, logicznym podejściem do zagadnienia. Niedawne dyskusje* w jasny sposób podkreśliły znaczenie poruszonych zagadnień.

Z historycznego punktu widzenia fizyka startuje od poziomu klasycznego, od eksperymentów robionych na skalę człowieka i od fundamentalnych definicji (długość, czas, masa, podstawowe prawa) opartych na tych eksperymentach. Z tego trwałego gruntu fizyk przechodzi do badania zagadnień

---

* W. Elsasser, Phys. Rev. **52** (1937), str. 987; Phil. Sci. **18** (1951), str. 300; w pracy: *Louis de Broglie, Physicien et Penseur*, Paris 1953, Albin Michel, str. 87.

N. Bohr, L. Rosenfeld, Phys. Rev. **78** (1950).

L. Rosenfeld, w pracy: *Louis de Broglie, Physicien et Penseur*, Paris 1953, Albin Michel, str. 13.

D. Bohm, *Quantum Theory*, New York 1951, Prentice-Hall.

atomowych i zaczyna odkrywać efekty kwantowe ze wszystkimi ograniczeniami tkwiącymi w zasadzie nieoznaczoności oraz mechanikę falową.

Konsekwencja wymaga, by nowe te prawa w granicy, przy bardzo dużej ilości kwantów, sprowadzały się do praw klasycznych.

Bohr wiele razy podkreśla fakt, że aparat pomiarowy jest zawsze opisany pojęciami klasycznymi, lecz wprowadził on również *zasadę odpowiedniości*, stosownie do której prawa kwantowe muszą prowadzić do praw klasycznych w przypadkach granicznych, kiedy stałą Plancka h można traktować jako wielkość zaniedbywalnie małą.

Bohm stwierdza, że „teoria kwantowa przyjmuje klasyczne pojęcia przy opisie tego poziomu". Jak zdefiniować można fizykę klasyczną? Elsasser uważa, że fizykę klasyczną charakteryzuje możliwość przeprowadzania eksperymentów niezakłócających, w których można dokonać wyraźnego podziału na obserwatora i układ obserwowany. Jako typowy przykład posłużyć może optyczna obserwacja układu mechanicznego, np. spadającego ciała. Doświadczenie nie zakłóca dostrzegalnie układu i może być powtarzane wiele razy z tymi samymi wynikami. Posiada ono fundamentalną własność *reproduktywności*. Nie łatwo jest ustalić precyzyjnie warunki, jakie muszą być spełnione, by można było dokonywać niezakłócających obserwacji. Zaakcentujemy jednak kilka punktów. Układ obserwowany musi zawierać dużo cząstek i dużą liczbę kwantów. Dokładność doświadczenia nie może być zbyt duża. Oczywistym warunkiem jest też to, że doświadczenie powinno przebiegać w warunkach dalekich od przypadku granicznego, w którym wchodzi w grę zasada nieoznaczoności:

$$\Delta p\, \Delta q \ggg h \quad \text{oraz} \quad \Delta E\, \Delta t \ggg h. \tag{16.1}$$

W celu jednoczesnego określania p, q, E oraz t, dopuszczalne błędy Δp, Δq, ΔE oraz Δt nie powinny być zbyt małe. Możemy teraz wprowadzić inny warunek, wynikający z rozważań przytoczonych w tej książce. Cena negentropijna ΔS obserwacji powinna być mała, do zaniedbania, gdy porównywana jest z całkowitą entropią S_0 obserwowanego układu:

$$\left.\begin{aligned} &\Delta S \ll S_0 \\ \text{i stąd} \quad &\\ &\Delta I \leqslant \Delta S \ll S_0, \end{aligned}\right\} \tag{16.2}$$

gdyż informacja ΔI jest zawsze mniejsza od ceny entropijnej ΔS. Są to warunki

konieczne, ale mogą nie być wystarczające. Zasadniczy schemat procedury pomiarowej jest następujący.

Układ obserwowany i przyrząd pomiarowy nie są ze sobą powiązane. Podczas obserwacji uwydatnia się powiązanie, i to z pewną wzajemną reakcją. Po obserwacji układ i przyrząd pomiarowy ponownie nie są ze sobą powiązane. W czasie obserwacji występują zmiany p, q, E oraz S układu; są to nieznane zmiany, które występują w czasie trwania doświadczenia równym Δt. Obserwację można rozpatrywć jako niezakłócającą, jeżeli te nieznane zmiany są bardzo małe w porównaniu z dokładnością doświadczenia. Warunki (16.1) i (16.2) są konieczne. Rozważmy np. przypadek (16.2). Doświadczenie wpływa na wzrost o ΔS całkowitej entropii obserwowanego układu i przyrządu pomiarowego. Gdy układy nie są powiązane ze sobą po doświadczeniu, część entropii ΔS związana będzie z układem obserwowanym, reszta zaś z urządzeniem pomiarowym, tak że tych dwóch składowych rozróżnić nie można. Dlatego musimy zażądać, aby całkowity przyrost ΔS był zawsze bardzo mały w porównaniu z entropią obserwowanego układu S_0.

Elsasser rozpatruje, jako przykład, pomiar cząsteczek w gazie. W zasadzie jest możliwy bardzo dokładny pomiar położenia wszystkich cząsteczek, lecz taka obserwacja wymagałaby strumienia światła o bardzo wysokiej częstotliwości. Wzajemne oddziaływanie światła i cząsteczek udzielałoby tym cząsteczkom dużych momentów pędu, a to powodowałoby daleko posunięte zmiany całego układu. Jasne jest, że nie jest to obserwacja niezakłócająca i że nie spełnia ona warunków (16.1) i (16.2). Innymi słowy, pomiary objętości, ciśnienia i gęstości mogą być dokonane przy zachowaniu warunków (16.1) i (16.2), co również prowadzi do niezbyt dużej dokładności.

16.2. OBSERWACJA JEST PROCESEM NIEODWRACALNYM

Należy podkreślić, że przyrost entropii ΔS wymagany w czasie obserwacji nie pozostaje w żadnym związku ze zmianami entropii pojawiającymi się w samym układzie tak długo, jak długo nie dokonuje się pomiaru.

Rozpatrzmy np. układ, w którym zachodzi przemiana adiabatyczna*.

* Przemianę adiabatyczną opisuje się jako ciąg stanów równowagi. Jeżeli układ znajduje się w cylindrze o długości L, zaopatrzonym w przesuwający się tłoczek, to ruch δx tłoczka musi być bardzo mały w czasie δt, dostatecznie długim, aby zaburzenie mogło rozchodzić się

Entropia wewnątrz układu pozostaje stała, ale przemiana adiabatyczna przebiega jako niezauważalna i nieobserwowalna, gdy układ nie jest sprzężony z przyrządem pozwalającym dokonać obserwacji. Przyjmijmy, że przemiana adiabatyczna powoduje zmiany wielkości fizycznych (p, V, T, itp.), lecz tak długo, dopóki zmiennych tych nie mierzymy, nie mamy sposobu dowiedzieć się, czy coś się zmienia. Aby zaobserwować p, V, T, musimy układ połączyć z fizycznym urządzeniem pomiarowym i przez to wprowadzamy małe zakłócenia. Sama obserwacja jest procesem nieodwracalnym, wnoszącym przyrost

---

tam i z powrotem wiele razy na całej długości L. Wymaga to, aby

$$\frac{\delta x}{\delta t} = v \lll c,$$

gdzie c jest prędkością rozchodzenia się zakłócenia. W przypadku promieniowania c jest prędkością światła; dla gazów c oznacza prędkość termiczną cząsteczek, dla ciał stałych lub ciekłych c oznacza prędkość rozchodzenia się głosu.

Przemianami adiabatycznymi układów skwantowanych zajmowało się wielu autorów. Bohm (*op. cit.*, rozdz. 20) stawia warunek

$$\frac{h}{(\Delta E)^2}\frac{\delta E}{\delta t} \ll 1,$$

gdzie E jest energią, zaś ΔE jest różnicą energii między dwoma stanami. Rozważmy falę stojącą na całej długości L z N węzłami, tak że

$$L = N\frac{\lambda}{2} = N\frac{c}{2\nu};$$

stąd

$$E = nh\nu = hcn\frac{N}{2L},$$

$$\Delta E = h\nu = hc\frac{N}{2L}$$

(jeśli istnieje n kwantów oscylacji $h\nu$ w naszej fali stojącej). Wówczas

$$\frac{\delta E}{\delta t} = -hcnN\frac{v}{2L^2}, \qquad v = \frac{\delta L}{\delta t}$$

i warunek Bohma dla przemiany adiabatycznej przyjmuje postać

$$\frac{2\nu n}{Nc} \ll 1,$$

a więc jest podobny do podanego wyżej warunku $v \lll c$.

entropii ΔS w laboratorium. Musimy zawsze odczytać położenie wskazówki na skali (pomiary odległości) i notować przedziały czasu. Tego typu podstawowe pomiary zostały omówione w rozdz. 15, a każdą obserwację fizyczną można do nich sprowadzić.

W przykładzie tym zobaczymy różnicę między zmianami entropii występującymi w układzie odizolowanym i stąd nieobserwowalnym a przyrostem entropii ΔS wymaganym do obserwacji i uwarunkowanym związkiem między obserwowanym układem i przyrządami pomiarowymi.

Zasadniczo obserwacja jest procesem nieodwracalnym. Z czysto termodynamicznego punktu widzenia dowiedliśmy, że obserwacja nie może być dokonana bez związanego z nią przyrostu entropii

$$\Delta S > 0 \tag{16.3}$$

w samym układzie fizycznym lub w urządzeniu używanym do doświadczenia i w czasie obserwacji sprzężonym z układem.

Jeżeli przyjmujemy bardziej ogólny punkt widzenia, opisany w poprzednich rozdziałach, to włączamy również informację ΔI, natomiast warunek (16.3) zapiszemy w postaci

$$\left.\begin{aligned} \Delta(S-I) &\geqslant 0, \\ \Delta S \geqslant \Delta I &> 0. \end{aligned}\right\} \tag{16.4}$$

W obu przypadkach w czasie obserwacji nieunikniony jest przyrost entropii i, podobnie jak przyrost entropii pociąga za sobą nieodwracalność, tak nieodwracalność występuje jako cecha charakterystyczna obserwacji.

Podobne spostrzeżenia poczyniło uprzednio wielu autorów, a szczególnie J. von Neumann, który omawia niektóre paradoksy mechaniki kwantowej. Nie należy mówić, że układ znajduje się w pewnym stanie, w sytuacji gdy nie mierzymy żadnych wielkości charakteryzujących ten stan. Proces pomiaru jest nieodwracalny i determinuje — co najmniej częściowo — przyszłe zachowanie się układu.

16.3. ZASADNICZE OGRANICZENIE DOKŁADNOŚCI POMIARÓW FIZYCZNYCH

Napotkaliśmy wiele trudności przy pomiarze odległości z bardzo dużą dokładnością, szczególnie wtedy, gdy wchodziły w grę bardzo małe odległości. Długość charakterystyczna Δx_T, odpowiadająca danej temperaturze T (por.

wzór (15.12)), zdefiniowana wzorem

$$\Delta x_T = \frac{hc}{kT} \approx 1{,}44\,\frac{1}{T} \quad \text{cm} \tag{16.5}$$

(gdzie T — temperatura w stopniach Kelvina) — może być traktowana jako granica rozdziału odległości, które nazywamy małymi i dużymi. Wzrastająca trudność w pomiarach coraz to mniejszych odległości wskazuje na to, że podstawowe założenia przestrzeni Euklidesa i czasu muszą być rozpatrywane jako idealizacja, która nie może i nie powinna być stosowana przy bardzo małych odległościach. Matematyczna definicja nieskończenie małych odległości odpowiada niemożliwym warunkom fizycznym. Wrócimy do tego zagadnienia w rozdz. 16.4.

Łatwo dopatrzyć się można w omówionych przykładach powiązania pomiędzy teorią kwantową i teorią informacji. Ograniczenia kwantowe trzeba uwzględniać w rozważaniach, ponieważ żyjemy w skwantowanym świecie. Kiedy rozpatrujemy eksperymenty na poziomie klasycznym, zauważamy, że stała Plancka h nie występuje w wynikach końcowych, które zawierają tylko stałą Bolzmanna k. Uznajemy przeto, że negentropijna zasada informacji jest rzeczywiście zasadą nową i nie można jej sprowadzić do relacji kwantowych i zasad nieoznaczoności. Stałej Plancka nie można wyeliminować, gdy doświadczenie przechodzi na poziom kwantowy, a obie teorie powiązane są na tym poziomie w sposób bardziej skomplikowany. Można to pokazać na podstawie zagadnień poruszonych w rozdz. 15.3 i 15.5. Duże odległości odpowiadają zagadnieniu klasycznemu (punkt 15.5.1). Pomiar bardzo małych odległości sięga już poziomu kwantowego. Przy pomiarze ekstremalnie małych odległości i przy bardzo niskiej temperaturze w grę wchodzą ograniczenia zarówno ze strony efektów kwantowych, jak i ze strony teorii informacji.

We wszystkich naszych ogólnych rozważaniach dążyliśmy do sprowadzenia zagadnień do podstawowych elementów i systematycznie pomijaliśmy degradację energii we wszelkiego rodzaju wzmacniaczach. Wzmacniacze są zawsze niezbędne w przyrządach wykorzystywanych przy doświadczeniach. Spełniają one następującą rolę: przekształcają efekt kwantowy na drodze zwiększania jego amplitudy w taki efekt, który może być zarejestowany przez aparat pracujący na poziomie klasycznym. Rola wzmacniacza wyraźnie zostaje wyjaśniona przez Elsassera. Kilka przykładów zilustruje ogólną sytuację: klisza fotograficzna może reagować na pojedynczy kwant, a wywoływanie jest

wzmacnianiem, co daje widzialny efekt w postaci czarnej plamki (efekt klasyczny). Fotokomórka może rejestrować pojedynczy kwant, jeżeli podłączona jest do wzmacniacza elektronicznego. Licznik Geigera może wykryć pojedynczą, naładowaną cząstkę, a wzmacnianie wynika z wyładowania jonowego. Komora Wilsona jest analogicznym przykładem takiego wzmacniacza. Tak długo, jak wzmacniacz jest wzmacniaczem bez sprzężenia zwrotnego, może być pominięty, gdyż nie wpływa on bezpośrednio na obserwację, chociaż zrozumiałe jest, że będzie wnosił związany z obserwacją przyrost entropii do układu. Wzmacniacz ze sprzężeniem zwrotnym odgrywa bardzo ważną rolę, jak to podkreślił Wiener w swojej teorii cybernetyki.

Jak wspomniano wyżej, warunki klasyczne występują przy pomiarze omawianym w punktach 15.3.1 i 15.5.1, gdy ΔL jest większe od Δx_T. Wyjątkowe sytuacje pojawiają się w przypadku omawianym w punktach 15.3.2 i 15.5.2, przy

$$\Delta L < \Delta x_T, \tag{16.6}$$

co odpowiada dużej dokładności, bardzo niskiej temperaturze i bardzo małym odległościom. W takich warunkach musimy odróżnić wymaganą do obserwacji energię ΔE od tej jej części, która przechodząc w ciepło musi ulec degradacji i wywołać przyrost entropii. W wielu praktycznych przypadkach całkowita energia ΔE ulega rozpraszaniu w postaci ciepła, ale jak wykazaliśmy w rozdz. 14 i 15, większej części tego rozproszenia można uniknąć, a przyrost entropii ΔS można (co najmniej teoretycznie) zredukować do wartości otrzymywanej w obserwacjach klasycznych i uczynić go nieznacznie większym od otrzymanej informacji ΔI. Już poprzednio przeanalizowaliśmy zagadnienie pomiaru odległości i doszliśmy do wniosku, że najbardziej ekonomiczną metodą pomiaru była ta, którą omówiono w rozdz. 15.2 i 15.3. Sprawność dana była wzorem (15.17):

$$\varepsilon = \frac{\Delta I}{\Delta S} = 1 + \frac{\ln(\ln 2)}{\ln n}. \tag{16.7}$$

Dla dokładności rzędu 100

$$\frac{L}{\Delta L} = \mathscr{A} = n = 100. \tag{16.8}$$

Wyrażenie (16.7) daje

$$\varepsilon = 1 + \frac{\ln(\ln 2)}{\ln 100} = 1 - 0{,}88 = 0{,}92. \tag{16.9}$$

Pomiary interferometryczne omówione w rozdz. 15.5 i 15.6 oraz obserwacje pod mikroskopem omówione w rozdz. 15.8 mają sprawność dużo mniejszą.

Warunki klasyczne (punkt 15.5.1) odpowiadają przypadkowi, kiedy wymagana do obserwacji energia ulega całkowitej zamianie w ciepło i niezbędna energia jest dana bezpośrednio relacją

$$\Delta E = T\,\Delta S. \tag{16.10}$$

Zachodzi to wówczas, gdy błąd L w pomiarze odległości nie jest mały i pozostaje większy od Δx_T.

W przypadku omawianym w punkcie 15.5.2 żądamy jednak zmniejszenia ΔL i osiągnięcia warunku (16.6). Sytuacja ulega tu całkowitej zmianie. Teoretycznie minimum przyrostu entropii ΔS można utrzymać na poziomie dostatecznie małym, lecz energia wymagana w doświadczeniu coraz bardziej wzrasta:

$$\Delta E > T\,\Delta S. \tag{16.11}$$

Punkt ten godny jest uwagi i omówimy go obecnie.

16.4. GRANICE STOSOWALNOŚCI GEOMETRII EUKLIDESOWEJ

Podsumujmy główne trudności związane z pomiarem bardzo małych odległości i zbadajmy zasadnicze ograniczenia, jakie trudności te nakładają na definicje fizyczne.

Aby pomierzyć odległość L z błędem ΔL, musimy posłużyć się promieniowaniem o długości fali

$$\lambda \leqslant 2\Delta L. \tag{16.12}$$

Warunek ten rozważaliśmy szczegółowo w rozdz. 15 i porównywaliśmy różne metody pomiarowe; doszliśmy do wniosku, że w niektórych przypadkach można stosować długie fale, ale ich cena energetyczna i entropijna jest dużo wyższa niż dla fal krótkich. Fala krótka (wzór (16.12)) odpowiada dużemu pędowi

$$p = \frac{h}{\lambda} \geqslant \frac{h}{2\Delta L} \tag{16.13}$$

i promieniowanie to będzie oddziaływać na układ obserwowany oraz wprowadzać będzie do układu zmianę pędu Δp, który może zmieniać się od $-p$

do $+p$ i stąd

$$\Delta p = 2p \geqslant \frac{h}{\Delta L}, \tag{16.14}$$

co jest zwykłą relacją nieoznaczoności.

Nowym punktem, na który chcemy zwrócić uwagę, jest to, że duży pęd p związany jest ząwsze z dużą energią i zgodnie z tym pomiar odległości z małym błędem ΔL wymaga zawsze dużo energii.

Załóżmy najpierw, że posługujemy się promieniowaniem elektromagnetycznym. Przypadek innego promieniowania rozważymy w rozdz. 16.5. Jeżeli posługujemy się światłem, to długości fali λ odpowiada kwant $h\nu$:

$$h\nu = \frac{hc}{\lambda} \geqslant \frac{hc}{2\Delta L}. \tag{16.15}$$

Najmniejsza ilość energii ΔE, przy pomocy której możemy dokonać doświadczenia, wynosi $h\nu$ i wobec tego:

$$\Delta E = h\nu, \quad \Delta E\, \Delta L \geqslant \frac{hc}{2}. \tag{16.16}$$

Nie ma wyraźnej granicy dolnej dla małych odległości lub długości, które mogą być jeszcze mierzone, lecz ilość energii wymaganej do przeprowadzenia pomiaru wzrasta jak $(\Delta L)^{-1}$ i szybko staje się kłopotliwą przeszkodą w pomiarach. Doświadczenia dokonywane w warunkach omawianych w rozdz. 15.5.2 będą z pewnością zakłócać bardzo obserwowany układ. Trzeba stosować bowiem wielkie częstotliwości, a to powodować będzie duże impulsy odbite, wytwarzać pary* itp.

Uwagi te jasno wskazują na to, że faktyczne koszty wyposażenia i maszyn niezbędnych do eksperymentów nuklearnych związane są ze wzrostem wydatku energii. Ostatecznie koszt pracy dużego cyklotronu lub synchrocyklotronu jest proporcjonalny do dużej energii potrzebnej do każdej obserwacji, a rozpraszanej w napromieniowywanej próbce**.

Matematycy mogą definiować nieskończenie małe odległości, ale jest to abstrakcja odpowiadająca sytuacji, która nie może wystąpić w rzeczywistości. Rozważmy np. laboratorium dysponujące określoną ilością energii. Jeżeli cała ta energia będzie wydatkowana na pomiar pewnej odległości, to odległość

* Chodzi o pary pozyton–elektron. (*Przyp. tłum.*).

** L. Brillouin, J. Appl. Phys. **24** (1953), str. 1152; **25** (1954), str. 887.

ta będzie najmniejszą odległością, jaką można obserwować w tym laboratorium.

Często nasuwało się przypuszczenie, że wiele trudności teorii kwantowej można by wyeliminować na drodze wyboru jakiejś minimalnej długości. Minimalną, ustaloną długość z trudem można by oszacować w oparciu o przytoczone uwagi. Wahamy się, by rozrzerzyć przykład laboratorium z określoną ilością dostępnej energii do rozmiarów wszechświata, gdyż są trudności ścisłego sprecyzowania jego rozmiarów i energii w nim zawartej. Tym niemniej możemy powiedzieć, że pomiary coraz to mniejszych odległości potęgują narastające trudności, a rosnący koszt obserwacji takich odległości może być nowym elementem w teorii. Równość (16.16) jest bardzo podobna do (15.3):

$$\Delta E\,\Delta L \approx \frac{nhc}{2} \tag{16.17}$$

i pokrywa się całkowicie z relacją nieoznaczoności (16.14). Warunki (16.14) mają jednak całkowicie odmienne znaczenie, niż wspomniane warunki związane z zasadą nieoznaczoności. W zasadzie nieoznaczoności ΔE odpowiada błędowi pomiaru energii E. W równaniu (16.16) ΔE jest ilością energii wymaganej w doświadczeniu i co najmniej częściowo ulega zamianie na ciepło podczas pomiaru L z błędem ΔL. Relacje te są więc tylko formalnie podobne: ich znaczenie fizyczne jest całkiem różne.

16.5. MOŻLIWOŚĆ ZASTĄPIENIA FOTONÓW CZĄSTKAMI CIĘŻKIMI

Rozpatrzyliśmy różnorodne urządzenia eksperymentalne przy założeniu, że do oświetlania użyto wiązki świetlnej. Zamiast fotonów o masie zerowej można rozważyć możliwość użycia innego typu fal, takich jak fale związane z cząstkami materialnymi o masie m. Zacznijmy znów od wzorów (16.12) i (16.13), które odnoszą się do dowolnego rodzaju promieniowania, i rozpatrzmy możliwość wykorzystania cząstki o masie spoczynkowej m_0. Musimy obliczyć energię kinetyczną tej cząstki. Cząstka posiada pęd p i energię kinetyczną E_k dane równaniami:

$$p = \frac{h}{\lambda} = m_0 \frac{v}{\sqrt{1-\beta^2}}, \qquad \beta = \frac{v}{c}; \tag{16.18}$$

$$E_k = m_0c^2\left[\frac{1}{\sqrt{1-\beta^2}} - 1\right] = c\sqrt{m_0^2c^2p^2} - m_0c^2. \tag{16.19}$$

W skrajnych przypadkach można stosować przybliżenia: dla fotonów lub cząstek o małej masie spoczynkowej mamy

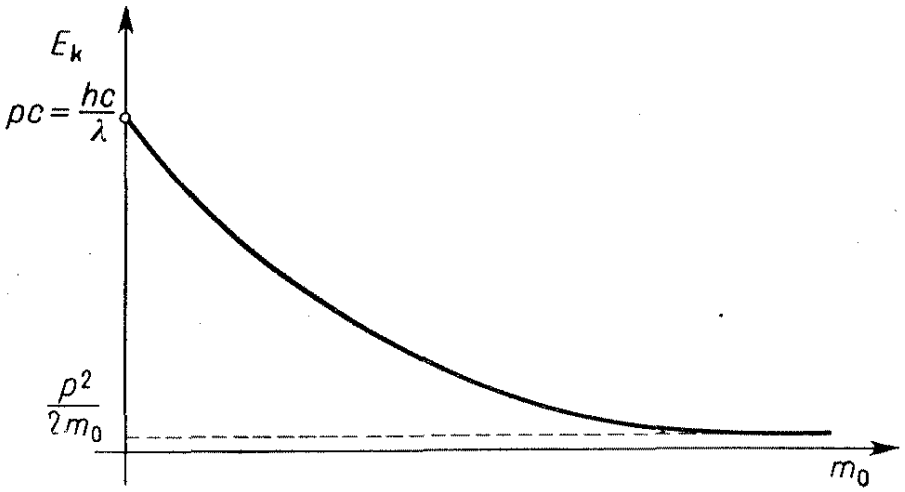
$$m_0 c \ll p = \frac{h}{\lambda}, \qquad E_k \approx pc - m_0 c^2, \tag{16.20}$$

natomiast dla czastek ciężkich mamy

$$m_0 c \gg p = \frac{h}{\lambda}, \qquad E_k \approx \frac{p^2}{2m_0}. \tag{16.21}$$

Dla danego p oraz zmieniającej się masy spoczynkowej m_0 krzywą przedstawiającą E_k jako funkcję m_0 pokazano na rys. 16.1.

Na pierwszy rzut oka stosowanie ciężkich cząstek interesująco wygląda,



Rys. 16.1. Wykres energii kinetycznej w funkcji m_0 przy ustalonym pędzie p. Krzywa ta zbliża się asymptotycznie do wartości $p^2/2m_0$ dla $m_0 c \gg p$

gdyż istnieje tu możliwość zmniejszenia ilości energii E_k wymaganej do doświadczenia. Jest to jednak niemożliwe przy pomiarach bardzo małych odległości d. Przypuśćmy np., że

$$\Delta L = d = 10^{-13}, \qquad p \approx 3 \cdot 10^{-14}$$

zgodnie z równaniem (16.14). Musimy porównać p z $m_0 c$, aby sprawdzić, które równanie przybliżone zastosować: (16.20) czy (16.21). Mamy

elektron: $m_0 \approx 10^{-27}$, $m_0 c \approx 3 \cdot 10^{-17}$ — przypadek (16.20);

proton: $m_0 \approx 2 \cdot 10^{-24}$; $m_0 c \approx 6 \cdot 10^{-14}$ — przypadek pośredni,

ciężki nukleon: $m_0 \approx \frac{5}{3} \cdot 10^{-22}$; $m_0 c \approx 5 \cdot 10^{-12}$ — przypadek (16.21).

Zmniejszenie wymaganej energii można by uzyskać stosując bardzo ciężkie jądra, lecz jądra te mają duże przekroje i byłyby nieodpowiednie przy mierzeniu odległości mniejszych od wymiarów ich przekroju.

W ogólności fotony, mezony, elektrony i nukleony są jedynymi cząstkami, które praktycznie stosować można w wiązce oświetlającej, i wszystkie prowadzą do podobnych wyników, jeżeli odległość d jest bardzo mała. Trudności omówione w rozdz. 16.3 i 16.4 są rzeczywiście trudnościami podstawowymi i nie można ich wyeliminować.

16.6. RELACJE NIEOZNACZONOŚCI W DOŚWIADCZENIU Z MIKROSKOPEM

Metody z teorii informacji wykorzystano w rozdz. 15.8 i 15.9 do omówienia obserwacji pod mikroskopem. Jest to typowe zagadnienie często wiązane z relacjami nieoznaczoności. Przyjmiemy tu, iż jeden foton wystarcza do uzyskania informacji o położeniu cząstki. Jest to oczywiście założenie upraszczające. Pojedynczy foton obserwowany w punkcie I z rys. 16.2 może pochodzić skądkolwiek i niekoniecznie musi wskazywać cząstkę w punkcie M, która rozprasza promieniowanie. Dyskusja przytoczona w rozdz. 15 wykazuje jasno, że dokładne określenie obrazu świetlnego w punkcie I wymaga dużej liczby fotonów. Dla jednego, pojedynczego fotonu spełniona jest relacja nieoznaczoności

$$\Delta x \, \Delta p_x \approx h, \tag{16.22}$$

natomiast dla wielu fotonów nieoznaczoność staje się większa:

$$\Delta x \, \Delta p_x \approx Ah, \tag{16.23}$$

przy czym współczynnik liczbowy A może stać się bardzo duży. Sytuacja taka jest typowa dla przyrządu doświadczalnego, w którym położenie x mierzone jest przy pomocy systemu optycznego bez stosowania materialnych ekranów, przesłon lub migawek. Rozważmy szczegółowo to zagadnienie.

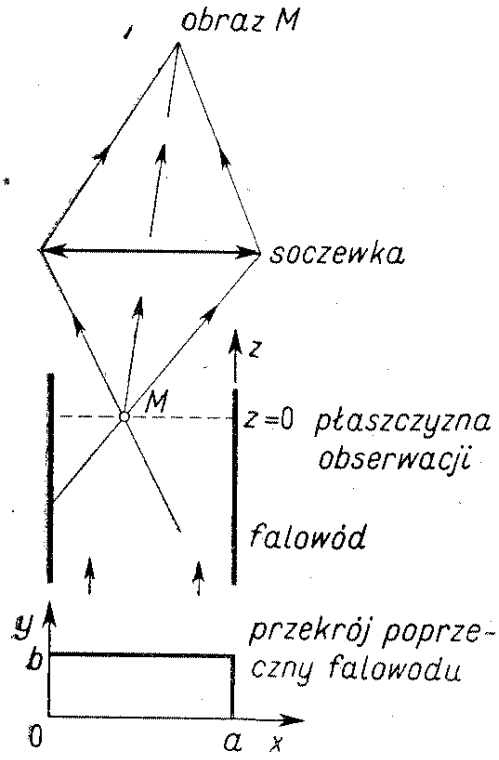
Warunki obserwacji podano na rys. 16.2. Chcemy mierzyć x w całym przedziale a z błędem Δx i nie interesuje nas y, b lub Δy. Przy równaniu (15.62) opisującym właściwe fale zwróciliśmy uwagę na współczynnik $\cos\left(\frac{\pi l x}{a}\right)$ i wartości l, podczas gdy pomijaliśmy współczynnik $\cos\left(\frac{\pi m y}{b}\right)$

i wartości m, jeżeli to było tylko dopuszczalne. Możemy np. wybrać $b<a$ i dla fali podstawowej przyjąć $l = 0$, $m = 1$, jako że wartości te dają mniejszą częstotliwość graniczną (wzór (15.65)). Fala taka przedstawiać będzie wiązkę przechodzącą przez falowód oraz padającą na cząstkę M, a więc zakładamy, że dla wiązki tej

$$l = 0, \qquad m = 1. \tag{16.24}$$

Wiązka padająca stanowi superpozycję dwu fal płaskich, które zgodnie ze wzorem (15.66) rozchodzą się w kierunkach tworzących z osią z kąty określone przez

$$\cos \varphi_y = \pm \frac{\lambda}{2b}.$$



Rys. 16.2. Sposób obserwacji cząstki przy pomocy falowodu i soczewki. Cząstka znajduje się przy ognisku soczewki M

Dotychczas zakładaliśmy, że m może przyjmować wartości dodatnie. W ogólności m może przyjmować wartości dodatnie i ujemne, tzn. $m = \pm 1$.

Kąty φ_x oraz φ_y określono w rozdz. 15.8; w ogólności mamy

$$\cos \varphi_x = l \frac{\lambda}{2a}, \qquad \cos \varphi_y = m \frac{\lambda}{2b}.$$

Przypuśćmy, że l oraz m są dodatnimi lub ujemnymi liczbami całkowitymi.

Światło rozproszone przez cząstkę M może zawierać fale własne ze wszystkimi możliwymi wartościami l oraz m. Większość tych fal można wyeliminować przy pomocy soczewki cylindrycznej o długości b ustawionej w kierunku L. Urządzenie takie odrzuca fale składowe z wyjątkiem fali, dla której $m = 0$. Proces rozpraszania opisać można jako przejście

$$\left.\begin{aligned} &\text{dla } l\text{:} \quad \text{od } 0 \text{ do } l, \\ &\text{dla } m\text{:} \quad \text{od } 1 \text{ do } 0. \end{aligned}\right\} \tag{16.25}$$

Jednakże, by uwzględnić znaki dodatnie i ujemne $\cos \varphi_x$ oraz $\cos \varphi_y$, musimy

rozpatrzyć też przejścia:

$$\left.\begin{array}{lll} \text{dla } l: & \text{od} \quad 0 \text{ do } \pm l, \\ \text{dla } m: & \text{od } \pm 1 \text{ do } 0. \end{array}\right\} \tag{16.25a}$$

Zmiana δl w czasie rozpraszania fotonu nadaje cząstce przyrost momentu pędu:

$$\delta p_x = \frac{h}{\lambda}\,\delta(\cos\varphi_x) = \frac{hl}{2a}. \tag{16.26}$$

Znak $\pm$ uwzględniony jest teraz w l i wartość l może przebiegać od $-l_M$ do -1 oraz od $+1$ do $+l_M$. Wartość $l = 0$ jest zabroniona, jako że l oraz m nie mogą jednocześnie być zerami.

Warunek (15.69) wymaga tego, abyśmy użyli n składowych w każdej fali w celu określenia t z błędem $\tau = \Delta t$. Ponieważ momenty pędu δp_x mogą być dodatnie lub ujemne, przeto obliczamy sumę kwadratów i oznaczamy ją przez $(\Delta p_x)^2$:

$$(\Delta p_x)^2 = \sum_{m,l} (\delta p_x)^2 = 4n\left(\frac{h}{2a}\right)^2 (1+4+9+\ldots+l_M^2 \tag{16.27}$$

lub

$$\Delta p_x = \left(\frac{h}{a}\right)\left[\frac{n}{3}\, l_M(l_M+1)\left(l_M+\frac{1}{2}\right)\right]^{1/2}.$$

Współczynnik 4 we wzorze (16.27) występuje ze względu na to, że wchodzą w grę cztery przypadki (równanie (16.25a)) odpowiadające kombinacjom $\pm l$ oraz ± 1. A więc, wykorzystując wartości l_M ze wzoru (15.68) otrzymamy

$$\Delta p_x\,\Delta x = h\left[\frac{\tau}{6\Delta t}\left(1+\frac{\Delta x}{a}\right)\left(\frac{a}{\Delta x}+\frac{1}{2}\right)\right]^{1/2}. \tag{16.28}$$

Jest to poprawiona zasada nieoznaczoności, która uwzględnia błąd określenia czasu i zakresu zmian x w doświadczeniu z mikroskopem.

Wartość τ może zmieniać się od $2\Delta t$ do ∞. Przypadek gdy $\Delta t = \tau$ nie może być wykorzystywany, gdyż nie uzyskuje się wówczas informacji.

Parametr a może zmieniać się od Δx do ∞. Tablica 16.1 zawiera wartości otrzymane dla kilku prostszych przypadków.

Gdy dokładność jest bardzo mała i określana głównie przez wymiary ekranów i przegród ($a = \Delta x$), to pojawia się zwykła nieoznaczoność. Możemy jednak

Tablica 16.1

$\frac{\tau}{\Delta t}$	$\frac{a}{\Delta x}$	$\left[\frac{\tau}{6\Delta t}\left(1+\frac{\Delta x}{a}\right)\left(\frac{a}{\Delta x}+\frac{1}{2}\right)\right]^{1/2}$
2	1	1
2	2	$\sqrt{5}/2 = 1{,}12$
2	3	$\sqrt{14}/3 = 1{,}25$
3	1	$\sqrt{1{,}5} = 1{,}22$
3	2	$\sqrt{7{,}5}/2 = 1{,}37$
duże wartości		$[a\tau/(6\,\Delta t\,\Delta x)]^{1/2}$

interesować się przypadkami, gdzie określania czasu i pomiaru odległości dokonuje się początkowo za pomocą systemu optycznego, co prowadzi do dużo większej nieoznaczoności. Przyjmijmy na przykład:

$$\frac{\tau}{\Delta t} = 6000, \qquad \frac{a}{\Delta x} = 1000, \qquad \Delta x\,\Delta p_x = h \cdot 10^3. \tag{16.29}$$

Dochodzimy do wniosku, że nieoznaczoność może stać się bardzo duża przy pomiarze położenia cząstki „swobodnej": taka swobodna cząstka może poruszać się na dużych odległościach, w ciągu długiego czasu, bez zakłóceń pochodzących od ekranów, przegród lub innych urządzeń mechanicznych. Oczywiście, takie mechaniczne urządzenia nie mogą być stosowane, gdy w grę wchodzą odległości mniejsze niż 1 Å, natomiast nasze rozważania mają większe znaczenie dla tego właśnie zakresu wielkości.

16.7. POMIAR PĘDU

W doświadczeniu z mikroskopem zakładaliśmy znajomość prędkości początkowej cząstki i próbowaliśmy mierzyć jej położenie. Wyniki nie były zadowalające. Spróbujemy odwrócić kolejność postępowania i zaczniemy od pomiaru położenia x w czasie t z dużą dokładnością. Pomiar taki nadaje cząstce duży, nieznany pęd (wzór (16.28)), ale nie ma to znaczenia, ponieważ i tak nie zna-

my prędkości początkowej. Wykorzystując zjawisko Dopplera* możemy mierzyć prędkość V_x oraz pęd p_x. Zwykłe rozumowanie prowadzi do relacji

$$\Delta x\, \Delta p_x \approx h. \tag{16.30}$$

Typowe rozważanie opiera się na założeniu, że pojedynczy foton $h\nu'$ może być wystarczający do dobrego pomiaru częstotliwości ν' w zjawisku Dopplera, przy warunku, że można użyć długiej paczki fal o całkowitym czasie trwania θ, tak aby

$$\Delta\nu'\theta \approx 1. \tag{16.31}$$

Chcemy obecnie prześledzić powyższe zagadnienie bardziej szczegółowo w oparciu o teorię informacji. Zmierzamy do wykazania, że pojedynczy foton nie wystarcza, i rozważymy zagadnienie, w jaki sposób liczba fotonów niezbędnych do doświadczenia związana jest z warunkami technicznymi postępowania doświadczalnego. Zmusza to nas do modyfikacji relacji nieoznaczoności, którą trzeba zapisać w postaci:

$$\Delta x\, \Delta p_x \approx nh, \qquad n \geqslant 1, \tag{16.32}$$

przy czym n zależy od warunków obserwacji. Powyższą granicę można obniżyć do granicy poprzedniej (16.30), gdy warunki będą tak wybrane, aby n było bliskie jedności.

Analiza ilości informacji wykazała, że zawsze musimy określać zakres zmienności i błąd mierzonych wielkości. W rozpatrywanym teraz zagadnieniu bierzemy:

$$\left.\begin{array}{lll} \text{czas obserwacji } t\text{:} & \text{błąd } \Delta t = \theta, & \text{zakres } \tau, \\ \text{prędkość cząstki } v_x\text{:} & \text{błąd } \Delta v_x, & \text{zakres } u. \end{array}\right\} \tag{16.33}$$

Tak jak w poprzednich przykładach wprowadzamy zakres τ, zakładając, że błyski o czasie trwania θ powtarzane są z okresem τ. Takie sygnały świetlne otrzymać można na drodze superpozycji n drgań o różnych częstotliwościach, gdzie

$$n = \frac{\tau}{2\Delta t} = \frac{\tau}{2\theta}. \tag{16.34}$$

* Zob. np. D. Bohm. *op. cit.*, str. 105.

Częstotliwości te zajmują pasmo $\delta\nu = \frac{1}{\theta}$. W wyrażeniu (16.33) określamy także zakres zmian prędkości u, jakie dopuszczamy w doświadczeniu. Będziemy projektować układ zdolny do mierzenia prędkości pomiędzy v_0 oraz v_0+u z błędem Δv. Oznacza to, że musimy użyć m rezonatorów nastrojonych na m różnych częstotliwości dopplerowskich ν', przy czym każdy rezonator musi być zdolny reagować w pasmie częstotliwości $\Delta\nu'$ odpowiadającym Δv:

$$m = \frac{u}{\Delta v}, \qquad \Delta\nu' = \nu\frac{\Delta v}{c}, \tag{16.35}$$

gdzie ν jest średnią częstotliwością początkową. Pomiar jest możliwy pod warunkiem, że początkowe pasmo $\delta\nu$ we wzorze (16.34) jest mniejsze od $\Delta\nu'$ ze wzoru (16.35). Zagadnienie doświadczenia wymagającego m równoczesnych obserwacji omówiono w rozdz. 14 i wykazano, że dla każdego rezonatora trzeba wybrać energię graniczną E_l daną wzorem

$$E_l = A_m kT. \tag{16.36}$$

Jeżeli obserwuje się energię E w rezonatorze, taką że $E<E_l$, to rezonator uznaje się za niewzbudzony, jeżeli natomiast $E>E_l$, to uważamy, że jest on wzbudzony. Jeśli założyć dużą pewność eksperymentu r, to zgodnie z rozdz. 14 mamy

$$A_m \approx \ln rm \qquad \text{dla} \qquad m \gg 1,\ r \gg 1. \tag{16.37}$$

Warunki te można spełnić wykorzystując jeden kwant $h\nu'$, jeśli

$$h\nu' > A_m kT.$$

W celu ustalenia momentu czasu posługujemy się n różnymi częstotliwościami ν', pokrywającymi pasmo $\delta\nu$ (wzór (16.34)); energia E_l odpowiadać będzie pojedynczemu kwantowi przy każdej częstotliwości ν', tak że bierzemy

$$h\nu' = E_l \approx A_m kT. \tag{16.38}$$

Ostatecznie energia rozproszona wynosi $nh\nu'$, a przyrost entropii

$$\Delta S = n\frac{h\nu'}{T} \approx nA_m k = kn \ln rm. \tag{16.39}$$

Otrzymuje się informację

$$\Delta I = k\ln\left(\frac{\Delta t}{\tau}\frac{\Delta v}{u}\right) = k \ln n\frac{m}{2} \tag{16.40}$$

i uogólniona II zasada jest spełniona:

$$\Delta(S-I) = k\left[(n-1)\ln m + n\ln r - \ln\frac{n}{2}\right] > 0. \tag{16.41}$$

Współczynnik liczbowy n może być mały, ale zawsze musi być większy od 2. Ponieważ założyliśmy dużą pewność i dużą dokładność ($r \gg 1$ oraz $m \gg 1$), przeto również ΔS jest dużo większe od ΔI.

Skoro posługujemy się n kwantami, otrzymujemy również odrzut energii n razy większy niż w zwykłym przypadku, a nieoznaczoność dana jest wzorem (16.32).

Weźmy np. $r = e^{10}$, $m = e^{10}$, $A_m = 20$, gdzie A_m obliczono ze wzoru (16.37). Zgodnie ze wzorem (16.38) wartości te dają

$$h\nu' \approx 20kT.$$

Jeżeli doświadczenie przeprowadzane jest w temperaturze pokojowej, to wzór (16.38) daje częstotliwości leżące blisko zakresu widzialnego.

W zastosowaniach do teorii pola posługiwać się będziemy wyrażeniem (16.30), które przedstawia dolną granicę (16.32).

16.8. NIEOZNACZONOŚĆ W POMIARACH POLA

Bohr i Rosenfeld* rozpatrzyli relacje nieoznaczoności w zastosowaniach do teorii pola, a w szczególności pola elektromagnetycznego. Wprowadzają oni do rozważań sondę o objętości V oraz ładunku ε, którą umieszczają w polu na czas τ. W chwili t jest mierzony pęd p_x, który w chwili $t+\tau$ przyjmuje już inną wartość p_x'. Średnia wartość składowej natężenia pola wzdłuż osi X na objętość V i za okres τ wyniesie

$$\overline{E}_x = \frac{p_x' - p_x}{\varepsilon\tau}. \tag{16.42}$$

Położenie sondy x mierzymy z błędem Δx, co wpływa na błąd Δp_x w pomiarze pędu p_x' oraz p_x. Przekonamy się, iż w tym przypadku stosować można relację (16.30).

* N. Bohr, L. Rosenfeld, Kgl. Danske Videnskal. Selskab. Mat. — fys. Medd. **12** (1953) art. 8.

W. Heitler, *The Quantum Theory of Radiation*, New York 1944, Oxford Univ. Press, rozdz. 2, str. 76.

Załóżmy, że najpierw mierzymy x, a potem, w chwili t, p_x (por. rozdz. 16.6). Następnie w chwili $t+\tau$ mierzymy p'_x i wreszcie x'. Ostatni pomiar daje często duży, nieznany pęd, ale to nie ma już znaczenia. Błąd pomiaru pędu p_x daje błąd pomiaru pola:

$$\Delta \overline{E}_x \approx \frac{h}{\varepsilon \tau \, \Delta x} \approx \frac{h}{e\tau Z \, \Delta x}, \tag{16.43}$$

gdzie $\nu = Ze$, e — ładunek elektronu.

Obserwacje przy pomocy mikroskopu pozwalają mierzyć pole dzięki oddziaływaniu tego pola na swobodną cząstkę. Pomiar odbywa się bez fizycznych przeszkód zakłócających pole. We wzorze (16.43) zamiast Δx przyjąć można średnicę cząstki, natomiast $Z\Delta x$ charakteryzuje cząstkę używaną jako sondę. Wartość $Z\Delta x$ może zmieniać się od 10^{-13} dla elektronu do 10^{-10} dla cząstek ciężkich ($Z = 96$, $\Delta x \approx 10^{-12}$). W celu określenia przedziału czasu τ niezbędna jest wiązka światła o częstotliwościach od 0 do ν_M:

$$\nu_M = \frac{1}{2\tau}, \tag{16.44}$$

a stąd

$$\nu \, \Delta \overline{E}_x \Delta x \approx 2h\nu_M. \tag{16.45}$$

Wzór ten posiada prostą interpretację, ponieważ iloczyn $\varepsilon \overline{E}_x \Delta x$ jest pracą wykonywaną przez pole $\overline{E}_x$ na drodze Δx. Pracy tej nie można pomierzyć z dużą dokładnością, gdyż może mieć miejsce emisja lub absorpcja kwantu $h\nu_M$ pochodzącego z wiązki światła, stosowanej do określania przedziału czasu.

Ostatecznie, rozważania nasze nie modyfikują podstawowej zależności (18.43), podanej przez Bohra i Rosenfelda, ale podkreślają niektóre trudności związane z relacjami nieoznaczoności. Zwykłe relacje nieoznaczoności odpowiadają absolutnej granicy teoretycznej, której nie daje się osiągnąć w doświadczeniach, gdzie z zasady nieoznaczoność bywa na ogół znacznie większa. W szczególności ma to miejsce w przypadku pomiaru położenia w pobliżu cząstki swobodnej.

Wiele zagadnień związanych z zastosowaniem uogólnionych relacji nieoznaczoności w fizyce omówimy w rozdz. 21 i 22. Zbadamy, jakie znaczenie ma dla praktyki zasadnicza niemożliwość osiągania dowolnie dużej dokładności w pomiarach oraz związany z tym brak pewności we wszystkich prawach fizycznych (16.1).

17

NEGENTROPIJNA ZASADA INFORMACJI W TELEKOMUNIKACJI

17.1. ANALIZA SYGNAŁÓW O SKOŃCZONEJ SZEROKOŚCI PASMA

W rozdziale 8 rozpatrzyliśmy podstawowe zagadnienie analizy sygnałów i udało się nam wykazać (rozdz. 8.6), że każdy sygnał o całkowitym czasie trwania τ i skończonej szerokości pasma

$$0 \leqslant |\nu| \leqslant \nu_M \tag{17.1}$$

ma tylko skończoną ilość stopni swobody

$$N = 2\nu_M \tau + 1. \tag{17.2}$$

Kiedy całkowity czas trwania τ jest duży, to wzór (17.2) możemy zastąpić wyrażeniem przybliżonym

$$N \approx 2\nu_M \tau. \tag{17.3}$$

Do tego ważnego wyniku doszliśmy dwiema metodami: metodą analizy fourierowskiej periodycznie powtarzanych sygnałów (rozdz. 8.6) i metodą próbkowania pojedynczego sygnału (rozdz. 7). W metodzie Fouriera stopnie swobody odpowiadają amplitudom pierwszych N harmonicznych w szeregu Fouriera. W metodzie próbkowania Shannona funkcja $f(t)$ przedstawiająca sygnał jest próbkowana w N równoległych chwilach ze stałym odstępem między tymi chwilami

$$\theta = \frac{\tau}{N} \approx \frac{1}{2\nu_M}. \tag{17.4}$$

Na podstawie próbek

$$f_m = f(m\theta) \tag{17.5}$$

można odtworzyć funkcję ciągłą według wzoru*

$$f(t) = \sum_m f_m \frac{\sin \pi(2\nu_M t - m)}{\pi(2\nu_M t - m)}.$$

Mówiliśmy również o innym podejściu, opartym na *komórkach informacyjnych* Gabora o objętościach określonych wzorem

$$\Delta\nu \, \Delta t = 1, \tag{17.6}$$

z których każda zawiera dwie składowe: jedną symetryczną i drugą antysymetryczną. Całkowita liczba takich komórek wynosi

$$n = \nu_M \tau, \tag{17.7}$$

co daje $N = 2\,n = 2\nu_M\,\tau$ oddzielnych składowych, jakie powinny być zadane dla określenia sygnału.

17.2. SYGNAŁY I SZUM TERMICZNY — PRZEDSTAWIENIE W HIPERPRZESTRZENI

Analiza z rozdz. 8 stosuje się do różnych rodzajów sygnałów. Jak możemy ją wykorzystać w przypadku szumu termicznego o skończonej szerokości pasma ν_M, który obserwowany jest przez czas τ?

W każdej komórce sygnałowej możemy określić średnie natężenie, ale faza jest przypadkowa. Mamy przeto n składowych, z których każda przedstawia oscylator harmoniczny o średniej energii kT (jeżeli zakładamy małe częstotliwości i pomijamy warunki kwantowe). Do tego wniosku dochodzimy wtedy, gdy zastosujemy analizę Fouriera dla całego systemu z jego n składowymi i przyjmiemy średnią energię kinetyczną $^1/_2\,kT$ dla każdej sinusoidy i cosinusoidy. Możemy zacząć także od N próbek Shannona i każdej

* Wzór (17.5) stanowi treść twierdzenia o próbkowaniu, podanego przez Shannona. Twierdzenie to w postaci podanej wyżej dotyczy sygnałów dolnopasmowych, czyli takich, których widmo zanika dla częstotliwości większych od ν_M. (*Przyp. tłum.*)

z nich przypisać średnią energię $^1/_2\,kT$. Każda z tych metod prowadzi w końcu do wniosku, że energia i moc wynoszą, odpowiednio,

$$\overline{E}_n = NkT/2 = \nu_M \tau kT, \tag{17.8}$$

$$P_n = \overline{E}_n/\tau = \nu_M kT. \tag{17.8a}$$

Odpowiada to sygnałom rozchodzącym się wzdłuż linii* w jednym kierunku, np. sygnał jest generowany od lewej strony linii i przechodzi na prawo. Wzór (17.8a) jest identyczny z (11.17), który wystąpił, gdy omawialiśmy to samo zagadnienie z punktu widzenia Nyquista.

Szum termiczny określony jest za pomocą mocy średniej i fluktuacji (por. rozdz. 9, 10 i 11). Gdyby moc szumu była dokładnie stała, można by go łatwo wyeliminować lub skompensować, lecz zakłócenia powstają rzeczywiście na skutek nie dających się przewidzieć fluktuacji, które są zawsze tego samego rzędu co moc średnia.

Rozważmy sygnał będący pewną dowolną funkcją o maksymalnej częstotliwości ν_M i czasie trwania τ. Funkcja ma N stopni swobody $X_1, X_2, \ldots, X_m, \ldots, X_N$, jak zdefiniowano to w rozdz. 17.1. Można przedstawić ją jako punkt o współrzędnych $X_1, X_2, \ldots, X_m, \ldots, X_N$ w N-wymiarowej przestrzeni S_N.** Każdy zbiór *komórek* będzie odpowiadał pewnemu układowi współrzędnych w S_N, natomiast zmiana układu współrzędnych w hiperprzestrzeni oznacza zmianę systemu komórek. Dla przedstawienia danej funkcji posłużymy się funkcjami ortogonalnymi. Całkowita energia będzie wówczas proporcjonalna do sumy kwadratów współrzędnych. Przy właściwym wyborze jednostek możemy napisać

$$E = \sum_{m=1}^{m=N} X_m^2 = R_N^2, \tag{17.9}$$

bez współczynnika liczbowego. Kiedy energia jest dana, punkt musi leżeć na hiperkuli o promieniu R_N. Dla dużych N objętość takiej kuli dana jest wzorem

$$V_N = C_N R_N^N \quad \text{przy} \quad C_N \approx (2\pi e/N)^{N/2}, \tag{17.10}$$

przy czym stała C_N zależy od liczby wymiarów N.

---

* Np. wzdłuż linii długiej. (*Przyp. tłum.*).

** Metodę tę bardzo szczegółowo omówił C. E. SHANNON, Proc. I.R.E. **37** (1949), str. 10.

Ciekawą własnością przestrzeni S_N o dużej liczbie wymiarów jest to, że objętość V_N jest skupiona głównie blisko powierzchni. Możemy dowieść tego wykazując, że małe przyrosty ΔR_N promienia mogą powodować duże przyrosty objętości:

$$\frac{\Delta V_N}{V_N} = N \frac{\Delta R_N}{R_N}. \tag{17.10a}$$

Jeżeli N przekracza $\dfrac{R_N}{\Delta R_N}$, wyrażenie to może stać się większe od jedności. Powinniśmy pamiętać, że praktycznie cała objętość takiej hiperkuli skupiona jest w cienkiej warstwie w pobliżu powierzchni.

17.3. PRZEPUSTOWOŚĆ KANAŁU Z SZUMEM(17.1)

Weźmy pod uwagę sygnał mający sens telegramu (17.2). Sygnał ten może być przedstawiony przez pewien punkt M w hiperprzestrzeni S_N. W obecności szumu punkt będzie przemieszczać się na pewną odległość w obrębie kuli o środku M. Promień takiej kuli równy jest $\overline{E}^{1/2}$ gdzie $\overline{E}_n$ jest średnią energią szumu. Nie wiemy, gdzie będzie wewnątrz tej kuli leżał nowy punkt M' w obecności nieznanego szumu. Dlatego nie powinniśmy wykorzystywać jako telegramu* innego sygnału, przedstawionego przez punkt leżący blisko punktu M. W związku z tym wyłania się zagadnienie dokładnego obliczenia, ile punktów można dostatecznie dobrze rozróżniać jeden od drugiego, czyli innymi słowy, jak wiele różnych sygnałów można rozróżnić w obecności szumu.

Niech całkowita energia sygnału będzie sumą energii sygnału E i średniej energii szumu $\overline{E}_n$. Stosownie do wzoru (17.9) suma ta określa promień

$$R' = (E+\overline{E}_n)^{1/2} \tag{17.11}$$

hiperkuli dla sygnału + szum. Telegram niezaszumiony leży w środku małej kuli o promieniu

$$r = \overline{E}_n^{1/2} \tag{17.12}$$

* Chodzi nam tutaj o sygnał, który pojawiłby się na wejściu odbiornika, gdyby w kanale nie było szumu; sygnał taki należałoby ściśle nazwać *odebranym telegramem niezaszumionym* w odróżnieniu od sygnału obserwowanego w obecności szumu, który nazywać można *odebranym telegramem zaszumionym*. W przypadkach nie budzących wątpliwości przymiotniki będziemy opuszczać. (*Przyp. red.*)

i musimy założyć, że szum dodany do danego telegramu zaciemnia kulę o promieniu r, wokół punktu M oznaczającego telegram. Powstaje pytanie: ile tych małych kul o promieniu r może być zawartych w dużej kuli o promieniu R'? Na pewno stosunek objętości daje górną granicę tej wielkości:

$$G = \frac{R'^N}{r^N} = [(E+\bar{E}_n)/\bar{E}_n]^{\frac{N}{2}} = \left(1+\frac{P}{P_n}\right)^{\frac{N}{2}}, \tag{17.13}$$

gdzie P oraz P_N oznaczają, odpowiednio, moce średnie telegramu i szumu. Zdefiniowaliśmy poprzednio (wzór (4.3)) przepustowość kanału jako granicę:

$$C = K \lim_{\tau\to\infty} \frac{\ln G(\tau)}{\tau}, \tag{17.14}$$

gdzie $G(\tau)$ jest całkowitą liczbą różnych telegramów ([17.3]) o długości τ, które można przesłać z daną pewnością przez kanał. Porównując wzory (17.13) i (17.14) otrzymujemy

$$C = K\left(\lim_{\tau\to\infty}\frac{N}{2\tau}\right)\ln\left(1+\frac{P}{P_n}\right);$$

ale N dane jest przez wzór (17.3), przeto

$$C = Kv_M \ln\left(1+\frac{P}{P_n}\right), \tag{17.15}$$

gdzie

K — stała zależna od przyjętych jednostek;
v_M — górna granica pasma częstotliwości;
P — średnia moc sygnału;
P_n — średnia moc szumu.

Równość (17.15) znana jest jako wzór Hartleya–Tullera–Shannona*.

17.4. OMÓWIENIE WZORU TULLERA–SHANNONA

Wzór (17.15) ze względu na swoją ważność wymaga dokładnego omówienia, aby usprawiedliwić wprowadzone przybliżenie i udowodnić, że granicę (17.13) można rzeczywiście osiągnąć. Przede wszystkim musimy uwzględnić

* Dokładne omówienie można znaleźć w pracy M. Leifera, W. F. Schreibera, Advances in Electronics **3** (1951), str. 306.

wpływ fluktuacji i nie dających się przewidzieć zmian szumu. Wszystko to ujęto ogólnym wzorem (9.32):

$$p(E, T) = B\mathrm{e}^{-\beta E}, \quad \beta = \frac{1}{kT}, \tag{17.16}$$

gdzie $p(T, E)$ jest prawdopodobieństwem mikrostanu (lub stanu kwantowego) o energii E przy temperaturze T. Współczynnik B wynika z warunku, że suma $p(E, T)$ musi być równa 1. Rozważanie powyższe można zastosować bezpośrednio do szumu. Niech $x_1, x_2, \ldots, x_m, \ldots, x_n$ będą zmiennymi przedstawiającymi tylko szum. Wtedy wzór (17.9) wyraża całkowitą energię szumu:

$$E_n = \sum_m x_m^2 = r^2. \tag{17.17}$$

Z drugiej strony — w równaniu (17.10) mamy objętość $\mathrm{d}V$ zawartą między r oraz $r+\mathrm{d}r$, a ilość stanów kwantowych jest proporcjonalna do tej objętości. Prawdopodobieństwo tego, że punkt leży między r oraz $r+\mathrm{d}r$, wynosi

$$p(r, T)\mathrm{d}r = B'\mathrm{e}^{-\beta r^2} r^{N-1}\,\mathrm{d}r \tag{17.18}$$

i stąd

$$\mathrm{d}V = NC_N r^{N-1}\,\mathrm{d}r,$$

przy czym B' oznacza nową stałą normalizacji.
Nowe prawdopodobieństwo $p(r, T)$ ma bardzo ostre maksimum przy pewnej wartości r, gdzie

$$\left.\begin{aligned} \frac{\delta}{\delta r}\left(r^{N-1}\mathrm{e}^{-\beta r^2}\right) &= (N-1-2\beta r^2) r^{N-2}\mathrm{e}^{-\beta r^2} = 0, \\ r_{mp}^2 &= \frac{N-1}{2\beta} = \frac{(N-1)kT}{2}. \end{aligned}\right\} \tag{17.19}$$

Łatwo udowodnić, że ze wzrostem N maksimum to staje się coraz ostrzejsze. Równanie (17.19) daje najbardziej prawdopodobną wartość r^2 i stąd najbardziej prawdopodobną wartością energii szumu jest

$$\overline{E}_{mp} = r_{mp}^2 = \frac{(N-1)kT}{2} \approx \frac{NkT}{2} = \overline{E}_n, \quad N \gg 1. \tag{17.20}$$

Jest to typowy przykład zgodności wartości najbardziej prawdopodobnej E_{mp} z przeciętną wartością $\overline{E}_n$, gdy ilość zmiennych N staje się bardzo duża. Przykład takiego zjawiska mieliśmy już w rozdz. 4.7 i 4.8.

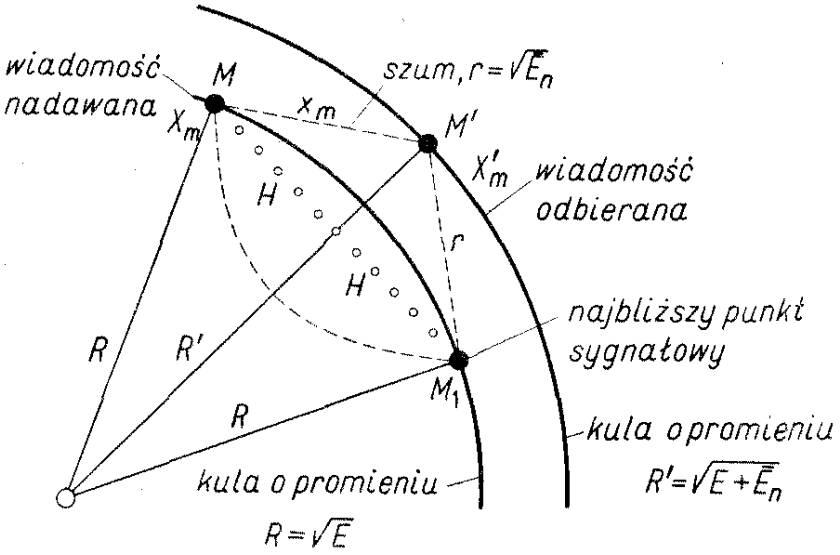
W naszym zagadnieniu telegram przedstawiony jest w hiperprzestrzeni N-wymiarowej przez punkt o współrzędnych X_m, a telegram *plus* szum przez punkt o współrzędnych

$$X'_m = X_m + x_m,$$

gdzie x_m jest wartością przypadkową.

Przeprowadzone rozważania dowodzą, że ten drugi punkt o współrzędnych X'_m musi praktycznie leżeć na powierzchni kuli o promieniu r_{mp} i o środku w punkcie początkowym M. Są to dokładnie warunki odpowiadające równaniom (17.12) i (17.13) przy wyprowadzeniu wzoru Tullera–Shannona.

Rozważania nasze dotyczyły tylko małych częstotliwości, gdy $h\nu$ jest dużo mniejsze od kT i można przyjąć, że stany kwantowe leżą nieskończenie blisko siebie. W takich okolicznościach często można założyć, że szum jest szumem



Rys. 17.1. Przekrój płaski hiperkuli N-wymiarowej pokazujący dwa najbliżej leżące punkty odpowiadające telegramom i efekt szumowy

gaussowskim (17.4). Rozkład w kierunku osi odpowiedniej zmiennej x_m jest niezależny od innych zmiennych i podlega prawu rozkładu $e^{-\beta x_m^2}$ zgodnie ze wzorami (17.16) i (17.17). Szum gaussowski nie odnosi się do układu kwantowego, jeżeli nie można założyć, że kwanty $h\nu$ są nieskończenie małe.

Rysunek 17.1 ilustruje powyższą sytuację z uwzględnieniem wyrażenia (17.16), z którego wynika, że punkty leżące w obrębie kuli muszą praktycznie leżeć na powierzchni tej kuli. Punkt M przedstawiający telegram niezaszumiony o energii E leży na kuli o promieniu $E^{1/2}$. Telegram ten *plus* szum

przedstawiony jest przez punkt M' leżący na kuli o promieniu $(E+\overline{E}_n)^{1/2}$, a odległość MM' wynosi $r = \overline{E}_n^{1/2}$ (por. wzory (17.12) i (17.19)). M_1 przedstawia możliwą pozycję innego sygnału odpowiadającego telegramowi, który mógłby być pomylony z sygnałem M, jeżeli punkt ten znajduje się w takiej samej odległości r' od M'. Odległość MM_1 przedstawia najmniejszą możliwą odległość między rozróżnialnymi punktami. Rysunek 17.1 jest przekrojem hiperprzestrzeni określonym trzema punktami M, M' oraz M_1. Odległość $MM_1 = 2H$ przyjmuje wartość graniczną, gdy MM' jest prostopadłe do OM, natomiast $M'M_1$ jest prostopadłe do OM_1. Oznacza to, że powierzchnia A trójkąta OMM' wynosi

$$A_{OMM'} = \frac{1}{2} H(\overline{OM'}) = \frac{1}{2} (\overline{MM'})(\overline{OM}), \tag{17.21}$$

gdzie

$$HR' = Rr \text{ oraz } H = \sqrt{\frac{E\overline{E}_n}{E+\overline{E}_n}} < r = \sqrt{\overline{E}_n}. \tag{17.22}$$

Odległość $2H$ pomiędzy punktami M oraz M_1 jest rzeczywiście mniejsza niż odległość $2r$, gdzie r jest promieniem kuli szumu okrążającej punkt M.

Musimy teraz zbadać rozkład punktów sygnałowych M w kuli R o objętości

$$V = C_N R^N = C_N E^{1/2N}, \tag{17.23}$$

aby zabezpieczyć transmisję z prawdopodobieństwem błędu dowolnie małym. Niech

$$\eta = e^{-\alpha} \ll 1 \tag{17.24}$$

wyraża prawdopodobieństwo błędu, który możemy tolerować. Rozkładamy punkty odpowiadające telegramowi ze stałą gęstością średnią na objętość V. Jeden z możliwych sposobów uzyskania stałej gęstości punktów na kuli polega na wykorzystaniu przypadkowego rozkładu w objętości (jak sugerował to Shannon). Niech C_1 będzie całkowitą liczbą takich punktów odpowiadających telegramowi. Każdy punkt wpadający w kulę o promieniu H otaczającą punkt M może być pomylony z punktem M' i dać błąd. Objętość tej kuli wynosi

$$u = C_N H^N. \tag{17.25}$$

Punkty wpadające w pozostałą objętość, $V-u$, nie wnoszą żadnego błędu.

Prawdopodobieństwo błędu jest oczywiście proporcjonalne do G_1 oraz u/V, a żądamy, aby

$$\varepsilon = \frac{G_1 u}{V} \leqslant \eta. \tag{17.26}$$

Oznacza to, że możemy posłużyć się liczbą punktów równą

$$G_1 \leqslant \eta V/u = \eta(R/H)^N = \eta[(E+\overline{E}_n)/\overline{E}_n]^{N/2} \tag{17.27}$$

zgodnie ze wzorami (17.12), (17.13) i (17.15). Można to zapisać w postaci

$$G_1 \leqslant \eta G, \tag{17.28}$$

gdzie G jest wielkością zdefiniowaną przez wzór (17.13). Maksymalna liczba G_1 odpowiada

$$C_1 = K \lim_{\tau\to\infty}[(\ln G_1)/\tau] = K \lim_{\tau\to\infty}\left(\frac{\ln G}{\tau} + \frac{\ln \eta}{\tau}\right)$$

lub, zgodnie ze wzorem (17.24),

$$C_1 = C - \lim_{\tau\to\infty}\frac{\alpha}{\tau}. \tag{17.29}$$

Przepustowość kanału można zbliżyć dowolnie do granicy teoretycznej C określonej poprzednio przez wzór (17.15). Niezbędne jest w tym celu uzyskanie rozkładu punktów odpowiadających telegramowi ze stałą gęstością po kuli o promieniu R.

17.5. PRZYKŁAD PRAKTYCZNY

W poprzednich rozważaniach zakładaliśmy, że sygnały (telegramy) mają długi całkowity czas trwania τ. Zbadajmy teraz, jak takie sygnały mogłyby być skonstruowane z niektórych standardowych sygnałów. Zaczniemy od systemu komórek z rozdz. 17.1. Nadaje się tu system komórek taki jak np. system odpowiadający szeregowi Fouriera lub odpowiadający metodzie próbkowania Shannona. Ze wzorów (17.7) i (17.8) mamy

$$n = \nu_M \tau \text{ komórek} \tag{17.30}$$

i każda z nich zaszumiona jest energią

$$\overline{e}_n = kT. \tag{17.31}$$

Całkowita średnia energia szumu

$$\bar{E}_n = n\bar{e}_n = nkT.$$

Dla danego telegramu z pewnej komórki wybierzemy poziom energetyczny E. Telegram ten zakłócany jest przez szum o średniej energii $\bar{e}_n$, która dodaje się do energii sygnału E wraz z fluktuacjami tego samego rzędu, przez co praktycznie pokrywa pasmo energii $2\bar{e}_n$. Zgodnie z tymi (raczej powierzchownymi) rozważaniami możemy próbować wybrać w każdej komórce zbiór poziomów energetycznych odległych wzajemnie o $2\bar{e}_n$ i użyć ich jako sygnałów elementarnych. Same telegramy wypełnią energie do pewnej wartości maksymalnej:

$$e_s = 0,\ 2\bar{e}_n,\ 4\bar{e}_n,\ \ldots,\ 2(m-1)\bar{e}_n, \tag{17.32}$$

a całkowite energie telegramu zaszumionego przybierają wartości

$$e_t = e_s + \bar{e}_n = \bar{e}_n,\ 3\bar{e}_n,\ 5\bar{e}_n,\ \ldots,\ 2(m-1)\bar{e}_n.$$

Mamy przeto m dokładnie określonych telegramów przypadających na każdą komórkę.

Cały telegram o czasie trwania τ otrzymamy podając sygnały odpowiadające n komórkom. Całkowita energia telegramu wynosi średnio dla telegramu niezaszumionego

$$\left.\begin{aligned} \bar{E}_s = n\bar{e}_s = n(m-1)\bar{e}_n, \\ \text{a dla telegramu zaszumionego} \\ \bar{E}_s + \bar{E}_n = nm\bar{e}_n. \end{aligned}\right\} \tag{17.33}$$

Definicje te zgadzają się z poprzednimi. Jeżeli czas trwania telegramu jest długi, to $\bar{E}_s$ będzie się bardzo mało różnić od wartości średniej i wszystkie telegramy praktycznie będą miały energie o wartościach danych przez wzór (17.33). Jest to znowu przypadek zgodności wartości najbardziej prawdopodobnej z wartością średnią.

Liczba rozróżnialnych telegramów uzyskanych w ten sposób wynosi

$$G' = m^n$$

i przepustowość kanału wyrazi się wzorem

$$C' = K \lim_{\tau \to \infty} [(\ln G')/\tau] = K\nu_M \ln m. \tag{17.34}$$

Zgodnie ze wzorem (17.33)

$$m = 1+(\bar{e}_s/\bar{e}_n) = 1+(\bar{E}_s/\bar{E}_n) = 1+(\bar{P}_s/P_n), \tag{17.35}$$

przy czym $\bar{P}_s = {}^1/_2 P_M$, gdzie P_M jest maksymalną mocą sygnału (telegramu). Tak więc przepustowość C' równa jest przepustowości granicznej danej przez wzór Tullera–Shannona (17.15). Model ten wykorzystamy do dalszych rozważań.

Rozważania te usprawiedliwiają wybór przedziału $2\bar{e}_n$ wprowadzonego do wzoru (17.32), gdyż prowadzi on do wzoru Shannona. Bardziej zwarte rozmieszczenie punktów odpowiadających telegramom prowadziłoby do zachodzenia na siebie dwóch pasm energetycznych, odpowiadających dwom następującym po sobie poziomom sygnałów. Rzadsze rozłożenie może zwiększyć niezawodność transmisji, lecz nie wykorzystuje się tu pełnej pojemności kanału.

Te dodatkowe uwagi można wykorzystać przy wykazaniu powiązania z analizą przedstawioną w rozdz. 17.4. Chcemy najpierw dowieść, że równoodległe poziomy energetyczne (wzór (17.32)) odpowiadają równomiernej gęstości punktów odpowiadających telegramom w hiperprzestrzeni $N = 2n$-wymiarowej.

Rozpatrzmy pewną komórkę, powiedzmy i-tą, która będzie zawierać 2 stopnie swobody: jeden typu sinusoidalnego, a drugi cosinusoidalnego. Nazwijmy je X_i i Y_i. Te $2n$ współrzędnych X_i, Y_i przedstawia hiperprzestrzeń N-wymiarową. Przyjmijmy równomierną gęstość $\varrho(X_i, Y_i)$ punktów odpowiadających telegramom na płaszczyźnie X_iY_i. Daje to średnio w małym obszarze $\mathrm{d}X_i\,\mathrm{d}Y_i$, liczbę punktów

$$\mathrm{d}G = \varrho(X_i, Y_i)\,\mathrm{d}X_i\,\mathrm{d}Y_i.$$

Wybierzmy teraz obszar pierścieniowy między r_i oraz $r_i+\mathrm{d}r_i$ taki, że

$$r_i^2 = X_i^2+Y_i^2 = E_i,$$

gdzie E_i oznacza odpowiednią wartość energii.

Przeciętna ilość punktów odpowiadających telegramom w tym pierścieniu wynosi

$$\mathrm{d}G = \varrho(X_i, Y_i)2\pi r_i\,\mathrm{d}r_i = \pi\varrho(X_i, Y_i)\,\mathrm{d}E_i \tag{17.36}$$

i otrzymujemy równomierną gęstość energii dla i-tej komórki:

$$\varrho'(E_i) = \pi\varrho(X_i, Y_i). \tag{17.37}$$

Odpowiada to założeniom (17.32).

Wzór (17.32) wyprowadzono zakładając równą odległość poziomów energetycznych z równymi prawdopodobieństwami aż do pewnego poziomu maksymalnego. Możemy wprowadzić inne założenie i rozważyć poziomy energetyczne z prawdopodobieństwem a priori

$$\pi_i = B\mathrm{e}^{-\gamma E_i} \tag{17.38}$$

gdzie E_i jest energią i-tej komórki. Daje to prawdopodobieństwo

$$\pi_i = B\mathrm{e}^{-\gamma r_i^2} = B\mathrm{e}^{-\gamma(X_i^2+Y_i^2)} \tag{17.39}$$

w N-wymiarowej hiperprzestrzeni. Zagadnienie to jest bardzo podobne do przedyskutowanego poprzednio (wzór (17.16)). Nowa stała γ zastępuje po prostu poprzednią β i w miejsce energii szumu bierzemy energię sygnału. Wszystkie punkty praktycznie leżą na kuli:

$$R^2 = (N-1)/2\gamma = \overline{E}, \tag{17.40}$$

gdzie $\overline{E}$ jest średnią wartością energii telegramów.

17.6. NEGENTROPIJNA ZASADA INFORMACJI W ZASTOSOWANIU DO KANAŁU Z SZUMEM

Podkreśliliśmy w poprzednich rozdziałach zasadę, że każda informacja otrzymywana o układzie fizycznym opłacana jest odpowiednią ilością ujemnej entropii branej z układu i odwrotnie, informacja zmagazynowana lub wysłana przez układ fizyczny odpowiada wzrostowi negentropii w układzie. Zasadę tę można sprawdzić na omówionych przykładach. Przytoczymy tu rozważania z artykułu Raymonda* odnoszące się do tego zagadnienia i wykażemy, w jaki sposób należy poprawić błędny współczynnik $^1/_4$.

Załóżmy, że telegramy są sygnałami elektrycznymi przechodzącymi przez kabel. Kabel ten znajduje się początkowo w temperaturze T. Ma on w normalnych warunkach entropię S_0, poziom szumu odpowiadający średniej

---

* R. C. Raymond, Am. Scientist **38** (1950), str. 275; zob. również C. A. Desoer, Man. Inst. Technol. Research Lab. Electronics, Quart. Progr. Rept. **46** (1951).

I. H. Felker, Proc. I.R.E. **40** (1952), str. 728; D. A. Bell, Am. Scientist **40** (1952), str. 682; L. Brillouin, J. Appl. Phys., **25** (1954), str. 595.

mocy P_n w pasmie o szerokości ν_M oraz pewien kierunek propagacji (wzory (11.77) i (17.8)):

$$P_n = kT\nu_M. \tag{17.41}$$

I tu Raymond stosuje w swoim wzorze (4) niepoprawny współczynnik 4, co doprowadziło do kłopotliwego współczynnika $^1/_4$ w wyrażeniu (5).

Gdy przesyłamy sygnał, wykorzystujemy moc średnią P w ciągu pewnego czasu τ i nie uwzględniamy ciepła. Entropia kabla nie ulega pierwotnie zmianie, ale jeśli nie odbierzemy nadanego sygnału i pozostawimy go w kablu, to będzie on przechodził tam i z powrotem i będzie stopniowo tłumiony na opornościach z powodu efektu Joule'a, aż całkowita energia $P\tau$ przemieni się w ciepło. Końcowy etap odpowiada entropii S większej niż entropia początkowa S_0. Negentropia ΔN wprowadzona na początku do kabla przez nadawany sygnał dana jest zależnością

$$\Delta N = S - S_0. \tag{17.42}$$

Jest to wielkość, którą musimy obliczyć.

Możemy najpierw założyć silne sprzężenie między sygnałami i układem fizycznym (kablem). Jest to założenie Raymonda, że „pojemność cieplna zamkniętego układu jest tak duża, że zmiana temperatury podczas pracy urządzenia komunikacyjnego jest do pominięcia". Prowadzi to do końcowej entropii

$$S = S_0 + P\frac{\tau}{T} \tag{17.43}$$

a stąd, wykorzystując wzór (17.41) otrzymujemy

$$\Delta N = P\tau/T = k\nu_M P\tau/P_n. \tag{17.44}$$

Negentropia ΔN przedstawia informację przepływającą w ilości $\Delta N/\tau$ na sekundę i stąd przepustowość kanału na sekundę wynosi

$$C_1 = \frac{\Delta N}{\tau} = k\nu_M \frac{P}{P_n}. \tag{17.45}$$

Jeżeli porównujemy ten wynik ze wzorem Shannona (17.15)

$$C_{Sh} = k\nu_M \ln\left[\left(1+\frac{P}{P_n}\right)\right], \tag{17.46}$$

widzimy, że oba wzory zgadzają się tylko dla bardzo małych wartości sto-

sunku P/P_n, gdy logarytm można rozłożyć w szereg i ograniczyć się do pierwszych wyrazów. Wzór Shannona otrzymano przy specjalnych założeniach, które rzeczywiście nie odpowiadają rozważanemu teraz modelowi. Shannon przyjął kabel bezstratny, bez oporności. Znaczy to, że nie ma sprzężenia pomiędzy prądami elektrycznymi przepływającymi przez kabel a samym fizycznym kablem. Zakłada się, że źródłem szumu jest źródło zewnętrzne i działające niezależnie. Nie ma tu prostego mechanizmu do wytworzenia końcowego przyrostu entropii, jaki znaleźliśmy w poprzednim przykładzie. Jeżeli przyjmujemy brak sprzężenia zwrotnego między sygnałami i kablem, musimy założyć inny mechanizm rozpraszania. Wzrost entropii może być spowodowany zniekształceniem sygnału lub nieregularną zmianą odstępów między impulsami prądowymi i zmianą ich natężenia.

Stan początkowy jest ułożonym ciągiem impulsów niosących wymaganą informację, a stan końcowy jest zdezorganizowanym ciągiem impulsów zakłócanych przypadkowo i nie niosących zrozumiałej informacji. Układ nasz posiada n oscylacyjnych stopni swobody (komórki z rozdz. 17.1), natomiast przypadkowy rozkład energii E_r związany jest z temperaturą T (wzór (17.8)):

$$E_r = nkT. \tag{17.47}$$

Jeżeli założymy, że energia wzrasta stopniowo o wartości dE_r, to przyrost entropii wyniesie

$$S - S_0 = \int_{E_1}^{E_2} \frac{dE_r}{T} = nk \int_{E_1}^{E_2} \frac{dE_r}{E_r} = nk \ln \frac{E_2}{E_1}. \tag{17.48}$$

Stan początkowy E_1 odpowiada tylko mocy P_n szumu o czasie trwania τ. Końcowy poziom energetyczny E_2 otrzymamy wtedy, gdy wszystka dodatkowa energia $P\tau$ zostanie w przypadkowy sposób rozdzielona na n oscylatorów i dodana do energii szumu:

$$E_1 = P_n\tau, \qquad E_2 = (P_n + P)\tau. \tag{17.49}$$

Mamy więc wyrażenie na negentropię:

$$\Delta N = S - S_0 = nk \ln \frac{P_n + P}{P_n}.$$

Daje to przepustowość kanału na jednostkę czasu:

$$C_2 = \frac{\Delta N}{\tau} = k \frac{n}{\tau} \ln \left(1 + \frac{P}{P_n}\right) = k\nu_M \ln \left(1 + \frac{P}{P_n}\right), \tag{17.50}$$

która jest identyczna z podaną przez Shannona (wzór (17.46)). Różnica między poprzednim rezultatem (17.45) i ostatnim (17.50) wynika stąd, że model pierwotny był utrzymywany w stałej temperaturze T, podczas gdy w drugim modelu obserwuje się wzrost temperatury proporcjonalnie do wzrostu przypadkowej energii E_r, a równanie (17.47) słuszne jest dla wszystkich momentów czasu. Przyrost temperatury zmniejsza różnicę entropii i wpływa na przepustowość kanału.

Logarytm występuje tu również, jak w dobrze znanym wyrażeniu $\ln V$ na entropię gazu doskonałego (wzór (12.12)). Sytuację podobną przypomina zagadnienie poruszone w rozdz. 13.7 (wzór (13.55)), kiedy to dowiedliśmy, że ilość informacji I rośnie jak logarytm pewnej wielkości F, podczas gdy zmiana entropii ΔS jest liniową funkcją F. W zagadnieniach telekomunikacyjnych otrzymujemy podobne wzory.

Ilość informacji I (lub przepustowość kanału C ze wzorów (17.46) i (17.50)) jest logarytmem $F = 1 + \frac{P}{P_n}$, podczas gdy zmiana entropii w modelu Raymonda (wzór (17.44)) proporcjonalna jest do $F-1$ i model ten przedstawia rzeczywisty układ fizyczny z rozpraszaniem.

Na zakończenie musimy podkreślić fizyczne znaczenie naszych rozważań. Wzór Shannona rzeczywiście przedstawia granicę, jaka odpowiada optymalnym warunkom, tj. maksymalnej sprawności przemiany negentropii w informację. Wynika to z faktu, że obliczenia oparte na informacji (rozdz. 17.3) i na negentropii (wzór (17.50)) dają jednakowe wyniki, pod warunkiem, że wszystkie wielkości mierzymy w tym samym układzie jednostek. Jeżeli informacja mierzona jest w bitach, a entropia w jednostkach termodynamicznych, wówczas każdy bit odpowiada $k \ln 2 \approx 10^{-16}$ jednostkom termodynamicznym. Prowadzi to znów do stwierdzenia (por. rozdz. 14), że jeden bit informacji nie może być otrzymany niższym kosztem negentropijnym od $k \ln 2$.

17.7. ZMODYFIKOWANY WZÓR GABORA I ROLA ZDUDNIEŃ

Gabor* po rozważeniu systemu *komórkowego*, który wyjaśniliśmy w rozdz. 8, zbadał bardzo szczegółowo, jakie poziomy energetyczne można użyć w każdej komórce, dla przedstawienia rozróżnialnych telegramów. To oryginalne

* D. Gabor, Phil. Mag. [7] **41** (1950), str. 1161.

rozważanie oparte jest na teorii kwantowej, lecz we wszystkich praktycznych zastosowaniach kwanty hv są tak małe, że można stosować teorię klasyczną. Czyniąc to, można się przekonać, że nowe składowe we wzorze Gabora odpowiadają *zdudnieniom* z szumem. Zdudnienia te wpływają na wzrost energii fluktuacji. Jeżeli nakładamy na sygnał o amplitudzie A_1 szum o amplitudzie A_2 ze zmienną fazą ψ, to mamy

$$\left.\begin{aligned} I &= A_1 \cos \omega t + A_2 \cos(\omega t + \psi), \\ I^2 &= A_1^2 \cos^2 \omega t + A_2^2 \cos(\omega t + \psi) + 2A_1 A_2 \cos(\omega t + \psi) \cos \omega t. \end{aligned}\right\} \tag{17.51}$$

Uśrednijmy to teraz za kilka okresów oscylacji, zakładając powolne zmiany A_2 oraz ψ:

$$I^2 = \frac{A_1^2}{2} + \frac{A_2^2}{2} + A_1 A_2 \cos\psi. \tag{17.52}$$

Moc za taki stosunkowo mały okres czasu wyniesie

$$\overline{P} = P_1 + \overline{P}_2 + 2(P_1 \overline{P}_2)^{1/2} \cos \psi. \tag{17.53}$$

Moc średnia za dłuższy okres czasu będzie zawierała tylko dwa składniki, gdyż $\cos \psi$ będzie zerem:

$$\overline{\overline{P}} = P_1 + \overline{P}_2. \tag{17.54}$$

Ostatni składnik wzoru (17.53) przedstawia *zdudnienia* między sygnałami i szumem. Grają one ważną rolę w fluktuacjach mocy Δp:

$$\left.\begin{aligned} \overline{P} &= \overline{\overline{P}} + \Delta p, \\ \overline{\Delta p^2} &= \overline{p_2^2} + 4p_1 \overline{p_2}\ \overline{\cos^2 \psi} = \overline{(p_2)^2} + 2p_1 \overline{P_2} \end{aligned}\right\} \tag{17.55}$$

i stąd dla szumu termicznego mamy

$$\overline{(P_2)^2} = \overline{P_2^2}, \qquad \overline{\cos^2 \psi} = 1/2.$$

Ze wzoru (17.55) widzimy, że fluktuacje wzrastają ze zwiększeniem mocy P_1. Wyniki te można wykorzystać w obliczeniach liczby rozróżnialnych poziomów energii m. Wybierzmy kroki $2(\overline{\Delta p^2})^{1/2}$, zgodnie ze wzorem Shannona (wzór (17.32)). Są one dwa razy większe od kroków początkowo proponowanych przez Gabora. Powiedzmy, że P_M jest maksymalną mocą transmisji

telegramów. Ilość kroków (włączając jeden dla $P = 0$) wynosi

$$m_G = 1 + \int_0^{P_M} \mathrm{d}P/(2\sqrt{\Delta p^2}) = 1 + \frac{1}{2}\int_0^{P_M} \mathrm{d}P/\sqrt{\bar{P}_2^2 + 2P\bar{P}_2} =$$

$$= \frac{1}{2}\sqrt{1+2\frac{P_M}{P_2}} + \frac{1}{2}. \tag{17.56}$$

Przy małych wartościach stosunku P_M/p_2 wyrażenie na m_G upraszcza się do $(1+{}^1/_2 P_M/P_2)$ — co jest identyczne ze wzorem (17.35). Dla dużych stosunków sygnał/szum — m_G rośnie wolniej z P_M niż wyrażenie (17.35), które odpowiada wzorowi Shannona.

Zgodnie z warunkami obserwacji niezbędnym będzie zbadać, czy dodatkowe fluktuacje spowodowane *zdudnieniami* mogą być istotne, czy też nie. Rozważanie to pokaże, czy stosować wzór Shannona, czy Gabora. Wzór Gabora (17.56) prowadzi do następującego wyrażenia na przepustowość kanału:

$$C_G = K\nu_M \ln m_G = K\nu_M \ln \frac{1}{2}\left(1+\sqrt{1+2\frac{P_M}{P_2}}\right), \tag{17.57}$$

gdzie P_M jest maksymalną mocą sygnału. Otrzymano to z wzoru (17.34) biorąc wartość Gabora na m_G.

Gdy wybierzemy długi czas τ, średnia $\cos\psi$ zdąża do zera i *zdudnienie* można pominąć, co potwierdza użycie wzoru Shannona w sensie granicznym.

18

PISANIE, DRUKOWANIE I CZYTANIE*

18.1. PRZESYŁANIE INFORMACJI — INFORMACJA ŻYWA

W poprzednim rozdziale rozważyliśmy zagadnienia telekomunikacyjne przyjmując konwencjonalne podejście rozwinięte przez Shannona–Wienera i grupę z zakładów badawczych Bell Telephone. Zagadnienie ograniczono do najprostszego przypadku telekomunikacyjnego: przesyłania wiadomości telegraficznej lub telefonicznej od jednej osoby do drugiej. Jest to dobrze zdefiniowane i bardzo ważne zagadnienie, którego wyniki rozciągnąć można na wiele podobnych sytuacji, np. gdy jedna osoba mówi bezpośrednio do drugiej. Niekoniecznie jednak teoria musi ograniczać się do tego specjalnego przypadku i można ją rozciągnąć na bardziej ogólne zagadnienia przy warunku, że zwrócono specjalną uwagę na sposób postępowania.

W rozdziale niniejszym zamierzamy przedyskutować dwa ogólne zagadnienia: wykazać, że nasza negentropijna zasada informacji stwarza nam możliwość badania wszystkich możliwych rodzajów komunikacji, i wyciągnąć niektóre ważne wnioski o podstawowych prawach fizycznych.

Pierwszym interesującym zagadnieniem jest *rozgłaszanie*. Tak nazywamy przypadek, gdy jedna osoba przekazuje wiadomość wielu ludziom:

(1) jeden spiker mówi do mikrofonu połączonego N różnymi kablami z N różnymi odbiornikami,

---

* Zagadnienia te zostały przedyskutowane przez autora na seminarium na Uniwesytecie im. Harvarda (październik 1951), w wykładach na Uniwersytecie Browna (luty 1952) i w artykule powitalnym do A.P.C. na zebraniu w Waszyngtonie (maj 1952). Zob. również L. Brillouin, J. Appl. Phys. **25** (1954), str. 595.

(2) jeden telegram wysłany jest do N stacji,
(3) mówca występuje przed audytorium,
(4) rozgłaszanie radiowe.

W przypadkach (1) i (2) mamy oczywiście przykłady wyżej podanej łączności *osoba–osoba* w N równoległych kanałach. Jeżeli kanały są identyczne, to po prostu musimy przekazać moc N razy większą niż wymaga się jej w pojedynczym kanale.

Przypadki (3) i (4) są mniej korzystne, ponieważ duża ilość mocy jest bezużyteczna. W przypadku (3) fale głosowe absorbowane są przez ściany, meble i inne przedmioty pochłaniające. W przypadku (4) fale radiowe zostają tłumione i rozpraszają się w przestrzeni. Z wyjątkiem tych bardzo dużych, dodatkowych strat ostatnie przypadki podobne są do (1) i (2). W ogólności można zagadnienia te sprowadzić do prostych kanałów, pracujących równolegle. Zagadnienia te mają następujące wspólne cechy:

(a) Moc i negentropia rozchodzą się razem z informacją.

(b) Odbiornik absorbuje energię i negentropię w tym samym cyklu, w którym odbierana jest informacja. Sytuację tę charakteryzujemy mianem *informacji żywej*. Informacja przesyłana jest wraz z energią potrzebną do jej wykrycia.

(c) W procesie transmisji część energii ulega degradacji, tracona jest negentropia i tym samym zatraca się informacja.

(d) Do wszystkich przykładów stosuje się uogólniona zasada Carnota, która stwierdza, że informacja musi być opłacona negentropią większą lub równą otrzymanej informacji.

18.2. ZAGADNIENIE CZYTANIA I PISANIA

Całkowicie inną sytuację spotykamy wtedy, gdy badamy inny typ transmisji informacji, a mianowicie pisanie i czytanie. Oczywiście nieuniknione jest pewne wydatkowanie energii w procesie pisania, ale nie znajdujemy śladu tej energii na zapisanym papierze. Nie ma zauważalnej negentropii i informacja jest tu całkowicie oddzielona od energii i negentropii. Nie jest to odosobniony przykład, ponieważ podobną sytuację znajdujemy we wszystkich przypadkach zapisu informacji (druk, zapis fonograficzny, taśma magnetyczna, dziurkowane karty, taśmy itp.). We wszystkich tych przypadkach mamy

zmagazynowaną informację, która nie jest związana ani z energią, ani z entropią. Co więcej, możemy odzyskać informację bez niszczenia zapisu. Wygląda to na możliwość nieograniczonego pomnażania informacji. Jako przykład weźmy autora piszącego książkę, która będzie wydana. Nie rozpatrujemy zagadnienia myślenia i pracy autora. Jest to element ludzki, którego nigdy bardziej szczegółowo nie włączymy do naszych rozważań (rozdz. 1 i 20). Nie robi to żadnej różnicy, czy rozważamy napisaną przez autora książkę, czy tylko zbieranie danych liczbowych, takich jak ceny na giełdzie, obserwacje pogody itp. Jeśli rękopis zawiera N liter, to zawiera on pewną ilość informacji. Nasze definicje nie wychodzą poza ten punkt. Przypuśćmy, że wydrukowano 1000 kopii książki i że każdy egzemplarz czytany jest przez 100 osób. Całkowita ilość informacji została pomnożona przez 10^5, podczas gdy przyrost entropii jest nieporównywalnie mały. Oczywiście wzrost informacji będzie zmniejszony z takich przyczyn, jak błędy drukarskie, błędy przesyłania, niedostateczna zrozumiałość itp., ale współczynnik pomnożenia może być w praktyce jeszcze bardzo duży. Wymienione zagadnienie wymaga dokładniejszego zbadania dla wykazania, jak do tych sytuacji stosuje się zasadę negentropii.

18.3. INFORMACJA MARTWA — JAK JĄ OŻYWIAĆ

Przytoczone wyżej sytuacje scharakteryzujemy jako przypadki zmagazynowania informacji martwej, nie związanej z energią lub negentropią. W celu odczytania informacji musimy przypisać jej energię. Do odczytu absolutnie nieodzowne jest dodatkowe źródło energii i to źródło energii prowadzi do negentropii, która zamienia się w nową informację. By czytać książkę, potrzebujemy źródła światła. Adapter nie będzie pracować bez silnika obracającego płytę. Niewidomy czytając książkę drukowaną alfabetem Braille'a musi przesuwać palce wzdłuż kartek i wywierać dostateczne ciśnienie, aby rozróżnić druk od innych nieregularności na kartce, ten zaś ruch oraz wywierane ciśnienie wymagają pracy.

Potrzebujemy światła i komórki fotoelektrycznej do czytania taśm perforowanych lub baterii i szczotek przewodzących prąd przez dziurki w kartach perforowanych. We wszystkich tych przypadkach ilość negentropii branej ze źródła zewnętrznego może się okazać większa od uzyskanej informacji.

Czytanie perforowanej taśmy jest typowym przykładem, dla którego możemy obliczyć entropię i informację oraz sprawdzić związek między tymi wielkościami.

Załóżmy, że mamy informację zarejestrowaną na metalizowanej taśmie perforowanej w systemie dwójkowym. Jak pokazano na rys. 18.1, dziurki oznaczają zera, a jedynki są przedstawione odcinkami niedziurkowanymi. Każda sekcja taśmy zawiera jeden bit informacji, czyli

$$\Delta I = k \ln 2. \tag{18.1}$$

Jest to informacja przypadająca na sekcję, wyrażona w jednostkach termodynamicznych, gdzie k jest stałą Bolzmanna.

Przyrząd czytający pokazany jest na rys. 18.2. Wiązka światła, pochodząca ze źródła, pada na taśmę metalową. Jeżeli jest dziurka, światło trafia do ko-

informacja wyrażona za pomocą cyfr dwójkowych	0	1	0	0	1	1	1	0	1	0	0	1
taśma perforowana	0		0	0				0		0	0	

Rys. 18.1. Taśma perforowana ze słowem zakodowanym w systemie binarnym. Dziurki na taśmie odpowiadają zerom

mórki fotoelektrycznej P_1. Jeżeli dziurki nie ma, wiązka odbija się i pada na fotokomórkę P_2. Możemy pomyśleć o wyeliminowaniu drugiej komórki P_2 i wnioskować, że jedynka pojawia się zawsze, gdy P_1 nie zadziała. Musimy jednak przypomnieć sobie, że zawsze żądaliśmy, żeby odczytowi ujemnemu nie nadawać znaczenia. Każdy *odcinek* taśmy będzie wymagał co najmniej jednego kwantu $h\nu$ i energia ta będzie absorbowana w P_1 lub P_2. Aby odróżnić kwant ten od promieniowania termicznego, $h\nu$ musi nieść dość dużą energię, tak aby

$$h\nu \geqslant kT \ln 2, \tag{18.2}$$

jak to udowodniono w rozdz. 14.3 (wzór (14.17)).

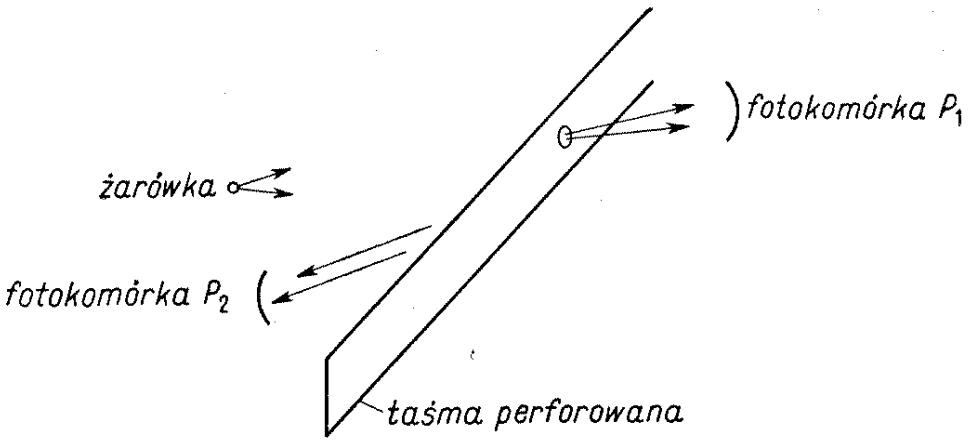
Gdy kwant zostanie zaabsorbowany, wówczas mamy utratę negentropii (przyrost entropii)

$$\Delta S = \frac{h\nu}{T} \geqslant k \ln 2 \tag{18.3}$$

i relację

$$\Delta(S-I) \geqslant 0, \tag{18.4}$$

co wskazuje na to, że spełniona jest uogólniona zasada Carnota. W tym prostym przypadku wykazujemy, że informacja pochodzi z negentropii wiązki świetlnej, i możemy teoretycznie osiągnąć granicę odpowiadającą sprawności



Rys. 18.2. Przyrząd służący do odczytu taśmy z rys. 18.1. Fotokomórka po prawej stronie służy do odbioru światła przechodzącego przez dziurki, a fotokomórka z lewej strony odbiera światło odbite od nieperforowanych odcinków taśmy

równej 1. Nie dostrzegano tego zagadnienia z powodu bardzo małej ilości żądanej negentropii:

$$k \ln 2 \approx 10^{-16} \tag{18.5}$$

(w układzie CGS, stopnie Celsjusza).

Przedstawiony sposób dziurkowania nie jest wygodny, gdyż jedynki nie są rzeczywiście zapisywane. Ciągły odcinek taśmy byłby odczytany jako ciąg jedynek, a to nie jest prawdą. Wygodniejsze byłoby przyjęcie dwu różnych poziomów dziurek, jak pokazano na rys. 18.3, i dwu fotokomórek przyjmujących światło, które może przejść przez jedną z dwu dziurek. Takie urządzenie wymaga jeszcze jednak absorbcji jednego kwantu na cyfrę binarną, tak że zależności (18.1) do (18.4) obowiązywałyby jeszcze.

Zamiast światła posłużyć się możemy szczotkami i baterią tak, aby przepuszczać prąd przez dziurki. Prąd odczytywany jest na amperomierzu lub rejestrowany za pomocą przekaźnika. Zgodnie z rozważaniami z rozdz. 14.2, granicą w tym przypadku jest znowu $k \ln 2$ na jeden odczyt.

W urządzeniu mechanicznym (takim jak płyty gramofonowe) małe impulsy dodatnie lub ujemne można wykryć tylko wtedy, jeżeli układ mecha-

niczny daje energię większą niż $kT \ln 2$ na impuls, który odpowiada przyrostowi entropii o $k \ln 2$. Podobne uwagi odnoszą się do impulsów rejestrowanych na taśmie magnetycznej. We wszystkich przypadkach strata negentropii na operację odczytywania jest większa lub najwyżej równa otrzymanej informacji i uogólniona zasada Carnota (równanie (18.4)) jest zawsze spełniona.

Uwagi te odnoszą się również do zagadnienia informacji zapisanej na światłodruku. Za każdym razem, gdy potrzebna jest informacja, musimy

informacja	0	1	0	0	1	1	1	0	1	0	0	1
taśma	0	0	0	0	0	0	0	0	0	0	0	0

Rys. 18.3. Lepsza metoda perforowania taśmy polegająca na wykorzystaniu dwu poziomów, jednego dla zer i drugiego dla jedynek

czytać światłodruk. Proces odczytywania absorbuje negentropię ze źródła światła i część tej negentropii zamieniana jest w informację potrzebną, powiedzmy, dla konstrukcji przyrządu dostosowanego do światłodruku. Powrócimy jeszcze do tego zagadnienia.

18.4. PISMO I DRUK

Możemy rozważyć teraz inne zagadnienie, a mianowicie: jak dużo negentropii traci się w procesie pisania? Negentropia jest tracona, lecz otrzymujemy nową kopię informacji, tak że mechanizm drukujący jest innym przykładem transformacji negentropii w informację.

Najprościej rozważyć zagadnienie druku fotograficznego. Kładziemy taśmę perforowaną na papier fotograficzny i oświetlamy każdy otworek. Za każdym razem co najmniej jeden kwant światła $h\nu$ służy do wydrukowania położenia otworu. Sytuacja jest podobna do przypadku z rozdz. 18.3 i prowadzi do podobnych wniosków: drukowanie nowej kopii kosztuje dokładnie tyle samo, co czytanie. Nowa kopia informacji opłacana jest negentropią. We wszystkich tych rozważaniach pomijaliśmy typowe nieodwracalne procesy: wywoływanie papieru fotograficznego, wzmacnianie impulsów elektrycznych itp. Wszystkie te procedury wymagają dużej ilości negentropii i całko-

wity koszt czytania lub drukowania będzie dużo wyższy niż wartość graniczna, którą obliczyliśmy.

Podstawowym punktem jest to, że wszelka informacja opłacana jest negentropią. Ilość wymaganej negentropii jest skończona i jest tak mała (10^{-16} na bit), że do tej pory całkowicie ją pomijaliśmy. Powiązanie między entropią i informacją jest z logicznego punktu widzenia absolutnie nieuniknione. W tym rozdziale bylibyśmy bezradni w rozpatrywaniu zagadnień bez tego powiązania. Bardzo mała wartość (10^{-16}) negentropii wymagana na bit informacji odgrywa jednak bardzo ważną rolę we współczesnym życiu i stwarza możliwość przekazywania informacji przy pomijalnie małym koszcie. Musimy teraz wspomnieć o prawie rozpadu: dowolny układ, w którym informacja została zarejestrowana, będzie się stawał coraz gorszy z upływem czasu. Co więcej, sama informacja będzie traciła znaczenie.

Z obu tych względów zarejestrowana informacja podlega ogólnemu prawu rozpadu, które zgadza się z uogólnioną zasadą Carnota. Rozpad będzie jednak dużo wolniejszy dla informacji martwej (zarejestrowanej) niż dla żywej.

18.5. OMÓWIENIE PRZYPADKU SPECJALNEGO

Pisanie i czytanie rozpowszechniły się w codziennym życiu tak szeroko, że nie zdajemy sobie sprawy, jak często się nim posługujemy.

Rozważmy np. procedurę wysyłania telegramu. Wydzielić tu można następujące etapy:

(1) Nadawca pisze telegram na poczcie.

(2) Operator czyta telegram i zamienia go na impulsy prądowe w nadajniku.

(3) Impulsy prądowe odbierane są przez odbiornik i odpowiedni mechanizm drukuje telegram.

(4) Operator czyta adres i wysyła kopię telegramu do adresata.

(5) Adresat czyta telegram.

Opuściliśmy jeszcze dwa etapy, ponieważ zawierają one procesy myślenia, a współczesna teoria nie jest w stanie rozważyć tego procesu:

(1a) Nadawca komponuje telegram tak, żeby przekazać informację adresatowi.

(6) Adresat rozumie znaczenie telegramu.

Ograniczając procedurę do etapów (1), (2), (3), (4) i (5) zauważamy, że etapy te zawierają następujące procesy:

(1) Pisanie.
(2) Czytanie, kodowanie, nadawanie.
(3) Dekodowanie, drukowanie.
(4) i (5) Czytanie.

Każda operacja pisania, czytania czy drukowania pociąga koszty w negentropii zgodnie z rozważaniami z rozdz. 18.3 i 18.4. Operacje kodowania i dekodowania wymagają urządzeń pamięciowych, z których muszą być odczytane zakodowane liczby. Operacja ta pociąga również za sobą pewien koszt negentropijny. Operację tansmisji i odbioru, rozważaną przez Shannona, analizowano w rozdz. 17. Kopie telegramu trafią w ręce różnych zainteresowanych ludzi — odpowiadają one informacji martwej, którą można odczytywać wiele razy, pod warunkiem, że dostarczona jest właściwa energia.

Urządzenia pamięciowe, wspomniane wyżej, rozważymy bardziej wyczerpująco w rozdz. 19 przy omawianiu maszyn matematycznych, gdzie czytanie i pisanie jest zmechanizowane i za każdym razem wykorzystujemy pewną ilość negentropii.

18.6. NOWA INFORMACJA A ROZWLEKŁOŚĆ

Pokazaliśmy w rozdz. 18.3, że operacja czytania wymaga przyrostu entropii ΔS, która jest większa lub co najmniej równa ilości informacji ΔI otrzymanej przez czytelnika. Stawiamy teraz następujące pytanie: czy informację tę można traktować jako nową informację, czy powinna być traktowana tylko jako rozwlekłość, nie niosąca żadnej informacji nowej, skoro jest ona tylko powtórzeniem informacji zawartej w napisanym dokumencie. Pytanie to leży w obrębie filozofii, ale możemy spróbować dać na nie praktyczną odpowiedź. Gdy osoba czyta po raz pierwszy książkę, na pewno uzyskuje informację odpowiadającą tej części książki, którą jest w stanie zrozumieć. Jest to dla czytelnika nowa informacja. Jeżeli czyta on tę książkę kilkakrotnie, to ostatecznie może zrozumieć ją całkowicie i wyciągnąć z niej całą informację. Czytając powtórnie książkę uzyskuje tylko rozwlekłość, a nie uzyskuje żadnej informacji nowej. Z ogólnego punktu widzenia możemy wyobrazić sobie idealnego uczonego, znającego wszystko, co jest znane ludzkości. Dla

tego idealnego uczonego każda nowa książka zawiera informację. Inni czytelnicy uzyskują jednak informację zbyteczną, gdyż jest ona już znana *idealnemu uczonemu.* Jest to jednak raczej nierealny punkt widzenia.

W celu rozróżnienia ilości możliwych poglądów dotyczących informacji i rozwlekłości związanej z czytaniem i drukowaniem kopii, wprowadzimy następujące określenia:*

Informacja absolutna: Jest to dowolna część informacji dostępnej człowiekowi na Ziemi, przy czym informacja ta liczy się tylko jeden raz, bez względu na ilość ludzi znających tę informację. Jakiś supertajny dokument niósłby taką samą informację jak wydanie New York Timesa, zawierające tę samą ilość liter.

Informacja rozprowadzona: Jest to iloczyn ilości absolutnej informacji i liczby ludzi, którzy biorą w niej udział. Ta druga definicja wygląda bardziej realnie niż filozoficzna *informacja absolutna* i taki punkt widzenia przyjęliśmy w tym rozdziale. Są jednak zagadnienia, w których wybór definicji informacji absolutnej może być wygodniejszy, jednak różnice między obydwoma poglądami na informację powinno się mieć na uwadze.

---

* L. Brillouin, Am. Scientist **38** (1950), str. 594.

19

ZAGADNIENIE OBLICZEŃ

19.1. MASZYNY MATEMATYCZNE

Musimy teraz rozważyć zagadnienia związane z maszynami matematycznymi i przedyskutować wpływ tych maszyn na informację. Często mówi się, że maszyny matematyczne wytwarzają nową informację. W rzeczywistości tak nie jest. Maszyny są jedynie w stanie przetwarzać informację: pobierają na ogół surowe dane wejściowe i podają wynik końcowy, ale całkowita ilość informacji nie wzrasta. W najbardziej idealnych okolicznościach informacja może pozostać stała podczas liczenia, lecz w normalnych warunkach wystąpią straty i informacja wyjściowa będzie mniejsza od informacji wejściowej.

Rozpatrzmy zagadnienie bardziej szczegółowo. Wyróżnia się w nim 2 fazy: *programowanie* i właściwe *liczenie*. Programowanie przeprowadzają naukowcy, którzy zamierzają korzystać z maszyny. Oni analizują zagadnienie, wybierają prawa matematyczne, które powinny być zastosowane, przygotowują program operacji arytmetycznych i wskazują na ich kolejność. Wszystkie te operacje są zapisane na materiale, z którego maszyna może wczytać dane, tj. na taśmie perforowanej, taśmie magnetycznej itp. Ta część pracy wymaga wysiłku umysłowego i wiedzy od uczonych. Można to nazwać informacją w sensie ogólnym i jest to typowy przypadek, do którego nie stosują się nasze definicje ilości informacji. Niejednokrotnie podkreślaliśmy już fakt, że statystyczna definicja informacji nie dotyczy procesów myślowych człowieka i nie możemy, przynajmniej w tej chwili, analizować tych procesów.

Na przygotowanej raz taśmie perforowanej z programem operacji możemy obliczyć ilość bitów i obliczyć, jak dużo informacji zawiera ostateczny program. Nie wydaje się, aby to miało duże znaczenie.

Następnym krokiem jest rozważenie maszyny ze specjalnym programem i zbadanie, jak ona rzeczywiście pracuje. Na wejście podaje się pewne dane liczbowe. Maszyna dokonuje obróbki tych danych zgodnie z programem i oblicza dane końcowe.

Jeżeli maszyna nie popełnia błędów liczenia, to działa dokładnie, tak jak urządzenie kodujące. Wykorzystuje informację zawartą w danych wejściowych, przetwarza ją w dane wyjściowe, które co najwyżej mogą zawierać wszystką informację podaną na wejściu, ale nie więcej.

Maszyna może być porównana do kanału transmisyjnego, włączając również w kanał kodowanie i dekodowanie. Przy założeniu idealnych warunków, telegram zostaje poprawnie przesłany przez kanał i informacja w nim zawarta nie jest tracona. Błędy i przybliżenia w obliczeniach wpływają na straty informacji. Pracę wykonaną przez maszynę najlepiej można porównać z pracą tłumacza: tłumaczowi dano słownik oraz reguły gramatyczne, które pozwalają mu przetłumaczyć na język angielski artykuł napisany w języku japońskim. Słownik i gramatyka odpowiadają programowi operacji. Artykuł po japońsku — to dane wejściowe, a tłumaczenie angielskie przedstawia dane wyjściowe. Całkowitą informację podano na wejście i można ją w najlepszym wypadku znaleźć na wyjściu.

W podobny sposób maszyna oblicza dane do prowadzenia ognia artyleryjskiego, gdy w programie zawarte są prawa ruchu pocisku i warunki atmosferyczne. Położenie i prędkość początkowa są danymi wejściowymi. Wyjściem jest punkt uderzenia i prędkość końcowa. Dane wejściowe logicznie zawierają już rozwiązanie, lecz nie jesteśmy w stanie odczytać go bezpośrednio. Jest to zakodowany sygnał, który wymaga dekodowania, a to właśnie można inaczej nazwać *tłumaczeniem*.

Jeżeli maszyna pracuje doskonale, to musi zachodzić odwracalność: z żądanego punktu uderzenia może ona obliczyć z powrotem dane wejściowe i otrzymać warunki początkowe. W podobny sposób tłumacz może znów przetłumaczyć tekst angielski na język japoński i powinien otrzymać oryginalny artykuł lub jego równoważnik. Idealne warunki postępowania odpowiadają odwracalności i stąd zachowaniu informacji.

Odwrócenie całego procesu można uzyskać drogą kolejnych kroków,

które mogą zawierać operacje nieodwracalne, takie jak zapisywanie danych w pamięci i odczytywanie ich w późniejszym czasie.

Każda z tych operacji pociąga za sobą pewien wydatek negentropii (rozdz. 18.3 i 18.4). Operacja maszyny działającej według pewnego programu będzie rzeczywiście kosztować pewną ilość negentropii (energia zdegradowana). Jest to ważne z technicznego punktu widzenia, lecz nie zmienia to naszego wniosku i ktoś może wyobrazić sobie, że ten koszt operacji będzie zmniejszany w przyszłych lepszych maszynach.

Powołajmy się tu na bardzo ważną rozprawę, napisaną więcej niż 100 lat temu przez Edgara A. Poego. Autor omawia automaty różnego rodzaju, natomiast bardziej szczegółowo omawia maszynę wymyśloną przez Babbage'a. Poniższy cytat określa bardzo dokładnie nasz punkt widzenia:

«Skoro maszyny te są genialne, to co powinniśmy pomyśleć o maszynie Babbage'a? Co pomyślimy o maszynie z drzewa i metalu, która może nie tylko obliczać tabele astronomiczne i nawigacyjne o dowolnym rozmiarze, ale wykonywać ścisłe operacje matematyczne z możliwością poprawiania swoich błędów? Co pomyślimy o maszynie, która nie tylko może wykonywać to wszystko, ale drukować będzie obliczane wyniki bez najmniejszej interwencji człowieka? Można na to odpowiedzieć, że maszyna taka, jaką opisaliśmy, pod każdym względem przewyższa gracza szachowego Maelzela, i pod żadnym względem nie ustępuje graczowi, jeśli dopuścimy twierdzenie (czego nigdy nie należy robić), że gracz szachowy jest *maszyną czystą* i wykonuje swoje ruchy bez pośrednictwa człowieka. Obliczenia arytmetyczne i algebraiczne ze względu na swój charakter są ustalone i określone. Jeżeli występują określone *dane*, to koniecznie i nieuniknienie pojawiają się pewne wyniki. Na wyniki te nic nie wpływa, z wyjątkiem danych wejściowych.

Rozwiązywane zagadnienie posuwa się lub przynajmniej powinno posuwać się do wyznaczonego końca kolejnymi bezbłędnymi krokami nie podlegającymi zmianom i modyfikacjom. Jeżeli tak jest, to bez trudu możemy wyobrazić sobie możliwość zbudowania takiego mechanizmu, który po jego uruchomieniu, zgodnie z danymi odpowiadającymi rozwiązywanemu zagadnieniu, wykonuje swoje ruchy regularnie, posuwając się naprzód w kierunku pożądanego rozwiązania i chociaż ruchy byłyby nawet złożone, to nie należy ich sobie wyobrażać inaczej jak tylko jako ruchy niezbędne i zdeterminowane.»*

Maszyna nie wytwarza żadnej nowej informacji, ale dokonuje ona bardzo cennych transformacji znanej już informacji. Ciekawym zagadnieniem jest znalezienie jakiejś miary tej transformacji oraz obliczenie jej wartości. Do chwili obecnej nie znaleziono metod równoważnych pracy maszyny matematycznej ani też pracy tłumacza.

---

* Edgar A. Poe, *Maelzel's Chess Player*, 1836.

19.2. MASZYNA MATEMATYCZNA JAKO ELEMENT MATEMATYCZNY

Ciekawie wygląda rola maszyny z punktu widzenia analitycznego, gdy wykonuje ona operacje w oparciu o dane wejściowe będące obserwacjami pewnego układu fizycznego, przy czym seria obserwacji daje funkcję ciągłą $f(t)$. Jeżeli układ fizyczny przepuszcza skończone pasmo ([19.1]), to również funkcja $f(t)$ ma skończone pasmo.

Cyfrowa maszyna matematyczna wykonuje obliczenia w oparciu o próbki sygnałów. Próbkowanie odpowiada zagadnieniom poruszanym w rozdz. 8.6 i 8.7. Jeżeli chcemy wykorzystać całkowitą informację, jaką niesie funkcja ciągła, to należy próbkować z okresem θ tak, aby

$$\theta = \frac{1}{2\nu_M}, \qquad f_m = f(m\theta). \tag{19.1}$$

Pierwotna funkcja $f(t)$ jest praktycznie określona tylko w skończonym przedziale czasowym τ, a do ścisłych rozważań matematycznych musimy określić zachowanie się tej funkcji od $-\infty$ do $+\infty$, jak zrobiliśmy to w rozdz. 8.6. Załóżmy znów, że funkcja periodyczna ma okres τ i spełnia warunki (8.52). Sygnał wejściowy maszyny przedstawiony jest przez N próbek f_m i jeżeli τ jest duże, to mamy

$$N = 2\nu_M \tau \tag{19.2}$$

(wzór (8.51)). Z próbek tych możemy obliczyć współczynniki Fouriera i odwrotnie (wzory (8.63) i (8.63a)):

$$A_n = \frac{1}{N} \sum_{m=1}^{N} f_m e^{-inm\omega_0\theta}, \qquad \omega_0 = \frac{2\pi}{\tau}, \tag{19.3}$$

$$f_m = \sum_{n=-n_M}^{+n_M} A_n e^{inm\omega_0\theta}. \tag{19.4}$$

Maszyna przetwarza próbki $f(m\theta)$ i wydaje wielkości wyjściowe ψ_μ:

$$\psi_\mu = F_\mu(f_1, \ldots, f_m, \ldots, f_N). \tag{19.5}$$

Ogólnie mówiąc, funkcje F_μ przedstawiające program obliczeń maszyny mogą być bardzo złożone. Jeżeli jednak ograniczymy rozważanie do zagadnień

liniowych, to operację dokonywaną przez maszynę można opisać za pomocą macierzy $(M_{\mu m})$. Wówczas

$$\psi_\mu = \sum_{m=1}^{N} M_{\mu m} f_m \tag{19.6}$$

i macierz $M_{\mu m}$ przedstawia program obliczeń ([19.2]). Mamy N różnych wartości m na wejściu, ale możemy mieć N_1 różnych wartości μ na wyjściu. Może być również pewna rozwlekłość w danych wejściowych i wyjściowych. Jeżeli nie chcemy zmieniać szerokości pasma całego układu, to ciągła funkcja wyjściowa musi mieć czas trwania τ_1:

$$\Psi(t) = \sum_{n=1}^{N_1} \psi_\mu g(t-\mu\theta), \qquad \tau_1 = N_1\theta, \tag{19.7}$$

aby odtworzyć ją na podstawie N_1 danych wejściowych zgodnie z rozdz. 8 (wzór (8.60)). Funkcję $\Psi(t)$ można znów rozłożyć na szereg Fouriera (jak we wzorze (8.58)):

$$\Psi(t) = \sum_{\beta=-\beta_M}^{+\beta_M} \alpha_\beta e^{i\beta\omega_1 t} \quad \text{dla} \quad \left.\begin{aligned} \beta_M \nu_1 &= \nu_M, \\ \nu_1 &= 1/\tau_1, \\ \omega_1 &= 2\pi/\tau_1. \end{aligned}\right\} \tag{19.8}$$

Relacja między współczynnikami Fouriera α_β i próbkami ψ_μ jest bardzo podobna do wzoru (19.3):

$$\alpha_\beta = \frac{1}{N_1} \sum_{\mu=1}^{N_1} \psi_\mu e^{-i\beta\omega_1\theta\mu}. \tag{19.9}$$

Ze wzorów (19.4)÷(19.9) możemy otrzymać bezpośrednie relacje między widmem wejściowym A_n i wyjściowym α_β przyjmując, że obliczenia zlinearyzowano i przedstawiono w postaci macierzy $(M_{\mu m})$:

$$\alpha_\beta = \frac{1}{N_1} \sum_{\mu=1}^{N_1} \sum_{m=1}^{N} \sum_{n=-n_M}^{n_M} M_{\mu m} A_n e^{i(nm\omega_0 - \beta\mu\omega_1)\theta}. \tag{19.10}$$

Jeżeli przyjmiemy, że N_1 danych wyjściowych równa się liczbie N danych wejściowych, a więc $N_1 = N$, to wzór (19.10) znacznie się upraszcza. Macierz

M jest kwadratowa, przy czym

$$\left.\begin{aligned} &\tau_1 = \tau, \quad \omega_1 = \omega_0, \quad n_M = \beta_M, \\ &\alpha_\beta = \frac{1}{N}\sum_{\mu=1}^{N}\sum_{m=1}^{N}\sum_{n=-n_M}^{n_M} M_{\mu m} A_n \, e^{i(nm-\beta\mu)\omega_0\Theta}. \end{aligned}\right\} \tag{19.11}$$

dla $-n_M \leqslant \beta \leqslant n_M$.

Wielkość $(n_m - \beta\mu)$ jest liczbą całkowitą dodatnią lub ujemną i prążki widma α_β sygnału wyjściowego są funkcją wymierną argumentu:

$$z = e^{-i\omega_0\theta} = e^{-i\frac{\pi}{\tau\nu_M}}. \tag{19.12}$$

Wzory te całkowicie podtrzymują rozwinięty w rozdz. 19.1 punkt widzenia: maszyna przetwarza informację, lecz nie wytwarza żadnej nowej informacji. Operację można odwrócić, gdy macierz nie jest osobliwa i może być odwrócona. Macierz odwrotną J można znaleźć z równania:

$$(M)(J) = (J)(M) = 1. \tag{19.13}$$

Oznaczmy przez $J_{m,\mu}$ elementy macierzy odwrotnej, która określa program odwrotnej operacji maszyny. Warunki (19.13) można zapisać w postaci:

$$\left.\begin{aligned} &\sum_{m=1}^{N} M_{\mu m} J_{m\mu'} = \delta\mu\mu' = \begin{cases} 1 \text{ dla } \mu = \mu', \\ 0 \text{ dla } \mu \neq \mu', \end{cases} \\ &\sum_{\mu=1}^{N} J_{m'\mu} M_{\mu m} = \delta_{mm'}. \end{aligned}\right\} \tag{19.14}$$

Operację odwrotną redukowaną przez maszyny przedstawić możemy relacją podobną do (19.11):

$$A_{n'} = \frac{1}{N}\sum_{m'=1}^{N}\sum_{\mu'=1}^{N}\sum_{\beta'=-n_M}^{n_M} J_{m'\mu'}\alpha_{\beta'} \, e^{i(\beta'\mu'-n'm')\omega_0\theta}. \tag{19.15}$$

Dowód na to jest łatwy i wykorzystuje się przy nim wzór (19.14) razem z (8.63b):

$$\frac{1}{N}\sum_{m=1}^{N} e^{ipm\omega_0\theta} = \delta_{p0} \begin{cases} 1 \text{ dla } p = 0, \\ 0 \text{ dla } p \neq 0 \text{ całkowite.} \end{cases} \tag{19.16}$$

Podstawiając wzór (19.11) do wzoru (19.15) otrzymujemy (dla $\beta = \beta'$):

$$A_{n'} = \frac{1}{N^2} \sum_{m',\mu',\beta',\mu,m,n} J_{m'\mu'} M_{\mu m} A_n \, e^{i(nm - m'n')\omega_0\theta} \, e^{i\beta'(\mu' - \mu)\omega_0\theta}. \quad (19.17)$$

Sumowanie względem β w ostatnim wykładniku i uwzględnienie wzoru (19.16) daje $\delta_{\mu'-\mu,0}$, skąd $\mu' = \mu$. Obliczamy wówczas sumę $\sum_{\mu} J_{m'\mu} M_{\mu m}$ i otrzymujemy $m' = m$ zgodnie ze wzorem (19.14). Ostatnim krokiem jest sumowanie składników $e^{i(n-n')m\omega_0\theta}$ względem m, a to daje zgodnie ze wzorem (19.16) $n = n'$, tak że ostatecznie otrzymujemy

$$A_{n'} = A_n,$$

c.b.d.o.

Otrzymaliśmy pełny opis maszyny pracującej według odwracalnego programu liniowego, gdzie nie ma straty informacji, ale i nie powstaje nowa informacja. Z pewnością wystąpi mała strata informacji spowodowana błędami przybliżeń (zaokrąglenia) w obliczeniach. Obliczanie zgodne z programem liniowym przedstawione przez macierz prostokątną M prowadzi do straty informacji, jeżeli $N_1 < N$, i wytwarza rozwlekłość, jeżeli $N_1 > N$. Programy nieliniowe typu (19.5) byłoby trudniej zanalizować.

W rozdziale 19.3 rozważymy nieco odmienne zagadnienie, które odpowiada innej możliwości użycia maszyny cyfrowej.

19.3. MASZYNA MATEMATYCZNA JAKO ELEMENT OBWODU. PRÓBKOWANIE I PROCES ODWROTNY DO PRÓBKOWANIA (SALZER I LINVILL)

Maszyna matematyczna może być włączona w układ sterowania i takie połączenie ma wiele ważnych zastosowań. Zagadnienie to zostało bardzo dokładnie zbadane przez Linvilla* i Salzera**. Dane pochodzą tu z obserwacji pewnego mechanizmu, a wyniki obliczeń wykorzystywane są do sterowania tego mechanizmu.

---

* W. K. Linvill, *Doctor thesis*, *MIT*, Cambridge, Mass. 1950; Proc. I.R.E. **40** (1952), str. 230.

** J. M. Salzer, *Doctor thesis*, *MIT*, Cambridge, Mass. 1951; Proc. I.R.E. **40** (1952), str. 231; **41** (1953), str. 901; **42** (1954), str. 457; I.R.E. Electonics Group, University of California, Los Angeles, May 15, 1952; również Hughes Aircraft Co., Memo. No. 338, 1954.

Maszyna jest częścią układu sterowania i powinna być analizowana jako element układu. Cały układ sterowania pracuje z sygnałami ciągłymi o skończonej szerokości pasma, a maszyna wykorzystuje próbki. W porównaniu z rozważanym poprzednio zagadnieniem jest tu następująca różnica: w rozdz. 19.2 przyjęliśmy, że maszyna miała do dyspozycji całkowity zbiór danych odpowiadających zakończonemu doświadczeniu o czasie trwania τ, oraz że opracowaliśmy dane po całkowitym zakończeniu doświadczenia. W urządzeniu sterującym musimy dokonywać obliczeń na bieżąco, wykorzystując tylko dane dostępne do chwili bieżącej t, aby móc sterować pracą urządzenia tak szybko, jak to możliwe. W takim zagadnieniu całkowity czas trwania doświadczenia nie jest określony z góry. Metoda z rozdz. 19.2 musi być zastąpiona przez metodę próbkowania Shannona (rozdz. 8.7).

W celu wykorzystania całkowitej informacji zawartej w pasmie przepuszczania, bez wprowadzania zbytecznej rozwlekłości, próbkujemy znów z okresem θ (jak we wzorze (19.1)). Pierwotna funkcja może być odtworzona według procedury podanej przez Shannona (zgodnie ze wzorem (8.69)) za pomocą funkcji impulsowych:

$$g_s(t) = \frac{\sin 2\pi\nu_M t}{2\pi\nu_M t} = \frac{1}{2\nu_M} \int_{\nu=-\nu_M}^{\nu_M} e^{i2\pi\nu t}\, d\nu \tag{19.18}$$

oraz

$$f(t) = \sum_m f_m g_s(t - m\theta) = \int_{\nu=-\nu_M}^{\nu_M} A(\nu)\, e^{i2\pi\nu t}\, d\nu. \tag{19.19}$$

Funkcja impulsowa Shannona g_s sięga na pewną odległość w przód i wstecz. Oznacza to, że nie można otrzymać wartości funkcji $f(t)$ w chwili t, jeżeli mamy tylko dostęp do wartości przeszłych ($m<0$).

Linvill wykorzystuje inną funkcję impulsową $g_L(t)$, będącą bardzo ostrym impulsem, jak pokazano na rys. 19.1. Taki impuls możemy opisać wzorem podobnym do (19.18):

$$g_L(t) = \frac{\sin 2\pi\nu_1 t}{2\pi\nu_1 t} = \frac{1}{2\nu_1} \int_{-\nu_1}^{\nu_1} e^{i2\pi\nu' t} d\nu' \tag{19.20}$$

z tym, że $\nu_1 \gg \nu_M$. Funkcję tak otrzymaną oznaczymy przez $\tilde{f}(t)$:

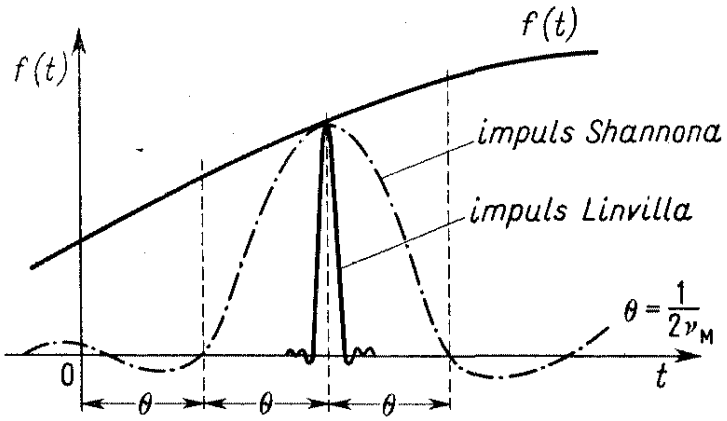
$$\tilde{f}(t) = \sum_m f_m g_L(t-m\theta) = \int_{-\nu_1}^{\nu_1} \tilde{A}(\nu')\, e^{i2\pi\nu' t}\, d\nu'. \tag{19.21}$$

Ze wzorów (19.18) i (19.19) obliczamy

$$A(\nu) = \frac{1}{2\nu_M} \sum_m f_m e^{-i2\pi\nu m\theta}, \quad -\nu_M \leqslant \nu \leqslant \nu_M, \tag{19.22}$$

podczas gdy wzory (19.20) i (19.21) dają

$$\tilde{A}(\nu') = \frac{1}{2\nu_1} \sum_m f_m e^{-i2\pi\nu' m\theta} - \nu_1 \leqslant \nu' \leqslant \nu_1. \tag{19.23}$$



Rys. 19.1. Funkcje impulsowe Shannona i Linvilla odtwarzające funkcję $f(t)$

Jeżeli porównamy wzory (19.22) i (19.23), zauważymy, że widmo $\tilde{A}(\nu')$ jest identyczne z $A(\nu)$ dla małych częstotliwości i odtwarza to widmo periodycznie w wyższych pasmach częstotliwości:

$$-\nu_M \leqslant \nu' \leqslant \nu_M: \quad \tilde{A}(\nu') = \frac{\nu_M}{\nu_1} A(\nu). \tag{19.24}$$

Jeżeli np. weźmiemy $\nu_M \leqslant \nu \leqslant 3\nu_M$ oraz podstawimy $\nu' = \nu + 2\nu_M$, to

$$\tilde{A}(\nu') = \frac{1}{2\nu_1} \sum_m f_m e^{-i2\pi\nu m\theta - i2\pi m} = \frac{\nu_M}{\nu_1} A(\nu), \tag{19.25}$$

ponieważ $2\nu_M\theta = 1$. W ogólności, jeżeli

$$(2k-1)\nu_M \leqslant \nu \leqslant (2k+1)\nu_M,$$

to bierzemy $\nu' = \nu + 2k\nu_M$ i otrzymujemy

$$\tilde{A}(\nu') = \frac{\nu_M}{\nu_1} A(\nu). \tag{19.26}$$

Z rezultatów tych wynika, że

$$\frac{\nu_1}{\nu_M} \tilde{A}(\nu') = \sum_{k=-\infty}^{\infty} A(\nu' - 2k\nu_M). \tag{19.27}$$

Metoda próbkowania Linvilla daje funkcję $f(t)$, którą można łatwo obliczyć z wartości przeszłych, gdyż jej impulsy są wyraźnie zlokalizowane. Różni się ona od pierwotnej funkcji dodatkowymi składowymi widma wyższego rzędu. Proces odwrotny do próbkowania sprowadza się po prostu do filtracji sygnału, która zachowuje tylko składowe widma w pasmie $|\nu| \leqslant \nu_M$. Operacja ta prowadzi zawsze do funkcji pierwotnej $f(t)$ z dokładnością do stałego współczynnika $\frac{\nu_M}{\nu_1}$.

19.4. OBLICZENIA NA PODSTAWIE PRÓBEK DANYCH W CHWILI t

Obliczenia dokonuje się w oparciu o dane znane w chwili t. Danymi wejściowymi są próbki

$$f_m = f(m\theta), \qquad m \leqslant m_t, \qquad m_t\theta = t. \tag{19.28}$$

Danymi wyjściowymi, dostępnymi w tym samym czasie, są próbki

$$\psi_\mu = \psi(\mu\theta), \qquad \mu \leqslant m_t - 1. \tag{19.29}$$

Jeżeli obliczenia dokonuje się według programu liniowego, możemy napisać relację (Salzer):

$$\psi_{mt} + \sum_{k=1}^{k_1} b_k(m_t)\psi_{m_t-k} = \sum_{h=0}^{h_1} a_h(m_t) f_{m_t-h}, \tag{19.30}$$

gdzie wykorzystujemy k_1 poprzednio obliczonych wartości wyjściowych oraz h_1+1 ostatnich danych wejściowych. Współczynniki b_k oraz a_h mogą zależeć od czasu (poprzez m_t). Ważny przypadek występuje dla *liniowego, czasowo-inwariantnego* programu obliczeń, ze stałymi b_k oraz a_k. Zamieniając

wartości wyjściowe ψ_{m_t-k} na kombinacje liniowe danych wejściowych, możemy zapisać

$$\psi_{m_t} = \sum_{h=0}^{h_2} c_h f_{m_t-h}. \tag{19.31}$$

Liczba składników h_2 zawartych w sumie jest duża i może być nieskończenie wielka, dając szereg nieskończony. Wzór ten można porównać z wynikami z rozdz. 19.2. Bierzemy:

$$m_t = \mu, \quad m_t - h = m, \quad h = \mu - m, \tag{19.32}$$

$$\psi_\mu = \sum_{m=\mu}^{\mu-h_2} c_{\mu-m} f_m. \tag{19.33}$$

Porównując ze wzorem (19.6) otrzymujemy macierz przekształcenia:

$$M_{\mu m} = \begin{cases} 0 & \text{dla} \quad m > \mu, \\ c_{\mu-m} & \text{dla} \quad \mu - h_2 \leqslant m \leqslant \mu, \\ 0 & \text{dla} \quad m < \mu - h_2. \end{cases} \tag{19.34}$$

Oto taka macierz dla $h_2 = 3$:

	0	1	2	3	4	5	6	7	$\rightarrow \mu$
0	C_0	C_1	C_2	C_3	0	0	0	0	
1	0	C_0	C_1	C_2	C_3	0	0	0	
2	0	0	C_0	C_1	C_2	C_3	0	0	
3	0	0	0	C_0	C_1	C_2	C_3	0	
4	0	0	0	0	C_0	C_1	C_2	C_3	
$\downarrow$									
m									

(19.35)

Wszystkie elementy macierzy leżące poniżej głównej przekątnej są zerami. Wszystkie wiersze mają takie same stałe współczynniki, przesunięte po prostu o μ pozycji w prawo. Ta bardzo specjalna macierz daje prostsze wyniki niż bardziej ogólna macierz z rozdz. 19.2.

19.5. FUNKCJA PRZENOSZENIA MASZYNY MATEMATYCZNEJ

Zbadajmy teraz zależność między widmem wejściowym i wyjściowym. Wyrażamy f oraz ψ w postaci całki Fouriera, podobnie jak we wzorze (19.19), i otrzymujemy

$$f_m = \int_{-\nu_M}^{\nu_M} A(\nu)\, e^{i2\pi\nu t}\, d\nu, \qquad t = m\theta, \tag{19.36}$$

$$\psi_m = \int_{-\nu_M}^{\nu_M} \alpha(\nu)\, e^{i2\pi\nu t}\, d\nu. \tag{19.37}$$

Podstawiamy teraz te wyrażenia do wzoru (19.30) i po obu stronach otrzymujemy całki względem $e^{2\pi i\nu t}\, d\nu$. Po lewej stronie występuje współczynnik

$$\alpha(\nu) + \sum_{k=1}^{k_1} b_k \alpha(\nu)\, e^{-i2\pi\nu k\theta}, \tag{19.38}$$

podczas gdy po prawej stronie

$$\sum_{h=0}^{h_1} a_h A(\nu)\, e^{-i2\pi\nu h\theta}. \tag{19.38a}$$

Obie funkcje (19.30), będąc równe, mają również równe transformaty Fouriera:

$$\alpha(\nu) + \sum_{k} b_k \alpha(\nu)\, e^{-2\pi i\nu k\theta} = \sum_{h} a_h A(\nu)\, e^{-i2\pi\nu h\theta}. \tag{19.39}$$

Z relacji tych otrzymujemy funkcję przenoszenia (19.3):

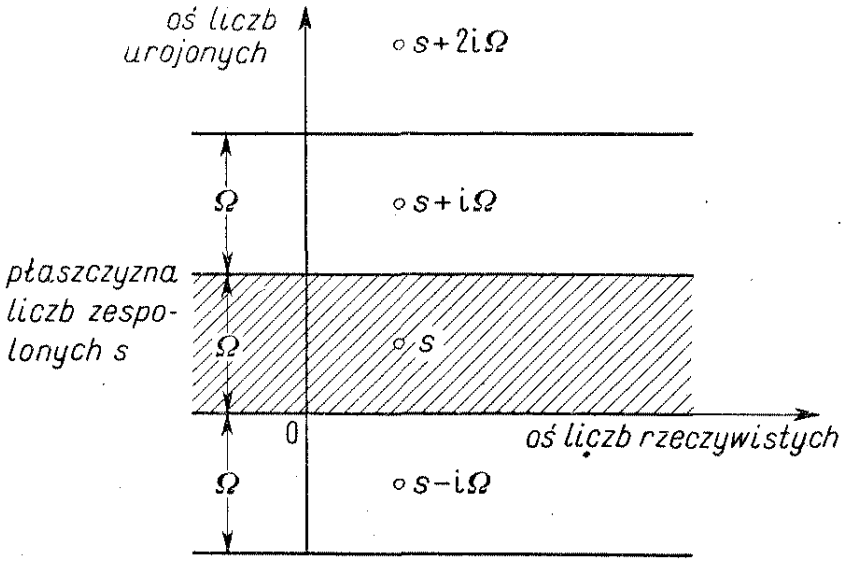
$$W(\nu) = \frac{\alpha(\nu)}{A(\nu)} = \frac{\sum_{h}^{h_1} a_h\, e^{-i2\pi\nu h\theta}}{1 + \sum_{k=1}^{k_1} b_k\, e^{-i2\pi\nu k\theta}} = \frac{a_0 + a_1 z + a_2 z^2 + \ldots + a_{h_1} z^{h_1}}{1 + b_1 z + b_2 z^2 + \ldots + b_{k_1} z^{k_1}}, \tag{19.40}$$

która dla rozważanej maszyny matematycznej jest stosunkiem widma wyjściowego do widma wejściowego. Jest ona funkcją wymierną względem zmiennej

$$z = e^{-i2\pi\nu\theta}, \text{ gdzie } \theta = \text{okresowi próbkowania.} \tag{19.41}$$

Godny uwagi jest fakt, że tego rodzaju obliczenie, jakie obrazuje równanie (19.30), można analizować w taki właśnie sposób i każdej częstotliwości

wyjściowej odpowiada widmo $\alpha(\nu)$ zależne tylko od widma wejściowego $A(\nu)$ związanego z samą częstotliwością. Funkcja przenoszenia postaci (19.40) została otrzymana przez Salzera i umożliwiła mu rozwiązanie różnorodnych zagadnień, w których maszyny matematyczne i zwykłe obwody były połączone w skomplikowany układ sterowania. Równie dobrze możemy wyjść



Rys. 19.2. Poziome pasma płaszczyzny zespolonej s, w których powtarzają się wartości $\tilde{A}(s)$

ze wzoru (19.31), gdzie przyjęliśmy, że rozwiązanie poprzedniego układu zostało otrzymane w celu wyeliminowania ψ_{m-k}. Relacja ta daje, tak jak przedtem:

$$W(\nu) = \frac{\alpha(\nu)}{A(\nu)} = \sum_{h=0}^{h_2} c_n \, e^{-i2\pi\nu h\theta}, \tag{19.42}$$

tj. wielomian lub nieskończony szereg względem z.

Wszystkie te rozważania zgadzają się z poglądem wyrażonym uprzednio, kiedy omawialiśmy rolę maszyny matematycznej. Tak więc maszyna matematyczna przedstawia automatyczne urządzenie do przetwarzania informacji i porównać ją można do narzędzia obrabiającego kawałek metalu. Nie wytwarza ona nowej informacji, a na wyjściu maszyny pojawia się dokładnie ta sama ilość informacji, jaką podano na wejście (w idealnym przypadku nie ma strat), tylko inaczej zakodowana.

Nasze rozważania oparliśmy na metodzie Fouriera, gdyż Linvill i Salzer stosowali transformaty Laplace'a. Obie procedury są równoważne. Wzory

Salzera można po prostu otrzymać przez zwykłą zamianę ([19.3]) $i\omega$ na s:

$$s = i\omega, \quad \omega = -is, \quad z = e^{-i\omega\theta} = e^{-s\theta}. \tag{19.43}$$

Dowiedliśmy, że ciąg próbek funkcji $f(t)$ ma widmo periodyczne $\tilde{A}(\nu)$ (rozdz. 19.3):

$$\left.\begin{aligned} &\tilde{A}(\omega+n\Omega) = \tilde{A}\omega, \\ &\text{gdzie} \\ &n \text{ — całkowite} \quad \text{oraz} \quad \Omega = 4\pi\nu_M = \frac{2\pi}{\theta} \text{ — rzeczywiste.} \end{aligned}\right\} \tag{19.44}$$

Własność tę łatwo przetransponować na zmienną s i można zapisać

$$\tilde{A}(s - in\Omega) = \tilde{A}\left(s - 2\pi i\,\frac{n}{\theta}\right) = \tilde{A}(s). \tag{19.45}$$

Na płaszczyźnie zespolonej s funkcja A przyjmuje te same wartości w kolejnych pasmach, jak to pokazano na rys. 19.2. Gdy funkcja jest znana w jednym z tych pasm, np. w pasmie zakreskowanym, to może być już ona określona na całej płaszczyźnie, gdyż powtarza się periodycznie w innych pasmach.

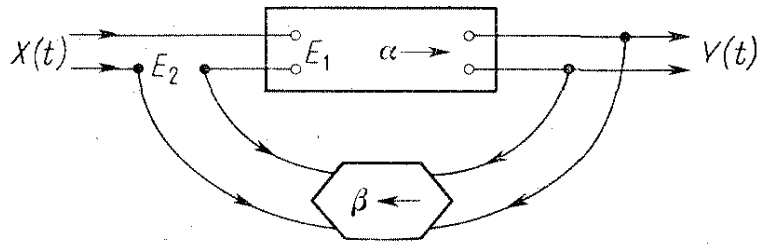
19.6. UKŁADY ZAWIERAJĄCE MASZYNĘ MATEMATYCZNĄ. ZAGADNIENIE STABILNOŚCI

Poprzednia analiza wykazała, że maszyna pracująca według danego programu liniowego może być rozważana jako przetwornik o pewnej funkcji przenoszenia $W(\nu)$, która zależy od częstotliwości ν. Gdy maszyna włączona jest w szereg z konwencjonalnym układem, to resztę układu charakteryzuje inna funkcja przenoszenia, $W_1(\nu)$, i cały układ ma wtedy funkcję przenoszenia

$$W_c(\nu) = W(\nu)\, W_1(\nu). \tag{19.46}$$

Uwaga ta umożliwia rozważanie zagadnienia układu z maszyną według dobrze znanych metod analizy obwodów. Charakterystyczne jest to, że funkcja przenoszenia $W(\nu)$ maszyny jest wymierna względem zmiennej z we wzorze (19.41), podczas gdy układy konwencjonalne dają funkcję przenoszenia wymierną względem ω.

Jednym z ważnych punktów dyskusji jest zagadnienie stabilności. Maszyna może przyjąć funkcję $f(t)$ i pomnożyć ją przez współczynnik C; pracuje więc jako wzmacniacz. Dalej, w ogólnym typie programu omawianym przez Salzera (por. rozdz. 19.4, równanie (19.30)) wykorzystuje się pewną ilość wcześniej obliczonych wartości wyjściowych do obliczania wartości wyjściowej w chwili t. Oznacza to, że program wymaga układu sprzężenia zwrotnego. Mając maszynę, która może pracować jako wzmacniacz ze sprzężeniem zwrot-



Rys. 19.3. Wzmacniacz o wzmocnieniu α i sprzężeniu zwrotnym β

nym, na pewno musimy rozważyć zagadnienie stabilności, gdyż wzmacniacz ze sprzężeniem zwrotnym może wytwarzać ciągłe oscylacje, gdy sprzężenie jest dostatecznie duże.

Zagadnienie to ma fundamentalne znaczenie i omówimy je wprowadzając do rozważań zwykły wzmacniacz o wzmocnieniu α i sprzężeniu zwrotnym β, tak jak pokazano na rys. 19.3.

Napięcie E_1 daje na wyjściu $G(t)$:

$$G = \alpha E_1, \tag{19.47}$$

gdzie α jest zespolone.

Część sygnału wyjściowego β podawana jest znowu na wzmacniacz i dodaje się do początkowej wartości sygnału wejściowego $I(t)$, tak że

$$\left.\begin{aligned} E_2 &= \beta G = \alpha\beta E_1, \\ E_1 &= I + E_2 = I + \alpha\beta E_1, \end{aligned}\right\} \tag{19.48}$$

gdzie β — wielkość zespolona, oraz

i z zależności tych otrzymujemy

$$\left.\begin{aligned} E_1 &= I/(1-\alpha\beta), \\ G &= I\alpha/(1-\alpha\beta). \end{aligned}\right\} \tag{19.49}$$

Stąd znajdujemy funkcję przenoszenia wzmacniacza ze sprzężeniem zwrotnym

$$W_a = G/I = \alpha/(1-\alpha\beta). \tag{19.50}$$

Wyrażenie to jest bardzo podobne do (19.40), otrzymanego dla maszyny matematycznej, jeżeli założymy, że

$$\alpha(\nu) = \sum_h a_h z^h, \qquad z = \mathrm{e}^{-\mathrm{i}2\pi\nu\theta}, \qquad -\alpha\beta = \sum_k b_k z^k. \tag{19.51}$$

Funkcja dana (19.50) ma pewną ilość biegunów, jeżeli

$$\alpha\beta = 1. \tag{19.52}$$

Warunek (19.55) może być spełniony dla określonych częstotliwości. Przy tych częstotliwościach dla $I = 0$ sygnał wyjściowy może mieć skończoną wartość, gdyż W jest nieskończone i układ jest samowzbudny. Rozkład biegunów na płaszczyźnie $s = \delta + \mathrm{i}\omega$ ma kapitalne znaczenie i zostanie omówiony w rozdz. 19.7.

Podsumowując to, co powiedziano wyżej, stwierdzamy [19.4]:

(1) Maszyna matematyczna jest podobna do układu aktywnego.

(2) Stabilność programu obliczeń można badać tymi samymi metodami co stabilność układów aktywnych.

(3) Można badać charakterystyki amplitudowo-fazowe danego programu liniowego, odpowiedź impulsową, wykresy Nyquista, położenie biegunów itp.

(4) Szerokość pasma zajmowanego przez maszynę jest określona bezpośrednio okresem próbkowania θ (wzór (19.1)), przy czym maksymalna częstotliwość

$$\nu_M = \frac{1}{2\theta}. \tag{19.53}$$

Salzer na przykładach liczbowych wykazał efektywność rozpatrywanej metody. Wyjaśnił również, jak dobierać obwody do maszyny lub odwrotnie. Zagadnienia związane z syntezą obwodu podobne są do zagadnień omawianych w zwykłej teorii obwodów.

19.7. STABILNOŚĆ PROGRAMU

Maszyna wykonująca operacje według danego liniowego programu opisana jest przez funkcję przenoszenia $W(s)$ i odpowiedź jest proporcjonalna do pobudzenia $A(s)$ pomnożonego przez W, jak to wykazano we wzorze (19.40):

$$\alpha(s) = A(s)\, W(s). \tag{19.54}$$

Tak długo dopóki moduł $|W|$, mający sens wzmocnienia, pozostaje skończony, odpowiedź jest skończona. Jeżeli $|W|$ staje się nieskończony, możemy otrzymać sygnał na wyjściu nawet wtedy, gdy nie ma pobudzenia na wejściu. Występuje to przy biegunach funkcji $W(s)$. Niech

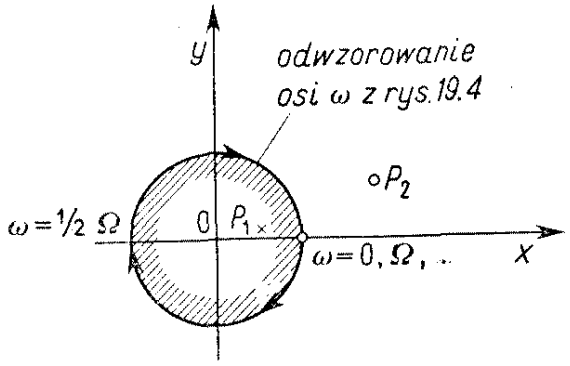
$$s = \sigma + i\omega \tag{19.55}$$

będzie takim biegunem. Odpowiedź ma postać

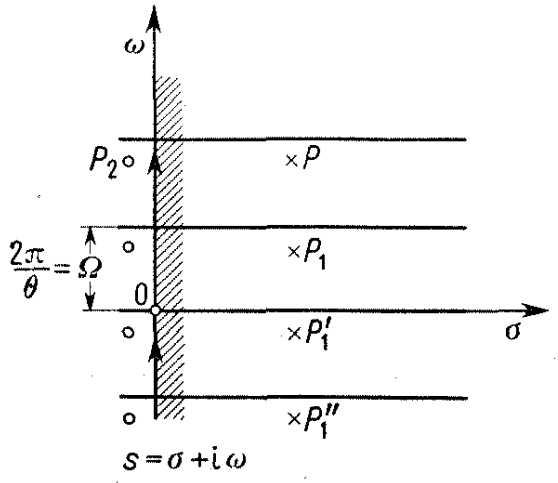
$$\alpha(s)\, e^{st} = \alpha(s)\, e^{\sigma t + i\omega t}. \tag{19.56}$$

Jeżeli σ jest dodatnie lub równe zeru, drgania narastają lub są tylko podtrzymywane. Gdy σ jest ujemne, powstające drgania są tłumione. Mogą one powstawać przejściowo, ale będą jednak zanikać.

Niestabilność odpowiada biegunom funkcji przenoszenia o dodatniej części rzeczywistej. Bieguny te leżą po prawej stronie osi urojonej zmiennej s.



Rys. 19.4. Na płaszczyźnie zespolonej s biegun P_1 leżący po prawej stronie osi urojonej odpowiada niestabilności układu. Biegun P_2 po lewej stronie nie powoduje niestabilności. W kolejnych pasmach bieguny stabilne zaznaczono kropkami, natomiast bieguny stabilne krzyżykami



Rys. 19.5. Odwzorowanie konforemne $z = e^{-s\theta}$ płaszczyzny zespolonej s na płaszczyznę zespoloną z. Oś urojona płaszczyzny s przechodzi w okrąg koła jednostkowego na płaszczyźnie z. Bieguny niestabilne P_1 z rys. 19.4 przechodzą do wnętrza koła jednostkowego, natomiast bieguny stabilne leżą na zewnątrz

Z powodu okresowości dany biegun P_1 pojawia się w każdym pasmie o szerokości Ω. Biegun P_2 po lewej stronie osi urojonej nie wprowadza niestabilności.

Rysunek 19.4 ilustruje ogólną sytuację: bieguny o ujemnych σ (stabilne) zaznaczone są kropkami, podczas gdy te, które powodują niestabilność, zaznaczono krzyżykami.

Odbicia P_1 w kolejnych pasmach o wysokości Ω zaznaczone są jako $P_1', P_1'', \ldots$ Zamiast płaszczyzny s możemy rozważać płaszczyznę zespoloną z pokazaną na rys. 19.5. Wzór

$$z = x + iy = e^{-s\theta} \tag{19.57}$$

przedstawia odwzorowanie konforemne płaszczyzny s w płaszczyznę z. Początek układu współrzędnych 0 na płaszczyźnie z oznacza $z = 0$ i odpowiada $s = +\infty$. Okrąg o promieniu 1 na płaszczyźnie z odpowiada osi urojonej na płaszczyźnie s:

$$s = i\omega, \qquad z = e^{-i\omega\theta},$$

a wnętrze koła jednostkowego z odpowiada prawej półpłaszczyźnie s. Biegun P_1 (i jego obrazy $P_1', P_1'', \ldots$) z prawej półpłaszczyzny s odpowiada biegunom P_1 w kole jednostkowym na półpłaszczyźnie z. Biegun P_2 z lewej półpłaszczyzny s jest przedstawiony przez punkt P_2 na zewnątrz koła jednostkowego na płaszczyźnie z.

Zagadnienie rozmieszczenia biegunów funkcji przenoszenia można teraz sprecyzować bardziej dokładnie. Naszą funkcję przenoszenia W ze wzoru (19.40) można zapisać, jak następuje:

$$W = \frac{P(s)}{Q(s)} = \frac{\sum a_h z^h}{1 + \sum b_k z^k}. \tag{19.58}$$

Wyeliminowano jednak wspólne dla P oraz Q pierwiastki, natomiast bieguny W odpowiadają pozostałym pierwiastkom równania:

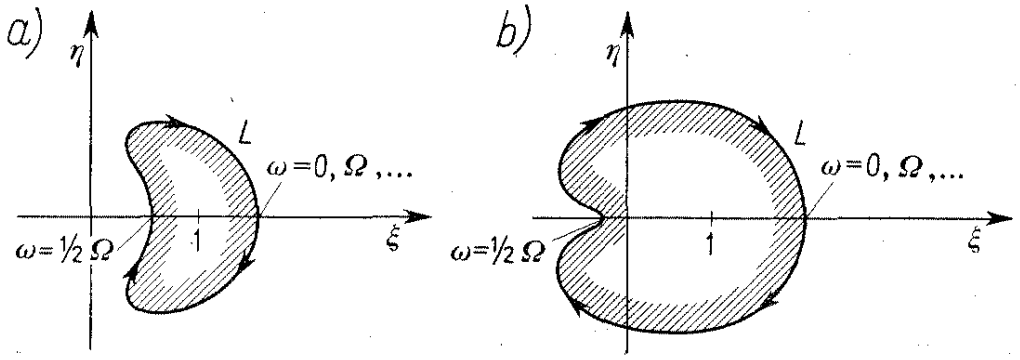
$$Q = 1 + \sum b_k z^k = 0. \tag{19.59}$$

Można rozważyć inne odwzorowanie konforemne, np. z płaszczyzny zespolonej z w płaszczyznę zespoloną Q, jak pokazano na rys. 19.6*. Początek układu $Q = 0$ z płaszczyzny Q odpowiada wszystkim pierwiastkom funkcji Q (wzór (19.59)), a więc przedstawia on wszystkie bieguny funkcji przenoszenia W. Punkt $Q = 1$ na płaszczyźnie Q odpowiada początkowi układu $z = 0$ na płaszczyźnie z. Koło jednostkowe z płaszczyzny z przechodzi w krzywą L (miejsce geometryczne punktów) na płaszczyźnie Q, a wnętrze koła jednostkowego przechodzi we wnętrze krzywej L.

* Są to tzw. wykresy Nyquista. Metody takie szeroko stosuje się w teorii automatycznego sterowania. (*Przyp. tłum.*).

Jeżeli biegun P_1 leży wewnątrz koła jednostkowego na płaszczyźnie z, to początek układu z płaszczyzny Q leży wewnątrz obszaru objętego krzywą L, jak pokazano na rys. 19.6b. Sytuacja taka odpowiada niestabilności. Rysunek 19.6a pokazuje krzywą L z początkiem układu współrzędnych na zewnątrz tej krzywej; jest to przypadek układu stabilnego. ([19.5])

Salzer rozważał różne przykłady i wykazał dużą przydatność praktyczną powyższej metody. Bardzo często orientacyjny szkic krzywej L wystarcza



Rys. 19.6. Płaszczyzna zespolona Q będąca odwzorowaniem konforemnym:
$$Q = \xi + i\eta = 1 + \sum_k b_k k^k$$

a) Przykład, w którym początek układu współrzędnych płaszczyzny zespolonej Q leży na zewnątrz trajektorii L. Jest to przypadek stabilności; b) Przykład, w którym początek układu współrzędnych płaszczyzny zespolonej Q leży wewnątrz trajektorii L. Jest to przypadek niestabilności

do zbadania stabilności. Wartość, o jaką trajektoria L omija początek układu, wskazuje na względną stabilność układu. Częstotliwość odpowiadająca największemu zbliżeniu się trajektorii do początku układu odpowiada najczęściej częstotliwości rezonansowej programu obliczeń.

Funkcja przenoszenia programu obliczeń W zawsze dana jest w postaci funkcji (19.40) lub (19.58). Jeżeli znamy funkcję przenoszenia operacji, to możliwości realizacji występują wtedy, gdy funkcję przenoszenia można zapisać w postaci (19.40) lub (19.58). Jest to podstawowy i zarazem bardzo ważny wynik rozważań Salzera.

19.8. PRZYKŁADY FUNKCJI PRZENOSZENIA

Przytoczymy kilka przykładów funkcji przenoszenia, odpowiadających typowym zagadnieniom obliczeniowym.

19.8.1. Maszyna matematyczna spełniająca rolę integratora

W tym przypadku ψ określamy jako całkę z sygnału wejściowego $f(t)$ tak, że możemy zapisać

$$\psi_m = \psi_{m-1} + \int_{(m-1)\theta}^{m\theta} f(t)\,\mathrm{d}t = \psi_{m-n} + \int_{(m-n)\theta}^{m\theta} f(t)\,\mathrm{d}t = \int_0^{m\theta} f(t)\,\mathrm{d}t. \quad (19.60)$$

Ostatnie wyrażenia wynikają przez iterację równości między pierwszymi dwoma wyrażeniami. Funkcja przenoszenia dana jest wzorem (19.40) lub (19.58):

$$W = \frac{\sum a_n z^n}{1 + \sum b_n z^n}, \quad (19.61)$$

a zagadnienie sprowadza się do znalezienia b_n oraz a_n.

Aby to zrobić, wykorzystujemy metodę przybliżoną, sprowadzając całkę (19.60) razem z (19.30) do postaci

$$\psi_m + \sum_n b_n \psi_{m-n} = \sum_n a_n f_{m-n}, \quad (19.62)$$

gdzie

$$f_k = f(k\theta).$$

Podamy wyniki obliczenia całki przy pomocy funkcji schodkowej, metodą trapezów, oraz w oparciu o regułę 1/3 i regułę 3/8 Simpsona.

Funkcja schodkowa

Obliczamy wartość przybliżoną całki (19.60), zastępując funkcję $f(t)$ funkcją schodkową oraz obliczając pole otrzymanych prostokącików o szerokości θ i wysokościach $f_1, f_2, \ldots, f_m$:

$$\psi_m - \psi_{m-1} = \int_{(m-1)\theta}^{m\theta} f(t)\,\mathrm{d}t = f_m \theta.$$

Porównując to ze wzorem (19.62) mamy

$$b_1 = -1, \quad a_0 = \theta, \quad b_j = a_k = 0 \quad \text{dla} \quad j > 1, \quad k > 0,$$

a stąd ze wzoru (19.61) wynika, że funkcja przenoszenia ma postać

$$W_0 = \frac{\theta}{1 - z}. \quad (19.63)$$

Metoda trapezów

Wykorzystanie metody trapezów w wyrażeniu (19.60) dla $n = 1$ daje

$$\psi_m - \psi_{m-1} = \int_{(m-1)\theta}^{m\theta} f(t)\,\mathrm{d}t = \frac{\theta}{2}(f_m + f_{m-1}),$$

co po uwzględnieniu wzoru (19.62) daje

$$b_1 = -1, \quad a_0 = a_1 = \frac{\theta}{2}, \quad b_j = a_k = 0 \quad \text{dla} \quad j>1,\ k>1.$$

W tym przypadku funkcja przenoszenia ma postać

$$W_1 = \frac{\theta}{2}[(1+z)/(1-z)]. \tag{19.64}$$

Reguła 1/3 Simpsona

W regule tej musimy przyjąć $n = 2$ we wzorze (19.60). Zatem

$$\psi_m - \psi_{m-2} = \int_{(m-2)\theta}^{m\theta} f(t)\,\mathrm{d}t = \frac{\theta}{3}(f_{m-2} + f_{m-1} + f_m),$$

co daje

$$b_2 = -1, \quad a_0 = a_2 = \frac{\theta}{3}, \quad a_1 = \frac{4\theta}{3},$$

$$b_j = a_k = 0 \quad \text{dla} \quad j\neq 2,\ k>2,$$

a funkcja przenoszenia ma postać

$$W_2 = \frac{\theta}{3}[(1+4z+z^2)/(1-z^2)]. \tag{19.65}$$

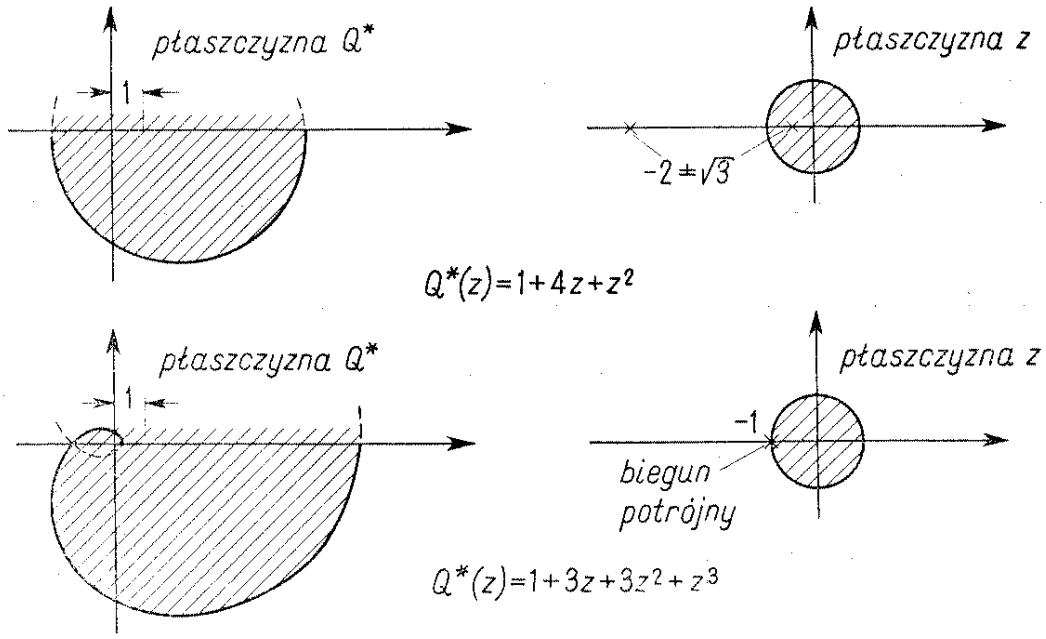
Reguła 3/8 Simpsona

W regule tej przyjmiemy $n = 3$ i otrzymamy w podobny sposób jak poprzednio

$$W_3 = \frac{3\theta}{8}[(1+3z+3z^2+z^3)/(1-z^2)]. \tag{19.66}$$

19.8.2. Maszyna wykonująca operacje różniczkowania

Odpowiedź ψ jest tu pochodną pobudzenia f. Skoro różniczkowanie jest operacją odwrotną do całkowania, to funkcja przenoszenia idealnego układu różniczkującego powinna być odwrotna do funkcji przenoszenia idealnego



Rys. 19.7. Krzywe pokazujące niestabilność programów różniczkowania odpowiadających funkcjom W_2^{-1} oraz W_3^{-1} na płaszczyźnie Q^* oraz z. Tutaj Q^* jest mianownikiem funkcji przenoszenia (licznik W). Obie trajektorie obejmują początek układu współrzędnych płaszczyzny Q^*

integratora. Oznacza to, że W_0^{-1}, W_1^{-1}, W_2^{-1}, W_3^{-1} przedstawiają odpowiednie programy różniczkowania. Okazuje się, że tylko W_0^{-1} i W_1^{-1} są stabilne, podczas gdy W_2^{-1} i W_3^{-1} są niestabilne. Sytuację tę pokazano na rys. 19.7, według artykułu Salzera. Pokazano bieguny na płaszczyźnie z oraz kształt trajektorii na płaszczyźnie Q.

19.8.3. Odtworzenie sygnału ciągłego na podstawie próbek

Odtworzenia sygnału ciągłego z sygnału spróbkowanego dokonać można na drodze ponownego zastąpienia impulsów Linvilla impulsami Shannona lub na drodze dokładnej filtracji wszystkich częstotliwości $\nu > \nu_M$. Pierwszy sposób nie powoduje przesuwów fazowych i nie zmienia natężenia sygnału o $\nu < \nu_M$ i tym samym funkcja przenoszenia równa się 1. Metoda druga wymaga, aby amplituda sygnału dla $\nu > \nu_M$ wynosiła zero, a więc $W = 0$ w tym zakresie częstotliwości. Jest to trudne do realizacji. Większość

filtrów przepuszcza sygnały o dużej amplitudzie w pasmie przepuszczania, ale powoduje przesuwy fazy. W rozważanym przez nas przypadku odtwarzania funkcji faza odgrywa istotną rolę i pod tym względem projektowanie filtru różni się od projektowania filtrów konwencjonalnych.

19.8.4. Opóźnienie czasowe

Jeżeli posłużymy się impulsami Shannona typu $\frac{\sin x}{x}$ przy założeniu, że funkcja jest stała dla dużych x, to powinniśmy obliczyć

$$\sum_{\substack{\text{na duże}\\ \text{wartości } x}} \frac{\sin x}{x} \approx \int_{X}^{\infty} \frac{\sin x}{x}\,dx \approx \frac{\cos x}{x}.$$

Wyrażenie to jest bardzo wolno zbieżne. Jeżeli żądamy, aby błąd w wyniku końcowym był mniejszy od 10^{-n}, to trzeba wziąć $x > 10^n = N$ składowych. Tak duża liczba składowych wpływa na opóźnienie, gdyż funkcja w chwili t nie będzie znana poprawnie aż do składnika N, co oznacza opóźnienie o $N\theta$.

20

INFORMACJA, ORGANIZACJA I INNE ZAGADNIENIA

20.1. INFORMACJA A ORGANIZACJA

Byliśmy w stanie przeanalizować ilość informacji zawartej w zapisanym zdaniu, w zbiorze liczb i w innych sytuacjach, do których stosować można przytoczone przez nas definicje ilości informacji. W analogiczny sposób można obliczać ilość informacji zawartej w światłodrukach, wykazując, że światłodruk można przedstawić za pomocą skończonej liczby symboli. Rozważmy na przykład schemat ideowy jakiegoś urządzenia elektrycznego. Wszystkie zaciski występujące na schemacie można ponumerować od 0 do N i przedstawić ten schemat liczbami wskazującymi połączenie odpowiednich zacisków; np. zespół liczb 1023—0216 oznacza, że trzeba połączyć zacisk 1023 z zaciskiem 0216. Przy skończonej ilości zacisków wystąpi skończona ilość P możliwych połączeń. Ilość informacji, jaką otrzymamy wykorzystując konkretne połączenia, zgodnie z definicją z rozdz. 1, wyniesie

$$I = K \ln P. \tag{20.1}$$

Definicja ta nie mówi nic o różnicy pomiędzy schematem ideowym modelu urządzenia a jakimś innym schematem, który odpowiadałby strukturze mniej wartościowej, np. ze zwarciami lub z innymi wadami. Ilość informacji, jaką obliczamy, jest informacją daną człowiekowi, który ma wykonywać połączenia; informacja ta nie mówi o tym, jaką rolę ma spełniać to urządzenie.

Nie rozważamy pytania, jak i dlaczego został wybrany ten właśnie schemat oraz jak on pracuje. O tym ostatnim myślał naukowiec lub inżynier, którzy projektowali urządzenie. Wymagało to naukowego rozważenia, przemyślenia

i innych procesów związanych z umysłem ludzkim, które nie mieszczą się w ramach przedstawionych przez nas metod badań.

Weźmy pod uwagę konkretny przykład, a mianowicie układ elektryczny złożony z n podstawowych elementów, z których każdy ma m zacisków, q zacisków zaś służy do połączeń układu ze źródłami zewnętrznymi. W sumie mamy

$$N = nm+q \text{ zacisków.} \tag{20.2}$$

Każdy z zacisków można połączyć z innym, co daje

$$P = N^N.$$

Informacja zawarta w jednym schemacie ideowym wynosi:

$$I = K \ln P = KN \ln N. \tag{20.3}$$

Przyjmując $K = \lg_2 e \approx 1{,}4$ otrzymujemy ilość informacji w bitach:

$$I = N \lg_2 N.$$

Niech na przykład

$$n = 999, \quad m = 10, \quad q = 10;$$

wówczas

$$N = 10000 \approx 2^{13{,}3}, \quad I = 133000 \text{ bitów.}$$

Jest to duża liczba w bitach, ale obliczając ją w jednostkach termodynamicznych należy przemnożyć ją przez 10^{-16}, co daje negentropię

$$-S = 1{,}33 \cdot 10^{-11},$$

a więc jest ona całkowicie do pominięcia.

Maszyna zbudowana według naszych założeń zawiera w sobie taką negentropię strukturalną, która mierzy informację lub stopień organizacji maszyny. Negentropia ta jest jednak do pominięcia w porównaniu z całkowitą entropią maszyny. Przykład ten zawiera w sobie niektóre charakterystyczne cechy całej teorii.

Negentropijna zasada informacji żąda, aby każda składowa informacji powiązana była z ujemnymi składowymi fizycznej entropii układu.

W praktyce, te ujemne składowe można pominąć w większości przypadków ze względu na mały współczynnik 10^{-16} wprowadzany przy zamianie jednostek

binarnych na termodynamiczne. Poszczególne organizacje mogą zawierać bardzo dużą ilość informacji wyrażonej w bitach. Rozpatrzmy np. sieć telefoniczną rozciągniętą na obszar Ameryki. Rząd liczby abonentów może wynosić kilkadziesiąt milionów, ale będziemy szczodrzy i przyjmiemy sto milionów:

$$N = 10^8.$$

Bezpośrednie zastosowanie równania typu (20.3) pokazuje, że w jakimś danym ustalonym czasie ilość informacji w całej sieci wynosi

$$I = N \lg_2 N \approx 40 \cdot 10^8 = 4 \cdot 10^9 \text{ bitów},$$

co w jednostkach termodynamicznych daje $4 \cdot 10^{-7}$, tj. bardzo małą wartość.

Trudno jest wyobrazić sobie maszynę zawierającą większą ilość informacji niż w podanym przykładzie. Gdy pomyślimy jednak o organizmach żywych, to znajdziemy całkiem inny rząd wielkości. Ilość komórek może być rzędu 10^{12}, natomiast trudno jest przewidzieć ilość cząsteczek chemicznych powiązanych wzajemnie ze sobą w organizmie żywym. Informacja zawarta w organizmie żywym może osiągnąć w skali entropijnej wartość, której nie można już pomijać. Zagadnienie to jest jeszcze jednym z niezbadanych zagadnień.

20.2. INFORMACJA ZAWARTA W PRAWIE FIZYCZNYM

Na różnych przykładach prześledziliśmy przydatność statystycznej definicji *ilości informacji*. Można próbować rozciągnąć tę definicję na zagadnienia bardziej ogólne i zasugerować nowe pole zastosowań.

W rozdziałach 14 do 16 prześledziliśmy informację otrzymywaną z eksperymentu fizycznego i doszliśmy do wniosku, że ilości informacji nie można określić, dopóki nie poda się zakresu obserwacji i dokładności pomiaru. Sytuację podobną spotykamy przy określaniu informacji zawartej w prawie fizycznym. Aby ściśle odpowiedzieć na to pytanie, musimy przede wszystkim sprecyzować warunki, w jakich prawo zostało odkryte, i w jaki sposób ma ono być wykorzystane.

Rozpatrzmy np. równanie gazów doskonałych:

$$f(p, v, T) = \frac{pv}{T} = R = \text{const.} \tag{20.4}$$

Musimy określić zakres zastosowań wyszczególniając, do jakich gazów można je stosować, w jakim zakresie zmieniają się p, v, T oraz jakiej dokładności można oczekiwać.

Prawo fizyczne jest podsumowaniem wyników doświadczeń i podlegać może dyskusji, dopóki nie wymienimy szczegółowo warunków doświadczenia. Mając to na uwadze, zagadnienie można rozważyć teraz w pojęciach ogólnych.

Przypuśćmy, że obserwowaliśmy układ fizyczny i mierzyliśmy zmienne $x_1, x_2, \ldots, x_n$. Każda zmienna x_i może przyjmować wartości z zakresu X_i, a błąd obserwacji wynosi α_i. Zatem całkowita liczba rozróżnialnych punktów pomiarowych w zakresie obserwacji X_i wynosi

$$\frac{X_i}{\alpha_i}. \tag{20.5}$$

Przypuśćmy, że po przeanalizowaniu wyników dochodzimy do wniosku, że funkcja

$$f(x_1, x_2, \ldots, x_n) = \text{const} \tag{20.6}$$

poprawnie odzwierciedla warunki doświadczenia z błędem δf nie przekraczającym pewnej wartości φ. ([20.1]) Nie jest to zupełnie wyczerpująco przedstawione zagadnienie. Musimy wprowadzić przede wszytkim zakres zmienności F funkcji f. Zakres ten jest zakresem a priori i zależy od początkowej wiedzy, jaką mamy przed rozpoczęciem doświadczeń. Jeżeli nic nie wiemy o układzie, to zakres F wyznaczają zakresy przyrządów pomiarowych. Funkcja f określona jest przeto w całym przedziale F i może przyjmować wartości różniące się o φ, liczba zaś możliwych (rozróżnialnych) wartości funkcji f wynosi F/φ. Występują one z równymi prawdopodobieństwami, przed rozpoczęciem obserwacji. Ilość rozróżnialnych w hiperprzestrzeni punktów o współrzędnych $x_1, x_2, \ldots, x_n$ wynosi

$$N = (X_1 X_2 \ldots X_n)/(\alpha_1 \alpha_2 \ldots \alpha_n). \tag{20.7}$$

Każdemu z tych punktów odpowiada pewna wartość funkcji, przy czym może to być jedna spośród F/φ wartości funkcji f. Całkowita liczba możliwości ([20.2]) będzie równa

$$\left.\begin{aligned} P_0 &= F/\varphi \quad \text{przed eksperymentem,} \\ P_1 &= 1^N = 1 \quad \text{po eksperymencie.} \end{aligned}\right\} \tag{20.8}$$

Ze wzorów (1.5) i (1.6) otrzymujemy ilość informacji

$$I = K \ln \frac{P_0}{P_1} = KN \ln \frac{F}{\varphi} = KN(\ln F - \ln \varphi). \tag{20.9}$$

Obszar zmienności zmiennych przyjęto zgodnie ze wzorem (20.5); jest to równoległościan. Ograniczenie to nie jest konieczne. Można wziąć w rozważanej hiperprzestrzeni $(x_1, x_2, \ldots, x_n)$ np. jakiś obszar o dowolnym kształcie i objętości U. Błędom odpowiadającym jednej obserwacji odpowiadać będzie mały obszar o objętości α. Wówczas zamiast wzoru (20.7) mamy

$$N = \frac{U}{\alpha} \tag{20.10}$$

i ilość informacji

$$I = \frac{KU}{\alpha}(\ln F - \ln \varphi) \tag{20.11}$$

odpowiada ilości informacji zawartej w prawie fizycznym. Wzór ten spełnia następujące wymagania: im dokładniejsze są dane (mniejsze α oraz φ) lub im większy jest zakres obserwacji U, tym większa będzie uzyskiwana ilość informacji. Występujący we wzorze (20.11) logarytm wyrażenia F/φ (jak zaznaczono w rozdz. 1) zapewnia addytywność niezależnych obserwacji.

Możemy wyjaśnić teraz sens wielkości F porównując informację otrzymaną w kolejnych krokach przy wzrastającej dokładności. Przyjmijmy, że pierwsza seria obserwacji daje funkcję f_1 z błędem φ_1, podczas gdy druga, bardziej dokładna seria obserwacji daje funkcję f_2 z błędem φ_2. Przypadek taki występuje np. wtedy, gdy prawo Boyle'a (20.4) zastępujemy prawem van der Waalsa. Wzór (20.11) daje wzrost ilości informacji odpowiadający takiej zamianie:

$$I = I_2 - I_1 = K \frac{U}{\alpha}(\ln \varphi_1 - \ln \varphi_2). \tag{20.12}$$

Wzór ten jest identyczny ze wzorem (20.11) z wyjątkiem tego, że F i φ zastąpiono przez φ_1 oraz φ_2. Wyeliminowano tu wielkość $\ln F$. W rzeczywistości gra ona rolę składowej stałej, przypominając stałą występującą w wielu wyrażeniach na entropię.

Przykład ten wykazuje, że naszą statystyczną definicję informacji stosować można do różnych przykładów i że prowadzi ona do ciekawych wyników.

MacKay* w swoich artykułach omówił zagadnienia informacji naukowej. Próbował on udowodnić, że należy stosować różne definicje w różnorodnych dziedzinach zastosowań. Wprowadza on pojęcie *informacji selektywnej* w miejsce wielkości, którą zdefiniowaliśmy jako *informację*. Wprowadził również pojęcia: *informacja strukturalna* i *informacja metryczna*. Te ostatnie określenia występują w połączeniu z zagadnieniami dokładności przyrządów doświadczalnych. Nasze badania i rozważania idą w kierunku wykazania, że wszystkie te zagadnienia można i powinno się rozważać stosując te same definicje i wzory. Jeśli byłoby inaczej, na pewno wykrylibyśmy nieścisłości, gdyż pola zastosowań różnych wzorów zachodzą na siebie i otrzymalibyśmy trzy odmienne, niezależne teorie informacji zamiast jednej. Wyczuwamy zdecydowanie, że pożądana jest tylko jedna teoria, która może pokryć adekwatnie całą dziedzinę badań.

20.3. INFORMACJA ZAWARTA W TABLICY LICZB

Zajmiemy się teraz zagadnieniem ilości informacji zawartej w tablicy liczb; przykładami takich tablic mogą tu być:

(1) tablica liczb przypadkowych,

(2) tablica przedstawiająca pewną funkcję.

Ogólna odpowiedź jest prosta: tablica zawiera N pozycji odpowiadających N wartościom zmiennej x. Każdej wartości x tablica przypisuje wartości innej zmiennej $y(x)$, przedstawionej liczbą składającą się z n cyfr dziesiętnych lub b cyfr binarnych. Liczby n oraz b, w przybliżeniu, związane są relacją

$$b \approx 3{,}3\, n. \tag{20.13}$$

W ogólności tablica zawiera

$$I = Nb \tag{20.14}$$

bitów informacji bez względu na prawo wiążące y z x.

W przypadku (1), odpowiadającemu tablicy liczb przypadkowych, zbiór liczb y powinien spełniać pewne statystyczne zależności, np. liczby nie mogą być skorelowane itp. Jednakże raz dana tablica nie podlega już działaniu czyn-

* D. M. MacKay, Phil. Mag. [7], **41** (1950), str. 289. *Symposium on Information Theory*, London, Trans. I.R.E. (P.G.I.T.) (1953), str. 60.

ników przypadkowych. Jest to tablica taka jak każda inna i można powiedzieć, że przedstawia ona pewną funkcję $y(x)$, która ogólnie mówiąc nie będzie funkcją ciągłą i nie będzie mieć pochodnej.

Jeżeli tablica przedstawia pewną funkcję, możemy rozróżnić następujące przypadki:

(2a) Jeżeli przyjąć, że początkowo nic nie wiemy o funkcji, to tablica zawiera informację Nb zgodnie ze wzorem (20.14).

(2b) Jeżeli coś o funkcji wiemy z góry, np. że funkcja ma I, II i n-tą pochodną z wyjątkiem pewnych punktów x, to tę zadaną z góry informację należy odjąć od informacji danej wzorem (20.14) i wówczas

$$I = Nb - I_0, \tag{20.15}$$

gdzie I_0 — informacja znana z góry.

Na przykład niech Δx będzie różnicą między dwiema kolejnymi wartościami x z tablicy. Załóżmy, że znany przyrost $|\Delta y|$ odpowiadający przyrostowi Δx nie może przekraczać b' jednostek binarnych. Wówczas

$$-2^{b'} \leqslant \Delta y \leqslant +2^{b'}. \tag{20.16}$$

Dla pierwszej wartości funkcji, przy $x=0$, mamy 2^b możliwych wartości y. Dla następnej wartości zakres redukuje się do wartości

$$2 \cdot 2^{b'} = 2^{b'+1}.$$

Ogólna liczba możliwości wynosi

$$P = 2^b (2^{b'+1})^{N-1},$$

a ilość informacji w bitach

$$I = \lg_2 P = b + (N-1)(b'+1) \tag{20.17}$$

i jest mniejsza od Nb, ponieważ $b+1$ jest mniejsze od b.

(2c) Możemy założyć, że z góry znamy dokładnie, jaka funkcja została stabelaryzowana. W tym przypadku tablica nie dostarcza żadnej nowej informacji.

Zwykle występuje przypadek (2b), kiedy dysponujemy częściową wiedzą o funkcji stabelaryzowanej, a tablica pożądana jest dla szczegółowej informacji.

Wprowadzone przez nas definicje wykazują, że najwięcej informacji zawiera tablica liczb przypadkowych, o których z góry nic nie wiemy.

20.4. UWAGI OGÓLNE

Teoria informacji zrodziła się na gruncie telekomunikacji i w tej dziedzinie ma najwięcej zastosowań. Staraliśmy się dowieść, iż teoria ta może być użyteczna w nauce „czystej", a szczególnie w fizyce.

Podobieństwo między informacją i entropią zostało zaakcentowane przez Shannona, a w rzeczywistości zwrócono na to uwagę po raz pierwszy w starej pracy Szilarda, która ukazała się w czasie, kiedy nie zdawano sobie sprawy z użyteczności teorii informacji.

Natura informacji i entropii fizycznej jest taka sama. Entropia jest miarą braku szczegółowej informacji o układzie fizycznym. Im większa będzie informacja, tym mniejsza będzie entropia. Informacja przedstawia ujemne składowe entropii układu i w oparciu o to sprecyzowaliśmy negentropijną zasadę informacji.

Za każdym razem, kiedy dokonujemy obserwacji układu fizycznego, musimy dysponować różnego rodzaju źródłami negentropii. Wykorzystując tę negentropię, zwiększamy entropię w laboratorium, w którym znajduje się obserwowany układ i przyrządy pomiarowe. W wyniku doświadczenia otrzymujemy pewną ilość informacji o układzie. Ilość uzyskanej informacji jest zawsze mniejsza od przyrostu entropii. Wynik ten uogólnia zasadę Carnota, a jego słuszność sprawdziliśmy na różnorodnych przykładach.

Wykazaliśmy, że teorii informacji nie można rozwijać jako oddzielnej całości. Powiązanie z termodynamiką jest tak ścisłe, że w konsekwencji wymagana jest fizyczna teoria informacji. Celem tej książki jest udowodnienie tego.

Teoria informacji pokazuje, że w większości zagadnień zmiany entropii układu fizycznego związane z wydobywaniem informacji są bardzo małe. Informację mierzy się zwykle w bitach (jednostkach binarnych), natomiast entropię — w jednostkach termodynamicznych. Stosunek tych dwu wielkości wynosi w przybliżeniu 10^{-16}; oznacza to, że bardzo duża liczba bitów np. 10^6 odpowiada jeszcze pomijalnie małej zmianie entropii 10^{-10}, niemniej jednak powiązanie między tymi wielkościami ze względu na logiczną konsekwencję jest nieuniknione.

Wiele zagadnień poruszanych w rozdz. 16 i 17 napotykało przeszkody. Nie można ich było rozwiązać bez uciekania się do „teorii fizycznej", a ich wyjaśnienie wymaga również tej nowej formy teorii.

Bardzo mała ilość negentropii, odpowiadająca dosyć dużej ilości informacji, jest głównym powodem tego, że przesyłanie informacji nie jest na ogół drogie. Pisanie, drukowanie i komunikacja elektryczna są tanie, jeżeli wyraża się je w jednostkach entropii. Koszt tego w dolarach jest również odpowiednio niski. Współczesne życie opiera się na tych faktach i byłoby całkowicie odmienne w świecie, w którym negentropia związana z informacją byłaby znacznie większa.

Fakt, że nasze rozważania dotyczyły małych wielkości, nie powinien nas zniechęcać. Początkowo wydawało się, że teoria względności wnosi tylko nieznaczne poprawki do mechaniki klasycznej.

Nowe zastosowania w fizyce jądrowej dowodzą, jakie fundamentalne znaczenie ma powiązanie masa—energia. Można się spodziewać, że prędzej czy później związek entropia—informacja wystąpi na pierwszy plan, i będzie można wykazać, gdzie znajdzie on pełne zastosowanie. Zagadnienie mierzenia skrajnie małych odległości, omówione w rozdz. 16, posłużyć może jako pierwszy przykład, a inne może już wkrótce się pojawią.

20.5. PRZYKŁADY ZAGADNIEŃ NIE MIESZCZĄCYCH SIĘ W RAMACH WSPÓŁCZESNEJ TEORII

Współczesna teoria informacji całkowicie pomija wartość przechowywanej, transmitowanej lub przetwarzanej informacji. Punkt ten skrupulatnie podkreślano w całej tej książce. Wydaje się, że inni autorzy nie zrozumieli znaczenia ograniczeń, wobec czego powstało wiele nieporozumień co do możliwości obecnego stanu teorii. Pomijając wartość informacji związanej bezpośrednio z człowiekiem, mogliśmy rozwinąć naukową teorię informacji w oparciu o statystykę i wykazać przydatność tej teorii. Jest wiele zagadnień, których nie można ująć w te ramy. Celem tego rozdziału jest krótkie omówienie niektórych zagadnień i wykazanie, jak wszystkie zmierzają do tego samego zagadnienia ogólnego, mianowicie, jak wprowadzić do teorii element *oceny wartości informacji*.

Element ten jest niezbędny za każdym razem, jeżeli zakłada się informację jako bazę dla produkcji i zastosowania praktycznego. Podsumujmy krótko obecny stan rzeczy. Załóżmy np., że wysyłamy telegram zapisany w kodzie binarnym, przy czym zawiera on n cyfr binarnych (0 lub 1). Jeżeli nie ma

rozwlekłości w kodowaniu, to telegram zawiera w całości n bitów informacyjnych.

Czy telegram przedstawia znaczenie dla użytkownika, czy nie — tego nie rozpatrujemy.

Na konferencji naukowej w Mediolanie w kwietniu 1954 r. Louis Couffignal (Paryż) postawił następujące zagadnienie: po wysłaniu telegramu autor dodaje jedną cyfrę o znaczeniu:

0: wszystko jest kłamstwem; nie zwracać uwagi na telegram.

1: telegram jest prawdziwy; można go wykorzystać.

Z punktu widzenia informacji oznacza to dodanie jednego bitu informacji tak, że w miejsce n cyfr binarnych otrzymuje się $n+1$. Praktycznie sytuacja jest inna:

0 (NIE) — niszczy zawartość informacyjną telegramu,

1 (TAK) — wprowadza po prostu rozwlekłość.

Można powiedzieć, że NIE (0) niesie informację negatywną, równą wartości bezwzględnej całkowitej informacji telegramu.

Obecna teoria rozpatruje informację zawsze pozytywną, natomiat pomija informację negatywną. Występuje tu wyraźnie rozgraniczenie między statystyczną miarą ilości informacji (współczesna teoria) a praktyczną przydatnością tej informacji dla użytkownika. Kody z wykrywaniem i korygowaniem błędów, omawiane w rozdz. 6, odpowiadają praktycznym zagadnieniom, bardzo podobnym do omówionego wyżej. Do telegramu zawierającego n znaków binarnych dodajemy 0 lub 1 tak, aby liczba jedynek była zawsze parzysta. Odbiornik sprawdza całkowitą liczbę jedynek:

jeżeli ilość jedynek jest nieparzysta, to może być 1, 3, 5, ... błędów;

jeżeli ilość jedynek jest parzysta, to błędu może nie być wcale lub może być ich 2, 4, 6 itp.

Przypadek pierwszy odpowiada sygnałowi NIE w zagadnieniu poruszanym przez Couffignala.

Dodatkowe sygnały do wykrywania lub korekcji błędu nie wprowadzają żadnej informacji. Zgodnie z obecnym stanem teorii sygnały te wnoszą po prostu rozwlekłość. Chodzi tu jednak o praktyczną *wartość* informacji.

Zamiast klasyfikacji na *prawdę* i *fałsz* (TAK lub NIE) możemy zdecydować się na podanie prawdopodobieństwa tego, że wiadomość jest prawdziwa. Przypuśćmy, że wybraliśmy skalę z $8 = 2^3$ wartościami i niech liczba m $(0 \leqslant m \leqslant 8)$ odpowiada prawdopodobieństwu $p = m/8$, co oznacza, że śred-

nio m z 8 stwierdzeń jest prawdziwych, a $(8-m)$ fałszywych. Z punktu widzenia informacji oznacza to właśnie dodanie trzech bitów informacyjnych do telegramu, a więc otrzymujemy całkowitą liczbę bitów $(n+3)$ zamiast n. O ile użytkownik bierze pod uwagę wartość informacji, to słuszne jest przyjęcie, że wartość tę wyraża się iloczynem pn. W wielu przypadkach informacja nie interesuje odbiorcy, jeżeli prawdopodobieństwo błędu przekracza ustalony próg, np. $^1/_2$; w tym przypadku wartość wynosi zero dla prawdopodobieństw 0, $^1/_8$, $^2/_8$, $^3/_8$ i nabiera coraz większego znaczenia, gdy prawdopodobieństwa wynoszą $^4/_8$, $^5/_8$, $^6/_8$, $^7/_8$ i 1.

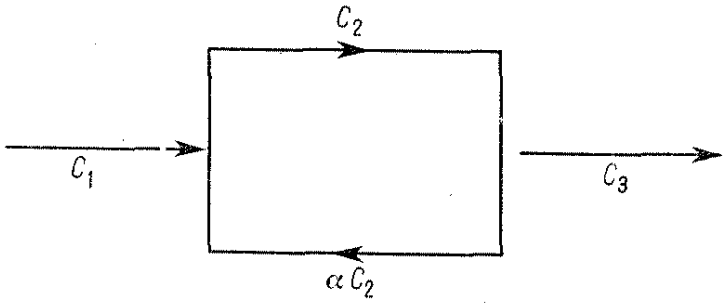
Zagadnienie pewności obserwacji naukowej omówiono w rozdz. 14. Pomiary fizyczne kryją w sobie możliwości błędów. Pewność metody doświadczalnej r zdefiniowano prawdopodobieństwem błędu $\frac{1}{r}$. Błąd ten powodowany jest fluktuacjami termicznymi, natomiast prawdopodobieństwo poprawnej obserwacji wynosi $\left(1-\frac{1}{r}\right)$. Zagadnienie to jest podobne do jednego z poprzednio rozważanych, ale zostało uzupełnione założeniem, że najmniejsza wartość rozważanego r wynosi 2, co daje 50% szans na poprawność obserwacji.

W rozdziale 19 omówiono maszyny matematyczne i wykazano, że obliczenia podobne są do tłumaczenia, kodowania i dekodowania. W najbardziej korzystnym przypadku obliczenia zachowują pełną informację, jaką podano na wejście (dane wejściowe), zwykle jednak część informacji ulega zatraceniu na skutek przybliżeń w procesie obliczeń. Z informacyjnego punktu widzenia obliczenia nie wprowadzają żadnej nowej informacji, lecz powtarzają ją w innym języku z możliwymi oczywiście stratami. Obliczenia, podobnie jak tłumaczenie czy dekodowanie, dodają z pewnością praktycznej wartości informacji i czynią ją bardziej zrozumiałą dla użytkownika. Zagadnienie informacji zawartej w programie obliczeń ma inny aspekt i do tej pory nie zostało zbadane.

Z punktu widzenia praktycznego wykorzystania informacji zagadnienie obiegu informacji w pewnych kanałach ze sprzężeniem zwrotnym (lub bez niego) jest bardzo istotne. Dodatnie sprzężenie zwrotne posiada inne właściwości niż ujemne. Według naszej definicji ilość informacji jest zawsze dodatnia i nigdy nie przyjmuje wartości ujemnej. W następstwie tego, co powiedziano, w kanałach sprzężenia zwrotnego mamy do czynienia nie z informacją

samą w sobie, ale jej wartością. Jak zdefiniować *wartość informacji* w tego typu zagadnieniach — to pytanie, które czeka na odpowiedź.

Tymczasem możemy wyciągnąć następujący wniosek (nie mający dużego znaczenia praktycznego). Rozważmy rys. 20.1, na którym kanał wejściowy C_1 dostarcza średnio C_1 bitów informacji na sekundę (przepustowość C_1).



Rys. 20.1. Schematyczne przedstawienie przepustowości kanału w obwodzie ze sprzężeniem zwrotnym

Informacja ta przechodzi przez kanał C_2, a pewna jej część α $(0<\alpha<1)$ wraca pętlą sprzężenia zwrotnego. Kanał C_2 musi przepuścić informację w ilości

$$C_1(1+\alpha+\alpha^2+\ldots) = \frac{C_1}{1-\alpha} \quad \text{bitów na sekundę,}$$

tak że jego przepustowość

$$C_2 = C_1 \frac{1}{1-\alpha},$$

natomiast przepustowość C_3 dana jest wyrażeniem

$$C_3 = C_2 - \alpha C_2 = C_1.$$

Nie ma to nawet nic wspólnego z dobrze znanymi zagadnieniami obwodów ze sprzężeniem zwrotnym. Zagadnienie tego typu można rozważać w oparciu o metody Linvilla i Salzera, które rozwinięto w rozdz. 19.

Duża liczba zagadnień wskazuje na potrzebę analizy praktycznej wartości informacji dla użytkownika. Miara ilości informacji — jak zdefiniowano — jest wielkością obiektywną, natomiast wartość informacji dla użytkownika jest względna. Nasuwa się myśl, czy zagadnienia te można omawiać z empirycznego punktu widzenia, próbując w każdym przypadku wprowadzić rozsądną definicję wartości. Porównując różnorodne zagadnienia w celu wy-

krycia wspólnych własności definicji, można by wyjaśnić, czy da się rozwinąć teorię, którą można stosować do wszystkich specjalnych przypadków. Rozważanie takie uczyniłoby prawdopodobnie teorię informacji bardziej przydatną w zastosowaniach praktycznych i otworzyłoby nowe kierunki badań o wielkim znaczeniu.

We wszystkich przypadkach jedno jest oczywiste: dowolne kryterium *wartości* prowadzi do oceny ilości odbieranej informacji. Jest to równoważne wyborowi definicji ilości informacji zgodnie z pewną miarą jakości. Pewną część informacji można uznać za ważną i zachować ją z całą jej wartością, podczas gdy pozostała część informacji nie będzie przedstawiać żadnej wartości i można ją odrzucić. Zagadnienie to pokrywa się z procesem filtracji. Filtr jest dobrze znanym elementem w teorii ogólnej i ma dwie ważne właściwości: jest nieodwracalny i zmniejsza ilość całkowitej informacji. Można podobnie stwierdzić, że względna wartość informacji dla danego użytkownika jest mniejsza lub równa wartości absolutnej. Jest to już pewien punkt wyjścia i wydaje się, że zagadnienie można by sprowadzić do ogólnego zdefiniowania jakości różnych metod filtracji i wykorzystać je do oszacowania względnej wartości informacji.

20.6. ZAGADNIENIA INFORMACJI SEMANTYCZNEJ

Teoria podana w tej książce oparta jest w głównej mierze na modelu statystycznym. Zgodnie z naszymi definicjami, zdania występujące bardzo rzadko niosą dużą ilość informacji. Zagadnienie wartości informacji, omówione krótko w rozdz. 20.5, nie jest rozważane w naszej teorii. Pominęliśmy też całkowicie inne zagadnienie, a mianowicie *myślenie*. To, czy telegram ma sens dla nadawcy lub odbiorcy, czy go nie ma, uznaliśmy za nieistotne. Niektóre zagadnienia semantyczne badali niedawno Ville* we Francji oraz Carnap i Bar-Hillel** w Stanach Zjednoczonych. Praca Carnapa i Bar-Hillela oparta jest

* J. Ville, *Inter. Congr. of Philosophy of Sci., Paris 1949*, Actualites Sci. et Ind. **1145** (1951), Hermann, str. 101—114.

** R. Carnap, Y. Bar-Hillel *An Outline of a Theory of Semantic Information*, Mass. Inst. Technol. Research Lab. Electronics Tech. Rept. No. 247 (1952); Brit. J. Phil. Sci. **4** (1953), str. 147—157.

na metodach i symbolach logiki symbolicznej. Podamy krótki zarys tej pracy, ograniczając się do zdań w bardzo prostym języku, częściowo ze względu na prostotę, częściowo z powodu ograniczeń metod logiki indukcyjnej, na których opierają się w większości ich rozważania. Autorzy badają kilka możliwych definicji miary ilości informacji zawartej w danym zdaniu i analizują zalety i wady każdej z nich. W ich analizie nie bierze się pod uwagę zagadnień wartości i interpretacji: pomijają oni to, czy odbiorca interpretuje telegram zgodnie z życzeniem nadawcy.

Według wszystkich podanych definicji, takie zdanie jak $17 \cdot 19 = 323$ nie niesie informacji. Powodem jest to, że odbiorca telegramu zna wszystkie możliwe wywody logiczne wynikające ze struktury języka, wobec czego zdanie takie jest rozwlekłe. Tylko takie zdania, w których co najmniej część treści nie jest uwarunkowana strukturą języka, uznać można za zdania niosące informację. Carnap i Bar-Hillel przyjmują za podstawę rozważań bardzo prosty język.

Składa się on ze skończonej liczby n rzeczowników (nazwanych indywidualiami), π przymiotników (nazwanych predykatami) oraz z pojedynczych słów MA lub JEST. I tak, jeżeli a jest rzeczownikiem, a P przymiotnikiem, to Pa czyta się jako: „a ma własności P" lub prościej: „a jest P". Język natomiast ma następujące elementy łączące, które służą do budowy zdań dłuższych.

$\sim$ NIE: negacja; $\sim Pa$ oznacza „a nie jest P";
$\vee$ LUB: alternatywa; $Pa \vee Qb$ oznacza „a jest P LUB b jest Q";
$\wedge$ I: koniunkcja; $Pa \wedge Qb$ oznacza „a jest P I b jest Q";
$\supset$ JEŻELI..., TO: implikacja;
$\equiv$ JEŻELI I TYLKO JEŻELI: tożsamość.

Wybór π przymiotników w języku jest nieco ograniczony. Nie powinny one w żaden sposób pokrywać się znaczeniowo: żaden przymiotnik nie powinien wpływać na użycie innego przymiotnika. Słownictwo języka uzupełnione jest wprowadzeniem dla każdego π przymiotników P jego antonimów lub negacji, które mogą być zaznaczone symbolem $\sim P$ lub nowym symbolem, a często obydwoma.

Zdania w tym języku dzielą się na zdania *logicznie fałszywe* (*wewnętrznie sprzeczne*, takie jak $Pa \wedge \sim Pa$), *faktualne* (*logicznie nieokreślone*, jak np. Pa) i *logicznie prawdziwe* (tzw. *tautologiczne*, np. $Pa \vee \sim Pa$).

Jeden z oczywistych sposobów określania informacji w takim zdaniu jest następujący: istnieje skończona liczba zdań, które można zbudować, oraz istnieją pewne powiązania logiczne między tymi zdaniami. Informację pewnego szczególnego zdania i można by zdefiniować jako odpowiednio wybraną funkcję liczby zdań uwarunkowanych przez i. Istnieją co najmniej dwa sposoby podania takiej definicji*, a czytelnik dla poznania szczegółów może odwołać się do książki Carnapa, jako że definicje te są w odległym związku z naszym punktem widzenia.

Artykuł R. Carnapa i Y. Bar-Hillela* podaje definicje bardziej zbliżone do tych, którymi się posługiwaliśmy. Definicje te opierają się na szczególnym typie zdania faktualnego, tzw. *opisie stanu***, oznaczonym przez Z. Opis stanu jest zdaniem składającym się z πn koniunkcji prostych (tzw. atomowych) zdań i każde z nich zawiera n rzeczowników powiązanych z przymiotnikiem lub jego negacją (jednak nie jednocześnie z obydwoma). Możliwych jest πn związków rzeczowników z przymiotnikami, a ponieważ każdy związek wystąpić może w dwu postaciach, tj. rzeczownik z przymiotnikiem lub z negacją przymiotnika, więc liczba opisów stanu wynosi $2^{\pi n}$. Opis stanu całkowicie oddaje możliwy stan wszechświata będącego przedmiotem rozważań, natomiast definicje informacji wymagają założenia prawdopodobieństwa a priori opisów stanu. W strukturze języka nie ma podstawy a priori do wyróżniania jakiegoś konkretnego zbioru prawdopodobieństw opisów stanu, chociaż niektóre własności przyjętego zbioru prawdopodobieństw ograniczają do pewnego stopnia wybór. W zagadnieniach takich, jakie poruszyliśmy w tej książce, własności układu fizycznego, który chcemy opisać, wskazują już na wybór prawdopodobieństw a priori. Na przykład można nazwać temperatury powyżej 20°C wysokimi, a niższe od 20°C niskimi. Oczywiście, prawdopodobieństwa temperatur wysokich na biegunie kuli ziemskiej będą inne niż na równiku.

Dla zdania i, które ogólnie rzecz biorąc nie jest opisem stanu, zakres wyboru i, oznaczony przez $R(i)$, jest zbiorem opisów stanu, w których i pozostaje słuszne. Na tym etapie można już wyprowadzić definicje informacji oparte o teorię zbiorów punktów. Interesują nas definicje ilościowe takie,

---

* R. Carnap, *Logical Foundations of Probability*, Chicago Ill. 1950, Univ. of Chicago Press.

* R. Carnap, Y. Bar-Hillel, *op. cit.*

** Tzw. *state description*. (*Przyp. tłum.*)

które dają liczbową miarę informacji. Wprowadźmy funkcję miary spełniającą następujące warunki:

(1) Dla każdego opisu stanu Z istnieje miara $m(Z)$ taka, że

$$0 \leqslant m(Z) \leqslant 1.$$

Jest to wielkość, którą można interpretować jako prawdopodobieństwo a priori stanu Z.

(2) Suma $m(Z)$ rozciągnięta na wszystkie Z wynosi 1.

(3) Dla fałszywego zdania f jest $m(f) = 0$.

(4) Dla zdania niefałszywego i definiujemy $m(i)$ jako sumę $m(Z)$ rozciągniętą na wszystkie Z mieszczące się w $R(i)$.

(5) Wszystkie rzeczowniki traktujemy jednakowo, tj. $m(i)$ jest inwariantne względem permutacji rzeczowników występujących w zdaniu i.

(6) Wszystkie przymiotniki traktujemy jednakowo, tj. $m(i)$ jest inwariantne względem permutacji przymiotników występujących w zdaniu i.

(7) Dowolne przymiotniki można zastąpić ich negacjami bez zmiany wartości $m(i)$.

(8) Jeżeli dwa zdania i oraz j nie mają wspólnych przymiotników, to mówimy, że są indukcyjnie niezależne:

$$m(i, j) = m(i)\, m(j).$$

(9) Liczba rzeczowników nie wymienionych w zdaniu i nie ma wpływu na wartość $m(i)$.

Taką funkcję miary nazwiemy: *własną m-funkcją*. Carnap zakłada, że właściwie można ją rozpatrywać jako prawdopodobieństwo, które nazywa *logicznym prawdopodobieństwem absolutnym*. Warunki (5), (6) i (7) wprowadzają związki między prawdopodobieństwami różnych opisów stanu, ale nie określają ich jednoznaczne. Własności $m(i)$ można wykorzystać przy dedukcji wielu twierdzeń i zależności, bez których nie można szczegółowo rozwinąć teorii. Z wielu relacji wykorzystamy następującą:

$$m(i) + m(\sim i) = 1.$$

Ponieważ $R(i)$ i $R(\sim i)$ nie zawierają wspólnych opisów stanu (opis stanu nie może być związany jednocześnie z i oraz z jego negacją), a każdy opis stanu mieści się w $R(i)$ lub $R(\sim i)$ (każdy opis stanu związany jest z i lub jego ne-

gacją), to

$$m(i)+m(\sim i) = \sum_{Z} m(Z) = 1.$$

Najprostszą definicję informacji, wyrażoną funkcją miary, Carnap i Bar-Hillel nazwali *miarą zawartości* (*cont*); wyraża się ona przez

$$cont(i) = m(\sim i) = 1-m(i).$$

Miara zawartości ma wiele ciekawych właściwości i jakościowo zgadza się z definicjami stosowanymi w tej książce. Sens jej jest następujący: Jeżeli $m(i)$ jest małe, wówczas $cont(i)$ jest względnie duże. Małe $m(i)$ oznacza uwzględnianie stosunkowo mało prawdopodobnych opisów stanu, tak że $cont(i)$ jest miarą rzadkości.

Druga definicja — *miara informacji* jest bardziej zbliżona do naszej definicji. Wyraża się zależnością

$$inf(i) = -\lg_2 m(i) = \lg_2\{1/[1-cont(i)]\}.$$

Jakościowo zgadza się ona, tak jak i $cont(i)$, z definicją podaną przez nas. Dalej można wykazać, że warunek (8), jaki spełnia funkcja $m(i)$, prowadzi do addytywności miary informacji dla koniunkcji zdań indukcyjnie niezależnych. Analiza szczególnego przypadku wykaże wyraźniej ścisły związek między obydwoma definicjami. Wybierzmy funkcję miary typu

$$m(Z) = \frac{1}{N} \qquad \text{dla wszystkich } Z,$$

gdzie N jest całkowitą liczbą opisów stanu. Załóżmy, że opisy stanu są równo prawdopodobne. Założenie to zgadza się z warunkami nałożonymi na funkcję miary. Jeżeli N_r będzie liczbą opisów stanu mieszczących się w $R(i)$, wówczas definicja $inf(i)$ daje

$$inf(i) = -\lg_2\left(\frac{N_r}{N}\right) = \lg_2\left(\frac{N}{N_r}\right).$$

Równość ta jest identyczna z tą, jaka występuje w całej książce, jeżeli utożsamić *opisy stanu* z możliwościami równoprawdopodobnymi w jakiejś danej sytuacji. Początkowo wchodzi w grę N możliwych przypadków N opisów stanu. Znajomość zdania i redukuje możliwą liczbę N do N_r tak, że informacja dos-

tarczona przez zdanie i zgodnie z poprzednimi definicjami wynosi

$$I = \lg_2 \left(\frac{N}{N_r} \right) \text{ bitów.}$$

Dla opisu stanu Z funkcja $R(Z)$ zawiera tylko Z i stąd

$$I = \lg_2 N \text{ bitów.}$$

Carnap i Bar-Hillel szczegółowo rozwijają dalej semantyczne definicje i stosują je do różnych zagadnień wnioskowania i dedukcji.

W tym krótkim streszczeniu próbowaliśmy wykazać związki między zagadnieniami poruszonymi przez wymienionych autorów a zagadnieniami, którym poświęcona jest ta książka.

21

NIEUNIKNIONE BŁĘDY, DETERMINIZM I INFORMACJA

21.1. INFORMACJA W NAUCE

Teoria informacji porusza wiele podstawowych kwestii związanych z pierwotnymi postulatami naukowymi i określa zakres ich znaczenia. Jako przykład porównajmy opinie czystego matematyka i fizyka o podstawach geometrii. Matematyk zaczyna od bezwymiarowości punktów, nieskończenie cienkich krzywych, nieskończenie cienkich płaszczyzn i ciągłej czasoprzestrzeni. Według nauki o atomach definicje te nie znajdują odzwierciedlenia w świecie rzeczywistym.

Weźmy pod uwagę bardzo cienką folię i prześwietlmy ją promieniami X. Odkrywamy w niej siatkę atomową z poszczególnymi atomami, porozdzielanymi wielkimi, pustymi obszarami. Folia ma skończoną grubość i nie jest ciągła. Nawet monomolekularna warstwa wykazuje podobne własności.

Matematyk bardzo dokładnie definiuje liczby niewymierne. Fizyk nigdy nie spotyka tego rodzaju liczb. Dowolny jego pomiar zapisany jest liczbą skończoną, wielocyfrową i z pewną nieoznaczonością. Matematyk natomiast drży na myśl o nieoznaczności i próbuje pomijać błędy doświadczenia.

Otwórzmy książkę czysto matematyczną i rozważmy dowolne twierdzenie. Zbudowane jest ono zawsze według tego samego schematu: zadane są pewne warunki A, B, C, o których przyjmuje się, że są ściśle spełnione, i można udowodnić, że wniosek Q musi być prawdziwy. Fizyka zaś zastanawia to, *jak* poznać, że A, B, C są ściśle spełnione. Obserwacja nie daje na to odpowiedzi. Jedyną rzeczą, jakiej możemy się dowiedzieć, jest to, że A, B, C są spełnione w przybliżeniu, z pewnymi błędami. Co wynika wówczas z twierdzenia? Otóż to, że bardzo małe błędy w A, B, C mogą spowodować bardzo mały błąd

w końcowym stwierdzeniu Q lub mogą całe to stwierdzenie obalić. Rozważanie nie jest zupełne, dopóki nie zostanie zbadane zagadnienie stabilności twierdzenia, ale jest to już inna sprawa.

21.2. INFORMACJA JEST SKOŃCZONA

W książce tej mówiliśmy dużo na temat zagadnienia błędów pomiarowych. Teoria informacji pozwala możliwie poprawnie postawić te zagadnienia i uzyskać logiczną odpowiedź. Jesteśmy teraz w stanie zdefiniować ilość informacji uzyskanej z pewnego doświadc enia i precyzyjnie ją mierzyć. Potrzebna nam jest tylko znajomość pola nieoznaczoności przed i po obserwacji. Stosunek tych dwu nieoznaczoności daje ilość informacji. Jeżeli nieoznaczoność końcowa jest bardzo mała (bardzo dokładny pomiar), to uzyskana informacja jest bardzo duża.

Matematyk marzy o pomiarach nieskończenie dokładnych, określając np. położenie punktu bez jakiegokolwiek błędu. Oznacza to, że doświadczenie dostarczałoby nieskończenie dużo informacji, co jest fizycznie niemożliwe.

Jednym z najważniejszych wyników przedstawionej tu teorii jest *negentropijna zasada informacji*. Stwierdza ona, że informacja uzyskiwana z doświadczenia musi być opłacona entropią. Za bardzo dużą ilość informacji płacić się będzie bardzo wysoką ceną entropijną. Nieskończenie duża ilość informacji jest nieosiągalna. Nieskończenie małych odległości nie można mierzyć i niemożliwe jest fizyczne określenie fizycznej ciągłości czasu i przestrzeni.

21.3. POGLĄD M. BORNA

Omówimy teraz ciekawą wypowiedź Maxa Borna i artykuł opublikowany w książce poświęconej Nielsowi Bohrowi z okazji jubileuszu 70 rocznicy urodzin. W artykule tym pod tytułem: „Ciągłość, determinizm i rzeczywistość"* Max Born** przytacza niektóre ogólne uwagi ściśle związane z tym rozdziałem, chociaż jego sposób rozumowania jest inny:

«Twierdzę, że matematyczna koncepcja punktu w ciągłej przestrzeni nie posiada prostego znaczenia fizycznego. Powiedzenie, że współrzędna x [...] ma wartość $\sqrt{2}$ cali lub $x = \pi$ cm (występują tu liczby niewymierne), nic nie znaczy.»

* Tytuł oryginału: *Continuity, Determinizm and Reality*. (*Przyp.. tłum.*).

** M. Born, Kgl. Danske Videnskab. Selskab. Mat.-fys. Medd. **30** (1955), str. 2.

«Współczesna fizyka odniosła wielkie sukcesy stosując metodologiczną zasadę, że koncepcje, które odnoszą się do dokładności leżących poza możliwościami doświadczenia, nie mają fizycznego znaczenia i należy je wyeliminować. [...] Najbardziej rzucającymi się w oczy udanymi przypadkami są podstawy teorii względności Einsteina oparte na odrzuceniu koncepcji eteru [...] i podstawy mechaniki kwantowej Heisenberga. [...]»

«[...] Myślę, że zasadę tę należy stosować również do koncepcji fizycznej ciągłości.»

Max Born wyjaśnia, że błędy doświadczalne powinny być brane pod uwagę poczynając od fizyki klasycznej, gdzie również często je pomijano.

Nie chce on odrzucać matematycznej koncepcji liczb rzeczywistych, ale precyzuje zagadnienie: „sytuacja wymaga opisu mglistości". Zamiast żądać dokładnej wartości pewnej wielkości fizycznej, należy określić prawdopodobieństwo wystąpienia wartości tej wielkości w danym przedziale. Ta ogólna zasada omówiona przez Borna często nazywana jest *operacyjnym punktem widzenia* Bridgmana.

21.4. OBSERWACJE I BŁĘDY DOŚWIADCZALNE

Rolę błędów doświadczalnych znano od dawna i rozważało ją wielu naukowców; zwykle przyjmowano jednak, że jest to zjawisko wtórne, źródło wielu kłopotów, które można było zaniedbywać przy lada okazjach i pomijać w teoriach. Zakładano, iż błędy te można było uczynić dowolnie małymi, jeżeli zastosować odpowiednio dokładne przyrządy. Zakładano również, że błędy nie odgrywają żadnej roli. Był to pogląd matematyków, którzy rozważali aksjomaty geometrii, i większość fizyków przyjęła ten rodzaj idealizacji bez zastrzeżeń, jednak tylko formalnie.

Współczesna fizyka musiała uwolnić się od tych nierealnych schematów i uznanie istnienia błędów przyjąć jako nieprzyjemną konieczność. Błędów nie można czynić „tak małymi, jak chcemy", a ich istnienie musi być uwzględnione w teorii.

Pierwszy znaleziony przykład odnosił się do termodynamiki statystycznej. Był on jednak zwykle zacierany, co doprowadziło do licznych (naszym zdaniem często nie mających znaczenia) dyskusji takich, jak ta: czy możliwe jest otrzymywanie nieodwracalnych praw termodynamicznych ze ściśle odwracalnych praw mechanicznych?

Powrócimy do tego zagadnienia, kiedy będziemy omawiali dokładnie zna-

czenie determinizmu, i wykażemy, iż odpowiada on metafizycznej wierze, a nie prawom fizycznym.

Podstawowa rola błędów doświadczalnych wraz z zasadą nieoznaczoności Heisenberga stała się bazą przyszłej fizyki.

W rozdziale 12 i 16 sformułowano dodatkowe prawo i nazwano je *negentropijną zasadą informacji*. Stwierdza ono, że obserwacja dostarcza pewnej ilości informacji i że informację tę można mierzyć ilościowo i porównywać z przyrostem entropii ΔS podczas doświadczenia. W ostatecznym wyniku mamy

$$\Delta S \geqslant \Delta I \quad \text{lub} \quad \Delta I + \Delta N \leqslant 0, \tag{21.1}$$

gdzie $\Delta N = -\Delta S$ jest entropią ujemną.

Warunki te oznaczają, że skończona ilość energii ΔE musi ulec degradacji i zamianie w ciepło:

$$\Delta E = T\Delta S \geqslant T\Delta I,$$

gdzie T oznacza temperaturę bezwzględną. Definicję wielkości ΔI podano w rozdz. 1 i można ją krótko ująć w prosty sposób: niech P_0 będzie liczbą równoprawdopodobnych możliwości przed obserwacją, natomiast P_1 liczbą tych możliwości po obserwacji. Wprowadzamy definicję

$$\Delta I = k \lg \frac{P_0}{P_1}, \tag{21.2}$$

gdzie k jest stałą Boltzmanna. Jeżeli dokładność jest bardzo duża, P_1 staje się bardzo małe i ilość informacji jest bardzo duża. Nieskończenie duża dokładność z nieskończenie małym błędem oznaczałaby nieskończony przyrost informacji; w takim razie nieskończenie duża ilość energii ΔE podlegałaby degradacji.

Jest zatem ogólnie wypowiedziane to, co ujęto równaniem (21.1) lub podobnymi relacjami, które teraz przytoczymy. Przy pożądanej, wyjątkowo dużej dokładności ktoś może odkryć nowe ograniczenia, które jeszcze bardziej utrudnią pomiar bardzo małych odległości. Ktoś inny może znów odkryć (rozdz. 16), że bardzo duża ilość energii ΔE jest nieodzowna do obserwacji x z bardzo małym błędem Δx:

$$\Delta E \Delta x \geqslant \frac{1}{2} hc. \tag{21.3}$$

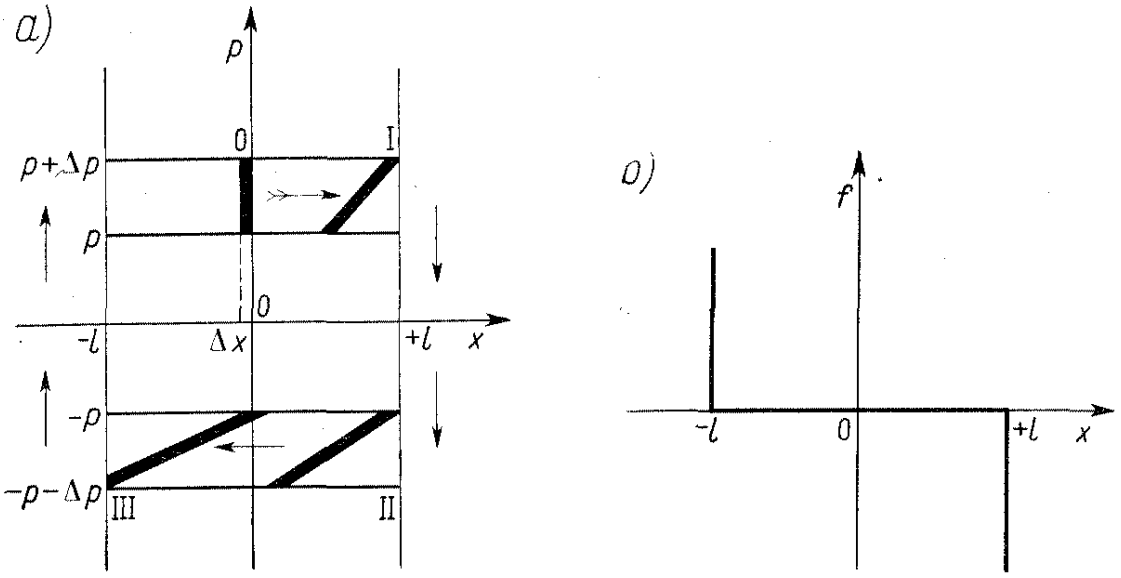
Ta bardzo duża energia ΔE ulega częściowo rozproszeniu zgodnie z warunkiem

(21.1) i kiedy Δx staje się bardzo małe (mniejsze niż 10^{-14} cm), warunek (21.3) wymaga wydatkowania energii dużo większej od tej, jakiej żąda wzór (21.1).

Born* rozważa również typowy przykład, który pokazuje związki między jego poglądami a podstawowymi założeniami termodynamiki. Przykład ten jest bardzo pouczający, bowiem pokazuje w sposób możliwie prosty najbardziej znamienne cechy, które chcemy podkreślić. W rozdziale 21.5 omówimy to zagadnienie bardziej szczegółowo.

21.5. PROSTY PRZYKŁAD: WYGNANIE DEMONA LAPLACE'A

Rozważmy piłkę poruszającą się w kierunku x między dwoma sztywnymi ścianami umieszczonymi przy $x = \pm l$. Wygląda na to, że ruch jest dokładnie określony, chcemy jednak rozważyć nieuniknione błędy przy pomiarze wa-



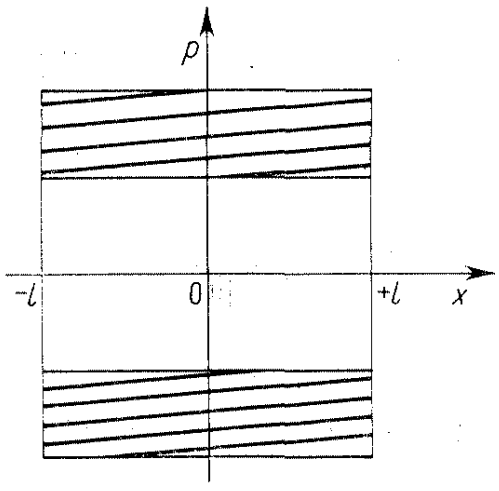
Rys. 21.1

runków początkowych. Gdy rzuciliśmy piłkę, znaleźliśmy punkt wyrzucenia $x = 0$ z błędem Δx i możemy określić początkowy pęd p z błędem Δp. To proste zagadnienie omówił M. Born. Zwrócił uwagę na fakt, że po czasie t, jaki upłynął, znamy położenie piłki z błędem $\Delta x + \left(\frac{m}{t}\right)\Delta p$ i może on wkrótce

* M. Born, Phil. Quart. **3** (1953), str. 139.

stać się większy niż odstęp między ścianami równy $2l$. Odtąd możemy tylko mówić, że piłka znajduje się gdzieś między ścianami. Jest to wniosek, który trudno nazwać zdeterminowanym.

Ten sam przykład rozpatrywał E. Borel* na wykładach mechaniki statystycznej w celu zilustrowania sensu słynnego twierdzenia Liouville'a. Jeżeli ruch rozpatrujemy na płaszczyźnie x, p (płaszczyźnie fazowej), twierdzenie to mówi, że obszar $\Delta x\,\Delta p$ określony przez błędy początkowe posuwa się wzdłuż trajektorii i podlega zniekształceniu w nie dający się przewidzieć sposób, ale pole tego obszaru zachowuje tę samą wartość. Rysunek 21.1 wyjaśnia zasadniczą sytuację: punkty wyjścia pokrywają początkowy prostokącik 0, ale punkty o większym p mają również większą prędkość $V = p/m$ i poruszają się szybciej w kierunku osi x; w takim razie po dość krótkim czasie t otrzymujemy równoległobok I, który zachowuje to samo pole co prostokąt 0. Występuje wówczas odbicie od pierwszej ściany i zmiana pędu p, co prowadzi do figur II i III; nowe odbicie od ściany $-l$ przyniesie z powrotem powierzchnię Liouville'a do górnego pasa. W czasie ruchu powierzchnia ta staje się coraz węższa i ulega rozdzieleniu na wiele warstw górnych i dolnych, jak pokazano na rys. 21.2. Nasza technika doświadczalna nie może być tak dokładna, aby wykrywać i rozróżniać wszystkie warstwy, i po chwili można powiedzieć, że tylko reprezentatywne punkty leżą w prostokątach $-l \leqslant x \leqslant +l$ oraz p, $p+\Delta p$ lub $-p$, $-p-\Delta p$. Co więcej, średnia gęstość tych punktów jest stała w obu prostokątach. Ten ostatni wynik jest postacią twierdzenia ergodycznego.



Rys. 21.2

W tym ogólnym dowodzie konieczne były następujące podstawowe założenia:

(1) warunki początkowe podano z nieuniknionym błędem Δx, Δp;

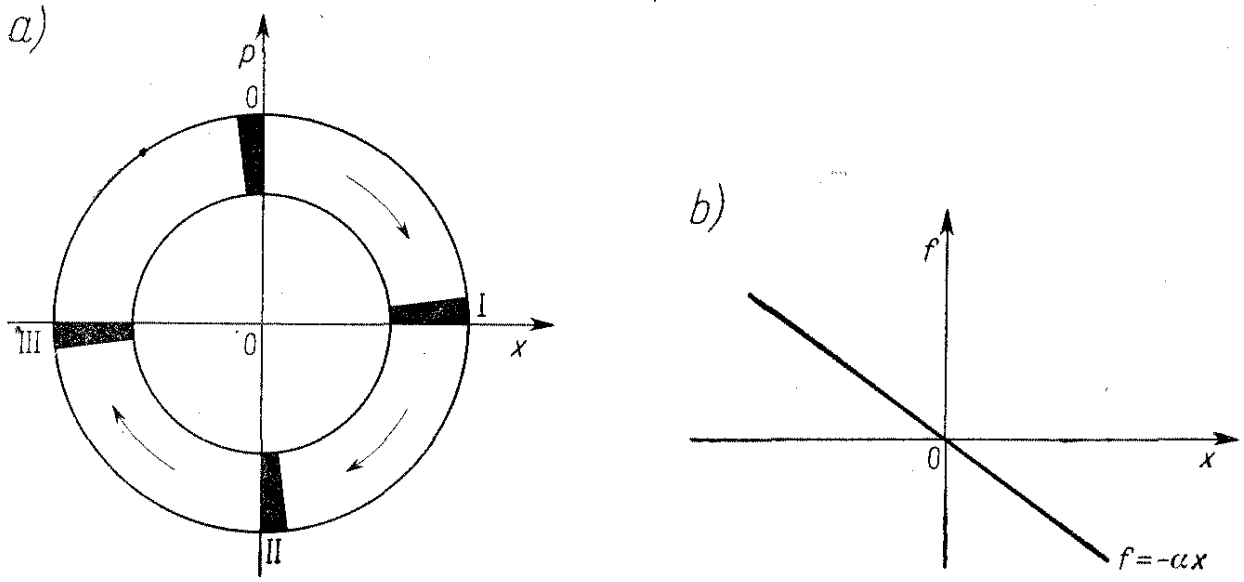
(2) stan końcowy obserwowany jest również z błędami Δx, Δp.

Oba te założenia są prostą konsekwencją zasady nieoznaczoności oraz warunków (21.1) i (21.2), jeżeli mamy na uwadze fakt, że ilość dostępnej energii

* E. Borel, *Mecanique statistique classique*, Paris 1925, Gauthier-Villars , str. 22.

w laboratorium (częściowo degradowanej w czasie eksperymentu) jest skończona. Ogranicza to przyrost entropii ΔS do pewnej wartości maksymalnej i nie można zakładać, że błędy Δx oraz Δp są nieskończenie małe.

Warunki (21.1) i (21.2) odrzucają możliwość naukowego udowodnienia słuszności determinizmu. Znając początkowe warunki z pewnymi błędami,



Rys. 21.3

można przewidzieć przyszłość tylko dla ograniczonego przedziału czasowego. Nie wiemy dokładnie, co zajdzie później. Powyższe stwierdzenia uzasadniają mechanikę statystyczną.

Ponad sto lat temu Laplace wymyślił demona, któremu przypisywał znajomość dokładnego położenia i prędkości wszystkich atomów we wszechświecie. Demon ten mógł dokładnie obliczać przyszłą ewolucję całego świata (dokładny determinizm). Przytoczone rozważanie odrzuca ideę demona Laplace'a. Zasada nieoznaczoności i negentropijna zasada informacji czynią schemat Laplace'a całkowicie nierealnym. Jest to schemat sztuczny, który bardziej można przypisać poetyckiej wyobraźni niż nauce doświadczalnej.

21.6. INNE PRZYKŁADY; OSCYLATOR ANHARMONICZNY I PROSTOWNIK

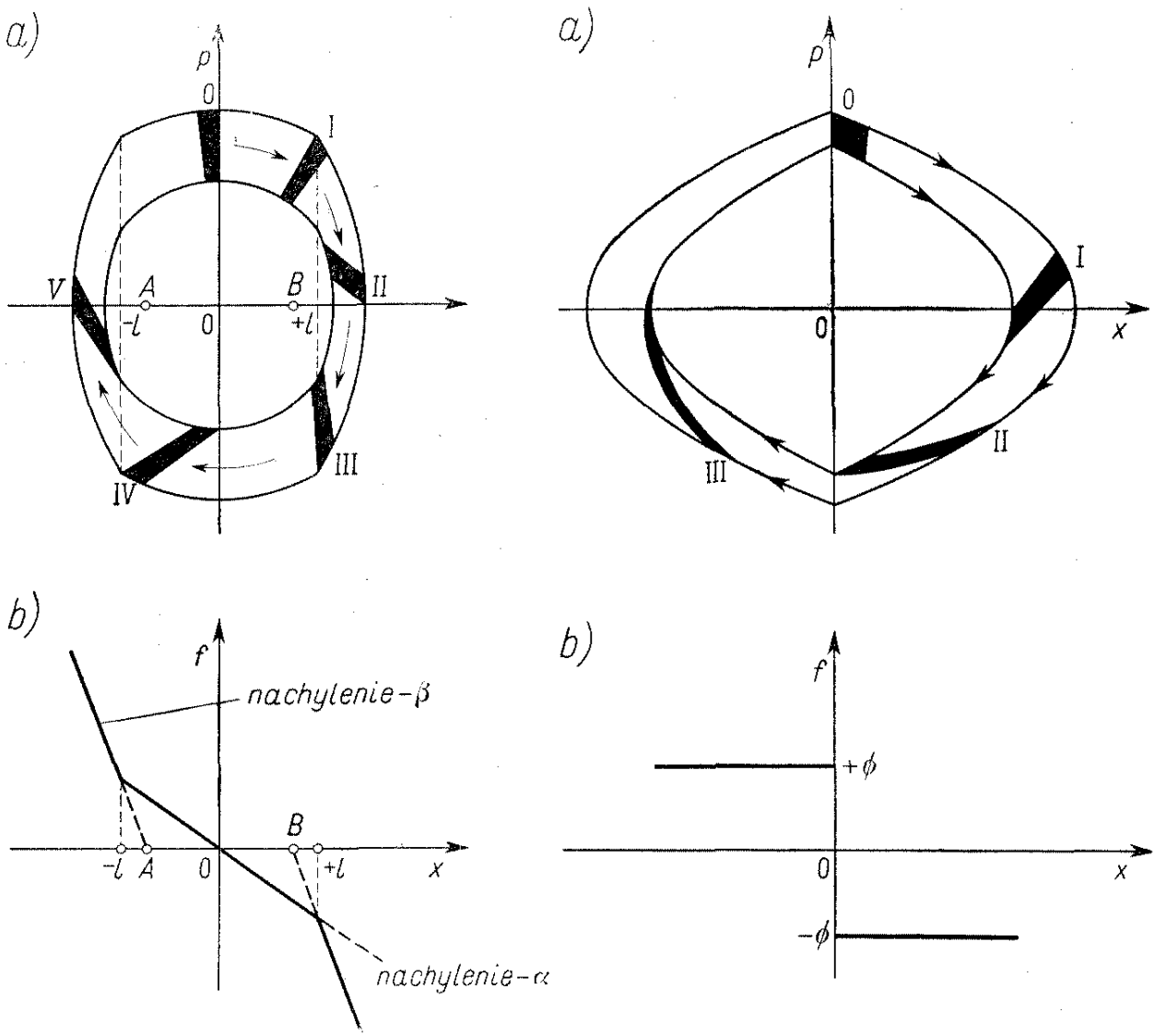
Celem poniższych rozważań jest dalsze uzasadnienie ważności poruszanego w tym rozdziale tematu.

Zacznijmy od oscylatora harmonicznego, którego trajektorie na płaszczy-

źnie x, p są elipsami:

$$\frac{1}{2m}p^2+\frac{1}{2}\alpha x^2 = E_c = \text{const.} \tag{21.4}$$

Jeżeli dokonamy zmiany jednostek, to elipsy można doprowadzić do postaci koła, np. dla $1/m = \alpha$.



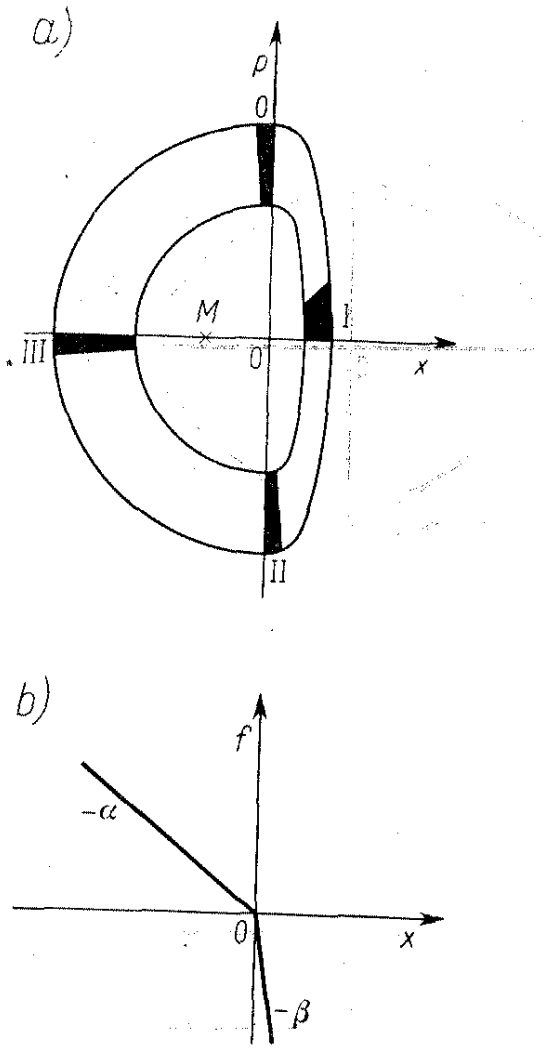
Rys. 21.4

Rys. 21.5

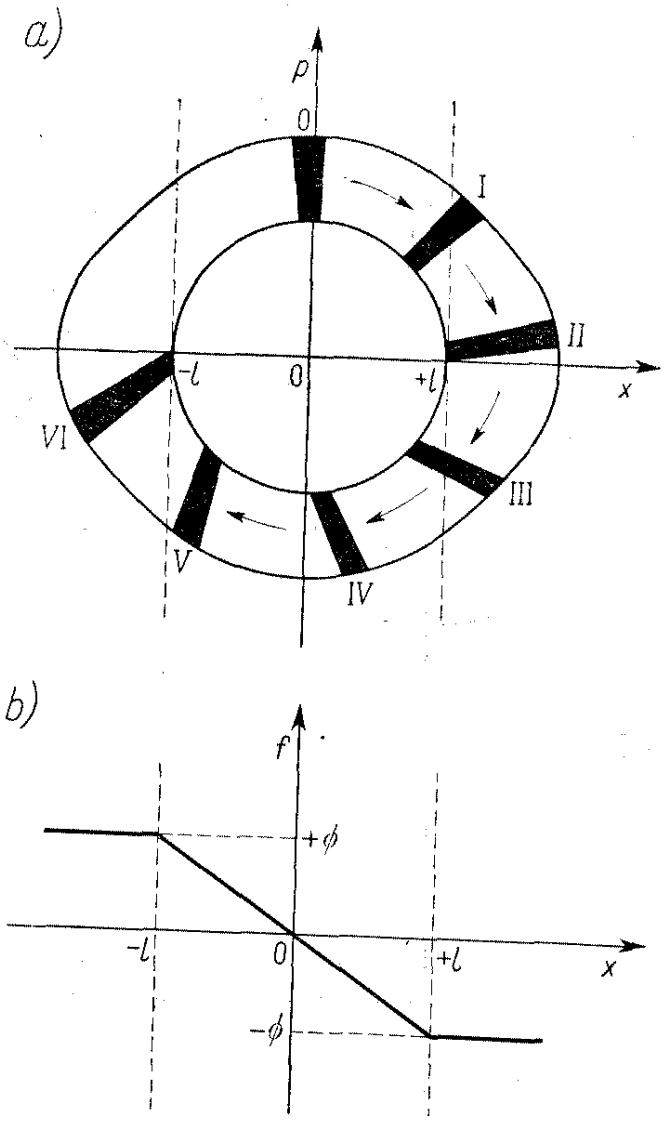
Oscylacje są izochroniczne i zachowują tę samą częstotliwość dla wszystkich amplitud. Wynikiem tego jest zachowanie nie tylko stałego pola, ale i kształtu powierzchni elementu Liouville'a*. Pokazano to na rys. 21.3, gdzie element

* Przez element Liouville'a rozumiemy obszar wyznaczony przez odcinki Δx, Δp we współrzędnych trajektorii x, p. (*Przyp. tłum.*).

ten występuje niezmieniony w punktach O, I, II i III, natomiast nie występuje tu rozciąganie się punktów reprezentatywnych modelu, co jest ciekawym wyjątkiem. Podobny przypadek można znaleźć dla wyidealizowanego prostownika (rys. 21.6). Skoro harmoniczny charakter drgań zostanie zakłócony,



Rys. 21.6



Rys. 21.7

to łatwo zauważyć, że obszar Liouville'a rozciąga się znowu. Przypadek pokazany na rys. 21.4 odpowiada wzrostowi sztywności dużych oscylacji. Dla tego typu oscylacji trajektorią jest elipsa (koło na rys. 21.4) o środku w punkcie O dla części środkowej, przechodząca z lewa i prawa w elipsy o środkach B i A. Na połączeniach obszar Liouville'a przyjmuje położenie nachylone względem promienia poprowadzonego z B lub A i nachylenie to wzrasta po każdym obrocie.

Zasadniczą sytuację pokazano na rys. 21.1 i rzeczywiście sprowadza się to do rys. 21.1, jeżeli wziąć $\alpha = 0$, a $\beta = \infty$.

Inną możliwość pokazano na rys. 21.5 z siłą przywracającą równowagę układu, która osiąga nasycenie $\pm\varphi$, gdy $|x|>l$. Trajektorie zewnętrzne są krzywymi parabolicznymi

$$\frac{1}{2m}p^2+\varphi x = E \tag{21.5}$$

i trajektorie te wykazują powolny ruch. Punkty zewnętrzne „wloką" się za punktami wewnętrznymi. Obszar Liouville'a ulega odkształceniu w przeciwnym kierunku niż na rys. 21.4. Rysunek 21.6 odpowiada pewnemu rodzajowi *prostownika* o małej sile przywracającej równowagę stanu $-\alpha x$ z lewej strony i bardzo dużej sile $-\beta x$ ze strony prawej. Trajektorie są elipsami centrowanymi względem O po obu stronach, drgania zaś pozostają synchroniczne dla wszystkich amplitud. Średnie położenie środka masy przesuwa się na lewą stronę, kiedy oscylacje narastają; odpowiada to efektowi prostowania. Nie występuje tu rozciąganie się obszaru Liouville'a.

Inny ciekawy przykład pokazano na rys. 21.7, który odpowiada przypadkowi granicznemu z rys. 21.5, gdy odległość $l = 0$. Trajektorie są dwustronnymi parabolami i zewnętrzna część obszaru Liouville'a mocno wlecze się z tyłu.

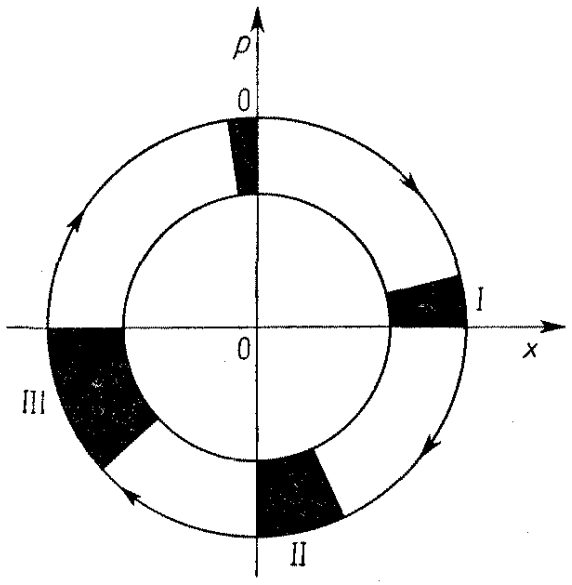
21.7. ANOMALIA OSCYLATORA HARMONICZNEGO

Sięgając wstecz do omawianych przykładów można zauważyć, że zasadnicza sytuacja przypomina sytuację omówioną w rozdz. 21.5. Obszar Liouville'a zachowuje stałą powierzchnię, ale w krótkim czasie tak się wydłuża, że pokrywa cały obszar możliwych ruchów. Jedyny wyjątek stanowi ruch dokładnie harmoniczny (rys. 21.3) lub bardziej sztuczny prostownik z rys. 21.6. W tych dwu przypadkach obszar Liouville'a zachowuje stały kształt i nie ulega zniekształceniu z upływem czasu. Czy jest to rzeczywista anomalia, czy odpowiada to sztucznej sytuacji?

Odpowiedź jest oczywista. Jest absolutnie nierealne mówić o oscylatorze harmonicznym z dokładnie określoną częstotliwością. Nigdy nie obserwujemy takiego przykładu. Drgania oscylatora zajmują węższe lub szersze pasmo.

Wiązać się to może z nieregularnymi fluktuacjami sił przywracających lub tłumieniem i wskutek tego nie ma absolutnie czystego prążka widmowego. W spektroskopii znaleziono co prawda kilka prążków dosyć czystych, które posłużyć mogą jako wzorce częstotliwości, ale nawet te wybrane prążki często mają skończoną i dającą się pomierzyć szerokość pasma. Z doświadczalnego punktu widzenia nie znamy pojedynczego prążka o zerowej szerokości pasma.

Teoretycznie rzecz biorąc, wszystkie oscylatory można sobie wyobrazić jako struktury fizyczne podlegające fluktuacjom, dla których warunki gra-



Rys. 21.8

niczne nie są absolutnie stałe. Nierealne jest mówić o dokładnych warunkach początkowych i nierealne jest również zakładać dokładną częstotliwość drgań. W takim razie częstotliwość ν określamy z błędem $\Delta\nu$. Przeczekajmy dostatecznie długi czas t i zobaczmy, jaka jest faza oscylacji. Zależność między czasem t i fazą φ jest następująca:

$$t = \left(n + \frac{\varphi}{2\pi}\right)\tau = \frac{1}{\nu}\left(n + \frac{\varphi}{2\pi}\right),$$

przy czym n jest całkowite, τ oznacza okres. Z powyższego wynika, że

$$\varphi = 2\pi(\nu t - n).$$

Zakładając, że t jest duże i ν dane z błędem $\Delta\nu$, stwierdzamy, że faza φ obarczona jest błędem

$$\Delta\varphi = 2\pi t\,\Delta\nu, \tag{21.6}$$

który dla dużych t może stać się tak duży, jak sobie tego życzymy. Sytuację tę zilustrowano na rys. 21.8. Twierdzenie Liouville'a stosuje się tylko wtedy, gdy siła przywracająca równowagę stanu jest stała w czasie; twierdzenie to nie dotyczy innego przypadku. Nie ma zachowania pola w elemencie powierzchniowym Liouville'a, który rozciąga się stopniowo, aż do pokrycia całkowitego obszaru możliwych ruchów. Nieoznaczoność częstotliwości przypomina tu znowu sytuację z innych przykładów. Dobrze znany jest fakt, że fluktuacje sił przywracających równowagę (stąd i fluktuacje częstotliwości) wynikają z różnych przyczyn fizycznych.

Obszerne badania wykazały, że rozmycia linii widmowych spowodowane są w rzeczywistości ruchami atomów (efekt döpplerowski) i zderzeniami atomów (skończone *ogony* fal).

Podsumowując zagadnienie należy zauważyć, że idealny oscylator (z $\Delta\nu = 0$) nadawałby się jako doskonały zegar i mógłby być wykorzystany do pomiaru czasu z nieskończenie dużą dokładnością, jednakże nieskończenie duża dokładność jest fizyczną niemożliwością.

21.8. ZAGADNIENIE DETERMINIZMU

Prawa mechaniki klasycznej są matematyczną idealizacją i nie należy ich traktować jako rzeczywistego odzwierciedlenia praw natury. Stosowanie ich w wielu zagadnieniach (np. astronomii) daje doskonałe rezultaty i obserwacje mieszczą się w granicach błędów eksperymentalnych. W innych zastosowaniach musiano je jednak skorygować (teoria względności, mechanika kwantowa).

Stary, klasyczny punkt widzenia pomijał rolę i znaczenie błędów doświadczalnych. Przyjmowano, że błędy są przypadkowe, a więc można je sprowadzać do tak małych, jak kto by sobie życzył, i wobec tego ostatecznie można je pomijać. Taki uproszczony obraz doprowadził do przyjęcia zupełnego determinizmu w mechanice klasycznej. Musimy jednak zrozumieć, że błędy doświadczalne są nieuniknione (por. rozdz. 21.4) i fakt ten sprawia, że ścisły determinizm jest niemożliwy. Błędy są nieodzowną częścią obrazu świata i muszą być uwzględnione w teorii.

Przyczynowość musi być zastąpiona prawdopodobieństwami statystycznymi; naukowiec może wierzyć lub nie wierzyć w determinizm. Jest to sprawa

wiary i należy do metafizyki. Rozważania fizyczne nie są w stanie tego udowodnić ani podważyć jego istnienia. Ten ogólny pogląd nazwać można stanowiskiem *rzeczowym*.

Max Born stawia tę sprawę jasno. Cytuje on Einsteina*, który mówi, że przed pojawieniem się mechaniki kwantowej zakładano, iż

«Wszystko da się sprowadzić do obiektów umieszczonych w czaso-przestrzeni i do ścisłych relacji między nimi. [...] Z niczym nie odwoływano się do naszej empirycznej wiedzy o tych obiektach. [...] Rozumiano pod tym fizyczny opis *rzeczywistego świata zewnętrznego*.»

Stanowisko takie jest nie do przyjęcia we współczesnej fizyce. Nie ma sposobu, aby wykazać istnienie takiego rzeczywistego świata, i niebezpiecznie jest mówić o czymkolwiek, czego nie możemy obserwować. Jeżeli rozumowanie ograniczymy do faktów dających się obserwować, możemy mówić o możliwych relacjach między pewnymi eksperymentami, ale nigdy nie należy rozważać tego, co się zdarzy w tym czasie, kiedy nie dokonujemy żadnej obserwacji. W takiej sytuacji musimy się przyznać szczerze, że po prostu niewiele wiemy (nie więcej, niż wiemy o tym, co zachodzi po drugiej stronie Księżyca).

Stanowisko takie przyjmuje Born; jest ono zgodne z filozofią nauki sformułowaną przez szkołę wiedeńską.

Czy pogląd taki zaakceptowali wszyscy fizycy? Odpowiedź jest daleka od jasności. „Czyści" matematycy mają duże trudności w pogodzeniu się z włączeniem błędów do teorii, a wielu fizyków-teoretyków nie jest jeszcze matematykami z prawdziwego zdarzenia.

Zasady nieoznaczoności Bohra i Heisenberga opierają się na sposobie rozumowania, jaki próbowaliśmy przedstawić. Jednakże, gdy ktoś śledzi dalszy rozwój teorii kwantowych, jest zdumiony z powodu wielu fantastycznych pomysłów opisywania fizyki terminami nieobserwowalnych tworów**.

Język fizyków obciążony jest żargonem i zrozumiały jest tylko dla specjalistów; wymyślano specjalne nazwy dla wyrazów szeregów aproksymujących, jak gdyby każdy oddzielny wyraz miał znaczenie (wymiana składników, powstawanie par, wirtualne powstawanie i absorpcja cząstek itp.).

W gruncie rzeczy znaczenie ma tylko końcowa suma. Mądrzy ludzie wiedzą, gdzie i jak posługiwać się tym językiem, i zdają sobie sprawę, że brak

* A. Einstein, *Scientific Papers presented to M. Born*, London 1953, Oliver and Boyd, str. 33.

** Mowa tu o demonach Maxwella i Laplace'a. (*Przyp. tłum.*).

jest w nim odzwierciedlenia rzeczywistości. Zdają oni sobie sprawę, że żargon jest tylko sztucznym sposobem opisywania skomplikowanych sytuacji. Jednak wielu fizyków może być wprowadzonych w błąd przez takie naprawdę niebezpieczne metody. Krótko mówiąc, teoria kwantów udaje, że służy zdrowej zasadzie *rzeczowych* opisów, ale szybko zapomina o tym i przechodzi na język bardzo niedbały.

Oprócz matematyków i teoretyków kwantowych wielu uczonych czuje wstręt do opisanej wyżej sytuacji i nie chcą oni zaniechać tradycyjnych idei. Wierzą ciągle w świat rzeczywisty i podlegający niezakłóconym ewolucjom, bez względu na to, czy je obserwujemy, czy też nie. Aby pogodzić taki pogląd z obecnymi odkryciami fizyki, mają oni nadzieję, że uda się im wynaleźć istnienie pewnej ilości „ukrytych zmiennych", których nie jesteśmy obecnie w stanie obserwować. Naszym zdaniem wprowadzenie takich ukrytych zmiennych może zrobić więcej szkody niż dobrego. Skoro zmiennych tych nie można obserwować — dopuśćmy twierdzenie, że zmienne te są nierealne i mogą istnieć tylko w wyobraźni autorów. Nie znaczy to, że jest to sarkazm. Nie można się obejść bez wyobraźni w badaniach naukowych, a wiele poważnych odkryć początkowo było czystym wytworem wyobraźni. Nabrały one znaczenia później, gdy doświadczenie pokryło się z wynikami przewidzianymi na drodze czystej wyobraźni. Ostatecznie nowe odkrycia doświadczalne stają się naukową bazą tej teorii, którą sprawdzono na drodze doświadczalnej.

Mechanika kwantowa doskonale opisała struktury atomowe i cząsteczkowe, ale nie zyskała powodzenia w teorii pola. Co się tyczy cząstek nowoodkrytych, są one jeszcze nie zgłębione i trudno przewidzieć teorię, która pomogłaby w ich zrozumieniu. Wiele nowych, odnoszących się do wymiarów mniejszych od 10^{-13}, praw trzeba będzie jeszcze odkryć. Prawa te dopiero stopniowo nabierają właściwego kształtu.

21.9. TEORIA INFORMACJI W ŚWIETLE OMÓWIONYCH PRZYKŁADÓW

Przykłady omówione w rozdz. 21.5 do 21.7 przeanalizujemy teraz z punktu widzenia teorii informacji. Położenie punktu na płaszczyźnie x, p znane jest początkowo z dokładnością co do obszaru nieoznaczoności o polu

$$A_0 = \Delta x_0 \, \Delta p_0 . \tag{21.7}$$

Jeżeli nie zmieniamy dokładności przyrządów pomiarowych, musimy pogodzić się w praktyce z tym, że obszar ten nie jest stały (co wykazał Liouville) i faktycznie stopniowo wzrasta. Zilustrowano to na rys. 21.1 ÷ 21.5, 21.7 i 21.8. Po długim okresie czasu obszar nieoznaczoności pokrywa praktycznie całą powierzchnię między zewnętrznymi i wewnętrznymi granicami energetycznymi, co widać z rys. 21.2. W takim razie pole A wzrasta, aż osiągnie końcową wartość A_∞ równą polu całego obszaru możliwych ruchów o energii E (z dokładnością ΔE.)

Informacja, jaką mamy o układzie, określona jest przez wzór (21.2). Liczba P_1 jest proporcjonalna do A i wzrasta od A_0 do A_∞. Liczba początkowa P_0 zależy od warunków doświadczenia. Załóżmy, że w jakimś układzie położenie może się zmieniać od $-L$ do $+L$, a pęd może przyjmować wartości od $-P_m$ do $+P_m$. Warunki takie prowadzą do wartości

$$P_0 = 4LP_m. \tag{21.8}$$

Przypomina to zagadnienie z rozdz. 20. Dla tego przykładu informacja, zgodnie ze wzorem (21.2), wyraża się przez

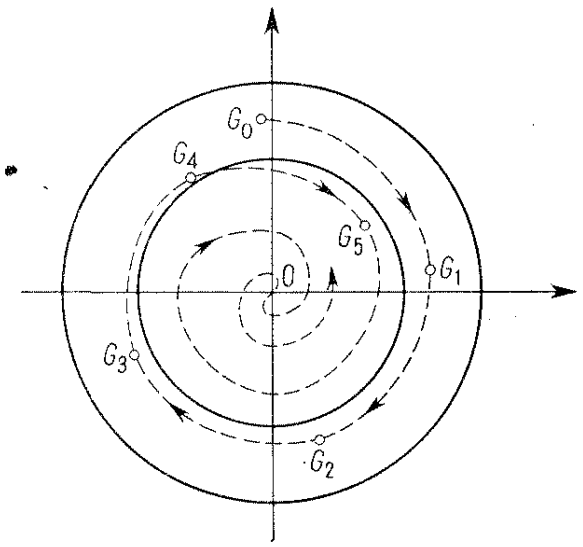
$$\Delta I = k(\lg P_0 - \lg P_1) = -k \lg A + C'; \tag{21.9}$$

we wzorze tym $k \lg P_0$ odgrywa rolę stałej addytywnej, która jest analogiczna do stałych w wielu wzorach na entropię. Z upływem czasu A wzrasta do A_∞ i informacja odpowiednio maleje. Naturalna ewolucja układu odpowiada stopniowej utracie informacji.

Rozważania nasze prowadzą bezpośrednio do prawa zmniejszania się informacji (wzrostu entropii), które wywołało tyle dyskusji od czasów Boltzmanna. Musieliśmy zrezygnować przy tym z wielu klasycznych założeń. Wykazaliśmy jednak niefizyczną naturę determinizmu i zachowaliśmy pogląd rzeczowy. W ten sposób otrzymaliśmy podstawowe prawo wzrostu entropii.

Możemy spojrzeć na to zagadnienie pod innym kątem i rozważać ruch środka ciężkości G powierzchni zawierającej punkt reprezentatywny. Pokazano to na rys. 21.9, który z grubsza odpowiada zagadnieniu z rys. 21.8. Kolejne położenia $G_0, G_1, G_2, \ldots$ opisują spiralę, która przechodzi przez początek układu O, gdy zaciemniona powierzchnia pokryje dokładnie (pierwszy raz) cały obszar między wewnętrznym i zewnętrznym kołem energetycznym. Gdy zaciemnione powierzchnie zachodzą na siebie, punkt G wyrusza z O i wraca znowu do O, gdy zaciemniona powierzchnia dwukrotnie pokryje pierścień; itd.

Jeżeli poruszające się cząstki są naładowane elektrycznie, to oscylacje środka ciężkości będą dostrzegalne dzięki promieniowaniu elektromagnetycznemu, a „ogon" fali będzie wykładniczo tłumiony z dudnieniami w chwilach



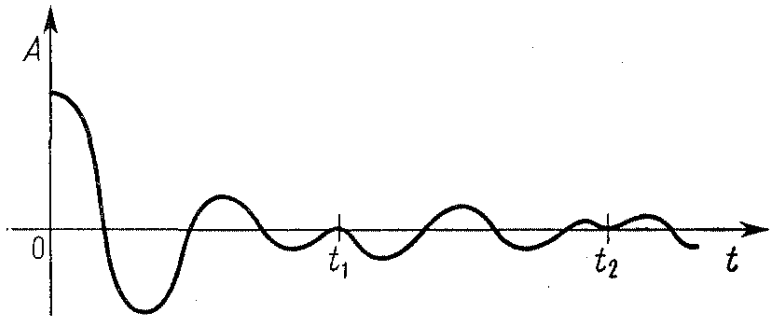
Rys. 21.9

t_1 oraz t_2, jak to pokazano na rys. 21.10. Po upływie długiego czasu oscylacje są tak małe, że nie można ich dłużej obserwować, i informacja zmaleje do swej najmniejszej wartości

$$\Delta I_\infty = C^t - k \lg A_\infty. \tag{21.10}$$

Rozważanie to pomaga naświetlić mechanizm strat informacji.

Dobrze znane argumenty przeciwko Boltzmannowi, wysunięte przez Loschmidta (odwracalność) czy też przez Zermelo i Poincarégo (rekurencja),



Rys. 21.10

można odrzucić jako odpowiadające niefizycznym założeniom, niemożliwym do zrealizowania w praktyce. Rozważmy na przykład odwracalność czasu. Cząstki emitowane są w momencie czasu $t = 0$ (z Δx oraz Δp), jak pokazano na rys. 21.1. W chwili t_1 wszystkie prędkości zostają nagle odwrócone

(jak to zrobić — to inne zagadnienie); cząstki wracają z powrotem i grupują się ponownie po czasie $2t_1$, odtwarzając pierwotny sygnał. Jeżeli można by tego rzeczywiście dokonać, to nieodwracalność zostałaby podważona. Ale podkreśliliśmy niemożliwość pomijania błędów doświadczalnych, a jest wiele miejsc, w których błędy te występują:

(1) Czas odwrócenia t_1 określany jest z błędem Δt_1, a więc sygnał jest odtwarzany po czasie $2t_1$ z błędem $2\Delta t_1$.

(2) Nie da się ściśle otrzymać odwrócenia. Cząstki pierwotnie poruszały się w kierunku x, a odbite cząstki będą miały małe składowe poprzeczne prędkości v_y, v_z. Powrócą one do początkowego punktu z przesunięciami $2t_1v_y$ oraz $2t_1v_z$ i błędy te wzrastają z t_1.

(3) Pole sił, przez które przechodzą cząstki, nie może być ściśle uniezależnione od czasu. Będzie ono wykazywać pewne fluktuacje (tak jak w zagadnieniu z rozdz. 21.7, rys. 21.7). Fluktuacje te unieważniają twierdzenie Liouville'a, a odwracalność czynią niemożliwą.

W ogólności koncepcja odwracalności w ruchach mechanicznych jest marzeniem i nie można jej sprawdzić wobec występowania takiej ilości błędów doświadczalnych, które stają się coraz większe, gdy odwrócony czas t_1 stopniowo wzrasta. W mechanice klasycznej kierunku czasu nie da się zmienić.

Tylko L. Hahn* wykrył znany przykład odwrócenia echa odbitego, o podobnych cechach, jakie wymieniliśmy tutaj. Echa stają się krótsze i słabsze, gdy czas odbicia wzrasta. Dokładna odwracalność i dokładne odwrócenie znajdują się poza ramami zagadnienia.

Potwierdzeniem powyższego wniosku jest dodatkowy fakt, że ruch odwrócony wykorzystywałby pola przyspieszające zamiast hamujących, a odwrócenie takie jest całkowicie niemożliwe (wniosek 5, rozdz. 21.10).

21.10. OBSERWACJA I INTERPRETACJA

Pogląd przedstawiony w tym rozdziale można porównać z dyskusją, jaka wywiązała się na konferencji w Brystolu**, gdzie filozofowie i fizycy porównywali opinie na temat obserwacji i interpretacji. Teoria należy do interpretacji

* E. L. Hahn, Phys. Rev. **80** (1950), str. 580.

** S. Körner, ed., *Observation and Interpretation. A Symposium of Philosophers and Physicists*, in: *Proc. Ninth Symposium Colston Research Society*, Bristol University, New York 1957, Academic Press.

doświadczeń, a całe zagadnienie sprowadza się do tego, jak dużo dodaje do faktów eksperymentalnych (i wtedy odpowiedzialna jest za niezrozumiałość) i jak odróżnić wytwory naszej wyobraźni od rzeczywistych obserwacji. Czy można wyeliminować element wyobraźni i zachować rzeczowy pogląd, którego broniliśmy poprzednio?

Przyjęliśmy w rozdz. 21.8, że teoria jest wytworem wyobraźni, budowanym początkowo na kilku dowolnych przypuszczeniach, a później porównywana jest z obserwacją. Jeżeli wyniki doświadczenia zgadzają się z teoretycznymi przewidywaniami, dowodzi to przydatności teorii, ale nie oznacza, że wszystkie pierwotne przypuszczenia (lub postulaty) są prawdziwe. Historia podaje liczne przykłady sprzeczności. Na przykład elastyczna teoria światła zakładała istnienie eteru o bardzo dziwnych własnościach. Zastąpiono ją elektromagnetyczną teorią Maxwella (przyjmującą jeszcze inny, również dziwny eter) i Einstein ostatecznie odrzucił eter jako rzecz bezsensowną. Obecnie przyjmujemy teorię kwantową, która na pewno kryje w sobie wiele nieobserwowalnych cech i z trudem uznać ją można za zakończony rozdział teorii. W tych kolejnych przeobrażeniach elementy wyobraźniowe były całkowicie odmienne, ale obszary pokrywane eksperymentami stawały się stopniowo coraz większe.

Teoria jest jakby mapą rzeczywistości i nie oznacza ona poprawnego i całkowitego opisu świata. Zaczynamy od planu mapy, później uwzględniamy wypukłość Ziemi i tak krok po kroku nanosimy charakterystyczne punkty, ale jest to zajęcie bez końca.

Po każdej fazie teorii niektórzy uczeni usiłują dowieść, że teoria jest już doskonała i nie można jej już w żaden sposób zmodyfikować. Jedną rzecz mogą tylko udowodnić, że teoria jest zgodna, logicznie zbudowana i zgadza się z materiałami eksperymentalnymi w pewnym ograniczonym obszarze. Wkrótce zostaną odkryte nowe, empiryczne fakty, które wymagają całkowitej zmiany teorii. Przypuszczalnie dowiedziono, że i teoria kwantowa jest ostatnim słowem*, ale dzisiaj ona nie jest w stanie wyjaśnić wszystkich dziwnych nowych cząstek, odkrywanych na bieżąco, nie mówiąc już o strukturach atomowych.

Na sympozjum w Brystolu L. Rosenfeld** omawiał podstawy teorii kwantowej i przedstawił pogląd szkoły kopenhaskiej, który jest ściśle związany

* Wiele rozważań na temat słynnego twierdzenia von Neumana można znaleźć w pracy: S. KÖRNER, *op. cit.*, str. 33, 47, 140; przedstawiają one przykład odbiegający od naszych uwag.

** S. KÖRNER, *op. cit.*, str. 41.

z ogólnym stanowiskiem przyjętym przez M. Borna i autora tej książki. D. Bohm, zwolennik innego poglądu, usiłował usprawiedliwić teorie przyjmujące zmienne nieobserwowalne i wywiązała się żywa dyskusja między obydwiema grupami. Wszyscy zgadzali się z tym, że dzieją się dziwne rzeczy z odległościami mniejszymi niż 10^{-13} cm, ale nie wyjaśniono tego jeszcze teoretycznie. Komunikat Rosenfelda winien być szczególnie polecony jako wyjątkowo przejrzysty i wskazujący na oszczędność w niesprecyzowanych postulatach. Dwa komunikaty konferencji w Brystolu (Feyerabend i Süssmann) poświęcono teorii pomiaru. Wydają się one bardziej interesujące od strony logicznego formalizmu niż od rzeczywistej strony eksperymentalnej. Ciekawe jest zwrócenie uwagi na fakt, że w komunikatach tych błędy eksperymentalne są praktycznie pominięte. Pierwszy opiera się na formalnej analizie von Neumanna, podczas gdy drugi podkreśla rolę fal Schrödingera i omawia *skoki kwantowe** powodowane sprzężeniem obserwowanego układu z przyrządem pomiarowym. Wydaje się, że wiele spośród uwag poczynionych w stosunku do tych komunikatów jest raczej nieuzasadnionych, gdyż pomijają one błędy doświadczalne, szumy tła w obserwacjach, szumy wzmacniaczy itp. Ekspertom od teorii informacji wiadome jest znaczenie tych czynników, które całkowicie przeoczono na sympozjum w Brystolu. Można by wyciągnąć wiele istotnych punktów z konferencji w Brystolu, odnoszących się do *naukowej predykcji* i opisu teorii kwantowej, ale wydaje się, iż dyskusja taka odciągnęłaby nas daleko od głównego tematu badań. Zasugerujmy czytelnikowi przestudiowanie tych interesujących komunikatów.

21.11. WNIOSKI

Krótki wniosek może okazać się pożyteczny i podsumuje cały ten rozdział**.

(1) Błędy doświadczalne są nieuniknione, a zakładanie nieskończonej dokładności jakiegoś pomiaru jest nienaukowe. Ograniczenie to podają już teorie klasyczne (M. Born), a sformułowano je precyzyjnie poprzez warunki kwantowe i teorię informacji (rozdz. 21.4).

* Tzw. *quantum jumps*. (*Przyp. tłum.*).

** Temat poruszony w tym rozdziale przedstawiono najpierw w Inf. and Control, **1** (1957), str. 1—5; **2** (1959), str. 45—63.

(2) Niektóre przykłady omówiono w rozdz. 21.5 i 21.7; wskazują one na rodzaj nieoznaczoności, nawet w modelach klasycznych.

(3) Nieuniknione błędy muszą być włączone do teorii, ponieważ są one zasadniczą częścią wiedzy o otaczającym nas świecie. Odrzuca to możliwość istnienia *dokładnego determinizmu* w przewidywaniu naukowym. Demon Laplace'a nie może dokonywać dokładnych pomiarów i nie jest on w stanie przewidzieć dokładnie przyszłości (rozdz. 21.5 i 21.8). Stanowisko takie nazwaliśmy rzeczowym punktem widzenia.

(4) Naturalna przemiana systemu odpowiada *utracie informacji*; stąd przyrost entropii (rozdz. 21.9). Podobne zagadnienia omówiono na sympozjum w Brystolu.

(5) Przypuszczenie o odwracalności praw mechaniki jest słuszne tylko wtedy, jeżeli pomijamy błędy i nieoznaczoności w czasie doświadczenia. Niemożliwość odwracalności czasu podkreślano dalej przez fakt, że zmiana znaku prędkości falowych zamieniłaby potencjały hamujące (lub pola) na potencjały przyspieszające, a taka zamiana jest fizycznie niemożliwa.

Znaczenie tego punktu widzenia podkreślili: L. Fantappié, G. i S. Arcidiacono i O. Costa de Beauregard*.

---

* L. Fantappié, *Principi di una teoria unitaria*, Humanitas Nova, Roma (1944), str. 90.

G. S. Arcidiacono, *Spazio, Tempo, Universo*, Roma 1961, Ed. Fuoco; i wcześniejsze artykuły cytowane w tej książce.

O. Costa de Beauregard, Revue de Synthese (1957), str. 7, Cahier de Physique, **12** (1958), str. 317; C. R. Acad. Sc. Paris **251** (1960), str. 2484.

22

ZAGADNIENIE SKRAJNIE MAŁYCH ODLEGŁOŚCI

22.1. TRUDNOŚCI W MIERZENIU SKRAJNIE MAŁYCH ODLEGŁOŚCI

W rozdziałach 16.4 i 16.5 omówiliśmy ważne zagadnienia występujące przy pomiarach bardzo małych odległości. Trudność wynika z tego powodu, że przy bardzo małych odległościach nie możemy rozważać stosowania sztywnej miarki, a musimy stosować promieniowanie o odpowiedniej długości fali*. Najbardziej korzystne warunki pojawiają się wtedy, gdy długość fali λ wynosi około $2\Delta x$. Odpowiada to częstotliwości ν i kwantowi energii

$$\Delta E_2 = h\nu = \frac{hc}{2\Delta x}. \tag{22.1}$$

Zakładamy, że w czasie doświadczenia posługujemy się światłem (falami elektromagnetycznymi), choć w istocie rzeczy inny rodzaj fali prowadziłby do tych samych wniosków. Kwant energii ΔE_2 przedstawia najmniejszą ilość energii, która jest niezbędna w doświadczeniu, aby pomierzyć odległość Δx. Energia ta nie musi całkowicie zmieniać się w ciepło. Nie ma wyraźnych granic małych odległości, jakie można mierzyć, ale strata entropii (lub energii) ogromnie wzrasta, kiedy odległości te stają się naprawdę małe.

Często sugerowano, iż wiele trudności związanych z teorią kwantową można by wyeliminować wprowadzając jakiegoś rodzaju minimalną długość. Poprzednie uwagi pozwalają przypuszczać, że ścisłe określenie takiej mini-

---

* L. Brillouin, Annals of Physics **5** (1958), str. 243; Physikertagung Essen **11** (1958), str. 140—151.

malnej długości byłoby trudne. Zwraca się uwagę na wzrost trudności w mierzeniu coraz to mniejszych odległości oraz na to, że wzrastający koszt obserwacji może być nowym parametrem w teorii. Ponieważ entropia zawsze wiąże się z prawdopodobieństwem, wydaje się, że bardzo małe odległości i czasy powinny pojawiać się z bardzo małym prawdopodobieństwem.

Nasuwa się przykład ilustrujący znaczenie warunku (22.1). Wprowadźmy długość ξ mierzoną w jednostkach Fermiego (10^{-13} cm):

$$\left.\begin{aligned} x &= 10^{-13}\xi; \\ x^2 &= 10^{-26}\xi^2 \\ \Delta x^2 &= 10^{-2}\,\xi^2 \text{ barnów}, \end{aligned}\right\} \tag{22.2}$$

ponieważ

$$1 \text{ barn} = 10^{-24}\,\text{cm}^2.$$

Jeżeli energię ΔE_2 mierzymy w elektronowoltach (eV), to przy napięciu V woltów

$$\Delta E_2 = e_0 \frac{U}{300} \text{ eV}, \tag{22.3}$$

przy czym e_0 wyrażono w j. ES. ład. Ze wzoru (22.1) otrzymujemy

$$U\xi \geqslant \frac{300}{e_0}\frac{hc}{2} \cdot 10^{13} \approx 6{,}2 \cdot 10^8 \text{ V}. \tag{22.4}$$

Pomiar odległości rzędu kilku jednostek Fermiego lub przekroju rzędu ułamka barna wymaga energii setek milionów elektronowoltów.

Wzór (22.1) dobrze zgadza się z rzędem wielkości napięcia w woltach stosowanego w cyklotronach lub synchrotronach. Rzeczywiście, bardzo duże energie są nie do uniknięcia przy pomiarach bardzo małych przekrojów.

Podsumowując to, co powiedziano wyżej, widzimy, że pomiar skrajnie małych odległości wymaga ogromnego wydatku energii. Koszt takiego eksperymentu staje się wkrótce bardzo wysoki i może przewyższyć sumę, na jaką stać laboratorium, instytucję lub nawet duże państwo. To praktyczne ograniczenie powinno być wprowadzone do teorii i wydaje się słuszne, by dokonać tego przy pomocy dogodnego współczynnika probabilistycznego, który przedstawiałby również trudności uzyskania dużej ilości energii.

22.2. MOŻLIWOŚĆ WYKORZYSTANIA POPRZEDNICH UWAG DO OBLICZANIA CAŁEK ROZBIEŻNYCH WYSTĘPUJĄCYCH W ZAGADNIENIACH FIZYKI TEORETYCZNEJ

Całki rozbieżne występują w różnorodnych zagadnieniach teorii kwantowej i jeszcze nie znaleziono sposobu uwolnienia się od tej trudności. Podamy propozycję i wyjaśnimy na prostym przykładzie, jak można z tego wybrnąć.

Weźmy pod uwagę całkę, która jest rozbieżna ze względu na zachowanie się funkcji $f(r)$ dla bardzo małej wartości r:

$$I = \lim_{a \to 0} \int_a^\infty f(r)\, \mathrm{d}r. \tag{22.5}$$

Przy $a = 0$ całka I staje się nieskończenie mała. Wróćmy znów do wzoru (21.1), który powiada, iż potrzeba bardzo dużej ilości energii ΔE_2, aby móc mierzyć bardzo małą odległość r:

$$\Delta E_2 = \frac{hc}{2r}. \tag{22.6}$$

Wydaje się słuszne założenie, że energia taka może być związana ze współczynnikiem wagowym

$$P = \exp\left(\frac{-\Delta E_2}{\gamma E_0}\right) = \exp\left\{\frac{-hc}{2\gamma r E_0}\right\}, \tag{22.7}$$

gdzie γ jest współczynnikiem liczbowym.

W termodynamice klasycznej E_0 jest rzędu kT, gdzie T jest temperaturą We wzorze (22.7) nie ma sensu mówić o temperaturze i zachowujemy energię E_0. Dla $r = 0$ współczynnik korekcji $P = 0$, natomiast $P = 1$ dla dużych odległości. Wprowadzamy współczynnik wagowy P do rozbieżnej całki (22.5) zapisujemy ją w postaci

$$J = \lim_{a \to 0} \int_0^\infty \mathrm{e}^{-\beta/r} f(r)\, \mathrm{d}r, \qquad \beta = \frac{hc}{2\gamma E_0}. \tag{22.8}$$

Wykładnik staje się zerem dla $r \ll \beta$ i równa się jedności dla $r \gg \beta$; β jest rzędu wymiarów promienia elektronu. Dokonajmy zamiany zmiennych:

$$J = \lim_{a \to 0} \int_0^{1/a} e^{-\beta\varrho} g(\varrho)\, d\varrho, \quad r = \frac{1}{\varrho}, \quad g = \frac{1}{\varrho^2} f\left(\frac{1}{\varrho}\right). \tag{22.9}$$

Ostatnia całka jest teraz rozbieżna dla ϱ równego nieskończoności. Wzór ten można porównać z granicą Cauchy'ego, zdefiniowaną przez van der Pola i Bremmera*, która przedstawia najbardziej efektywne przejście graniczne Cesarò dla całki jak poniższa:

$$J_C = \lim_{\xi \to 0} \int_0^{\infty} e^{-\xi\varrho} g(\varrho)\, d\varrho. \tag{22.10}$$

Oba wzory są bardzo podobne, lecz dla $\xi \to 0$ we wzorze (22.10), β ze wzoru (22.9) przyjmuje małą skończoną wartość:

$$\beta = \frac{hc}{2\gamma E_0} \approx \frac{10^{-16}}{\gamma E_0} \quad \text{(w jednostkach CGS)}. \tag{22.11}$$

Mamy teraz zdecydować, jaką wartość przyjąć dla E_0, która na pewno przedstawia wielkość charakterystyczną dla cząstki znajdującej się przy $r = 0$. Można przyjąć, że

$$E_0 = m_0 c^2; \tag{22.12}$$

jest to energia spoczynkowa cząstki o masie m_0. Hipoteza ta prowadzi do wzoru

$$\beta = \frac{h}{2m_0 c}\frac{1}{\gamma} \approx \frac{1}{\gamma} \begin{cases} 1{,}21 \cdot 10^{-10} \text{ dla elektronu } (m_0 = 0{,}91 \cdot 10^{-27}), \\ 0{,}66 \cdot 10^{-13} \text{ dla nukleonu } (m_0 = 1{,}67 \cdot 10^{-24}), \end{cases} \tag{22.13}$$

gdzie $1/\gamma$ jest współczynnikiem liczbowym. Długość β dla elektronu jest prawie $1/2\gamma$ razy większa od długości fali Comptona. Wartości β nie są nierozsądne; związane są z nimi poprawki w obliczaniu całki (22.5) dla r rzędu β lub mniejszego.

Podobieństwo między naszym wzorem (22.9) a granicą w sensie Cesarò – Cauchy'ego jest godna uwagi. Jak wiadomo, jeżeli przejście graniczne Cesarò

* B. Van der Pol, H. Bremmer, *Operational Calculus*, Cambridge 1955, Cambridge Univ. Press., str. 103.

zastosujemy do szeregu zbieżnego, to otrzymamy taki sam wynik jak przy sumowaniu zwykłym. W tym przypadku wzór (22.9) zmieniłby nieco wartość całki, ale byłaby to zmiana niewielka.

22.3. PRZYKŁAD: MASA ELEKTROMAGNETYCZNA ELEKTRONU

Zastosujmy poprzednią metodę do obliczenia klasycznej masy elektromagnetycznej cząstki naładowanej, mającej masę mechaniczną m_0. Energia elektrostatyczna wyraża się wzorem

$$E_{st} = \frac{1}{8\pi}\int_0^\infty e^{-\beta/r}E^2\,dr = \frac{1}{8\pi}\int_0^\infty e^{-\beta/r}\left(\frac{e_0}{r^2}\right)^2 4\pi r^2\,dr =$$

$$= \frac{e_0^2}{2}\int_0^\infty e^{-\beta/r}\frac{dr}{r^2} = \frac{e_0^2}{2\beta}, \tag{22.14}$$

gdzie E jest natężeniem pola elektrycznego wytwarzanego przez ładunek e_0, natomiast $e^{-\beta/r}$ jest współczynnikiem wagowym. Wynik powyższy przypomina wzór na energię elektrostatyczną cząstki o promieniu β.

Obliczmy teraz związany z energią E_{st} przyrost masy δm i zastąpmy β przez wartość ze wzoru (22.13). Otrzymujemy

$$\left.\begin{aligned} E_{st} &= c^2\,\delta m, \\ \frac{\delta_m}{m_0} &= \frac{\gamma e_0^2}{hc} = \frac{\gamma\alpha}{2\pi} = \frac{\gamma}{2\pi\cdot 137}, \end{aligned}\right\} \tag{22.15}$$

gdzie α jest tzw. *stałą subtelnej struktury*. Rząd wielkości $\delta m/m_0$ jest ściśle związany z wynikiem Schwingera dotyczącym efektu Lamba–Rutherforda. Schwinger* przytacza wzór

$$\frac{\delta_m}{m_0} \approx \frac{2\pi e_0^2}{hc} = \frac{1}{137}; \tag{22.16}$$

jest on identyczny z naszym wzorem (22.15), jeżeli przyjmiemy $\gamma = 2\pi$. Inni autorzy wykorzystują wzory na nieskończoności logarytmiczne.

* SCHWINGER, Phys. Rev. **73** (1948), str. 146.

22.4. OBRONA NASZYCH ZAŁOŻEŃ: RUCH DRŻĄCY SCHRÖDINGERA*

Założyliśmy poprzednio, że współczynnik wagowy P ze wzorów (22.7) i (22.8) można traktować jako współczynnik prawdopodobieństwa związany z właściwościami elektronu w otoczeniu jego środka**. W dalszym ciągu można wykazać, że tego rodzaju prawdopodobieństwo związane jest z istnieniem pewnego rodzaju ruchu Browna dla elektronu. Wartość średnia $\overline{r^2}$ przy tego rodzaju ruchu Browna jest rzędu

$$\beta^2 = \frac{h^2}{4\gamma^2 m_0^2 c^2}. \tag{22.17}$$

Wynik ten można interpretować jako rodzaj ruchu Browna środka elektronu wokół jego arbitralnie określonego położenia spoczynkowego $r = 0$. Faktycznie środek elektronu nie jest w spoczynku, ale porusza się przypadkowo wokół położenia mającego sens położenia średniego i nie będącego położeniem spoczynkowym. Wzory przytoczone tutaj są identyczne z tymi, które otrzymał Schrödinger z bardzo wnikliwego rozważania równania elektronu podanego przez Diraca. Schrödinger odkrył, że elektron pozostaje w pewnym ruchu, który określił jako *ruch drżący* (niem. *Zitterbewegung*). Średnia wartość kwadratu amplitudy $\overline{|\xi|^2}$ składowej tego ruchu w kierunku osi x wynosi

$$\overline{|\xi|^2} \leqslant \left(\frac{h}{4\pi m_0 c}\right)^2. \tag{22.18}$$

Odpowiada to faktowi wykrytemu przez samego Diraca, że prędkość chwilowa elektronu równa jest prędkości światła c, podczas gdy średnia prędkość pozostaje skończona i wynosi zero dla elektronu w spoczynku.

Wartość średnia $\overline{r^2}$ powinna być trzy razy większa, jeżeli ruchy są niezależne we wszystkich trzech kierunkach x, y, z, a więc

$$\overline{r^2} \leqslant \frac{3h^2}{16\pi^2 m_0^2 c^2}. \tag{22.19}$$

Porównując wzory (22.17) i (22.19) widzimy, że oba te wyniki są bardzo

* Autor posługuje się określeniem oryginalnym — *Zitterbewegung*. (*Przyp. tłum.*).

** W dalszym ciągu rozdz. 22.4 zostanie dokładniej wyjaśniony sens pojęcia *otoczenie środka elektronu*. (*Przyp. tłum.*).

podobne i stają się identyczne, jeżeli przyjmiemy

$$4\gamma^2 \geqslant \frac{16}{3}\pi^2, \qquad \gamma \geqslant \frac{2\pi}{\sqrt{3}}. \tag{22.20}$$

Rzędy wielkości są całkowicie zgodne.

Równania Diraca zawężają elektron do punktu, ale ruch drżący rozciąga ten elektron na całą objętość V. Środek elektronu przedstawia tylko średnie położenie „elektronu punktowego".

22.5. OMÓWIENIE I MOŻLIWOŚĆ UOGÓLNIEŃ

Sytuacja, jaką opisaliśmy w tym rozdziale, prowadzi nas do dalszych rozważań.

(1) Po pierwsze: jak należy rozumieć ruch drżący elektronu? Ruch Browna jakiejś cząstki w cieczy wynika z przypadkowych zderzeń tej cząstki z cząsteczkami cieczy. Możemy założyć, że ruch drżący elektronu powstaje na skutek pędów spowodowanych przechodzeniem elektronu w stany mas ujemnych. Wynika to oczywiście z omówionego przez Schrödingera równania Diraca. Jąk wiele z tego ruchu może pozostać, jeżeli przejścia do mas ujemnych są wzbronione przez sztuczne założenie, że wszystkie ujemne stany energetyczne są zajęte? Co dalej? — możemy zapytać. Dlaczego nie włączono tu innych zjawisk, takich jak powstanie i absorpcja pary fotonów i powstanie nowej cząstki o dziwnych właściwościach? Wszystko to może odegrać ważną rolę przy energiach rzędu m_0c^2. Istnieje wielka różnorodność wzajemnych oddziaływań, które są w stanie wywołać ruch drżący, natomiast równanie Diraca na pewno tłumaczy tylko ograniczoną część tych efektów.

(2) Następnym punktem jest zagadnienie dotyczące tego, że elektron nie jest wyjątkiem, ale tylko pierwszym przykładem ogólnej zasady. Heisenberg uogólnił równania Diraca na wszystkie cząstki elementarne. Uogólnienie to opiera się o macierze Diraca, ale zawiera już nowe nieliniowe składniki, które nadają równaniom Heisenberga charakter podobny raczej do równań pola samozwartego. Ruch drżący elektronu wynika z macierzy Diraca i zapewne zostanie odkryty dla innych cząstek, nie wiemy jednak, ile spośród brakujących w równaniach Diraca nowych efektów udało się uwzględnić z powodzeniem Heisenbergowi.

Należy się spodziewać, że na pytanie to najbliższa przyszłość da odpowiedź. Jest rzeczą godną uwagi, że wiele zagadnień współczesnej fizyki stawało przed nami znowu, gdy omawialiśmy rolę teorii informacji i znaczenie błędów w obserwacjach doświadczalnych. «Nigdy nie mówimy o wielkościach, których nie można mierzyć» — powiedział P. W. Bridgman. Odległości w pobliżu cząstki nie można dokładnie pomierzyć i naszym zdaniem powód tego leży w tym, że nie możemy utrzymać cząstki w spoczynku.

Fizyka klasyczna opierała się na założeniach, że błędy doświadczalne były przypadkowe i powinny być pominięte w teorii. Współczesna fizyka rozumie, że błędy są nieuniknione i niemożliwe jest posuwanie się do granicy błędów nieskończenie małych. Zasada nieoznaczoności i negentropijna zasada informacji potwierdzają to, że im mniejszy błąd, tym wyższa jest cena, jaką musimy płacić za obserwacje. Nie ma ścisłej granicy dokładności, lecz cena osiągnięcia tej granicy czyni ją już nieosiągalną.

W fizyce klasycznej zakłada się pełny determinizm. Nowoczesny rzeczowy punkt widzenia stwierdza, że niemożliwe jest udowodnienie tego założenia. Jeżeli błędy pomiarowe są bardzo małe, to wiąże się z tym dłuższy czas osiągnięcia końcowego rozkładu statystycznego, ale ten końcowy stan równowagi statystycznej będzie zawsze osiągalny. Wiele przykładów można przytoczyć, natomiast rzeczowy punkt widzenia można porównać z podobnymi ideami przedstawionymi przez szkołę wiedeńską M. Borna i przez szkołę kopenhaską.

PRZYPISY

([1.1]) Ściśle biorąc, nie ma sensu mówić o jednostce dla wielkości bezwymiarowej, gdyż jest to 1. Jednakże w przypadku, gdy interesująca nas wielkość jest logarytmem pewnej innej bezwymiarowej wielkości, na ogół wybór podstawy logarytmów jest w pewnej mierze rzeczą umowną. Wobec tego dogodnie jest obok liczby będącej logarytmem podawać symbol informujący, jaką podstawę logarytmów wybrano. Tego rodzaju sytuacja od dawna występuje w telekomunikacji, gdzie wprowadza się logarytm stosunków wzmocnień lub mocy. Jeśli wybrać logarytmy naturalne, to obok liczby określającej logarytm naturalny stosunku mocy pisze się „jednostkę" *neper*, w przypadku zaś stosowania logarytmów dziesiętnych — „jednostkę" *decybel*. Taka sama sytuacja ma miejsce w przypadku ilości informacji.

([1.2]) Powiązania pomiędzy informacją a entropią stosowaną w termodynamice będą szczegółowo omawiane w rozdz. 12.

([1.3]) W terminologii stosowanej na stronicach 22–24 chodzi tu o sytuację, gdy pierwotnie wiemy, że system może się znajdować w jednym spośród P_0 stanów, natomiast po dokonaniu obserwacji wiemy, iż może się on znajdować w jednym spośród P_1 stanów, przy czym $1 \leqslant P_1 \leqslant P_0$.

([1.4]) Oznacza to, że każda spośród G cyfr może przyjmować jedną spośród N postaci.

([1.5]) Słowo „informacja" użyte tu jest w sensie potocznym.

([1.6]) W terminologii wprowadzonej na stronicach 35–36 „słowo" odpowiada stanowi, natomiast zbiorowi stanów odpowiada zbiór wszystkich sensownych w danym języku słów. Zatem wprowadzone przez autora pojęcie „ilość informacji zawartej w słowie" ma sens ilości informacji uzyskiwanej dzięki temu, że poznajemy konkretne słowo w sytuacji, gdy pierwotnie znamy tylko zbiór możliwych słów.

([1.7]) Zagadnienia zastosowań teorii informacji do badania języka poruszane były przez wielu autorów. W polskim piśmiennictwie zagadnienia te poruszane są w szczególności w książkach: M. A. Jagłom, I. M. Jagłom, *Prawdopodobieństwo i informacja* (Warszawa 1963, KiW) oraz J. Seidler, *Teoria kodów* (Warszawa 1965, PWN), rozdz. I.

([1.8]) Pod nazwą „słowo" rozumiemy tu ogólnie zespół liter mający sens w danym języku.

Szczególnym przypadkiem tak ogólnie rozumianego „słowa" jest zarówno słowo, jak i zdanie w sensie gramatyki klasycznej.

([1.9]) Dla języka polskiego rozkłady te można znaleźć w pracy: T. EKIERT, J. SEIDLER, *Badanie struktury słów języka polskiego*, Prace IPPT, Zakład Łączności, 1961, praca nr 11.

([1.10]) Pod nazwą „litera" rozumiemy tu ogólnie element, z którego budowane są słowa w sensie ogólnym. Niektórzy autorzy literę w sensie ogólnym nazywają *wiadomością elementarną*, a słowo w sensie ogólnym — *wiadomością* czy też *informacją*. Na przykład litera w sensie ogólnym może być cyfrą w sensie potocznym; słowem w sensie ogólnym jest wówczas ciąg cyfr mający np. sens liczby zapisanej za pomocą cyfr.

([1.11]) Rozwój badań w ciągu ostatniego dziesięciolecia potwierdził tę tezę autora. Okazało się, że powiązanie pomiędzy ilością informacji zdefiniowaną wzorem (1.13) a wartością informacji uzyskiwanej dzięki obserwacji (czy też wyborowi), mierzoną za pomocą funkcji mających sens strat (czy też ryzyka) — są nieliczne i nie mają większego znaczenia praktycznego (zob. np.: J. SEIDLER, rozdz. VI książki cytowanej w notce ([1.7])). O wiele bardziej perspektywiczne wydają się inne miary zmniejszenia pierwotnego zakresu wchodzących w grę możliwości, jak np. dyspersja (zob. J. SEIDLER, *Optymalne wydobywanie informacji i sterowanie w sytuacjach niedeterministycznych*, w pracy zbiorowej *Współczesne metody statystyczne w teorii sterowania*, Wrocław 1968, Ossolineum).

([1.12]) Stwierdzenie to słuszne jest przy założeniu, że interesujące nas stany (czy zdarzenia) można traktować jako realizacje wielkości przypadkowych w sensie rachunku prawdopodobieństwa, a więc że istnieją i są znane ich rozkłady prawdopodobieństw. Jak wiadomo, taki punkt widzenia, zwany bayesowskim, budzi często wątpliwości (zob. np. praca J. SEIDLERA cytowana w notce ([1.11])).

([2.1]) Jak wykazał J. SEIDLER w rozdz. VI książki cytowanej w notce ([1.7]), własność (B) jest bardzo istotna. Jeśli postulować, iż miara ilości informacji ma własność (B), to wynika stąd ogólna definicja (1.13) ilości informacji.

([2.2]) Pojęcie „zmienna losowa" wyjaśnione jest w standardowych podręcznikach rachunku prawdopodobieństwa, zob. np.: M. FISZ, *Rachunek prawdopodobieństwa i statystyka matematyczna*, Warszawa 1958, PWN.

([2.3]) Pod pojęciem „informacja dla zmiennej" autor rozumie ilość informacji, jaką uzyskujemy obserwując konkretną wartość, którą przyjęła zmienna losowa w sytuacji, gdy pierwotnie znamy tylko prawdopodobieństwa wartości, jakie zmienna ta może przyjmować.

([2.4]) W sensie niezależności zmiennych losowych; zob. np. M. FISZ, książka cytowana w notce ([2.2]).

([2.5]) Ściślej: ilości informacji, jakie uzyskujemy poznając wartość zmiennej losowej y w sytuacji, gdy pierwotnie znamy tylko warunkowe prawdopodobieństwa wartości, jakie zmienna y przyjmuje.

([3.1]) Pod słowem „korelacja" autor rozumie ogólnie powiązania statystyczne pomiędzy zmiennymi losowymi czy też zdarzeniami przypadkowymi. Pojęcie to wiąże się jedynie luźno z pojęciem współczynnika korelacji czy też funkcji korelacji.

([3.2]) W polskim piśmiennictwie podobne zagadnienia ujmowane są w książkach cytowanych w notce ([1.7]).

(3.3) Empiryczne częstości występowania liter autor nazywa tu prawdopodobieństwami.

(3.4) W tym paragrafie wyrazy „słowo" i „litera" rozumiemy jak w sensie potocznym (por. notki (1.8) oraz (1.10).

(3.5) Doświadczenie można uprościć biorąc zamiast zdań słowa.

(3.6) Tego typu grę dla języka polskiego opisali T. Ekiert i J. Seidler w pracy cytowanej w notce (1.9).

(3.7) Rozważania w tej części mają charakter uwag na temat cytowanej pracy Shannona oraz rozważań wstępnych.

(3.8) *Kodem* nazywamy tu zasadę przyporządkowania słowom będącym ciągami liter wybieranych z pewnego alfabetu — telegramów będących ciągami zbudowanymi z symboli wybranych tak, by można je było przesyłać przez kanał czy też zarejestrować w pamięci. Zob. również hasło „kodowanie" w Wielkiej Encyklopedii Powszechnej PWN. W innych publikacjach, np. w książce J. Seidlera cytowanej w notce (1.7), stosowana jest nieco inna terminologia, a mianowicie: *słowo*↔*wiadomość*, *litera* ↔ *wiadomość elementarna*, *telegram* ↔ *ciąg kodowy*, *symbol* ↔ *sygnał elementarny*. W szczególnym przypadku symbole mogą być znów literami, a więc telegramy mogą być budowane z tych samych elementów co słowa. Przyporządkowanie telegramu konkretnemu słowu nazywamy *kodowaniem*, urządzenie zaś, które dokonuje kodowania — urządzeniem kodującym bądź krótko: *koderem*. Odtwarzanie słowa na podstawie telegramu nazywamy *dekodowaniem*, natomiast urządzenie dokonujące dekodowania — *dekoderem*.

(3.9) Autor zakłada tu, że kodowania słów dokonuje się przez przyporządkowywanie z osobna poszczególnym literom symboli czy też bloków symboli, tworzących następnie telegram. Jest to najprostszy sposób kodowania, który nazywamy *kodowaniem poszczególnych liter*. Jak to będzie wykazane, najogólniejszy i na ogół najbardziej sprawny sposób kodowania polega na tym, że *całym słowom* przyporządkowujemy od razu *całe telegramy*.

(3.10) Kod nazywamy *odwracalnym*, jeśli na podstawie telegramu przyporządkowywanego słowu można bezbłędnie odtworzyć to słowo.

(4.1) Ściślej: ilość informacji, jaką uzyskujemy poznając słowo w sytuacji, gdy pierwotnie znamy tylko prawdopodobieństwa wchodzących w grę słów.

(4.2) Zob. notka (3.8).

(4.3) Autor wprowadza tu istotne założenie, iż telegramy mają określony rozkład prawdopodobieństwa. Sytuacja taka ma miejsce wówczas, gdy źródło słów wybieranych z pewnymi prawdopodobieństwami dołączymy do urządzenia kodującego wytwarzającego telegramy.

(4.4) Jeśli do źródła słów dołączymy urządzenie kodujące (na którego wyjściu otrzymujemy telegramy) i zespół złożony ze źródła słów i urządzenia kodującego potraktujemy jako wtórne źródło wiadomości, to telegramy będą odgrywać wówczas rolę wtórnych słów. Jeśli natomiast litery, z których budujemy słowa, mogą być przesyłane bezpośrednio przez kanał, to słowa stają się telegramami. Wobec tego rozróżnienie pomiędzy słowem a telegramem nie jest istotne; podobnie jest z literą i symbolem. Zagadnienia budowy telegramów i słów można więc rozpatrywać łącznie.

(4.5) Wyrazy „słowo" i „litera" rozumiemy w tym rozdziale w sensie potocznym (por. notki (1.8) i (1.10)).

(4.6) Wynika stąd, że „wiadomość" (znaczenie w języku naturalnym) odpowiada „słowu",

o którym mowa była w poprzednich rozdziałach, natomiast „słowu" w języku naturalnym odpowiada „telegram" (por. notkę (4.4)).

(4.7) Przy założeniu, że każdej wiadomości odpowiada jedno słowo oraz że na podstawie słowa można odtworzyć jednoznacznie wiadomość (zrozumieć znaczenie słowa), kod jest odwracalny i częstość występowania danego słowa jest taka sama jak częstość występowania wiadomości, które dane słowo przenosi.

(4.8) Zgodnie z notką (4.5) wyrazu „słowo" używamy tu w sensie potocznym; w tej sytuacji słowo jest telegramem w sensie definicji podanej w notce (3.8); por. również z notką (4.4).

(4.9) Znakami nazywamy więc elementarne składniki, z których budowane są słowa w sensie ogólnym. Jest to więc w zasadzie synonim poprzednio wprowadzonej nazwy „litera" w sensie ogólnym.

(4.10) Jak poprzednio, symbolem nazywamy tu elementarny składnik telegramu. W przypadku gdy $n > v$, jako symbol musimy brać blok symboli pierwotnych.

(4.11) Wprowadzamy tu pojęcie bloku symboli jako zespołu przyporządkowywanego pojedynczej literze czy też blokowi liter. Ma więc miejsce następująca odpowiedniość: *litera* ↔ *symbol, znak* (blok liter) ↔ *blok symboli, słowo* ↔ *telegram*. Rozróżnienie pomiędzy literą, blokiem liter a słowem jest podobne jak pomiędzy symbolem, blokiem symboli a telegramem. Oba te rozróżnienia nie są istotne i mają charakter umowny — por. notka (4.4).

(5.1) W sensie potocznym.

(5.2) *Kodem binarnym* nazywamy kod, w którym telegramy budowane są z symboli binarnych, tj. symboli mogących przyjmować jedną z dwu postaci, np. 0, 1.

(5.3) Podobnego typu porównania można znaleźć w rozdz. 1 książki J. SEIDLERA cytowanej w notce (1.7).

(5.4) Literom rozpatrywanym poprzednio odpowiadają teraz słowa, zatem blokom symboli odpowiadają telegramy (por. notkę (4.11)).

(5.5) Zamiast o słowach można by tu mówić również o literach.

(5.6) Jest to zasada tzw. *modulacji delta*, która (zwłaszcza w ostatnich latach) zaczęła znajdować szerokie zastosowanie techniczne. Przegląd zagadnień związanych z modulacją delta można znaleźć w pracy: T. BARTKOWSKI, *Modulatory systemu delta*, Zeszyty Naukowe Politechniki Gdańskiej, Gdańsk 1963, nr 37.

(6.1) Cyfra binarna odgrywa tu rolę symbolu, którego ciągi tworzą telegram.

(6.2) Od czasu napisania książki nastąpił istotnie ogromny rozwój teorii kodów korekcyjnych. Rozwinęła się przede wszystkim teoria *kodów liniowych*, które są uogólnieniem omawianego tu kodu wykrywającego błąd. Odznaczają się one tym, że wszystkie *pozycje* dzielą się na *kontrolne* i *informacyjne*, a sygnał w pozycji kontrolnej jest wybierany tak, by suma obejmująca ten sygnał oraz część sygnałów z pozycji informacyjnych była parzysta. Szczególnie duże znaczenie praktyczne mają specjalne rodzaje kodów liniowych: kody *iteracyjne*, *rekurencyjne* i *cykliczne*. Wprowadzenie do zagadnień związanych z kodami liniowymi można znaleźć w książce J. SEIDLERA cytowanej w notce (1.7), szczegóły zaś o kodach cyklicznych w książkach: W. W. PETERSON, *Error Correcting Codes* (New York 1961) oraz А. П. УДАЛОВ, Б. А. СУПРУН, *Избыточное кодирование при передаче информации двоичными кодами* (Москва 1964). Zagadnienia kodów rekurencyjnych omawiają J. M. WOZENCRAFT i B. REIFFEN w pracy *Sequential Decoding*, New York 1962.

([8.1]) Zgodnie z ogólnym charakterem książki, rozważania tego rozdziału mają częściowo charakter heurystyczny i oparte są częściowo na argumentacji fizykalnej. Poruszane w tym rozdziale zagadnienia, ujęte w sposób sformalizowany, można znaleźć przede wszystkim w książce: A. A. ХАРКЕВИЧ, *Спектры и анализ* (Москва 1952) oraz w rozdz. VII monografii: J. SEIDLER, *Statystyczna teoria odbioru sygnałów* (Warszawa 1963, PWN).

([8.2]) Autor zakłada tu milcząco, że chodzi o drgania oscylatora harmonicznego (zob. rozdz. 9.7). Na ogół związek pomiędzy energią drgań a ich częstotliwością zależy od charakteru drgań i może być inny niż $\Delta E = n\,\Delta\nu$.

([8.3]) W tej części rozdziału, w przeciwieństwie do poprzednich, autor zakłada milcząco, że funkcja $f(t)$ może przyjmować wartości zespolone.

([8.4]) Wniosek ten wynika stąd, że $E^2 = \int_{-\infty}^{+\infty} f^*(t) f(t)\,\mathrm{d}t$.

([8.5]) Autor snuje tu rozważania oparte na odległych analogiach pomiędzy własnościami obserwowanych w warunkach makroskopowych sygnałów a własnościami drgań oscylatora harmonicznego, podlegającego prawom mechaniki kwantowej. Fakt, że mimo zupełnie innego charakteru zjawisk otrzymuje się podobnego typu zależności liczbowe, jest niewątpliwie ciekawy.

([8.6]) Autor zakłada w tej części rozdziału, że funkcja $f(t)$ jest funkcją rzeczywistą (zob. notka ([8.3])).

([8.7]) Sformułowanie to ma charakter założenia, gdyż na ogół, np. dla funkcji nieciągłych, jest ono w sposób oczywisty fałszywe.

([8.8]) Jest to założenie dość ogólnikowe, w którym autor omija jedno z trudnych zagadnień analizy widmowej polegające na tym, że nie istnieją, poza funkcjami identycznie równymi zeru, funkcje, które znikałyby na zewnątrz pewnego przedziału czasu i których widmo byłoby równe zeru na zewnątrz pewnego przedziału częstotliwości. Wynika to z twierdzenia podanego przez Wienera (zob. R. PALEY, N. WIENER, *Fourier Transforms in the Complex Domain*, New York 1934), które brzmi: jeżeli funkcja oraz jej widmo opadają na zewnątrz pewnych przedziałów szybciej niż wykładniczo, to funkcja jest identycznie równa zeru. W tej sytuacji o zagadnieniach własności iloczynu: czas trwania *razy* szerokość widma — można mówić tylko wprowadzając definicje tych wielkości nie wymagające zupełnie znikania funkcji czy też widma na zewnątrz pewnych przedziałów. Definicje tego typu są zestawione np. w rozdz. VII książki J. SEIDLERA cytowanej w notce ([8.1]).

([8.9]) Rozważania z rozdz. 8.8 mają charakter bardzo pobieżnego skrótu. Aby zdefiniować płaszczyznę ν, t, trzeba wprowadzić pojęcie sygnału analitycznego (zob. np. książka J. SEIDLERA cytowana w notce ([8.1]) oraz J. SEIDLER, *Współczesne metody optymalizacji systemów telekomunikacyjnych*, Warszawa 1965, WKiŁ).

([8.10]) Autor wprowadza tu funkcję autokorelacji przebiegów zdeterminowanych z uśrednieniem po czasie. Pokrywa się ona z najczęściej stosowaną funkcją autokorelacji dla procesów przypadkowych, uśrednieniem statystycznym tylko dla procesów ergodycznych (zob. np. rozdz. VII książki J. SEIDLERA cytowanej w notce ([8.1])).

([8.11]) Autor ma tu na myśli wariant twierdzenia Wienera–Chinczyna dla funkcji zdeterminowanych. Najczęściej podawana postać tego twierdzenia dotyczy związków pomiędzy funkcją statystycznej autokorelacji przypadkowego procesu stochastycznego a gęstością jego widma (zob. np. rozdz. VII książki J. SEIDLERA cytowanej w notce ([8.1])).

([8.12]) Definicje szumu białego można znaleźć np. w rozdz. VII książki J. SEIDLERA cytowanej w notce ([8.1]).

([8.13]) W przeciwieństwie do poprzednich rozważań, które dotyczyły funkcji zdeterminowanych, autor nawiązuje tu do zagadnień dotyczących procesów przypadkowych (zob. np. rozdz. VII książki J. SEIDLERA cytowanej w notce ([8.1])).

([8.14]) Wyczerpującą dyskusję powiązań pomiędzy zagadnieniami filtracji a analizą widmową można znaleźć w rozdz. 8 książki A. BLANCK-LAPIERRE, R. FORTET, *Théorie des fonctions aléatoires*, Paris 1953, Maison.

([8.15]) Podobnie jak zespoły liczb x_1, x_2, x_3 możemy *interpretować jako punkty* w przestrzeni trójwymiarowej z prostokątnym układem osi odpowiadających wektorom $\mathbf{d}_1, \mathbf{d}_2, \mathbf{d}_3$, tak samo zespoły liczb $y_1 = h_1 b_1 = h_2 b_2 = h_3 b_3$ mające sens częstotliwości możemy traktować jako punkty trójwymiarowej przestrzeni z prostokątnym układem osi współrzędnych odpowiadających wektorom $\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3$ z tym, że długości tych wektorów są odwrotnościami długości wektorów $\mathbf{d}_1, \mathbf{d}_2, \mathbf{d}_3$. Przestrzeń punktów o współrzędnych y_1, y_2, y_3 ma sens przestrzeni częstotliwości.

([8.16]) W przestrzeni częstotliwości (zob. notka ([8.15])).

([9.1]) Wstęp do zagadnień termodynamiki można znaleźć w języku polskim w podręcznikach: L. LANDAU, E. LIFSZYC, *Fizyka statystyczna*, Warszawa 1959, PWN (przekł. z ros.); W. WEIZEL, *Fizyka teoretyczna*, t. 2, cz. 1 i 2, Warszawa 1957, PWN (przekł. z niem.).

([9.2]) Wstęp do teorii oscylatorów kwantowych można znaleźć np. w podręcznikach cytowanych w notce ([9.1]).

([11.1]) W ciągu minionego dziesięciolecia ukazało się wiele podręczników traktujących o szumach termicznych w oporach, lampach i półprzewodnikach (zob. np. В. И. ТИХОНОВ, *Статистическая радиотехника* (Москва 1966), oraz R. KING, *Electrical Noise* (London 1966)).

([15.1]) Autor omawia w tym rozdziale powiązania pomiędzy pewnością, dokładnością a informacją dla kilku konkretnych metod pomiarowych. Jak wykazał J. SEIDLER w rozdz. VI książki cytowanej w notce ([1.7]), istnieją ogólne związki pomiędzy znormalizowanym błędem, pewnością a ilością informacji, które można wyprowadzić bez odwoływania się do zjawisk fizycznych leżących u podstaw danej metody pomiarowej. Związki te w powiązaniu z omawianymi w rozdz. 6.7 ogólnymi związkami pomiędzy ilością informacji uzyskiwanej przy pomiarze a zmianą entropii układu pomiarowego — prowadzą w wielu przypadkach do podobnych zależności jak te, które autor wyprowadza tu na „zupełnie innej drodze", analizując szczegółowo konkretne metody pomiarowe.

([16.1]) Wspominane w notce ([15.1]) ogólne zależności pomiędzy dokładnością, pewnością a ilością informacji można też traktować jako jeden z wyrazów zasady nieoznaczoności.

([17.1]) Autor podaje tutaj, oparty w dużej mierze na przesłankach heurystycznych, szkic dowodu twierdzenia Shannona o kodowaniu dla kanału ciągłego z białym szumem gaussowskim. Bardziej sformalizowany dowód oraz liczne cytaty nowszej literatury na ten temat można znaleźć w rozdz. IV książki J. SEIDLERA cytowanej w notce ([1.7]).

([17.2]) Ściśle należałoby powiedzieć: nie powinniśmy wykorzystywać telegramu, który po przejściu przez kanał daje odebrany telegram niezaszumiony taki, iż punkt reprezentujący odebrany telegram niezaszumiony leży blisko punktu M.

([17.3]) Chodzi tu o odebrane telegramy niezaszumione (zob. też notka ([17.2])).

([17.4]) Gaussowskim nazywamy szum, dla którego łączny rozkład prawdopodobieństwa dowolnego zespołu próbek jest rozkładem gaussowskim (zob. np. rozdz. VIII książki J. SEIDLERA cytowanej w notce ([1.7])).

([19.1]) O szerokości pasma przepuszczania układu jest sens mówić tylko wówczas, gdy układ jest liniowy bądź może być linearyzowany.

([19.2]) Autor czyni tu założenie (z którego w bardzo istotny sposób korzysta w dalszych rozważaniach), że maszyna dokonuje liniowego przekształcenia ciągu próbek. W ten sposób zagadnienie zostaje bardzo zawężone i w dalszym ciągu rozdziału faktycznie omawiane są zagadnienia liniowych układów impulsowych. Podstawową metodą analizy takich układów jest tzw. *transformacja zet*. Dokładne omówienie zagadnień związanych z liniowymi układami impulsowymi oraz zagadnienie transformacji *zet* można znaleźć w wielu podręcznikach. (Zob. np.: E. MISHKIN, L. BRAUN, *Adaptacyjne układy sterowania automatycznego*, Warszawa 1965, WNT, rozdz. 5).

([19.3]) Autor wprowadza tu tzw. *funkcję zet przenoszenia*. Szczegóły na ten temat można znaleźć w książce E. MISHKINA, L. BRAUNA cytowanej w notce ([19.2]).

([19.4]) Wszystkie te uwagi odnoszą się jedynie do maszyny realizującej program liniowy. Innymi słowy, są to uwagi słuszne dla liniowego układu impulsowego.

([19.5]) Jak zwrócono na to uwagę w notce ([19.2]), maszyna z programem liniowym jest równoważna liniowemu układowi impulsowemu. Rozważania, które tu autor podaje, są więc faktycznie rozważaniami na temat stabilności układów liniowych, zawierających układy impulsowe. Szczegółowe rozwiązanie można znaleźć w wielu podręcznikach, np. w podręczniku E. MISHKINA, L. BRAUNA cytowanym w notce ([19.2]).

([20.1]) Stosując dość specjalną terminologię, autor rozpatruje tu następujące zagadnienie: Mamy zbiór wiadomości będących funkcjami pewnych argumentów z zadanym rozkładem a priori; należy obliczyć ilość informacji, jaką uzyskamy dokonując wyboru konkretnej funkcji, z tym jednak, że funkcje te możemy obserwować jedynie z ograniczoną dokładnością. Tak więc rozważany jest faktycznie problem ilości informacji zawartej w funkcji mierzonej ze skończoną dokładnością. Zagadnienie to można ująć w sposób ogólny i ścisły posługując się pojęciem entropii. (Zob. np. rozdz. V książki J. SEIDLERA cytowanej w notce ([1.7])).

([20.2]) W tym przypadku *możliwością* nazywamy możliwość wystąpienia (wyboru) funkcji określonej na zbiorze N punktów o współrzędnych $x_1, \ldots, x_n$, przy czym w każdym z tych punktów funkcja może przyjmować jedną spośród k dopuszczalnych dyskretnych wartości.

SKOROWIDZ NAZWISK

Arcidiacono G. S. 403

Babbage C. 345
Balser M. 101, 105
Bar-Hillel Y. 378, 379, 382, 383
Beauregard (de) O. Costa 403
Bell D. A. 328
Bell D. E. 25
Bernamont J. 192
Bohm D. 298, 301, 313
Bohr H. 114
Bohr N. 298, 315, 316
Boltzmann L. 73, 163, 168
Borel E. 389
Born M. 385, 388
Bremmer H. 407
Bridgman P. W. 192, 201
Brillouin L. 115, 152, 154, 157, 163, 199, 216, 306, 328, 334, 342, 404

Callen H. B. 197
Carnap R. 378, 379, 382, 383
Casimir H. B. G. 198
Cheatham T. P. 142
Chinczyn A. 142
Churchill R. V. 117
Clausius R. 160
Couffignal L. 375
Courtines M. 192, 243

Darrow K. K. 157, 208
Demers P. 215, 217, 220
Desoer C. A. 328
Devé H. 144
Devey G. 47
Dicke R. H. 287
Dickson L. E. 81

Einstein A. 172, 174, 396
Elias P. 93
Elssasser W. 298, 300, 303
Emde F. 118

Fantappié L. 403
Felker I. H. 328
Fine H. B. 81
Fine P. C. 157
Francia (di) G. T. 290

Gabor D. 128, 139, 221, 232, 236, 331, 333
Gibbs J. W. 73, 163, 168
Gilbert E. N. 101, 105
Golay M. J. E. 101
Greene R. F. 197
Groot (de) S. R. 198

Haas-Lorentz (de) G. L. 197
Hahn E. L. 400
Hamming R. W. 99, 101, 105
Harker D. 154

Harrison C. W. 93
Heitler W. 315

Ising G. 243, 246
Iskenderian H. P. 163

Jacobson H. 217, 221
Jahnke E. 118
James R. W. 154
Jeans J. H. 214
Johnson J. B. 192
Jordan P. 209

Kelvin (Thomson William) 160, 162
Körner S. 400
Kretzmer E. R. 93

Labrouste H. 144
Labrouste Y. 144
Lawson J. L. 190
Le Corbeiller P. 126
Lee Y. W. 142
Leifer M. 321
Levy M. 144
Lewis G. N. 215
Linvill W. K. 349, 351, 352, 355, 377
Lorentz H. A. 197
Luhn P. 91

MacDonald D. K. C. 106
MacKay D. M. 371
Mandelbrot B. 59, 73, 74, 75
Maxwell J. C. 73, 163, 214, 215, 217
Mayer J. C. 157, 207
Mayer M. G. 157, 207
Milne-Thomson L. M. 80
Montgomery C. G. 287

Neumann (von) J. 209, 302
Nyquist H. 175, 192

Oliver B. M. 91, 92
Onsager L. 198

Patterson A. L. 154
Pepinsky R. 155
Perrin J. 174
Planck M. 73, 163, 172, 203
Poe F. A. 345
Pratt F. 47
Purcell E. M. 287

Raymond R. C. 328. 329, 331
Reed I. S. 101
Rosenfeld L. 298, 315, 316, 401
Rothstein J. 213

Salzer J. M. 349, 352, 356, 357, 358, 361, 364, 377
Schreiber W. F. 321
Schrödinger E. 160, 208, 409
Schwinger 408
Shannon C. E. 27, 36, 44, 48, 51, 52, 53, 56, 59, 66, 68, 73, 76, 87, 89, 91, 104, 133, 136, 213, 319, 324, 329, 332, 333, 341, 373
Silverman R. A. 101, 105
Slater J. C. 215
Smoluchowski M. 214
Sokolnikoff E. S. 81, 117
Sokolnikoff I. S. 81, 117
Stern O. 208
Szilard L. 214, 215, 217, 232, 373

Tait P. G. 160
Tolman R. C. 157

Uhlenbeck G. E. 190

Van der Pol B. 407
Ville J. 378

Wagner C. A. 105
Walker R. M. 112
Weaver W. 27, 48, 53, 56, 59, 66, 73
Welton T. A. 197
Wheeler H. A. 126
Whittaker E. T. 139
Wiener N. 142, 304, 334
Wiesner J. B. 142

Zemansky M. W. 157
Zipf G. K. 47, 76

SKOROWIDZ RZECZOWY

alfabet 57
amplituda maksymalna 228, 229
anomalia oscylatora 393
apertura 270
— kwadratowa 294
— soczewki 270
aproksymacja 262
Avogadry liczba 163, 174

bit, definicja 23
— informacji 249
błąd absolutny, względny i znormalizowany 264
błądzenie przypadkowe 175

całka Fouriera 120
— rozbieżna 406
Carnota zasada 220, 234, 285
cena entropijna obserwacji 239, 284
cont 382
czas trwania impulsu 126
— — efektywny 229
— — nominalny 126
częstotliwość duża 288
— krytyczna 250
— mała 319

decybel 412
degradacja energii 158, 303
demon Laplace'a 388
— Maxwella 214–220, 222
determinizm 384, 385, 395
— dokładny 403
długość charakterystyczna 271, 276
— fali Comptona 407
— minimalna 307
dokładność, definicja 254, 264, 281
— duża i niska temperatura 277, 278, 304
— — i wielkie częstotliwości 269–273
— —, warunki 271
— graniczna pomiarów 174
— i błąd proporcjonalny 284
— i nieoznaczoność 312
— i pewność, różnice 254
—, ograniczenia 302–305
— pomiarów 242, 264–286
dopasowanie kanału do źródła wiadomości 77
— kodu do kanału 67, 73
— obciążenia do linii 289
Dopplera zjawisko 313
dyfrakcja 270

efekt Lamba – Rutherforda 408
— śrutowy 179
efektywność kodów korekcyjnych 101
— sposobów klasyfikacji 106
efektywny czas trwania impulsu 229
element Liouvillle'a 391

energia całkowita szumu 322, 326
— — układu 159
— graniczna 314
— kinetyczna elektronu 251
— rezonatora 246
— ruchu cieplnego 169, 230
entropia a informacja 24, 25
—, definicja 158
— gazu idealnego 235
— kabla 329, 330
— układu 219

Fabry–Perota interferometr 275
falowód 270
Fermiego jednostka 405
Fourriera szereg 114, 282, 347
funkcja autokorelacji 141–144
— — Wienera 155
— impulsowa Shannona 350
— —, przykłady 361
— korelacji skrośnej 145
— przenoszenia maszyny matematycznej 354
— — wzmacniacza 377
— własna *m* 381
— z ograniczonym pasmem częstotliwości 141
— zwana wiadomością 132, 141

Gabora wzór 331, 333
Gibasa zjawisko 120
gracz szachowy Maelzela 345
granica Cauchy'ego 407
graniczna wartość entropii 248

Hamminga kody 96
Hartleya–Tullera–Shannona wzór 321
Heisenberga zasada nieoznaczoności *p.* zasada
hiperkula 319, 320
hiperprzestrzeń 320–324

impedancja charakterystyczna 289
implikacja 379
impuls Linvilla 351
impuls Shannona 351
indywidualia 379
ilość informacji, definicja 22, 32, 368
informacja absolutna 342
— a rozwlekłość 341
— , ilość — definicja 22, 32, 368
— martwa 336
— metryczna 371
— oparta o alfabet 26
— rozprowadzona 342
— semantyczna 378
— strukturalna 371
— swobodna 202
— warunkowa 44
— związana 202
— żywa 335

jądro Dirichleta 155
— Fejera 155
— przekształcenia 155
jednostka dwójkowa 24
— Fermiego 405
język naturalny 73

kanał 95
— z szumem 104
Kartezjusza reguła 81
kod 52, 53
— binarny 82, 95
— Hamminga 96
— korygujący błędy 96
— najbardziej efektywny 72
— odwracalny 53
— wykrywający błędy 95
koder 53
kodowanie bloków liter 93
— liter 82
— sekwencyjne 57
— w mózgu 75
komórka informacyjna 139, 318
— neutralna 106
— podstawowa sieci 153
— sieci 147

koniunkcja 379
korelacja w języku 47
koszt słowa 74
kryterium Isinga 244, 248
kwant 220, 233, 404

Lagrange'a, tożsamość 292
liczba Avogadry 163, 174
— kwantowa 244–246
— słów w słowniku 76
licznik Geigera 304

masa spoczynkowa 307, 308
maszyna Babbage'a 345
— cieplna 161
— Gabora 237
— matematyczna 343–350
— — jako element matematyczny 346
— — — — obwodu 349
— — spełniająca rolę integratora 362
— — wykonująca operacje różniczkowania 364
mediana 244–246
metoda Linvilla 350
— Nyquista 192
— Pattersona 155
— próbkowania 136, 151–153
— trapezów 363
miara informacji 382
— zawartości (*cont*) 382
mikrostan 164

negacja 379
negentropia 160
— , minimum 242, 253
— potrzebna do obserwacji 226, 320
negentropijna zasada informacji 204, 214, 242, 297, 298, 317, 328, 335, 385, 387
neper 412
nieoznaczoność 215, 285, 307
— częstotliwości 395
relacje 309, 310, 313, 315
nieoznaczoność w pomiarach pola 315
nierówność Schwarza 130

obszar Liouville'a 392, 393
odczyt negatywny 251
— pozytywny 250
odwzorowanie konforemne 359, 360
ognisko soczewki 287
— w falowodzie 291, 292, 295
ogony fal 395
opis stanu 380, 382
— — rzeczowy 397
oscylator anharmoniczny 390
— harmoniczny 244, 318, 390
— kwantowy 169
otoczenie środka elektronu 409

pasmo częstotliwości 314
— interferencyjne 272
— nominalne 125
perpetuum mobile drugiego rodzaju 125
pewność duża lub mała 254
— obserwacji 243, 247, 248
— odczytu negatywnego 250
— urządzenia 243
płaszczyzna ogniskowania 291
pobudzenie cieplne 171, 185, 230
— — i ruchy Browna 171–174
— — przy wysokich częstotliwościach 198
— — w obwodzie elektrycznym 185–201
poziomy kwantyzacji 94
prawo Boyle'a 235
— Mandelbrota 75
— van der Waalsa 370
— Zipfa 75, 76
prawdopodobieństwo 230
— poziomu energetycznego 244
— telegramu 90
— wystąpienia aktywnego przedziału 241
— zderzeń cząsteczki 240
predykaty 379
predykcja naukowa 402
— warunkowa 44

prędkość fazowa 287, 288
— grupowa 287, 288
program czasowo inwariantny 352
— liniowy obliczeń 352
programowanie 343
prostownik 215, 390
— idealny 215
próbkowanie 346, 349
przemiana adiabatyczna 300, 301
— nieodwracalna i odwracalna 158, 159
przepustowość kanału 55, 56, 320–326, 330, 331, 377
punkty próbkowania 151, 152

rekurencja 399
reproduktywności własność 299
rezonator 244, 314
— wielkiej częstotliwości 247–249
rozgłaszanie 324, 325
rozwlekłość języka 46, 49, 72
ruch cieplny 169
— — w krysztale 169
— drżący 409
— przypadkowy 175, 215
ruchy Browna 182, 243

sieć krystaliczna 147
— odwrotna 152
skoki kwantowe 402
sprawność całkowita 231, 232
— obserwacji 273, 304
— postępowania demona 231
— przemiany 231–238
stabilność programu 358
stała Boltzmanna 221, 387
— Plancka 198, 221
— subtelnej struktury 408
stopień jakości energii 158
— swobody funkcji 150
— — wiadomości 132
symbole informacyjne 97
— kontrolne 97
system dolnoprzepustowy i górnoprzepustowy 124
— pasmowoprzepustowy 124
— telekomunikacyjny 124
szereg Fouriera 282, 347
szerokość pasma 126
— — nominalna 126
— — szumu 318
— — przez maszynę matematyczną 358
szum cieplny 186–190
— gaussowski 323

telegram 56–58, 78, 88
— nieskończenie długi 72
— niezaszumiony 320
— odebrany zaszumiony 320
tożsamość Lagrange'a 292
transformacja Laplace'a 355
— informacji w negentropię 217
twierdzenie Wienera–Chinczyna 141

układ fal 291
— odizolowany 216
ukryte zmienne 397
uogólniona zasada termodynamiki 315

wartość informacji 377
warunek Bragga 154
widmo ciągłe 121, 122, 227
— dyskretne 121, 122, 227
— Fouriera 261
Wilsona komora 304
współczynnik sprawności cieplnej 162
wykresy Nyquista 197, 359–361
wzmacniacz 304, 357
— ze sprzężeniem zwrotnym 304, 357
wzmocnienie 357
wzór Boltzmanna–Plancka 164
— Gabora 331, 333
— Gibsa 167
— Hartleya–Tullera–Shannona 321
— Parsevala 116, 229

wzór Plancka 171
— Shannona 30
— Stirlinga 29, 31

zasada Carnota 220, 234
— — uogólniona 297
— dopełnienia 298
— negentropijna informacji 204, 214, 242, 297, 298, 317, 328, 335, 385, 387
— nieoznaczoności 126, 298, 299
— odpowiedniości 299
zasada termodynamiki 157
zdania faktualne 379
— logicznie fałszywe 379
— tautologiczne 379
zdarzenia łączne 42
zdolność rozdzielcza 270
zdudnienia 331
— szumu i sygnału 332
zegar elektroniczny 283, 284
zjawisko Dopplera 313
— Gibsa 117, 120

WYŻSZA SZKOŁA EKONOMICZNA W KATOWICACH BIBLIOTEKA TECHNICZNA

POLSKIE TOWARZYSTWO CYBERNETYCZNE
PAŃSTWOWE WYDAWNICTWO NAUKOWE

INFORMACJA I STEROWANIE

Seria poświęcona tematyce cybernetycznej