SCIENCE AND INFORMATION THEORY

LEON BRILLOUIN

Adjunct Professor of Physics, Columbia University,
Member of the National Academy of Sciences

ACADEMIC PRESS INC • PUBLISHERS • NEW YORK

1956

Л. БРИЛЛЮЭН

НАУКА И ТЕОРИЯ ИНФОРМАЦИИ

Перевод с английского
А. А. ХАРКЕВИЧА

ГОСУДАРСТВЕННОЕ ИЗДАТЕЛЬСТВО
ФИЗИКО-МАТЕМАТИЧЕСКОЙ ЛИТЕРАТУРЫ
МОСКВА 1960

13-4-5

АННОТАЦИЯ

Эта книга, написанная видным физиком, посвящена применениям идей и методов теории информации в ряде областей науки, в особенности же в физике. Одной из главных тем книги является рассмотрение взаимосвязи теории информации и термодинамики, приводящее к формулировке *негэнтропийного принципа информации.*

Кроме того, в книге разбирается целый ряд важных общефизических проблем. Книга будет полезна физикам, инженерам, аспирантам и студентам, имеющим дело с теорией информации.

Леон Бриллюэн.

Наука и теория информации.

Редактор *В. Д. Козлов.*

Техн. редактор *В. Н. Крючкова.* Корректор *С. Н. Емельянова.*

Сдано в набор 4/VIII 1959 г. Подписано к печати 19/XII 1959 г. Бумага $84 \times 108^1/_{32}$. Физ. печ. л. 12,25. Условн. печ. л. 20,09. Уч.-изд. л. 19,45. Тираж 16 000 экз. Т-11094. Цена книги 11 руб. 75 коп. Заказ 1548.

Государственное издательство физико-математической литературы.
Москва, В-71, Ленинский проспект, 15.

Ленинградский Совет народного хозяйства. Управление полиграфической промышленности. Типография № 1 «Печатный Двор» имени А. М. Горького.
Ленинград, Гатчинская, 26.

СОДЕРЖАНИЕ

ИЗ ПРЕДИСЛОВИЯ АВТОРА

За последние годы народилась новая научная теория — теория информации. Она немедленно привлекла большой интерес и быстро распространилась. Эта новая теория возникла первоначально в результате практического и утилитарного обсуждения некоторых основных проблем: Каким образом возможно определить количество информации, содержащейся в передаваемом сообщении или телеграмме? Как измерить количество информации, сообщаемой при помощи телеграфных сигналов? Как сравнить эти две величины [1]) и оценить эффективность кодирующих устройств? Все эти вопросы и многие другие подобные заботят инженера связи и могут теперь обсуждаться с количественной точки зрения.

Из этого обсуждения возникла новая теория, имеющая как практический, так и математический характер. Теория основана на вероятностных соотношениях. Точно сформулированная, она может применяться во многих основных научных рассуждениях. Она позволяет решить проблему демона Максвелла и показать непосредственную связь между информацией и энтропией. Термодинамическая энтропия есть мера недостатка информации о некоторой физической системе. Когда в лаборатории производится какой-либо эксперимент, он оплачивается увеличением энтропии, и обобщенный принцип Карно устанавливает, что цена, уплаченная в виде

---

[1]) То есть количество информации и объем сигнала. (*Прим. перев.*)

увеличения энтропии, всегда больше, чем полученное количество информации. Информация соответствует отрицательной энтропии, для которой автор ввел термин ***негэнтропия***. Обобщенный принцип Карно может также быть назван ***негэнтропийным принципом информации***. Этот принцип накладывает новые ограничения на физические эксперименты и не зависит от известного соотношения неопределенности квантовой механики.

Настоящая книга основана на лекциях, читанных инженерам фирмы International Business Machine Corporation, а позднее в различных университетах, в частности в Калифорнийском университете в Беркли.

Январь 1956 г. *Леон Бриллюэн*

ВВЕДЕНИЕ

Новая территория была завоевана для науки с появлением в недавнее время теории информации. Это открытие создало новую область, немедленно привлекшую разведчиков и исследователей. Это интересное явление в истории науки, и такое внезапное расширение области научного исследования заслуживает более пристального рассмотрения. Как это случилось? Как далеко это идет? И где оно может продолжать распространяться? Означает ли это вторжение науки на территорию, принадлежащую по традиции философии, или это есть открытие новой страны, своего рода «ничейной земли», которая ускользала от прежних исследований? [1]). Мы разберем все эти вопросы и дадим на них ответ.

Прежде всего, что такое *информация*? Заглянем в словарь Вебстера: «Сообщение, или получение знаний или сведений. Факты, приготовленные для сообщения, в отличие от тех, которые воплощены в мысли или знании. Данные, новости, сведения, знания, полученные путем изучения или наблюдения . . .». Мы можем установить, что информация есть сырой материал и состоит из простого собрания данных, тогда как знание предполагает некоторое размышление и рассуждение, организующее данные путем их сравнения и классификации. Следующий шаг приводит к научному знанию и формулировке научных законов.

Каким образом можно сформулировать научную теорию информации? Прежде всего, нужно начать с точного определения. Наука начинается, когда значения слов четко разграничены. Слова могут быть выбраны из существующего словаря, либо могут быть созданы новые слова, но все они

---

[1]) Едва ли автор не относит философию к области науки. Просто он употребляет слово «наука» в смысле, соответствующем нашему термину «точные науки». (*Прим. перев.*)

должны получить новое определение, исключающее недоразумения и двусмысленность в пределах того раздела науки, где они применяются.

Может случиться, что одно и то же слово имеет различные значения в двух различных отраслях науки: слово *корень* имеет одно ясно определенное значение для изучающего алгебру и другое столь же специфическое значение для ботаника. Однако опасность смешения при столь удаленных областях невелика. Алгебраические корни не растут, а корни ботаника никогда не бывают мнимыми. Эта единственность значения слов характерна для научного метода. Так как сходные определения введены учеными всех стран, то перевод облегчается однозначным соответствием научных словарей. Если бы такое положение преобладало в повседневной практике, то международное взаимопонимание было бы много легче осуществить!

Рядовой человек испытывает беспокойство, когда обычные слова применяются в новом научном определении, и он склонен называть эту практику научным жаргоном. Но жаргоны применяются, как правило, в любой специальной области — в богословии и в философии так же, как и в технике. Рядовой читатель не может понять язык специалистов, так как он недостаточно знаком с обсуждаемыми вопросами.

Точное определение слов в научном языке обычно основано на двух различных методах. В математике определение начинается с некоторого числа тщательно отобранных и сформулированных постулатов; более сложные сущности выводятся из этих постулатов и выражаются через них. Новые определения равносильны словесному переводу формул, данных в символической форме и основанных на постулатах. Экспериментальные науки ввели другой тип определения, часто называемый *операционным* (operational). Сила, масса, скорость и т. д. определяются кратким описанием эксперимента, необходимого для измерения этих величин. Операционная точка зрения в экспериментальных науках настойчиво рекомендуется многими выдающимися учеными, и имя П. Бриджмена часто упоминается в этой связи. Как правило, считается целесообразным вводить в научный язык только те величины, которые могут быть определены операционно. Слова, не поддающиеся операционному определению, обычно в конце концов признаются не заслуживающими доверия и исключаются из научного словаря. Вспомним, например, *эфир*, и как теория относительности лишила этот термин смысла.

Возвращаясь к теории информации, мы должны начать с точного определения слова *информация*. Мы рассматриваем задачу с некоторым числом возможных ответов, если мы не имеем специальной информации о действительном положении. Если окажется, что мы располагаем некоторой информацией о задаче, то число возможных ответов уменьшается, а полная информация может даже оставить нам лишь единственный возможный ответ. Информация есть функция отношения числа возможных ответов до и после (получения информации), и мы выбираем логарифмический закон для обеспечения аддитивности информации, содержащейся в независимых ситуациях. Эти задачи и определения обсуждаются в главе 1 и составляют основу новой теории.

Методы этой теории могут с успехом применяться ко всем техническим проблемам, касающимся информации, как-то: кодирование, связь, вычислительные устройства и т. д. Во всех этих проблемах мы фактически перерабатываем информацию, или передаем ее из одного места в другое, и данная теория очень полезна для формулировки правил и установления точных пределов того, что может быть, а что не может быть сделано.

Но мы не в состоянии исследовать процесс мышления, и мы не можем в настоящее время ввести в нашу теорию какой-либо элемент, включающий человеческую оценку информации. Это исключение человеческого элемента является очень серьезным ограничением, но это есть та цена, которую мы должны были уплатить за возможность построения этой области научного знания. Введенные ограничения позволяют нам дать количественное определение информации и трактовать информацию как физически измеримую величину. Определение не может делать различия между очень важной информацией и новостью, не имеющей большой ценности для того, кто ее узнает.

Определение может показаться на первый взгляд искусственным, но в действительности оно практично и научно. Оно основано на собрании статистических данных о каждой обсуждаемой проблеме, и эти данные, поскольку они доступны, одинаковы для всех наблюдателей. Поэтому наше определение информации есть абсолютно объективное определение, не зависящее от наблюдателя. С другой стороны, *ценность* информации является, очевидно, субъективным элементом, относящимся к наблюдателю. Информация, содержащаяся в некотором предложении, может иметь очень большое значение для меня и быть совершенно неинтересной для моего соседа. Заметка в газете может быть прочитана с некоторым интересом многими читателями, но теорема Эйнштейна не имеет ценности для рядового человека, тогда как она привлекает большое внимание физика.

Все эти элементы человеческой оценки игнорируются в настоящей теории. Это не означает, что они должны игнорироваться всегда, но в настоящее время они еще не исследованы и не классифицированы. По всей вероятности эти проблемы будут очередными в программе научных исследований, и можно надеяться, что они смогут обсуждаться с применением научных методов.

Теперешняя теория простирается на «ничейную землю» абсолютной информации, на проблемы, которые до сих пор не обсуждали ни ученые, ни философы. Когда мы дойдем до проблем ценности, то мы начнем уже вторгаться на территорию, принадлежащую философии. Сможем ли мы пересечь эту границу и раздвинуть пределы науки в этом направлении? Будущее ответит на этот вопрос [1]).

Определение абсолютной информации имеет большое практическое значение. Исключение человеческого элемента как раз дает возможность ответить на целый ряд вопросов. Инженера, который конструирует телефонную систему, не интересует, будет ли она использована для передачи сплетен, биржевых цен или дипломатических сообщений. Техническая задача всегда одна и та же: передать информацию, какова бы она ни была, правильно и точно. Конструктор вычислительной машины не знает, будет она применяться для составления астрономических таблиц или для коммерческих расчетов. Исключение человеческой оценки информации — это как раз путь к ее научному обсуждению, не подверженному влиянию предвзятых мнений и эмоций.

Физика вступает в игру, когда мы обнаруживаем примечательное сходство между информацией и энтропией. Это сходство давно уже было отмечено Л. Силардом в его старой

---

[1]) См. предыдущую сноску.

работе 1929 г., которая является предшественницей данной теории. В этой работе Силард является поистине пионером на неизведанной территории, которую мы сейчас изучаем во всех направлениях. Он исследовал проблему демона Максвелла, которая является одним из важных вопросов, обсужденных в этой книге. Связь между информацией и энтропией была вновь открыта Шенноном в другом классе задач, и мы отводим много глав этому сравнению. Мы показываем, что информация должна рассматриваться как отрицательное слагаемое энтропии системы; короче говоря, информация есть негэнтропия.

Энтропия физической системы часто описывалась как мера случайности строения системы. Мы можем теперь сформулировать это положение несколько иначе: всякая физическая система не полностью определена. Мы знаем лишь значения некоторых макроскопических переменных, и мы не в состоянии определить в точности положения и скорости всех молекул, образующих систему. Мы имеем лишь скудную частичную информацию о системе, и большая часть информации о детальной структуре отсутствует. Энтропия есть мера недостатка информации; она выражает общее количество отсутствующей информации об ультрамикроскопической структуре системы.

Эта точка зрения определяется как негэнтропийный принцип информации и непосредственно ведет к обобщению второго начала термодинамики, так как энтропия и информация должны рассматриваться совместно и не могут трактоваться порознь. Этот негэнтропийный принцип информации будет подтвержден различными примерами в диапазоне от теоретической физики до повседневной жизни. Существенно показать, что любое наблюдение или эксперимент над физической системой автоматически ведет к увеличению энтропии лаборатории. После этого возможно сравнить потерю негэнтропии (увеличение энтропии) с количеством полученной информации.

Эффективность эксперимента может быть определена как отношение полученной информации к связанному с ней увеличению энтропии. Эта эффективность согласно обобщенному принципу Карно всегда меньше единицы. Примеры показывают, что эффективность может быть близка к единице в некоторых специальных случаях, но может быть также чрезвычайно малой в других случаях.

Этот подход очень полезен при сравнении основных экспериментов, применяемых в науке, в частности в физике. Он ведет к новому исследованию эффективности различных методов наблюдения, а также их точности и надежности.

Интересным результатом этого рассуждения является заключение о том, что измерение чрезвычайно малых расстояний физически невозможно. Математик определяет бесконечно малые, но физик абсолютно не в состоянии измерить их, и они представляют собой чистую абстракцию, лишенную физического смысла. Если мы примем операционную точку зрения, мы должны исключить бесконечно малые из физических теорий, но, к сожалению, мы не имеем представления о том, как выполнить такую программу.

В общем, мы надеемся, что научная теория информации отмечает начало новой важной главы научных исследований, в особенности в области физики, а также биологии.

ГЛАВА 1

ОПРЕДЕЛЕНИЕ ИНФОРМАЦИИ

1. Определение информации

Теория информации, разработанная за последние годы, нашла широкое применение в различных областях: в связи, в технике вычислений, в чистой физике и при обсуждении основного процесса научного наблюдения. Некоторая величина может быть определена как мера количества информации в данной операции, и мы покажем, что эта величина тесно связана с физической энтропией, применяемой в термодинамике.

Определение информации выводится из статистических рассуждений. Обсудим сначала очень простой пример. Рассмотрим ситуацию, в которой могут произойти P_0 различных событий, но при условии, что эти P_0 возможных исходов считаются априори равновероятными. Это есть исходная ситуация, когда мы не имеем специальной информации о рассматриваемой системе. Если мы получаем бóльшую информацию, мы сможем уточнить, что только один из P_0 исходов в действительности реализуется. Чем больше неопределенность исходной задачи, тем больше P_0 и тем больше количество информации, необходимое для выбора.

Итак, мы имеем:

начальная ситуация: $I_0 = 0$ при P_0 равновероятных исходах,

конечная ситуация: $I_1 \neq 0$ при $P_1 = 1$, т. е. при одном избранном исходе.

Символ I означает информацию, и определение информации есть

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

где K — постоянная, а $\ln$ — означает натуральный логарифм.

Применение логарифма в (1.1) оправдывается тем, что мы желаем, чтобы информация обладала свойством аддитивности.

Рассмотрим две независимые задачи, первая из которых имеет P_{01} *априори* равновероятных решений, а вторая — P_{02}. Каждое решение первой задачи может быть спарено с некоторым решением второй. Таким образом, общее число начальных исходов равно

$$P_0 = P_{01} P_{02},$$

что дает:

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

где

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

Общее количество информации, необходимое для решения обеих задач, равно сумме раздельных информаций I_{11} и I_{12}.

2. Системы единиц

Прежде чем идти далее, нужно обсудить вопрос о *единицах*. В теории информации принято рассматривать информацию I, как безразмерную величину (отвлеченное число), а стало быть, постоянная K есть отвлеченное число. Наиболее удобная система основана на двоичных единицах (binary digits, сокращенно bits). Поясним это на примере.

Рассмотрим задачу с n различными независимыми выборами, каждый из которых сводится к двоичной альтернативе: 0 или 1. Общее число возможностей

$$P = 2^n,$$

и из (1.1) получаем информацию

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

Мы желаем отождествить I и n (число двоичных единиц) и берем:

$$K = \frac{1}{\ln 2} = \log_2 e. \tag{1.3a}$$

Таким образом, информация I в двоичных единицах равна

$$I = \log_2 P. \tag{1.3b}$$

В качестве примера рассмотрим колоду из 32 различных карт, из которых должна быть выбрана одна. На основании (1.3b) мы скажем, что $I=5$ двоичных единиц, так как $32=2^5$. Но возьмем теперь две колоды, по 32 карты в каждой. Если мы выберем две карты, по одной из каждой колоды, то желательно, чтобы мы могли сказать, что информация при этом вдвое больше. Общее число возможностей равно

$$P=P_1 P_2,$$

где

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

и, следовательно,

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

Из (1.1) находим:

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

и, таким образом, $I=10$ двоичных единиц.

Итак, логарифмическое определение информации представляется разумным.

Другая система единиц может быть введена, если мы сравним *информацию* с термодинамической *энтропией* и будем измерять обе величины в одних и тех же единицах. Энтропия (см. главу 9) имеет размерность энергии, деленной на температуру. Для энтропии имеется формула Больцмана, очень сходная с (1.1) и содержащая коэффициент

$$k=1{,}38 \cdot 10^{-16} \textit{ эрг/градус}. \tag{1.4}$$

Эта константа известна как постоянная Больцмана (см. (9.15)). Если мы введем эту постоянную k вместо K в (1.1), то мы будем измерять информацию в единицах энтропии.

Мы можем сделать еще один шаг и выбрать единицы так, чтобы как энтропия, так и информация были безразмерными величинами и представлялись отвлеченными числами[1]). Для этого нужно измерять температуру в единицах энергии. Обычная стоградусная шкала применима, если k имеет численное значение (1.4) и рассматривается как отвлеченное число. При этом отношение единиц наших двух систем также

---

[1]) D. E. Bell, J. Appl. Phys. **23**, 372 (1952).

есть отвлеченное число, равное

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

Это численное значение играет важную роль во всех применениях теории.

3. Обобщение и примеры

Определение меры информации может быть распространено на случай, когда в начальной ситуации имеется P_0 возможных исходов, а в конечной ситуации — P_1 исходов:

в начале: $I_0 = 0$ при P_0 равновероятных исходах,

в конце: $I \neq 0$ при P_1 равновероятных исходах.

В этом случае

$$I_1 = K \ln \frac{P_0}{P_1} = K \ln P_0 - K \ln P_1. \tag{1.5}$$

Это определение сводится к (1.1), когда $P_1 = 1$ и представляет собой очевидное обобщение (1.1).

Рассмотрим численный пример, в котором мы предположим все цифры равновероятными. Пусть G — количество цифр с основанием N. Вообще,

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

Мы можем выразить (1.6) в двоичных единицах:

$$I = G \log_2 N, \tag{1.7}$$

так как

$$(\log_2 N)(\ln 2) = \ln N \quad \text{и} \quad K = \frac{1}{\ln 2}.$$

Так, если $N = 2$, мы будем иметь 2^G различных исходов и

$$I = G \text{ двоичных единиц информации.} \tag{1.8}$$

Легко показать, что мы получим то же количество информации, если изменим основание N и соответственно число цифр. Пусть, например, выбрано основание $N' = 8$. Тогда для записи рассматриваемого числа потребуется всего $G' = G/3$ цифр, и

$$I = G' \log_2 N' = \frac{G}{3} \cdot 3 = G, \tag{1.9}$$

что согласуется с (1.8).

Рассмотрим десятичное число из G_{10} цифр:

$$N = 10 \quad \text{и} \quad I = G_{10} \log_2 10 = G_{10} 3{,}32 \text{ дв. ед.} \tag{1.10}$$

Одна десятичная цифра дает 3,32 дв. ед. информации; двоичная запись числа требует в 3,32 раза больше цифр, чем десятичная.

4. Информация и алфавит

Большая часть используемой нами информации сообщается посредством языка. В устной речи элементарными символами являются основные звуки (называемые часто фонемами), а в письменной речи слова составлены из букв. Рассмотрим письменное предложение и подсчитаем количество информации, содержащейся в этом предложении. Эта сложная задача имеет большое практическое значение; она была подробно обсуждена К. Шенноном[1]) и многими другими. Как мы увидим, полное и строгое решение задачи до сих пор неизвестно из-за отсутствия полных статистических данных о языке.

Мы можем рассматривать буквы как символы, которые необходимо выбирать для построения предложения. Полный алфавит содержит 27 символов: 26 букв плюс промежуток между словами. Если эти 27 символов равновероятны априори, то мы можем сказать, что информация, содержащаяся в предложении из G букв, составляет:

$$I = G \log_2 27 \text{ дв. ед.} \tag{1.11}$$

или

$$i = \log_2 27 = 4{,}76 \text{ дв. ед. на букву.}$$

Это соответствует прямому применению формулы (1.7). Однако полученное таким образом решение неудовлетворительно, так как различные буквы встречаются в языке с неодинаковыми априорными вероятностями (таблица 1.1).

Пусть p_j — априорная вероятность j-й буквы ($j = 1, 2, \ldots$ $\ldots, 27$), как указано в таблице 1.1. Средняя информация

---

[1]) C. E. Shannon, W. Weaver, The Mathematical Theory of Communication, U. of Illinois Press, Urbana, Ill., 1949.

Таблица 1.1

Вероятность появления p и значения $-\log_{10} p$ для букв английского языка

Символ	Вероятность p	$-\log_{10} p$	Символ	Вероятность p	$-\log_{10} p$
Промежуток	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

на одну букву, согласно Шеннону, запишется как

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

Обсудим применимость этой формулы.

5. Информация, содержащаяся в совокупности символов с различными априорными вероятностями

Предположим, что мы применяем M различных символов $1, 2, \ldots, j, \ldots, M$, которые имеют *априорные* вероятности $p_1, p_2, \ldots, p_j, \ldots, p_M$ соответственно. Никаких других условий или ограничений относительно применения этих символов не накладывается. Мы желаем вывести шенноново равенство (1.12), согласно которому последовательность из G символов содержит информацию

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

где i — средняя информация на символ и

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

Начнем с простой задачи, которую затем легко обобщить. Рассмотрим алфавит из двух «букв» — точки и тире, как в телеграфии, или 0 и 1. Рассмотрим G позиций (cells — ячеек), и пусть N_0 позиций содержат 0, а N_1 содержат 1, так что $N_0 + N_1 = G$. Таким образом, все позиции заполнены. Тогда вероятность позиции, содержащей 0, будет

$$p_0 = \frac{N_0}{G}, \tag{1.14}$$

и вероятность того, что позиция содержит 1, есть

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

причем

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

так как вероятность того, что в некоторой позиции содержится либо 0, либо 1, равна единице.

Теперь найдем число способов заполнения каждой из G позиций либо 0, либо 1 (но никоим образом не обоими символами одновременно). Эта задача в точности та же, что в статистике Ферми. Число способов заполнения G позиций равно числу способов заполнения N_0 позиций нулями, так как, если мы уже распределили N_0 нулей, то остальные N_1 позиций должны содержать по единице каждая. Но число способов заполнения N_0 позиций нулями равно числу перестановок (с повторяющимися элементами) из G по N_0:

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

Это есть число сообщений из G символов двухбуквенного алфавита, один из которых встречается N_0 раз, а другой — N_1 раз. Для одного из таких сообщений согласно (1.1) информация равна

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

Если сообщение длинно и G, N_0 и N_1 достаточно велики, то логарифмы факториалов могут быть выражены приближенно на основании формулы Стирлинга:

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

Эта формула дает, как известно, очень хорошее приближение для $Q > 100$. Итак, если $G \gg 1$, $N_0 \gg 1$ и $N_1 \gg 1$, то

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

или, учитывая, что $G = N_0 + N_1$,

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

Снова применяя равенство $G = N_0 + N_1$, перепишем (1.20) в виде

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

Подставляя (1.14) и (1.15) и деля на G, получаем:

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

где i — информация на символ сообщения. Это и есть формула Шеннона для случая двух символов. Заметим, что p_0 и p_1 меньше единицы, их логарифмы отрицательны, а потому формула (1.22) дает положительное значение для i.

Обобщение на случай более чем двух символов получается легко. Обозначим через $N_1, N_2, \ldots, N_j, \ldots, N_M$ числа символов M различных типов и выберем число позиций

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

Определим вероятность j-го символа

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

Имеем:

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

Общее число сообщений P, которые можно получить, распределяя символы случайным образом по G позициям (так чтобы на одной позиции никогда не оказывалось более одного символа), составляет:

$$P = \frac{G!}{\prod_{j=1}^{j=M} N_j!}. \tag{1.26}$$

Формула (1.26) является прямым обобщением (1.17). Мы получаем для информации, содержащейся в одном из сообщений:

$$I = K \ln P = K \left[\ln (G!) - \sum_{j=1}^{j=M} \ln (N_j!)\right] \approx$$

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

Равенства в этом выражении соответствуют (1.18), (1.19) и (1.20), полагая по-прежнему, что G и N_j достаточно велики, так что применима формула Стирлинга. Метод, при помощи которого получены (1.21) и (1.22), дает теперь:

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

а это и есть формула Шеннона.

В качестве примера рассмотрим сообщение из 10 000 букв 27-буквенного алфавита, выбранных случайным образом с одинаковыми априорными вероятностями. Тогда

$$G = 10\,000 \text{ и } I = 10\,000 \log_2 27 = 47\,600 \text{ дв. ед.} \tag{1.29}$$

и

$$i = \frac{I}{10\,000} = 4{,}76 \text{ дв. ед. на букву.} \tag{1.29a}$$

Если, однако, мы составим сообщение такой же длины, но выберем буквы в соответствии с их действительными априорными вероятностями, то нужно воспользоваться (1.28):

$$i = \frac{I}{G} = \frac{I}{10\,000} \approx K \sum_{j=1}^{j=27} p_j \ln p_j = 4{,}03. \tag{1.30}$$

Последнее значение легко подсчитать по данным таблицы 1.1.

Более подробный анализ структуры языка будет дан в следующих разделах. Будет показано, что вышеприведенное значение представляет собой верхнюю грань и что действительное количество информации на одну букву много меньше 4, вероятно, между 1 и 2 дв. ед. на букву.

6. Общие замечания

Одно общее замечание можно сделать сразу: любое ограничение, любое дополнительное условие, накладываемое на возможную свободу выбора, ведет к уменьшению информации.

Рассмотрим систему, представляющую P различных возможностей, когда все переменные свободны. Когда мы налагаем на переменные ограничения, уменьшающие свободу выбора, эти условия исключают некоторые из ранее существовавших возможностей. Новое число возможностей P' с ограничениями должно быть, очевидно, меньше, чем первоначальное P, и поэтому мы должны получить новое значение информации $I' < I$:

без ограничений: P возможностей, $I = K \ln P$;

с ограничениями: P' возможностей, при $P' < P$ и

$$I' = K \ln P' < I. \tag{1.31}$$

Пример предыдущего раздела, относящийся к применению букв, поясняет это. Когда буквы применяются свободно (одинаковые априорные вероятности), информация составляет 4,76 дв. ед. на букву. Когда мы накладываем ограничение в соответствии с таблицей 1.1 и учитываем априорные вероятности различных букв, информация на букву падает до 4,03 дв. ед. Дополнительные ограничения приведут к дальнейшему уменьшению этой величины.

Другой способ объяснения этого общего положения состоит в том, чтобы представить ограничение как некоторую предварительную информацию I_c о сообщении, которое нам предстоит выбрать; в таком случае

$$I' = I - I_c$$

есть количество информации, которое остается получить, если I_c известно.

Теперь нужно сказать несколько слов по поводу нашего способа рассуждения, чтобы показать пределы его применимости. Мы выбрали в (1.1) статистическое определение термина *информация*. Это математическое определение очень полезно при обсуждении многих научных и технических проблем. Оно позволит нам сделать некоторые выводы, имеющие действительное практическое значение и большую общность. Но это весьма точное определение ограничивает нас. Чтобы получить это определение, мы должны исключить и игнорировать многие из обычных значений слова *информация*.

Мы определяем *информацию* как результат выбора; мы не рассматриваем *информацию* как основание для предсказания, как результат, который может быть использован для того, чтобы сделать другой выбор. Мы полностью игнорируем человеческую оценку информации. Последовательности из 100 букв приписывается определенное значение информации, и мы не исследуем, имеет ли эта последовательность какой-либо смысл, и если да, имеет ли этот смысл какое-либо практическое значение. В соответствии с нашим определением совокупность 100 букв, выбранных случайным образом (согласно правилам таблицы 1.1), фраза в 100 букв из газеты, пьесы Шекспира или теоремы Эйнштейна имеют в точности одинаковое количество информации. Другими словами, мы определяем информацию как нечто отличное от *знания*, для которого у нас нет количественной меры.

Мы не делаем различия между полезной и бесполезной информацией и предпочитаем полностью игнорировать ценность информации. Наше статистическое определение информации основано исключительно на редкости. Если ситуация встречается редко, она содержит информацию. Является ли эта информация ценной или никчемной — это нас не касается. Понятие ценности относится к возможному использованию живым наблюдателем. Это находится за пределами досягаемости нашей теории, и мы не намерены обсуждать процессы мышления и любые другие проблемы, связанные с использованием информации живыми существами[1]).

Наше определение информации является, тем не менее, в высшей степени полезным и практичным. Оно в точности соответствует задаче инженера связи, который должен передать всю информацию, содержащуюся в поданной телеграмме, вне зависимости от ценности этой информации для адресата.

Информация есть абсолютная величина, имеющая одно и то же численное значение для любого наблюдателя. Человеческая ценность информации, с другой стороны, должна быть относительной величиной, имеющей различные значения для различных наблюдателей, в соответствии с их способностью понять информацию и использовать ее в дальнейшем. Теорема Эйнштейна имела бы, вероятно, для математика гораздо большую ценность, чем заметка в газете. С другой стороны, для рядового читателя газетная новость может иметь ценность, а может и не иметь, тогда как теорема почти наверное не будет иметь для него ценности.

Возьмем более тривиальный пример. Информация при выборе одной из 32 карт составляет 5 дв. ед. (см. п. 2). Она всегда измеряется этим числом — 5 дв. ед., — хотя карта может быть тузом, семеркой или королем. Ценность этих карт зависит, однако, от правил игры.

В соответствии с нашим определением информация измеряется всегда положительной величиной. Ценность информа-

---

[1]) Едва ли следует настаивать на терминах: *живой наблюдатель, живые существа*. Функция *использования* информации может быть возложена и на автомат в объеме, определяемом его логической организацией (т. е. в соответствии с алгоритмом, определяющим действие автомата). (*Прим. перев.*)

ции может и в известных случаях должна считаться отрицательной. Профессор читает длинную лекцию, затем внезапно обнаруживает, что он сделал ошибку, и заключает: «Простите, все это было неверно». Эта последняя фраза имеет отрицательную ценность и уничтожает ценность всей предшествующей информации[1]).

Приведенные примеры ясно показывают пределы применимости данной теории; эти пределы нужно иметь в виду при любом применении теории.

---

[1]) Это весьма спорный пример. Можно полагать, напротив, что предшествующее изложение, будучи основаио на иеобнаруженной ошибке, имело нулевую или весьма малую цеиность. Указание же на ошибку позволяет восстановить истину и, следовательно, не разрушает, а восстанавливает ценность предшествующей информации. Уж если говорить об отрицательной ценности информации, то ее естественио было бы приписывать ложной информации, на основании которой человек (или автомат) может предпринять действия, нецелесообразные при данных действительных обстоятельствах. С этой точки зрения понятие о ценности информации не так уже неосязаемо. Ценность информации можно определить через *приращение вероятности достижения цели*, для которой собирается информация. Так, для человека, преследующего преступника и имеющего целью поймать его, информация, полученная от прохожих, о направлении, в котором скрылся преступник, будет иметь положительную ценность (которую можно выразить числом), если направление указано правильно, и отрицательную ценность, если указано ложное (например, противоположное) направление. (*Прим. перев.*)

ГЛАВА 2

ПРИМЕНЕНИЕ ОПРЕДЕЛЕНИЙ И ОБЩЕЕ ОБСУЖДЕНИЕ

1. Определения

В первой главе было показано, что мера информации как степени неопределенности, существовавшей до того, как выбор был сделан, является точной, но по необходимости ограниченной. Так, например, *ценность* информации не могла быть включена в эту меру.

Было также показано, что если возможные выборы имели неравные априорные вероятности, то эти вероятности могут рассматриваться как ограничения нашего выбора, что ведет в конечном счете к уменьшению количества информации.

Так, если априорные вероятности для символов (1), (2), ..., (j), ... равны соответственно $p_1, p_2, \ldots, p_j, \ldots$, то количество информации на символ, как было показано ранее (см. (1.12)), равно

$$i = -K \sum_j p_j \ln p_j . \tag{2.1}$$

Это выражение было получено из формулы для информации на символ в случае, когда выбор не ограничен:

$$i = \frac{K}{G} \ln m^G = K \ln m \tag{2.2}$$

для m различных равновероятных символов, где G — общее число использованных символов (мы заменили M раздела 5 главы 1 на m). Исходя из (2.2), мы вывели формулу (2.1). Однако многие авторы, в особенности Шеннон, брали за исходное соотношение (2.1), из которого (2.2) следует непосредственно.

Действительно, положим, что мы имеем m различных символов с одинаковыми априорными вероятностями. Тогда

$$p_1 = p_2 = \ldots = p_j = \ldots = p_m = \frac{1}{m}, \tag{2.3}$$

и из (2.1) получаем:

$$i = -K \sum_{m \text{ членов}} \frac{1}{m} \ln \frac{1}{m} = -K \ln \frac{1}{m} = K \ln m, \tag{2.4}$$

а это и есть равенство (2.2).

Рассмотрим теперь некоторые свойства соотношения (2.1).

2. Свойство A

Если выбор разбит на два последовательных выбора, то первоначальная информация будет выражаться взвешенной суммой двух частичных информаций. Рассмотрим пример. Вероятность перехода из точки O в точку A равна $1/2$, из O в B — $1/3$ и из O в C — $1/6$, как показано схематически на рис. 2.1. Рассмотрим, однако, диаграмму рис. 2.2. На этом

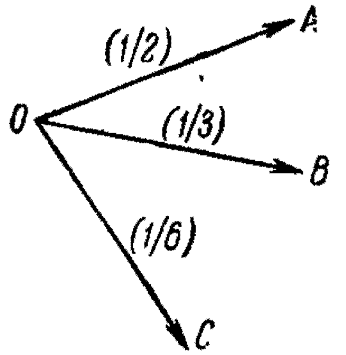

Рис. 2.1. Вероятности перехода из точки O в точки A, B и C.

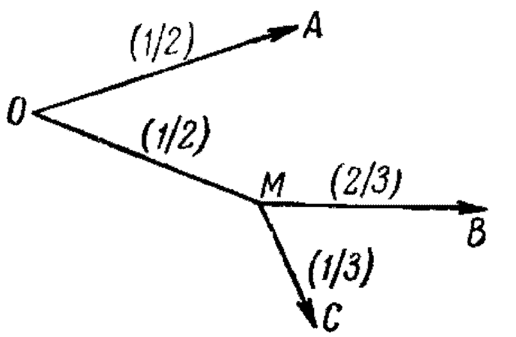

Рис. 2.2. Вероятности перехода из точки O в точки A, B и C с промежуточной точкой M между точкой O и точками B и C.

рисунке выбрана промежуточная точка M таким образом, что вероятность перехода из O в M также равна $1/2$. Это значит, что вероятность перехода из M в B должна быть равна $2/3$, так как первоначальная вероятность перехода из O в B равна $1/3$. Аналогично, вероятность перехода из M в C равна $1/3$.

Обозначим информацию, соответствующую рис. 2.1 через $I\left(\frac{1}{2}, \frac{1}{3}, \frac{1}{6}\right)$. Для рис. 2.2 информация для путей OA, OM есть $I\left(\frac{1}{2}, \frac{1}{2}\right)$, а информация для путей MB, MC есть $I\left(\frac{2}{3}, \frac{1}{3}\right)$. Полная информация для рис. 2.1 и 2.2 должна быть одинакова: $I\left(\frac{1}{2}, \frac{1}{3}, \frac{1}{6}\right)$. Из рис. 2.1 имеем:

$$I\left(\frac{1}{2}, \frac{1}{3}, \frac{1}{6}\right) = -K\left(\frac{1}{2}\ln\frac{1}{2} + \frac{1}{3}\ln\frac{1}{3} + \frac{1}{6}\ln\frac{1}{6}\right), \tag{2.5}$$

или

$$I\left(\frac{1}{2}, \frac{1}{3}, \frac{1}{6}\right) = \\ = -K\left[\frac{1}{2}\ln\frac{1}{2} + \frac{1}{2}\cdot\frac{2}{3}\ln\left(\frac{1}{2}\cdot\frac{2}{3}\right) + \frac{1}{2}\frac{1}{3}\ln\left(\frac{1}{2}\cdot\frac{1}{3}\right)\right]. \tag{2.6}$$

Раскрывая и группируя, получаем:

$$I\left(\frac{1}{2}, \frac{1}{3}, \frac{1}{6}\right) = \\ = -K\left[\left(\frac{1}{2}\ln\frac{1}{2} + \frac{1}{2}\ln\frac{1}{2}\right) + \frac{1}{2}\left(\frac{2}{3}\ln\frac{2}{3} + \frac{1}{3}\ln\frac{1}{3}\right)\right] = \\ = -K\left(\frac{1}{2}\ln\frac{1}{2} + \frac{1}{2}\ln\frac{1}{2}\right) + \frac{1}{2}\left[-K\left(\frac{2}{3}\ln\frac{2}{3} + \frac{1}{3}\ln\frac{1}{3}\right)\right] = \\ = I\left(\frac{1}{2}, \frac{1}{2}\right) + \frac{1}{2}I\left(\frac{2}{3}, \frac{1}{3}\right). \tag{2.7}$$

Последнее выражение соответствует рис. 2.2, для которого полная информация должна равняться взвешенной сумме информаций для путей OA, OM и MB, MC. Это — общий результат, являющийся следствием того факта, что информация — логарифмическая функция.

3. Свойство B

Свойство B выражается неравенством, которым нам предстоит пользоваться в дальнейшем. Чтобы получить это неравенство, рассмотрим функцию

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

где p_j — вероятность j-го выбора, а q_j могут рассматриваться как другая совокупность вероятностей, так что

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

Запишем

$$q_j = p_j + u_j = p_j \left(1 + \frac{u_j}{p_j}\right).$$

Тогда u_j удовлетворяют условию

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

Теперь, заменяя q_j в Ψ, получаем:

$$\Psi = \sum_j p_j \ln\left[p_j\left(1 + \frac{u_j}{p_j}\right)\right] =$$

$$= \sum_j p_j \ln p_j + \sum_j p_j \ln\left(1 + \frac{u_j}{p_j}\right). \tag{2.10}$$

Сравнивая графики $\ln\left(1 + \frac{u}{p}\right)$ и u/p (рис. 2.3), мы видим,

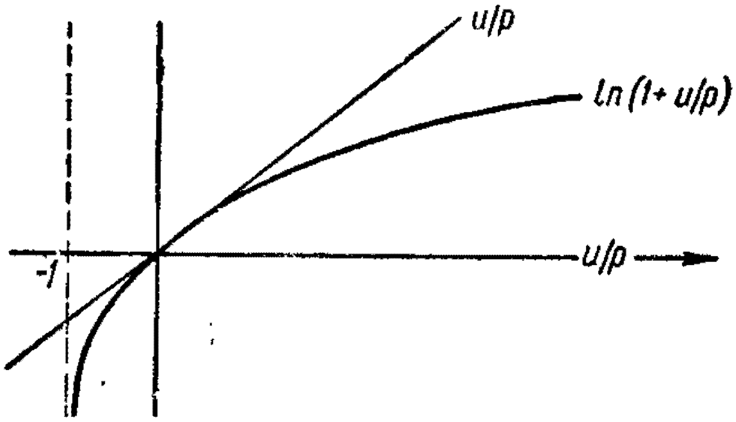

Рис. 2.3. Графики u/p и $\ln(1 + u/p)$ в функции от u/p.

что

$$\frac{u}{p} \geqslant \ln\left(1 + \frac{u}{p}\right) \tag{2.11}$$

везде, где существует $\ln\left(1 + \frac{u}{p}\right)$. Применяя (2.11) к (2.10), получаем:

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

Поэтому, используя (2.9), находим:

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

Умножая (2. 13) на $-K$, получаем неравенство, выражающее свойство B:

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

4. Свойство C

Информация I имеет максимум, когда все m вероятностей равны:

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

Для доказательства заметим, прежде всего, что информация является функцией только $(m-1)$ независимых переменных, так как

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

Возьмем в качестве $(m-1)$ независимых переменных p_1, p_2, ..., p_{m-1}. Информация $I(p_1, p_2, \ldots, p_{m-1})$ имеет максимум при условии, что, во-первых, все первые частные производные I равны нулю. Для получения второго условия составим определитель из вторых производных I для точки, определяемой первым условием:

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

где

$$I_{ij}=\frac{\partial^2 I}{\partial p_i\,\partial p_j}=I_{ji}\ \text{в экстремальной точке.}$$

Второе условие максимальности I состоит в том, что I_{11} и определители порядка 2, 3, ..., $(m-1)$, получаемые путем последовательного добавления следующих строк и столбцов в (2.17), должны образовать знакопеременную последовательность.

Из второй формулы (2. 16) имеем:

$$\frac{\partial p_m}{\partial p_j}=-1,$$

и следовательно,

$$\frac{\partial I}{\partial p_j}=-K(\ln p_j+1-\ln p_m-1)=-K\ln\left(\frac{p_j}{p_m}\right).$$

Применяя первое условие, находим, что для получения максимума необходимо:

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

откуда с помощью (2.16)

$$p_j=\frac{1}{m},\ j=1,\ 2,\ \ldots,\ m.$$

Вторые производные I в этой точке равны

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

$$\frac{\partial^2 I}{\partial p_i\,\partial p_j}=-K\frac{1}{p_m}=-Km,$$

и второе условие требует, чтобы определители вида

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

порядка 1, 2, ..., $(m-1)$ попеременно изменяли знак при возрастании порядка. Так оно и есть, так как значение такого определителя порядка n, как легко показать, есть

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

и свойство C, таким образом, доказано.

Свойство C может быть в простейших случаях показано графически. В случае двух возможностей, при $p_1+p_2=1$, мы имеем:

$$I=-K(p_1 \ln p_1+p_2 \ln p_2). \tag{2.18}$$

На рис. 2.4 I построена как функция p_1 (или p_2). I должно равняться нулю при $p_1=0$ и $p_2=1$. В силу симметрии I должна иметь максимум при $p_1=\frac{1}{2}=p_2$, и это максимальное значение равно

$$I=-K\left(\frac{1}{2} \ln \frac{1}{2}+\frac{1}{2} \ln \frac{1}{2}\right)=-\left(\frac{1}{\ln 2}\right)(-\ln 2)=1 \text{ дв. ед.}$$

Для трех возможностей

$$p_1+p_2+p_3=1, \tag{2.19}$$

и

$$I=-K(p_1 \ln p_1+p_2 \ln p_2+p_3 \ln p_3)$$

построим двумерную диаграмму (рис. 2.5). Двух измерений достаточно, так как только две из трех вероятностей p_1, p_2,

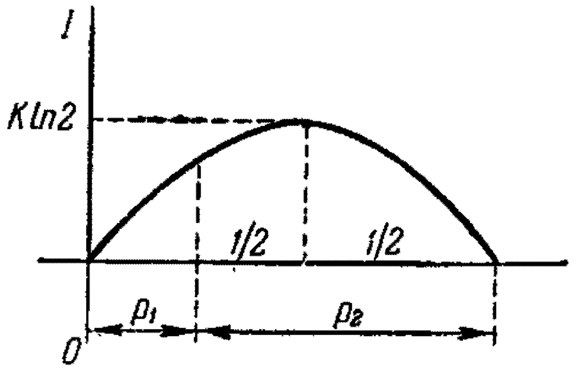

Рис. 2.4. Информация как функция $p_1=1-p_2$.

Максимум достигается при $p_1=p_2={}^1/_2$ и равен $K \ln 2=1$ дв. ед.

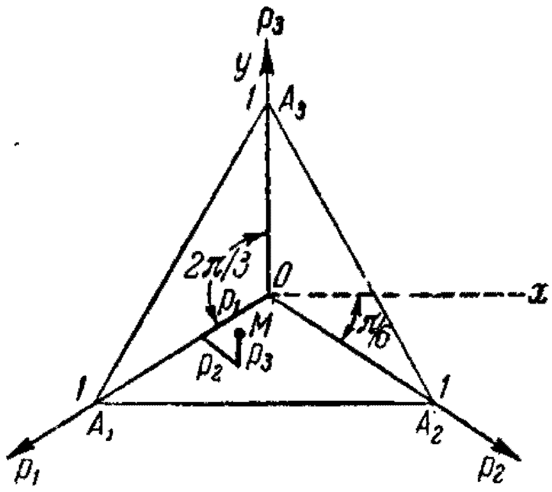

Рис. 2.5. Представление p_1, p_2 и p_3 на двумерной диаграмме.

Это возможно, так как только две из этих трех вероятностей независимы:

$$p_1+p_2+p_3=1.$$

p_3 независимы. Возьмем три направления под углом $2\pi/3$ друг к другу. Отложим отрезок p_1 по первому направлению, p_2 — по второму и p_3 — по третьему и получим, таким образом, точку M. По условию (2.19) точка M должна находиться внутри равностороннего треугольника A_1, A_2, A_3. Точка A_1 получается, например, когда $p_1 = 1$, $p_2 = p_3 = 0$. Линия A_1A_2 соответствует $p_1 + p_2 = 1$, $p_3 = 0$. Это легко показать, применяя на плоскости прямоугольные координаты и проектируя контур $Op_1p_2p_3M$ сперва на ось x, а затем на ось y. Координаты точки M таковы:

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

Уравнение A_1A_2 есть $y = -\frac{1}{2}$, что на основании (2.20) дает $p_3 = 0$, и следовательно, $p_1 + p_2 = 1$. Аналогичные результаты справедливы в силу симметрии и для остальных сторон треугольника.

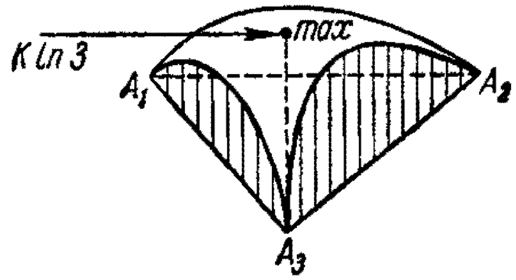

Рис. 2.6. Поверхность I (информации), построенная как функция p_1, p_2, p_3 при $p_1 + p_2 + p_3 = 1$.

Максимум I достигается при $p_1 = p_2 = p_3 = 1/3$ и равен $K \ln 3 = (\ln 3)/(\ln 2)$ дв. ед.

Мы можем воспользоваться треугольником $A_1A_2A_3$ как основанием. Каждая точка внутри треугольника соответствует совокупности чисел p_1, p_2 и p_3, удовлетворяющих условию (2.19). Отложим информацию по вертикали (в направлении, перпендикулярном к плоскости треугольника) и получим поверхность, показанную в перспективе на рис. 2.6. Вдоль каждой стороны треугольника получается кривая, сходная с кривой рис. 2.4. Наивысшая точка поверхности (т. е. максимум информации) получается при $p_1 = p_2 = p_3 = 1/3$, т. е. над центром треугольника $A_1A_2A_3$.

5. Зависимые события

Рассмотрим теперь проблему *зависимых событий*. Пусть даны две переменные x и y, могущие принимать значения

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

и пусть в некоторый момент мы выбираем значение для каждой из двух переменных. Так, мы можем иметь для момента t: $x=i$, $y=j$. Определим $p(i, j)$ как вероятность того, что $x=i$ и $y=j$ в один и тот же момент. Тогда

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

Если p_i есть вероятность того, что $x=i$ (вне зависимости от значения y), то

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

Аналогично, вероятность того, что $y=j$ (не рассматривая x), есть

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

Мы имеем следующие соотношения:

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

Перейдем к выражениям для информации, соответствующей различным вероятностям. Совместная информация есть

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

Информация для x в отдельности

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

и для y в отдельности

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

Сравним сумму $I(x)+I(y)$ с $I(x, y)$. Из (2.26) и (2.27) имеем:

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

Легко показать, используя свойство B, что

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

В самом деле, заменяя в (2.14) q_j на $p_i\ p_j$, получаем:

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

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

Ясно, что существование совместной вероятности $p(i, j)$ налагает ограничение, которое, как указывалось в главе 1, вызывает уменьшение информации. Равенство получается только, если $p_{i,j}=p_i p_j$, что означает, что события независимы.

6. Условная информация

Стремясь построить равенство, Шеннон применяет понятие *условной информации* путем введения вероятности $p_i(j)$ того, что $y=j$, если уже известно, что $x=i$. Мы называем $p_i(j)$ *условной вероятностью*. Ясно, что

$$p_i p_i(j)=p(i, j) \quad \text{или} \quad p_i(j)=\frac{p(i, j)}{p_i}. \tag{2.30}$$

Условная информация есть информация, связанная с $p_i(j)$; она обозначается $I_x(y)$ и определяется как

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

Легко показать, что $I(x, y)=I(x)+I_x(y)$, так как, раскрывая правую часть этого равенства, мы имеем из (2.30):

$$I(x)+I_x(y)=-K\left[\sum_{i,j}p(i,\ j)\ln p_i+\sum p(i,\ j)\ln p_i(j)\right]=$$
$$=-K\sum_{i,j}p(i,\ j)\ln[p_ip_i(j)]=I(x,\ y). \qquad (2.32)$$

Объединяя теперь (2.29) и (2.32), получаем:

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

или

$$I_x(y)\leqslant I(y). \qquad (2.33)$$

Условная информация $I_x(y)$ есть информация относительно y, если x уже известно. Знание x представляет собой ограничение, уменьшающее информацию при выборе y. Равенство получается только тогда, когда x и y являются независимыми событиями.

ГЛАВА 3

ИЗБЫТОЧНОСТЬ В ЯЗЫКЕ

1. Корреляция и зависимые события

Для того чтобы ввести в рассмотрение корреляцию, мы повторим некоторые определения и соотношения, относящиеся к зависимым событиям.

Мы рассматривали две переменные x и y, которые могут, например, принимать значения

$$x = 1,\ 2,\ 3,\ \ldots,\ i,\ \ldots,\ m,$$

$$y = 6,\ 9,\ 8,\ \ldots,\ j,\ \ldots,\ n,$$

и определили $p(i, j)$ как вероятность того, что в данный момент t $x = i$ и $y = j$. Соответствующая информация есть $I(x, y)$. Мы определили также другие вероятности. Величина $p_i = \sum_j p(i, j)$ равна вероятности того, что $x = i$ для любого y. С этой вероятностью связана информация $I(x)$. Величина $p_j = \sum_i p(i, j)$ — вероятность того, что $y = j$ для любого x. Связанная с этой вероятностью информация есть $I(y)$.

В предыдущей главе было найдено, что

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

Это неравенство следует из того факта, что при наложении на систему ограничений информация убывает.

Для получения равенства мы взяли *условную вероятность* $p_i(j)$, т. е. вероятность того, что $y = j$, когда известно, что $x = i$. С этой вероятностью связана условная информация $I_x(y)$.

Связь между вероятностями выражается соотношением

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

и, определив условную информацию, как

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

мы получили следующее равенство:

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

2. Корреляция в языке

Корреляция имеется во всех языках [1], так как, если появляется некоторая буква, то вероятность появления за ней некоторой другой буквы не равна ее априорной вероятности. Так, в английском языке, если дана буква t, то вероятность того, что непосредственно последующей буквой будет h, гораздо больше вероятности, что следующей буквой будет n. Аналогично, если дано сочетание tio, то вероятность появления n в качестве следующей буквы очень велика. Корреляция подобного рода определяется Шенноном [2],[3] как *избыточность* (redundancy).

Применим теперь соотношения предыдущего раздела к исследованию информационного содержания сообщения. Эти соотношения могут применяться, если корреляция имеется

---

[1]) В дополнение к приведенным ниже литературным ссылкам укажем следующие книги: Fletcher Pratt, Secret and Urgent, Blue Ribbon Books, NY, 1939, 1942 (содержит вероятностные данные об английском языке); G. Dewey, Relative Frequency of English Speech Sounds, Harvard U. P., Cambridge Mass., 1923; G. K. Zipf, Human Behavior and the Principle of Least Effort, Addison — Wesley, Cambridge, Mass., 1949 [в этой книге и других публикациях Ципф дает много кривых, вроде представленной на рис. (3.1), для различных языков].

[2]) C. E. Shannon, W. Weaver, The Mathematical Theory of Communication (стр. 3—28) U. of Illinois Press, Urbana, 1949.

Шеннонова «энтропия информации» отождествляется с нашей «мерой информации», но не с физической энтропией. Как мы увидим позднее, мера информации связана с «отрицательной энтропией».

[3]) C. E. Shannon, Prediction and Entropy of Printed English, BSTJ **30**, 50—64 (1951).

только между соседними символами. Для каждой пары букв мы связываем переменную x с первой буквой, а y — со второй. Тогда средняя информация на символ (или букву) будет условная информация

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

Мы предполагаем, что условия стационарны и что вероятности не меняются с течением времени.

Если между буквами нет корреляции, то, так как $p_i(j) = p_j$ и $\sum_i p_i = 1$, (3.1) переходит в обычную формулу:

$$I(y) = -K\sum_j p_j \ln p_j. \qquad (3.2)$$

Рассмотрение корреляции между отдельными буквами может быть легко распространено на сочетания букв, вроде упомянутого выше сочетания tio. Мы обозначаем сочетание $(N-1)$ букв как $b_i(N-1)$, а вероятность появления этого сочетания как $p(b_i(N-1))$. Обсудим теперь вероятность того, что за этим сочетанием последует некоторая буква j, иначе говоря, образуем новое сочетание из N букв:

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

Обозначим вероятность появления этого нового сочетания, рассматриваемого как целое, через

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

Условная вероятность $p_{b_i(N-1)}(j)$ есть вероятность того, что j последует за данным сочетанием $b_i(N-1)$. Это определение сходно с примененным в предыдущем разделе, и мы имеем аналогичную связь между вероятностями:

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

Определим теперь *среднюю информацию на символ* в последовательности, в которой учитывается корреляция на расстоянии до N символов:

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

Эта формула аналогична (3.1). Суммирование по i распространяется на все возможные сочетания из $(N-1)$ букв.

3. Избыточность в языке

Применяя наше определение, например, к английскому языку, мы хотим получить точное значение средней информации на букву. Это значение выражается как предел

$$I = \lim_{N\to\infty} F_N. \qquad (3.5)$$

При возрастании N учитывается большее количество ограничений, так что информация убывает. Последовательность F_N должна, таким образом, монотонно убывать, и ее предельное значение и есть действительная средняя информация на букву.

Вероятности отдельных букв, так же как двух- и трехбуквенных сочетаний, для английского языка были определены, но для сочетаний большего числа букв существующие данные недостаточны. Имеющиеся данные сведены в таблице, в которой F означают информацию, выраженную в двоичных единицах на букву:

1. Все буквы равновероятны (27 букв, включая интервал) $F_0 = 4{,}76$.
2. С учетом вероятностей отдельных букв $F_1 = 4{,}03$.
3. С учетом данных о двухбуквенных сочетаниях $F_2 = 3{,}32$.
4. С учетом данных о трехбуквенных сочетаниях $F_3 = 3{,}1$.

Избыточность выражается через F следующим образом:

$$R = 1 - \frac{F_{\lim}}{F_0}. \qquad (3.6)$$

Однако $F_{\lim}$ неизвестно, так как имеется очень мало сведений о сочетаниях более чем из трех букв. Для оценки значения $F_{\lim}$ мы можем привлечь другого рода статистику, а именно, частоту слов. Таблицы частоты слов были составлены как для целей кодирования, так и для эффективного обучения языку.

Частота слова характеризуется его порядковым номером в списке, составленном в порядке частоты появления. На

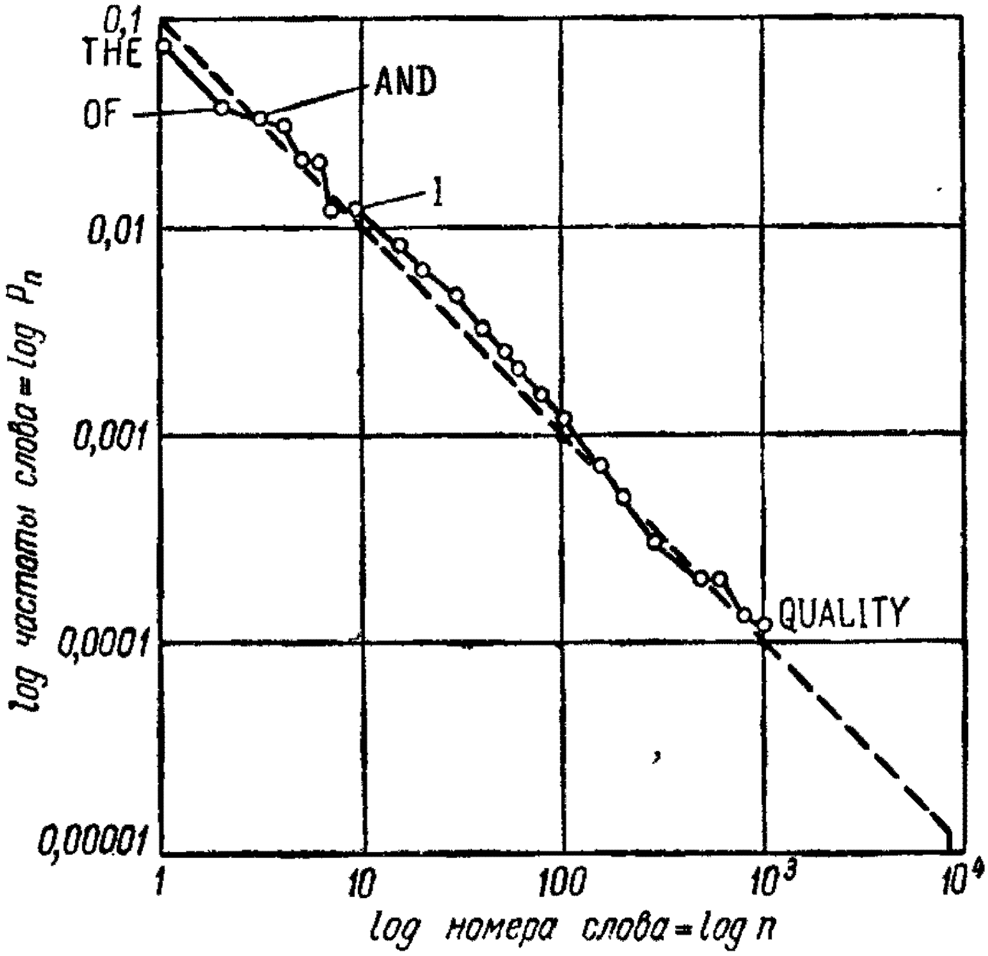

Рис. 3.1. График (в логарифмическом масштабе по обеим осям) частоты слов в зависимости от порядкового номера для первых 8727 слов.

рис. 3.1 логарифм частоты построен в зависимости от логарифма порядкового номера, начиная с the и кончая словом за № 8727 для того, чтобы сумма частот равнялась единице. Как видим, график представляет практически прямую линию (в двойном логарифмическом масштабе) и может быть хорошо аппроксимирован формулой

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

Применяя это выражение, находим, что

$$-K \sum_n p_n \ln p_n = 11,82 \text{ дв. ед. на слово.} \tag{3.8}$$

Далее, если считать, что среднее английское слово (включая интервал между словами) содержит 5,5 буквы, то получаем предельное значение $F_{\text{word}} = 2{,}14$ дв. ед./букву. Эта величина не та же, что F_{lim}, так как имеются еще дополнительные ограничения, обусловленные связями между последовательными словами.

4. Некоторые эксперименты

Для получения более подробных данных предложена игра: берется предложение и участнику, не знающему предварительно предложения, предлагается угадывать буквы (включая интервалы), начиная с первой. После каждой попытки дается ответ «да» или «нет», и игра продолжается. Число попыток для каждой буквы записывается, суммируется и делится на общее число букв в предложении. Это дает число двоичных единиц на букву, так как каждое угадывание с ответом «да» или «нет» представляет как раз одну двоичную единицу информации.

Некоторые примеры такого рода были опубликованы Шенноном[1]). В одном предложении из 102 букв общее число попыток, потребовавшееся для восстановления предложения, составило 198. Это соответствует 1,94 дв. ед. на букву.

В другом примере правила игры были иные. После каждого угадывания давался ответ «да», если буква была угадана правильно, в противном случае сообщалась действительная буква. В предложении из 129 букв 89 были угаданы правильно, а 40 были сообщены. Пользуясь информацией F_1, равной 4 дв. ед. на букву (как указано ранее), можно получить общее число двоичных единиц

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

или 1,93 дв. ед. на букву.

Третий пример дает около 1,9 дв. ед. на букву. Число 1,9 должно рассматриваться как завышенное значение, так как во многих случаях угадываемые буквы были очевидны, и правильное угадывание, сопровождаемое утвердительным

---

[1]) C. E. Shannon, BSTJ **30**, 54—58 (1951).

ответом, едва ли могло считаться за реальную единицу информации[1]).

На рис. 3.2 представлен данный Шенноном график, на котором число дв. ед. на букву отложено как функция числа букв, вплоть до 15 букв, с экстраполяцией до 100 букв.

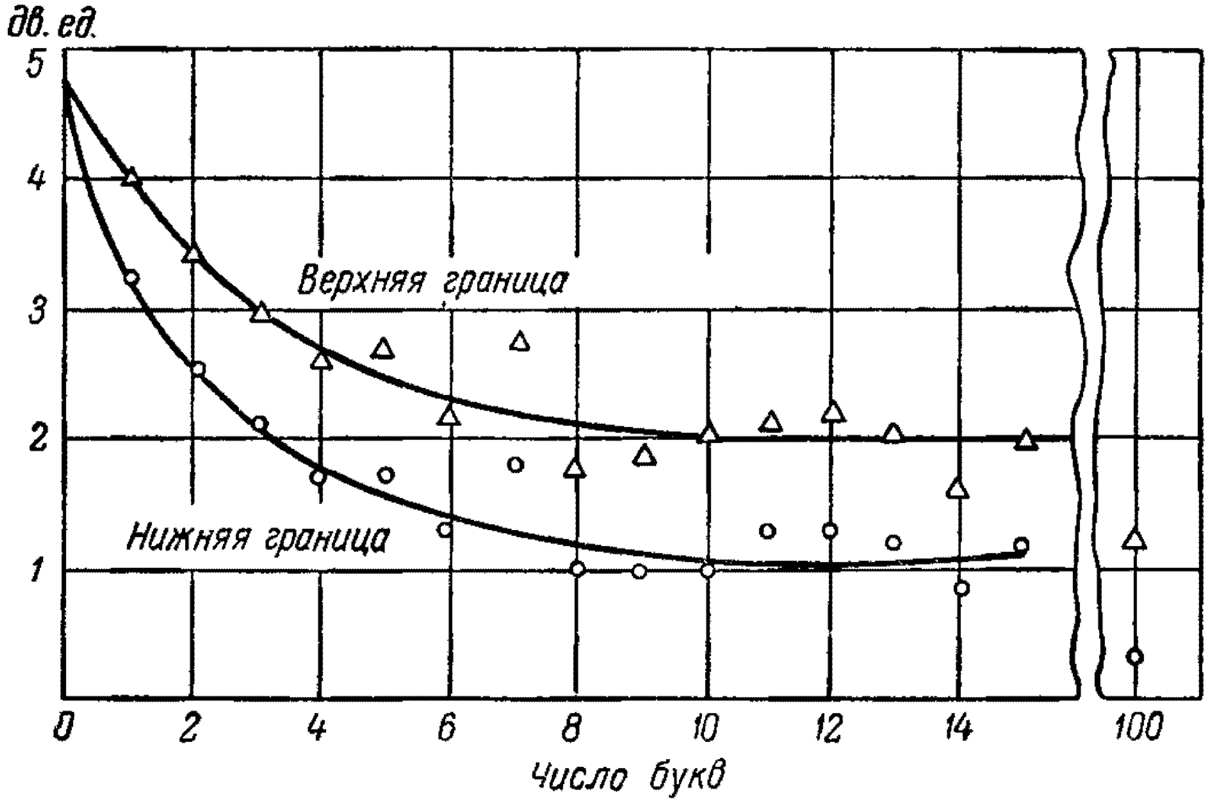

Рис. 3.2. Число двоичных единиц на букву, построенное как функция числа букв до 15 букв с экстраполяцией до 100 букв.

Принимая экстраполированное значение 1,4 дв. ед. на букву и сравнивая его с $F_0 = 4{,}76$ дв. ед. на букву, мы видим, что английский язык имеет значительную избыточность.

---

[1]) Основная идея эксперимента состоит в том, что испытуемый худо или хорошо, сознательно или нет учитывает *все* статистические взаимосвязи языка, на чем и основана возможность приближения к $F_{\lim}$. Методика опыта могла бы быть усовершенствована следующим образом: перед каждым очередным угадыванием испытуемый делит все возможные буквы на две группы с примерно равными вероятностями (по собственной оценке). При этом выбор одной из двух групп дает около одной единицы информации. Если выбор очередной буквы не вызывает сомнений, то количество информации, связанное с этим выбором, считается равным нулю. Распределение на группы делается каждый раз заново и зависит, разумеется, от всего предшествующего текста. Такая процедура представляет, как нетрудно видеть, перенос в эксперимент методики кодирования по Шеннону—Фано. (*Прим. перев.*)

По-видимому, возможно сокращение раза в три, которое могло бы сделать язык более точным. Нужно, однако, учитывать и другие соображения, как, например, разборчивость и возможность понимания речи при наличии шума.

5. Кодирующие устройства

Соображения предшествующих разделов естественно ведут к проблемам кодирования и построения кодирующих устройств. Так, например, кодирующее устройство должно обладать памятью для уменьшения избыточности поступающих букв; иначе говоря, кодирующее устройство должно «вспоминать» предыдущие буквы для того, чтобы создать кодовое обозначение следующей поступающей буквы. Если мы обозначим: α_n — состояние памяти в данный момент, x_n — следующая буква, $y_n = f(x_n, \alpha_n)$ — кодовое обозначение буквы x_n, то память должна принять новое состояние, также зависящее от x_n и α_n:

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

Результат наших вероятностных рассуждений по поводу зависимых событий кратко формулируется в 7-й теореме Шеннона:

«Кодирующая система может уменьшить, а в лучшем случае сохранить имеющуюся информацию».

Обратимое кодирующее устройство, могущее применяться как для кодирования, так и для декодирования, сохраняет информацию неизменной. Информация может быть отнесена либо к целому предложению, либо к единице времени, если сообщение создается непрерывно некоторым источником.

Результат Шеннона сходен с положением в термодинамике, где обратимый процесс не изменяет энтропии, тогда как при необратимом процессе энтропия всегда возрастает. Методы кодирования мы будем обсуждать в следующей главе.

ГЛАВА 4

ПРИНЦИПЫ КОДИРОВАНИЯ; ПРОПУСКНАЯ СПОСОБНОСТЬ КАНАЛА

1. Введение

В предыдущих главах мы определили информационное содержание некоторого сообщения и обсудили роль избыточности, уменьшающей действительное количество информации, приходящейся на одну букву письменного сообщения. Если i есть средняя информация на букву, и если источник выдает буквы со скоростью m букв в секунду, то поток (скорость создания) информации равен

$$I' = mi \text{ дв. ед. в секунду.} \tag{4.1}$$

Может оказаться удобным заменить алфавит другой совокупностью символов для того, чтобы уменьшить избыточность и сократить число символов, передаваемых за одну секунду. Общая процедура, которую нужно применить для выполнения такой операции кодирования, описана в конце предыдущей главы. Если применяемый преобразователь обратим, то поток информации I' сохраняется, и при кодировании не происходит потери информации. Аналогичный преобразователь (действующий в обратном направлении) может применяться для декодирования сообщения и превращения его снова в письменный текст. Если наш преобразователь необратим, он может потерять часть информации, что поведет к ошибкам и недоразумениям. Итак, первое требование к преобразователю состоит в том, что он должен быть обратимым.

Как построить наиболее эффективный преобразователь? Общего ответа на этот вопрос нет; ответ зависит от того, каково предполагаемое использование кодированных сигналов. Мы должны, прежде всего, обсудить свойства *канала*, по которому сигналы должны передаваться (если речь идет о проблемах связи), или запасаться (если мы имеем в виду запоминающее устройство вычислительной машины).

Наилучшим является тот метод кодирования, который эффективно согласовывает поток информации с пропускной способностью канала. Сначала обсудим проблему канала и его пропускной способности.

2. Определение канала и его пропускной способности

Если мы желаем передавать информацию или запасать ее в запоминающем устройстве, то мы, прежде всего, выбираем некоторую физическую среду, кабель, линию радиосвязи, ртутную линию, магнитную ленту и т. п. и должны изучить физические возможности этой среды. Так, если мы применяем кабель с полосой пропускания $\Delta\nu$, то кратчайший сигнал, который может быть передан, имеет длительность Δt, причем (см. главу 8)

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

Затем возникает техническая проблема построения системы символов, согласованных с физической средой и удовлетворяющих нас с точки зрения применения их для связи: точки, тире и паузы, сигналы различной интенсивности или полярности и т. д. Если этот выбор сделан, то тем самым определен канал.

Для обсуждения вопроса со статистической точки зрения достаточно рассмотреть применяемую совокупность n дискретных символов $S_1, S_2, \ldots, S_j, \ldots, S_n$ и длины (или длительности) $t_1, t_2, \ldots, t_j, \ldots, t_n$ этих символов. В дальнейшем мы можем не думать о физическом строении системы передачи и о технических или практических основаниях для выбора различных символов S_j и их длительностей t_j. Для всех последующих теоретических и практических рассуждений *канал* определен просто совокупностью символов S_j, имеющих длительности t_j. Мы можем игнорировать физическую природу символов; только их длительность играет роль в обсуждении. Мы не вводим никаких ограничений длительности; возможно, что некоторое количество символов имеет одинаковую длительность. Предполагается, что символы S_j полностью независимы друг от друга и определены так, что между ними отсутствует какая-либо корреляция. Предполагается, что произвольная последовательность символов способна представлять некоторое сообщение. Полная длительность сообщения T есть попросту сумма длительностей t_j отдельных символов, составляющих сообщение.

Рассмотрим все возможные сообщения, имеющие полную длительность T. Эти сообщения будут соответствовать всем возможным комбинациям символов S_j, имеющим общую длительность T. Обозначим через $N(T)$ число таких отдельных сообщений, различающихся между собой выбором символов и порядком их размещения. Применение различных символов не связано никакими условиями, ограничениями или корреляцией. Все N сообщений длительности T рассматриваются как равновероятные априори. Тогда, если мы выберем одно из этих сообщений, то получим информацию

$$I = K \ln N,$$

в соответствии с (1.1), где выбор единиц определяется выбором постоянной K.

Скорость, с которой информация передается по каналу, есть I/T. Эта скорость меняется с изменением общей длительности T, но можно показать, что во всех практических случаях она стремится к пределу при неограниченном возрастании T. Шеннон [1]) выбирает этот предел в качестве определения пропускной способности канала:

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

Если мы применяем единицы энтропии, то $K = k$ — постоянной Больцмана. В большинстве теоретико-информационных задач мы будем измерять информацию в двоичных единицах, к которым переходим, положив

$$K = \log_2 e \tag{4.4}$$

или

$$C = \lim_{T \to \infty} \frac{\log_2 N(T)}{T}, \tag{4.5}$$

---

[1]) C. E. Shannon, W. Weaver, The Mathematical Theory of Communication, p. 7, U. of Illinois Press, Urbana, Ill., 1949.

и пропускная способность будет выражена в двоичных единицах в секунду, если T измерено в секундах.

В последующих разделах этой главы мы будем обсуждать пригодность этого определения и условие существования предела C. Затем мы покажем, что пропускная способность канала соответствует наибольшему среднему потоку информации в канале, когда все возможные комбинации символов разрешены.

Наконец, мы исследуем вопрос о том, как кодировать сообщения для того, чтобы полностью использовать пропускную способность канала. Это приведет к практической постановке задачи кодирования: дана совокупность сообщений, написанных на некотором языке с известным алфавитом; задача состоит в том, чтобы согласовать язык с каналом и приблизиться, по возможности, к скорости передачи информации, соответствующей пропускной способности C канала. Мы докажем, что эта задача разрешима, и обсудим некоторые общие условия, которые должны быть выполнены при кодировании для достижения теоретического предела.

3. Символы, слова и сообщения при кодировании последовательностей

Предположим, что мы выбрали символы и что мы хотим воспользоваться комбинациями этих символов для построения слов или сообщений. Это представляет типичный случай кодирования последовательностей. Наши символы могут соответствовать буквам, которые, сочетаясь, образуют слова, или они могут представлять цифры, последовательность которых дает некоторое число. Более общая задача включает случай цифр или букв, применяемых для представления целых предложений. В случае речи каждое слово представлено комбинацией элементарных звуков, или фонем, которые играют роль символов.

Каждый символ имеет определенную *цену*, и способ подсчета этих цен зависит от физической системы передачи. В наших первых задачах мы будем полагать, что цена соответствует длительности сигнала, представляющего символ. Это есть главная проблема связи, где необходимая мощность мала и где задача сводится к передаче возможно большего количества сообщений за некоторое время T. Каждый символ

S_j имеет определенную длительность t_j, и многие символы могут иметь одинаковую длительность; это имеет место, в частности, если применяются импульсы переменной интенсивности или различной частоты. Аналогично обстоит дело и с телеграфными сигналами, состоящими из комбинаций посылок и пауз, когда все комбинации имеют одинаковую длину.

Мы не предполагаем каких-либо ограничений, наложенных на возможные последовательности символов. Если мы применяем буквы, то это значит, что мы допускаем последовательности, непроизносимые и не имеющие смысла ни в каком языке. Среди наших символов мы можем выбрать один какой-либо для обозначения пробела, или можно использовать для этой цели некоторую комбинацию символов. Рассортируем теперь наши символы по цене (или длительности). Мы имеем:

$$\left.\begin{array}{llll} a_1 & \text{символов длительностью} & t_1, & \\ a_2 & \text{»} \qquad\qquad \text{»} & t_2, & \\ a_j & \text{»} \qquad\qquad \text{»} & t_j, & a_j \geqslant 0, \\ a_m & \text{»} \qquad\qquad \text{»} & t_m = \sigma. & \end{array}\right\} \tag{4.6}$$

Если σ означает наибольшую возможную длительность символа, а n есть полное число символов, то

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

Величины a_j — неотрицательные целые числа, и мы можем считать длительности t_j также целыми. Это предположение отвечает действительности, так как длительности всегда определяются лишь с определенной погрешностью.

Следуя программе, намеченной в п. 2, мы подсчитаем теперь число $N(t)$ различных сообщений общей длительностью t. Это число N удовлетворяет соотношению:

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

которое означает, что сообщения длительностью $(t-t_1)$ могут быть дополнены любым из a_1 символов длительностью t_1, для образования сообщения общей длительностью t, и аналогично для t_2, t_3, ..., t_m.

Равенство (4.8) представляет собой линейное уравнение в конечных разностях. Оно имеет σ независимых решений (см. приложение в конце этой главы) вида

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

а общее решение выражается линейной комбинацией

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

с постоянными коэффициентами A_k. Подставляя частное решение $N(t) = X^t$ в (4.8), мы убеждаемся, что X_k являются корнями уравнения

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

или

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

Это называется характеристическим уравнением уравнения (4.8). Умножая на X^σ, чтобы избавиться от отрицательных степеней X, получаем алгебраическое уравнение σ-й степени:

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

Здесь имеется только одна перемена знака и, следовательно, по правилу Декарта, один положительный корень. Так как наша задача требует по меньшей мере двух символов, то многочлен отрицателен при $X = 1$, так что

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

Многочлен становится положительным при очень большом X. Поэтому мы имеем:

$$\left.\begin{array}{r} \text{один вещественный положительный корень,} \\ \text{больший единицы: } X_1 > 1, \\ \sigma - 1 \text{ отрицательных или комплексных корней:} \\ X_2, X_3, \ldots, X_\sigma. \end{array}\right\} \qquad (4.13)$$

4. Обсуждение

Предыдущая постановка задачи дана впервые Шенноном, подробное обсуждение можно найти в диссертации Мандельброта [1]).

Мы имеем общее решение (4.10), содержащее σ постоянных коэффициентов A_k. Эти коэффициенты определяются из начальных условий, которые мы сейчас обсудим. Рассмотрим (4.8) при малых значениях t. Один из членов соответствует

$$t_j = t \quad \text{или} \quad t - t_j = 0,$$

и мы должны положить $N(0) = 1$, так как очевидно, что мы можем представить некоторые из сообщений общей длительностью $t = t_j$ при помощи a_j символов длительностью t_j. Бо́льшие значения $t_{j+1}, t_{j+2}, \ldots, t_m$ не могут применяться. Это утверждение справедливо для всех значений j. Это условие удовлетворяется, если взять

$$\begin{gathered} N(0) = 1, \\ N(-1) = N(-2) = \ldots = N(-\sigma + 1) = 0. \end{gathered} \qquad (4.14)$$

Мы получаем, таким образом, σ начальных условий, определяющих σ произвольных коэффициентов A_k. Значения функции $N(t)$ для $t = -\sigma, -\sigma - 1, \ldots$ никогда не встретятся в (4.8) и не играют роли в нашей задаче.

---

[1]) B. Mandelbrot, Contribution à la théorie mathématique des jeux de communication, Ph. D. thesis, Paris, Dec. 16, 1952, Publ. de l'Inst. de statistique de l'Univ. de Paris, vol. 2, № 1, 2, pp. 80—102, 1953; см. также Trans. IRE (PGIT) 3, 124 (1954).

Выделим положительный корень X_1 и запишем:

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

Первый член растет экспоненциально с t и является преобладающим. Остальные члены имеют колебательный характер. Если корень X_k отрицателен, он дает попеременно положительные и отрицательные члены X_k^t по мере того, как t увеличивается каждый раз на единицу (t всегда целое). Если X_k — комплексный корень, то его комплексно-сопряженное X_k^* также является корнем, и эта пара корней, вместе взятых, при возрастании t дает колебание.

Решение (4.15) устойчиво, если все эти колебательные члены имеют убывающую амплитуду, и неустойчиво, если амплитуда колебаний возрастает. Мы говорим, что наше решение

$$\left.\begin{array}{l} \text{абсолютно устойчиво, если } |X_j| < 1,\ j \geqslant 2, \\ \text{относительно устойчиво, если } |X_j| < X_1,\ j \geqslant 2, \\ \text{неустойчиво, если } |X_j| \geqslant X_1 \text{ хотя бы для} \\ \qquad\qquad\qquad\qquad \text{одного } j \geqslant 2. \end{array}\right\} \qquad (4.16)$$

В первом случае амплитуда колебаний убывает, во втором случае амплитуда колебаний возрастает медленнее, чем преобладающий член.

Начальное условие (4.14) дает:

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

В первых двух случаях (абсолютная или относительная устойчивость) мы можем во многих задачах игнорировать колебания и удерживать в (4.15) только преобладающий член:

$$N'(t) = A_1 X_1^t = A_1 e^{\beta t}, \qquad (4.18)$$

где $\beta = \ln X_1$.

Итак, $N(t)$ представляет число сообщений, полная длина которых равна t. Но нас может интересовать число $N_1(t)$ сообщений, длина которых меньше или равна t. Имея в виду, что t всегда целое, имеем:

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

Мы не включили в сумму слагаемое $s=0$, так как сообщение нулевой длины не имеет смысла. Используя соотношение

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

получаем:

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

где

$$B_k=\frac{A_k X_k}{X_k-1}=\frac{A_k}{1-X_k^{-1}}.$$

Мы можем переписать этот результат следующим образом:

$$N_1(t)=B_1 X_1^t+\sum_{k=2}^{k=\sigma} B_k X_k^t-B_0, \quad (4.22)$$

где

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

Число N_1 сообщений, длина которых меньше чем t, или равна t, содержит преобладающий член с X_1^t, колебательные члены ($k=2, 3, \ldots, \sigma$) и постоянную B_0. Отбрасывая колебательные члены, имеем:

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

что является хорошим приближением в устойчивых случаях для больших t. Мы получаем, таким образом,

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

где

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

5. Примеры

Покажем действие метода на нескольких примерах.

A. Все символы имеют одинаковую длину

В этом случае мы берем $a_1 = n$ — общему числу символов длины $t_1 = \sigma = 1$. Равенство (4.12) дает:

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

Имеется один вещественный положительный корень $X = n$, и следовательно,

$$N = n^t \text{ и } N_1 = \frac{n^t - 1}{1 - n^{-1}}.$$

Этот канал обладает абсолютной устойчивостью без каких-либо колебаний.

B. Символы двух различных длин: 1 и σ

B. 1. Степень характеристического уравнения равна σ. В общем случае

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

B. 2. Простейшие соотношения получаем при $\sigma = 2$:

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

Корни этого уравнения равны

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

Так как $|X_2| < X_1$, то относительная устойчивость обеспечена. Для абсолютной устойчивости необходимо

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

Это условие выполняется для канала с двумя символами:

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

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

В. 3. Рассмотрим случай, когда длинные сигналы втрое длиннее коротких:

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

Характеристическое уравнение имеет вид

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

Берем

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

и получаем:

$$Y^3 + pY + q = 0, \tag{4.32}$$

где $p = a_1/a_3$ и $q = -1/a_3$. Это — хорошо известное уравнение. Так как величина $27q^2 + 4p^3$ положительна, то имеется один положительный и два комплексных корня:

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

где

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

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

Мы выбираем для A и B вещественные кубические корни, так что Y_1 вещественно, тогда как Y_2 и Y_3 комплексно-сопряженные. Относительная устойчивость получается при

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

Легко подсчитать:

$$|Y_2|^2 = Y_2 \tilde{Y}_2 = Y_2 Y_3 = A^2 + B^2 - AB,$$

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

Но

$$AB = -\frac{p}{3} < 0,$$

и таким образом, условие (4.34) относительной устойчивости выполнено. Для абсолютной устойчивости требуется:

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

— условие, выполняемое, когда $|A|$ и $|B|$ не слишком малы. Имеем:

$$A^3 + B^3 = (A + B)(A^2 + B^2 - AB) = -q = \frac{1}{a_3},$$

откуда

$$Y_1 |Y_2|^2 = \frac{1}{a_3}.$$

Но Y_1 меньше, чем 1 ($X_1 > 1$), так что из (4.35) следует легко выполнимое уравнение:

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

C. По одному символу каждой длины

В этом случае

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

при длинах

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

Характеристическое уравнение (4.12) принимает вид

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

или

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

Умножая на $(X - 1)$, получаем:

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

Это уравнение имеет лишний корень $X = 1$. Уравнение (4.37) имеет только один положительный корень в промежутке между 1,62 (случай $\sigma=2$, (см. (4.30)) и 2 (для $\sigma\to\infty$). Основываясь на (4.38), мы запишем:

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

Имеется один отрицательный корень, когда σ четно, а при нечетном σ нет отрицательных корней. Это легко показать, применяя к (4.38) декартово правило знаков:

$$X'=-X\begin{cases}\sigma \text{ четно: } -X'^{\sigma+1}-2X'^{\sigma}+1=0, & \text{одна перемена знака,}\\ \sigma \text{ нечетно: } X'^{\sigma+1}+2X'^{\sigma}+1=0, & \text{нет перемены знака.}\end{cases} \tag{4.40}$$

Все остальные корни комплексны. Когда σ четно, отрицательный корень находится между $X'=0$ и $X'=1$ (что соответствует $X=-1$) и, следовательно, не влияет на устойчивость решения.

При больших σ комплексные корни близки к корням σ-й степени из единицы; запишем:

$$X=e^{\frac{2n\pi i}{\sigma}+\varepsilon}, \tag{4.41}$$

где n — целое, а ε — малая величина. Это значение X удовлетворяет (4.38), если ε выражено приближенным соотношением:

$$e^{\sigma\varepsilon}\approx\frac{1}{2-e^{2n\pi i/\sigma}},\quad \sigma\gg 1.$$

Это выражение дает для ε малое значение, если σ велико. При этом комплексные корни по абсолютной величине будут близки к единице и, следовательно, относительная устойчивость обеспечена.

6. Вычисление пропускной способности канала

Мы можем теперь найти пропускную способность канала согласно (4.3). Рассмотрим случай кодирования последовательностей (см. п. 3) в предположении, что система устойчива. При очень большом T решение сводится к преобладающему члену (4.18):

$$N(T) \to A_1 X_1^T = Ae^{\beta T}, \text{ при } T \to \infty, \tag{4.42}$$

откуда

$$\ln N \to \beta T + \ln A_1,$$

где $\beta = \ln X_1$. Теперь с помощью (4.3) получаем для пропускной способности канала

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

Шеннон рассматривает также задачи с ограничениями, наложенными на возможные последовательности символов S_i, но мы удовольствуемся предшествующим обсуждением задачи без ограничений.

Определив таким образом пропускную способность канала, Шеннон вводит свое определение меры информации, которое мы обсуждали в предыдущих главах. Затем он доказывает основную теорему для канала без шумов:

«Если канал имеет пропускную способность C дв. ед. в секунду и получает сообщение от источника в количестве i дв. ед. на символ, то наилучшая система кодирования позволит использовать канал со скоростью $R = C/i$ символов в секунду».

Доказательство этой теоремы последует из рассуждений дальнейших разделов.

7. Согласование кода с каналом

Мы намереваемся обсудить более подробно физический смысл важной теоремы Шеннона и уточнить свойства систем кодирования, позволяющих полностью использовать пропускную способность канала. Будет показано, что наилучшим является такое кодирование, которое дает наиболее вероятное распределение символов [1]), и будут даны правила для нахождения таких оптимальных кодирующих систем.

---

[1]) Нужно заметить, что в этом разделе автор весьма нечетко разграничивает вероятностные и статистические категории. Кроме того, он пользуется неупотребительной терминологией, применяя, например,

Повторим некоторые результаты предыдущих глав. Мы предполагали, что в нашей системе связи применяется алфавит из n различных символов, и мы подсчитали общее число сообщений, т. е. различных комбинаций этих символов, когда числа $N_1, N_2, \ldots, N_j, \ldots, N_n$ различных символов, т. е. число повторений данного символа, были наперед заданы. Пусть

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

есть общее число символов в сообщении и пусть

$$p_j = \frac{N_j}{G_0}, \text{ где } \sum_{j=1}^{j=n} p_j = 1, \tag{4.45}$$

представляет относительную частоту j-го символа. Полное число различных сообщений рассматриваемого типа равно (см. (1.26))

$$P = \frac{G_0!}{\prod\limits_j N_j!}.$$

Применяя формулу Стирлинга, получаем приближенно:

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

Эта формула была использована для определения количества информации, получаемого при выборе одного сообщения из P возможных (см. (1.28)), и для информации i на символ мы получили:

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

для обозначения частоты (в смысле математической статистики) термины density и даже proportion. Под распределением автор понимает здесь и далее распределение частот, т. е. то, что носит название распределения выборки. Под «наиболее вероятным распределением» он подразумевает распределение выборки, совпадающее с распределением вероятностей. (*Прим. перев.*)

А. Символы одинаковой длины

Рассмотрим сначала простой пример, предположив, что все символы имеют одинаковую длину t_1. Длинный интервал T содержит

$$G = \frac{T}{t_1}$$

позиций (или ячеек) для различных символов. Для каждой ячейки мы располагаем n различными символами, так что полное число возможных сообщений равно

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

Из определения (4.3) получаем для пропускной способности канала

$$C = \frac{K \ln n}{t_1}. \tag{4.48a}$$

При рассмотрении пропускной способности не делается никаких предположений относительно априорных частот n различных символов. Когда мы определяли информацию, априорные частоты $p_1, p_2, \ldots, p_j, \ldots, p_n$ различных символов предполагались известными, и мы получили для информации на символ выражение (4.47). Эта величина имеет максимум, когда все p_j равны:

$$p_1 = p_2 = \ldots = p_j = \ldots = p_n = \frac{1}{n},$$

и максимальное значение равно

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

так как сумма имеет n одинаковых членов. Это есть информация на символ, когда все символы равновероятны. Каждый символ имеет длительность t_1, и скорость передачи информации есть

$$\frac{i}{t_1} = \frac{K \ln n}{t_1} \quad \text{дв. ед. в секунду.} \tag{4.49}$$

Это выражение совпадает с формулой (4.48a) для пропускной способности. Сходство (4.48a) и (4.49) обращает на себя внимание. При вычислении пропускной способности мы использовали все возможные распределения символов. При выводе (4.49) мы исходили из произвольной априорной совокупности частот, которые позднее подобрали (положив $p_j = \frac{1}{n}$) так, чтобы максимизировать информацию, и получили одинаковые результаты.

На основании (4.48а) и (4.49) скорость передачи равна

$$R = \frac{C}{i} = \frac{1}{t_1}, \tag{4.50}$$

и мы не можем, разумеется, передавать больше, чем $1/t_1$ символов (длительностью t_1) в секунду. Итак, в нашей упрощенной задаче наилучшим является такое кодирование, при котором все символы встречаются одинаково часто, в соответствии с теоремой Шеннона.

Отметим, что условие максимума информации приводит к наиболее вероятному распределению. Действительно, $N(t)$ есть полное число возможных сообщений, а P, выражаемое равенством

$$\ln P = -\sum_{j=1}^{j=n} p_j \ln p_j = \frac{1}{K},$$

есть число возможных сообщений с распределением символов, определяемых вероятностями p_j. Делая P максимальным (или, что то же, делая максимальным i), делаем максимальной информацию. Но P/N (N считается постоянным) есть вероятность того, что имеет место распределение P (определяемое совокупностью p_j), а если P имеет максимум, то мы имеем наиболее вероятное распределение.

Для рассматриваемого в этом разделе упрощенного случая (все символы имеют одинаковую длину) мы приходим к заключению, что максимум информации получается при наиболее вероятном распределении. Можно также заметить, что это распределение является асимптотическим; это основывается на выводе равенства (4.48), при котором мы не учитывали априорного распределения. Это асимптотическое распределение можно себе представить как среднее распределение в том смысле, что при усреднении за большое время только асимптотическое распределение имеет смысл. Мы увидим в следующем разделе, что эти замечания приложимы и в случае, когда символы имеют неодинаковую длительность.

8. Общая задача: символы различной длины

Мы переходим к рассмотрению общей задачи с символами различной длины; докажем, что предшествующий результат сохраняет силу: наиболее эффективным является такое кодирование, которое дает различным символам наиболее вероятное априорное распределение частот. Мы выведем также это наиболее вероятное распределение и покажем его связь с (4.11) и (4.12).

Для случая символов разной длины рассуждение оказывается несколько более сложным. Пусть сообщение имеет полную длину T и пусть применяется $N_1, N_2, \ldots, N_j, \ldots, N_n$ символов соответственно типа $1, 2, \ldots, j, \ldots, n$. Полная длительность сообщения есть

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

Введем общее число символов G_0 и относительные частоты p_j:

$$G_0 = \sum_{j=1}^{j=n} N_j, \quad p_j = \frac{N_j}{G_0}, \tag{4.52}$$

откуда

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

Перепишем (4.51) в следующем виде:

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

Логарифм полного числа P различных сообщений, которые можно составить, переставляя различные символы, по-прежнему выражается формулой (4.46):

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

Информация, полученная при выборе одного из P сообщений, равна $K \ln P$, в соответствии с нашим определением (1.1). Все эти сообщения имеют одинаковую длительность T, так как они состоят из одинаковых чисел $N_1, N_2, \ldots, N_n$ различных символов и могут считаться априори равновероятными. Мы желаем найти наиболее вероятное распределение, или, что то же, сделать максимальной информацию, содержащуюся в сообщении. Это значит, что мы делаем максимальным $\ln P$, учитывая условия (4.53) и (4.54). Максимум находится методом множителей Лагранжа. Запишем условие[1])

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

где два произвольных множителя α и β будут определены ниже. Переменными являются $G_0, p_1, p_2, \ldots, p_j, \ldots, p_n$. Дифференцируя (4.55), получаем:

$$d(\ln P) = -(dG_0)\sum_{j=1}^{j=n} p_j \ln p_j - G_0 \sum_{j=1}^{j=n} (1 + \ln p_j)\, dp_j.$$

Подставляя это выражение в (4.56), находим:

$$-dG_0\left(\sum_{j=1}^{j=n} p_j \ln p_j + \beta \sum_{j=1}^{j=n} p_j t_j\right) + \\ + \sum_{j=1}^{j=n} dp_j\left[-\alpha - \beta G_0 t_j - G_0(1 + \ln p_j)\right] = 0. \tag{4.57}$$

Мы должны рассматривать теперь dG_0 и dp_j как произвольные независимые величины. Полный дифференциал может быть равен нулю, только если все скобки равны нулю:

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

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

---

[1]) Условие (4.56) устанавливает, что $\ln P$ имеет экстремум, но не доказывает, что этот экстремум есть максимум. Доказательство того, что это действительно есть максимум, может быть проведено методом, примененным при доказательстве свойства C (см. раздел 3 главы 2).

Подстановка (4.59) в (4.58) дает:

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

Учитывая условие (4.53), мы должны выбрать

$$1+\frac{\alpha}{G_0}=0 \quad \text{или} \quad \alpha=-G_0, \tag{4.60}$$

и теперь (4.59) сводится к

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

Произвольная постоянная β определяется из условия (4.53):

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

Запишем

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

и вспомним, что мы можем иметь более чем один символ определенной длительности. Пусть имеется a_j символов длительности t_j. Тогда (4.62) запишется в виде

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

а это совпадает с (4.11) (см. п. 3).

Если рассмотреть бесконечно длинные сообщения, то снова обнаружится совпадение асимптотического распределения п. 3 и распределения, дающего максимум информации. Это последнее (см. распределение п. 7) является также наиболее вероятным распределением, равно как и среднее распределение. Распределение (4.61) дает число сообщений (4.55):

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

$$=G_0\beta\sum_{j=1}^{j=n} t_j p_j, \tag{4.64}$$

или

$$\ln P = \beta T, \quad \text{или} \quad P = e^{\beta T} = X^T,$$

в соответствии с (4.54) и определением X. Наиболее вероятное распределение P в свою очередь совпадает с величиной $N(T)$, соответствующей всем возможным распределениям, и пропускная способность, в соответствии с (4.43), равна

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

Для длинных сообщений наиболее вероятное распределение настолько вероятнее любого другого, что все другие возможные распределения могут быть просто отброшены. Наилучшим является такое кодирование, при котором достигается это наиболее эффективное распределение символов, т. е.

$$\frac{\ln p_j}{t_j} = -\beta \quad \text{(постоянная)}, \tag{4.66}$$

в соответствии с (4.61). Наибольший вещественный корень уравнения (4.62) соответствует наибольшему значению X, получаемому в качестве решения уравнения (4.12), и дает на основании (4.64) самое вероятное распределение и тем самым наибольшее значение P.

Применяя наиболее эффективный код и распределение символов, соответствующее наиболее вероятному, мы получаем для информации на символ:

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

где $\bar{t}$ есть средняя длительность символа, когда вероятности различных символов выражены соотношением (4.61). Из (4.65) и (4.67) получаем для скорости передачи

$$R = \frac{C}{i} = \frac{1}{\bar{t}}. \tag{4.68}$$

Эта формула очень похожа на (4.50).

9. Проблема согласования

Условие (4.61) или (4.66) соответствует системе кодирования, наиболее эффективно выполняющей согласование с каналом. Оно играет в проблемах кодирования такую же важную роль, как условие согласования характеристических сопротивлений в проблемах теории цепей. В одной специальной задаче Шеннон применяет процедуру, приводящую к весьма сходному правилу. Мы доказали, что результат является совершенно общим и применим ко всем задачам кодирования. Специальные применения и примеры будут обсуждены в следующей главе.

Заметим в заключение, что совпадение среднего и наиболее вероятного распределения является хорошо известным результатом для весьма большого числа событий. Это правило играет важную роль в статистической термодинамике, где с его помощью объясняется совпадение различных определений энтропии, данных Гиббсом, Максвеллом, Больцманом и Планком. Этот вопрос очень ясно изложен у Лоренца[1]).

10. Проблемы статистики слов

Результаты пп. 3, 4, 7 и 8 были применены Мандельбротом к проблеме статистики слов, обсуждавшейся в предыдущей главе с эмпирической точки зрения. Какова роль слова в языке? Значение слова можно найти в словаре. Оно представляет собой вид сообщения. Это сообщение кодируется при помощи букв при письме. Оно кодируется при помощи фонем в речи. Каким образом слово кодируется в нашем мозгу? Это, конечно, открытый вопрос, но мы можем предположить некоторого рода кодирование, основанное на элементарных символах или сигналах, каждый из которых характеризуется определенной ценой. Это предположение позволяет нам воспользоваться предыдущим анализом, в котором нужно просто заменить *длительность — ценой.* Предположим далее, что код практически согласован с частотой слов, т. е. с вероятностями их применения в языке. Эти предположения были приняты Мандельбротом в его исследовании.

[1]) H. A. Lorentz, Leçons au College de France (Les theories statistiques en thermodynamique, Teubner, Leipzig, 1916).

Пусть S_1, $S_2, \ldots, S_j, \ldots, S_n$ — основные символы или сигналы кода, а t_1, $t_2, \ldots, t_j, \ldots, t_n$ — их относительные цены. Число $N(t)$ слов ценой t выражено формулой (4.15), а если кодирование устойчиво, мы можем воспользоваться упрощенным выражением (4.18). Число слов ценой меньше t или равной t выражается величиной $N_1(t)$, определяемой по формуле (4.22), которая для устойчивого кода принимает упрощенный вид

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

Расположим теперь все слова словаря в порядке возрастания цены t. N_1' будет означать порядковый номер слова в этом списке. В среднем соотношение между ценой t и номером N_1' выражается как

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

Рассмотрим теперь частоту (или вероятность) слова. Между ценой и вероятностью могло бы и не существовать никакой связи, однако мы предположим, что вероятность слова и его цена надлежащим образом согласованы. Это означает, что редкие слова имеют длинное дорогостоящее кодовое обозначение, а наиболее употребительные слова — короткое и экономное кодовое обозначение.

Проблема согласования была обсуждена в пп. 8 и 9; было выяснено, как следует выбрать наилучшим образом вероятности символов, имеющих различные цены t_j. Результат выражается формулой:

$$\ln p_j = -\beta_1 t_j, \quad \text{или} \quad p_j = e^{-\beta_1 t_j}. \tag{4.61}$$

Применим теперь сходные рассуждения не к символам или буквам, а к словам. Слово за номером N_1' имеет цену $t_{N_1'}$, причем обе величины связаны соотношениями (4.23) и (4.24). Наше новое условие хорошего согласования состоит в том, чтобы вероятность $p_{N_1'}$ слова была связана с его ценой $t_{N_1'}$ общим условием (4.61), содержащим новую произвольную постоянную β_1. Сопоставляя эти формулы, мы можем исключить цену $t_{N_1'}$ и связать порядковый номер N_1'

слова непосредственно с его вероятностью $p_{N'_1}$:

$$-\beta_1 t_{N'_1} = \ln p_{N'_1} = -\beta_1 t_0 - \frac{\beta_1}{\beta} \ln (N'_1 + B), \qquad (4.69)$$

откуда

$$p_{N'_1} = P (N'_1 + B)^{-\gamma}, \qquad (4.70)$$

где $P = e^{-\beta_1 t_0}$, $\gamma = \beta_1/\beta$. Понятие о кодировании слов в мозгу и о цене кодовых обозначений потребовалось для вывода соотношений (4.23) и (4.24), но мы постарались исключить неизвестные цены, предположив надлежащее согласование.

Формула (4.70) выражает закон, предложенный Мандельбротом. Он включает, как частный случай, закон, высказанный Ципфом:

$$p_{N'_1} = \frac{P}{N'_1}, \quad P \simeq \frac{1}{10}, \quad B = 0, \quad \gamma = 1, \ (\beta_1 = \beta). \qquad (4.71)$$

Еще одной важной величиной является полное число слов в словаре. Если мы применяем словарь из R слов, то

$$1 \leqslant N'_1 \leqslant R,$$

а сумма вероятностей должна равняться единице:

$$\sum_{N'_1=1}^{N'_1=R} p_{N'_1} = P \sum_{N'_1=1}^{N'_1=R} (N'_1 + B)^{-\gamma} = 1. \qquad (4.72)$$

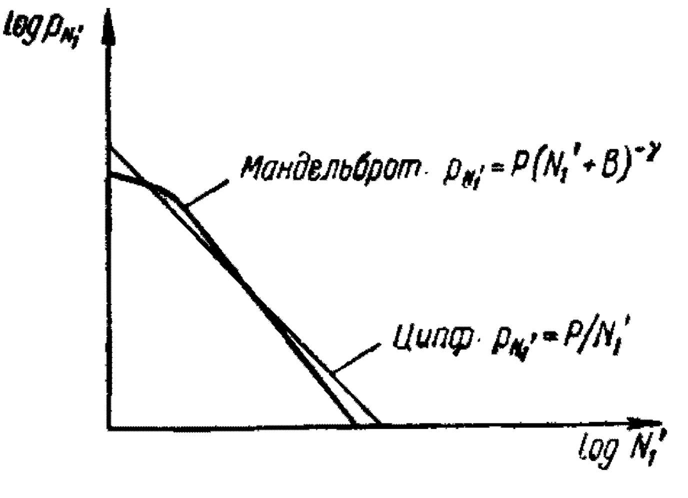

Рис. 4.1. Сравнение распределения частоты слова в функции порядкового номера по Мандельброту с данными Ципфа.

Это равенство определяет P. Если P задано (Ципф принимает $P=1/10$), то оно определяет общее число применяемых слов. Именно таким образом Шеннон нашел, следуя Ципфу, число слов в словаре, равное 8727. Для большинства языков $1 \leqslant \gamma \leqslant 1{,}2$, хотя γ может в исключительных случаях достигать 1,6, как, например, в детской речи.

Мандельброт в двух ранее указанных статьях привел много примеров, а также рассмотрел некоторые другие свойства языков. На рис. 4.1 представлены для сравнения законы,

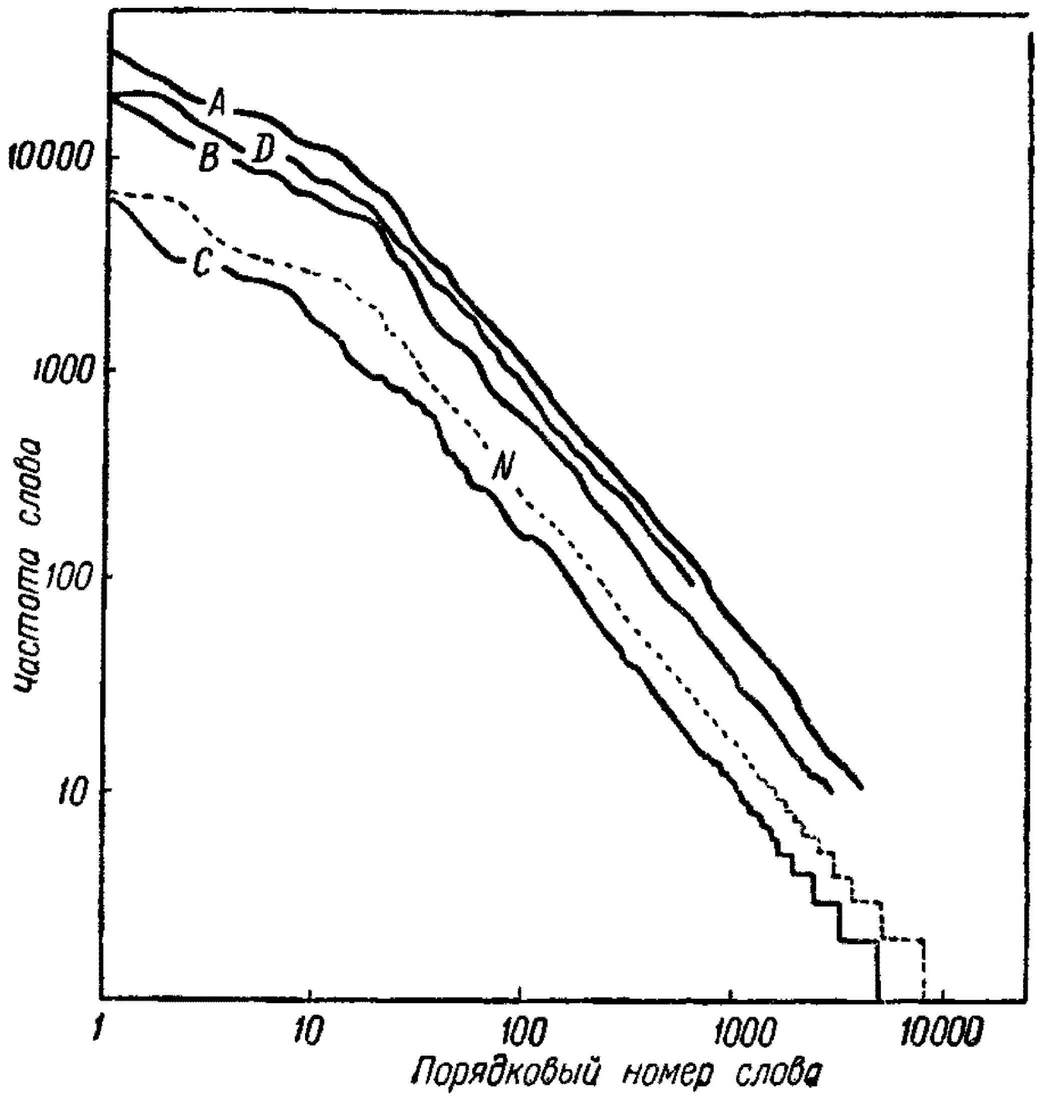

Рис. 4.2. Экспериментальные кривые частоты слова в зависимости от порядкового номера.

Кривые A, B, C и D — для норвежского языка, кривая N — для немецкого языка.

предложенные Ципфом и Мандельбротом. На рис. 4.2 даны некоторые экспериментальные кривые, построенные Ципфом.

11. Решение проблемы согласования

Мы подчеркивали в п. 9 необходимость кодирующего устройства, согласующего канал с языком, применяемым при передаче. Задача состоит теперь в выяснении того, как этот результат может быть получен практически.

В п. 4 мы начали с обсуждения свойств канала. Мы предположили, что в канале имеется n символов $S_1, \ldots, S_j, \ldots, S_n$

и что их относительные длительности равны $t_1, \ldots, t_j, \ldots, t_n$. Мы нашли, что средняя информация на символ максимальна (см. п. 8), когда вероятность p_j символа связана с его длительностью t_j соотношением

$$p_j = e^{-\beta t_j}, \tag{4.61}$$

где β — постоянная.

Рассмотрим теперь сообщение, написанное на языке, применяющем некоторое число ν знаков $\sum_\lambda$. Эти знаки могут быть буквами, группами букв, словами и т. д. В китайском языке, например, каждый знак соответствует слову. Для упрощения задачи предположим, что каждый знак $\sum_\lambda$ имеет определенную вероятность π_λ, не зависящую от предшествующих знаков. Другими словами, мы игнорируем корреляцию между знаками, во избежание усложнения задачи.

Рассмотрим сперва простейший возможный код, применяющий ν символов (или групп символов, если $n < \nu$) и присваивающий кратчайшие символы наиболее вероятным знакам. При этом обычно условие (4.61) не выполняется и согласование получается неудовлетворительное.

Мы получим значительно лучшее решение, кодируя группы знаков посредством групп символов, однако обсуждение такой системы требует предварительного исследования вероятностей групп знаков. Мы имеем:

$$\left.\begin{array}{l} \nu \text{ знаков: } \sum_1, \sum_2, \ldots, \sum_\lambda, \ldots, \sum_\nu; \\ \text{вероятности: } \pi_1, \pi_2, \ldots, \pi_\lambda, \ldots, \pi_\nu. \end{array}\right\} \tag{4.73}$$

Определим показатели θ_λ вероятностей π_λ:

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

Группа символов $\sum_1, \sum_2, \ldots, \sum_\lambda$ имеет вероятность

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

Мы имеем, таким образом, закон аддитивности показателей вероятностей θ. Найдем теперь число $\mathfrak{N}(\Theta)$ групп знаков, имеющих определенную вероятность $\pi = e^{-\Theta}$. Задача в точности совпадает с задачей, рассмотренной в п. 3, в которой нужно только заменить длительности t_j и T показателями вероятностей θ_λ и Θ. Для длинных групп знаков мы перепишем результаты, выражаемые формулами (4.18) и (4.23):

$$\mathfrak{N}(\Theta) = a_1 e^{\gamma\Theta}, \tag{4.76}$$

где $\mathfrak{N}(\Theta)$ — среднее число групп знаков при $\sum_\theta = \Theta$, γ — коэффициент, введенный вместо прежнего β. Аналогично

$$\mathfrak{N}_1(\Theta) = b_1 e^{\gamma\Theta} - b_0, \tag{4.77}$$

где $\mathfrak{N}_1(\Theta)$ — среднее число групп знаков при $\sum_\theta \leqslant \Theta$.

Сравним теперь статистические свойства групп знаков и групп символов. Будем кодировать $\mathfrak{N}_1(\Theta)$ групп знаков посредством $N_1(t)$ групп символов, присваивая, таким образом, кратчайшие кодовые обозначения наиболее часто встречающимся группам знаков. Приравняем $\mathfrak{N}_1(\Theta)$ и $N_1(t)$ и воспользуемся соотношениями (4.77) и (4.23):

$$\mathfrak{N}(\Theta) = b_1 e^{\gamma\Theta} - b_0 = B_1 e^{\beta t} - B_0 = N_1(t). \tag{4.78}$$

Эта упрощенная формула применима к длинным группам знаков и символов. При этом экспоненциальные члены в (4.78) будут много больше постоянных b_0 и B_0, которые можно опустить. Практическое условие запишется в виде

$$b_1 e^{\gamma\Theta} = B_1 e^{\beta t},$$

$$\gamma\Theta = \beta t + \ln\frac{B_1}{b_1}. \tag{4.79}$$

Так как Θ и t велики, последний член можно отбросить и записать:

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

Это соотношение выражает пропорциональность между показателем вероятности Θ и длительностью t.

При таком кодировании символ S_j длительностью t_j будет иметь среднюю вероятность

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

а это и есть экспоненциальный закон (4.61) оптимального кодирования. Условие (4.78) оставляет нам значительную свободу в выборе частного способа кодирования, однако мы смогли доказать возможность выполнения наших условий согласования языка с каналом. Хорошее согласование требует кодирования длинных групп знаков и символов, а это означает задержку процесса кодирования на передающей стороне и декодирования на приемной стороне.

Более сложные задачи, учитывающие корреляцию между последовательными символами (и избыточность), могут быть решены методом, сходным с тем, который был применен для решения нашей упрощенной задачи.

Вернемся к условию (4.78). Мы предоставили длинные группы символов редким группам знаков и короткие группы символов часто встречающимся группам знаков. При применении групп символов различной длины возникает трудность, состоящая в том, что для обозначения места стыка соседних групп необходим особый символ. Этого можно избежать, применяя группы символов одинаковой длины. На первый взгляд это решение кажется неэкономичным. Можно, однако, показать, что оно не хуже того, которое обсуждалось выше. Возьмем вместо (4.78)

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

Так как Θ и t велики, мы можем пренебречь постоянной b_0:

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

В этом выражении можно отбросить последний член, и мы придем снова к условию (4.80). Это иллюстрирует многообразие способов кодирования, дающих хорошее согласование.

Приложение

Однородное линейное уравнение в конечных разностях с постоянными коэффициентами может быть записано в виде[1])

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

где p — постоянные. Это уравнение линейно по отношению к неизвестной функции $u(x)$; x изменяется, получая единичные приращения. Уравнение имеет n частных решений $u_1(x), \ldots, u_n(x)$, являющихся независимыми, если между ними

[1]) L. M. Milne—Thomson, The Calculus of Finite Differences, Chapt. 12, MacMillan, London, 1951.

нет линейных соотношений. Эти решения можно найти, положив

$$u(x)=\rho^x. \tag{A.2}$$

Подстановка в (А.1) дает:

$$\rho^x f(\rho)=0,$$
$$f(\rho)=\rho^n+p_{n-1}\rho^{n-1}+\ldots+p_1\rho+p_0=0, \tag{A.3}$$

и мы получаем решение (А.1), если ρ есть решение характеристического уравнения (А.3). Пусть корни обозначены ρ_1, ρ_2, ..., ρ_n; все корни предполагаются различными. Можно показать, что эти решения независимы, и общее решение есть

$$u(x)=\sum_{k=1}^{n} A_k\rho_k^x, \tag{A.4}$$

где A_k — n произвольных коэффициентов. Для случая кратного корня ρ_i степени ν нужно взять в качестве независимых решений величины

$$\rho_i^x,\ x\rho_i^x,\ \ldots,\ x^{\nu-1}\rho_i^x.$$

Можно также рассмотреть уравнения с правой частью. Задача состоит в нахождении корней алгебраического уравнения (А.3). Напомним основные свойства алгебраического уравнения с вещественными коэффициентами

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

А. Если $f(\alpha)$ и $f(\beta)$ имеют противоположные знаки, то имеется по меньшей мере один корень между α и β. Число таких корней нечетно.

В. *Правило Декарта.* Число положительных вещественных корней уравнения (А.5) либо равно числу перемен знака $f(x)$, либо меньше этого числа на положительное четное целое. Число отрицательных вещественных корней аналогичным образом связано с $f(-x)$.

С. Полное число корней равно n. Если имеется комплексный корень $a+ib$, то комплексно-сопряженное число $a-ib$ также является корнем.

ГЛАВА 5

ПРОБЛЕМЫ КОДИРОВАНИЯ

1. Алфавитное кодирование, двоичная система

В предыдущих главах были обсуждены условия наиболее эффективного кодирования; мы пришли к заключению, что идеальная процедура кодирования выражается условием

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

где p_j — *априорная* вероятность символа S_j, имеющего длительность t_j, β — постоянная.

Условие (5.1) легко объяснить: наиболее вероятные символы имеют вероятность, близкую к единице, а следовательно, $(-\log p_j)$ является малой положительной величиной, и соответствующие символы, согласно (5.1), имеют относительно малую длительность. Другими словами, наиболее вероятные сигналы должны быть короткими, тогда как маловероятные сигналы могут быть длинными. Этот результат почти очевиден, и (5.1) выражает идеальное условие, к которому следует по возможности стремиться. Это условие может быть применено к алфавитным кодам, если использовать хорошо известные вероятности различных букв в языке.

В таблице 5.1, повторяющей таблицу 1.1, даны вероятности p_j и их отрицательные логарифмы для букв английского алфавита. Неважно, какие логарифмы мы применяем; изменение основания логарифма приводит лишь к умножению константы β на постоянный множитель. Рассматривая таблицу 5.1, можно отметить, что логарифмы вероятностей букв алфавита лежат между 0,7 и 3. Мы можем построить двоичный код с таким же интервалом длин сигналов. Это относится к каналу, способному передавать только один вид импульса: отсутствие импульса означает 0, наличие импульса означает 1.

Таблица 5.1

Вероятность p и значение $-\log_{10}p$ для букв английского языка

Символ	Вероятность	$-\log_{10}p$	Символ	Вероятность	$-\log_{10}p$
Промежуток между словами	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

Так как сложные сигналы имеют различные длины, мы должны предусмотреть способ распознавания конца сигнала. Для обозначения окончания группы импульсов, представляющих букву, можно воспользоваться символом 0, повторяя его два, три или четыре раза. Это означает, что в самих сигналах можно применять только одиночные нули. В таблице 5.2 дан пример такого кода.

Таблица 5.2

Возможный двоичный код для букв английского языка

Буква	Код	Буква	Код	Буква	Код
Промежуток между словами	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

Длина кодовой комбинации составляет от трех до восьми символов на букву, что является, в общем, подходящим соотношением. В длинном сообщении средняя длина на букву получается умножением длительности t_j на ее вероятность p_j и суммированием:

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

где значения p_j берутся из таблицы 5.1. Значение 4,65 является также средним числом символов на букву, и для применяемой нами двоичной системы имеем:

информация на букву = (информация на символ) $\times$ (символов на букву) $= 4{,}65 \times \log_2 2 = 4{,}65$.

Сравним этот результат с некоторыми другими кодами. Если взять символы одинаковой длины, то потребуется пять двоичных цифр для получения 32 комбинаций, из которых можно выбрать 27 для обозначения букв алфавита. Символы одинаковой длины, равной 5, несколько хуже, чем наша система переменных длин. Однако полученная нами величина все еще слишком велика, так как она превосходит теоретическое значение 4,03 дв. ед. на букву (см. (1.30)). Код, представленный в таблице 5.2, много лучше обычного кода Морзе, применяемого в телеграфии.

2. Алфавитное кодирование, троичная система

Многие технические устройства позволяют получать положительные и отрицательные импульсы. Так обстоит дело в магнитной записи и в телеграфных системах, применяющих ток изменяемого направления. Для этих систем может рассматриваться троичный код с элементами $-1, 0 + 1$. Пример (вероятно, не наилучший) такого кода дан в таблице 5.3. Мы можем использовать один из сигналов (например, -1) для обозначения промежутка между буквами. Это позволяет кодировать самые буквы двоичным кодом.

Этот код дает в среднем 3,3 символа на букву, как легко подсчитать, беря вероятности из таблицы 5.1. Число двоичных единиц на букву равно

$$3{,}3 \log_2 3 = 5{,}25, \tag{5.3}$$

Таблица 5.3

Возможный троичный код для букв английского языка

Буква	Код	Буква	Код	Буква	Код
Промежуток между словами	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

т. е. больше, чем для кодов предыдущего раздела. Трудность, встречающаяся при составлении всех кодов с переменной длиной, состоит в том, что мы вынуждены вводить специальные сигналы, указывающие окончание кодовой комбинации. Код таблицы 5.3 не является наилучшим, так как символ — 1 встречается реже, чем 0 и 1. Можно было бы улучшить код в этом отношении и несколько уменьшить число двоичных единиц на символ.

В конце этой главы мы обсудим возможность кодирования слов вместо кодирования букв и покажем, что такая система могла бы быть значительно более экономичной.

3. Алфавит и числа

Еще одна задача состоит в том, чтобы, кроме букв алфавита, кодировать еще и числа. Решение зависит от назначения кодирования. Можно предположить, что при машинописи буквы встречаются чаще, чем цифры. Для счетной машины справедливо обратное. Вот некоторые примеры.

А. Воспользуемся кодом *IBM* [1]) (шестизначный двоичный код), так что буквы и десять десятичных цифр имеют

[1]) *IBM* — International Business Machine Corporation. (*Прим. перев.*)

одинаковую длину. Отношение вероятностей в среднем равно:

$$\frac{\text{вероятность буквы}}{\text{вероятность десятичной цифры}} = \frac{27}{10} = 2{,}7, \tag{5.4}$$

что является подходящим соотношением для пишущей машинки. Однако код неэффективен, так как применяет только 37 из 64 возможных комбинаций; остальные 27 комбинаций остаются неиспользованными.

В. Возьмем семь двоичных цифр, дающих 128 комбинаций, и закодируем все числа от 0 до 99 плюс 27 букв. Одна комбинация останется неиспользованной. Это очень эффективный код, дающий в среднем 3,5 дв. ед. на число (оптимальное значение равно 3,32). Мы представляем себе каждое из этих 100 чисел в виде двух десятичных цифр (число 7, например, представляется в виде 07), так что в общем в числах встречается 200 десятичных цифр. Отсюда

$$\frac{\text{вероятность буквы}}{\text{вероятность десятичной цифры}} = \frac{27}{200} = 0{,}135. \tag{5.5}$$

Это соотношение является подходящим для счетной машины.

С. Возьмем 10 двоичных цифр, дающих 1024 комбинации, и закодируем все числа от 0 до 999 плюс 24 буквы. Две буквы останутся незакодированными. Мы можем использовать одно и то же кодовое обозначение для буквы *O* и для нуля, а также для единицы и *L*. Здесь мы получаем 3,33 дв. ед. на число, т. е. практически лучшее возможное значение. Имеется 1000 трехзначных десятичных чисел, и следовательно, в общем 3000 десятичных цифр, так что

$$\frac{\text{вероятность буквы}}{\text{вероятность десятичной цифры}} = \frac{26}{3000} = 0{,}00866. \tag{5.6}$$

Этот код был бы подходящим для вычислительной машины, где буквы встречаются лишь изредка.

В кодах В и С дополнения до 99 или 999 играют роль обычного дополнения 9 и легко могут быть получены применением кодов *с излишком 14* или *с излишком 12* соответственно[1]). Существенно, что прежде, чем выбирать код, нужно знать относительную вероятность цифр и букв.

---

[1]) Эта терминология относится к применяемым в вычислительных машинах двоично-кодированным системам счисления. См. по этому поводу: А. И. Китов, Электронные цифровые машины, Сов. радио, 1956, § 6. (*Прим. перев.*)

4. Двоичное кодирование слов

В главе 3 мы обсуждали исследование Шеннона избыточности английского языка. В одном из своих примеров Шеннон рассматривал частоту различных слов в языке и указал закон

$$p_m = \frac{1}{10\,m}, \tag{5.7}$$

устанавливающий вероятность появления слова, имеющего порядковый номер m в списке, составленном в порядке убывающей частоты. Этот закон нельзя применять до бесконечности, так как сумма будет расходиться, и Шеннон предположил, что применяется только 8727 слов. Это число выбрано так, чтобы сумма вероятностей равнялась единице[1]):

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

В этих предположениях была подсчитана информация, оказавшаяся равной

$$\begin{aligned} I &= 11{,}82 \text{ дв. ед. на слово,} \\ i &= 2{,}14 \text{ дв. ед. на букву,} \end{aligned} \tag{5.9}$$

так как среднее слово содержит 5,5 буквы при алфавите из 27 букв (26 букв плюс промежуток между словами). Цифра 2,14 дв. ед. на букву представляет значительную экономию по сравнению с обычным алфавитным кодированием, которое требует около 5 дв. ед. на букву.

Посмотрим, как можно использовать это преимущество. Рассмотрим сначала сигналы одинаковой длины. Применяя двоичные цифры 0 и 1 и беря постоянную длину, равную 13 цифрам, можем составить $2^{13} = 8192$ различных сигналов, которыми можно закодировать 8192 слова, образующих словарь, достаточный для многих применений. Этот очень простой код дает 13 дв. ед. на слово или 2,36 дв. ед. на

---

[1]) Для подсчета конечного числа слов Шеннон пользовался экспериментально найденными вероятностями наиболее часто встречающихся слов и обращался к формуле (5.7) лишь при больших m ($m > 100$). Если бы формула (5.7) применялась с самого начала, то конечное число слов было бы гораздо больше 8727 и составляло бы около 12 000.

букву, что уже очень близко к теоретическому пределу (см. (5.9)).

Можно получить лучший результат, применяя метод, обсужденный в п. 1 этой главы и иллюстрированный таблицей 5.2. Исключим кратные нули в главном сигнале и используем их только в окончании. Каждый сигнал начинается единицей и кончается группой 00, 000, 0000, . . . Несколько таких комбинаций дано в таблице 5.4.

Таблица 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
				11111.00

В четвертом столбце комбинация 1001 опущена, так как она содержит два нуля подряд. По этой же причине опущены комбинации 10001.00; 10010.00; 10011.00 и 11001.00 в пятом столбце.

В последующем рассуждении мы опустим последние два нуля, появляющиеся в конце каждого сигнала. Предположим, что мы нашли все сигналы некоторого столбца (например, семь сигналов четвертого столбца с четырьмя цифрами в главной части каждого сигнала). Введем следующие обозначения:

n — номер столбца,

$N_{n,1}$ — число сигналов, оканчивающихся единицей,

$N_{n,0}$ — число сигналов, оканчивающихся одним нулем,

$N_{n,00}$ — число сигналов, оканчивающихся группой нулей.

Например, для четвертого столбца (не считая последних двух нулей после точки)

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

Подсчитаем теперь числа $N_{n+1,1}$, $N_{n+1,0}$ и $N_{n+1,00}$ для сигналов с $n+1$ цифрами в главной части. Сигналы $N_{n,0}$ и $N_{n,1}$ могут быть дополнены либо нулем, либо единицей. Сигнал $N_{n,00}$ может быть дополнен только нулем, и мы получаем следующие соотношения:

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

Применяя эти соотношения шаг за шагом, получаем результаты, сведенные в таблице 5.5.

Таблица 5.5

Двоичное кодирование слов

	Номер столбца, n											
	1	2	3	4	5	6	7	8	9	10	11	12
N_{n1}	1	1	2	3	5	8	13	21	34	55	89	144
N_{n0}	0	1	1	2	3	5	8	13	21	34	55	89
N_{n00}	0	0	1	2	4	7	12	20	33	54	88	143
ΣN	1	2	4	7	12	20	33	54	88	143	232	376
Общее число кодированных слов	1	3	7	14	26	46	79	133	221	364	596	972

	Номер столбца						
	13	14	15	16	17	18	19
N_{n1}	233	377	610	987	1597	2584	4181
N_{n0}	144	233	377	610	987	1597	2584
N_{n00}	232	376	609	986	1596	2583	4180
ΣN	609	986	1596	2583	4180	6764	10 945
Общее число кодированных слов	1581	2567	4163	6746	10 926	17 690	28 635

Если взять от 1 до 17 столбцов для главной части, число двоичных цифр будет изменяться от 3 до 19 (включая два нуля, которые должны быть добавлены в конце каждого сигнала), что даст словарь, содержащий около 11 000 слов.

Подсчет среднего числа двоичных единиц на слово может быть выполнен без особых трудностей, если принять результаты Шеннона для вероятности последовательных слов. Такой подсчет дает около 12 дв. ед. на слово, что очень близко к теоретическому пределу 11,82. Такого рода кодирование является, по-видимому, наиболее экономичным из числа практически выполнимых.

Для практического применения код должен быть дополнен буквенным кодом. Буквы во всяком случае будут нужны для всех слов, не вошедших в словарь, в особенности для имен, географических названий и т. д. Если такие слова встречаются редко, можно оставить последние 27 символов кода для алфавита. Если же специальные слова встречаются очень часто, кодирование слов становится невыгодным.

Воспользуемся соотношениями (5.10) для вычисления вероятностей p_{n1}, p_{n0} и p_{n00} сигналов, главная часть которых имеет окончание 1, 0 или 00. Имеем:

$$\begin{gathered} N_{n1} = N_n p_{n1}; \quad N_{n0} = N_n p_{n0}; \quad N_{n00} = N_n p_{n00}; \\ N_n = N_{n1} + N_{n0} + N_{n00}. \end{gathered} \tag{5.11}$$

Равенства (5.10) и (5.11), взятые совместно, определяют последовательные значения вероятностей. Когда n велико (например, $n > 10$), эти вероятности делаются независимыми от n, и легко получить

$$p_1 \simeq p_{00} = \frac{1}{2}(3 - \sqrt{5}) = 0{,}382, \quad p_0 = 0{,}236, \tag{5.12}$$

и равенства (5.10) принимают вид:

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

откуда

$$\begin{gathered} N_n = D_0 (1{,}618)^n, \\ N_1 + N_2 + \ldots + N_n = D_1 (1{,}618)^n, \end{gathered} \tag{5.14}$$

где $D_0 = 1{,}163$, $D_1 \approx 3$. Сравнение с точными цифрами таблицы 5.5 показывает, что эти соотношения дают хорошее приближение при $n > 10$. При помощи этих формул можно провести более детальное исследование кода, предложенного в данном разделе.

5. Алфавитное кодирование слов

Другой способ кодирования слов основан на применении букв. Если мы кодируем слова двумя буквами: *aa*, *ab*, *ac*, ..., *zz* то всего может быть закодировано $(26)^2 = 676$ слов, что дает $2 \cdot 4{,}7 = 9{,}4$ дв. ед. на слово, так как алфавит из 26 букв дает 4,7 дв. ед. на букву. Это соответствует словарю Basic English [1]). Для большей гибкости языка мы можем применить трехбуквенные кодовые слова, что дает $(26)^3 = 17\,576$ слов, т. е. очень большой словарь без избыточности, дающий $3 \cdot 4{,}7 = 14{,}1$ дв. ед. на слово. Оба примера показывают, какова избыточность при обычном начертании слов, требующем в среднем по 5,5 буквы на слово вместо двух или трех букв, как в этих новых кодах. Конечно, кодированные слова обычно невозможно произнести.

Менее эффективный код, содержащий примерно 10 000 слов, был составлен Луном (P. Luhn) из Корпорации IBM. Этот код применяет произносимые трех- и четырехбуквенные слова, оканчивающиеся таким образом, что разделительного знака между словами не требуется. Получается практичный код, дающий примерно 3,8 буквы на слово, вместо 5,5, требуемых в английском языке.

6. Кодирование, основанное на группах букв и на корреляции

В конце главы 3 упоминалась общая процедура кодирования, предложенная Шенноном. Метод основан на применении запоминающегося устройства и на использовании корреляционных связей при кодировании следующих друг за другом букв. Эта процедура является прямым приложением математического анализа главы 3, но он может применяться только к языку, в котором вероятности диграмм и триграмм (т. е. групп из двух, трех и более букв) известны из предшествующего статистического исследования языка.

Оливер [2]) недавно рассмотрел практические методы этого рода. Он полагал, что при передаче применяются 27 импуль-

---

[1]) Basic English (бэйзик инглиш) — упрощенный английский язык, обходящийся «прожиточным минимумом» основных слов. Существует специальная литература на этом языке, в том числе переработки известных литературных произведений. (*Прим. перев.*)

[2]) B. M. Oliver, Efficient coding, BSTJ 31, 724 (1952).

сов с интенсивностями 0, 1, 2, ... , 27, и исследовал возможность сведения к минимуму средней мощности сигнала. Первая возможность состоит в расположении квантованных импульсов в алфавитном порядке:

$$\left.\begin{array}{llcccccc} \text{символ:} & \text{промежуток} & A & B & C & D & E & \ldots \\ \text{высота импульса:} & 0 & 1 & 2 & 3 & 4 & 5 & \ldots \end{array}\right\} \quad (5.15)$$

Очевидно, что это — неудачное решение. Лучшее кодирование основано на вероятностях таблицы 5.1:

$$\left.\begin{array}{llcccccc} \text{символ:} & \text{промежуток} & E & T & O & A & N & \ldots \\ \text{высота импульса:} & 0 & 1 & 2 & 3 & 4 & 5 & \ldots \end{array}\right\} \quad (5.16)$$

Это, конечно, более эффективный код, так как наиболее вероятным буквам соответствуют импульсы малой интенсивности.

Предложенное Оливером практическое устройство показано на рис. 5.1. Импульсы отклоняют луч электронно-лучевой

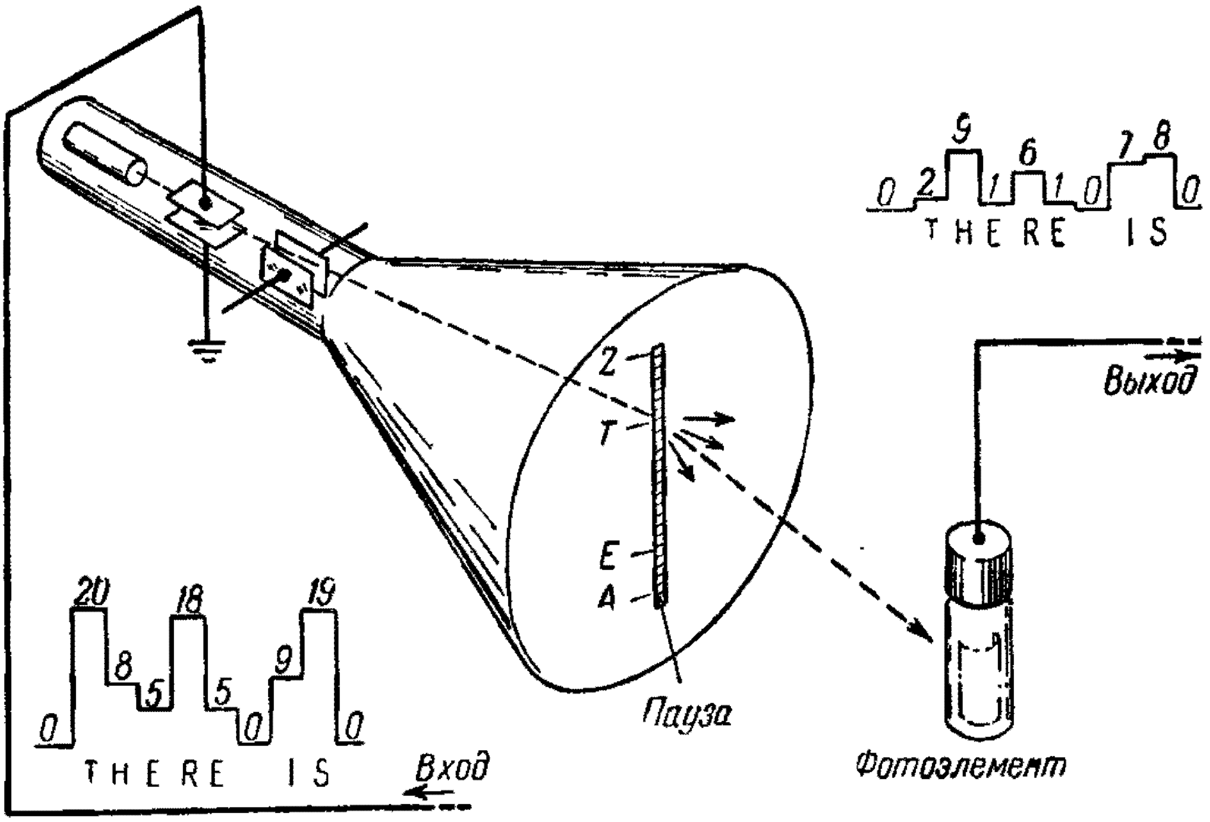

Рис. 5.1. *Монограммер* Оливера для кодирования одиночных букв.

трубки; на экране расположена маска переменной прозрачности (изменяющейся согласно (5.16)), и в результате на выходе фотоэлемента получаются импульсы, соответствующие (5.16). Система названа ***монограммером.***

Устройство *диграммера* показано на рис. 5.2. Здесь использованы обе пары отклоняющих пластин электронно-лучевой трубки. На одну пару пластин воздействует предшествующая буква, а на другую — приходящая в данный

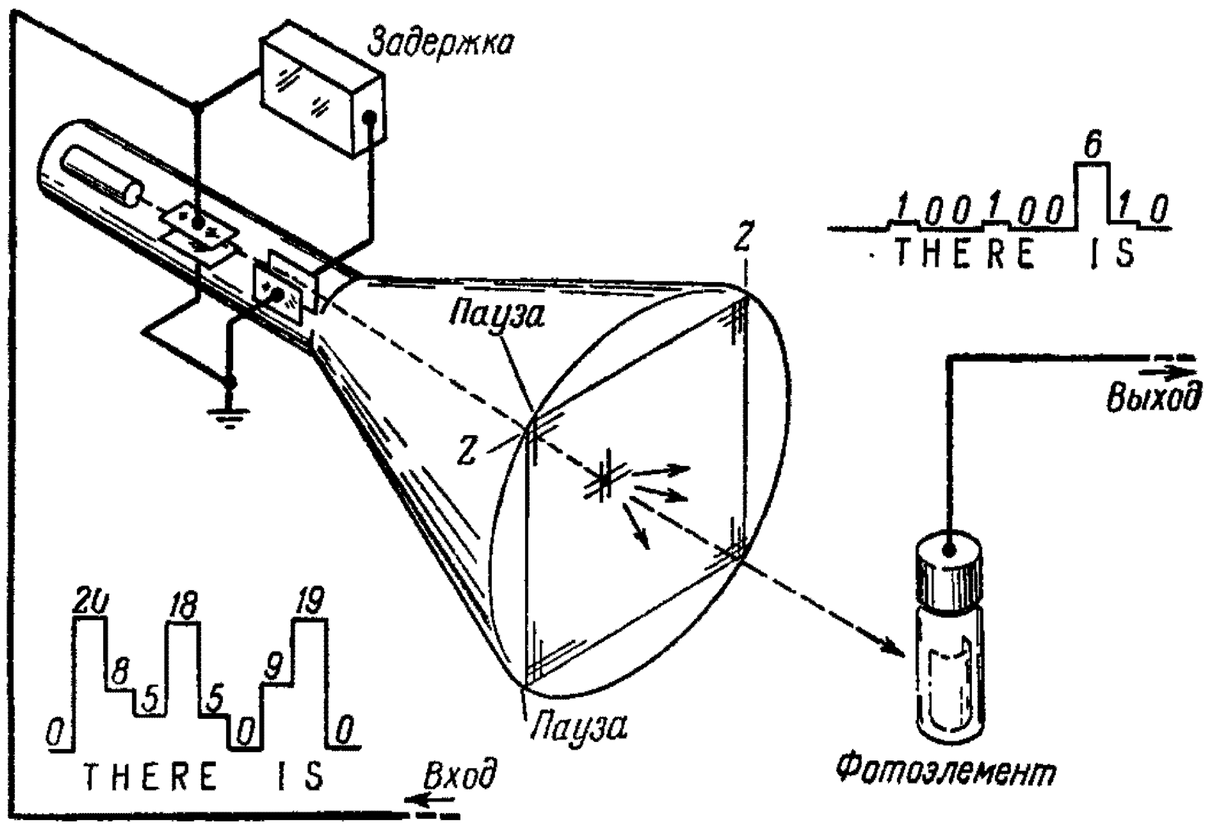

Рис. 5.2. *Диграммер* Оливера для кодирования пар букв.

момент буква. Каждой группе букв (i, j) соответствует точка на экране, и прозрачность маски, наложенной на экран, соответствует вероятности $p\ (i,\ j)$. Таким путем можно получить дополнительную экономию средней мощности, требуемой для передачи. Обсуждены также устройства, основанные на триграммах и тетраграммах, но они едва ли являются практичными.

Другой метод кодирования предложен рядом авторов [1]). Его можно пояснить применительно к телефонии: нужно передать непрерывную кривую и мы располагаем статистическими данными относительно типа применяемых кривых. Используя эти данные и несколько уже переданных предшествующих значений, мы можем предсказать наиболее вероятное следующее значение и передать только поправку, т. е. уклонение от предсказанного значения. Сила этого

[1]) P. Elias, Harvard University thesis PIRE **39**, 839 (1951); E. R. Kretzmer, BSTJ **31**, 751 (1952); C. W. Harrison, ibid., p. 764.

метода состоит в том, что он может быть применен в двух измерениях (телевидение) при условии, что имеются соответствующие статистические данные. Применительно к телевидению рассмотрим систему, применяющую $64 = 2^6$ уровней квантования. Если все уровни равновероятны, нам потребуется 6 дв. ед. на каждый элемент изображения. Однако различные уровни имеют различную вероятность, и это простое обстоятельство сразу уменьшает информацию до 5 дв. ед. на элемент. Используя корреляцию между соседними элементами, можно снизить эту величину до $2{,}4 \div 2{,}6$ дв. ед. на элемент. Это представляет потенциальную возможность сокращения пропускной способности телевизионного канала вдвое. Как и в других аналогичных проблемах, нужно, однако, заметить, что передача с малой избыточностью в значительно большей мере страдает от помех, так как устранение избыточности уменьшает также возможность предсказания и исправления ошибок при приеме.

ГЛАВА 6

КОДЫ, ОБНАРУЖИВАЮЩИЕ И ИСПРАВЛЯЮЩИЕ ОШИБКИ

1. Коды, обнаруживающие ошибки

В предыдущих главах обсуждались методы устранения избыточности и построения кодов, требующих наименьшего числа двоичных единиц на букву. Те же методы могут применяться не только к буквам, но и к любого рода символам в более общих задачах теории информации.

Рассмотрим теперь вопрос о том, не является ли избыточность благоприятной в некоторых практических случаях. Когда мы запасаем информацию в устройстве с высоким уровнем шума, то ложные шумовые сигналы могут вызвать ошибки, и некоторая избыточность была бы очень полезна для обнаружения и даже для исправления ошибок. Именно так обстоит дело, когда мы читаем книгу с опечатками или когда мы получаем телеграмму, в которой некоторые буквы переданы неправильно. Вместо того чтобы использовать для обнаружения ошибок высокую избыточность языка, мы можем построить специальный проверочный механизм для применения совместно с кодом малой избыточности. Таким образом мы получим более эффективную систему, способную (в отличие от системы, основанной на избыточности языка) обнаруживать ошибки в начертании имен, географических названий и т. п.

При применении двоичного кода простейшим методом проверки является проверка на четность. При известных условиях мы можем быть почти уверены, что в группе из некоторого числа (например, $n=6$) двоичных знаков не может быть больше одной ошибки. Для осуществления проверки на четность мы объединяем наши сигналы в группы по шесть знаков и добавляем к каждой группе седьмой знак, выбирая его так, чтобы общее число единиц было четным. Например:

сигнал: 101 100, проверочный знак 1,

сигнал: 100 010, проверочный знак 0.

Это позволяет незамедлительно обнаружить на приемной стороне одиночную ошибку; однако узнать, в каком знаке произошла ошибка, нельзя. Двойная ошибка не будет обнаружена.

2. Коды, исправляющие одиночную ошибку

Мы пойдем далее и исследуем возможность построения методов проверки, позволяющих определить точное местоположение ошибки и, следовательно, исправить ее.

В интересной работе Хэмминга[1]) обсуждаются коды, дающие возможность определить положение ошибки и исправить ее. Это делается путем нескольких проверок на четность, и Хэмминг дает практические примеры применения своего метода.

Рассмотрим совокупность n двоичных цифр и предположим, что среди n позиций никогда не встретится более чем одна ошибка. Из числа n позиций мы выбираем m позиций для передачи информации, а остальные k позиций используем для проверки[2]). Эти k двоичных цифр представляют 2^k различных двоичных чисел, которые должны указывать:

1) либо, что ошибки нет,

2) либо, если ошибка есть, то в какой позиции.

---

[1]) R. W. Hamming, BSTJ **29**, 147 (1950). (Русский перевод в сборнике «Коды с обнаружением и исправлением ошибок», ИЛ, Москва, 1956.)

[2]) Излагаемая здесь концепция, согласно которой знаки делятся на информационные и проверочные, сильно сужает круг понятий, относящихся к помехоустойчивым кодам. Более широкая точка зрения состоит в том, что вся совокупность возможных для данного кода комбинаций делится на разрешенные и запрещенные. Ошибка превращает разрешенную комбинацию в запрещенную, вследствие чего ошибка обнаруживается. Для исправления ошибки используются различия между комбинациями: правильной комбинацией является та, которая наименее отличается от принятой ошибочной. В геометрических терминах различие определяется *расстоянием* между кодовыми комбинациями. Это расстояние может измеряться как в евклидовой метрике, так и в метрике Хэмминга, в которой единичное расстояние

Всего получается $n+1$ указаний, откуда следует условие:

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

Проверка наиболее эффективна в случае знака равенства в (6.1).

В таблице 6.1 даны наибольшие числа символов, которые могут быть проверены при различных значениях k.

Таблица 6.1

Наибольшее число символов, которые могут быть проверены применением от одного до пяти проверочных символов

Число проверочных символов k	Всего символов n	Информационных символов $m = k - n$
1	1	0
2	3	1
3	7	4
4	15	11
5	31	26

Второй случай в таблице 6.1 соответствует предположению, что в последовательности из трех двоичных цифр никогда не может быть больше, чем одна ошибка. Одна из этих цифр служит для передачи информации, а две другие — для проверки, согласно таблице 6.2.

Проверочные цифры B и C выбираются так, чтобы $(A+B)$ и $(A+C)$ были четными. Приемник вычисляет обе суммы, и если одна из них оказывается нечетной, проверочный сигнал в ней содержит ошибку. Если, например, $(A+B)$ нечетна, тогда как $(A+C)$ четна, то сигнал B неверен. Если имеется ошибка в сигнале A, то обе суммы будут нечетными. Отсюда ясно, какой сигнал должен быть исправлен.

Третий случай — код «четыре из семи»: четыре двоичных цифры для информации, три для проверки. Ошибки могут

соответствует различию в одном двоичном знаке, а расстояние k единиц—различию в k двоичных знаках. Так, например, описанное в п. 1 добавление одного проверочного знака (0 или 1) увеличивает на единицу расстояние между разрешенными комбинациями (обладающими четным числом единиц). (*Прим. перев.*)

быть обнаружены и локализованы в предположении, что в группе из семи двоичных цифр никогда не встречается более

Таблица 6.2

Схема кода с одной цифрой для информации и двумя для проверки

Информация	Проверка	
A	B	C
X	X	
X		X

Примечание. Проверки должны быть выбраны так, чтобы сумма символов «X» по каждой строке давала четное число.

одной ошибки. Проверка может осуществляться, как показано в таблице 6.3.

Таблица 6.3

Код с четырьмя двоичными цифрами для информации и тремя для проверки

Информация				Проверка		
A	B	C	D	E	F	G
X	X	X		X		
X	X		X		X	
X		X	X			X

(То же примечание, что и к табл. 6.2.)

Сигналы A, B, C и D могут быть либо 0, либо 1. Проверочные сигналы нужно выбрать так, чтобы были четными суммы

$$\left.\begin{aligned} &A+B+C+E, \\ &A+B+D+F, \\ &A+C+D+G. \end{aligned}\right\} \qquad (6.2)$$

Приемник сопоставляет эти три суммы. Если все они четны, то ошибки нет (в силу нашего предположения о том, что в семи знаках не может быть больше одной ошибки). Если одна из сумм нечетна, то ошибка имеется в проверочном знаке этой суммы. Так, если первая сумма нечетна, то сигнал E неверен. Если две суммы нечетны, то ошибка имеется в том сигнале, который является для обеих сумм общим, но не входит в третью. Так, если первые две суммы нечетны, то ошибка должна быть в B. Если все три суммы нечетны, то ошибка в A.

Примененный здесь метод несколько отличается от выбранного Хэммингом, но результаты в точности совпадают.

Рассмотрим четвертый случай таблицы 6.1 в предположении, что не может быть более одной ошибки в 15 двоичных цифрах. Это означает, что число информационных позиций $m=11$, а число проверочных позиций $k=4$. Система проверки показана в таблице 6.4.

Таблица 6.4

Код с одиннадцатью двоичными цифрами для информации и четырьмя для проверки

	Информация											Проверка			
Позиции	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				X

(То же примечание, что и к предыдущим таблицам.)

Это значит, что проверочные сигналы нужно выбрать так, чтобы четными были суммы

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

Невыполнение этого условия в одной сумме указывает, что проверочный сигнал неверен. Ошибки в информационных сигналах дают нечетность в 2, 3 или 4 контрольных суммах.

Общий метод, примененный в этих примерах, теперь ясен. Если мы имеем k проверочных позиций и некоторое число m информационных сигналов $A, B, C, D, \ldots$, мы включаем A во все k проверочных сумм. Каждый из следующих k сигналов встречается в $(k-1)$ из k проверочных сумм. Затем мы используем $(k-2)$ из k проверочных сумм для следующих $\frac{1}{2}k(k-1)$ сигналов и т. д. Каждая проверочная сумма содержит только по одному проверочному сигналу, и каждый данный проверочный сигнал встречается лишь однажды. Вообще k проверочных позиций дают:

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

возможностей проверки, и это является наибольшим числом информационных сигналов, которые могут быть проверены. Одиночные ошибки могут появиться в k проверочных позициях; отсутствие ошибки дает еще одну возможность.

Таким образом,

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

в соответствии с (6.1), так что метод является общим.

Если имеется 6 информационных позиций (как в коде IBM), то требуется, по-прежнему, 4 проверочные позиции (см. таблицу 6.1), и мы можем произвольно выбрать 6 из 11 позиций от A до K таблицы 6.4.

3. Коды, исправляющие одиночную ошибку и обнаруживающие двойную

Предыдущие схемы могут быть распространены на случай обнаружения двойных ошибок. Начнем с кода, исправляющего одиночную ошибку. К этому коду мы можем добавить еще одну позицию проверки на четность. Могут быть следующие случаи:

А. Ошибок нет, все проверки на четность удовлетворяются.

В. Одиночная ошибка. Одна из первоначальных проверок дает нечет и указывает местоположение ошибки. Последняя проверка также дает нечет.

С. Двойная ошибка. Последняя проверка дает чет, а некоторые из предыдущих — нечет, что указывает на имеющееся где-то искажение. Метод является общим и может быть применен к любому примеру, обсужденному в предыдущих разделах. Хэмминг указал геометрическую интерпретацию, применяющую представление в многомерном пространстве. Это представление дает элегантное доказательство общности описанных методов.

Новые исследования по теории исправляющих кодов содержатся в работе Джилберта[1]). Обширный материал по кодам, исправляющим многократные ошибки, можно найти в Trans. IRE[2])

4. Эффективность исправляющих кодов

Исправляющие коды, описанные в предыдущих разделах, могут быть использованы следующим образом. Возьмем в качестве примера код «4 из 7» (см. таблицу 6.3). Начнем с некоторого сообщения, кодированного двоичными цифрами, например:

$$0000100010100100 \tag{6.5}$$

и разделим его на группы по четыре цифры в каждой:

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

Дополним теперь каждую из этих четырехзначных групп еще тремя знаками, соответствующими проверкам на четность, согласно (6.2):

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

Это сообщение передается, приемник применяет проверки на четность для исправления ошибок, затем устраняет излишние последние три знака в каждой семизначной группе и восстанавливает исходное сообщение в виде (6.6) и (6.5).

Этот метод годится в том случае, если на n символов никогда не приходится больше одной ошибки (в нашем примере $n=7$). Это условие никогда в точности не выполняется. Обычно имеется совокупность случайных ошибок со средней вероятностью p. Мы покажем, что исправляющий код может

---

[1]) E. N. Gilbert, BSTJ 31, 504 (1952).

[2]) Trans. IRE (PGIT) 1—4, см. в особенности в т. 4 статьи Golay, Elias, Reed, Silverman, Balser и др. См. также Conv. Rec. IRE National Convention, 1954.

значительно уменьшить число ошибок и дать малую вероятность P неисправленных ошибок.

Пусть p означает вероятность ошибки в некотором определенном знаке. Это означает переход 0 в 1 или 1 в 0. Возьмем код (k, n, m), т. е. код с k проверочными знаками и m информационными позициями, а всего с n двоичными позициями, согласно таблице 6.1. В данной последовательности из n двоичных цифр имеем следующие вероятности:

$$\left.\begin{array}{ll}\pi_0 = (1-p)^n & \text{— вероятность отсутствия ошибки;} \\ \pi_1 = np(1-p)^{n-1} & \text{— вероятность одной ошибки;} \\ \pi_2 = \dfrac{n(n-1)}{2} p^2 (1-p)^{n-2} & \text{— вероятность двух ошибок в двух разных позициях;} \\ \pi_l = \dfrac{n!}{l!\,(n-l)!} p^l (1-p)^{n-l} & \text{— вероятность } l \text{ ошибок в } l \text{ разных позициях.}\end{array}\right\} \quad (6.8)$$

Рассмотрим случай одной ошибки, чтобы показать, как получаются величины π_l. В этом случае имеем $(n-1)$ правильных знаков с вероятностью $(1-p)^{n-1}$ и один неверный знак (с вероятностью p), который может находиться в любой из n позиций. Это дает приведенное выше значение π_1.

Дальнейшие вероятности вычисляются в предположении, что ошибки встречаются всегда в разных позициях. Это предположение оправдывается тем, что две ошибки, наложенные друг на друга, взаимно компенсируются и не оставляют следа: если мы имеем 1, которая в результате ошибки перешла в 0, то вторая ошибка снова дает нам 1. Таким образом, наложенные ошибки уменьшают число видимых ошибок.

Величины π_l действительно являются вероятностями, так как их сумма равна единице:

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

(по формуле бинома).

Если ошибок нет или если имеется одна ошибка на всю последовательность, то сообщение будет принято правильно, но если будет иметься $2, 3, \ldots, l, \ldots$ ошибок то окончательная последовательность будет неверна. Вероятность неправильной последовательности равна поэтому

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

Пусть p мало; пренебрежем членами с p^3, p^4, ... Преобладающий член в выражении для P зависит от π_2:

$$P \simeq \pi_2 \simeq \frac{1}{2} n(n-1) p^2, \tag{6.11}$$

и эту величину нужно сравнить с ожидаемым числом ошибок без применения проверок. Это означало бы применение только m информационных позиций с вероятностью ошибки

$$1 - (1-p)^m = mp - \frac{1}{2} m(m-1) p^2 \ldots \simeq mp.$$

Отношение mp/P может служить мерой качества кода. Можно также ввести $p' = \frac{P}{m}$, что означает новую вероятность ошибки в одном исходном знаке после исправления. В таблице 6.5 даны значения mp, P, mp/P и p' для нескольких кодов.

Таблица 6.5

Различные исправляющие коды и их характеристики

Код			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	$26p$	$465p^2$	$\frac{26}{465p}$	$\frac{465}{26}p^2$

Положим, например, что в обычном коде получается в среднем одна ошибка на 100 знаков, т. е. $p = 10^{-2}$. Код «4 из 7» дает $P = 21 \cdot 10^{-4} = \approx \frac{1}{475}$, т. е. одну ошибку на последовательность из 475 знаков (после исправления). Мера качества составляет $4\frac{10^{2}}{21} = 19$. Вероятность p' ошибки в первоначальном знаке равна $p' = 5{,}25 \cdot 10^{-4}$.

5. Пропускная способность двоичного канала с шумом

Сравним полученные результаты с теоретической формулой Шеннона [1]). Рассматривая общую проблему дискретного канала с шумом, он получил для случая двоичного канала

$$C = 1 + p\log_2 p + (1-p)\log_2 (1-p), \qquad (6.12)$$

C — пропускная способность в двоичных единицах на символ, p — вероятность ошибки в одном двоичном знаке. Эта формула может быть проверена элементарным путем. Мы передаем сообщение из N двоичных знаков, из которых будет Np знаков, переданных неверно. Для исправления требуется только знать положение неверных знаков. Если неверный знак был принят как 1, то его нужно просто заменить на 0, и наоборот. Информация о положении ошибок передается путем добавления специальных сигналов к двоичному коду (0 для правильного, 1 для неверного) с вероятностью p для 1 и $(1-p)$ для 0. Для этих добавочных сигналов требуется канал с пропускной способностью

$$-p\log_2 p - (1-p)\log_2 (1-p)$$

двоичных единиц на символ. Эта дополнительная информация, будучи передаваема по тому же каналу, уменьшает его пропускную способность. Исходный канал без шума имеет $C = 1$, а канал с шумом имеет, следовательно, пропускную способность, выражаемую формулой (6.12). Теоретически возможно найти метод кодирования, реализующий такую пропускную способность.

---

[1]) C. E. Shannon, W. Weaver, The Mathematical Theory of Communication, p. 38, U. of Illinois Press, Urbana, 1949. (Русский перевод в сборнике «Теория передачи электрических сигналов при наличии помех», ИЛ, Москва, 1953.)

В нашем предыдущем примере при $p = 10^{-2}$ формула (6.12) дает пропускную способность около 0,92, что означает, что только один дополнительный двоичный знак на каждые девять исходных знаков был бы достаточен для исправлений. Наш код «4 из 7» не дает полного исправления и требует 7 знаков на 4 исходных знака, что дает увеличение на 3/4 вместо 1/9.

Мы очень далеки от теоретического предела. Джилберт приходит к такому же заключению, не предлагая никакого практического решения. Его статья содержит интересное обсуждение свойств большого числа различных кодов.

Проблема исправляющих кодов имеет большое значение для связи, и за последнее время этой проблеме посвящено много работ. Метод Хэмминга (см. раздел 2) создан для автоматического исправления неверного символа. Более экономичный метод был предложен Вагнером (MIT). Его принцип состоит в нахождении символа, неверного с наибольшей вероятностью, и в исправлении этого символа. Этот метод обсужден Силвермэном и Болсером[1]) и представляется многообещающим. Для обнаружения символа, который имеет наибольшую вероятность быть неверным, авторы используют некоторые априорные данные о взаимосвязях передаваемого сообщения. Наличие взаимосвязей означает избыточность и указывает, что число применяемых для передачи символов слишком велико. Вычисления числа необходимых символов, выполненные Силвермэном и Болсером для их метода и метода Хэмминга, по-видимому, не учитывают этого обстоятельства. Однако основная идея представляет большой интерес, и к ожидаемой дальнейшей дискуссии о различных методах исправления следует отнестись со вниманием.

---

[1]) Silverman, Balser, PIRE **42**, 1428 (1954); Trans. IRE (PGIT) **4**, 50 (1954). (Русский перевод в сборнике «Коды с обнаружением и исправлением ошибок», ИЛ, 1956.)

ГЛАВА 7

ПРИЛОЖЕНИЯ К НЕКОТОРЫМ СПЕЦИАЛЬНЫМ ПРОБЛЕМАМ

1. Задача о расположении с применением смешанной ячейки

Если мы желаем классифицировать данные (буквы, ошибки, растения и т. д.), расположив их в определенном порядке по набору ячеек, то возникает вопрос об оптимальном размере ячейки. Если ячейка мала по сравнению с размерами элементов, то возникает неопределенность; если ячейка слишком велика, то расположение будет неэффективным. Рассмотрим эту задачу и исследуем роль различных возможных правил расположения. Мы можем применить *смешанную* ячейку (*miscellaneous* cell) для элементов, принадлежность которых к какой-либо определенной ячейке неясна; мы можем также применить систему *перекрестных ссылок* (*cross-references*). Интересно сравнить эффективность обоих методов[1]). Наше рассуждение основывается на общей формуле (1.13):

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

выражающей информацию на символ для случая, когда делается выбор из некоторого числа событий $a, b, \ldots, j, \ldots$, имеющих априорные вероятности $p_a, p_b, \ldots, p_j, \ldots$ При применении двоичной системы $K = \log_2 e$,

$$I = -\sum_j p_j \log_2 p_j, \tag{7.1a}$$

и информация измеряется в двоичных единицах.

---

[1]) Эта проблема впервые поставлена и обсуждена Мак-Дональдом: D. K. C. MacDonald, J. Appl. Phys. **23**, 529 (1952).

В качестве упрощенной схемы рассмотрим расположение отрезка длиной l на линии полной длины L; положим, что $L \gg l$, так что краевыми эффектами можно пренебречь. Линия L разделена на N смежных сегментов длиной m, так что

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

Длина m представляет размер ячейки, и мы ищем оптимальную длину m, т. е. такое ее значение, которое дает наибольшую информацию. Эта схема представлена на рис. 7.1. Расположим случайным образом на линии L сегмент AB длиной l. Отрезок CD представляет одну из ячеек длиной m. Если точка A

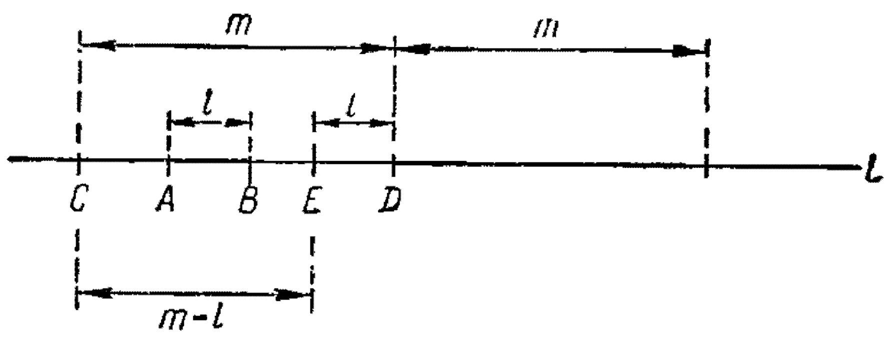

Рис. 7.1. Модель расположения с ячейками длиной $m = CD$ и длиной располагаемых элементов $l = AB$.

попадает между C и E, то сегмент AB полностью находится внутри ячейки. Если A попадает между E и D, то сегмент перекрывает точку D — место стыка двух ячеек.

Мы имеем:

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

Предполагается, что попадание AB в любую ячейку имеет одинаковую вероятность. Вероятность того, что AB окажется полностью внутри i-й ячейки, есть

$$p_i = \frac{m - l}{L} \tag{7.4}$$

(это выражение сводится к $p_i = \frac{m}{l}$ для точечных элементов). Полная вероятность того, что AB окажется полностью внутри какой-либо ячейки, равна

$$P = \sum_{i=1}^{i=N} \frac{m - l}{L} = \frac{N(m - l)}{L} = 1 - \frac{l}{m}. \tag{7.5}$$

Случай $m=0$ тривиален и не рассматривается, так что наличие m в знаменателе не приведет к недоразумениям.

Дополнительная вероятность того, что AB перекроет границу между ячейками и, следовательно, должен быть помещен в (дополнительную) *смешанную* ячейку, равна, очевидно

$$1-P\equiv q=\frac{l}{m}. \tag{7.6}$$

Количество информации (на элемент данных) выражается формулой (7.1):

$$\begin{aligned}\frac{I}{K}&=-\left(\sum_{i=1}^{i=N}p_i\ln p_i+q\ln q\right)=\\&=-\left[N\left(\frac{m-l}{L}\right)\ln\left(\frac{m-l}{L}\right)+\frac{l}{m}\ln\frac{l}{m}\right]=\\&=-\left[\left(1-\frac{l}{m}\right)\ln\left(\frac{m-l}{L}\right)+\frac{l}{m}\ln\frac{l}{m}\right].\end{aligned} \tag{7.7}$$

Желая получить максимум информации, мы приравниваем производную $\frac{dI}{dm}$ нулю:

$$\frac{I}{K}\frac{dI}{dm}=\frac{-l}{m^2}\ln\left(\frac{m-l}{L}\right)-\left(1-\frac{l}{m}\right)\left(\frac{1}{m-l}\right)+\frac{l}{m^2}\ln\frac{l}{m}+\frac{l}{m^2}=0,$$

или

$$-\frac{l}{m}\ln\left(\frac{m-l}{L}\right)-1+\frac{l}{m}\ln\frac{l}{m}+\frac{l}{m}=0,$$

что дает максимум I при

$$\frac{l}{m}\left\{1+\ln\left[\frac{l}{m}\left(\frac{L}{m-l}\right)\right]\right\}=1,$$

или

$$m=l\left(1+\ln\frac{lL}{m(m-l)}\right). \tag{7.8}$$

Очевидно, что оптимальное значение m имеет тот же порядок величины, что и l, но точное решение зависит (впрочем, некритически, в силу логарифмической зависимости) от выбора L, а точнее говоря, от отношения полной «длины» системы расположения к элементу l. Пусть, например, $L=e^5l$, т. е. $L\approx 150l$; это значит, что «типичная» ошибка занимает менее 1% всей системы. Тогда, обозначая $x = m/l$, имеем:

$$x = 6 - \ln[x(x-1)], \tag{7.9a}$$

решение которого $x \approx 3{,}7$; это значит, что размер ячейки должен быть примерно вчетверо больше длины элемента.

Если же $L = e^{10}l$ (т. е. $L \approx 22\,000l$), так что типичная ошибка занимает около 1/200% всей системы, то

$$x = 11 - \ln[x(x-1)], \tag{7.9b}$$

что дает $x \approx 8$.

2. Расположение с перекрестными ссылками

Рассмотрим теперь схему с перекрестными ссылками для данных, элементы которых не располагаются полностью внутри *одной* какой-либо ячейки. Прежде всего, предположим (как это может быть в системе, подверженной локализованным ошибкам), что элемент может перекрыть только две соседние ячейки. Наш анализ будет зависеть от того, как мы находим перекрытие и как мы его интерпретируем.

Если бы мы занимались, к примеру, системой расположения букв, то перекрестная ссылка на соседнюю ячейку означала бы просто, что букве присвоено более широкое место в пространстве расположения. В терминах схемы раздела 1 мы имели бы длину $(2m - l)$, в которую может попасть A. Таким образом (по-прежнему пренебрегая краевыми эффектами), мы получили бы

$$\frac{I}{K} = -\left\{\left[\sum_{i=1}^{i=N} \frac{m-l}{L} \ln\left(\frac{m-l}{L}\right)\right] + \frac{l}{m} \ln\left(\frac{2m-l}{L}\right)\right\}. \tag{7.10}$$

Заметим, что множитель во втором члене не изменился, в предположении, что априорная вероятность l/m перекрытия осталась неизменной, тогда как количество информации, получаемое в результате перекрестных ссылок, $\ln\left(\frac{2m-l}{L}\right)$ теперь, конечно, иное; находя максимум I по m, и, пользуясь прежним обозначением $\frac{m}{l} = x$, получим условие оптимума:

$$x = \frac{2x-1}{2x+1} \ln \frac{2x-1}{x-1}, \tag{7.11}$$

что дает $x \approx 1{,}06$; интересно заметить, что это значение не зависит от L.

В том же случае, когда наша классификационная система служит для непосредственного накопления статистической информации, перекрытие позволяет нам локализовать элемент. Это приводит к соотношению

$$\frac{I}{K} = -\left[\left(1 - \frac{l}{m}\right) \ln \frac{m-l}{L} + \frac{l}{m} \ln \frac{l}{L}\right], \tag{7.12}$$

и условие оптимума будет теперь

$$x = -\ln(x-1), \tag{7.13}$$

откуда $x \simeq 1{,}28$ — значение, также не зависящее от L.

Сравним теперь количество информации на элемент для различных методов расположения, всякий раз при оптимальных условиях. В первом случае двойного расположения при оптимальном значении m/l мы имеем:

$$\left(\frac{I}{K}\right)_{\text{opt}} = \ln \frac{L}{l} + \ln\left(\frac{l}{m-l}\right) - \frac{2m+l}{2m-l}, \tag{7.14}$$

или

$$\left(\frac{I}{K}\right)_{\text{opt}} = \ln\left(\frac{L}{l}\right) = 0{,}03, \tag{7.14a}$$

тогда как во втором случае

$$\left(\frac{I}{K}\right)_{\text{opt}} = \ln \frac{L}{l} + 0{,}28. \tag{7.15}$$

Сравнение этого последнего значения со случаем *смешанной* ячейки, рассмотренной в разделе 1, показывает, что система двойного расположения оказывается лучшей при условии

$$\frac{m-l}{l} + \ln \frac{m-l}{l} \geqslant -0{,}28,$$

т. е. когда

$$\frac{L}{l} \geqslant 1{,}11, \tag{7.16}$$

а при таком малом значении L задача расположения лишается смысла.

Заметим в заключение, что не представляет труда рассмотрение схемы, в которой элементы данных имеют некоторое статистическое распределение по длине. Так, если плотность вероятностей l есть $p(l)$, то, например, для расположения со смешанной ячейкой (см. (7.7)) имеем:

$$\frac{l}{K} = -\int_0^\infty \left\{\left(1-\frac{l}{m}\right)\ln\left[\left(1-\frac{l}{m}\right)\frac{m}{L}\right] + \frac{l}{m}\ln\frac{l}{m}\right\} p(l)\,dl. \quad (7.17)$$

Если взять

$$p(l) = \begin{cases} \frac{1}{l_0} & \text{при} \quad 0 \leqslant l \leqslant l_0, \\ 0 & \text{при} \quad l > l_0, \end{cases}$$

то после некоторых преобразований получаем условие максимальной информации:

$$\frac{m}{l_0} = 1 + \ln\left(\frac{Ll_0}{m(m-l_0)}\right) - \left(\frac{m}{l_0}\right)^2 \ln\left(\frac{m}{m-l_0}\right),$$

что для случая $\frac{L}{l_0} = e^5$, рассмотренного ранее, дает:

$$\frac{m}{l_0} = 2{,}2. \quad (7.18)$$

Во всех рассмотренных примерах мы имели дело с элементами, которые могут быть классифицированы (расположены) вдоль линии, в одном измерении. Сходные соображения могут быть развиты для элементов, зависящих от двух, трех и более переменных; такие элементы должны быть классифицированы на плоскости, в объеме или в многомерном пространстве. В этих условиях существенно изменится положение с перекрытием, так как один элемент может теперь перекрывать несколько различных ячеек.

3. Наивыгоднейшее число элементов на одну элементарную ячейку

На вопрос о наивыгоднейшем числе элементов можно получить различные ответы в зависимости от подхода. Станем на точку зрения физики и, рассматривая цепи с определенными постоянными времени, будем стремиться получить наибольшую эффективность. Нижеследующее рассуждение принадлежит Р. М. Уокеру (IBM Watson Laboratory). Предполагается, что шум отсутствует. Проблемы шума обсуждаются отдельно в последующих главах. Наши линии, цепи и детали обладают некоторой определенной постоянной времени τ. При возбуждении до энергии E_n цепь будет затухать по экспоненциальному закону

$$E = E_n e^{-t/\tau}. \tag{7.19}$$

Положим, что мы используем n равноотстоящих энергетических уровней

$$0,\ E_0,\ 2E_0,\ \ldots,\ E_n = (n-1) E_0. \tag{7.20}$$

Если возбужден наивысший уровень E_n и затухание происходит согласно (7.19), то, очевидно, мы должны ждать, пока E упадет до значения ниже $E_0/2$, прежде чем мы сможем отсчитать следующую энергетическую ступень, которая может быть 0 или E_0. Это ведет к условию

$$\frac{E_0}{2} = (n-1) E_0 e^{-t/\tau},$$

$$\frac{t}{\tau} = -\ln\frac{1}{2(n-1)} = \ln[2(n-1)]. \tag{7.21}$$

Время t, определяемое равенством (7.21), представляет кратчайший интервал между двумя следующими друг за другом импульсами, если мы хотим различать импульсы типа (7.20).

На протяжении большого времени T мы имеем общее число символов

$$G = \frac{T}{t} = \frac{T}{\tau \ln[2(n-1)]}. \tag{7.22}$$

Полагая, что все импульсы имеют одинаковые априорные вероятности, получим наибольшее возможное количество информации (см. (1.6)):

$$I = KG \ln n = \frac{KT}{\tau} f(n), \tag{7.23}$$

где

$$f(n) = \frac{\ln n}{\ln[2(n-1)]}. \tag{7.24}$$

Можно построить график $f(n)$ как функцию целого числа n, изменяющегося от 2 (двоичный код) до ∞ (непрерывное изменение); этот график дан на рис. 7.2. Функция $f(n)$ равна единице при $n=2$, достигает минимума, равного 0,77, при

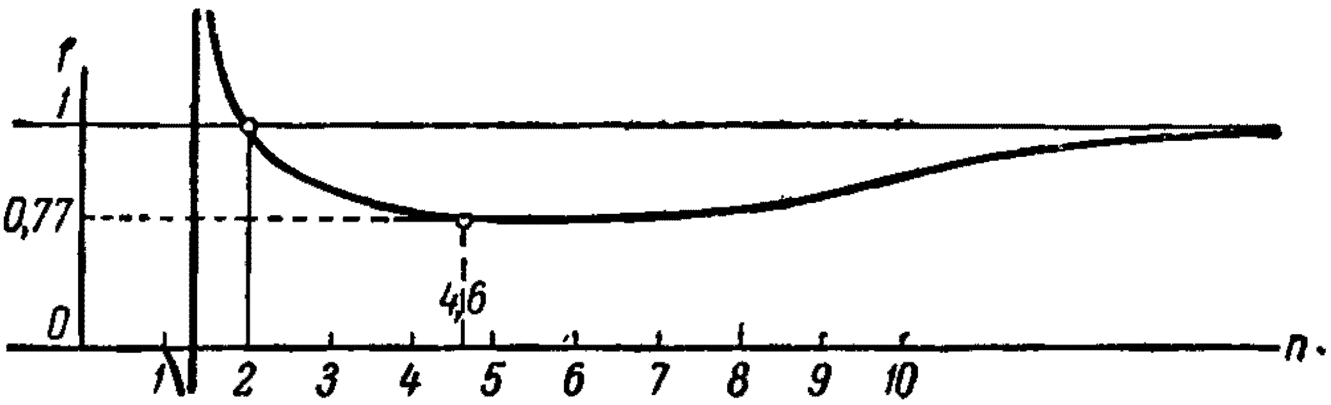

Рис. 7.2. График $f(n) = (\ln n)/\ln[2(n-1)]$ в функции n.

$n=4{,}6$ и, возрастая, стремится снова к единице при очень больших значениях n.

Этим доказано, что количество информации, передаваемое за данное время, максимально в случаях двоичного телеграфного кода и телефонной связи с непрерывным изменением.

ГЛАВА 8

АНАЛИЗ СИГНАЛОВ: МЕТОД ФУРЬЕ И ПРОЦЕДУРЫ ОТСЧЕТОВ

1. Ряд Фурье

В последующих главах нам понадобятся некоторые математические результаты, относящиеся к анализу различного рода сигналов. Настоящая глава посвящена систематическому обзору математических методов и их взаимоотношений.

Начнем с разложения Фурье периодических функций[1]). Пусть $f(t)$ периодическая функция с периодом τ. Эта функция может быть представлена суммой гармонических составляющих. В комплексной экспоненциальной форме

$$f(t)=\sum_{n=-\infty}^{n=+\infty} C_n e^{in\omega_0 t}, \quad \omega_0=\frac{2\pi}{\tau}, \tag{8.1}$$

где

$$C_n=\frac{1}{\tau}\int_0^\tau f(t)\, e^{-in\omega_0 t}\, dt. \tag{8.2}$$

Для вещественной $f(t)$ имеем:

$$C_{-n}=C_n^*, \tag{8.3}$$

где звездочкой обозначено комплексное сопряжение. Для вычисления интегралов (8.2) было предложено много практических методов. Рассмотрим метод конечных интервалов ввн-

---

[1]) Ясное и строгое изложение основ теории рядов Фурье дано у Бора. (H. Bohr, Almost Periodic Functions, Chapter 1, Chelsea, New York, 1947.)

ду его тесной связи со многими проблемами, рассматриваемыми в дальнейшем. Выберем N равноотстоящих точек на интервале $[0, \tau]$, соответствующем одному периоду

$$t_m = m\theta, \quad \theta = \frac{\tau}{N}, \text{ где } m \text{ — целое,} \tag{8.4}$$

и вычислим вместо интеграла (8.2) сумму

$$\gamma_{n'} = \frac{1}{N} \sum_{m=1}^{N} f(t_m)\, e^{-in'\omega_0 t_m}. \tag{8.5}$$

Заменяя $f(t_m)$ рядом Фурье (8.1), получаем:

$$\gamma_{n'} = \frac{1}{N} \sum_{m=1}^{N} \sum_{n=-\infty}^{n=+\infty} C_n e^{j(n-n')m\omega_0\theta}, \quad \omega_0\theta = \frac{2\pi}{N}. \tag{8.6}$$

Но

$$\sum_{m=1}^{N} e^{ipm\frac{2\pi}{N}} = \begin{cases} N & \text{при } p = qN, \\ 0 & \text{при } p \neq qN, \end{cases} \quad q \text{ — целое,} \tag{8.7}$$

так как во втором случае суммирование ведется по последовательности углов $pm\,2\pi/N$, равномерно распределенных между 0 и $2\pi p$. Не равные нулю члены суммы (8.6) соответствуют

$$p = n - n' = qN,$$

где q — положительное или отрицательное (или равное нулю) целое, так что

$$\gamma_{n'} = \sum C_n = \sum_{q} C_{n'+qN}. \tag{8.8}$$

Коэффициент γ'_n, вычисленный из (8.5), есть сумма коэффициентов C_n с индексами $n = n' + qN$. Кроме того, коэффициенты $\gamma_{n'}$ периодичны относительно n' с периодом N. Эти соотношения представляют частный пример общего результата[1]).

---

[1]) L. Brillouin, Wave Propagation in Periodic Structures, 1st ed. McGraw-Hill, NY, 1946, 2nd ed. Dover Publications, NY, 1952. См., в частности, гл. 1, стр. 7.

Интересен случай, когда для вещественной функции можно пренебречь гармониками выше некоторого предела n_M:

$$C_n = 0 \text{ при } |n| > n_M, \quad C_{-n} = C_n^*. \tag{8.9}$$

Тогда можно положить:

$$N = 2n_M + 1, \tag{8.10}$$

откуда

$$\gamma_n = C_n \text{ при } |n| \leqslant n_M. \tag{8.11}$$

Первые N членов γ_n дают в точности первые коэффициенты Фурье C_n, а следующие члены γ_n просто повторяют эти коэффициенты C_n с периодом N (рис. 8.1) в области более высоких частот.

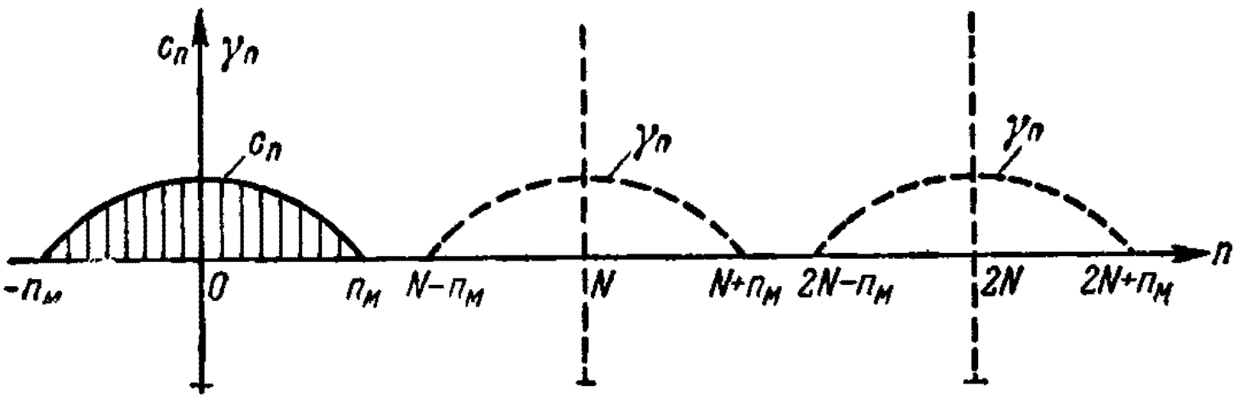

Рис. 8.1. График, показывающий равенство коэффициентов Фурье C_n и коэффициентов γ_n конечных сумм при $-n_M \leqslant n \leqslant n_M$ и периодичность γ_n как функции n для бо́льших значений n.

Очень важный результат известен под названием *равенства Парсеваля*, справедливого при условии, что функция $f(t)$ вещественна, ограничена и интегрируема на интервале $[0, \tau]$, соответствующем одному периоду:

$$\frac{1}{\tau}\int_0^\tau f^2\,dt = \sum_{n=-\infty}^{n=+\infty} |C_n|^2. \tag{8.12}$$

Это равенство легко доказывается заменой одного из $f(t)$ в подынтегральном выражении его разложением Фурье (8.1):

$$\frac{1}{\tau}\int_0^\tau f^2\,dt = \frac{1}{\tau}\int_0^\tau f \sum_n C_n e^{in\omega_0 t}\,dt = \sum_n C_n C_n^* = \sum_n |C_n|^2.$$

Во многих приложениях величины, фигурирующие в (8.12), представляют энергию, и, таким образом, равенство Парсеваля показывает, что энергия за период τ может быть подсчитана как путем интегрирования f^2, так и путем суммирования энергий составляющих Фурье.

2. Явление Гиббса и сходимость рядов Фурье

Условия сходимости рядов Фурье обсуждены во многих руководствах [1]). Разложение Фурье может применяться к функциям, имеющим конечное число разрывов на протяжении периода τ. Пусть имеется разрыв в точке t_0 со значениями справа и слева $f(t_0-0)$ и $f(t_0+0)$ и с правой и левой производными в t_0. Ряд Фурье при $t=t_0$ сходится к значению

$$\frac{1}{2}[f(t_0-0)+f(t_0+0)]. \tag{8.13}$$

Чтобы сделать наглядным этот интересный результат, рассмотрим специальный, но типичный пример. Пусть дана «пилообразная» функция (рис. 8.2) с периодом 2π и разрывами в точках $t=\pm\pi, \pm 3\pi, \ldots$

$$f(t)=t, \quad -\pi<t<t+\pi. \tag{8.14}$$

Величина скачка при $t=\pi$ равна -2π. Соответствующий ряд Фурье есть

$$f(t)=2\left(\frac{\sin t}{1}-\frac{\sin 2t}{2}+\frac{\sin 3t}{3}\ldots\right). \tag{8.15}$$

Возьмем теперь сумму S_p первых p членов и сравним ее с исходной функцией. Разность Δ равна

$$\Delta=f(t)-S_p=2\sum_{n=p+1}^{\infty}(-1)^{n+1}\frac{\sin nt}{n}. \tag{8.16}$$

Нас интересует положение в окрестности разрыва при $t=\pi$.

---

[1]) R. V. Churchill, Fourier Series and Boundary Value Problems, McGraw-Hill, NY, 1941; I. S. Sokolnikoff, E. S. Sokolnikoff, Higher Mathematics for Engineers and Physicists, p. 68, McGraw-Hill, NY, 1941.

Возьмем

$$t=\pi-\theta, \quad \sin nt=(-1)^{n+1}\sin n\theta,$$

тогда

$$\Delta=f-S_p=2\sum_{n=p+1}^{n=+\infty}\frac{\sin n\theta}{n}.$$

Последовательные значения n различаются на $\Delta n=1$, а θ предполагается весьма малым. Введем новую переменную

$$\eta=n\theta, \quad \Delta\eta=\theta, \quad \theta\ll 1.$$

Мы можем заменить сумму интегралом и записать:

$$\Delta=2\int\limits_{\eta=(p+1)\theta}^{\infty}\frac{\sin\eta}{\eta}\,d\eta=-2\,\mathrm{si}\,(p+1)\,\theta, \tag{8.17}$$

где $\mathrm{si}\,(x)$ означает интегральный синус в обычном определении [1]). Функция si имеет значение $-\pi/2$ при $x=0$ и совершает затухающие колебания, стремясь к нулю при неограниченном увеличении x. Первые экстремумы функции $\mathrm{si}\,(x)$:

$$\frac{(p+1)\,\theta}{\pi}=\frac{x}{\pi}=\begin{cases}1\\2\\3\\4\end{cases}\quad \mathrm{si}\,(x)\begin{cases}+0{,}281,\\-0{,}153,\\+0{,}104,\\-0{,}079.\end{cases} \tag{8.18}$$

Ход последовательных приближений показан на рис. 8.2. Отметим следующее:

А. Все парциальные суммы S_p представлены кривыми, проходящими через точку $t=\pi$, $S_p=0$. Следовательно, точка посредине скачка есть точка сходимости, что согласуется с условием (8.13).

В. Парциальные суммы дают в окрестности разрыва большие колебания, не соответствующие действительному ходу функции. При достаточно большом числе членов парциальной суммы эти колебания представлены кривой Δ (см. (8.17)).

---

[1]) Е. Янке и Ф. Эмде, Таблицы фувкцнй с формуламн и кривыми, Гостехиздат, 1948.

Колебания всегда имеют одну и ту же величину; расположение экстремумов дается (8.18). Парциальная сумма дает выброс при $\theta = \frac{\pi}{p}$ и превосходит истинное значение функции на 0,281, что соответствует относительной погрешности

$$\frac{\Delta}{f} = \frac{\Delta}{\pi} = \frac{2}{\pi}\,0{,}281 = 18\,\%. \tag{8.19}$$

По мере возрастания числа членов p, удерживаемых в парциальной сумме, протяженность возмущенной области убывает, но максимумы и минимумы сохраняют неизменную амплитуду. Это и есть *явление Гиббса.* На рис. 8.2 видно, что величина выброса все время составляет 18 % при 4, 6 или 10 членах парциальной суммы.

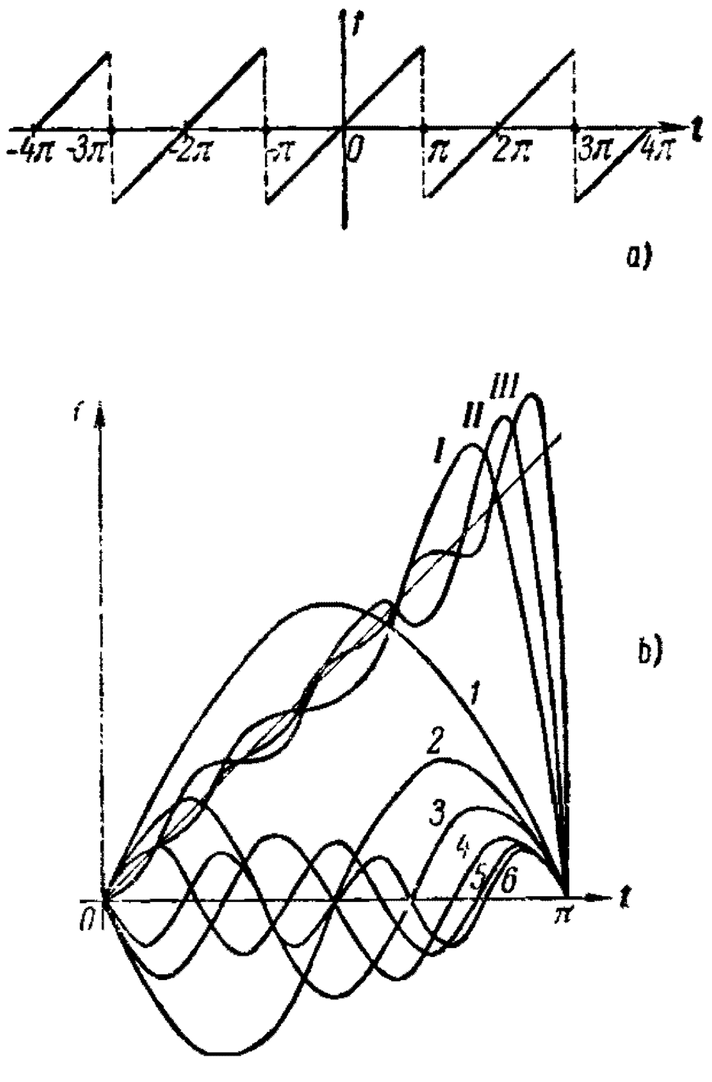

Рис. 8.2. *a*) График разрывной функции с периодом 2π

$$f(t) = t, \quad -\pi < t < \pi.$$

b) Первые шесть членов ряда Фурье для $f(t)$, построенных как функции времени. Показаны также суммы первых четырех (кривая *I*), первых шести (кривая *II*) и первых десяти (кривая *III*) членов.

Результат, полученный в этом специальном примере, является общим и приложим к любой функции с разрывами. Рассмотрим функцию $F(t)$ с периодом 2π со скачками на $D_1, D_2, D_3, \ldots$, при $t_1, t_2, t_3, \ldots$ Можно построить непрерывную функцию

$$G = F + \frac{D_1}{2\pi} f(t - t_1 + \pi) + \frac{D_2}{2\pi} f(t - t_2 + \pi) + \ldots$$

путем добавления к функции F некоторого количества функций (8.14), умноженных на подходящие коэффициенты. Ряд Фурье для непрерывной функции G быстро сходится; его коэффициенты убывают, как $1/n^2$ или еще быстрее. Разрывы функций f дают коэффициенты, убывающие только, как $1/n$ (см. (8.15)), и поведение ряда Фурье для F вблизи разрывов определяется в основном этими медленно сходящимися членами, отражающими выше обсужденные особенности.

Следует отметить еще одно важное обстоятельство: парциальная сумма S_p дает наилучшее возможное приближение исходной функции в смысле среднеквадратичного уклонения.

3. Интеграл Фурье

Рассмотрим интеграл Фурье на примере, очень близком к практическим задачам импульсной техники. Возьмем периодическую (с периодом τ) последовательность импульсов

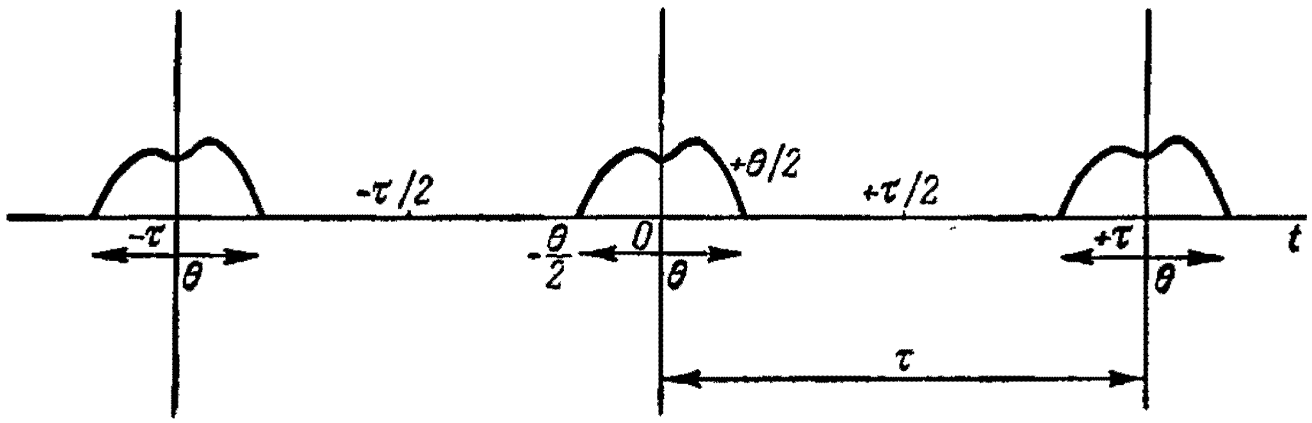

Рис. 8.3. Импульс длительностью θ, повторяемый периодически с интервалом τ.

длительностью θ и разложим эту последовательность в ряд Фурье в соответствии с (8.1) и (8.2) (рис. 8.3):

$$f(t)=\sum_{n=-\infty}^{n=+\infty} C_n e^{in\omega_0 t},\quad \omega_0=2\pi\nu_0=\frac{2\pi}{\tau},$$

$$C_n=\frac{1}{\tau}\int\limits_{-\tau/2}^{\tau/2} f(t)\,e^{-in\omega_0 t}dt,\qquad C_{-n}=C_n^*, \tag{8.20}$$

где $f(t)$ — вещественная функция. Так как импульс равен нулю вне короткого интервала θ, то пределы интегрирования могут быть взяты равными $\pm\theta/2$. Предположим, что расстояние τ между соседними импульсами постепенно увеличивается и стремится в пределе к бесконечности. Дискретная сумма для $f(t)$ может быть записана в виде

$$f(t)=\sum_{-\infty}^{\infty} C_n e^{i2\pi\nu_n t}\,\Delta n=\sum_{-\infty}^{\infty} C_n \tau e^{-i2\pi\nu t}\,\Delta\nu,\qquad \nu_n=n\nu_0,$$

$$\Delta n=1,\quad \Delta\nu=\nu_0=\frac{1}{\tau}.$$

Когда τ очень велико, сумма переходит в интеграл

$$\tau\to\infty:\ f(t)=\int_{\nu=-\infty}^{\infty} C(\nu)\,e^{i2\pi\nu t}\,d\nu, \tag{8.21}$$

и

$$\lim_{\tau\to\infty}(C_n\tau)=C(\nu)=\int_{t=-\infty}^{\infty} f(t)\,e^{-i2\pi\nu t}dt=C^*(-\nu). \tag{8.22}$$

В интеграле (8.22) интегрирование в действительности происходит на интервале θ, соответствующем одиночному импульсу, так как $f(t)\neq 0$ только при $-\theta/2<t<\theta/2$. Для

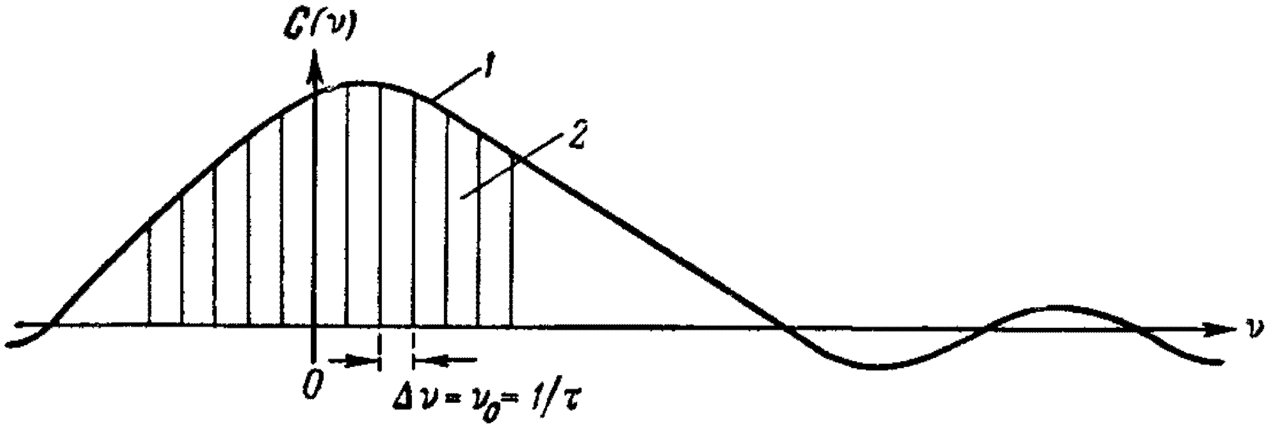

Рис. 8.4. Непрерывный спектр $C(\nu)$ интеграла Фурье (кривая *1*) как огибающая дискретного спектра $C_n\tau$ (*2*).

больших τ мы получаем для $C_n\tau$ дискретный спектр, огибающая которого выражается непрерывным спектром интеграла Фурье. Чем больше τ, тем короче интервал $\Delta\nu=\frac{1}{\tau}$ между линиями дискретного спектра рис. 8.4. Равенства (8.21) и (8.22) показывают полную взаимность между переменными t и ν: если импульс $f(t)$ имеет спектр $C(\nu)$, то импульс $C(t)$ имеет спектр $f(\nu)$. Если импульс симметричен, то и спектр симметричен:

$$f(t)=f(-t)=\int\limits_{\nu=-\infty}^{\infty} C(\nu)\cos 2\pi\nu t\,d\nu, \tag{8.23}$$

$$C(\nu)=C(-\nu)=\int\limits_{t=-\infty}^{\infty} f(t)\cos 2\pi\nu t\,dt. \tag{8.24}$$

Это — основные формулы, нужные для нашего исследования. Приведем несколько примеров.

Прямоугольный импульс: Спектр:

$$f(t)=\begin{cases} 1 \text{ при } |t|<\frac{1}{2}\theta_1, \\ 0 \text{ при } |t|>\frac{1}{2}\theta_1, \end{cases} \qquad C(\nu)=\theta_1\frac{\sin\pi\nu\theta_1}{\pi\nu\theta_1}. \tag{8.25}$$

Сглаженный импульс: Прямоугольный спектр:

$$f(t)=2\nu_M\frac{\sin 2\pi\nu_M t}{2\pi\nu_M t}, \qquad C(\nu)=\begin{cases} 1, |\nu|<\nu_M, \\ 0, |\nu|>\nu_M. \end{cases} \tag{8.26}$$

Треугольный импульс:

$$f(t)=\begin{cases} \left(1-\frac{|t|}{\theta_1}\right), & |t|<\theta_1, \\ 0 & |t|>\theta_1, \end{cases} \qquad C(\nu)=\theta_1\left(\frac{\sin\pi\nu\theta_1}{\pi\nu\theta_1}\right)^2. \tag{8.27}$$

Гауссов импульс: Гауссов спектр:

$$f(t)=e^{-\pi(t/\theta_1)^2}. \qquad C(\nu)=\theta_1 e^{-\pi(\nu\theta_1)^2}. \tag{8.28}$$

Соответствующие графики даны на рис. 8.5. Первые два случая являются примерами взаимно-обратных функций и демонстрируют t, ν - симметрию в (8.23) и (8.24).

Для интеграла Фурье справедливо равенство Парсеваля, сходное с равенством (8.12), относящимся к ряду Фурье. Мы имеем:

$$\int|f|^2\,dt=\int f^*f\,dt=\int f^*dt\int C(\nu)\,e^{i2\pi\nu t}\,d\nu.$$

Здесь $f(t)$ было заменено соответствующим интегралом Фурье на основании (8.21). Интегрируя по t и замечая, что

$$\int f^*(t)\,e^{i2\pi\nu t}\,d\nu=C^*(\nu),$$

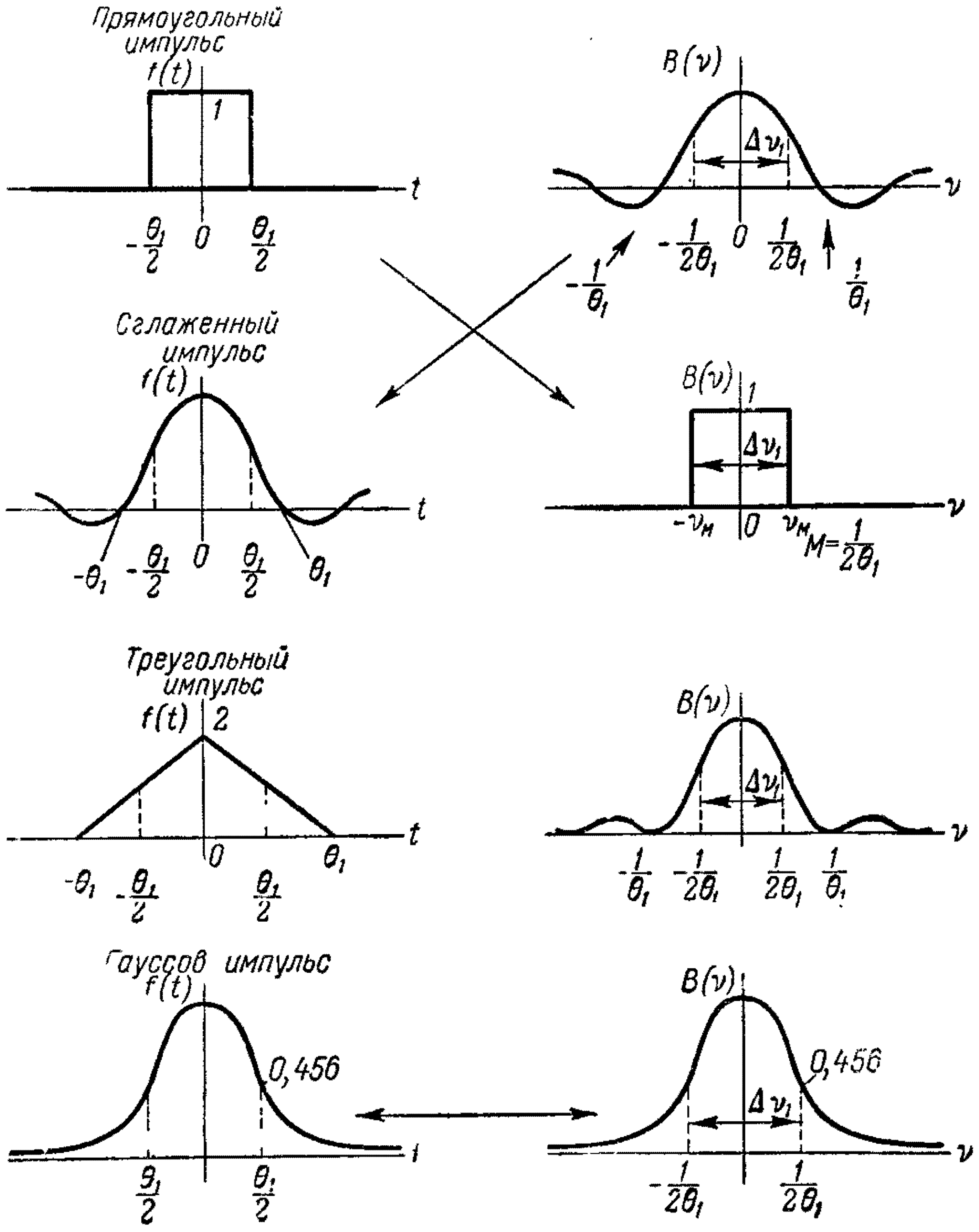

Рис. 8.5. Примеры функций и их спектров.

Эти графики соответствуют формулам (8.25) – (8.28). Следует отметить t, ν-симметрию (указанную стрелками) для первых двух случаев и для импульса и спектра в последнем случае.

получаем:

$$\int_{-\infty}^{\infty} |f(t)|^2\,dt = \int_{-\infty}^{\infty} |C(\nu)|^2\,d\nu = E. \qquad (8.29)$$

В большом числе физических задач эти интегралы выражают полную энергию колебания $f(t)$, и величина F может быть подсчитана как непосредственно по функции $f(t)$, так и по ее фурье-спектру.

4. Роль конечной ширины полосы частот

Одним из важных свойств любой системы связи является то, что она всегда использует конечную полосу частот. Система может быть типа фильтра нижних частот, если применяются все частоты от нуля до некоторой наивысшей частоты ν_M, или типа полосового фильтра, если применяются частоты в интервале от ν_1 до ν_2. Система типа фильтра верхних частот является чисто гипотетической, так как не существует физической системы без верхнего предела частоты. Исследование полосовой системы может быть сведено к случаю нижних частот, модулирующих несущую частоту. Мы можем поэтому ограничить свое рассмотрение случаем нижних частот. Вопрос ставится так: дана верхняя граничная частота ν_M; каковы сигналы, которые мы можем передавать? Или иначе: каковы свойства функций $f(t)$, спектр которых не простирается выше ν_M?

Это условие означает, что мы используем полосу частот $\Delta\nu_1$:

$$|\nu| \leqslant \nu_M, \quad -\nu_M \leqslant \nu \leqslant \nu_M, \quad \Delta\nu_1 = 2\nu_M. \qquad (8.30)$$

Прежде всего, нужно отметить, что сигнал длительности θ_1 требует для его воспроизведения полосы

$$\Delta\nu_1 \geqslant \frac{1}{\theta_1}, \quad \Delta\nu_1 \cdot \theta_1 \geqslant 1. \qquad (8.30a)$$

Практически полоса ограничена не резко. Для частот внутри полосы $C(\nu)$ велико по сравнению с его значениями вне полосы. Аналогично функция $f(t)$ представляет сигнал относительно большой интенсивности на протяжении длительности θ_1 и относительно малой интенсивности в остальное время. Произведение $\Delta\nu_1\theta_1$ имеет нижнюю грань. Качественно это видно из примеров (8.25) — (8.28), в которых сигналы, более четко определенные во времени, имеют более широкие спектры.

Задача установления количественной связи между шириной полосы и длительностью сигнала представляется эмпирической. Нижеследующий опыт из области телевидения дает основание для выбора количественных определений.

Установлено, что требуется полоса шириной около 6 *мгц* для передачи сигналов при 525 строках, 30 кадрах в секунду и 360 импульсах на строку. Это дает для длительности импульса

$$\theta_1 = \frac{1}{525 \cdot 30 \cdot 360} = \frac{1}{5{,}7} \cdot 10^{-6}$$

и

$$\Delta\nu_1\theta_1 = 6 \cdot 10^6 \cdot \frac{1}{5{,}7} \cdot 10^{-6} = 1{,}05.$$

Этот пример показывает, что ширина полосы и длительность должны удовлетворять неравенству

$$\Delta\nu_1\theta_1 \geqslant 1. \tag{8.30a}$$

Рассмотрим теперь примеры (8.25) — (8.28) предыдущего раздела. Мы имеем здесь короткие симметричные импульсы с максимальным значением при $t=0$ и действительной спектральной плотностью, принимающей наибольшее значение при $\nu=0$.

Нижеследующие определения $\Delta\nu_1$ и θ_1 находятся в соответствии с (8.30a)[1]:

$$f(0)\,\theta_1 = \int_{-\infty}^{\infty} f(t)\,dt, \quad \theta_1 \text{ — номинальная длительность,} \tag{8.31}$$

$$C(0)\,\Delta\nu_1 = \int_{-\infty}^{\infty} C(\nu)\,d\nu, \quad \Delta\nu_1 \text{ — номинальная ширина полосы.} \tag{8.32}$$

Но интеграл в (8.31) равен $C(0)$ из (8.24). Аналогично интеграл в (8.32) представляет $f(0)$ в (8.23), так что

$$\theta_1 = \frac{C(0)}{f(0)}, \quad \Delta\nu_1 = \frac{f(0)}{C(0)}, \quad \theta_1\Delta\nu_1 = 1. \tag{8.33}$$

---

[1]) H. A. Wheeler, PIRE **27**, 359 (1939); P. Le Corbeiller, Electronic Circuits and Tubes, pp. 586—590, McGraw-Hill, NY, 1947.

Эти определения дают равенство вместо неравенства (8.30). Они могут быть легко проверены на наших предыдущих примерах. Эти специальные примеры не отражают, однако, общей ситуации, и соотношения (8.31) и (8.32) не могут применяться для практического и общего определения длительности сигнала и ширины полосы. Так, при подсчете длительности сигнала θ_1 отрицательные значения f должны давать положительный вклад, тогда как в (8.31) они входят

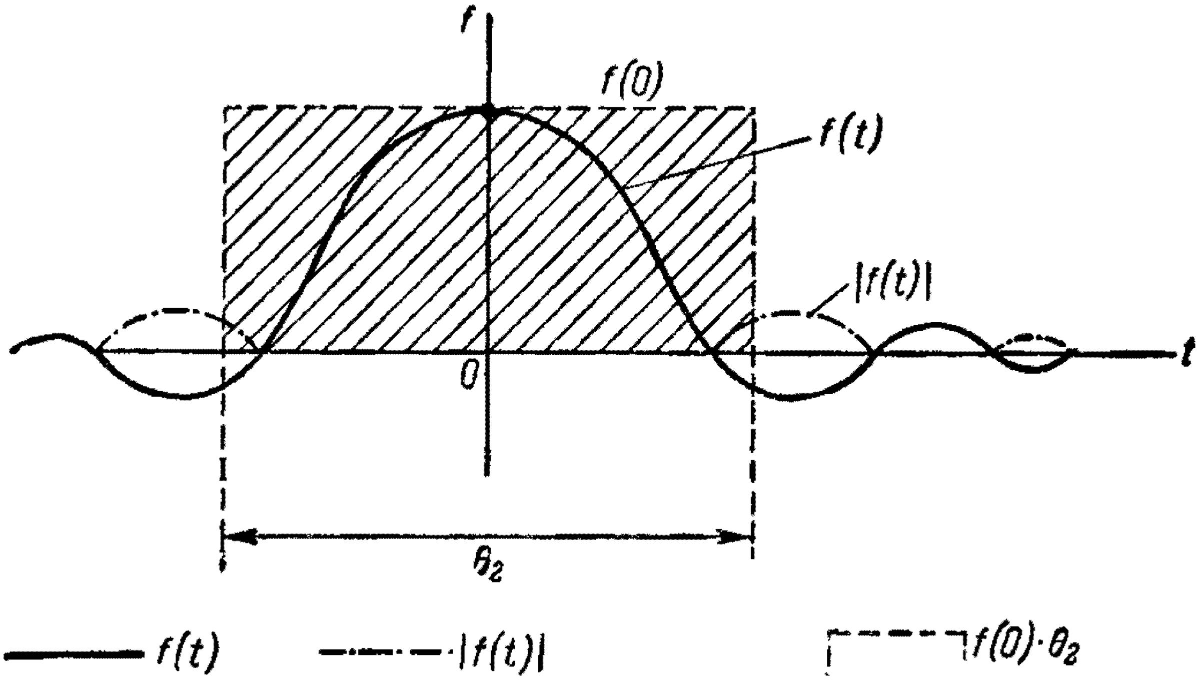

Рис. 8.6. Длительность θ_2 импульса получена путем приравнивания произведения $f(0)\cdot\theta_2$ площади под кривой абсолютных значений $|f(t)|$.

с отрицательным знаком. Мы можем испробовать следующие видоизмененные определения, основанные на абсолютных значениях:

$$|f(0)|\,\theta_2=\int_{-\infty}^{\infty}|f(t)|\,dt\geqslant\left|\int_{-\infty}^{\infty}f(t)\,dt\right|=|C(0)| \qquad (8.34)$$

и аналогично

$$|C(0)|\,\Delta\nu_2=\int_{-\infty}^{\infty}|C(\nu)|\,d\nu\geqslant\left|\int_{-\infty}^{\infty}C(\nu)\,d\nu\right|=|f(0)|. \qquad (8.35)$$

Эти определения дают:

$$\theta_2\geqslant\theta_1,\quad \Delta\nu_2\geqslant\Delta\nu_1,\quad \theta_2\Delta\nu_2\geqslant\theta_1\Delta\nu_1=1, \qquad (8.36)$$

что и представляет собой условие (8.30а), которому мы стремимся удовлетворить. Рис. 8.6 иллюстрирует метод: отрезок θ_2

выбирается так, чтобы прямоугольник с основанием θ_2 и высотой $|f(0)|$ имел площадь $|f(0)|\theta_2$, равную площади под кривой $|f(t)|$ («выпрямленной» кривой $f(t)$). Аналогично обстоит дело и с полосой частот. Длительность θ_2 может быть уменьшена, если выбрать начало отсчета времени в точке, в которой $|f|$ максимальна. Аналогично ширина полосы $\Delta\nu_2$ может быть уменьшена надлежащим выбором начала отсчета частот, для чего нужно умножить спектр на $e^{2\pi i\nu t}$. Эта операция есть попросту *демодуляция* сигнала. Но и эти новые значения θ_2' и $\Delta\nu_2'$ всегда удовлетворяют общему условию (8.36).

5. Соотношение неопределенности для времени и частоты

Общее условие (8.30а) представляет собой специальный случай принципа неопределенности квантовой механики. Согласно этому принципу переменная q и ее *сопряженный импульс* p не могут быть одновременно измерены с любой точностью, но существует соотношение между ошибкой Δq в определении q и ошибкой Δp в определении p. Это соотношение таково:

$$\Delta q\,\Delta p \geqslant \alpha, \tag{8.37}$$

где α берется иногда равной постоянной Планка h, а иногда равной постоянной Дирака $\hbar = \frac{h}{2\pi}$. Оба значения настолько малы, что любое из них приемлемо с точки зрения эксперимента.

Известно, что в механике энергия E есть импульс, сопряженный переменной времени t, так что согласно (8.37)

$$\Delta E\,\Delta t \geqslant \alpha. \tag{8.38}$$

Энергия колебаний частоты ν есть $h\nu$, и, следовательно,

$$\Delta E = h\Delta\nu,$$

и если воспользоваться эмпирическим условием (8.30а), то получим:

$$\Delta E\Delta t = h\Delta\nu\Delta t \geqslant h. \tag{8.39}$$

Поэтому следовало бы положить $\alpha = h$, а не $\alpha = \hbar$. Мы еще обсудим этот пункт в конце данного раздела.

Предварительно, однако, приведем рассуждение, при помощи которого в волновой механике вводится принцип неопределенности[1]). Вычисляется волновая функция $f(t)$, но не предполагается, что она имеет какой-либо прямой физический смысл. Вместе с тем предполагается, что квадрат $|f|^2$ дает плотность вероятностей для переменной t. Точнее говоря, вероятность того, что значение t окажется между t и $t+dt$, определяется как

$$p(t)\,dt = \frac{1}{E}\,|f(t)|^2\,dt, \tag{8.40}$$

где E есть полная энергия волны, определяемая (8.29), из которого следует:

$$\int_{-\infty}^{\infty} p(t)\,dt = 1. \tag{8.40a}$$

Аналогично, вероятность того, что частота заключена между ν и $\nu+d\nu$, равна

$$P(\nu)\,d\nu = \frac{1}{E}\,|C(\nu)|^2\,d\nu \tag{8.41}$$

и

$$\int_{-\infty}^{\infty} P(\nu)\,d\nu = 1. \tag{8.41a}$$

Теперь среднее время появления сигнала есть

$$t_0 = \int_{-\infty}^{\infty} tp(t)\,dt, \tag{8.42}$$

и средняя частота

$$\nu_0 = \int_{-\infty}^{\infty} \nu P(\nu)\,d\nu. \tag{8.43}$$

Уклонения t_1 и ν_1 от этих средних значений определяются

---

[1]) См., например, P. Gabor, MIT Lectures, 1951.

соотношениями

$$t = t_0 + t_1, \quad \nu = \nu_0 + \nu_1.$$

Средние квадраты этих уклонений равны

$$\overline{t_1^2} = \int_{-\infty}^{\infty} t_1^2 p(t)\, dt, \quad \overline{\nu_1^2} = \int_{-\infty}^{\infty} \nu_1^2 P(\nu)\, d\nu. \tag{8.44}$$

Мы определяем средние ошибки Δt по времени и $\Delta\nu$ по частоте как

$$\Delta t = \sqrt{a\overline{t_1^2}}, \quad \Delta\nu = \sqrt{a\overline{\nu_1^2}} \tag{8.45}$$

и выбираем постоянную a так, чтобы удовлетворялось условие (8.30а). Иначе говоря, мы выбираем a так, чтобы было

$$\Delta t\, \Delta\nu = \sqrt{a^2\overline{t_1^2}\,\overline{\nu_1^2}} \geqslant 1, \text{ или } a^2\overline{t_1^2}\,\overline{\nu_1^2} \geqslant 1. \tag{8.45a}$$

С помощью (8.40), (8.41) и (8.44) получим вместо (8.45а).

$$a^2 \int_{t=-\infty}^{\infty} t_1^2 f^*(t) f(t)\, dt \int_{\nu=-\infty}^{\infty} \nu_1^2 C^*(\nu) C(\nu)\, d\nu \geqslant E^2. \tag{8.46}$$

Без потери общности можно положить $t_0 = 0$ и $\nu_0 = 0$; тогда $t_1 = t$ и $\nu_1 = \nu$.

Первый шаг в определении a состоит в доказательстве тождества

$$\int_{\nu=-\infty}^{\infty} \nu^2 C^*(\nu) C(\nu)\, d\nu = -\frac{1}{4\pi^2} \int_{t=-\infty}^{\infty} f^*(t) \frac{d^2 f}{dt^2}\, dt, \tag{8.47}$$

что можно сделать при условии, что как $f(t)$, так и $C(\nu)$ дифференцируемы и достаточно быстро убывают. Интеграл Фурье (8.21) дает:

$$\frac{d^2 f}{dt^2} = \int_{\nu=-\infty}^{\infty} C(\nu)(-4\pi^2\nu^2)\, e^{2\pi i\nu t}\, d\nu,$$

и, таким образом,

$$\int\limits_{t=-\infty}^{\infty} f^* \frac{d^2 f}{dt^2}\, dt = -4\pi^2 \int\limits_{\nu=-\infty}^{\infty} \nu^2 C(\nu)\, d\nu \int\limits_{t=-\infty}^{\infty} f^* e^{2\pi i \nu t}\, dt.$$

Интеграл по t есть не что иное, как $C^*(\nu)$ (см. (8.22)), и тождество (8.47) этим доказано.

Применяя (8.47), можно переписать (8.46) в виде

$$-\frac{a^2}{4\pi^2} \int\limits_{t=-\infty}^{\infty} t^2 f^* f\, dt \int\limits_{t=-\infty}^{\infty} f^* \frac{d^2 f}{dt^2}\, dt \geqslant \left| \int\limits_{t=-\infty}^{\infty} f^* f\, dt \right|^2. \quad (8.48)$$

Для дальнейшего воспользуемся неравенством Шварца[1])

$$4 \int F^* F\, dt \int G^* G\, dt \geqslant \left| \int (F^* G + F G^*)\, dt \right|^2, \quad (8.49)$$

где

$$F = tf, \quad G = \frac{df}{dt}.$$

Вычисляя второй интеграл по частям и учитывая, что f стремится к нулю на бесконечности, получаем:

$$\int\limits_{-\infty}^{\infty} G^* G\, dt = \int\limits_{-\infty}^{\infty} \frac{df^*}{dt} \frac{df}{dt}\, dt = f^* \frac{df}{dt} \Bigg|_{-\infty}^{\infty} - \int\limits_{-\infty}^{\infty} f^* \frac{d^2 f}{dt^2}\, dt =$$

$$= -\int\limits_{-\infty}^{\infty} f^* \frac{d^2 f}{dt^2}\, dt. \quad (8.50)$$

Интегрируем также правую часть неравенства Шварца по частям:

$$\int\limits_{-\infty}^{\infty} t \left(f^* \frac{df}{dt} + f \frac{df^*}{dt} \right) dt = (t\, |f|^2) \Bigg|_{-\infty}^{\infty} - \int\limits_{-\infty}^{\infty} |f|^2\, dt = -E,$$

и неравенство Шварца принимает теперь вид

$$-4 \int\limits_{-\infty}^{\infty} t^2 f f^*\, dt \int\limits_{-\infty}^{\infty} f^* \frac{d^2 f}{dt^2}\, dt \geqslant E^2.$$

---

[1]) Доказательство неравенства Шварца дано в приложении в конце этой главы.

Сравнение с (8.48) показывает, что нужно взять

$$a^2 = 16\pi^2 \quad \text{или} \quad a = 4\pi,$$

т. е. мы определяем:

$$\Delta t = \sqrt{4\pi \overline{t_1^2}}, \quad \Delta\nu = \sqrt{4\pi \overline{\nu_1^2}}.$$

Интересно сравнить эти определения с теми, которые выражены соотношениями (8.31) и (8.32). В предшествующем разделе для примеров (8.25), (8.27) и (8.28) мы имели длительность θ_1 и полосу $1/\theta_1$. Новые определения дают:

$$(8.25):\quad E = \theta_1, \quad \Delta t = \sqrt{\frac{\pi}{3}}\,\theta_1, \quad \Delta\nu = \infty, \quad \Delta t\Delta\nu = \infty;$$

$$(8.27):\quad E = \frac{2}{3}\theta_1, \; \Delta t = \sqrt{\frac{2\pi}{5}}\,\theta_1, \; \Delta\nu = \frac{1}{\theta_1}\sqrt{\frac{3}{\pi}},$$

$$\Delta t\Delta\nu = \sqrt{\frac{6}{5}};$$

$$(8.28):\quad E = \frac{\theta_1}{\sqrt{2}}, \quad \Delta t = \sqrt{2}\,\theta_1, \quad \Delta\nu = \frac{1}{\theta_1}\sqrt{2}, \; \Delta t\Delta\nu = 2.$$

В вычислениях, предоставляемых читателю, встречаются интегралы

$$\int_{-\infty}^{\infty}\left(\frac{\sin x}{x}\right)^2 dx = \pi, \quad \int_{-\infty}^{\infty}\frac{\sin^4 x}{x^2}\,dx = \frac{\pi}{2}, \quad \int_{-\infty}^{\infty} x^2 e^{-x^2} dx = \sqrt{\pi}.$$

Во многих статьях и руководствах по квантовой механике принимается $a = 2$. Тогда

$$\Delta_1 t = \sqrt{2\overline{t_1^2}} = \frac{\Delta t}{\sqrt{2\pi}}, \quad \Delta_1\nu = \frac{\Delta\nu}{\sqrt{2\pi}},$$

что дает:

$$\Delta_1 t\Delta_1\nu \geqslant \frac{1}{2\pi}.$$

Если такие же определения приняты для $\Delta_1 q$ и $\Delta_1 p$, то соотношение неопределенности принимает вид

$$\Delta_1 q\Delta_1 p \geqslant \hbar.$$

Эти определения представляются непрактичными в свете предшествующего обсуждения. Они ведут к тому, что нижняя грань $\Delta\nu\Delta t$ делится на 2π. Для случая телевидения, упомянутого в предыдущем разделе, это означало бы, что достаточна полоса около 1 *Мгц*. Разумеется, инженеры телевидения заменили бы полосу 6 *Мгц* шестью полосами по 1 *Мгц*, если бы это было возможно.

6. Степени свободы сообщения

Рассмотрим теперь следующий вопрос: некоторая функция времени $f(t)$ имеет спектр, не содержащий частот выше ν_M, и существует на промежутке времени τ. Сколько параметров (или степеней свободы) необходимо для определения такой функции? Мы докажем, что для такой функции имеется только

$$N = 2\nu_M\tau \tag{8.51}$$

(предполагая, что $\nu_M\tau \gg 1$) независимых параметров, и обсудим различные возможные выборы этих параметров, а также некоторые общие свойства рассматриваемых функций.

Заметим прежде всего, что функция не определена полностью, если заданы ее значения только на конечном интервале, например $0 < t < \tau$. Для корректной постановки задачи нужно задать значения функции до 0 и после τ. При этом не должно добавляться что-либо к информации, содержащейся в сообщении $f(t)$.

Можно применить два различных определения.

А. Периодическая функция $f(t)$, бесконечно повторяющая ход f между 0 и τ:

$$f(t + q\tau) = f(t) \tag{8.52}$$

где q — целое.

В. Функция одиночного сообщения (single message function):

$$f(t) = 0 \quad \text{при} \quad t < 0 \quad \text{или} \quad t > \tau. \tag{8.53}$$

Последний пример рассмотрен Шенноном.

Начнем с определения А и исследуем периодическую функцию с периодом τ. Представим эту функцию рядом

Фурье:

$$f(t) = \frac{1}{2} a_0 + \sum_{n=1}^{n_M} (a_n \cos n\varphi + b_n \sin n\varphi), \tag{8.54}$$

где

$$\varphi = \frac{2\pi t}{\tau} = 2\pi\nu_0 t, \quad \nu_0 = \frac{1}{\tau}. \tag{8.55}$$

Положим, что высшая частота ν_M равна в точности частоте одной из гармоник:

$$n_M = \nu_M \tau \quad \text{или} \quad n_M \nu_0 = \nu_M. \tag{8.56}$$

Ряд Фурье содержит конечное число членов вплоть до члена с номером n_M. Для каждой гармоники имеем две составляющие a_n и b_n, а всего

$$N = 2n_M + 1 = 2\nu_M \tau + 1 \tag{8.57}$$

составляющих, включая и постоянную a_0. Если длительность сигнала достаточно велика, то (8.57) практически переходит в (8.51). Коэффициенты a_n и b_n представляют один возможный выбор параметров. Вместо вещественного ряда Фурье (8.54) мы можем воспользоваться комплексной формой (8.1) и (8.2):

$$f(t) = \sum_{n=-n_M}^{+n_M} c_n e^{in\omega_0 t}, \quad \omega_0 = 2\pi\nu_0 = \frac{2\pi}{\tau}, \tag{8.58}$$

$$c_n = \frac{1}{2}(a_n - ib_n) = \frac{1}{\tau}\int_0^\tau f(t')\, e^{-in\omega_0 t'}\, dt' = c^*_{-n},$$

где звездочка означает комплексно-сопряженную величину.

Вместо ряда Фурье мы можем применить к нашим периодическим функциям $f(t)$ метод отсчетов (sampling method). Выберем N равноотстоящих точек на протяжении периода τ, например:

$$t_m = m\theta, \quad \theta = \frac{\tau}{N} = \frac{1}{2\nu_M + \nu_0}, \tag{8.59}$$

причем

$$N = 2n_M + 1 = \tau(2\nu_M + \nu_0) = 2\nu_M\tau + 1,$$

и назовем отсчетом (sampled value) мгновенное значение

$$f_m = f(m\theta), \tag{8.59a}$$

причем в силу условия периодичности (8.52)

$$f_{m+qN} = f_m,$$

где q — целое. Мы можем восстановить исходную функцию $f(t)$, если известны $2n_M + 1$ отсчетов на протяжении периода τ. Запишем:

$$f(t) = \sum_{m=1}^{N} f_m g(t - m\theta), \tag{8.60}$$

где $g(t)$ означает импульсную функцию, периодически повторяемую в моменты $t = q\tau$. Этой импульсной функции мы придаем следующие свойства:

$$\begin{aligned} &g(t) = 1 \quad \text{при} \quad t = 0, \quad \text{или} \quad t = q\tau \qquad (q \text{ — целое}),\\ &g(t) = 0 \quad \text{при} \quad t = m\theta, \qquad m \neq qN. \end{aligned}$$

Импульсная функция равна нулю во всех остальных точках отсчета на протяжении периода τ. Такую функцию нетрудно построить с конечным спектром, ограниченным частотой ν_M. Решение задачи содержится в тождестве Лагранжа:

$$\frac{1}{2} + \sum_{n=1}^{n_M} \cos n\varphi = \frac{1}{2} \sum_{n=-n_M}^{+n_M} e^{in\varphi} = \frac{\sin\left(n_M + \frac{1}{2}\right)\varphi}{2\sin\left(\frac{\varphi}{2}\right)}, \tag{8.61}$$

где берется $\varphi = \omega_0 t$. Эта функция равна $n_M + \frac{1}{2}$ при $t = 0$ или $t = q\tau$, где знаменатель равен нулю. Она колеблется и обращается в нуль в точках

$$\left(n_M + \frac{1}{2}\right)\omega_0 t = \pi,\ 2\pi,\ \ldots,\ m\pi,$$

$$t = \frac{\tau}{N},\ \frac{2\tau}{N}, \ldots,\ \frac{m\tau}{N}, \quad m \neq qN,$$

если m не является целым. кратным N.

Мы можем взять для нашей импульсной функции $g(t)$ выражение

$$g(t)=\frac{\sin\left(N\frac{\omega_0 t}{2}\right)}{N\sin\left(\frac{\omega_0 t}{2}\right)}=\frac{1}{N}\sum_{-n_M}^{+n_M} e^{in\omega_0 t}. \tag{8.62}$$

Сравнивая (8.60), (8.62) и (8.58), получаем:

$$f(t)=\frac{1}{N}\sum_{-n_M}^{+n_M}\sum_{m=1}^{N} f_m e^{in\omega_0 (t-m\theta)},$$

откуда

$$c_n=\frac{1}{N}\sum_{m=1}^{N} f_m e^{-inm\omega_0\theta}, \tag{8.63}$$

— равенство, непосредственно связывающее коэффициенты Фурье c_n [1]) с отсчетами f_m. Обратное соотношение получается из (8.58):

$$f_m=f(m\theta)=\sum_{n'=-n_M}^{+n_M} c_{n'} e^{in'm\omega_0\theta}. \tag{8.63a}$$

В комплексном ряде Фурье мы имеем $N=2n_M+1$ комплексных амплитуд c_n, которые являются комплексно-сопряженными для $\pm n$. Это дает в общем N независимых действительных переменных, и наши N точек отсчета обеспечивают равное число степеней свободы. Легко непосредственно проверить (8.63) и (8.63a):

$$\left.\begin{aligned} &c_n=\frac{1}{N}\sum_{m=1}^{N}\sum_{n'=-n_M}^{n_M} c_{n'} e^{i(n'-n)m\omega_0\theta}, \\ &\text{но} \\ &\frac{1}{N}\sum_{m=1}^{N} e^{-ipm\omega_0\theta}=\begin{cases} 1 & \text{для } p=0; \\ 0 & \text{для } p\neq 0 \text{ (целое)} \end{cases} \end{aligned}\right\} \tag{8.63b}$$

---

[1]) (8.63) тождественно (8.5) и (8.11), так как мы имеем здесь дело с функцией, имеющей конечную ширину спектра, как в (8.9).

так как для $p=0$ имеется N членов, равных единице, тогда как для $p \neq 0$ N гармонических членов имеют фазы, равномерно распределенные между 0 и 2π, что дает в результате нуль. В общем,

$$c_n = \sum c_{n'} \delta(n - n'),$$

что и требовалось доказать.

Вместо отсчитывания в моменты $m\theta$ можно взять моменты $m\theta + \theta'$ с постоянным смещением $\theta' < \theta$. Это дает другую систему отсчетов $f'_m = f(m\theta + \theta')$, которую можно применять вместо f_m. Равенства (8.63) и (8.63а) заменяются на

$$c_n = \frac{1}{N} \sum_m f'_m e^{-in\omega_0 (m\theta + \theta')},$$

$$f'_m = \sum_{n'} c_{n'} e^{in'\omega_0 (m\theta + \theta')}.$$

Между f_m и f'_m существует система N линейных зависимостей.

7. Метод отсчетов Шеннона

Шеннон применяет метод отсчетов в связи с задачей В (см. (8.53)) о функции одиночного сообщения, предполагаемой равной нулю при $t < 0$ или $t > \tau$. Мы можем решить эту задачу следующим образом: возьмем периодическую функцию с длинным периодом $\tau_1 > \tau$ и положим, что $f(\tau)$ имеет свои исходные значения на интервале от 0 до τ и равна нулю от τ до τ_1. При этом, конечно, предполагается плавный переход к нулевой линии на обоих концах, во избежание появления частот выше предела ν_M. Новая функция отсчитывается на новом интервале τ_1, в N_1 равноотстоящих точках в соответствии с процедурой, описываемой равенствами (8.59) и (8.62). Новый интервал отсчетов θ_1 аналогично (8.59) равен

$$\theta_1 = \frac{1}{2\nu_M + \nu_1}, \quad \nu_1 = \frac{1}{\tau_1}, \quad \omega_1 = \frac{2\pi}{\tau_1}. \tag{8.64}$$

Если τ и τ_1 велики, то ν_0 и ν_1 пренебрежимо малы по сравнению с высшей частотой ν_M, и мы имеем:

$$\theta_0 \approx \theta_1 \approx \frac{1}{2\nu_M}.$$

Полное число точек отсчета составляет теперь:

$$N_1 = 2\nu_M \tau_1 + 1. \tag{8.65}$$

Из числа этих N_1 точек мы имеем N точек, попадающих в первоначальный интервал $0 < t < \tau$, и $N_1 - N$ точек в интервале $\tau < t < \tau_1$. Первая система точек дает ненулевые отсчеты

$$f_m = f(m\theta_1) \neq 0, \quad 0 < m < N, \tag{8.66}$$

остальные точки дают нули:

$$f_m = f(m\theta_1) = 0, \quad N < m < N_1.$$

Импульсная функция $g(t)$ из (8.62) принимает вид

$$g(t) = \frac{\sin\left(N_1 \frac{\omega_1 t}{2}\right)}{N_1 \sin\left(\frac{\omega_1 t}{2}\right)}. \tag{8.67}$$

Будем устремлять τ_1 к бесконечности, а ν_1 — к нулю. В пределе получим интервал отсчетов

$$\theta_{\text{lim}} = \frac{1}{2\nu_M}$$

и будем иметь бесконечное число отсчетов f_m, равных нулю. Не равны нулю только те отсчеты, которые относятся к интервалу

$$0 < m\,\theta_{\text{lim}} < \tau.$$

Число этих отсчетов равно

$$N_{\text{lim}} = \frac{\tau}{\theta_{\text{lim}}} = 2\nu_M \tau, \tag{8.68}$$

что совпадает с (8.51), тогда как импульсная функция $g(t)$

приводится к виду

$$g(t) \simeq \frac{\sin\left(N_1 \frac{\omega_1 t}{2}\right)}{N_1 \frac{\omega_1 t}{2}} \approx \frac{\sin 2\pi\nu_M t}{2\pi\nu_M t}, \tag{8.69}$$

так как

$$N_1\omega_1 = (2\nu_M\tau_1 + 1)\frac{2\pi}{\tau_1} = 4\pi\nu_M + \frac{2\pi}{\tau_1} \approx 4\pi\nu_M.$$

Таким образом, функция представляется в пределе равенством (8.60), принимающим форму:

$$f(t) = \sum_m f_m \frac{\sin \pi(2\nu_M t - m)}{\pi(2\nu_M t - m)}. \tag{8.70}$$

Легко показать, что разложение (8.70) принимает значения f_m во всех точках отсчета. Возьмем, например, точку q:

$$t_q = \frac{q}{2\nu_M}.$$

Член $m = q$ в сумме (8.70) дает составляющую f_q, а все остальные члены равны нулю:

$$m \neq q: \quad \sin\pi(q - m) = 0.$$

Сумма (8.70) не дает в точности $f(t) = 0$ для $t < 0$ и $t > \tau$, но представляет функцию, быстро убывающую в обе стороны с небольшими колебаниями частоты ν_M. Такое представление применялось Шенноном.

Функция одиночного сообщения $f(t)$, определенная в соответствии с (8.53), имеет только N степеней свободы, как показывает вышеописанный метод отсчетов. Разложение этой функции по Фурье приводит к интегралу Фурье вместо ряда Фурье. Число членов разложения Фурье бесконечно, но оно содержит лишь N независимых переменных f_m. Мы можем проверить это, так как наша функция $g(t)$ совпадает со сглаженным импульсом (8.26) и имеет спектр

$$B(\nu) = \begin{cases} \frac{1}{2\nu_M} & \text{при} \quad |\nu| < \nu_M, \\ 0 & \text{при} \quad |\nu| > \nu_M, \end{cases} \tag{8.70a}$$

откуда

$$g(t)=\frac{1}{2\nu_M}\int\limits_{\nu=-\nu_M}^{+\nu_M}\cos 2\pi\nu t\,d\nu=\frac{1}{2\nu_M}\int\limits_{\nu=-\nu_M}^{+\nu_M}e^{i2\pi\nu t}d\nu \quad (8.71)$$

и

$$f(t)=\sum_m \frac{f_m}{2\nu_M}\int\limits_{-\nu_M}^{+\nu_M}e^{i2\pi\nu(t-m\theta)}\,d\nu,\ \text{где}\ \theta=\frac{1}{2\nu_M}. \quad (8.72)$$

Общие принципы этой проблемы были открыты независимо несколькими учеными. Первоначальная идея принадлежит, по-видимому, Уиттекеру (1915 г.) [1, 2]. Дальнейшие подробности и приложения даны в разделах 6 и 7 главы 17.

8. Информационные ячейки Гэйбора

Два рассмотренных представления являются частными случаями более общего метода, рассмотренного Гэйбором [3]).

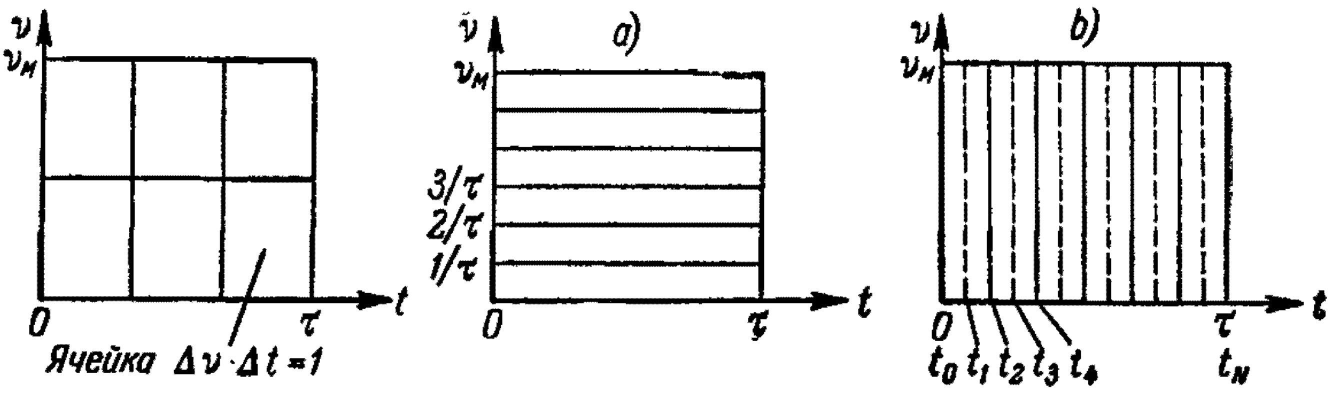

Рис. 8.7. Различный выбор ячеек единичной площади на плоскости ν, t: *a*) ряды Фурье, *b*) метод отсчетов.

На плоскости ν, t область, занимаемая функцией, есть прямоугольник $\nu_M\tau$ (рис. 8.7). Этот прямоугольник может быть разделен на элементарные ячейки единичной площади

$$\Delta\nu\Delta t=1, \quad (8.73)$$

и в каждой ячейке функция имеет две составляющие — сим-

[1]) E. T. Whittaker, Proc. Roy. Soc., Edinburgh **35**, 181 (1915).

[2]) В отчетливой и современной форме разложеиие (8.70) дано В. А. Котельниковым в 1933 г. (*Прим. перев.*)

[3]) D. Gabor, Theory of Communication, J. Inst. El. Eng. **93**, III, 429 (1946); **94**, III, 369 (1947); см. также Phil. Mag. [7] **41**, 1161 (1950); La Cybernetique, p. 115, Editions Revue d'Optique, Paris, 1951.

метричную и антисимметричную, — две элементарные функции, использующие ширину полосы $\Delta\nu$ и длительность Δt. Амплитуды обеих функций должны быть заданы и полностью определяют нашу функцию $f(t)$. Число ячеек равно

$$n = \nu_M \tau.$$

Число составляющих

$$N = 2n = 2\nu_M \tau, \tag{8.74}$$

как в (8.51). Форма ячеек не играет роли; имеет значение только их площадь. Ряд Фурье (8.58) использует горизонтальные ячейки, а метод отсчетов (8.61) — вертикальные. Гэйбор рекомендует в качестве элементарных — функции с гауссовой огибающей и с заполнением по синусу или по косинусу,

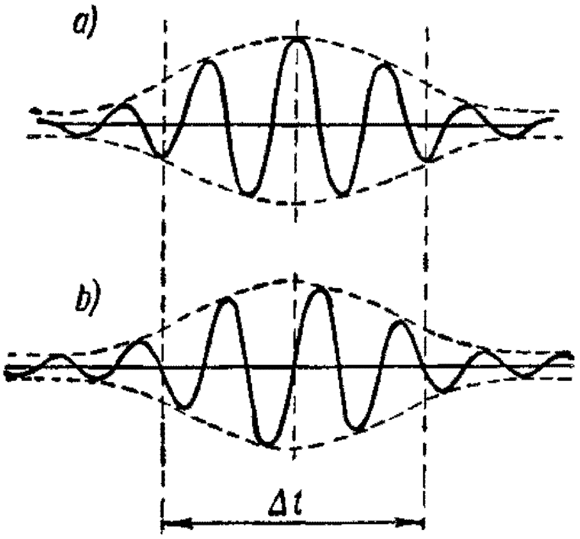

Рис. 8.8. Гэйборовы синусная (*b*) и косинусная (*a*) функции сообщения.

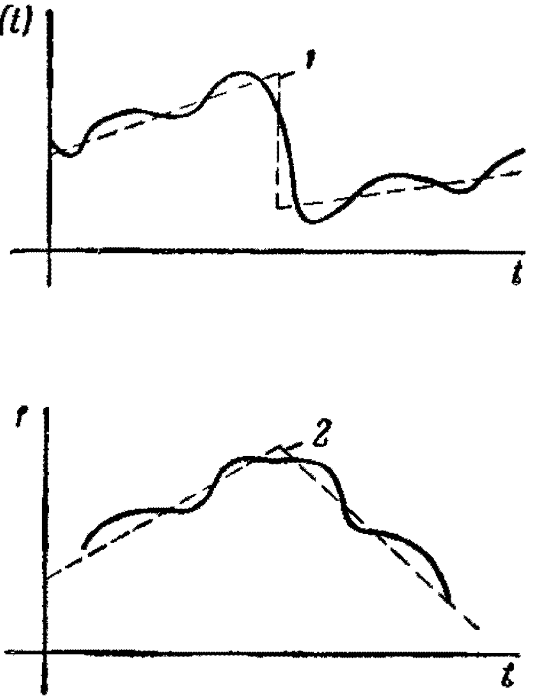

Рис. 8.9. Представление функции с конечной шириной полосы.

1 — разрывная функция, *2* — функция с разрывной производной.

как показано на рис. 8.8. Ячейка (8.73) соответствует ячейкам квантовой механики, где энергия E и время t — две сопряженные переменные, и ячейка определяется соотношениями

$$\Delta E \, \Delta t = h, \quad \text{где} \quad E = h\nu \quad \text{и} \quad \Delta\nu \, \Delta t = 1. \tag{8.75}$$

Функции с конечной шириной полосы представляют собой весьма специальный класс. Они не могут иметь разрывов, острых углов; они могут иметь только сглаженный, закругленный характер, как показано примерно на рис. 8.9.

9. Автокорреляция и спектр; формула Винера — Хинчина

Рассмотрим снова задачу А раздела 6 для вещественной периодической функции $f(t)$ с периодом τ (см. (8.52)), спектр которой не содержит частот выше ν_M. Разложим $f(t)$ в комплексный ряд Фурье (8.1) и (8.58):

$$f(t)=\sum_{-n_M}^{+n_M} C_n e^{in\omega_0 t}, \quad \omega_0=\frac{2\pi}{\tau}=2\pi\nu_0,$$

$$C_{-n}=C_n^*, \quad n_M=\frac{\nu_M}{\nu_0}=\nu_M\tau. \tag{8.76}$$

Определим автокорреляционную функцию $R(t_0)$ для функции $f(t)$:

$$R(t_0)=\overline{f(t)\,f(t+t_0)}=\frac{1}{\tau}\int_0^\tau f(t)\,f(t+t_0)\,dt. \tag{8.77}$$

Заменяя $f(t+t_0)$ его разложением Фурье (8.76), получаем

$$R(t_0)=\frac{1}{\tau}\sum_{-n_M}^{n_M} C_n e^{in\omega_0 t_0}\int_0^\tau f(t)\,e^{in\omega_0 t}dt=\sum_{-n_M}^{+n_M} C_n C_n^* e^{in\omega_0 t_0}, \tag{8.78}$$

так как интеграл равен как раз C_n^* в соответствии с (8.2).

Мы получили, таким образом, разложение Фурье для $R(t_0)$; заметим, что оно содержит только квадраты модулей

$$I_n=C_n C_n^*=|C_n|^2=|C_{-n}|^2=I_{-n} \tag{8.79}$$

коэффициентов Фурье функции $f(t)$. Фазы выпали в процессе усреднения согласно (8.77) и (8.79). Можно переписать (8.78) следующим образом (используя симметрию (8.79)):

$$R(t_0)=I_0+2\sum_{n=1}^{n_M} I_n\cos n\omega_0 t_0=R(-t_0). \tag{8.80}$$

Функция автокорреляции есть четная функция времени сдвига t_0. Она, как и исходная функция, периодична с тем же периодом τ и имеет то же частотное ограничение $|\nu|\leqslant\nu_M$. Равенство (8.80) представляет специальный случай формулы

Винера—Хинчина[1]), связывающей автокорреляцию и спектр. Исходная функция $f(t)$ имеет конечное число N степеней свободы (см. (8.57)):

$$N = 2n_M + 1,$$

среди которых мы можем различить:

$$\begin{array}{ll} 1 \text{ постоянную составляющую } & C_0 = \gamma_0, \\ n_M \text{ амплитуд гармоник} & \gamma_n, \\ n_M \text{ фаз гармоник} & \varphi_n, \end{array}$$

$$C_n = \gamma_n e^{i\varphi_n}, \quad C_{-n} = \gamma_n e^{-i\varphi_n}, \tag{8.81}$$

откуда

$$I_n = \gamma_n^2, \quad n = 0,\ 1,\ \ldots,\ n_M. \tag{8.82}$$

Новое разложение (8.80) содержит только $n_M + 1$ степеней свободы, так как мы потеряли n_M фаз. От разложений Фурье (8.78) и (8.80) автокорреляционной функции мы можем перейти к процедуре отсчетов, сходной с использованной в (8.59) и (8.63). Выберем N равноотстоящих точек отсчета на протяжении полного периода τ, простирающегося от $-\tau/2$ до $+\tau/2$:

$$t_{0m} = m\theta, \qquad \theta = \frac{\tau}{N}, \quad N = 2n_M + 1, \tag{8.83}$$

$$m = -n_M,\ -n_M + 1,\ \ldots,\ -1,\ 0,\ 1,\ \ldots,\ n_M - 1,\ n_M,$$

и рассмотрим отсчеты

$$R_m = R(m\theta) = R_{-m}. \tag{8.84}$$

Условие симметрии (8.80) оставляет нам $n_M + 1$ независимых отсчетов, соответствующих

$$m = 0,\ 1,\ \ldots,\ n_M, \tag{8.85}$$

что находится в соответствии с предыдущим рассуждением, согласно которому число степеней свободы равно $n_M + 1$. Функция $R(t_0)$ может быть восстановлена по отсчетам, как в (8.60):

$$R(t_0) = \sum_{m=1}^{n_M} R_m [g(t_0 - m\theta) + g(t_0 + m\theta)] + R_0 g(t_0), \tag{8.86}$$

где функция g определена равенством (8.62).

---

[1]) N. Wiener, Acta Math. **55**, 117 (1930); A. Khintchine, Math. Ann. **109**, 604 (1934); Y. W. Lee, T. P. Cheatham, J. B. Wiesner, MIT Report № 141, 1949.

Эта процедура использует только n_M+1 степеней свободы исходной функции из общего числа $2n_M+1$. Она с успехом применяется во всех задачах, в которых фазы φ_n неизвестны либо по причине их случайности (проблемы шума), либо вследствие того, что измерительный процесс не в состоянии обнаружить фазовые углы (например, исследование строения кристаллов рентгеновыми лучами). Вместо комплексного ряда Фурье (8.76) для $f(t)$ мы можем воспользоваться рядом Фурье в вещественной форме (8.54):

$$\left.\begin{aligned} &f(t)=\frac{1}{2}A_0+\sum_{n=1}^{n_M} A_n\cos(n\omega_0 t+\varphi_n), \\ &A_0=2\gamma_0,\quad A_n=2\gamma_n,\quad I_0=\frac{A_0^2}{4},\quad I_n=\frac{A_n^2}{4}, \end{aligned}\right\} \tag{8.87}$$

и согласно (8.80)

$$R(t_0)=\frac{1}{4}A_0^2+\frac{1}{2}\sum_{n=1}^{n_M} A_n^2\cos(n\omega_0 t_0). \tag{8.88}$$

Белый (тепловой) шум, например, дает постоянное значение A_n^2, не зависящее от n.

10. Линейные преобразования и фильтры

Мощный метод исследования функций основан на линейных преобразованиях[1]). Рассмотрим функцию и ее интеграл

---

[1]) H. Labrouste, I. Labrouste, Comptes Rend. **184**, 259 (1927); Ann. Inst. Phys. Globe Univ. Paris **7**, 190 (1929); **9**, 99 (1931); **11**, 93 (1933); Terrestrial Magnetism and Atm. Elec. **41**, 15, 105 (1936); H. Devé, Tables numeriques, Evreux, Paris, 1930; M. Levy, Comptes Rend. **198**, 2222 (1934); **199**, 1031 (1934); **200**, 646 (1935); Ph. D. thesis, Paris, 1951. H. Labrouste занимался специально случаем численного вычисления по дискретным значениям функции. Он пользовался общим преобразованием вида

$$F(t)=\sum_n \varphi_n f(t+n\tau),$$

где n — целое. M. Levy обобщил результаты для интегрального преобразования вида (8.91).

Фурье

$$f(t)=\int c(\nu)\, e^{2\pi i\nu t}\, d\nu \tag{8.89}$$

и функцию преобразования

$$\varphi(\theta)=\int \gamma(\nu)\, e^{2\pi i\nu\theta}\, d\nu, \tag{8.90}$$

достаточно быстро убывающую. Преобразованная функция определяется как

$$F(t)=\int t(t+\theta)\,\varphi(\theta)\, d\theta \tag{8.91}$$

и может быть представлена интегралом Фурье

$$F(t)=\int C(\nu)\, e^{i\,2\pi\nu t}\, d\nu, \tag{8.92}$$

где

$$C(\nu)=\int F e^{-i\,2\pi\nu t}\, dt=\int f(t')\, e^{-i\,2\pi\nu t'}\, dt' \int \varphi(\theta)\, e^{+i\,2\pi\nu\theta}\, d\theta.$$

(Здесь использовано (8.91) и положено $t'=t+\theta$.) Интегралы равны соответственно $c(\nu)$ и $\gamma^*(\nu)$, так что

$$C(\nu)=c(\nu)\,\gamma^*(\nu). \tag{8.93}$$

Таким образом, применение функции преобразования φ эквивалентно фильтрации при помощи фильтра с передаточной функцией $\gamma(\nu)$. Лабруст и Леви рассмотрели ряд функций преобразования, позволяющих выделить синусоидальные составляющие из сложных кривых, полученных путем эксперимента. Различные импульсы (8.25)—(8.28), например, дают передаточные функции, соответствующие их спектрам, и могут применяться для преобразования. Функция (8.91) есть просто взаимная корреляция между функциями f и φ. Когда φ совпадает с f, мы получаем автокорреляцию функции f, и равенства (8.91) и (8.93) сходны с (8.77) и (8.79).

Чтобы показать связь между результатами различных разделов этой главы, рассмотрим пример. Предположим, что мы рассматриваем экспериментальную кривую $f(t)$, относительно которой известно, что частоты выше некоторой ν_M записаны не очень надежно. Устраним, прежде всего, эти частоты, что можно сделать при помощи преобразования (8.91) с функцией преобразования (8.26); нужно вычислить

$$F(t) = 2\nu_M \int f(t+\theta) \frac{\sin(2\pi\nu_M\theta)}{2\pi\nu_M\theta} d\theta. \tag{8.94}$$

Функция $\varphi(\theta)$ резко ограничена на частоте ν_M; ее спектр

$$\gamma(\nu) = \begin{cases} 1 & \text{при } |\nu| \leqslant \nu_M, \\ 0 & \text{при } |\nu| > \nu_M. \end{cases}$$

Преобразование представляет идеальный фильтр нижних частот и подавляет все высокие частоты. Оно сглаживает экспериментальную кривую и устраняет острые углы и разрывы, которые могут быть сомнительными. Новая функция $F(t)$ имеет отчетливую границу при ν_M.

Если наблюдение велось за время τ, функция содержит только

$$N = 2\nu_M \tau$$

степеней свободы (см. (8.51)). Мы рассматриваем ее как функцию одиночного сообщения (раздел 6, В) и применяем метод отсчетов Шеннона (раздел 7) с интервалом отсчетов (8.68):

$$\theta_{\lim} = \frac{1}{2\nu_M}.$$

Операция отсчитывания заменяет непрерывную кривую последовательностью импульсов с отсчитанными амплитудами $F_M = F(m\theta_{\lim})$. Последовательность импульсов имеет теперь бесконечно протяженный спектр, как указано в разделе 1 (см. (8.9), (8.11) и рис. 8.1). Спектр состоит из исходного спектра ($|\nu| \leqslant \nu_M$), повторяемого периодически вдоль шкалы частот с периодом $2\nu_M$.

Отсчеты могут передаваться по системе связи одним из методов современной импульсной техники. Эти отсчеты воспроизводятся на приемной станции, и задача состоит в восстановлении непрерывной кривой $F(t)$. Это можно сделать электрическим путем при помощи полосового фильтра, выделяющего одну из одинаковых частотных полос рис. 8.1. То же самое можно сделать и путем вычисления суммы (8.70):

$$F(t) = \sum_m F_m \frac{\sin \pi (2\nu_M t - m)}{\pi (2\nu_M t - m)}, \tag{8.94a}$$

что снова соответствует математическому описанию действия идеального фильтра нижних частот. Для вычисления функции в момент t нам нужны в принципе все отсчеты до и после t. Если мы хотим восстановить $F(t)$ с точностью до 10^{-4}, мы должны располагать по меньшей мере 10^4 членами до и после момента t, так как члены разложения убывают очень медленно, всего лишь как m^{-1} или $(t'-t)^{-1}$. Однако вычисление вышеприведенной суммы является, по-видимому, наилучшим путем к восстановлению кривой, если требуется высокая точность.

11. Анализ Фурье и метод отсчетов в трех измерениях

Методы, рассмотренные в предыдущих разделах, могут быть распространены с одного на два или на три измерения. Исследуем кратко эти проблемы и покажем связь с рентгеноструктурным анализом кристаллов или с дифракцией нейтронов.

Периодичность в пространстве проявляется в кристаллической решетке и может быть характеризована тремя переносами $\boldsymbol{d}_1, \boldsymbol{d}_2, \boldsymbol{d}_3$. Функция $F(\boldsymbol{r})$, где $\boldsymbol{r}$ — вектор (x, y, z) периодична с этими тремя периодами, если она обладает следующим свойством:

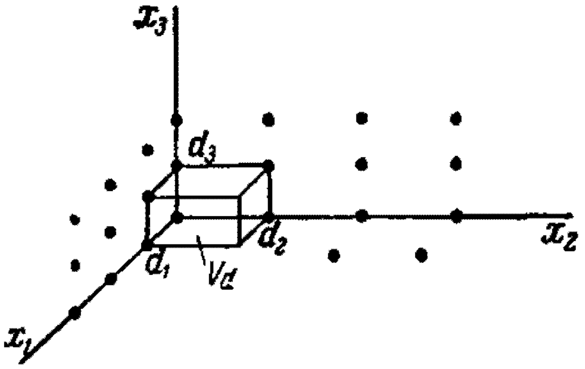

Рис. 8.10. Прямоугольная решетка, определяемая ортогональными векторами $\boldsymbol{d}_1$, $\boldsymbol{d}_2$, $\boldsymbol{d}_3$.

$$F(\boldsymbol{r}+\rho_1\boldsymbol{d}_1+\rho_2\boldsymbol{d}_2+\rho_3\boldsymbol{d}_3)=F(\boldsymbol{r}), \quad (8.95)$$

где ρ_1, ρ_2, ρ_3 — положительные или отрицательные целые. Общий случай косоугольной системы векторов $\boldsymbol{d}_1, \boldsymbol{d}_2, \boldsymbol{d}_3$ вызывает некоторые затруднения, которые отпадают в случае *ортогональных векторов*. Мы ограничимся этим более простым случаем и выберем прямоугольные координаты x_1, x_2, x_3 в направлениях $\boldsymbol{d}_1$, $\boldsymbol{d}_2$, $\boldsymbol{d}_3$ (рис. 8.10). Составляющие векторов таковы:

$$\left.\begin{array}{ll} \boldsymbol{r}: & (x_1 \;\; x_2 \;\; x_3), \\ \boldsymbol{d}_1: & (d_1, \;\; 0 \;\; 0), \\ \boldsymbol{d}_2: & (0 \;\; d_2 \;\; 0), \\ \boldsymbol{d}_3: & (0 \;\; 0 \;\; d_3), \end{array}\right\} \quad (8.96)$$

и периодичность (8.95) запишется как

$$F(x_1 + \rho_1 d_1,\ x_2 + \rho_2 d_2,\ x_3 + \rho_3 d_3) = F(x_1,\ x_2,\ x_3). \quad (8.96a)$$

Три вектора $\boldsymbol{d}_1$, $\boldsymbol{d}_2$, $\boldsymbol{d}_3$ определяют основную ячейку решетки с объемом

$$V_d = d_1 d_2 d_3.$$

Периодическая функция F может быть разложена в тройной ряд Фурье:

$$F = \sum_{h_1 h_2 h_3} C_{h_1 h_2 h_3} e^{2\pi i (h_1 b_1 x_1 + h_2 b_2 x_2 + h_3 b_3 x_3)}, \quad (8.97)$$

где

$$b_1 = \frac{1}{d_1},\quad b_2 = \frac{1}{d_2},\quad b_3 = \frac{1}{d_3};\quad h_1 h_2 h_3 \text{ — целые.}$$

Три обратных вектора $\boldsymbol{b}_1, \boldsymbol{b}_2, \boldsymbol{b}_3$ также ортогональны и определяют ячейку обратной решетки (рис. 8.11):

$$V_b = b_1 b_2 b_3 = \frac{1}{V_d}.$$

Обратная решетка может быть определена и для косоугольных структур; ее исследование в этом случае приводит к сходным формулам. Коэффициенты Фурье $C_{h_1 h_2 h_3}$ вычисляются с помощью следующего интеграла:

$$C_{h_1 h_2 h_3} = \frac{1}{V_d} \int_0^{d_1}\int_0^{d_2}\int_0^{d_3} F(x_1,\ x_2,\ x_3)\, e^{-2\pi i (h_1 b_1 x_1 + h_2 b_2 x_2 + h_3 b_3 x_3)} \times$$
$$\times\, dx_1 dx_2 dx_3. \quad (8.98)$$

Для вещественной функции F

$$C_{-h_1,\,-h_2,\,-h_3} = C^*_{h_1,\,h_2,\,h_3}.$$

Эти формулы являются прямым обобщением (8.1), (8.2) и (8.3). Соответствие обозначений таково:

период τ:	$d_1,\ d_2,\ d_3$,
основная частота ν_0:	$b_1,\ b_2,\ b_3$,
номер гармоники n:	$h_1,\ h_2,\ h_3$.

В трехмерном спектре могут встретиться различного рода ограничения.

А. *Три отдельных предела в трех направлениях*

$$|h_1 b_1| \leqslant \rho_1, \quad |h_2 b_2| \leqslant \rho_2, \quad |h_3 b_3| \leqslant \rho_3, \tag{8.99}$$

откуда (см. рис. 8.11)

$$|h_1| \leqslant \frac{\rho_1}{b_1} = \rho_1 d_1, \quad |h_2| \leqslant \rho_2 d_2, \quad |h_3| \leqslant \rho_3 d_3.$$

Величины ρ_1 ρ_2 ρ_3 представляют наивысшие частоты в направлениях 1, 2, 3. Задача представляет собой попросту наложение трех отдельных одномерных задач типа А (см. раздел 6

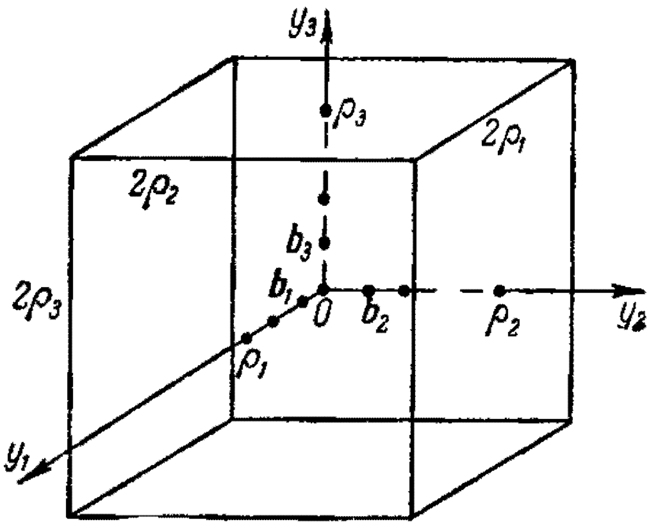

Рис. 8.11. Обратная решетка определяется векторами $\boldsymbol{b}_1$, $\boldsymbol{b}_2$, $\boldsymbol{b}_3$.

Точки решетки лежат внутри показанного на рисунке параллелепипеда, если ρ_1, ρ_2, ρ_3 — наивысшие частоты в трех измерениях.

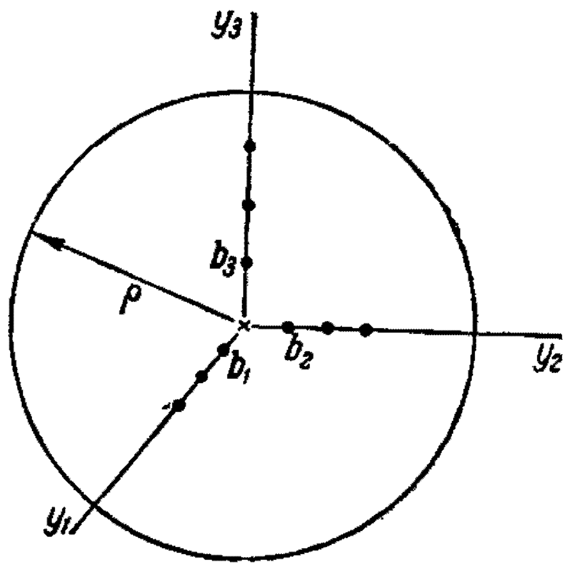

Рис. 8.12. Обратная решетка определена векторами $\boldsymbol{b}_1$, $\boldsymbol{b}_2$, $\boldsymbol{b}_3$.

Наибольшая частота ρ не зависит от направления, и точки решетки вне сферы радиуса ρ исключены.

этой главы). В этом примере имеем такое соответствие обозначений:

наивысшая частота ν_M ... $\rho_1 \rho_2 \rho_3$.

В. Рассмотрим *единый предел частоты* ρ при условии

$$h_1^2 b_1^2 + h_2^2 b_2^2 + h_3^2 b_3^2 \leqslant \rho^2 \tag{8.100}$$

(рис. 8.12). Это условие соответствует кратчайшей длине волны $1/\rho$. Обсудим подробнее его физический смысл. Каждая составляющая h_1 h_2 h_3 в разложении (8.97) выражает синусоидальное распределение в пространстве. Все точки, лежащие в плоскости

$$h_1b_1x_1 + h_2b_2x_2 + h_3b_3x_3 = 0,$$

имеют ту же фазу, что и в начале. Другие параллельные плоскости с одинаковой фазой находятся из соотношения

$$h_1b_1x_1 + h_2b_2x_2 + h_3b_3x_3 = m,$$

где m — целое. Длина волны Λ есть расстояние между двумя соседними плоскостями $\Delta m = \pm 1$, и легко найти:

$$\frac{1}{\Lambda^2} = (h_1b_1)^2 + (h_2b_2)^2 + (h_3b_3)^2 \leqslant \rho^2, \tag{8.100a}$$

где ρ определено (8.100). Большие значения h соответствуют коротким волнам. Вектор $(h_1b_1,\ h_2b_2,\ h_3b_3)$ ортогонален плоскостям волн, и его длина есть $1/\Lambda$.

Функции с ограниченным спектром (условие A или B) обладают только конечным числом степеней свободы. Условие A является более простым, так как оно непосредственно обобщает проблему A раздела 6. Аналогия с (8. 57) такова:

$$\left.\begin{aligned} N_1 &= 2\rho_1 d_1 + 1 \text{ степеней свободы вдоль } x_1, \\ N_2 &= 2\rho_2 d_2 + 1 \text{ степеней свободы вдоль } x_2, \\ N_3 &= 2\rho_3 d_3 + 1 \text{ степеней свободы вдоль } x_3, \end{aligned}\right\} \tag{8.101}$$

откуда общее число степеней свободы

$$N = N_1N_2N_3 \approx 8V_d\rho_1\rho_2\rho_3. \tag{8.102}$$

Последнее приближение справедливо, если $\rho_1\rho_2\rho_3$ много больше, чем $b_1b_2b_3$:

$$\frac{\rho_1}{b_1} = \rho_1 d_1 \gg 1, \quad \rho_2 d_2 \gg 1, \quad \rho_3 d_3 \gg 1.$$

Процедура отсчитывания, описанная в разделе 6 (см. (8.59)—(8.64)) непосредственно применима при прямоугольной решетке с N точками отсчета внутри ячейки $d_1d_2d_3$:

$$\left.\begin{aligned} &x_{1s} = m_1\xi_1, \\ &x_{2s} = m_2\xi_2, \quad \xi_i = \frac{d_i}{N_i} = \frac{1}{2\rho_i + b_i} \approx \frac{1}{2\rho_i}, \\ &x_{3s} = m_3\xi_3, \quad 0 \leqslant m_i < N_i, \quad i = 1,\ 2,\ 3. \end{aligned}\right\} \tag{8.103}$$

Здесь имеется соответствие:

$$\text{интервал отсчетов } \theta, \ldots, \xi_1, \xi_2, \xi_3.$$

Исходная функция легко восстанавливается по отсчетам при помощи тройной суммы

$$F = \sum_{m_1 m_2 m_3} F_{m_1 m_2 m_3} g_1(x_1 - m_1\xi_1) g_2(x_2 - m_2\xi_2) \times \\ \times g_3(x_3 - m_3\xi_3), \tag{8.104}$$

где

$$F_{m_1 m_2 m_3} = F(m_1\xi_1, \ m_2\xi_2, \ m_3\xi_3).$$

Функция g была определена равенством (8.62):

$$g_i(x_i) = \frac{\sin(N_i \pi b_i x_i)}{N_i \sin(\pi b_i x_i)}, \quad i = 1, 2, 3, \tag{8.105}$$

если принять во внимание соответствие обозначений. Задача, таким образом, полностью решена, и метод отсчетов применяется без всяких затруднений.

Вторая задача, B, с единственным пределом ρ (см. (8.100)) является более сложной. Мы легко найдем конечное число степеней свободы, но метод отсчетов может быть лишь намечен и остается предметом обсуждения.

Составляющие Фурье h_1 h_2 h_3 в разложении (8.97) могут быть представлены наглядно, как точки на прямоугольной решетке:

$$y_1 = h_1 b_1, \quad y_2 = h_2 b_2, \quad y_3 = h_3 b_3, \tag{8.106}$$

где h_1, h_2, h_3 — положительные или отрицательные целые. Эта решетка известна как *обратная решетка* (reciprocal lattice). Каждая точка (h_1, h_2, h_3) этой обратной решетки соответствует одной составляющей Фурье. Основная ячейка этой решетки имеет объем

$$V_b = b_1 b_2 b_3,$$

откуда получаем:

$$\frac{1}{V_b} = V_d \tag{8.107}$$

точек обратной решетки на единицу объема. Предел (8.100) соответствует сферической области в обратной решетке. Она имеет объем $\left(\frac{4}{3}\right)\pi\rho^3$ и содержит

$$N \approx V_d \frac{4}{3} \pi\rho^3 \tag{8.108}$$

точек в соответствии с (8.107). Это выражение дает число составляющих Фурье, а следовательно, и число независимых степеней свободы для нашей функции. Сходство между (8.102) и (8.108) очевидно. Первое равенство соответствует прямоугольной области в обратной решетке, второе — сферической.

Мы могли бы построить процедуру отсчитывания с использованием N точек отсчета в элементарной ячейке $d_1d_2d_3$ прямой решетки, но эта проблема не была пока должным образом обсуждена. Прямоугольная (или кубическая) решетка точек отсчета не подходит, так как не обладает требуемой сферической симметрией, и соответствует другой задаче (см. (8.99)). Можно представить себе систему точек решетки, размещенных в центрах плотно уложенных сфер. Это приводит к гранецентрированной кубической решетке [1]), так как другие гексагональные структуры обладают меньшей симметрией [2]).

Рассмотрим систему точек отсчета, расположенных в гранецентрированной кубической решетке с ребром δ. Куб δ^3 содержит четыре точки отсчета, и, следовательно, объем u на точку отсчета равен

$$u = \frac{1}{4} \delta^3. \tag{8.109}$$

В элементарной ячейке V_d кристаллической решетки имеем:

$$N' = \frac{V_d}{u} = \frac{4V_d}{\delta^3} \tag{8.110}$$

---

[1]) См., например, Л. Ф. Тот, Расположения на плоскости, на сфере и в пространстве, Физматгиз, М., 1958 (стр. 266). (*Прим. перев.*)

[2]) L. Brillouin, Wave Propagation in Periodic Structures, 1-st ed. McGraw-Hill, NY, 1946, 2-nd ed., Dover Publications NY, 1952, см., например, главу 1, стр. 7.

точек отсчета. Это число точек отсчета должно равняться числу N степеней свободы (см. (8.108)), откуда находим:

$$N' = N, \quad \frac{4}{3}\pi\rho^3 = \frac{4}{\delta^3}, \quad \rho\delta = \left(\frac{3}{\pi}\right)^{1/3}. \tag{8.111}$$

Расстояние между плоскостями точек отсчета, параллельными осям x_1, x_2, x_3, равно $\delta_1 = \delta/2$; расстояние между плоскостями, перпендикулярными к главным диагоналям, равно $\delta_2 = \delta/\sqrt{3}$. Сравнение этих расстояний с кратчайшей длиной волны $\lambda = \rho^{-1}$ дает:

$$2\rho\delta_1 = \left(\frac{3}{\pi}\right)^{1/3} \approx 0{,}986, \quad 2\rho\delta_2 = \left(\frac{3}{\pi}\right)^{1/3}\frac{2}{\sqrt{3}} \approx 1{,}14 \tag{8.112}$$

вместо условия

$$2\rho\xi = 1$$

в трехмерной задаче (см. (8.103)).

Для последовательного доведения до конца метода отсчетов нужно рассмотреть импульсную функцию $g(r)$, соответствующую данной задаче. Импульсная функция должна содержать только волны длиннее $1/\delta$. Остается выяснить, совместимо ли в трех (или двух) измерениях это условие с другим предположением относительно (8.60), а именно, что импульсная функция $g(r)$ равна единице при $r = 0$ и нулю во всех остальных точках отсчета. Если это второе условие не может быть выполнено, то применение метода отсчетов существенно усложнится. Единственное практическое решение состояло бы в применении процедуры, сходной с описанной в конце раздела 10, а именно, в произвольном введении границы типа А (см. (8.99)) путем тройного интегрирования вроде (8.94) с последующим использованием прямоугольной решетки точек отсчета (8.103) и (8.104). Это даст по крайней мере последовательный и практичный метод.

12. Исследование кристаллов рентгеновыми лучами

Вопросы, обсуждавшиеся в предыдущих разделах, тесно связаны с анализом кристаллов при помощи рентгеновых лучей. Кристалл представляет собой трехмерную периодическую структуру с основной ячейкой $\boldsymbol{d}_1$, $\boldsymbol{d}_2$, $\boldsymbol{d}_3$, и периодическая функция в этой решетке выражает плотность электронов. Интенсивность рентгеновых лучей, отраженных в различных направлениях, определяется квадратами модулей $|C_{h_1h_2h_3}|^2$ коэффициентов Фурье в тройном разложении (8.97), но соответствующие фазы не могут быть найдены по экспериментальным данным [1]). Более того, существует ограничение числа коэффициентов, которые могут быть получены из экспериментов с рентгеновыми лучами данной длины волны λ. Синусоидальное распределение электронных плотностей с длиной волны Λ отражает рентгеновы лучи, падающие под углом φ, если выполнено условие Брэгга [2]):

$$\lambda = 2\Lambda \cos\varphi. \tag{8.113}$$

Если поворачивать кристалл в пучке рентгеновых лучей, то можно обнаружить все значения Λ, удовлетворяющие условию

$$\Lambda \geqslant \frac{1}{2}\lambda = \frac{1}{\rho}. \tag{8.114}$$

В методе Дебая—Шерера получаются сходные ограничения. Мы приходим к задаче B раздела 11. Функция F при этих условиях имеет не более чем N степеней свободы (см. (8.108)), и мы получаем только $N/2$ данных из опытов с дифракцией рентгеновых лучей. Паттерсон [3]) предложил рассматривать функцию

$$P(u_1, u_2, u_3) = \frac{1}{V_d}\int_0^{d_1}\int_0^{d_2}\int_0^{d_3} F(x_1+u_1, x_2+u_2, x_3+u_3)\times$$
$$\times F(x_1x_2x_3)\,dx_1dx_2dx_3. \tag{8.115}$$

Это — попросту автокорреляционная функция (8.77) в трех

---

[1]) См. предыдущую ссылку, глав 6 и 7; R. W. James, The Optical Principles of the Diffraction of X-Rays, гл. 7, G. Bell, London, 1950.

[2]) Называемое также условием Брэгга—Вульфа, так как оно было независимо найдено Ю. В. Вульфом. (*Прим. перев.*)

[3]) A. L. Patterson, Phys. Rev. **46**, 372 (1934); Z. Krist. **90**, 517 (1935); D. Harker, J. Chem. Phys. **4**, 381 (1936); R. W. James, The Optical Principles of the Diffraction of *X*-Rays, G. Bell, London, 1950.

измерениях. Заменяя $F(x_1+u_1, \ldots)$ в (8.115) ее разложением Фурье (8.97), получаем:

$$P(u_1, u_2, u_3) = \sum_{h_1h_2h_3} C_{h_1h_2h_3} e^{2\pi i \Sigma h_i b_i u_i} \frac{1}{V_d} \int\limits_0^{d_1} \int\limits_0^{d_2} \int\limits_0^{d_3} F(x_1x_2x_3) \times$$

$$\times e^{2\pi i \Sigma h_i b_i x_i} dx_1 dx_2 dx_3 =$$

$$= \sum_{h_1h_2h_3} C_{h_1h_2h_3} C^*_{h_1h_2h_3} e^{2\pi i (h_1b_1u_1 + h_2b_2u_2 + h_3b_3u_3)}. \quad (8.116)$$

Эта формула является прямым обобщением (8.78) и содержит только интенсивности CC^*, но не фазы. Метод Паттерсона, предложенный независимо, представляет собой непосредственное развитие автокорреляционной функции Винера. Соотношения симметрии, обсужденные в разделе 9, применимы и здесь, и функция Паттерсона содержит только $\frac{1}{2}N$ независимых параметров. Для представления этой функции при помощи отсчетов необходимо иметь в силу симметрии только $\frac{1}{2}N$ отсчетов на протяжении половины элементарной кристаллической ячейки.

Другой мощный метод исследования предложен Пепинским[1]) под названием *конволюции*. Практически он соответствует преобразованиям, обсужденным в разделе 10, обобщенным для трех измерений. Этот метод оказался исключительно полезным при исследовании сложных структур. Мы рекомендовали в конце раздела 10 применение импульсов (8.25)—(8.28) в качестве ядер линейных преобразований. Пепинский называет функцию (8.26) *ядром Дирихле* и обратную функцию (8.27) — *ядром Фейера* с принятием $C(t)$ для функции и $f(\nu)$ для спектра (f и C заменяют друг друга). Пепинским предложены и другие ядра и обсуждены многие важные приложения.

Итак, исследование кристаллов рентгеновыми лучами представляется в виде трехмерной задачи, сходной с анализом сигнала в одном измерении.

[1]) R. Pepinsky, Report to the Office of Naval Research, Feb. 1, 1952.

Приложение

Неравенство Шварца

Для доказательства неравенства (8.49) рассмотрим интеграл

$$I = \int |F + cG|^2 dt = \int (F^* + cG^*)(F + cG)\, dt > 0, \qquad (8A.1)$$

где c — вещественная постоянная. Раскрывая скобки, получаем:

$$I = c^2 \int G^*G\, dt + c \int (F^*G + FG^*)\, dt + \int F^*F\, dt.$$

Это есть соотношение относительно c:

$$Ac^2 + Bc + C > 0. \qquad (8A.2)$$

Вещественного корня не должно быть; поэтому

$$B^2 - 4AC \leqslant 0,$$

или

$$4AC \geqslant B^2, \qquad (8A.3)$$

а это и есть неравенство Шварца.

ГЛАВА 9

ОСНОВЫ ТЕРМОДИНАМИКИ

1. Введение

В предыдущих главах обсуждался вопрос о том, как математически определить информацию и как измерить ее количественно. Ограниченность такого определения была подчеркнута в первой главе; последующие главы были посвящены развитию и применениям общих определений.

Была обнаружена тесная связь между теорией информации и термодинамикой; информация представляется связанной с энтропией. Обсуждение этого нового аспекта теории мы начнем с краткого обзора основ термодинамики.

2. Два начала термодинамики; энтропия и негэнтропия

А. Сохранение энергии.

В. Принцип Карно. По Кельвину этот принцип означает *деградацию энергии.*

Энергия высокого качества (high grade) — это механическая и электрическая энергия.

Энергия среднего качества — это химическая энергия.

Энергия низкого качества — это тепло.

Общее количество энергии остается неизменным в замкнутой изолированной системе. Преобразования или химические реакции внутри системы не изменяют качества энергии при обратимых преобразованиях. Необратимые преобразования понижают качество.

Качество энергии может быть точно определено через *отрицательную энтропию* ($-S$). Согласно второму началу энтропия всегда возрастает, или по крайней мере остается неизменной, так что огрицательная энтропия всегда убывает.

Определение энтропии

Если система получает количество тепла Δq, то это соответствует увеличению энтропии системы на

$$\Delta S = \frac{\Delta q}{T}, \quad \Delta q = T\Delta S, \tag{9.1}$$

T — абсолютная температура (по Кельвину). В градусах Цельсия

$$T\,^\circ\mathrm{K} = t\,^\circ\mathrm{C} + 273{,}15.$$

Несколько типичных случаев послужат примерами. Рассмотрим две соприкасающиеся системы A и B, могущие обмениваться работой и теплом. Будем считать, что система AB изолирована от своего окружения. Некоторое преобразование приводит к следующим условиям:

Система	Температура	Совершаемая работа (work output)	Получаемое тепло (heat input)	Приращение энтропии (entropy input)
A	T_A	W_A	q_A	$\Delta S_A = \frac{q_A}{T_A}$
B	T_B	W_B	q_B	$\Delta S_B = \frac{q_B}{T_B}$

$$\left.\begin{array}{l}\text{Первое начало: } W_A - q_A + W_B - q_B = 0,\\ \text{Второе начало: } \Delta S_A + \Delta S_B \geqslant 0.\end{array}\right\} \tag{9.2}$$

Полная энергия системы AB остается неизменной.

Примеры необратимых преобразований

Поток тепла от горячего тела к холодному: $T_A > T_B$.
Работа не совершается: $W_A = 0$; $W_B = 0$.
Переданное тепло: $q_A = -q_B$, $q_B > 0$

$$\Delta S_A + \Delta S_B = q_B\left(\frac{1}{T_B} - \frac{1}{T_A}\right) > 0. \tag{9.3}$$

Общая энтропия возросла.

Трение, вязкое демпфирование: $T_A = T_B = T$ — однородная температура. Работа совершается телом A, тепло выделяется в B:

$$W_A > 0, \quad q_A = 0, \quad W_B = 0, \quad q_B > 0,$$
$$W_A - q_B = 0,$$
$$\Delta S_A + \Delta S_B = \frac{q_B}{T} = \frac{W_A}{T} > 0. \tag{9.4}$$

Общая энтропия возросла.

Полная энергия, содержащаяся в системе, может быть определена, если мы знаем построение системы из ее составных частей (энергия которых предполагается известной) и если мы учтем совершенную при построении работу и выделившееся тепло. Полная энергия U есть функция всех физических и химических параметров, определяющих систему.

Если система может быть построена путем последовательных *обратимых преобразований*, то *полная энтропия* S системы определяется как

$$S = S_1 + S_2 + \ldots + S_n, \tag{9.5}$$

т. е. как сумма энтропий, связанных с каждым обратимым шагом преобразования.

Негэнтропия[1] ($N = -S$) представляет *качество* энергии, и должна всегда убывать. В этом состоит смысл кельвинова принципа *деградации энергии*, и неудивительно, что важность негэнтропии была признана близким другом Кельвина, который писал[2]:

«Весьма желательно иметь слово, выражающее *способность* тепла, запасенного в некотором складе, произвести работу; термин для возможности, утеря которой называется *рассеянием*. К сожалению, превосходное слово *энтропия*, которое Клаузиус ввел в связи с этим, было им применено в отрицательном смысле по отношению к понятию, которое нам естественно хочется выразить. Но мы лишь смутили бы учащегося, если бы попытались выдумать новый термин для этой цели. Вместе с тем надобность в такого рода тер-

---

[1]) Сокращение для *negative entropy* — отрицательной энтропии. (*Прим. перев.*)

[2]) P. G. Tait, Sketch of Thermodynamics, p. 100, Edmonston Douglas, Edinburgh, 1868.

мине будет очевидна из прекрасных примеров, следующих ниже».

Изолированная система обладает *негэнтропией*, если она обнаруживает возможность совершения механической или электрической работы: если система не имеет однородной температуры, а состоит из различных частей с различными температурами, она содержит некоторое количество негэнтропии. Негэнтропия может быть использована для получения некоторой механической работы, совершаемой системой, или она может быть просто рассеяна и потеряна вследствие теплопроводности.

Разность давлений в различных частях системы является другим случаем негэнтропии. Разность электрических потенциалов представляет еще один пример. Сосуд со сжатым воздухом в комнате при атмосферном давлении, откачанный сосуд в такой же комнате, заряженная батарея, любое устройство, которое может производить энергию высокого качества (механическую работу) или деградировать в результате какого-либо необратимого процесса (теплопроводность, электрическое сопротивление, трение, вязкость), есть источник негэнтропии.

Важность рассмотрения отрицательной энтропии была подчеркнута Шредингером в его интереснейшей книге «Что такое жизнь» [1]). Если живой организм нуждается в пище, то лишь из-за негэнтропии, которую он может получить из пищи и которая необходима для восполнения потерь на совершенную механическую работу или вследствие процессов деградации в живом организме. Энергия, содержащаяся в пище, не имеет существенного значения, так как она сохраняется и никогда не теряется; важную роль играет именно негэнтропия.

3. Невозможность вечного движения; тепловые машины

Машина, применяющая только *один* резервуар при данной температуре T, не может превратить тепло в работу, так как это было бы в противоречии с равенством (9.4) нашего примера и означало бы уменьшение энтропии. Машина может

---

[1]) E. Schrödinger, What is life?, Cambridge U. P., NY, 1945. [Э. Шредингер, Что такое жизнь с точки зрения физики, ИЛ, 1946.]

производить работу при наличии двух резервуаров A и B при двух различных температурах

$$T_A > T_B.$$

Действие машины состоит из последовательных циклов, включающих в себя:

A. Передачу количества тепла q_A от резервуара тепла A машине (получаемое тепло).

B. Производство машиной работы W.

C. Передачу резервуару B количества тепла q_B (при низкой температуре). Так как мы считаем *получаемое тепло* положительным, то здесь мы имеем отрицательное получаемое тепло, т. е. $q_B < 0$. По завершении полного цикла действия машина возвращается в исходное состояние. Это значит, что ее полная энергия имеет снова исходное значение:

$$W - q_A - q_B = 0. \tag{9.6}$$

Если действие обратимо, то и полная энтропия машины не должна измениться:

$$\Delta S_A + \Delta S_B = 0, \tag{9.7}$$

откуда

$$\frac{q_A}{T_A} + \frac{q_B}{T_B} = 0,$$

или

$$q_B = -\frac{T_B}{T_A} q_A, \tag{9.8}$$

q_B отрицательно, как пояснено выше.

Другими словами, некоторое количество энтропии

$$\Delta S = +\Delta S_A = -\Delta S_B \tag{9.9}$$

спадает от высокой температуры T_A до низкой температуры T_B. Количества тепла, получаемые при этих двух температурах, равны соответственно:

$$q_A = +T_A \Delta S, \quad q_B = -T_B \Delta S. \tag{9.10}$$

Общее количество тепла $q_A + q_B = (T_A - T_B)\Delta S > 0$ положительно. Машина получила больше тепла от источника высокой температуры, нежели возвратила при низкой температуре. Баланс энергии уравнивается совершенной машиной работой (см. (9.6)). *Тепловой коэффициент полезного действия* определяется как отношение совершенной работы к полученному теплу:

$$\frac{W}{q_A}=\frac{T_A-T_B}{T_A}<1. \tag{9.11}$$

Примеры

Тепловая машина: тепло течет от высокой температуры к низкой, совершается работа. Холодильник есть обращенная машина: здесь *получаемая работа* дает поток тепла от низкой температуры к высокой.

Различные качества энергии в этих примерах легко распознать. Работа всегда может быть полностью превращена в тепло (см. (9.4)). Тепло может быть снова превращено в работу лишь частично (см. 9.11)).

Если измерять количество тепла в единицах энергии, то q дано в эргах (система CGS), а не в калориях. Выражая температуру в градусах Кельвина, получим энтропию S в эргах на градус. Энтропию S можно сделать безразмерной величиной, если считать, что температура T имеет размерность энергии (см. главу 1, раздел 1). В этом случае коэффициент преобразования есть

$$k=1{,}38\cdot 10^{-16}.$$

Величина kT имеет размерность энергии, и k может рассматриваться как отвлеченное число.

Если рассматриваются только обратимые преобразования, то возможно определить полную энергию U некоторой физической системы и ее полную энтропию S. Обе эти величины являются функциями физических переменных, определяющих состояние системы.

В качестве примера применения термодинамических методов рассмотрим случай диэлектрика с диэлектрической проницаемостью $\varepsilon(T)$, зависящей от температуры. Например,

$$D=\varepsilon E,\quad \varepsilon(T)=\varepsilon_0+\frac{C}{T-\theta}. \tag{9.12}$$

где D — электрическое смещение, E — напряженность электрического поля, C — постоянная Кюри, θ — температура Кюри.

Теоретическое исследование обнаруживает зависимость теплоемкости от поляризации, а следовательно, от смещения D. Подробный анализ дает соотношение между этими различными свойствами: полная энергия U и энтропия S вещества выражаются формулами:

$$U(T, D) = U_0(T) + \frac{1}{2} D^2 \left[\frac{1}{\varepsilon} - T \frac{\partial}{\partial T} \left(\frac{1}{\varepsilon} \right) \right], \qquad (9.13)$$

$$S(T, D) = S_0(T) - \frac{1}{2} D^2 \frac{\partial}{\partial T} \left(\frac{1}{\varepsilon} \right), \qquad (9.14)$$

где U_0 и S_0 соответствуют отсутствию поляризации ($D = 0$).

Предполагается, что вещество твердо и несжимаемо, и электрострикцией пренебрегается. Аналогичные формулы применимы к проблемам магнетизма [1]), что имеет важные приложения. На этих соотношениях основан известный метод охлаждения размагничиванием Дебая—Джиóка (Debye—Giauque), который позволяет достичь температур, составляющих долю градуса абсолютной шкалы.

4. Статистическое толкование энтропии

Величина, названная *энтропией*, формально определена равенством (9.1), и мы показали на нескольких примерах, что она находится в обратном соответствии *качеству энергии*. Второе начало устанавливает, что энтропия должна всегда возрастать, тогда как энергия всегда теряет свое качество (деградация энергии Кельвина).

Ясное истолкование этих общих свойств дано Гиббсом, Максвеллом, Больцманом и завершено Планком. Основание этого объяснения найдено в атомной структуре материи и квантовых законах устойчивости атомов, молекул и кристаллов. Атомы так малы, что количество их в любом кусочке вещества, с которым мы экспериментируем, ошеломляюще велико. Грамм-атом вещества содержит около $6 \cdot 10^{23}$ атомов. Вместе с тем мы совершенно не в состоянии следить за движением отдельных атомов. Единственные свойства, кото-

---

[1]) См., например, L. Brillouin, H. P. Iskenderian, Elec. Commun. **25**, 300 (1948).

рые мы можем практически наблюдать, — это средние статистические величины.

Статистическая теория термодинамики, основанная на этих общих замечаниях, представляет собой одну из наиболее разработанных глав физики. Как будет показано, энтропия связана с вероятностью. Можно создать искусственно замкнутую изолированную систему с очень маловероятной структурой. Если предоставить эту систему самой себе, то она будет эволюционировать в сторону более вероятной структуры. Вероятность имеет естественную тенденцию к возрастанию, как и энтропия. Точное соотношение дается известной формулой Больцмана — Планка:

$$S = k \ln P, \tag{9.15}$$

где $k = 1{,}38 \cdot 10^{-16}$ эрг/градус — постоянная Больцмана, S — энтропия рассматриваемой системы, P — статистический вес (число *элементарных комплексий*[1])).

Рассмотрим, прежде всего, точный смысл числа P, чтобы пояснить связь с теорией вероятностей. Согласно квантовой теории атомы и молекулы могут существовать лишь в виде конечного числа различных структур. На атомном уровне нет непрерывности, есть только дискретные стабильные (или метастабильные) структуры, и система атомов внезапно перескакивает от одной системы к другой, поглощая или излучая при этом энергию. Каждая из этих дискретных конфигураций квантованной физической системы называется микросостоянием (*комплексия* Планка).

Рассмотрим теперь физическую систему: кусок кристалла, или газ в сосуде при определенных условиях в отношении объема, давления, полной энергии, температуры, химического строения. *Макроскопические переменные* — это те величины, которые мы можем измерить в лаборатории. Их недостаточно, чтобы полностью определить состояние системы. Имеется огромное число *микроскопических* переменных, которые невозможно в точности измерить: положения и скорости всех отдельных атомов, квантовые состояния этих атомов или

[1]) Вместо термина *комплексия* в русской литературе предпочитают термин *микросостояние*. Что касается числа микросостояний, то для этой величины имеется несколько наименований. Больцман называл ее Permutabilität; Планк пользовался термином *термодинамическая вероятность*. В нашей литературе принят термин *статистический вес*, которым мы дальше и будем пользоваться. (*Прим. перев.*)

молекулярных структур и т. д. Все эти неизвестные величины дают системе возможность принимать большое число разнообразных квантовых структур, или микросостояний. Общее число P этих микросостояний мы и ввели в (9.15) и связали с энтропией системы. Это число чрезвычайно велико, но, конечно, и (9.15) имеет определенный смысл.

В нашем рассуждении понадобится еще одно важное соотношение — известное соотношение между энергией и частотой

$$E_1 - E_2 = h\nu, \tag{9.16}$$

где E_1 — начальная энергия, E_2 — окончательная энергия системы, излучающей колебание с частотой ν, $h = 6{,}6 \cdot 10^{-27}$ CGS — постоянная Планка.

Излучение испускается, когда физическая система теряет энергию, и поглощается, когда система получает энергию. Так как уровни энергии дискретны, частоты излучения (или поглощения) образуют прерывистый спектр, типичный для рассматриваемых атомных или молекулярных систем.

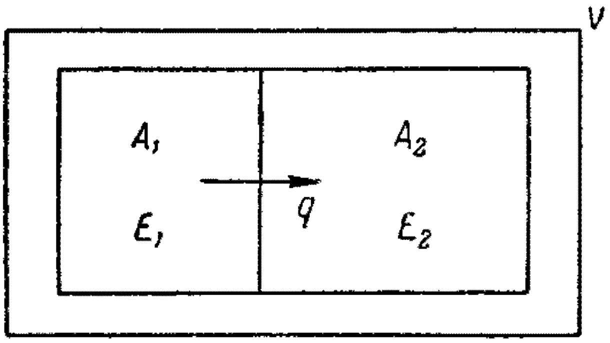

Рис. 9.1. Два тела A_1 и A_2 с энергиями E_1 и E_2 соответственно соприкасаются. Они могут обмениваться только теплом. Система A_1A_2 заключена в изолирующий контейнер V.

5. Примеры статистических рассуждений

Несколько примеров покажут, насколько наши статистические определения соответствуют экспериментальным фактам. Рассмотрим два тела, A_1 и A_2, соприкасающихся так, что они могут обмениваться только теплом. Система A_1A_2 помещена в изолирующий контейнер V, как показано на рис. 9.1. При этих условиях полная энергия

$$E_1 + E_2 = E \tag{9.17}$$

системы в целом остается постоянной, но мы должны рассмотреть возможность передачи q калорий от A_1 к A_2:

$$\Delta E_1 = -q, \quad \Delta E_2 = +q. \tag{9.18}$$

Найдем наиболее вероятное распределение энергии между A_1 и A_2. Для A_1 статистический вес P_1 есть функция E_1, для A_2 статистический вес P_2 есть функция E_2. Статистический вес сложной системы равен

$$P = P_1(E_1) P_2(E_2), \tag{9.19}$$

так как каждое микросостояние A_1 может сочетаться с любым из P_2 (E_2) микросостояний A_2. Наиболее вероятно такое распределение энергий, при котором P максимально. Если P имеет максимум, то он не будет меняться при передаче небольшого количества тепла от A_1 к A_2 (см. (9.18)). Положим, что P_1 и P_2 могут рассматриваться как непрерывные функции E_1 и E_2 соответственно и что эти функции дифференцируемы. Тогда

$$\Delta P = \frac{dP}{dq} q = \left(-\frac{dP_1}{dE_1} P_2 + P_1 \frac{dP_2}{dE_2}\right) q = 0. \tag{9.20}$$

Здесь приняты во внимание знаки в (9.18).

Предыдущее условие дает:

$$\frac{1}{P_1} \frac{dP_1}{dE_1} = \frac{1}{P_2} \frac{dP_2}{dE_2}, \tag{9.21}$$

или

$$\frac{d \ln P_1}{dE_1} = \frac{d \ln P_2}{dE_2}. \tag{9.22}$$

Применим теперь формулу Больцмана—Планка (9.15) для определения энтропий S_1 и S_2 тел A_1 и A_2 и получим:

$$\frac{1}{k} \frac{dS_1}{dE_1} = \frac{1}{k} \frac{dS_2}{dE_2}. \tag{9.23}$$

В нашем примере единственный вид энергии, которым обмениваются A_1 и A_2, — это тепло; поэтому в соответствии с (9.1)

$$\frac{dS_1}{dE_1} = \frac{dS_1}{dQ_1} = \frac{1}{T_1}. \tag{9.24}$$

Первый член (9.23) выражает величину, обратную абсолютной температуре тела A_1. Второй член в том же равенстве соответствует $1/T_2$, так что окончательно мы получаем:

$$\frac{1}{T_1} = \frac{1}{T_2}, \quad \text{или} \quad T_1 = T_2. \tag{9.25}$$

Наиболее вероятное распределение энергий между A_1 и A_2 получается при равенстве температур. Полученная система формул согласуется с общеизвестными экспериментальными результатами.

Решающим пунктом в нашем рассуждении является предположение о том, что $P_1(E_1)$ и $P_2(E_2)$ — непрерывные функции с регулярными производными. Это предположение справедливо только для больших материальных систем, содержащих огромное число атомов, когда энергии E_1 и E_2 очень велики по сравнению с квантовыми скачками энергии.

6. Флуктуации энергии; формула Гиббса

Рассмотрим такое очень большое тело, предположив, например, что A_1 — большой сосуд, наполненный водой и тщательно изолированный в тепловом отношении от своего окружения. Это соответствует физическому устройству, известному под названием *термостат*, т. е. аппарат, сохраняющий практически постоянную температуру. Температура T_1 системы действительно постоянна, если обмениваемые количества тепла q остаются очень малыми по сравнению с полной энергией E_1 термостата A_1.

Для такого термостата из (9.15) и (9.24) имеем:

$$\frac{d\ln P_1}{dE_1} = \frac{1}{kT_1}, \tag{9.26}$$

или путем прямого интегрирования

$$\ln P_1 = \frac{E_1}{kT_1} + C_1, \tag{9.27}$$

где C_1 — постоянная интегрирования, откуда

$$P_1 = P_0 e^{\frac{E_1}{kT_1}}, \tag{9.28}$$

где $P_0 = e^{C_1}$. Такое соотношение между статистическим весом P_1 и энергией E_1 типично для термостата.

Введем теперь в термостат A_1 небольшое тело A_2 и выясним состояние этого тела, поддерживаемого термостатом при постоянной температуре T_1. Наиболее вероятная энергия E_2 дана соотношениями (9.21), (9.24) и (9.25), но могут встретиться и другие значения энергии, хотя они менее вероятны.

Тело A_2 не будет сохранять все время наиболее вероятную энергию; будут наблюдаться *флуктуации энергии.*

Из (9.17) получаем:

$$E_1 = E - E_2. \tag{9.29}$$

Подставляя в (9.28), запишем:

$$P_1 = B_1 e^{-\frac{E_2}{kT_1}}, \quad \text{где} \quad B_1 = P_0^{\frac{E}{kT_1}}. \tag{9.30}$$

Статистический вес P системы в целом по-прежнему выражен (9.19), так что

$$P = P_2(E_2)\, B_1 e^{-\frac{E_2}{kT_1}}. \tag{9.31}$$

Это значит, что в термостате с *температурой* T_1 каждое из P_2 микросостояний A_2, соответствующих некоторой энергии E_2, имеет коэффициент *вероятности*

$$p(T) = B_1 e^{-\frac{E_2}{kT_1}}. \tag{9.32}$$

Последняя формула типична для определения, первоначально введенного Гиббсом. Действительно, вместо того чтобы начать, как это сделал Больцман, с формулы (9.15), для определения состояния тела A_2, поддерживаемого при температуре T_1, Гиббс воспользовался формулой вида (9.32). Можно показать, что оба подхода эквивалентны. Мы еще воспользуемся формулой Гиббса (9.32) при обсуждении многих проблем, в которых важную роль играют флуктуации.

Это рассуждение показывает применение статистического подхода к объяснению термодинамики. Тело при температуре T находится в состоянии постоянного движения. В случае газа молекулы движутся внутри сосуда беспорядочно и очень часто соударяются. Средняя кинетическая энергия молекул на степень свободы равна

$$\bar{E}_{\text{kin}} = \frac{1}{2} kT, \tag{9.33}$$

и это *тепловое движение* представляет энергию, содержащуюся в газе при температуре T.

В кристалле атомы занимают фиксированные положения в пространстве и образуют правильную решетку, которая и составляет кристалл. Тепловое движение проявляется в колебаниях атомов около их положений равновесия. До тех пор пока частота этих колебаний мала, кинетическая энергия колебания выражается равенством (9.33), и среднюю полную энергию каждого осциллятора легко найти, как сумму кинетической и потенциальной энергий:

$$\bar{E}_{\text{total}}=\bar{E}_{\text{kin}}+\bar{E}_{\text{pot}}, \quad \bar{E}_{\text{pot}}=\bar{E}_{\text{kin}}=\frac{1}{2}kT, \tag{9.34}$$

$$\bar{E}_{\text{total}}=kT \tag{9.35}$$

для низкочастотного осциллятора.

7. Квантованный осциллятор

Эти простые формулы уже не применимы к осциллятору высокой частоты, когда квант $h\nu$ уже не является малой величиной по сравнению с kT. В этом случае мы должны учитывать то обстоятельство, что в квантованном осцилляторе энергия может принимать только дискретные значения:

$$E=0,\ h\nu,\ 2h\nu,\ 3h\nu,\ \ldots,\ nh\nu,\ \ldots \tag{9.36}$$

В термостате T эти уровни энергии имеют вероятности, определяемые (9.32):

$$p_0=B,\quad p_1=Be^{-\frac{h\nu}{kT}},\quad p_2=Be^{-\frac{2h\nu}{kT}},\ \ldots,\ p_n=Be^{-\frac{nh\nu}{kT}}. \tag{9.37}$$

Коэффициент B определяется из условия, что сумма всех этих вероятностей должна равняться единице:

$$1=\sum_{i=0}^{\infty} p_i=B(1+e^{-x}+e^{-2x}+\ldots+e^{-nx}+\ldots)=$$

$$=\frac{B}{1-e^{-x}}, \tag{9.38}$$

где

$$x=\frac{h\nu}{kT}. \tag{9.39}$$

Отсюда

$$B=1-e^{-x}, \tag{9.40}$$

что дает:

$$p_0 = 1 - e^{-x},\ p_1 = e^{-x} - e^{-2x}, \ldots, p_n = e^{-nx} - e^{-(n+1)x}. \quad (9.41)$$

Из (9.41) легко получить очень простое выражение для суммы вероятностей энергий выше определенного уровня:

$$E \geqslant mh\nu, \quad n \geqslant m,$$

а именно:

$$\sum_{n=m}^{\infty} p_n = p_m + p_{m+1} + p_{m+2} + \ldots = e^{-mx}. \quad (9.42)$$

Найдя вероятности p_n дискретных уровней энергии, мы можем затем подсчитать среднюю энергию E квантованного резонатора

$$\overline{E} = \overline{n}h\nu,$$

где $\overline{n}$ — среднее число квантов — равно

$$\overline{n} = \sum_{n=0}^{\infty} np_n = B\left(0 + e^{-x} + 2e^{-2x} + \ldots + ne^{-nx} + \ldots\right) =$$

$$= B\frac{e^{-x}}{(1-e^{-x})^2}. \quad (9.43)$$

Сумму можно вычислить, если заметить, что она представляет собой производную по x от

$$-\sum_{n=0}^{\infty} e^{-nx} = -\frac{1}{1-e^{-x}}.$$

Подставляя значение B из (9.40), находим:

$$\overline{n} = \frac{1}{e^x - 1}, \quad (9.44)$$

где

$$x = \frac{h\nu}{kT},$$

так что

$$\overline{E} = \frac{h\nu}{e^{\frac{h\nu}{kT}} - 1}. \quad (9.45)$$

Это — известная формула Планка, являющаяся основой теории излучения *черного тела.* В задаче о квантованном резонаторе (см. (9.38) — (9.44)) мы можем отметить различие между наиболее вероятным состоянием, соответствующим наибольшему значению $p = p_0 = B$, $n = 0$, и средним числом квантов $\bar{n}$ (см. (9.44)).

Для очень малых значений x,

$$x \ll 1, \quad h\nu \ll kT, \quad \frac{1}{e^x - 1} \approx \frac{1}{x},$$

имеем:

$$\bar{E} = \frac{h\nu}{x} = kT. \tag{9.46}$$

Это есть случай низких частот, и результат согласуется с ранее полученным (см. (9.35)).

8. Флуктуации

Полученные формулы понадобятся нам при обсуждении вопроса о точности физических измерений. При любой данной температуре T результатом теплового движения является беспорядочное движение всех частей прибора, что делает невозможным измерение какой-либо величины с беспредельной точностью.

Тепловое движение есть причина броунова движения, шума в электрических цепях, шумового фона в акустике и т. д. Все эти явления будут рассматриваться в связи с применением теории информации к проблемам физики. Мы отметим очень интересную связь между информацией и энтропией и придем к систематическому обсуждению фундаментальных ограничений, встречаемых в физических экспериментах.

Тепловая энергия не влияла бы на точность измерений, если бы она оставалась постоянной. Измерения затрудняются всегда существующими флуктуациями, проявляющимися в том, что истинная энергия системы в данный момент не может быть предсказана и может сильно отличаться от среднего значения.

В предыдущем разделе мы отмечали различие между наивероятнейшим и средним значениями энергии. Исследуем флуктуации в квантованном осцилляторе. В некоторый момент времени система имеет

$$n = \bar{n} + m \tag{9.47}$$

квантов с флуктуацией m относительно среднего значения $\bar{n}$. Подсчитаем среднее значение n^2:

$$n^2 = (\bar{n} + m)^2 = (\bar{n})^2 + 2\bar{n}m + m^2. \tag{9.48}$$

При усреднении учтем, что $\overline{m} = 0$, так как $\bar{n}$ — среднее значение n, так что

$$\overline{n^2} = (\bar{n})^2 + \overline{m^2}. \tag{9.49}$$

Если $\overline{n^2}$ известно, то средний квадрат флуктуации $\overline{m^2}$ легко найти из (9.49). Так как вероятности всех квантованных состояний известны (см. (9.37)), то нетрудно подсчитать $\overline{n^2}$. Используя (9.43), запишем:

$$\overline{n^2} = \sum_{n=0}^{\infty} n^2 p_n = B \sum_{n=0}^{\infty} n^2 e^{-nx} = -B \frac{d}{dx} \sum_{n=0}^{\infty} n e^{-nx} =$$
$$= -B \frac{d}{dx}\left[\frac{e^{-x}}{(1-e^{-x})^2}\right]. \tag{9.50}$$

Выполняя дифференцирование и подставляя значение B из (9.40), получаем:

$$\overline{n^2} = B\frac{e^{-x}+e^{-2x}}{(1-e^{-x})^3} = \frac{e^{-x}+e^{-2x}}{(1-e^{-x})^2} = \frac{e^x+1}{(e^x-1)^2}. \tag{9.51}$$

Сравнение (9.49) с (9.51) с использованием (9.44) дает:

$$\overline{m^2} = \overline{n^2} - (\bar{n})^2 = \frac{e^x+1-1}{(e^x-1)^2} = (\bar{n})^2 + \bar{n}. \tag{9.52}$$

Этот знаменитый результат, полученный впервые Планком и Эйнштейном, связывает флуктуации со средними значениями.

Для низкочастотного осциллятора ($h\nu \ll kT$) среднее число квантов $\bar{n}$ очень велико, и мы имеем просто

$$\overline{m^2} = (\bar{n})^2. \tag{9.53}$$

Это же соотношение может быть выражено через флуктуацию энергии ε:

$$E = \bar{E} + \varepsilon, \quad \overline{\varepsilon^2} = \overline{m^2}(h\nu)^2 = (\bar{E})^2, \tag{9.54}$$

где $\bar{E} = \bar{n}h\nu$ — средняя энергия осциллятора.

При низких частотах

$$h\nu \ll kT, \quad (\overline{\varepsilon^2})^{1/2} = \overline{E} = kT, \tag{9.55}$$

корень квадратный из среднего квадрата флуктуации равен средней энергии, или, грубо говоря, флуктуации имеют тот же порядок величины, что и средняя энергия.

Для высоких частот ($h\nu \gg kT$) среднее значение $\overline{n}$ мало, $(\overline{n})^2$ еще меньше, так что

$$\overline{m^2} \approx \overline{n} \ll 1, \tag{9.56}$$

или

$$\overline{\varepsilon^2} \approx \overline{n}\,(h\nu)^2 = \overline{E}h\nu. \tag{9.57}$$

При высоких частотах флуктуации энергии в среднем много больше, так как $\overline{E}$ меньше, чем $h\nu$, а поэтому

$$\overline{E}h\nu \gg \overline{E^2}. \tag{9.58}$$

Флуктуации относительно больше при высоких частотах (см. (9.57)), чем при низких (см. 9.54)). Этот результат типичен для квантовых явлений.

Рассмотрим теперь систему из N гармонических низкочастотных осцилляторов. Полная средняя энергия $\overline{E}_{\text{total}}$ равна

$$\overline{E}_{\text{total}} = N\overline{E} = NkT. \tag{9.59}$$

Наблюдаются флуктуации $\varepsilon_{\text{total}}$

$$E_{\text{total}} = \overline{E}_{\text{total}} + \varepsilon_{\text{total}}, \tag{9.60}$$

которые независимы в каждом из N осцилляторов; поэтому

$$\overline{\varepsilon^2_{\text{total}}} = N\overline{\varepsilon^2} = N\,(kT)^2 \tag{9.61}$$

в соответствии с (9.54) и (9.55).

ГЛАВА 10

ТЕПЛОВОЕ ДВИЖЕНИЕ И БРОУНОВО ДВИЖЕНИЕ

1. Тепловое движение

Напомним одно из важных заключений предыдущей главы, а именно, что $\bar{E}_{\text{kin}}=\frac{1}{2}kT$ для каждой степени свободы при низкой частоте, и продолжим обсуждение некоторых практических приложений.

Броуново движение

Наблюдая мелкие частицы, погруженные в жидкость, и рассматривая их под микроскопом, мы видим, что они непрерывно движутся случайным образом. Первые исследователи полагали сначала, что эти частицы наделены жизнью; однако вскоре было установлено, что их движение полностью обусловлено случайной бомбардировкой со всех сторон молекулами жидкости. Это было впервые объяснено Эйнштейном (в 1905 г.) и проверено экспериментально Перреном. Более того, опыты Перрена позволили ему определить размеры атомов, поскольку он мог найти непосредственно число Авогадро — число атомов или молекул в одном моле вещества.

Предельная точность измерений

Рассмотрим проблему измерения тока при помощи амперметра. Стрелка прибора вращается на оси и имеет, следовательно, одну степень свободы, как показано на рис. 10.1. Таким образом, кинетическая энергия случайных колебаний стрелки равна $\bar{E}_{\text{kin}}=kT/2$, и величина этих колебаний сразу может быть вычислена. Мы можем уменьшить размах колебаний, охлаждая прибор; это окажет лишь частичное действие, так как для получения устойчивого состояния слабого теплового движения нужно охлаждать также всю схему.

Проблема шума (например, в сопротивлениях) усиленно изучалась; Найквист был одним из первых исследователей в этой области. Рассматривая последовательный LRC-контур

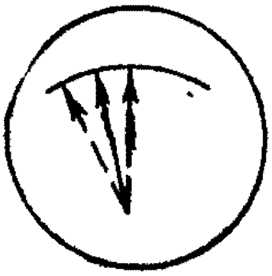

Рис. 10.1. Стрелка амперметра вращается на оси и имеет поэтому одну степень свободы.

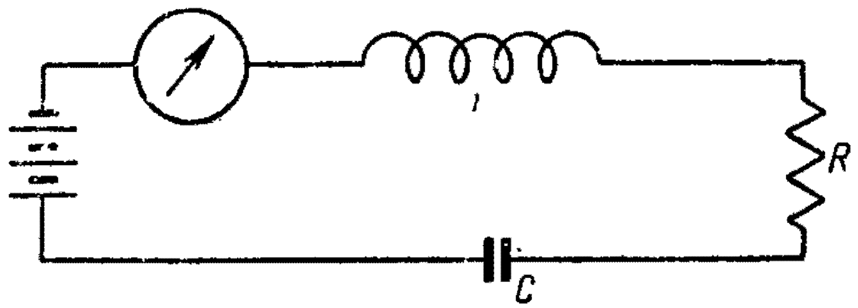

Рис. 10.2. Последовательный контур, содержащий источник э.д.с., амперметр, индуктивность L, сопротивление R и емкость C. Шум генерируется в сопротивлении.

(рис. 10.2), все элементы которого имеют одинаковую температуру T, и замечая, что контур имеет одну степень свободы (ток I в цепи), получаем:

$$\frac{1}{2} kT = \bar{E}_{\text{kin}} = \frac{1}{2} LI^2. \tag{10.1}$$

Следовательно, существуют беспорядочные токи, величина которых определяется вышеприведенным равенством. Единственная возможность уменьшения этих токов состоит в охлаждении всего контура. Распределение этих токов по частоте будет исследовано ниже.

2. Случайное блуждание

Начнем со статистической задачи, имеющей большое значение для всех примеров беспорядочного движения. Частица может перемещаться вдоль прямой, совершая через равные промежутки времени равные шаги либо вправо, либо влево. Обозначим единицу смещения (шаг) через l, а моменты, в которые происходят перемещения, через 1, 2, 3 и т. д.

Рис. 10.3 дает общую схему. Диаграмма воспроизводит известный треугольник Паскаля; число возможных путей, приводящих к данной точке после n шагов, есть

$$\frac{n!}{q!\,(n-q)!},$$

если расстояние точки от начала равно

$$m = -n + 2q,$$
$$q = \frac{n+m}{2}, \quad n - q = \frac{n-m}{2}\,^{1)}. \tag{10.2}$$

Вероятность $\pi_{n,\,m}$ достижения точки m выражается отноше-

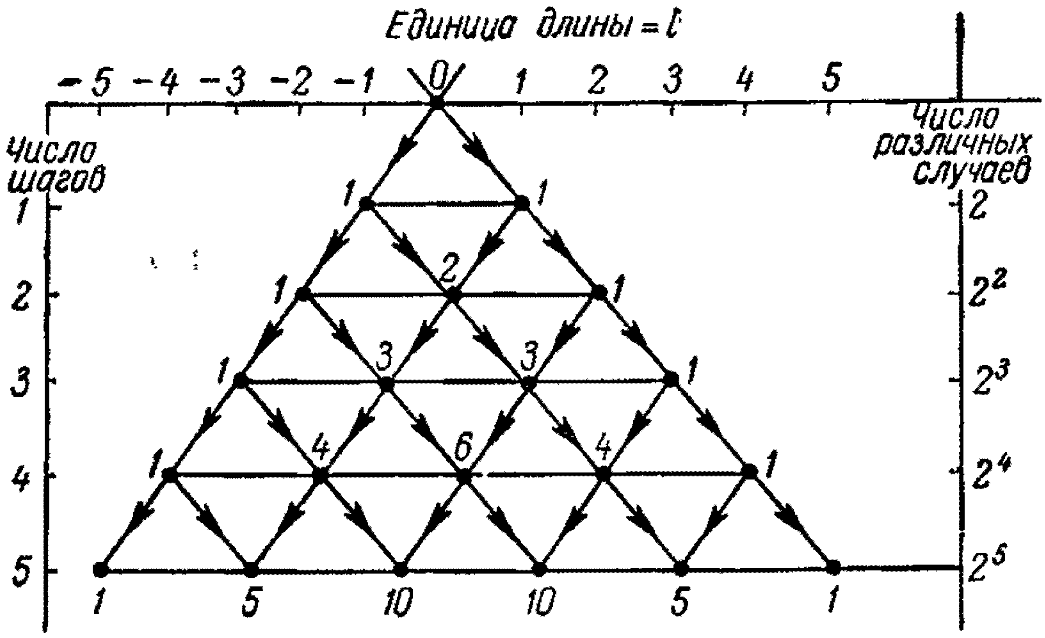

Рис. 10.3. Случайное блуждание с шагами одинаковой длины воспроизводит треугольник Паскаля для биноминальных коэффициентов. Цифры означают число путей, ведущих к данной точке. Общее число случаев равно 2^n.

нием числа путей, ведущих к точке, к общему числу возможных путей, равному, очевидно, 2^n, так как каждый путь ветвится на два ($\pm l$) на каждом шаге:

$$\pi_{n,\,m} = \frac{1}{2^n}\,\frac{n!}{q!\,(n-q)!} = \frac{n!}{2^n\left(\frac{n+m}{2}\right)!\left(\frac{n-m}{2}\right)!}. \tag{10.3}$$

---

1) Смысл этих соотношений станет ясен, если заметить, что q означает число шагов вправо, а $(n-q)$ — число шагов влево. (*Прим. перев.*)

Точки m, которые могут быть достигнуты после n шагов, имеют в силу (10.2) ту же четность, что и n. Среднее смещение $\overline{R_n}$ после n шагов равно, очевидно, нулю, так как движения вправо и влево равновероятны. Средний квадрат смещения $\overline{R_n^2}$ возрастает с числом шагов. Схема рис. 10.3 дает:

$$\overline{R_n^2} = \overline{m^2} l^2 = l^2 \sum_{m=-n}^{+n} m^2 \pi_{n,\, m}, \tag{10.4}$$

$$\left.\begin{aligned} &\overline{R_0^2} = 0, \\ &\overline{R_1^2} = \frac{1}{2}(l^2 + l^2) = l^2, \\ &\overline{R_2^2} = \frac{1}{4}[(2l)^2 + 2\cdot 0 + (2l)^2] = 2l^2, \\ &\overline{R_3^2} = \frac{1}{8}[(3l)^2 + 3l^2 + 3l^2 + (3l)^2] = \frac{24}{8} l^2 = 3l^2. \end{aligned}\right\} \tag{10.5}$$

Эти результаты подсказывают закон:

$$\overline{R_n^2} = n l^2, \tag{10.6}$$

который можно доказать методом индукции; именно, нужно показать, что если

$$\overline{m_n^2} = \sum_{m=-n}^{m=n} \pi_{n,\, m} m^2 = n, \tag{10.7}$$

то

$$\overline{m_{n+1}^2} = \sum_{m=-n-1}^{n=n+1} \pi_{n+1,\, m} m^2 = n + 1. \tag{10.7a}$$

Для этого заметим сначала, что если положение $m < n+1$ достигается за $n+1$ шагов, то положение после n шагов должно было быть либо $m+1$, либо $m-1$, причем обе эти возможности равновероятны. Поэтому

$$\pi_{n+1,\, m} = \frac{1}{2}(\pi_{n,\, m+1} + \pi_{n,\, m-1}), \quad m < n+1, \tag{10.8}$$

и

$$\pi_{n+1,\, n+1} = \pi_{n+1,\, -n-1} = \frac{1}{2^{n+1}}. \tag{10.8a}$$

Напомним, что оба индекса в $\pi_{n,\,m}$ имеют одинаковую четность, так что при суммировании по какому-либо индексу не равные нулю слагаемые дают только значения индекса через один. Учитывая это обстоятельство, и используя (10.8) и (10.8а), получаем для суммы (10.7а):

$$\overline{m_{n+1}^2}=2\frac{(n+1)^2}{2^{n+1}}+\frac{1}{2}\sum_{m=-n+1}^{m=n-1}(\pi_{n,\,m+1}+\pi_{n,\,m-1})\,m^2=$$

$$=\frac{(n+1)^2}{2^n}+\frac{1}{2}\sum_{m=-n+1}^{m=n-1}\pi_{n,\,m+1}m^2+\frac{1}{2}\sum_{m=-n+1}^{m=n-1}\pi_{n,\,m-1}m^2.$$

Положим в первой сумме $k=m+1$, а во второй $l=m-1$. Тогда

$$\overline{m_{n+1}^2}=\frac{(n+1)^2}{2^n}+\frac{1}{2}\sum_{k=-n+2}^{k=n}\pi_{n,\,k}(k-1)^2+$$

$$+\frac{1}{2}\sum_{l=-n}^{k=n-2}\pi_{n,\,l}(l+1)^2=\frac{(n+1)^2}{2^n}+$$

$$+\frac{1}{2}\sum_{k=-n}^{k=n}\pi_{n,\,k}(k-1)^2-\frac{1}{2}\pi_{n,\,-n}(-n-1)^2+$$

$$+\frac{1}{2}\sum_{l=-n}^{l=n}\pi_{n,\,l}(l+1)^2-\frac{1}{2}\pi_{n,\,n}(n+1)^2.$$

Заметим теперь, что

$$\pi_{n,\,n}=\pi_{n,\,-n}=\frac{1}{2^n},$$

так что остаются только две суммы, которые могут быть объединены:

$$\overline{m_{n+1}^2}=\frac{1}{2}\sum_{k=-n}^{k=n}\pi_{n,\,k}[(k-1)^2+(k+1)^2]=$$

$$=\frac{1}{2}\sum_{k=-n}^{k=n}\pi_{n,\,k}(2k^2+2)=$$

$$=\sum_{k=-n}^{k=n}\pi_{n,\,k}k^2+\sum_{k=-n}^{k=n}\pi_{n,\,k}=n+1, \qquad (10.9)$$

на основании (10.7) и вследствие того, что величины $\pi_{n,k}$ являются вероятностями, сумма которых по k равна единице. Итак, равенство (10.7а) доказано, что и требовалось.

Распределение точек расширяется со временем, и мы получаем для среднеквадратичного смещения (при одинаковых шагах длиной l)

$$(\overline{R^2})^{1/2} = n^{1/2}l. \tag{10.10}$$

Можно доказать (см. Приложение), что при увеличении числа шагов распределение (10.3) сходится к гауссову распределению.

Мы увидим позднее, как этот результат прилагается к броунову движению. Мы можем также применить его для округления ошибок в длинном вычислении. При этом мы можем взять экстремальный случай $|l| = {}^1/_2$, при котором среднеквадратичная ошибка

$$\sqrt{\overline{R_n^2}} \approx \frac{1}{2}\sqrt{n}.$$

Иначе говоря, мы полагаем, что последний знак лежит в интервале $\pm{}^1/_2$. С другой стороны, наибольшая ошибка пропорциональна ${}^1/_2 n$, что соответствует весьма маловероятному случаю сложения всех ошибок.

Наши рассуждения можно распространить на случай шагов разной длины. При этом мы имеем случайное распределение около среднего:

$$\overline{R_n^2} = \overline{l_n^2}, \quad \text{или} \quad \sqrt{\overline{R_n^2}} = \sqrt{\overline{l^2}}\sqrt{n}. \tag{10.11}$$

Последняя формула больше подходит для округления ошибок.

3. Дробовой эффект

Мы можем применить формулу случайного блуждания к исследованию шума ламп (дробовой эффект) и радиоактивных превращений. При радиоактивном превращении вещество эмитирует (испускает) частицы со средней скоростью N частиц в секунду. Это есть эмиссия дискретных частиц без корреляции между следующими друг за другом частицами.

Действительная эмиссия может быть записана в виде

$$P = Nt + \nu, \tag{10.12}$$

где ν есть поправка к средней эмиссии, соответствующая шуму. Это описание близко соответствует электронной эмиссии горячего катода, и поэтому такого же типа шум мы имеем в электронной лампе. Эмиссия как функция времени показана на рис. 10.4.

Прежде всего, нужно показать, как свести это к задаче о случайном блуждании. Это можно сделать различными способами в зависимости от действительных условий эмиссии.

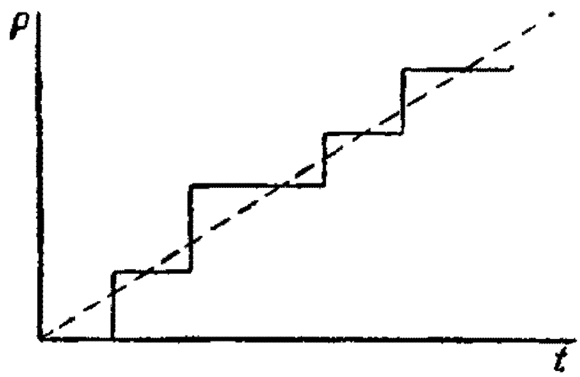

Рис. 10.4. Эмиссия как функция времени есть ступенчатая функция. Средняя эмиссия представляется прямой. «Ступеньки» соответствуют члену ν в (10.12).

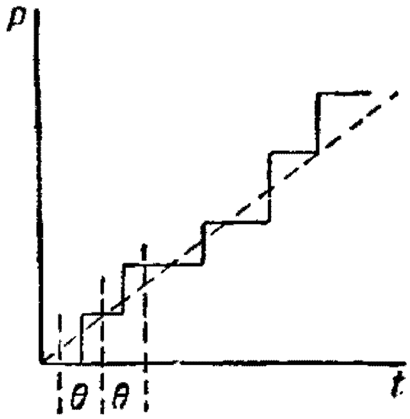

Рис. 10.5. Ступенчатая функция рис. 10.4 с интервалом θ, выбранным так, что на протяжении θ вероятность испускания частицы равна $^1/_2$.

Начнем с очень простой задачи, соответствующей следующим предположениям: существует временной интервал θ настолько малый, что на протяжении этого интервала возможны два равновероятных события — либо частица испускается, либо нет (рис. 10.5). Таким образом, на протяжении этого интервала изменение эмиссии есть

$$\Delta t = \theta, \quad \Delta P = \begin{cases} 0 \\ 1 \end{cases} = \frac{1}{2} + \begin{cases} -\frac{1}{2} \\ +\frac{1}{2} \end{cases}, \tag{10.13}$$

откуда следует, что

$$N = \frac{1}{2\theta}$$

и что

$$\Delta\nu = \begin{cases} -\frac{1}{2} \\ \ \ \frac{1}{2} \end{cases}.$$

Так, если взять $|l| = \frac{1}{2}$, то мы сводим задачу к постоянному нарастанию Nt, на которое наложено случайное блуждание. Средний квадрат флуктуаций по-прежнему выражается как

$$\overline{\nu_n^2} = nl^2 = \frac{1}{4}n, \tag{10.14}$$

где n выражает полное время $t = n\theta$. На протяжении этого промежутка времени t средняя эмиссия равна

$$\bar{P} = Nt = n\theta\frac{1}{2\theta} = \frac{1}{2}n,$$

откуда

$$\overline{\nu_n^2} = \frac{1}{2}\bar{P}. \tag{10.15}$$

Средний квадрат флуктуации в рассматриваемом случае равен половине среднего числа эмитированных частиц.

Более важная и более близкая к действительности физическая задача получается при иных предположениях, соответствующих *полностью случайной эмиссии*. Мы выбираем очень малый временной интервал θ и предполагаем, что на протяжении этого интервала приращение эмиссии ΔP есть

$$\Delta P = \begin{cases} 0 \text{ с вероятностью } 1 - N\theta, \\ 1 \text{ с вероятностью } N\theta \ll 1. \end{cases}$$

Интервал θ настолько мал, что эмиссия двух, трех и более частиц считается невозможной. Средняя скорость эмиссии есть N. Условие может быть записано в следующем виде:

$$\Delta P = N\theta + \begin{cases} -N\theta, \text{ вероятность } (1 - N\theta), \\ 1 - N\theta, \text{ вероятность } N\theta. \end{cases} \tag{10.16}$$

Мы разделили здесь среднее нарастание $N\theta$ и флуктуации, приводящие к задаче о случайном блуждании

$$\Delta\nu = \begin{cases} -N\theta, \text{ вероятность } (1 - N\theta), \\ 1 - N\theta, \text{ вероятность } N\theta. \end{cases} \tag{10.16a}$$

Это есть случай неравных шагов, и мы применяем соответствующую формулу (10.11). Средний квадрат длины шага равен

$$\overline{l^2}=\overline{\Delta\nu^2}=(N\theta)^2(1-N\theta)+(1-N\theta)^2N\theta=N\theta(1-N\theta),$$

или, так как $N\theta$ очень мало,

$$\overline{l^2}=N\theta.$$

Из этого соотношения и (10.11) получаем средний квадрат флуктуации на всем промежутке t, на протяжении которого средняя эмиссия есть Nt:

$$t=n\theta \quad (n \text{ — велико}),$$
$$\overline{P}=Nt=nN\theta,$$
$$\overline{\nu^2}=n\overline{l^2}=nN\theta=\overline{P}.$$

Средний квадрат флуктуации в случае полностью случайной эмиссии в точности равен среднему числу эмитированных частиц.

4. Броуново движение

Вернемся к обсуждению броунова движения; будем следовать трактовке Лоренца[1]). Напомним некоторые численные значения:

$$k=1{,}38\cdot 10^{-16}\ \text{эрг/молекула}\ {}^\circ K,$$
$$kT=4{,}14\cdot 10^{-14}\ \text{эрг/молекула}\approx 1/40\ ev,$$
$$\text{при}\ T=300^\circ K\approx 27^\circ C.$$

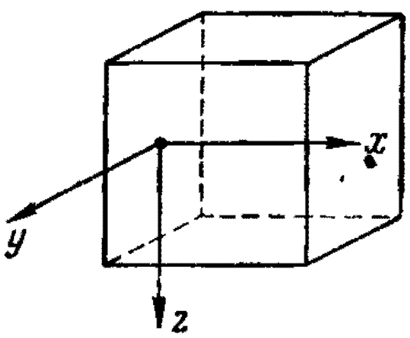

Рис. 10.6. Координаты x, y и z частицы, погруженной в вязкую жидкость.

Рассмотрим частицу, погруженную в среду с координатами, показанными на рис. 10.6. Мы ограничимся рассмотрением движения только вдоль оси x. Так как движение случайно, то движение в направлении оси x не зависимо от движений в направлениях y и z.

[1]) H. A. Lorentz, Les théories statistiques en thermodynamique, p. 49, Teubner, Leipzig, 1916; G. L. de Haas—Lorentz, Die Brownsche Bewegung, Vieweg, Braunschweig, 1913.

Вводя коэффициент вязкого трения w, запишем уравнение движения частицы с массой m:

$$F_x = m\frac{dv}{dt} + wv, \tag{10.17}$$

где F_x означает случайную силу, действующую на частицу. Сила F_x должна быть случайной, так как на протяжении любого промежутка времени (разумно выбранной малости) будет происходить очень большое число случайных столкновений частицы с молекулами жидкости.

Рассмотрим теперь временной интервал τ, настолько малый, чтобы изменение скорости Δv частицы на протяжении этого интервала было мало. Тогда полный импульс силы, действующей на частицу за время τ, равен

$$X = \int_0^{\tau} F_x\,dt. \tag{10.18}$$

Скорость частицы равна v_0 при $t=0$ и v_1 при $t=\tau$; интегрирование (10.17) по интервалу τ дает:

$$m(v_1 - v_0) + wv_0\tau = X. \tag{10.19}$$

Перегруппировывая члены, имеем:

$$mv_1 = v_0(m - w\tau) + X. \tag{10.20}$$

Так как F_x полностью случайна, можно считать:

$$\overline{X} = 0, \tag{10.21}$$

но надо учесть, что $\overline{X^2}$ не равно нулю. Возведем (10.20) в квадрат:

$$v_1^2 = v_0^2\left(1 - \frac{2w\tau}{m} + \frac{w^2\tau^2}{m^2}\right) + \frac{X^2}{m^2} + \frac{2Xv_0}{m}\left(1 - \frac{w\tau}{m}\right), \tag{10.22}$$

и заметим, что τ очень мало, так что

$$\frac{w^2\tau^2}{m^2} \ll \frac{2w\tau}{m}.$$

Усредняя затем (10.22) и учитывая, что мы положили X

равным нулю, найдем среднее значение v_1^2:

$$\overline{v_1^2}=\overline{v_0^2}\left(1-\frac{2w\tau}{m}\right)+\frac{\overline{X^2}}{m^2}. \tag{10.23}$$

Наша система (частица + жидкость) имеет температуру T, и к ней применимы соотношения кинетической теории, именно:

$$\frac{1}{2}\,m\overline{v^2}=\frac{1}{2}\,kT.$$

Поэтому

$$\overline{v_1^2}=\overline{v_0^2}=\frac{kT}{m}. \tag{10.24}$$

Подставляя (10.24) в (10.23) и разрешая относительно $\overline{X^2}$, получаем для этой величины

$$\overline{X^2}=2w\tau kT. \tag{10.25}$$

Итак, средний квадрат импульса силы не равен нулю. Далее, переписывая (10.23) в виде

$$\frac{1}{2}\,m\overline{v_1^2}=\frac{1}{2}\,m\overline{v_0^2}\left(1-\frac{2w\tau}{m}\right)+\frac{\overline{X^2}}{2m},$$

мы замечаем, что $\frac{1}{2}\,m\overline{v_0^2}(2w\tau/m)$ представляет потерю кинетической энергии частицы при ее движении в вязкой среде, тогда как $\overline{X^2}/2m$ представляет компенсирующую энергию, доставляемую частице случайными столкновениями с молекулами.

Итак, мы наблюдаем передачу энергии в двух направлениях:

Кинетическая энергия частицы $\rightarrow$ движение молекул.

Движение молекул $\rightarrow$ кинетическая энергия частицы.

Далее, мы видим, что если вязкое трение велико, то и компенсирующий перенос энергии велик. Большое вязкое трение означает сильную связь между частицей и окружающей ее вязкой жидкостью, а следовательно, и мощные столкновения, соответствующие большим случайным импульсам X. Работа случайных сил производится за счет энергии теплового движения (т. е. теплоты) и снова превращается в теплоту через посредство вязкого трения.

Для нахождения среднеквадратичного смещения частицы мы несколько упростим рассуждение, предположив, что при $t=0$ $v_0=0$, а при $t=\tau$ $v=v_1$. Предположим также, что после момента τ силы, действующие на частицу, отсутствуют; иначе говоря, мы выделяем одно из столкновений и рассматриваем его, а также каждое из других столкновений как индивидуальное событие. Тогда после τ движение затухает, и мы можем записать:

$$v=v_1 e^{-\frac{wt}{m}}. \tag{10.26}$$

Разрешая относительно x, находим:

$$x=\int_{\tau}^{\infty} v\,dt=v_1\int_{\tau}^{\infty} e^{-\frac{wt}{m}}\,dt=\frac{v_1 m}{w}\left[e^{-\frac{wt}{m}}\right]_{\tau}^{\infty}.$$

Но так как τ мало, мы можем взять пределы для экспоненциальной функции 0 и ∞; поэтому из (10.20) при $v_0=0$ получаем:

$$x=\frac{mv_1}{w}=\frac{X}{w}. \tag{10.27}$$

Величина x представляет смещение частицы вследствие импульса X, полученного в момент τ. Это есть вид случайного блуждания, так как $\overline{x}=0$ потому, что $\overline{X}=0$. Но $\overline{X^2}\neq 0$; возводя в квадрат и усредняя (10.27), получаем:

$$\overline{x^2}=\frac{\overline{X^2}}{w^2}. \tag{10.28}$$

Подстановка (10.25) в (10.28) дает:

$$\overline{x^2}=\frac{2kT}{w}\tau. \tag{10.29}$$

Для большего промежутка времени t, содержащего много малых интервалов τ, квадраты смещений суммируются как независимые величины, и мы имеем:

$$\overline{x^2}=\frac{2kTt}{w}. \tag{10.30}$$

Этот результат был проверен экспериментально путем наблюдения зависимости от времени положений коллоидальных частиц в суспензии.

5. Тепловое движение в электрической цепи

Обратимся теперь к проблеме, очень сходчой с броуновым движением: к случайному течению зарядов в электрической цепи. Рассмотрим цепь рис. 10.7. Она состоит из индуктивности, сопротивления и амперметра для измерения

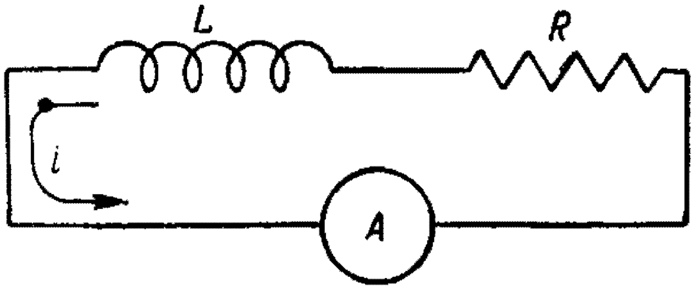

Рис. 10.7. Цепь, использованная для рассуждения о шуме, обусловленном тепловым движением. Индуктивность, сопротивление и амперметр включены последовательно, и вся схема поддерживается при температуре T.

тока; вся система находится при некоторой температуре T. Система имеет одну степень свободы и, следовательно,

$$\overline{E}_{\text{kin}} = \frac{1}{2} kT = \frac{1}{2} L\overline{i^2}. \tag{10.31}$$

Ток флуктуирует случайным образом, $\bar{i} = 0$. Если записать уравнение напряжений

$$L\frac{di}{dt} + Ri = F, \tag{10.32}$$

где F означает случайную э.д.с., действующую в цепи, то, как видим, это уравнение совпадает с уравнением движения частицы, совершающей броуново движение (см. (10.17)). Более того, механизм обмена энергией сходен в обоих случаях.

Электрон отдает свою кинетическую энергию молекулам сопротивления, а столкновения молекул дают кинетическую энергию электрону. Компенсация выражается теперь через средний квадрат импульса случайной э.д.с., и мы получаем соотношение, сходное с (10.25):

$$\overline{X^2} = 2R\tau kT. \tag{10.33}$$

Заряд q занимает место смещения x; аналогично (10.30) имеем:

$$\overline{q^2} = \frac{2kT}{R} t. \tag{10.34}$$

Итак, имеется два источника шума в электрической цепи: тепловой шум, обусловленный тепловым возбуждением и флуктуациями заряда в таких элементах цепи, как сопротивления, и дробовой эффект в лампах. В дальнейшем мы рассмотрим вопрос о спектральном составе теплового шума.

Можно построить несколько более подробную модель, чтобы учесть возрастание со временем среднеквадратичного значения импульса силы в случае броунова движения (или э.д.с. в случае электрической цепи).

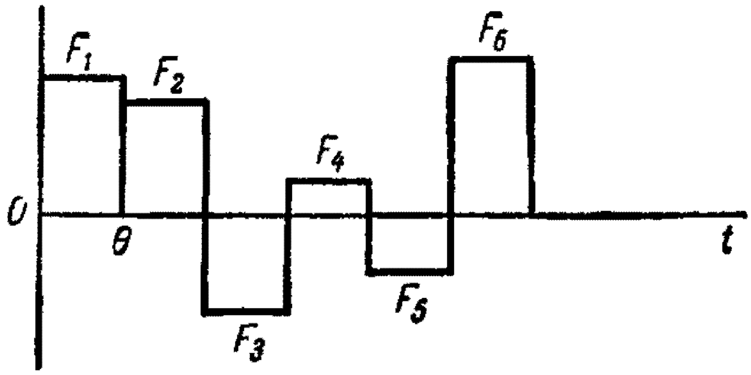

Рис. 10.8. Случайная сила есть ступенчатая функция, имеющая постоянное значение на протяжении каждого интервала от $(n-1)\,\theta$ до $n\theta$.

Предположим, что сила изменяется во времени случайно, как показано на рис. 10.8, но что она остается постоянной на протяжении каждого интервала θ (θ — очень мало, $\tau = n\theta$). Сила случайна, корреляция отсутствует и поэтому

$$\left.\begin{aligned} \overline{F_iF_k} &= 0, \quad i \neq k, \\ \overline{F_i^2} &= \overline{F^2} \neq 0. \end{aligned}\right\} \qquad (10.35)$$

Находим:

$$X = \int\limits_{t=0}^{t=\tau=n\theta} F\,dt = \theta\,(F_1 + F_2 + \ldots + F_n)$$

и

$$X^2 = \theta^2\,(F_1^2 + F_2^2 + \ldots + F_n^2 + 2F_1F_2 + 2F_1F_3 + \ldots). \qquad (10.36)$$

Усредняя, получаем:

$$\overline{X^2} = \theta^2\,\overline{(F_1^2 + F_2^2 + \ldots)}.$$

Но из (10.35), учитывая, что средние значения взаимных членов равны нулю, имеем:

$$\overline{X^2} = n\theta^2\overline{F^2}. \qquad (10.37)$$

Подставляя $\tau = n\theta$, получаем:

$$\overline{X^2} = \tau\,\overline{(F^2\theta)}, \tag{10.38}$$

и мы показали, что средний квадрат импульса силы (или э.д.с.) возрастает пропорционально времени, что согласуется с (10.25) и (10.33) и поясняет механизм этого возрастания.

Приложение

В разделе 2 обсуждалась задача о случайном блуждании и было получено выражение для вероятности $\pi_{n,m}$ достижения точки m после n шагов (см. (10.3)):

$$\pi_{n,m} = \frac{n!}{2^n \frac{n+m}{2}!\,\frac{n-m}{2}!}.$$

Исследуем теперь соотношения, когда число шагов n становится очень большим, тогда как m остается конечным. Применяя формулу Стирлинга

$$n! \to \left(\frac{n}{e}\right)^n \sqrt{2\pi n}, \quad n \gg 1,$$

после некоторых преобразований получим:

$$\pi_{n,m} \to \frac{n^n}{(n+m)^{\frac{n+m}{2}}(n-m)^{\frac{n-m}{2}}}\,\frac{\sqrt{2\pi n}}{2\pi\sqrt{\frac{n+m}{2}\cdot\frac{n-m}{2}}} =$$

$$= \frac{2\sqrt{n}}{\sqrt{2\pi(n^2-m^2)}}\cdot\frac{1}{\left(1+\frac{m}{n}\right)^{\frac{n+m}{2}}\left(1-\frac{m}{n}\right)^{\frac{n-m}{2}}}.$$

Члены знаменателя имеют вид $(1+\varepsilon)^{1/\varepsilon}$ и стремятся к e при уменьшении ε, но нам нужно лучшее приближение. Воспользуемся разложением

$$\ln(1+\varepsilon)^{1/\varepsilon} = \frac{1}{\varepsilon}\ln(1+\varepsilon) = \frac{1}{\varepsilon}\left(\varepsilon - \frac{\varepsilon^2}{2} + \ldots\right) = 1 - \frac{\varepsilon}{2} + \ldots,$$

откуда

$$(1+\varepsilon)^{1/\varepsilon} \to e^{1-\frac{\varepsilon}{2}},$$

что дает при $\varepsilon = m/n$:

$$\left[\left(1+\frac{m}{n}\right)^{\frac{n}{m}}\right]^{\frac{m}{2}+\frac{m^2}{2n}} \to \left(e^{1-\frac{m}{2n}}\right)^{\frac{m}{2}+\frac{m^2}{2n}} = e^{\frac{m}{2}+\frac{m^2}{4n}+\ldots}.$$

Аналогично

$$\left(1-\frac{m}{n}\right)^{\frac{n-m}{2}} \to e^{-\frac{m}{2}+\frac{m^2}{4n}},$$

и, наконец,

$$\pi_{n,m} \to \frac{2\sqrt{n}}{\sqrt{2\pi(n^2-m^2)}}\, e^{-\frac{m^2}{2n}} \to \frac{2}{\sqrt{2\pi n}}\, e^{-\frac{m^2}{2n}}$$

— гауссово распределение вероятностей. Нужно помнить, что m имеет всегда ту же четность, что и n. Для данного n расстояние между последовательными точками есть $\Delta m = 2$ (см. рис. 10.3). Рассмотрим, к примеру, случай, когда n и m четны:

$$n = 2h, \quad m = 2y, \quad \Delta y = 1, \quad h \gg y.$$

Пренебрежем членом с m^2 в знаменателе $\pi_{n,m}$ и запишем:

$$\pi_{h,y} = \frac{1}{\sqrt{\pi h}}\, e^{-\frac{y^2}{h}}.$$

Сравним этот результат со стандартной формулой нормального распределения:

$$P = \frac{1}{\sqrt{2\pi\sigma}}\, e^{-\frac{y^2}{2\sigma^2}}, \quad \overline{y^2} = \sigma^2,$$

где σ — стандартное отклонение. В нашей формуле для $\pi_{h,y}$

$$\sigma^2 = \frac{h}{2}, \quad \overline{m^2} = 4\overline{y^2} = 2h = n,$$

что соответствует общему соотношению (10.7).

ГЛАВА 11

ТЕПЛОВОЙ ШУМ В ЭЛЕКТРИЧЕСКОЙ ЦЕПИ; ФОРМУЛА НАЙКВИСТА

1. Модель со случайными импульсами

Применим некоторые результаты предыдущей главы к обсуждению формулы Найквиста, определяющей частотное распределение шума, обусловленного случайными тепловыми э.д.с. [1]). Напомним, что согласно (10.33) и (10.38) средний квадрат импульса э.д.с. F за время τ равен

$$X = \int_0^\tau F dt, \quad \overline{X^2} = 2RkT\tau, \tag{11.1}$$

$$\overline{X^2} = \tau\theta\overline{F^2}. \tag{11.2}$$

Соотношение (11.2) получено путем рассмотрения модели со случайной э.д.с., причем каждая э.д.с. F постоянна на протяжении малого временного интервала θ и

$$\tau = n\theta.$$

Объединение (11.1) и (11.2) дает:

$$\overline{F^2}\theta = 2RkT. \tag{11.3}$$

Хотя в нашей модели мы имеем случайную последовательность прямоугольных импульсов, спектр мощности этой последовательности совпадает со спектром одиночного прямоугольного импульса, так как при случайном следовании

---

[1]) J. L. Lawson, G. E. Uhlenbeck, Threshold Signals, Radiation Laboratory Series 24, Mc Graw-Hill, NY, 1950. (Русский перевод: «Пороговые сигналы», Сов. радио, 1952.)

импульсов (без корреляции) результат получается путем суммирования квадратов амплитуд частотных составляющих.

Для одиночного импульса э.д.с. высотой F и длительностью θ согласно (8.23) и (8.25) имеем:

$$f(t)=\int\limits_{\nu=-\infty}^{+\infty} C(\nu)\cos 2\pi\nu t\,d\nu=\int\limits_{\nu=0}^{\infty} 2C(\nu)\cos 2\pi\nu t\,d\nu, \quad (11.4)$$

где спектральная плотность

$$C(\nu)=F\theta\,\frac{\sin \pi\nu\theta}{\pi\nu\theta}. \quad (11.5)$$

Длительность импульса есть θ и *номинальная* ширина полосы $1/\theta$ (глава 8, раздел 4):

$$\Delta\nu=2\nu_M=\frac{1}{\theta}, \quad -\nu_M\leqslant\nu\leqslant+\nu_M, \quad \nu_M=\frac{1}{2\theta}. \quad (11.6)$$

В грубом приближении можно представить, что спектр прямоугольного импульса высотой F и длительностью θ простирается приблизительно от $-\nu_M$ до $+\nu_M$ с постоянной спектральной плотностью

$$C(\nu)=F\theta. \quad (11.7)$$

Если мы не желаем вводить отрицательные частоты, то можно взять

$$0\leqslant\nu\leqslant\nu_M, \quad \nu_M=\frac{1}{2\theta}, \quad (11.8)$$

и удвоить спектральную плотность (как в (11.4)):

$$C'(\nu)=2C(\nu)=2F\theta. \quad (11.9)$$

Рассмотрим теперь последовательность импульсов длительностью θ и со средним $\overline{F^2}$. Средний квадрат э.д.с. согласно (11.3) равен

$$\overline{E^2}=\overline{F^2}=2RkT\,\frac{1}{\theta} \quad (11.10)$$

и занимает приблизительно полосу частот $0\leqslant\nu\leqslant\nu_M$. Поэтому

$$\overline{E_\nu^2}=4RkT\nu_M. \quad (11.11)$$

распределена почти равномерно в полосе частот от 0 до ν_M. Можно сделать пропорциональный переход к меньшему интервалу частот $\Delta\nu$, т. е. к интервалу между ν и $\nu + \Delta\nu$,

$$\overline{e^2} = 4RkT\Delta\nu. \tag{11.12}$$

Это и есть формула Найквиста[1]).

2. Метод Найквиста

Рассмотрим теперь вывод этой формулы самим Найквистом. Прежде всего, мы должны построить цепь, способную передавать непрерывный диапазон частот, например длинную линию (рис. 11.1). Пусть линия имеет длину l и

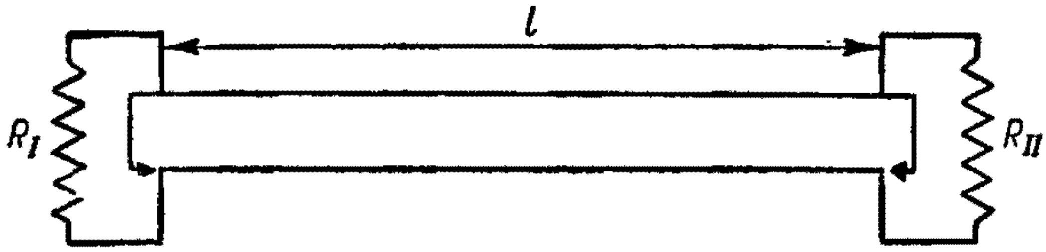

Рис. 11.1. Длинная линия длины l, нагруженная на концах сопротивлениями $R_{\rm I}$ и $R_{\rm II}$, с ключами для замыкания сопротивлений накоротко.

волновое сопротивление z. Замкнем линию на концах на сопротивления $R_{\rm I}$ и $R_{\rm II}$ и пусть $R_{\rm I} = R_{\rm II} = z$. Вся схема находится при некоторой температуре T. Каждое сопротивление является источником тепловой э.д.с., так что сигналы посылаются из $R_{\rm I}$ и поглощаются в $R_{\rm II}$, и обратно. Эти сигналы покрывают широкую полосу частот. Когда достигнуто равновесие, мы замыкаем накоротко оба конца линии, и так как должно поддерживаться тепловое равновесие, то имеющиеся в линии сигналы уже более не поглощаются (и не посылаются), а лишь отражаются от обоих концов линии. Условие равновесия для замкнутой накоротко линии состоит в том, что

---

[1]) Вот некоторые основные источники: P. W. Bridgman, Note on the Principle of Detailed Balancing, Phys. Rev. **31**, 101 (1928); J. B. Johnson, ibid. **29**, 367 (1927); **32**, 97 (1928); H. Nyquist, ibid. **32**, 110 (1928); J. Bernamont, Fluctuations de potentiel aux bornes d'un conducteur metallique, Ann. Phys. [11], 7 (1937); M. Courtines, Les fluctuations dans les appareils de mésures, Congr. intern. d'electricité, vol. 2, Paris (1932).

имеющиеся сигналы образуют стоячие волны с узлами на обоих короткозамкнутых концах (рис. 11.2). Это — *собственные* частоты линии, определяемые соотношением

$$\frac{n\lambda}{2} = l, \tag{11.13}$$

где каждое значение n означает одну собственную частоту и, следовательно, одну степень свободы линии.

Так как линию можно представить себе в виде гармонического осциллятора для каждой собственной частоты, то средняя энергия для каждой собственной частоты (на каждую

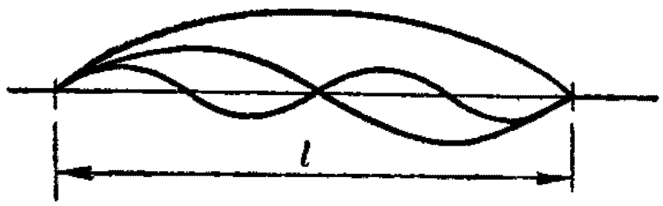

Рис. 11.2. Стоячие волны на линии при замкнутых накоротко сопротивлениях.

степень свободы) должна равняться сумме потенциальной и кинетической энергии, каждая из которых равна $\frac{1}{2}kT$. Таким образом, средняя энергия на степень свободы равна kT.

Чтобы получить энергию в линии, приходящуюся на частотный интервал $\Delta\nu$, мы должны подсчитать соответствующее число степеней свободы Δn. Из (11.13):

$$n = \frac{2l}{\lambda} = 2l\frac{\nu}{c}, \tag{11.14}$$

так как $\lambda = c/\nu$, где c — скорость волны. Поэтому

$$\Delta n = \frac{2l}{c}\Delta\nu \tag{11.15}$$

есть число степеней свободы в частотном интервале $\Delta\nu$. Полная средняя энергия короткозамкнутой линии, приходящаяся на интервал $\Delta\nu$, равна

$$\bar{E}_{\text{total}} = \Delta nkT = \frac{2l}{c}\Delta\nu kT. \tag{11.16}$$

Эта энергия была первоначально доставлена линии двумя равными сопротивлениями $R_{\rm I}$ и $R_{\rm II}$. Полная энергия, доставленная каждым сопротивлением, равна $\frac{l}{c}kT\,\Delta\nu$ в частотном интервале $\Delta\nu$.

Чтобы найти э.д.с., развиваемую в каждом сопротивлении, подсчитаем среднюю мощность P, т. е. энергию в единицу времени:

$$\overline{P}=\frac{\frac{l}{c}kT\,\Delta\nu}{\frac{l}{c}},$$

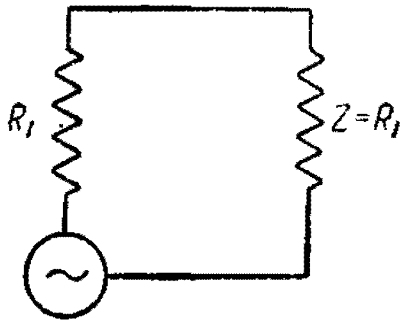

Рис. 11.3. Сопротивление $R_{\rm II}$ вместе с линией заменено волновым сопротивлением $Z=R_{\rm I}$.

так как $\frac{l}{c}$ есть время пробега сигнала от $R_{\rm I}$ к $R_{\rm II}$. Средняя мощность, выделяемая в каждом сопротивлении и посылаемая в линию, равна, таким образом,

$$\overline{P}=kT\,\Delta\nu. \tag{11.17}$$

Но каждое сопротивление, как источник тепловой э.д.с., работает на нагрузку $2R$ (рис. 11.3), так как нагрузка состоит из сопротивления R плюс волновое сопротивление линии $Z=R$.

Таким образом,

$$e_\nu=2Ri_\nu,$$

и мощность, выделяемая в Z, равна

$$\overline{P}=R\overline{i_\nu^2}=\frac{\overline{e_\nu^2}}{4R}, \tag{11.18}$$

откуда согласно (11.17)

$$\frac{\overline{e_\nu^2}}{4R}=kT\,\Delta\nu, \tag{11.19}$$

или

$$\overline{e^2}=4kTR\,\Delta\nu, \tag{11.20}$$

что совпадает с ранее полученным результатом.

3. Обсуждение и приложения

Для проверки нашего вывода применим его к последовательному LRC-контуру и подсчитаем средний квадрат случайного тока, обусловленного тепловыми э.д.с. Контур показан на рис. 11.4. Тепловая э.д.с. выражается формулой (11.20) или (11.12), и мы имеем:

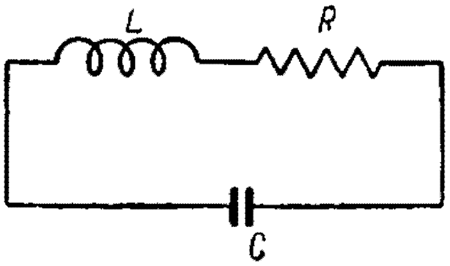

Рис. 11.4. Сопротивление R, индуктивность L и емкость C в последовательном включении образуют гармонический осциллятор. Шум генерируется в сопротивлении.

$$\left.\begin{aligned} \overline{i_\nu^2} &= \frac{\overline{e_\nu^2}}{|Z|^2} = \frac{\overline{e_\nu^2}}{R^2 + \left(L\omega - \frac{1}{\omega c}\right)^2}, \\ \omega &= 2\pi\nu. \end{aligned}\right\} \qquad (11.21)$$

Если положить

$$\omega_0 = 2\pi\nu_0 = \frac{1}{\sqrt{LC}} \qquad (11.22)$$

и воспользоваться (11.20), то (11.21) принимает вид

$$\overline{i_\nu^2} = \frac{4RkT\Delta\nu}{R^2 + L^2\left(\omega - \frac{\omega_0^2}{\omega}\right)^2}. \qquad (11.23)$$

Введя добротность контура

$$Q_0 = \frac{\omega_0 L}{R}, \qquad (11.24)$$

получаем:

$$\overline{i_\nu^2} = \frac{2}{\pi}\frac{kT}{L}\frac{Q_0}{\omega_0}\frac{\Delta\omega}{1 + \frac{Q_0^2}{\omega_0^2}\left(\omega - \frac{\omega_0^2}{\omega}\right)^2}. \qquad (11.25)$$

Кинетическая энергия в контуре при некоторой частоте ν равна

$$\overline{E}_{k\nu} = \frac{1}{2} L\overline{i_\nu^2}. \qquad (11.26)$$

Полная кинетическая энергия может быть получена путем суммирования по всем частотам от $\nu = 0$ до $\nu = \infty$; сумма заменяется интегралом

$$\overline{E}_{\text{kin}} = \sum \overline{E}_{k\nu} = \sum \frac{1}{2} L\overline{i_\nu^2} = \frac{kT}{\pi}\frac{Q_0}{\omega_0}\int\limits_0^\infty \frac{d\omega}{1 + \frac{Q_0^2}{\omega_0^2}\left(\omega - \frac{\omega_0^2}{\omega}\right)^2}. \qquad (11.27)$$

Этот интеграл нетрудно вычислить, и можно показать[1]), что

$$\frac{Q_0}{\omega_0}\int_0^\infty \frac{d\omega}{1+\frac{Q_0^2}{\omega_0^2}\left(\omega-\frac{\omega_0^2}{\omega}\right)^2}=\frac{\pi}{2}, \tag{11.28}$$

и полная (средняя) кинетическая энергия равна

$$\bar{E}_{\text{kin}}=\frac{1}{2}kT. \tag{11.29}$$

Имеется равное количество $\frac{1}{2}kT$ средней потенциальной энергии, зависящей от заряда конденсатора, так что окончательно полная средняя энергия равна

$$\bar{E}=\bar{E}_{\text{kin}}+\bar{E}_{\text{pot}}=kT, \tag{11.30}$$

что и является общеизвестным результатом для низкочастотного гармонического осциллятора.

4. Обобщения формулы Найквиста

Мы обосновали формулу Найквиста для случая электрического контура с сопротивлением R. Шум генерируется только в сопротивлении, где имеется связь между электрическим током и тепловым движением в веществе проводника. В индуктивностях и емкостях такой связи нет, и они не дают добавочного шума. В полном сопротивлении

$$Z=R+iX \tag{11.31}$$

только составляющая R дает шум.

---

[1]) Полагая $\omega=\omega_0 s$ и разбивая интеграл на два — от 0 до 1 и от 1 до ∞, — имеем:

$$Q_0\int_0^1 \frac{ds}{1+Q_0^2\left(s-\frac{1}{s}\right)^2}+Q_0\int_1^\infty \frac{ds}{1+Q_0^2\left(s-\frac{1}{s}\right)^2}=$$

$$=Q_0\int_1^\infty \frac{ds/s^2+ds}{1+Q_0^2\left(s-\frac{1}{s}\right)^2}=Q_0\int_0^\infty \frac{dx}{1+Q_0^2x^2}=\frac{\pi}{2},$$

где $x=s-\frac{1}{s}$ (с пределами 0 (для $s=1$) и ∞ (для $s=\infty$)).

Вместо электрической системы мы могли бы рассмотреть механическую систему с вязким трением. Мы уже пользовались аналогией механических и электрических колебаний в разделе 5 главы 10. Эта аналогия также применима и к формуле Найквиста. Объединяя результаты главы 10 (см. (10.18), (10.25), (10.30), (10.33) и (10.34)) и настоящей главы, мы можем установить более общие соотношения для любой системы с затуханием, выражаемым величиной R:

$$\overline{X^2} = 2RkTt, \tag{11.32}$$

где X — импульс силы за время t;

$$\overline{x^2} = \frac{2kT}{R}t, \tag{11.33}$$

где x — смещение при свободном броуновом движении за время t;

$$\overline{E}_{\text{kin}} = \frac{1}{2}kT \tag{11.34}$$

есть средняя кинетическая энергия на степень свободы и

$$\overline{f^2} = 4RkT\Delta\nu \tag{11.35}$$

есть средний квадрат сил с частотами между ν и $\nu + \Delta\nu$.

Все эти результаты имеют весьма широкое применение при низких частотах. Де-Хаас-Лоренц рассмотрела большое количество примеров и приложений формул (11.32), (11.33) и (11.34) [1]). Вопрос о применении (11.35) к ряду неэлектрических задач тщательно исследовал Каллен [2]), доказавший полную общность этой формулы и рассмотревший случай задач с различного рода ограничениями. Эти результаты показывают важные применения термодинамики к необратимым процессам [3]).

---

[1]) H. A. Lorentz, Les théories statistiques en thermodynamique, Teubner, Leipzig, 1916; G. L. de Haas-Lorentz, Die Brownsche Bewegung, Vieweg, Braunschweig, 1931.

[2]) H. B. Callen, Ph. D. thesis MIT, 1948; R. F. Greene, H. B. Callen, Phys. Rev. **83**, 1231 (1951); H. B. Callen, T. A. Welton, ibid. **83**, 34 (1951); H. B. Callen, R. F. Greene, ibid. **86**, 702 (1952).

[3]) L. Onsager, Phys. Rev. **37**, 405 (1931); **38**, 2265 (1931); H. B. G. Casimir, Rev. Mod. Phys. **17**, 343 (1945); С. Р. де Гроот, Термодинамика необратимых процессов, перевод с английского, Гостехиздат, 1956.

При очень высоких частотах ситуация существенно меняется. Прежде всего, нужно учесть, что многие новые степени свободы могут начать играть заметную роль. Так, в электрическом контуре появятся и будут возбуждаться тепловым движением высшие формы колебаний, обусловленные распределенными емкостями и индуктивностями. В механической системе будут возбуждаться различные виды поперечных колебаний. К тому же, существуют квантовые ограничения. Предыдущая теория может применяться только при низких частотах, для которых

$$h\nu \ll kT, \tag{11.36}$$

где h — постоянная Планка. Для высоких частот формула (11.34), как известно, уже не применима.

Средняя энергия осциллятора равна уже не kT, а

$$\overline{E(\nu)} = \frac{h\nu}{e^{\frac{h\nu}{kT}} - 1}. \tag{11.37}$$

Аналогичные поправки должны быть внесены и в другие формулы.

Физическая картина ясна: вязкое сопротивление указывает на связь данного вида движения с тепловым движением. Эта связь вызывает затухание движения и превращает его энергию высокого качества в тепло (энергию низкого качества). Но это преобразование энергии не может все время продолжаться, так как оно привело бы систему в состояние полного покоя, а мы знаем, что в системе при температуре T должна существовать та или иная форма броунова движения. Поэтому та самая связь, которая обусловливает затухание, должна также порождать случайные силы, действующие на систему и поддерживающие ее в состоянии теплового движения [1]). Это и является исходным положением при выводе формул (11.32) или (11.35). Объяснение в случае броунова движения частицы в воде очень просто: следствием столкновений частицы с молекулами воды является вязкое трение. Случайные

[1]) Смысл этого не очень ясного рассуждения состоит в противопоставлении макродвижения Броуновой частицы микродвижению молекул. Последнее и есть энергия в тепловой форме, тогда как первое трактуется, как механическая энергия (см. раздел 4 главы 10). (*Прим. перев.*)

столкновения молекул воды с частицей создают случайные силы, поддерживающие тепловое движение частицы. Положение аналогично во всех примерах, которые можно себе представить.

5. Тепловое движение в выпрямителе[1])

Пусть выпрямитель включен последовательно с сопротивлением R и реактивностью X, как показано на рис. 11.5. Активные сопротивления в цепи (сопротивление R, а также

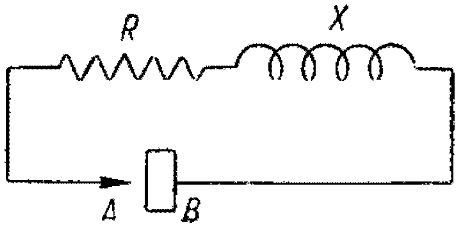

Рис. 11.5. Активное сопротивление R, реактивность X и выпрямитель AB в последовательном соединении. Шум генерируется в сопротивлении. Имеется выпрямленная э.д.с., но нет выпрямленного тока.

активное сопротивление выпрямителя) будут создавать случайные э.д.с., обусловленные тепловым движением. Выпрямитель не может выпрямить случайные токи, вызванные этими э.д.с., так как выпрямленный ток мог бы произвести работу вопреки второму началу термодинамики.

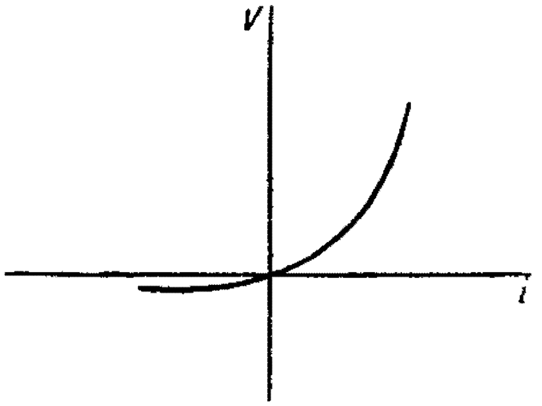

Рис. 11.6. Вольт-амперная характеристика выпрямителя.

Характеристика выпрямителя показана на рис. 11.6; так как она идет круче в правой части, то можно заключить, что флуктуации влево должны быть больше, чем вправо. В противном случае среднее значение тока не равнялось бы нулю, а имелся бы выпрямленный ток. Однако может быть выпрямленное напряжение, значение которого мы подсчитаем.

Для этого запишем выражение случайной э.д.с. (генерируемой в R и сопротивлении выпрямителя r) в данный момент:

$$\varphi = \varphi_0 + \sum_{\nu} \varphi_{\nu}, \tag{11.38}$$

[1]) L. Brillouin, Phys. Rev. 78, 627 (1950).

где φ_ν — комплексные члены разложения Фурье для э.д.с. Так как φ_ν — комплексные экспоненциальные функции, то в среднем

$$\overline{\varphi}_\nu = 0 \tag{11.39}$$

для всех ν и

$$\overline{\varphi} = \varphi_0. \tag{11.40}$$

Если бы в схеме не было выпрямителя, то мы имели бы $\varphi_0 = 0$. Мы имеем в виду вычислить именно φ_0 — постоянное среднее значение, обусловленное наличием выпрямителя.

Тепловые э.д.с. создают случайные токи, которые также можно представить рядом Фурье:

$$i = \sum_\nu i_\nu. \tag{11.41}$$

В этом разложении член i_0 отсутствует, так как выпрямленного тока не может быть. Как и ранее, i_ν представляются комплексными экспоненциальными функциями и равны в среднем нулю, так что для всех ν

$$\overline{i}_\nu = 0 \text{ и, следовательно, } \overline{i} = 0. \tag{11.42}$$

Падение напряжения на выпрямителе равно

$$\begin{aligned} V_A - V_B = f = ri + bii^* = \\ = r\sum_\nu i_\nu + b\left(\sum_\nu i_\nu i_\nu^* + \sum_\nu \sum_{\mu \neq \nu} i_\nu i_\mu^*\right), \end{aligned} \tag{11.43}$$

так как случайные токи очень малы, следовательно, выпрямитель работает на участке характеристики, близком к началу координат, и этот участок может быть аппроксимирован параболой.

Усредняя, получаем:

$$\overline{f} = b\sum_\nu i_\nu i_\nu^*. \tag{11.44}$$

Члены с перекрестными произведениями дают в среднем нуль, так как и они выражаются комплексными экспоненциальными функциями с показателем, не равным нулю.

Далее, мы составляем обычное уравнение напряжений:

$$\varphi = \varphi_0 + \sum_\nu \varphi_\nu = R \sum_\nu i_\nu + j \sum_\nu X_\nu i_\nu + f, \tag{11.45}$$

где $j = \sqrt{-1}$, и, усредняя, получаем:

$$\bar{\varphi} = \varphi_0 = \bar{f} = b \sum_\nu \overline{i_\nu i_\nu^*}. \tag{11.46}$$

Для нахождения φ_0 воспользуемся снова (11.45) и приравняем подобные члены:

$$\varphi_\nu = (R + r + jX_\nu)\, i_\nu. \tag{11.47}$$

Член ri_ν зависит от f (см. (11.43)). Членами с перекрестными произведениями (также зависящими от f) $i_\nu i_\mu^*$ $(\mu \neq \nu)$ пренебрежем по сравнению с членами i_ν, так как первые квадратичны относительно малых токов. Из (11.47) имеем:

$$\overline{i_\nu i_\nu^*} = \frac{\overline{\varphi_\nu \varphi_\nu^*}}{(R+r)^2 + X_\nu^2}. \tag{11.48}$$

Мы можем теперь воспользоваться (11.20) для нахождения $\overline{\varphi_\nu \varphi_\nu^*}$, так как только активные сопротивления в цепи дают случайные э.д.с.:

$$\overline{\varphi_\nu \varphi_\nu^*} = 4\,(R + r)\, kT\, \Delta\nu, \tag{11.49}$$

так что для выпрямленной э.д.с., даваемой выпрямителем, получаем окончательно:

$$\varphi_0 = 4\, kT\, (R + r) \sum_\nu \frac{\Delta\nu}{(R+r)^2 + X_\nu^2}. \tag{11.50}$$

Если заменить сумму интегралом и принять, что реактивность имеет индуктивный характер, т. е. $X = 2\pi L\nu$, то

$$\varphi_0 = \frac{bkT}{L}. \tag{11.51}$$

В заключение еще раз подчеркнем, что выпрямленный ток отсутствует и что выпрямленное напряжение равно падению напряжения на выпрямителе, так что в цепи не может быть совершена работа. Применение формулы к специальным примерам всегда обнаруживает такого рода полную компенсацию постоянного среднего напряжения. Наше рассмотрение является примером общего метода детального равновесия [1]).

[1]) P. W. Bridgman, Phys. Rev. **31**, 101 (1928).

ГЛАВА 12

НЕГЭНТРОПИЙНЫЙ ПРИНЦИП ИНФОРМАЦИИ

1. Связь между информацией и энтропией

Количественное определение информации, основанное на статистических представлениях, было дано в главе 1.

Мы рассматриваем ситуацию, в которой имеется P_0 различных возможных случаев или событий с одинаковыми априорными вероятностями. Для уменьшения числа возможных случаев до P_1 требуется информация I_1, и логарифм отношения P_0/P_1 есть мера I_1 (см. (1.6)):

$$\left.\begin{array}{lll}\text{Начальное положение } I_0=0 & P_0 \text{ возможностей,} \\ \text{Конечное положение } I_1>0 & P_1 \text{ возможностей,}\end{array}\right\} \qquad (12.1)$$

где

$$I_1=K\ln\frac{P_0}{P_1}.$$

В термодинамических единицах $K=k=$ постоянной Больцмана.

Мы введем теперь различие между двумя видами информации:

1. Свободная информация (free information) I_f, возникающая, когда возможные случаи рассматриваются как абстрактные и не имеющие определенного физического значения.

2. Связанная информация (bound information) I_b, возникающая, когда возможные случаи могут быть представлены как микросостояния физической системы. Таким образом, связанная информация есть специальный случай свободной.

Основание для введения этого различия состоит в том, что мы намереваемся обсуждать связь между информацией и энтропией (и ее отрицанием — негэнтропией) и предпочитаем применять термин *энтропия* только в обычном термодинамическом смысле. Таким образом, только информация, возникающая в определенных физических задачах, т. е. связанная информация, будет представляться связанной с энтропией.

Для установления соотношения между связанной информацией и энтропией будем рассматривать равновероятные случаи как микросостояния. Тогда имеем следующую картину:

Связанная информация	Статистический вес	Энтропия
Начальное положение $I_{b_0} = 0$	P_0	$S_0 = k \ln P_0$
Конечное положение $I_{b_1} \neq 0$	$P_1 < P_0$	$S_1 = k \ln P_1$

(12.2)

Очевидно, что в этой схеме система не изолирована: энтропия убывает с получением информации, уменьшающей число микросостояний, и эта информация должна доставляться внешним агентом, энтропия которого будет возрастать. Связь между уменьшением энтропии системы и требуемой информацией очевидна, так как согласно (12.1)

$$I_{b_1} = k(\ln P_0 - \ln P_1) = S_0 - S_1, \tag{12.3}$$

или

$$S_1 = S_0 - I_{b_1}. \tag{12.3a}$$

Связанная информация появляется в качестве отрицательного слагаемого полной энтропии физической системы, и из этого мы заключаем, что

$$\text{связанная информация} = \text{убыванию энтропии } S = \text{увеличению негэнтропии } N, \tag{12.4}$$

где негэнтропия определена как отрицательная энтропия. Это положение составляет *негэнтропийный принцип информации*. Мы обсудим это соотношение на ряде примеров и покажем, как информация может быть превращена в негэнтропию, и обратно.

В случае свободной информации предпочтительно не говорить о связи между информацией и энтропией, так как соотношение между энтропией и числом случаев определено только, если рассматриваемые случаи являются микросостояниями физической системы.

2. Негэнтропийный принцип информации; обобщение принципа Карно

В ситуации, описанной (12.2), мы уменьшили энтропию системы при посредстве некоторого внешнего агента. Если мы изолируем систему, то согласно принципу Карно при любой естественной дальнейшей эволюции системы

$$\Delta S_1 \geqslant 0, \text{ или } \Delta(S_0 - I_{b1}) \geqslant 0. \tag{12.5}$$

Возрастание энтропии может произойти за счет либо слагаемого S_0, или $-I_{b_1}$, или того и другого. Когда система изолирована и предоставлена самой себе, то она, естественно, стремится к наиболее вероятной средней структуре, соответствующей физическим условиям, определяемым фиксированными значениями некоторых макроскопических параметров (объем, энергия, химическое строение и т. д.). Система может быть переопределена даже в своих *начальных* условиях, соответствующих некоторой энтропии S_0. Дополнительная информация I_{b_1}, требуемая формулами (12.2) и (12.3), соответствует большему переопределению. Мы можем уточнить много микроскопических параметров, изменение которых не может быть в подробностях прослежено. Например, мы можем построить систему из некоторых известных количеств химических веществ, но затем эти химикалии вступают в реакцию, и химический состав системы может с течением времени меняться, причем мы не в состоянии наблюдать детали химического строения. Подобного рода переопределение структуры может содержаться как в слагаемом S_0, так и в $-I_{b_1}$, и общее возрастание энтропии иногда затруднительно разделить между обоими слагаемыми.

Интересен случай, когда S_0 соответствует общей структуре, получаемой в результате свободной эволюции изолированной системы, так что S_0 не переопределена. В этой задаче S_0 остается постоянной и (12.5) сводится к

$$\Delta I_{b_1} \leqslant 0 \tag{12.5a}$$

и

$$N = -S, \quad \Delta N_1 \leqslant 0, \quad \Delta(N_0 + I_{b1}) \leqslant 0, \tag{12.6}$$

или $\Delta I_{b1} \leqslant 0$, если начальное состояние не переопределено. Пользуясь негэнтропией как мерой качества энергии (глава 9, раздел 2), получаем принцип Кельвина деградации энергии, выраженный символически равенством (12.6). Равенство получается для обратимых преобразований, тогда как необратимые преобразования дают неравенства.

Тот факт, что $\Delta I_{b1} \leqslant 0$ (см. (12.5а)), находится в согласии с полученным ранее результатом, относящимся к свободной информации, а именно:

$$\Delta I_f \leqslant 0. \tag{12.7}$$

Это соотношение получено при доказательстве того, что информация (свободная, так как символы предполагались абстрактными и не рассматривались как микросостояния физической системы) имеет максимум при равновероятных случаях.

Раздел 3 этой главы будет посвящен обсуждению различных примеров, показывающих, что оба вида информации могут быть превращены в негэнтропию и что информация, как связанная, так и свободная, может быть получена только за счет негэнтропии некоторой физической системы. Мы можем переписать (12.4) в форме обратимой реакции:

$$I \rightleftarrows N. \tag{12.8}$$

Это утверждение приводит к обобщению принципа Карно [1]), если мы включим оба вида информации. Если

$$\Delta I_f \leqslant 0, \quad \Delta N_1 \leqslant 0,$$

то

$$\Delta N_1 + \Delta I_f \leqslant 0,$$

или

$$\Delta (S_1 - I_f) = \Delta (S_0 - I_{b1} - I_f) = \Delta (S_0 - I) \geqslant 0 \tag{12.9}$$

для случая, где в рассмотрение входит как связанная, так и свободная информация. Принцип деградации выражается через негэнтропию неравенством

$$\Delta (N_0 + I) \leqslant 0. \tag{12.10}$$

Величина N_0 есть негэнтропия некоторой физической системы. Информационное слагаемое включает как связанную информацию, могущую увеличить N_0 (в примере, к которому относится (12.2)), I_{b1} увеличивает N_0 до N_1), так и свободную информацию. Сумма негэнтропии и информации может

[1]) L. Brillouin, J. Appl. Phys. 24, 1152 (1953).

оставаться неизменной при обратимом преобразовании, в противном случае она будет убывать. Эти неравенства относятся только к средним значениям. Во всех приложениях принципа Карно не исключена возможность возникновения случайных флуктуаций.

Рассмотрим пример с участием как связанной, так и свободной информации. Пусть индивидуум располагает информацией, которая свободна, пока остается в его уме, и не связана поэтому непосредственно с какой-либо физической системой. Проследим потери информации при передаче ее другому индивидууму и на каждом шагу будем устанавливать, теряется ли свободная или связанная информация.

A. Индивидуум располагает информацией, которая является свободной.

B. Он говорит об этом собеседнику, и информация теперь связана: она преобразована в звуковые волны, или электрические импульсы, или другие физические возмущения, которые могут применяться для связи. При кодировании для передачи возможны ошибки. Это будет потерей свободной информации, так как ошибки происходят до преобразования в физическое возмущение.

C. Вследствие искажений и теплового шума в системе связи некоторая доля информации может быть утрачена. Это есть потеря связанной информации.

D. Собеседник тугоух и не воспринимает часть слов. При этом теряется связанная информация, но информация, полученная собеседником, теперь свободна, так как она находится в его уме.

E. Спустя некоторое время собеседник забывает часть информации, и эта потеря есть потеря свободной информации [1]. На каждом этапе имеется потеря информации; в наиболее благоприятном случае потери отсутствовали бы. Ситуация

---

[1]) Этот пример только показывает шаткость определений *свободной* и *связанной* информации. Получается, что мозг человека нематериален (т. е. не является физической системой). Вся физическая концепция автора сильно страдает при этом. Между тем довольно ясно, например, что процесс запоминания есть процесс некоторого упорядочивания в физической системе, каковой является мозг, а процесс забывания есть результат постоянно действующих (и также физических) дезорганизующих факторов. (*Прим. перев.*)

сходна с той, которая может получиться, когда маклер передает по телефону информацию о биржевых курсах.

Нужно отметить, что потери на этапах B, D и E представляют собой пограничные случаи. Иногда затруднительно различить свободную и связанную информации, в частности, в случаях, в которых участвует человек-наблюдатель. Мы должны, конечно, игнорировать все человеческие оценки. Так, например, мы исключаем случай, когда ученый собирает данные, а затем на основе их изучения открывает научный закон. Наши определения здесь не применимы. Мы смогли построить удовлетворительную теорию информации только путем исключения процесса мышления. Можно, конечно, надеяться, что в дальнейшем окажется возможным расширить теорию и включить в нее исследование такого рода проблем.

Возвращаясь к формулам (12.6) и (12.10), мы можем резюмировать взаимосвязь между термодинамикой и теорией информации следующим образом:

$$\left.\begin{array}{l}\text{негэнтропия } N \text{ соответствует информации } I,\\ \text{температура } T \text{ означает тепловой шум, нарушающий передачу информации,}\\ \text{энергия сохраняет свое обычное значение,}\\ \Delta Q = T\Delta S \text{ есть тепло, участвующее в некотором процессе,}\\ \Delta W = T\Delta N = T\Delta I \text{ соответствует механической работе, которая может быть произведена.}\end{array}\right\} \quad (12.11)$$

При высокой температуре T шум сильнее и работа, необходимая для передачи определенной информации, больше, так как необходимо перекрыть помехи.

3. Несколько типичных физических примеров

Рассмотрим несколько классических задач физики, чтобы показать, как возникает связанная информация и какую роль она играет. Это будет подходящим случаем для рассмотрения некоторых старых споров о значении энтропии и для разъяснения некоторых трудных вопросов.

Возьмем случай идеального одноатомного газа, заключенного в теплоизолированный сосуд объема V.

Когда достигнуто равновесное состояние, энтропия равна[1])

$$S = kn\left\{\frac{5}{2} + \ln\left|g\frac{V}{n}\left(\frac{4\pi mE}{3h^2n}\right)^{3/2}\right|\right\}, \tag{12.12}$$

где n — число атомов газа, m — масса атома, k и h — постоянные Больцмана и Планка соответственно, E — полная энергия и g — число неразличимых состояний на основном уровне атома.

Если основное состояние не вырождено, то $g = 1$. Это есть случай атома без момента количества движения в основном состоянии ($j = 0$). Если имеется момент количества движения j (спин плюс орбитальный момент), то имеется $g_i = 2j + 1$ состояний, предполагаемых равновероятными. Основное состояние предполагается достаточно низким, а температура не слишком высокой, так что возбуждение на высшие энергетические уровни не может иметь места.

Теоретическая формула хорошо согласуется с экспериментальными данными, но многие теоретические члены, совершенно необходимые для логической полноты, в действительности слишком малы, чтобы их можно было наблюдать в большинстве термодинамических экспериментов[2]).

Перепишем (12.12) следующим образом:

$$\left.\begin{aligned}
&\text{для } g = 1: \quad \text{нет вырождения, } j = 0, \quad S_1,\\
&\text{для } g = 2: \quad j = \frac{1}{2}, \quad S_2 = S_1 + kn\ln 2,\\
&\text{для } g = 3: \quad j = 1, \quad S_3 = S_1 + kn\ln 3.
\end{aligned}\right\} \tag{12.13}$$

Случай $g = 2$ соответствует атомам со спином $1/2$ и с двумя ориентациями $\left(\pm\frac{1}{2}\right)$. При $g = 3$ мы имеем $j = 1$ и три ориентации (-1, 0, $+1$). Все ориентации равновероятны, если внешнее поле отсутствует; небольшое внешнее поле расщепляет энергетические уровни и устраняет вырождение.

---

[1]) См., например, J. E. Mayer, M. G. Mayer, Statistical mechanics, стр. 115, формула (5.19), а также стр. 132—140.

[2]) O. Stern, Rev. Mod. Phys. **21**, 534 (1949); E. Schrödinger, Proc. Roy. Irish Acad. **53A**, 189 (1950); K. Darrow, BSTJ **21**, 51 (1942); **22**, 108 (1943).

Разности между S_1, S_2 и S_3 очень малы и не могут непосредственно наблюдаться, так как мы не в состоянии изменить значение j для данного атома. Однако при радиоактивных превращениях мы можем наблюдать превращение атома с массой m и спином j в другой атом с массой m' и спином j', и в этом случае эти величины имели бы физический смысл.

Это рассуждение показывает, что можно говорить о различии в энтропии ΔS только тогда, когда преобразование физически осуществимо. В противном случае это не имеет практического смысла. Предположим, что мы имеем дополнительную информацию о состоянии газа; например, нам удалось узнать, что в некоторый предшествующий момент газ занимал меньший объем V'. Так обстояло бы дело, если бы газ содержался в сосуде V', который мы внезапно соединили бы с другим объемом V'':

$$V = V' + V'', \quad V' < V.$$

Начальная энтропия S' меньше энтропии S после расширения на величину, определяемую (12.12):

$$S - S' = kn(\ln V - \ln V') = I, \tag{12.14}$$

где I есть информация. Когда мы впускаем газ в объем V'', между обоими сосудами возникают колебания плотности, и постепенно устанавливается равновесное состояние с плотностью, однородной во всем объеме V. Возрастание энтропии и потеря информации происходят совместно. Мы можем сказать, что газ постепенно «забывает» информацию [1]).

В качестве другого примера рассмотрим диффузию газа. Чтобы упростить задачу, сохранив ее существенные черты, предположим, что имеются два газа с атомами одинаковой массы m при $g = 1$; n_1 атомов первого газа занимают первоначально объем V_1, n_2 атомов второго газа — объем V_2:

$$n_1 + n_2 = n, \quad V_1 + V_2 = V,$$

---

[1]) Дж. фон Нейман также рассматривает наблюдателя, забывающего информацию, и устанавливает, что этот процесс также означает возрастание энтропии. Здесь мы снова видим, что невозможно во всех случаях провести ясное различие между свободной и связанной информацией. Этим подчеркивается необходимость обобщения принципа Карно (см. раздел 2). Дальнейшее обсуждение этого вопроса см. у Неймана: J. von Neumann, Mathematische Grundlagen der Quanten-Mechanik, Springer, Berlin, 1932; см. также P. Jordan, Philosophy of Sci. **16**, 269 (1949).

$$n_1 = p_1 n, \quad n_2 = p_2 n, \quad V_1 = p_1 V, \quad V_2 = p_2 V, \quad p_1 + p_2 = 1. \tag{12.15}$$

Эти соотношения требуют, чтобы начальные концентрации составляющих и конечная концентрация были равны:

$$\frac{V_1}{n_1} = \frac{V_2}{n_2} = \frac{V}{n}.$$

Кроме того, мы предполагаем равнораспределение энергий:

$$\frac{E_1}{n_1} = \frac{E_2}{n_2} = \frac{E}{n}, \tag{12.16}$$

и все эти условия, вместе взятые, требуют, чтобы в начальном состоянии давления и температуры обоих газов не различались.

Введем для скобок в (12.12) обозначение

$$Q = \left(\frac{4\pi mE}{3h^2 n}\right)^{3/2}.$$

В силу наших предположений Q имеет одинаковое значение для обоих газов и сохраняет это значение на протяжении всего процесса смешивания газов. Отсюда имеем:

до смешивания

$$S_{12} = k(n_1 + n_2)\left(\frac{5}{2} + \ln\frac{V}{n} + \ln Q\right) = S, \tag{12.17}$$

после смешивания

$$\begin{aligned} S'_{12} &= kn_1\left(\frac{5}{2} + \ln\frac{V}{n_1} + \ln Q\right) + kn_2\left(\frac{5}{2} + \ln\frac{V}{n_2} + \ln Q\right) = \\ &= S_{12} + kn_1(\ln n - \ln n_1) + kn_2(\ln n - \ln n_2), \\ S'_{12} &= S_{12} - kn(p_1 \ln p_1 + p_2 \ln p_2). \end{aligned} \tag{12.18}$$

Необратимое возрастание энтропии равно

$$\Delta S' = S'_{12} - S_{12} = -kn(p_1 \ln p_1 + p_2 \ln p_2) > 0. \tag{12.19}$$

Эта величина положительна, так как для $p_1 < 1$ и $p_2 < 1$ логарифм отрицателен. Мы имели в исходном состоянии некоторую информацию I (связанную информацию) о системе:

$$I = -kn(p_1 \ln p_1 + p_2 \ln p_2), \tag{12.20}$$

и эта информация была потеряна в процессе диффузии газа. Можно сравнить (12.20) с формулой Шеннона (2.1) для информации. Газ в окончательно смешанном состоянии соответствует ситуации с двумя символами с априорными вероятностями p_1 и p_2. Роль символа здесь играет один из видов атомов. Выбор некоторой частной комбинации символов дает информацию $-k(p_1 \ln p_1 + p_2 \ln p_2)$ на символ (Шеннон). Начальное состояние газа соответствует отбору атомов 1 и помещению их в V_1, тогда как атомы 2 помещаются в V_2. Для этого требуется информация, выражаемая (12.20).

Рассмотрим теперь несколько видоизмененную задачу, предположив, что атомы имеют спин 1/2, причем атомы вида 1 имеют ориентацию $+\frac{1}{2}$, а вида 2: $-\frac{1}{2}$. Столкновения вызовут переходы между двумя видами атомов, и в окончательном состоянии будем иметь:

$$p_1 = \frac{1}{2}, \quad p_2 = \frac{1}{2} \tag{12.21}$$

с энтропией

$$S''_{12} = S_{12} + kn \ln 2. \tag{12.22}$$

Мы снова теряем информацию и увеличиваем энтропию

$$\Delta S'' = S''_{12} - S'_{12} = kn(\ln 2 + p_1 \ln p_1 + p_2 \ln p_2) > 0. \tag{12.23}$$

Окончательная энтропия S''_{12} равна энтропии S_2 (см. (12.13)) для газа со спином 1/2 с двумя равновероятными ориентациями. Наше рассуждение ясно показывает необходимость члена, зависящего от g, в формуле для энтропии и объясняет его физическое значение.

Возрастание энтропии соответствует потере информации, как и в предыдущих примерах. Сравним нашу формулу с результатами главы 2. Мы нашли (см. (2.18)), что выбор одной возможности из двух с априорными вероятностями p_1 и p_2 дает информацию

$$I_p = -K(p_1 \ln p_1 + p_2 \ln p_2) \tag{12.24}$$

на символ. Эта величина достигает максимума при

$$p_1 = p_2 = \frac{1}{2}, \quad I_{\max} = k \ln 2. \tag{12.25}$$

Если мы переходим от случая p_1, p_2 к случаю 1/2, 1/2, то информация изменяется на

$$\Delta I = I_{\max} - I_p = k(\ln 2 + p_1 \ln p_1 + p_2 \ln p_2) \quad (12.26)$$

на символ. Это в точности совпадает с (12.23).

4. Несколько общих замечаний

Из рассмотрения вышеприведенных задач можно сразу вывести несколько интересных заключений.

Получение информации о физической системе соответствует понижению энтропии этой системы. Низкая энтропия означает неустойчивое состояние, которое рано или поздно путем естественной эволюции перейдет в устойчивое состояние с высокой энтропией. Второе начало ничего не говорит нам о требуемом для этого времени, и поэтому мы не знаем, как долго система будет помнить информацию. Но если классическая термодинамика не в состоянии дать ответ на этот очень важный вопрос, то мы можем получить ответ из рассмотрения молекулярной или атомной модели с помощью кинетической теории: скорость затухания всякого рода волн, скорость диффузии, скорость протекания химических реакций и т. д. могут быть вычислены для подходящих моделей, и время установления может меняться от малых долей секунды до годов или столетий.

Такого рода задержки используются во всех практических приложениях: не так много времени требуется, чтобы серия импульсов (представляющих, например, точки и тире), передаваемая по электрическому кабелю, затухла и была забыта; однако этот короткий промежуток времени достаточно велик для передачи даже на большие расстояния, что и делает возможной дальнюю связь.

Система, способная удерживать информацию на некоторое время, может быть использована в качестве запоминающего устройства вычислительной машины. Примеры, обсужденные в предыдущем разделе, интересны не только в теоретическом отношении, они указывают также подход к практическим задачам. Рассмотрим, например, задачи о диффузии и распределении спинов (см. (12.15) — (12.26)). Они тесно связаны с магнитными запоминающими устройствами, так как спин всегда ассоциируется с магнитным моментом. Следовательно, ситуация вроде описываемой равенством (12.15) означает, что объем V' намагничен до насыщения в определенном направлении, тогда как объем V'' намагничен в противоположном направлении. Информация, запасенная в этой системе, соответствует уменьшению энтропии. Наше рассуждение показывает, как это состояние постепенно разрушается диффузией и столкновениями, увеличивающими энтропию и стирающими информацию.

Энтропия обычно описывается как мера беспорядочности в физической системе. Более точно утверждение, что *энтропия есть мера недостатка информации* о действительной структуре системы. Этот недостаток информации приводит к тому, что возможно большое разнообразие различных микроструктур, которые мы практически не в состоянии отличить друг от друга. Так как каждая из этих различных микроструктур действительно может быть реализована в некоторый данный момент, недостаток информации соответствует действительному беспорядку в скрытых степенях свободы.

Эта картина хорошо разъясняется на примере идеального газа. Когда мы задаем общее число атомов n, их массу m, кратность вырождения g и полную энергию E (см. (12.12)), мы не устанавливаем положений и скоростей каждого индивидуального атома. Это есть недостаток информации, приводящий к энтропии, выражаемой равенством (12.12). Так как мы не задаем положений и скоростей атомов, то мы не в состоянии различить две пробы газа, если различие состоит только в различии положений и скоростей атомов. Поэтому мы можем описывать ситуацию как беспорядочное движение атомов.

Источник современных представлений об энтропии и информации содержится в старой статье Силарда [1]), основополагающая работа которого в свое время не была хорошо понята. Связь между энтропией и информацией была вновь открыта Шенноном [2]), но он определял энтропию со знаком, обратным обычному термодинамическому определению.

---

[1]) L. Szilard, Z. Physik **53**, 840 (1929).

[2]) C. E. Shannon, W. Weaver, The Mathematical Theory of Communication, U. of Illinois Press, Urbana Ill., 1949.

Поэтому то, что Шеннон называет энтропией информации, в действительности означает негэнтропию. Это ясно видно на двух примерах (стр. 27 и 61 книги Шеннона), где Шеннон доказывает, что при некоторых необратимых процессах (необратимый преобразователь или фильтр) его энтропия информации убывает. Чтобы привести это в соответствие с нашими определениями, нужно изменить знак и читать *негэнтропия* [1]).

Связь между энтропией и информацией ясно изложена в недавних статьях Ротштейна [2]) в полном согласии с точкой зрения, представленной в настоящей главе.

---

[1]) Что-нибудь из двух! (*Прим. перев.*)

[2]) J. Rothstein, Science **114**, 171 (1951); Phys. Rev. **85**, 135 (1952).

ГЛАВА 13

ДЕМОН МАКСВЕЛЛА И НЕГЭНТРОПИЙНЫЙ ПРИНЦИП ИНФОРМАЦИИ

1. Демон Максвелла; исторический обзор

Проблема демона Максвелла представляет прекрасный случай для применения теории информации и ясно показывает связь между информацией и энтропией.

Сортирующий демон родился в 1871 г. и впервые появился в максвелловской «Теории теплоты» (стр. 328) в качестве «существа, способности которого настолько изощрены, что оно может следить за каждой молекулой на ее пути, и в состоянии делать то, что в настоящее время для нас невозможно... Предположим, что имеется сосуд, разделенный на две части A и B перегородкой с небольшим отверстием, и что существо, которое *может видеть отдельные молекулы*, открывает и закрывает это отверстие так, чтобы дать возможность только более быстрым молекулам перейти из A в B и только более медленным перейти из B в A. Это существо, таким образом, без затраты работы повысит температуру в B и понизит в A, вопреки второму началу термодинамики» [1]).

Этот парадокс рассматривало несколько поколений физиков. Смолуховский [2]) первый отметил возможное влияние броунова движения на дверцу, которое привело бы к ее случайному открыванию и закрыванию и серьезным образом нарушило бы действие всей системы. Это имело бы особое значение во всяком автоматическом устройстве, вроде

---

[1]) Весь этот отрывок приведен у Джинса (J. H. Jeans, Dynamical Theory of Gases, 3[d] ed., p. 183, Cambridge U. P., NY, 1921.

[2]) M. von Smoluchowski, Zs. Physik 13, 1069 (1912).

пружинного клапана, и полностью исключило бы возможность продолжительного действия подобного устройства. Смолуховский заключил, что броуново движение является только кажущимся нарушением второго начала вследствие случайности и непредсказуемости, а также кратковременности движения. Постоянное действие любой системы[1]) невозможно, и невозможность вечного двигателя второго рода остается несомненной; система может двигаться, но лишь беспорядочным образом.

Вопрос о действии пружинного клапана представляет интерес: если бы не его собственное броуново движение, пружинный клапан мог бы создать разность давлений между двумя сообщающимися сосудами. Этот случай сходен с задачей о выпрямителе, обсужденной в конце главы 11. Идеальный выпрямитель, воздействующий на отдельные электроны, выпрямил бы тепловое движение электронов, вопреки второму началу. Но идеального выпрямителя не существует, и единственный возможный результат состоит в нарушении симметрии теплового движения, как пояснено в главе 11.

В своей примечательной работе Силард[2]) впервые показал, что демон воздействует на информацию о деталях движения газа и фактически превращает информацию в негэнтропию.

Мы обсудим позднее некоторые интересные вопросы, поставленные Силардом.

Льюис[3]) рассматривал проблему, очень сходную с исследованной в предыдущей главе: разделение газов и диффузия. Он рассматривал энтропию, зависящую от смешивания или разделения газов, и пришел к заключениям, сходным с нашими.

Слэйтер[4]) поставил вопрос о том, в какой мере в проблеме может играть роль принцип неопределенности.

Максвеллов демон должен измерять одновременно положение и скорость данного атома. Обе величины не могут быть измерены одновременно с беспредельной точностью

---

[1]) Основанной на использовании броунова движения. (*Прим. перев.*)

[2]) L. Scilard, Zs. Physik **53**, 840 (1929).

[3]) G. N. Lewis, Zs. Physik **71**, 569 (1930).

[4]) J. C. Slater, Introduction to Chemical Physics, Mc Graw-Hill, NY, 1939.

вследствие известного ограничения

$$\Delta p \, \Delta q \geqslant h, \tag{13.1}$$

где $p = mv$. Эта неопределенность может играть роль для легких атомов (малая масса m) и при больших давлениях, когда положения атомов должны определяться с большой точностью. Но ограничения, накладываемые неопределенностью, не имеют значения для тяжелых атомов при низком давлении, как показано Димерсом [1]).

По существу, мы имеем дело с гораздо более фундаментальным вопросом, поставленным впервые Димерсом [1]) и Бриллюэном [2]):

Может ли демон действительно видеть отдельные атомы?

Мы предполагаем, что вся система изолирована и находится вначале при данной температуре T. Демон находится в замкнутом объеме в состоянии равновесия при постоянной температуре, где излучение должно быть излучением черного тела, и увидеть что-либо внутри черного тела невозможно. Повышение температуры бесполезно. При красной температуре излучение имеет максимум в красной области и имеет в точности одинаковую интенсивность вне зависимости от того, имеются ли в объеме миллионы молекул или их там нет вовсе. Не только интенсивность, но и флуктуации остаются неизменными. Демон будет воспринимать тепловое излучение и его флуктуации, но он никогда не увидит молекул.

Неудивительно, что Максвелл не подумал о включении излучения в систему, находящуюся в равновесии при температуре T. Излучение черного тела было вряд ли известно в 1871 г.; это было за тридцать лет до того, как термодинамика излучения была хорошо понята и появилась теория Планка.

Демон не может видеть молекулы, а стало быть, он не может управлять дверцей и не в состоянии нарушить второе начало. Демон не в состоянии видеть молекулы, но он мог бы попытаться обнаружить их каким-либо другим

---

[1]) P. Demers, Can. J. Research **22**, 27 (1944); **23**, 47 (1945).

[2]) L. Brillouin, Am. Scientist **37**, 554, см. стр. 575 (1949); **38**, 594 (1950); см. также J. Appl. Phys. **22**, 334 (1951).

методом: он может измерять силы Ван-дер-Ваальса, или поля, обусловленные электрическими диполями или магнитными моментами. Здесь имеются другие соображения: все эти поля создают короткодействующие силы, убывающие по меньшей мере как r^{-2}, где r — расстояние. Поэтому демон мог бы обнаружить молекулы только в непосредственной близости от стенки или дверцы. Это слишком поздно для того, чтобы можно было привести в действие дверцу, не совершая никакой работы. Те самые силы, на которых основывается демон для обнаружения молекул, будут также действовать на на дверцу или на клапан. Поэтому для передвижения дверцы потребуется работа, и задача сильно усложнится. Для предположений Максвелла существенно, что молекула может быть обнаружена задолго до того, как она достигает стенки на таком расстоянии, что всеми короткодействующими силами можно с уверенностью пренебречь.

Это можно сделать только при помощи какого-либо излучения, и световое излучение является простейшим для обсуждения.

2. Изгнание демона

Рассмотрим возможности демона. Мы можем снабдить его электрическим фонарем, так что он сможет увидеть молекулы. Фонарь есть источник излучения, не находящийся в равновесном состоянии. Он вводит в систему негэнтропию. Из этой негэнтропии демон получает информацию. Используя эту информацию, он действует дверцей и восстанавливает негэнтропию, завершая, таким образом, цикл:

$$\text{негэнтропия} \to \text{информация} \to \text{негэнтропия}. \qquad (13.2)$$

Димерс[1]) рассматривал процедуру в целом, не останавливаясь на промежуточном этапе, на котором демон получает информацию, используемую в последней операции.

Димерс показал на различных примерах, что потери негэнтропии фонаря больше, чем окончательный выигрыш. Другими словами, начальное увеличение энтропии больше, чем конечное уменьшение, и общий баланс удовлетворяет принципу Карно.

---

[1]) P. Demers, Can. J. Research **22**, 27 (1944); **23**, 47 (1945).

Силард отметил промежуточный этап с информацией. Его рассуждение о преобразовании информации в негэнтропию было очень тщательно, но он основывал исследование первого этапа на рассмотрении весьма искусственной модели.

Начнем с грубого рассуждения[1]), которое пояснит наиболее существенные черты явлений. Более тщательное исследование, использующее некоторые результаты Джейкобсона[2]), будет приведено ниже.

Для обсуждения баланса энтропии мы должны определить изолированную систему, к которой второе начало с уверенностью применимо. Наша система составлена из следующих частей:

1) заряженная батарея и электрическая лампочка;

2) газ при постоянной температуре T_0, заключенный в максвелловом замкнутом объеме, с перегородкой, разделяющей сосуд на две части, снабженной отверстием;

3) демон, оперирующий дверцей в отверстии. Вся система изолирована и замкнута[3]).

Батарея нагревает нить лампы до высокой температуры

$$T_1 > T_0. \tag{13.3}$$

Это условие необходимо для получения видимого света:

$$h\nu_1 > kT_0, \tag{13.4}$$

который может быть распознан на фоне излучения черного тела в замкнутом объеме при температуре T_0[4]). На протяжении опыта батарея дает полную энергию E, но не энтропию. Нить излучает энергию E и теряет энтропию. Изменение энтропии нити равно

$$S_f = -\frac{E}{T_1} \tag{13.5}$$

---

[1]) L. Brillouin, Am. Scientist **37**, 554, см. стр. 565 (1949); **38**, 594 (1950); см. также J. Appl. Phys. **22**, 334 (1951).

[2]) H. Jacobson, Trans. NY Acad. Sci. **14**, 6 (1951).

[3]) Мы можем заменить демона автоматическим устройством с «магическим глазом», открывающим дверцу в нужные моменты. Это лишь вопрос конструирования какого-либо хитроумного приспособления, не затрагивающий общих условий задачи.

[4]) См. (9.45). Условие (13.4) необходимо для того, чтобы сделать среднюю энергию $\overline{E}$ квантованного резонатора ν_1 существенно меньше, чем kT_0.

и равно негэнтропии, введенной в газ. Без вмешательства демона энергия E поглощается в газе при температуре T_0, и мы наблюдаем общее возрастание энтропии

$$S = \frac{E}{T_0} + S_f > 0. \tag{13.6}$$

Теперь рассмотрим работу демона. Он может обнаружить молекулу, если по меньшей мере один квант энергии $h\nu_1$ рассеян молекулой и поглощен глазом демона (или фотоэлементом, если он пользуется подобным устройством). Это означает в конечном счете увеличение энтропии:

$$S_d = \frac{h\nu_1}{T_0} = kb, \tag{13.7}$$

где

$$\frac{h\nu_1}{kT_0} = b > 1,$$

в соответствии с условием (13.4).

Раз информация получена, она может быть использована для уменьшения энтропии системы. Энтропия системы согласно формуле Больцмана равна

$$S_0 = k \ln P_0, \tag{13.8}$$

где P_0 означает статистический вес системы. После получения информации система определена более полно. P_0 уменьшается на величину p, и

$$P_1 = P_0 - p,$$

$$\Delta S_i = S - S_0 = k\Delta(\ln P) = -k\left(\frac{p}{P_0}\right). \tag{13.9}$$

Очевидно, что $p \ll P_0$ во всех практических случаях.

Общий баланс энтропии выражается соотношением

$$\Delta S_d + \Delta S_i = k\left[b - \left(\frac{p}{P_0}\right)\right] > 0, \tag{13.10}$$

так как $b > 1$ и $\frac{p}{P_0} < 1$. В конечном счете энтропия изолированной системы возрастает, как этого и требует второе начало. Все, что демон может сделать, это восстановить небольшую часть негэнтропии и использовать информацию для уменьшения деградации энергии.

На первом этапе процесса (см. (13.7)) мы имеем увеличение энтропии ΔS_d и, следовательно, изменение негэнтропии ΔN_d:

$$\Delta N_d = -kb < 0, \qquad (13.7\text{a})$$

т. е. уменьшение. Из этой потерянной негэнтропии некоторое количество превращается в информацию, и на последнем этапе процесса эта информация снова превращается в негэнтропию, получающую изменение

$$\Delta N_i = k\left(\frac{p}{P_0}\right) > 0, \qquad (13.9\text{a})$$

т. е. увеличение. Это подтверждает общую схему (13.2).

3. Обсуждение

Рассмотрим подробнее исходную задачу Максвелла. Мы можем предположить, что по прошествии некоторого времени демон сумел получить разность температур ΔT:

$$\left.\begin{aligned} T_B > T_A, \quad T_B &= T + \frac{1}{2}\Delta T, \\ T_B - T_A = \Delta T, \quad T_A &= T - \frac{1}{2}\Delta T. \end{aligned}\right\} \qquad (13.11)$$

Далее, демон выбирает быструю молекулу в A с кинетической энергией $\frac{3}{2}kT(1+\varepsilon_1)$ и направляет ее в B. Затем он выбирает медленную молекулу в B с кинетической энергией $\frac{3}{2}kT(1-\varepsilon_2)$ и дает ей проникнуть в B. Для того чтобы пронаблюдать эти две молекулы, демону требуется два световых кванта, и, следовательно, имеется увеличение энтропии (аналогично выраженному равенством (13.7)):

$$\Delta S_d = 2kb, \qquad (13.12)$$

где $b = h\nu_1/kT > 1$. Обмен молекулами приводит к переносу из A в B энергии

$$\Delta Q = \frac{3kT(\varepsilon_1 + \varepsilon_2)}{2}, \qquad (13.13)$$

что, на основании (13.11), соответствует уменьшению полной энтропии:

$$\Delta S_i = \Delta Q\left(\frac{1}{T_B} - \frac{1}{T_A}\right) = -\Delta Q\,\frac{\Delta T}{T^2} =$$
$$= -\frac{3}{2}\,k\,(\varepsilon_1 + \varepsilon_2)\,\frac{\Delta T}{T}. \quad (13.14)$$

Величины ε_1 и ε_2 обычно будут малы, но могут изредка достигать значений в несколько единиц. ΔT много меньше T, и поэтому

$$\Delta S_i = -\frac{3}{2}\,k\eta, \quad \eta \ll 1,$$

и

$$\Delta S_d + \Delta S_i \doteq k\left(2b - \frac{3}{2}\,\eta\right) > 0, \quad (13.15)$$

что удовлетворяет принципу Карно.

Димерс[1]) рассмотрел другой пример, предположив, что демон находится в замкнутом объеме при более низкой температуре $T_2 \ll T_0$. Можно представить себе обстановку опыта, в которой демон может различать кванты $h\nu$, испускаемые молекулами при температуре T_0. Тогда вместо (13.4) имеем условие

$$h\nu > kT_2, \quad T_2 < T_0, \quad (13.16)$$

и рассуждение продолжается аналогичным образом.

Применяется ли более высокая температура T_1 (см. (13.3)), или более низкая температура T_2 (см. (13.16)), всегда необходима некоторая разность температур; в противном случае демон не сможет действовать. А при наличии разности температур мы не нуждаемся в помощи демона. Любая тепловая машина будет работать. Например, если мы хотим создать разность температур и применяем лампочку при температуре T_1, то самый простой путь к цели состоит в том, чтобы нагревать излучением половину газа, в то время как вторая половина остается при температуре T_0. Эта процедура была бы значительно эффективнее, чем деятельность демона!

Наше первоначальное рассуждение было грубым, так как основывалось на предположениях, соответствующих обычным

[1]) P. Demers, Can. J. Research **22**, 27 (1944); **23**, 47 (1945).

условиям, а именно: $p \ll P_0$ в (13.9) или $\Delta T \ll T$ в (13.11). Мы исследуем проблему более тщательно в следующем разделе и покажем, что принцип Карно всегда удовлетворяется, даже в самых исключительных условиях.

Тем не менее мы обнаружили в (13.10) очень важный физический закон: всякое физическое измерение требует соответственного увеличения энтропии, и имеется нижний предел, ниже которого измерение становится невозможным. Этот предел соответствует изменению энтропии порядка k — постоянной Больцмана. Более тщательное исследование покажет, что точное значение предела есть $k \ln 2$ или приблизительно $0{,}7k$ на одну двоичную единицу полученной информации. Как говорит Гэйбор[1]: «Мы ничего не можем получить даром, даже наблюдения». Этот весьма важный закон является прямым следствием нашего общего негэнтропийного принципа информации и будет обсужден в дальнейшем. Удивительно, что столь общий результат до недавнего времени ускользал от внимания.

Общей особенностью всего этого рассуждения является то, что хотя квантовые условия были учтены, постоянная Планка h отсутствует в окончательном результате, зависящем только от постоянной Больцмана k. Это показывает, что результат не зависит от принципа неопределенности и рассуждение может быть проведено в рамках классической теории без введения квантовых условий.

4. Действие демона как преобразование информации в негэнтропию

Значение общей проблемы таково, что она требует более пристального и подробного изучения. В этом разделе мы дополним рассуждение, намеченное Джейкобсоном. Выберем пример более простой, чем придуманный Максвеллом. Вместо «температурного демона» будем рассматривать «демона давления». Действуя дверцей, он пропускает атомы из B в C и не позволяет атомам ускользнуть из C обратно в B. Через некоторое время он получит в C большее давление, а в B — меньшее. Эту операцию легче обсуждать, и «демон

---

[1]) D. Gabor, MIT Lectures, 1951.

давления» не нуждается в точном измерении скорости, так что мы можем не учитывать соотношение неопределенности.

Для устранения второстепенных параметров мы примем условия опыта, показанного на рис. 13.1. Сосуды B и C соединены трубкой с площадью сечения A, равной площади отверстия и дверцы. Луч света от внешнего источника проходит через стеклянные окошки, рассеянный атомами свет используется для их обнаружения. Обнаружение должно

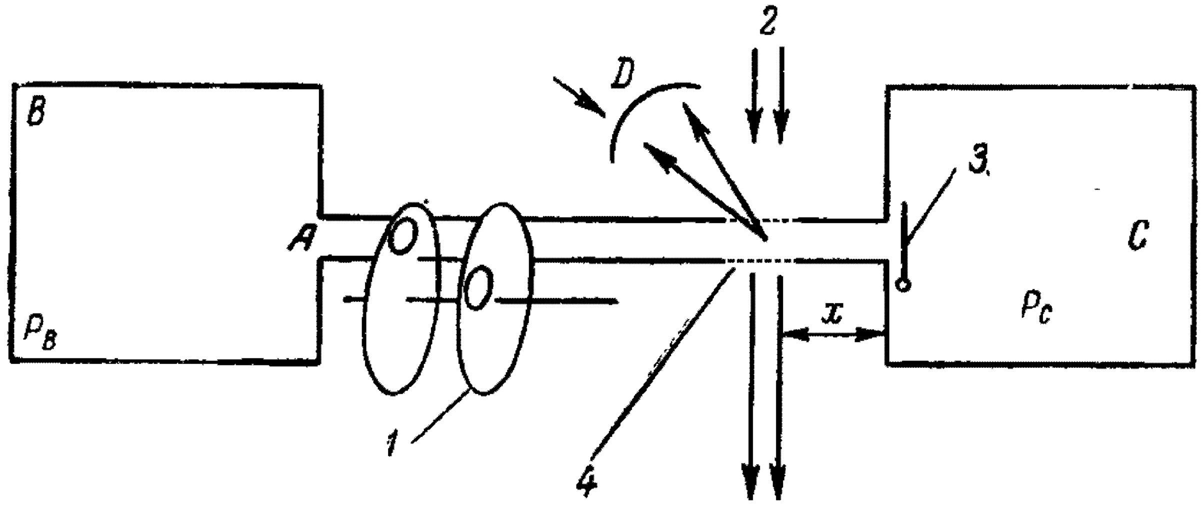

Рис. 13.1. Сосуды B и C содержат частицы под давлениями P_B и P_C соответственно. Частицы могут переходить из B в C вдоль трубы с площадью поперечного сечения A.

Два вращающихся затвора (*1*) с отверстиями площади A установлены так, что частица, проходящая через левый затвор, пройдет также и через правый затвор в том случае, если она имеет приблизительно определенную, наперед заданную скорость. Луч света (*2*) проходит через стеклянные окошки (*4*) в трубе для обнаружения частицы. Свет, рассеянный частицей, поглощается фотоэлементом D, который при этом открывает дверцу (*3*), так что частица может пройти в C.

происходить на некотором расстоянии x до дверцы, и скорости атомов должны быть известны хотя бы приблизительно для того, чтобы определить момент, когда нужно открыть дверцу. Можно, например, применить систему вращающихся затворов, пропускающих атомы, если они имеют скорость, приближающуюся к средней, и задерживающих атомы с отличными скоростями. В такого рода устройстве площадь A определяется точно, а скорость атомов — приблизительно.

Найдем количество информации, которое нужно демону для успешного действия. Рассмотрим длинный промежуток времени t, подразделенный на малые интервалы τ. Демон должен знать, открыть ли ему дверцу на протяжении интервала τ или держать ее закрытой. На каждом интервале τ требуемая информация такова: имеется ли частица (или частицы), которая должна попасть в дверцу площади A слева? Это — типичный вопрос с ответом «да» или «нет». Имеется априорная вероятность p_1 для «да» и p_2 для «нет». На протяжении длинного промежутка t число малых интервалов, в течение которых дверца будет открыта или закрыта, пропорционально в среднем соответствующим вероятностям:

$$t = G\tau, \quad N_1 = Gp_1, \quad N_2 = Gp_2, \quad p_1 + p_2 = 1. \tag{13.17}$$

Наши определения информации (глава 1) непосредственно применимы к такого рода задаче и дают информацию на интервал τ:

$$I_\tau = -k(p_1 \ln p_1 + p_2 \ln p_2) \tag{13.18}$$

(в термодинамических единицах).

Рассмотрим теперь нашу задачу с позиций кинетической теории. Среднее число молекул, попадающих в дверцу площади A за время τ (при температуре T_0) пропорционально плотности газа. Мы имеем в среднем:

$$\left.\begin{aligned} &u = bA\tau \text{ молекул, ударяющих слева,} \\ &v = ru = cA\tau \text{ молекул, ударяющих справа,} \end{aligned}\right\} \tag{13.19}$$

и

$$r = \frac{c}{b} = \frac{P_C}{P_B}, \tag{13.20}$$

где P_B и P_C — давления в B и C соответственно. Действие затвора, отбирающего движущиеся слева молекулы со средней скоростью, отсеет часть молекул, движущихся в этом направлении (вследствие отражения от второго затвора), и поэтому в (13.20) нужно внести поправку:

$$r \geqslant \frac{P_C}{P_B} = r_p. \tag{13.21}$$

Соотношение между p_1 и u таково:

$$p_1 = 1 - e^{-u}, \quad p_2 = e^{-u}, \tag{13.22}$$

и если $u \ll 1$, то $p_1 \approx u$ и $p_2 \approx 1 - u$. Вывод этого соотношения дан в приложении 1 в конце главы.

Из (13.22) получаем для информации на интервал τ:

$$I_\tau = k[-(1 - e^{-u}) \ln (1 - e^{-u}) + ue^{-u}], \tag{13.23}$$

которое при малых u сводится к

$$I_\tau \approx k(-u\ln u + u) = ku(1 - \ln u), \quad u \ll 1, \qquad (13.24)$$

если пренебречь высшими степенями u.

Теперь мы можем подсчитать убыль энтропии ($\Delta S < 0$) в результате действия дверцы. Число частиц в сосуде C возрастает за время τ на величину u (т. е. на число частиц, поступающих слева направо) за вычетом vp_1 (число частиц, уходящих справа налево), так как дверца открыта лишь долю p_1 всего времени:

$$\Delta n_\tau = u - rup_1 \approx u(1 - ru), \quad u \ll 1, \qquad (13.25)$$

и изменение энтропии равно

$$\Delta S_\tau = -k\ln\left(\frac{P_C}{P_B}\right)\Delta n_\tau = -ku(1 - ru)\ln r_p. \qquad (13.26)$$

Это соотношение получается следующим образом: перепишем выражение (12.12) для энтропии идеального одноатомного газа в виде

$$S = k\left[\alpha n + n\ln\left(\frac{V}{n}\right)\right],$$

где α зависит от массы молекул и от их средней энергии. В нашей задаче масса для обоих сосудов B и C одинакова, и если средняя энергия вначале одинакова, то она и останется одинаковой на протяжении всего опыта в силу способа, которым демон переводит молекулы из B в C[1]). Энтропия системы равна

$$S = S_B + S_C = k\left[\alpha(n_B + n_C) + n_B\ln\frac{V_B}{n_B} + n_C\ln\frac{V_C}{n_C}\right].$$

При переводе молекул из B в C объемы остаются неизменными и

$$\Delta n_C = \Delta n_\tau = -\Delta n_B,$$

[1]) Речь идет о том, что «демон давления» сортирует молекулы только по направлению их движения, а не по скорости. Поэтому хотя число молекул в C возрастает, но средняя энергия молекулы не меняется. (*Прим. перев.*)

так что

$$\Delta S = k\left[\alpha(\Delta n_B + \Delta n_C) + \Delta n_B \ln\frac{V_B}{n_B} + \Delta n_C \ln\frac{V_C}{n_C} + \right.$$
$$\left. + n_B\frac{n_B}{V_B}\left(-\frac{V_B}{n_B^2}\right)(\Delta n_B) + n_C\frac{n_C}{V_C}\left(-\frac{V_C}{n_C^2}\right)(\Delta n_C)\right] =$$
$$= -k\Delta n_\tau \ln\left(\frac{V_B}{n_B}\cdot\frac{n_C}{V_C}\right).$$

Но из известного соотношения кинетической теории

$$P_B = \frac{n_B \bar{v}_B^2 m}{3V_B}, \quad P_C = \frac{n_C \bar{v}_C^2 m}{3V_C},$$

и с учетом нашего предположения о равенстве средних энергий (13.26) получается непосредственно.

Формула (13.26) показывает, что уменьшение энтропии может быть получено только, если $ru < 1$, и что наиболее благоприятные условия состоят в том, что

$$ru \ll 1, \quad \text{откуда} \quad u \ll 1 \quad \text{и} \quad r_p \approx r. \qquad (13.27)$$

Это означает, что число молекул, ускользающих обратно, остается малым.

Мы в состоянии теперь подсчитать эффективность действия демона на втором этапе преобразования (13.2):

информация $\rightarrow$ негэнтропия.

Мы определяем эффективность отношением

$$\varepsilon_{II} = -\frac{\Delta S_\tau}{I_\tau} = (1 - ru)\frac{\ln r}{1 - \ln u} \qquad (13.28)$$

в соответствии с (13.24), (13.26) и (13.27). Возьмем

$$ru = e^{-\beta} \ll 1, \quad \beta > 1, \qquad (13.29)$$

и

$$u = \frac{e^{-\beta}}{r}, \quad \ln u = -\beta - \ln r,$$

так что

$$\varepsilon_{II} = \frac{(1 - e^{-\beta})\ln r}{1 + \beta + \ln r}. \qquad (13.30)$$

Эффективность всегда меньше единицы и может приближаться к единице, когда β велико, а r очень велико; в этом случае u наверняка очень мало, что подтверждает наши предположения (13.24) и (13.27). Таким образом, высокая эффективность на втором этапе преобразования получается при большой разности давлений между C и B.

5. Негэнтропия, требуемая при наблюдении

Нам предстоит рассмотреть первый этап и подсчитать количество негэнтропии, требуемое для получения информации.

Для того чтобы знать, попадет ли молекула в дверцу на протяжении некоторого интервала времени τ, мы должны использовать последовательность импульсов излучения, длительностью τ. Если молекула отсутствует, излучение проходит через прибор (см. рис. 13.1) без поглощения и

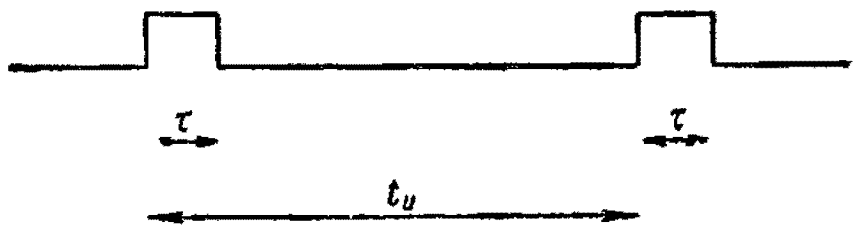

Рис. 13.2. Прямоугольные импульсы длительностью τ, повторяемые через интервалы t_u.

энтропия не возрастает. Если имеется молекула (одна или более), излучение рассеивается и поглощается фотоэлементом. При этом энтропия возрастает на E/T_0, если T_0 — температура системы, а E — поглощенная энергия.

Мы приняли (см. (13.19)), что среднее число молекул, попадающих в дверцу за время τ, равно u, и нашли вероятность p_1 (одного или более) попадания за это время (см. (13.22)). Это означает, что существует средний интервал t_u между молекулами (или группами молекул)

$$t_u = \frac{\tau}{p_1}. \tag{13.31}$$

Это есть средний интервал между последовательными моментами открывания дверцы. Фотоэлементом будет зарегистрирована последовательность импульсов длительностью τ со средним интервалом t_u, как показано на рис. 13.2. Сколько квантов нам требуется, и какова должна быть их величина для точной регистрации?

Одиночный импульс представляется интегралом Фурье (см. (8.25)—(8.28)) с непрерывным спектром, занимающим эффективную полосу до частоты

$$\nu_{\max} = \frac{1}{2\tau}. \tag{13.32}$$

Таким образом, импульс имеет бесконечное число составляющих и, следовательно, бесконечную энергию[1]).

Иное положение имеем для последовательности импульсов, повторяемых (в среднем) через интервал t_u. Рассмотрим

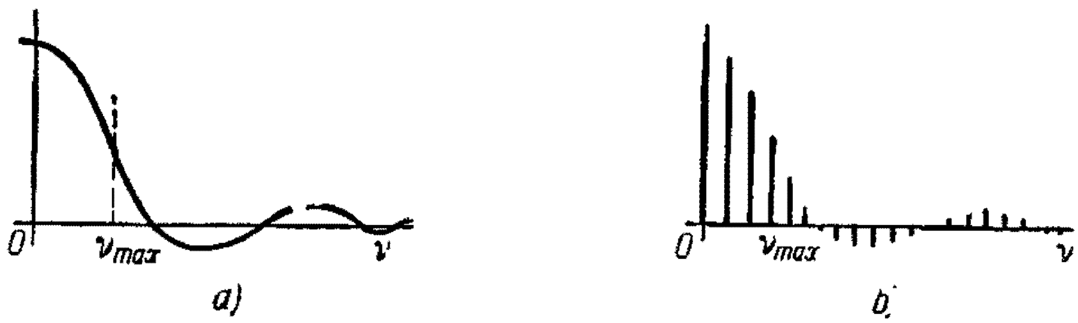

Рис. 13.3. *a*) непрерывный спектр одиночного импульса; *b*) дискретный спектр периодической последовательности импульсов. Огибающая совпадает с кривой непрерывного спектра.

задачу в том упрощающем предположении, что интервал повторения есть в точности t_u, что делает процесс периодическим с периодом t_u[2]). Такая последовательность импульсов может быть разложена в ряд Фурье с дискретным спектром, содержащим только определенные частоты, как показано на рис. 13.3. Эти частоты равны 0, ν_u, $2\nu_u$, $3\nu_u$, вплоть до $n\nu_u$, где

$$\nu_u = \frac{1}{t_u} = \frac{p_1}{\tau}. \tag{13.33}$$

[1]) Это совершенно непонятное утверждение. Конечный импульс имеет конечную энергию. Что же касается спектральных составляющих, то хотя в сплошном спектре их и бесконечное количество, но они бесконечно малы (амплитуда равна спектральной плотности, умноженной на $d\nu$. (*Прим. перев.*)

[2]) Более общее рассуждение дано в приложении II в конце главы. Замена одиночного импульса (интеграл Фурье) последовательностью импульсов (ряд Фурье) является обращением рассуждения главы 8, раздела 3, где вводится интеграл Фурье (рис. 8.3 и 8.4).

Наибольшая частота должна быть достаточно высока, чтобы можно было построить короткие импульсы длительностью τ. Из этого условия вытекает ограничение, сходное с (13.32), или приблизительно

$$\nu_{\max} = n\nu_u \approx \frac{1}{2\tau}. \tag{13.34}$$

Здесь нужно заметить, что наша упрощенная периодическая модель применима лишь до тех пор, пока t_u больше 2τ, так что пауза $(t_u - \tau)$ не короче импульса τ. В общей сложности мы получаем $n+1$ составляющих (включая нулевую частоту)

$$n \approx \frac{1}{2\tau\nu_u} = \frac{1}{2p_1} = \nu_{\max}\, t_u. \tag{13.34a}$$

Рассмотрим пример, положив амплитуду для нулевой частоты равной $A/2$ и равной A для всех остальных частот. Положим, далее, что каждая гармоника имеет только косинусную составляющую. Получим сигнал вида

$$f(t) = A\left(\frac{1}{2} + \cos\Phi + \cos 2\Phi + \ldots + \cos n\Phi\right) = \\ = A\,\frac{\sin\left(n+\frac{1}{2}\right)\Phi}{2\sin\frac{1}{2}\Phi} \tag{13.35}$$

в соответствии с тождеством Лагранжа, которым мы уже пользовались (см. (8.61)), где

$$\Phi = 2\pi\nu_u t = \frac{2\pi t}{t_u}.$$

Мы получили, таким образом, последовательность сглаженных (уже не прямоугольных) импульсов, показанных на рис. 13.4. Их форма выражается равенством (13.35) с максимальным значением $\left(n+\frac{1}{2}\right)A$ при $t=0$ и с нулями по обе стороны при

$$\pm\left(n+\frac{1}{2}\right)\Phi_1 = \pi, \quad \pm t_1 = \frac{t_u}{(2n+1)}.$$

Полный промежуток времени между двумя первыми нулями есть $2t_1$, но эффективная или *номинальная* длительность (см. главу 8, разделы 4 и 5) импульса соответствует приблизительно половине этого интервала. Итак, мы принимаем определение

$$\tau = t_1 = \frac{t_u}{(2n+1)}, \text{ или } n + \frac{1}{2} = \frac{t_u}{2\tau} \tag{13.36}$$

— соотношение, заменяющее наше приближенное условие (13.34 а). Период повторения импульсов есть t_u.

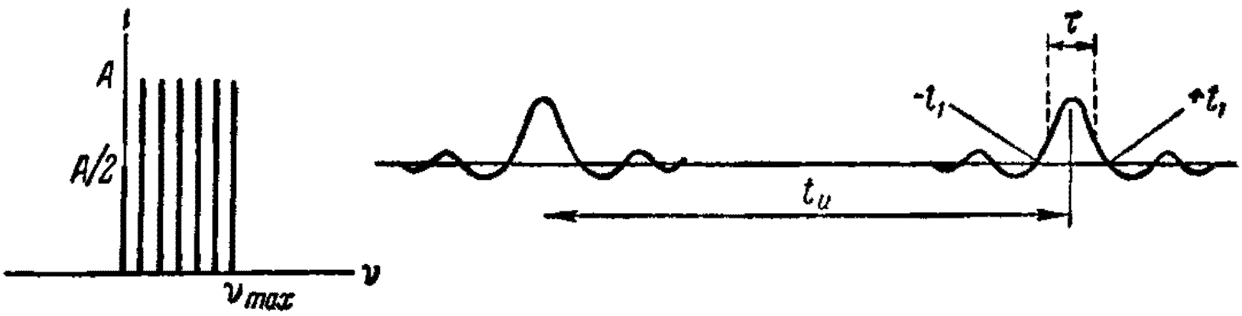

Рис. 13.4. Последовательность сглаженных импульсов эффективной длительности τ, повторяемых с интервалами t_u (см. (13.35)), и соответствующий спектр.

Подсчитаем среднюю энергию для последовательности импульсов с помощью равенства Парсеваля (см. (8.12)):

$$\overline{f^2} = \frac{1}{t_u}\int_0^{t_u} f^2\,dt = \frac{1}{4}A^2 + \sum_1^n \frac{1}{2}A^2 = \frac{1}{2}A^2\left(n+\frac{1}{2}\right).$$

Составляющая нулевой частоты в (13.35) дает слагаемое, пропорциональное ее квадрату, тогда как члены $A^2\cos^2 m\Phi$ дают $\frac{1}{2}A^2$ каждый. Средняя мощность (энергия в секунду) последовательности импульсов выражается как $b\overline{f^2}$, где значение численного множителя b зависит от выбора системы единиц:

$$\overline{P} = b\overline{f^2} = \left(n+\frac{1}{2}\right)\frac{1}{2}bA^2. \tag{13.37}$$

Почти вся энергия сосредоточена в коротких импульсах τ. Это легко проверить: наибольшее значение $\left(n+\frac{1}{2}\right)A$ при $t=0$ соответствует согласно (13.37) пиковой мощности

$$P_M = b\left(n+\frac{1}{2}\right)^2 A^2 = 2\left(n+\frac{1}{2}\right)\overline{P},$$

и из (13.36) находим:

$$P_M\tau = \bar{P}t_u, \tag{13.37a}$$

что доказывает, что энергия в импульсе равна полной энергии за время t_u.

Теперь мы должны исследовать тепловое возбуждение в различных степенях свободы, и мы предположим сперва, что все частоты сравнительно низки, т. е.

$$h\nu < kT,$$

где T — температура системы. Мы имеем (см. (13.34)) $n+1$ степеней свободы на интервал t_u, что согласуется с рассуждением главы 8 (см. (8.57)). Составляющая нулевой частоты имеет среднюю энергию $\frac{1}{2}kT$, а n колебательных членов имеют среднюю энергию kT каждый. В общем, средняя тепловая энергия за интервал t_u равна

$$\bar{E}_T = \left(n + \frac{1}{2}\right)kT = \frac{kTt_u}{2\tau} \tag{13.38}$$

в соответствии с (13.36). Эта энергия равномерно распределена по промежутку t_u.

Вероятность поглощения света фотоэлементами пропорциональна энергии, имеющейся на протяжении каждого интервала: на протяжении импульса τ ее величина согласно (13.37а) и (13.38) составляет:

$$E_1 = P_M\tau + \bar{E}_T\frac{\tau}{t_u} = \bar{P}t_u + \frac{1}{2}kT.$$

На протяжении паузы $(t_u - \tau)$ мы имеем:

$$E_2 = \bar{E}_T\frac{(t_u - \tau)}{t_u} = \frac{1}{2}kT\left(\frac{t_u}{\tau} - 1\right).$$

Это слагаемое E_2 дает ложное поглощение, приводящее к нежелательному нерегулярному открыванию дверцы. Для уменьшения такого рода ошибок в работе системы положим:

$$E_1 = \alpha E_2, \quad \alpha \gg 1, \tag{13.39}$$

откуда

$$\bar{P}t_u + \frac{1}{2}kT = \frac{1}{2}\alpha kT\left(\frac{t_u}{\tau} - 1\right).$$

Так как α велико, а τ/t_u мало, это условие сводится к

$$\bar{P} \approx \frac{\alpha k T}{2\tau}, \tag{13.40}$$

и $\bar{P} t_u$ выражает энергию излучения, поглощенную фотоэлементом за время t_u. Средняя энергия, поглощенная за время τ, есть $\bar{P}\tau$, и соответствующее среднее увеличение энтропии на интервал τ составляет:

$$\Delta S_\tau = \frac{\bar{P}\tau}{T} = \frac{1}{2}\alpha k. \tag{13.41}$$

Все сделанные приближения основаны на предположении

$$t_u \gg \tau, \quad u \ll 1.$$

Мы можем теперь обсудить эффективность на первом этапе действия демона, когда негэнтропия превращается в информацию. Пользуясь (13.18) и (13.24) для информации на интервал τ, получаем эффективность

$$\varepsilon_{\mathrm{I}} = \frac{I_\tau}{\Delta S_\tau} \leqslant \frac{2(-p_1 \ln p_1 - p_2 \ln p_2)}{\alpha} \approx \frac{2u(1-\ln u)}{\alpha}. \tag{13.42}$$

Это выражение справедливо только для очень малых значений u. Случай, когда u близко к единице, труден для исследования. Возможно, что в этой области значений эффективность выше. Для меньших значений u эффективность уменьшается.

Подытоживая результаты, мы видим, что

ε_{I} — эффективность на первом этапе — получается наибольшей при $u \approx 1$,

$\varepsilon_{\mathrm{II}}$ — эффективность на втором этапе — получается наибольшей при очень малых u.

Подсчитаем общую эффективность; с помощью (13.28) и (13.42) имеем:

$$\varepsilon_0 = \varepsilon_{\mathrm{I}} \cdot \varepsilon_{\mathrm{II}} = \frac{2n(1-ru)\ln r}{\alpha}. \tag{13.43}$$

Как и в (13.29), берем

$$ru = e^{-\beta} = \eta < 1$$

и находим:

$$\varepsilon_0 = \frac{2\eta(1-\eta)}{\alpha} \frac{\ln r}{r} \leqslant \frac{\ln r}{2r\alpha}, \qquad (13.44)$$

так как оптимальное значение η есть $1/2$. Последнее выражение имеет при $r=e$ максимум, равный $u=\eta/r=e/2$; это — небольшая величина, для которой все наши приближения действительны:

$$\varepsilon_{0\,\max} = \frac{1}{(2e\alpha)} \approx \frac{1}{(5{,}44\alpha)}, \quad \alpha \gg 1.$$

Общая эффективность падает до нуля при $r=1$ (равные давления в B и C), а также при очень большом r (высокое давление в C). Мы показали, таким образом, что общая эффективность всегда меньше единицы и что как ε_{I}, так и $\varepsilon_{\mathrm{II}}$ также меньше единицы, даже при весьма искусственных условиях, когда отношение давлений r велико.

6. Задача Силарда: полностью информированная тепловая машина

Силард опубликовал в 1929 г.[1]) очень примечательную работу, посвященную проблеме демона Максвелла, и впервые обнаружил связь между информацией и энтропией. Мы исследуем некоторые интересные вопросы, поставленные Силардом, и обсудим недавнюю работу Гэйбора[2]), посвященную этой же проблеме.

Силард рассматривает следующий случай: замкнутый цилиндр объема V может быть разделен на две части V_1 и V_2 путем вдвигания определенным образом расположенной заслонки. Цилиндр содержит только одну молекулу. В тот момент, когда наблюдатель вдвигает заслонку, он может так или иначе знать, находится молекула в V_1 или в V_2. Положим, что молекула находится в V_1. Тогда заслонка передвигается, как поршень, вдоль цилиндра и объем V_1 медленно расширяется до исходной величины V, причем вся система поддерживается термостатом при постоянной температуре T. Молекула может удариться о поршень много раз на протяжении этой медленной операции, и результатом

---

[1]) L. Scilard, Zs. Physik **53**, 840 (1929).

[2]) D. Gabor, MIT Lectures, 1951.

этих соударений является среднее давление, сходное с давлением идеального газа. При этот совершается некоторая работа. Затем заслонка удаляется вбок, возвращается в исходное положение, и операция может быть повторена. Эта система дает механическую работу. Она использует только температуру, но требует информации о положении молекулы, и мы должны выяснить, как связана эта информация с изменениями энтропии. Наше точное определение информации и результаты, полученные в предыдущих примерах, позволяют нам упростить и уточнить рассуждения.

В схеме Силарда этапы следуют в таком порядке:

А. Заслонка вдвигается в некотором положении.

В. Выясняется, где находится молекула — в V_1 или в V_2.

С. В соответствии с этим поршень (заслонка) движется вверх или вниз.

Первый этап А не требует обсуждения. Что касается В, то мы должны придумать экспериментальное устройство для определения местонахождения молекулы. Мы можем использовать луч света B_1, проходящий через V_1, и другой луч B_2, проходящий через V_2. Два фотоэлемента C_1 и C_2 могут воспринимать рассеянный свет из V_1 и V_2 соответственно. Включим сперва луч B_1. Если C_1 получает рассеянный свет, это означает, что молекула находится в V_1. Если C_1 не обнаруживает света, то еще ничего не известно, так как молекула может не находиться в V_1, молекула может не рассеять свет, рассеянный световой квант может не попасть в C_1. Затем мы повторим опыт рассеяния со вторым лучом B_2 в V_2. В конце концов квант $h\nu$ рассеивается либо от V_1, либо от V_2, и мы определяем местоположение молекулы. Поглощение кванта $h\nu$ в одном из фотоэлементов соответствует увеличению энтропии на

$$\Delta S = \frac{h\nu}{T} \geqslant k \tag{13.45}$$

в соответствии с предыдущим рассуждением (см. (13.7)). Более подробное исследование (см. (14.17)) несколько понижает предел и дает:

$$\Delta S \geqslant k \ln 2 = 0{,}7k. \tag{13.45a}$$

Подсчитаем теперь полученную информацию. Вероятность p_1 застать молекулу в V_1 равна, очевидно, V_1/V, а вероятность того, что молекула находится в V_2, равна V_2/V:

$$p_1 = \frac{V_1}{V}, \quad p_2 = \frac{V_2}{V}, \quad p_1 + p_2 = 1.$$

Рассмотрим эти два случая порознь. С помощью формул (1.5) и (1.6) можно подсчитать:

$$\left.\begin{aligned} &\text{молекула в } V_1:\\ &\text{информация } i_1 = k\ln\left(\frac{V}{V_1}\right) = -k\ln p_1,\\ &\text{молекула в } V_2:\\ &\text{информация } i_2 = k\ln\left(\frac{V}{V_2}\right) = -k\ln p_2, \end{aligned}\right\} \qquad (13.46)$$

и средняя информация на одну операцию равна

$$\bar{I} = p_1 i_1 + p_2 i_2 = -k(p_1\ln p_1 + p_2\ln p_2) > 0, \qquad (13.46a)$$

что соответствует формуле Шеннона (см. (2.1)). Каждая отдельная операция может дать информацию больше или меньше увеличения энтропии ΔS, потребовавшегося для получения этой информации, но *средняя информация* (13.46а) меньше, чем *среднее увеличение энтропии* на одну операцию:

$$\Delta\bar{S} \geqslant \bar{I}. \qquad (13.47)$$

Знак равенства получается при равенстве объемов V_1 и V_2:

$$V_1 = V_2 = \frac{1}{2}V, \quad p_1 = p_2 = \frac{1}{2}, \quad \bar{I} = k\ln 2,$$

что соответствует наибольшей средней информации. Мы можем иметь флуктуации информации, полученной при отдельных операциях. Однако в длинной серии испытаний средняя информация меньше, чем уплаченная за нее цена, выраженная через негэнтропию ΔN. Обобщенный принцип Карно (глава 13) удовлетворяется в среднем с положительными и отрицательными флуктуациями

$$\Delta N = -\Delta S, \quad \Delta(N + I) \leqslant 0. \qquad (13.47a)$$

Рассмотрим теперь изменение энтропии газа и последний этап C действия машины. Энтропия идеального газа равна (см. (12.12)):

$$S = S_0 + k \ln V.$$

Когда заслонка вдвигается, мы получаем некоторое уменьшение энтропии:

$$\left.\begin{aligned} &\text{либо } \Delta\sigma_1 = k \ln\left(\frac{V_1}{V_0}\right) = k \ln p_1 < 0, \\ &\qquad \text{если молекула находится в } V_1, \\ &\text{или } \Delta\sigma_2 = k \ln p_2 < 0, \\ &\qquad \text{если молекула находится в } V_2. \end{aligned}\right\} \tag{13.48}$$

Вероятности, соответствующие этим двум возможностям, равны p_1 и p_2. Заметим, что уменьшение энтропии ($-\Delta\sigma_1$ или $-\Delta\sigma_2$) всегда в точности равно получаемой в данном случае информации (i_1 или i_2), так как (13.46) и (13.48) совпадают, за исключением знака. Средняя энтропия газа убывает за одну операцию на

$$\overline{\Delta\sigma} = p_1 \Delta\sigma_1 + p_2 \Delta\sigma_2 = -I. \tag{13.49}$$

Когда мы двигаем поршень и увеличиваем объем до его первоначальной величины V, мы восстанавливаем исходное значение энтропии. Совершенная работа W равна теплу Q, полученному из резервуара тепла (в предположении медленного обратимого процесса); применяя закон Бойля к одной молекуле, для случая 1 (молекула находится в V_1) имеем:

$$W = Q = \int_{V_1}^{V} p\,dV = kT \int_{V_1}^{V} \frac{dV}{V} = kT \ln\left(\frac{V}{V_1}\right), \quad pV = kT.$$

Увеличение энтропии на протяжении процесса расширения составляет:

$$\frac{Q}{T} = k \ln\left(\frac{V}{V_1}\right) = -\Delta\sigma_1 > 0;$$

аналогично обстоит дело, когда молекула находится в V_2.

Последняя операция подтверждает справедливость равенств (13.48) и показывает, каким образом получается первоначальное уменьшение энтропии. В общем, за полный цикл действия получается (см. (13.47)) возрастание энтропии на $\Delta S - I$.

Таким образом, мы показали на этом примере, что информация соответствует негэнтропии (см. (13.47) и (13.49)), но мы должны были учитывать флуктуации, так как предполагалось, что заслонка вначале вдвигается без предварительного знания местонахождения молекулы.

7. Рассуждение Гэйбора

Гэйбор[1]) исследовал сходную задачу, но с отличным способом действия машины. Изменяя порядок первых двух операций А и В, он предлагает действовать следующим образом:

А. Убедиться, что молекула находится в V_1 (по Силарду — этап В).

В. Вдвинуть заслонку (по Силарду — этап А).

С. Двигать поршень.

Гэйбор предполагает, что для получения информации применяется луч света, освещающий объем V_1. Эскиз всего устройства машины дан на рис. 13.5. Единственная молекула движется в цилиндре V, поддерживаемом при температуре T. Нижняя часть цилиндра (объем V_1) имеет прозрачные стенки и освещается лучом от нагретой нити, излучающей определенную частоту ν. Система зеркал и линз поддерживает циркуляцию света без поглощения до тех пор, пока некоторое количество света не будет рассеяно молекулой и поглощено фотоэлементом. Когда это случится, механизм приводится в действие посредством реле: поршень, лишенный трения, вводится в цилиндр; два зеркала, скользя вниз, перекрывают световой луч. Молекула постепенно поднимает поршень и совершает механическую работу.

Изменение энтропии при введении поршня есть убывание:

$$\Delta S_1 = k \ln\left(\frac{V_1}{V}\right) = k \ln p_1 < 0, \tag{13.50}$$

а при расширении происходит увеличение энтропии, восстанавливающее ее первоначальную величину. Когда расширение закончено, поршень выдвигается и возвращается в исходное

[1]) См. предыдущую ссылку.

положение (эта часть механизма на рис. 13.5 не показана). Уменьшение энтропии (13.50) может быть сделано сколь угодно большим путем увеличения объема V. Трудность

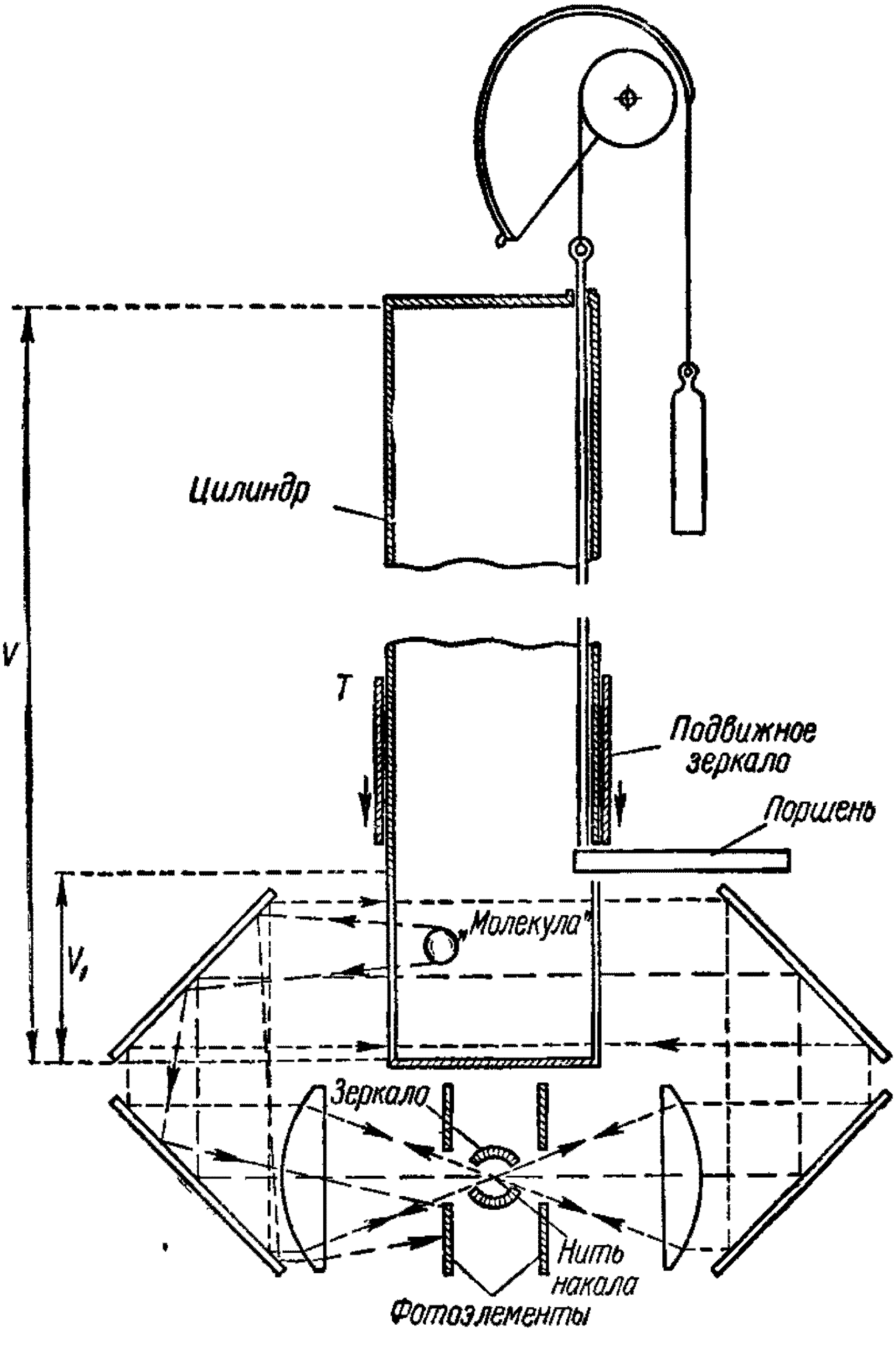

Рис. 13.5. Машина Гэйбора.

Свет циркулирует через V_1, пока не появится молекула, которая рассеет фотон в один из фотоэлементов. Когда это случится, реле вдвинет поршень в цилиндр, подвижные зеркала перекроют световой луч, и молекула поднимет поршень.

состоит в доказательстве того, что количество света, требуемое для наблюдения, также возрастает с увеличением V таким образом, что первоначальное увеличение энтропии всегда больше, чем $|\Delta S_1|$.

Рассмотрим первый этап действия, когда мы ожидаем, что молекула вступит в объем V_1 и рассеет свет. Среднее время ожидания t_u пропорционально V, а вспышка рассеянного света имеет длительность τ, пропорциональную V_1. В общем, наше наблюдение должно дать двоякого рода информацию:

А. Молекула находится в V_1 вместо V.

В. Это происходит на протяжении интервала τ всего времени t_u.

Ситуация отличается от той, которую мы имели в опыте рис. 13.1, где положение определялось геометрией прибора, и единственное требуемое наблюдение относилось ко времени. Мы можем подсчитать общее количество информации с помощью (12.1):

$$I = k \ln\left(\frac{P_0}{P_1}\right),$$

где P_0 — число равновероятных возможностей в общем случае, P_1 — число равновероятных возможностей, когда дана информация I. Здесь мы имеем, очевидно,

$$\frac{P_0}{P_1} = \frac{t_u}{\tau}\left(\frac{V}{V_1}\right) = \left(\frac{V}{V_1}\right)^2, \quad \text{так как} \quad \frac{t_u}{\tau} = \frac{V}{V_1}. \tag{13.51}$$

Следовательно, количество информации в такого рода опыте равно

$$I = k \ln\left(\frac{V}{V_1}\right)^2 = 2k \ln\left(\frac{V}{V_1}\right) = -2k \ln p_1. \tag{13.52}$$

Информация I вдвое больше первоначального увеличения энтропии ΔS_1 (см. (13.50)), определяющего негэнтропию, получаемую с помощью информации I. Эффективность преобразования информации в негэнтропию равна $1/2$.

Подсчитаем теперь начальное увеличение энтропии, требуемое для получения этой информации. Во всем объеме полной длины L (пропорциональной V) мы должны образовать луч света, покрывающий только длину L_1 (пропорциональную V_1). Это можно сделать надлежащим наложением некоторого числа собственных колебаний длины L. Эти собственные колебания имеют длины волн

$$2L, \quad L, \quad \frac{2L}{3}, \quad \ldots, \quad \frac{2L}{M},$$

и кратчайшая длина волны должна быть порядка $2L_1$ для того, чтобы длина L_1 отчетливо определилась. Отсюда имеем для числа собственных колебаний в объеме

$$M = \frac{L}{L_1} = \frac{V}{V_1}. \tag{13.53}$$

Для каждого из этих колебаний мы имеем вспышки τ с повторением через интервал t_u. Ситуация аналогична обсужденной в разделе 5 (см. (13.39) и (13.42)) и требует увеличения энтропии

$$\frac{\alpha k t_u}{2\tau}, \quad \text{где} \quad t_u \gg \tau \quad \text{и} \quad \alpha \gg 1,$$

на каждый вид колебания. В общем, для полной энтропийной стоимости наблюдения получаем:

$$\Delta S = \frac{M \alpha k t_u}{2\tau} = \frac{\alpha k}{2}\left(\frac{V}{V_1}\right)^2. \tag{13.54}$$

Сравним этот результат с формулой (13.52) для информации:

$$\Delta S - I = k\left(\frac{\alpha F}{2} - \ln F\right) > 0, \quad \alpha \gg 1, \tag{13.55}$$

где

$$F = \left(\frac{V}{V_1}\right)^2 = \left(\frac{1}{p_1}\right)^2.$$

Это выражение всегда положительно. Когда p_1 неограниченно убывает, F стремится к бесконечности, а следовательно, и скобка в (13.55). В этом заключается объяснение задачи Силарда. Дальнейшие подробности даны в главе 15.

Приложение I

Для доказательства соотношения (13.22) воспользуемся рассуждением Джейкобсона. Возьмем очень короткий временной интервал θ, настолько малый, что на его протяжении не более чем одна молекула может попасть в площадь A:

вероятность одного попадания: $\alpha\theta \ll 1$,
вероятность отсутствия попадания: $1 - \alpha\theta$.

Рассмотрим теперь интервал $\tau = m\theta$; будем иметь

$\pi_0 = (1-\alpha\theta)^m$ — вероятность того, что на протяжении τ не будет попадания,

$\pi_1 = m\alpha\theta(1-\alpha\theta)^{m-1}$ — вероятность одного попадания на протяжении какого-либо из m интервалов θ,

$\pi_2 = \dfrac{m(m-1)}{2}(\alpha\theta)^2(1-\alpha\theta)^{m-2}$ — вероятность двух попаданий за время τ,

$\pi_n = \dfrac{m!}{n!(m-n)!}(\alpha\theta)^n(1-\alpha\theta)^{m-n}$ — вероятность n попаданий за время τ.

Очевидно,

$$\pi_0 + \pi_1 + \ldots + \pi_n + \ldots + \pi_m = [\alpha\theta + (1-\alpha\theta)]^m = 1.$$

Когда θ становится очень малым, а m — очень большим,

$$\pi_0 = (1-\alpha\theta)^m = (1-\alpha\theta)^{\frac{\alpha\tau}{\alpha\theta}} \to e^{-\alpha\tau}.$$

Но π_0 совпадает с нашим прежним p_2 (см. (13.22)), и $\alpha\tau = u$, что и подтверждает формулу (13.22). Кроме того,

$$p_1 = \pi_1 + \pi_2 + \ldots + \pi_m = 1 - e^{-\alpha\tau} = 1 - e^{-u}$$

есть вероятность одного или более попаданий.

Приложение II

Мы можем обобщить рассуждение, связанное с (13.33). Вместо предположения о равноотстоящих импульсах, положим, что m импульсов расположены случайно на протяжении очень большого времени t:

$$t = \frac{m\tau}{p_1},$$

где $m \gg 1$, p_1 — вероятность появления активного интервала τ. Положим также, что эта последовательность импульсов повторяется в следующие промежутки времени, длительностью t, что дает периодичность с большим периодом t. Это периодическое расположение имеет спектр с частотами

$0,\ \nu_0,\ 2\nu_0,\ \ldots,\ n\nu_0$ (см. (13.34)—(13.36)), где

$$\nu_0 = \frac{1}{t} = \frac{p_1}{m\tau}, \quad n\nu_0 = \nu_{\max} \approx \frac{1}{2\tau}.$$

Наибольшее значение n равно

$$n = \frac{1}{2\tau\nu_0} = \frac{m}{2p_1}.$$

Рассуждение, представленное равенствами (13.37)—(13.41), непосредственно применимо в данном случае и дает среднее увеличение энтропии на интервал τ:

$$\Delta S_\tau = \frac{\alpha k}{2},$$

как в (13.41).

ГЛАВА 14

НЕГЭНТРОПИЙНЫЙ ПРИНЦИП ИНФОРМАЦИИ В ОБЩЕЙ ФИЗИКЕ

1. Проблема измерения в физике

До сих пор мы смогли сделать несколько определенных заключений, которые можно подытожить следующим образом:

А. Информация может быть превращена в негэнтропию, и обратно. Если преобразование обратимо, то потерь нет, но необратимое преобразование всегда сопровождается потерей.

В. Любой опыт, дающий информацию о физической системе, приводит в среднем к увеличению энтропии системы или ее окружения. Это среднее увеличение всегда больше получаемого количества информации (или равно ему). Другими словами, информация всегда должна оплачиваться негэнтропией, причем оплата всегда больше, чем получаемое количество информации (или равна ему). Всякое наблюдение всегда сопровождается увеличением энтропии и связано, таким образом, с необратимым процессом.

С. Наименьшее возможное количество негэнтропии, требуемое при наблюдении, имеет порядок k. Более подробное исследование дает значение:

$$\text{минимум негэнтропии} = k \ln 2 = 0{,}7\, k \approx 10^{-16}\ \text{CGS}\ {}^{\circ}\text{K} \qquad (14.1)$$

для точного предела, что согласуется с данными Силарда. Этот минимум соответствует как раз одной двоичной единице.

D. Эти замечания приводят к разъяснению проблемы демона Максвелла, который попросту представляет собой устройство, превращающее негэнтропию в информацию, а затем снова в негэнтропию.

Исследуем более тщательно заключения В и С. Это приведет к обсуждению проблемы надежности и точности физического эксперимента. Читатель мог заметить, что в рассуждениях главы 13 мы ввели произвольный множитель α (см. (13.39) и (13.54)), который мы полагали достаточно большим, чтобы уменьшить ошибки наблюдения и ненадлежащее действие системы. Этот множитель α обеспечивает *надежность* экспериментального устройства. Мы исследуем также вопрос о наименьшей негэнтропии, требуемой для наблюдения, и подтвердим правило (14.1).

Общие результаты В и С были нами впервые обнаружены при обсуждении проблемы демона Максвелла. Наблюдатель, будь то физик или демон, нуждается в источнике негэнтропии, так как всякое наблюдение всегда производится за счет негэнтропии окружения. Так, он нуждается в батареях, источниках питания, сжатых газах и т. п. — все это представляет собой источники негэнтропии. Ему нужен также свет в его лаборатории, чтобы читать показания амперметров и других измерительных приборов.

Каково наименьшее возможное количество негэнтропии, требуемое для наблюдения? Положим, что мы хотим прочесть показание амперметра, поддерживаемого при температуре T. Стрелка с ее пружиной образует колебательную систему с очень низкой собственной частотой. Стрелка совершает колебательное броуново движение со средней кинетической энергией $kT/2$ и со средней полной энергией

$$\overline{E}_{\text{total}} = kT. \tag{14.2}$$

Айзинг[1]) полагает, что для получения правильного отсчета необходима энергия

$$E_{1} \geqslant 4\overline{E}_{\text{total}} = 4kT. \tag{14.3}$$

После того как отсчет сделан, эта энергия будет рассеяна вследствие трения, джоулева эффекта, вязкого затухания прибора и т. д. Это означает увеличение энтропии для всей

---

[1]) G. Ising, Phil. Mag. [6] **51**, 827 (1926); M. Courtines, Les fluctuations dans les appareils de mesures, Congr. intern. d'électricité, vol. 2, Paris (1932).

системы на

$$\Delta S = \frac{E_1}{T} \geqslant 4k. \tag{14.4}$$

Таким образом, по Айзингу для надежного наблюдения требуется негэнтропия, равная $4\,k$. Мы покажем, что $k \ln 2 = 0{,}7\,k$ есть предел, при котором мы имеем как раз 50% шансов, что наблюдение правильно, и 50% — что наблюдаемое отклонение обусловлено просто тепловым движением.

2. Наблюдения над осциллятором

Рассмотрим гармонический осциллятор частоты ν с квантованными уровнями энергии $E_n = nh\nu$. Вышеприведенная задача об амперметре соответствует предельному случаю очень низкой частоты. Мы получили выраженные формулами (9.31)—(9.44) некоторые общие результаты, относящиеся к статистическим свойствам такого осциллятора, поддерживаемого при температуре T. Энергетический уровень $nh\nu$ имеет вероятность

$$p_n = Be^{-nx} = e^{-nx} - e^{-(n+1)x},$$

где

$$x = \frac{h\nu}{kT},$$

$$\sum_{n=0}^{\infty} p_n = \sum_{n=0}^{\infty} Be^{-nx} = 1, \quad B = 1 - e^{-x}. \tag{14.5}$$

Отсюда для вероятностей состояний выше $(q-1)$ и ниже q получаем соответственно:

$$\left.\begin{aligned} P_{n \geqslant q} &= p_q + p_{q+1} + \ldots = e^{-qx}, \\ P_{0 < n < q} &= p_0 + p_1 + \ldots + p_{q-1} = 1 - e^{-qx}. \end{aligned}\right\} \tag{14.5a}$$

Мы можем определить медиану квантового числа m условием

$$P_{n < m} = P_{n \geqslant m},$$

которое дает:

$$e^{-mx} = \frac{1}{2}, \quad \text{или} \quad mx = \ln 2 \approx 0{,}7,$$

так что

$$E_m = mh\nu = kT \ln 2. \tag{14.6}$$

Эту медиану можно сравнить со средним значением

$$\bar{n} = \sum_{n=0}^{n=\infty} np_n = \frac{1}{e^x - 1}. \tag{14.7}$$

Для низких частот

$$h\nu \ll kT, \quad x \ll 1,$$

так что (14.7) дает:

$$\bar{n}x = 1 \quad \text{или} \quad \bar{E} = \bar{n}h\nu = kT, \tag{14.8}$$

и мы отмечаем, что медиана меньше среднего. Для высоких частот получаются иные соотношения, которые будут обсуждены в следующем разделе.

Далее мы подсчитаем среднее квантовое значение для состояний $n \geqslant q$:

$$\bar{n}_q = \frac{\sum\limits_{n=q}^{\infty} np_n}{P_{n \geqslant q}} = \frac{Be^{-qx} \sum\limits_{r=0}^{\infty} (q+r)\, e^{-rx}}{e^{-qx}}, \quad n = q + r,$$

и с помощью (14.5) и (14.7) получим:

$$\bar{n}_q = q + \bar{n}, \quad \text{так как} \quad \bar{r} = \bar{n}. \tag{14.9}$$

Среднее квантовое значение $\bar{n}'_q$ для группы состояний $n < q$ получается непосредственно из соотношения

$$P_{n<q}\bar{n}'_q + P_{n \geqslant q}\,\bar{n}_q = \bar{n},$$

которое дает:

$$\bar{n}'_q = \bar{n} - \frac{q}{e^{+qx} - 1}, \quad \text{при} \quad n < q. \tag{14.10}$$

Применим эти результаты к медиане:

$$\left.\begin{aligned} q = m:\ \bar{n}_m &= m + \bar{n}, \quad n \geqslant m; \\ \bar{n}'_m &= \bar{n} - m, \quad n < m; \end{aligned}\right\} \tag{14.11}$$

для низких частот

$$\left.\begin{aligned}\bar{n}_m &= \bar{n}(1+\ln 2) \approx 1{,}7\ \bar{n};\\ \bar{n}'_m &= \bar{n}(1-\ln 2) \approx 0{,}3\ n.\end{aligned}\right\} \qquad (14.12)$$

Мы можем теперь рассмотреть вопрос о том, как производится наблюдение и как оценивается его надежность. Положим, что мы наблюдаем в некоторый момент времени t энергию, соответствующую $n \geqslant q$ квантам. Вероятность того, что это высокое квантовое число может обусловливаться тепловыми флуктуациями, обозначим $P_{n \geqslant q}$. Вероятность того, что резонатор получил нормальным образом $n < q$, а затем поглотил некоторое количество добавочных квантов (от внешнего источника) для достижения уровня выше q, обозначена через $P_{n<q}$. Если считать оба случая равновероятными, то мы должны взять q равным медиане

$$q = m = \frac{\ln 2}{x}.$$

Если, с другой стороны, мы возьмем значение Айзинга для «надежного» наблюдения (см. (14.3)), то

$$q_1 = 4\,\frac{1}{x}$$

и

$$\frac{P_{n \geqslant q}}{P_{n<q}} = \frac{e^{-4}}{1-e^{-4}} = \frac{1}{e^4-1} \approx \frac{1}{54}, \qquad (14.13)$$

что дает для ошибки вследствие теплового движения вероятность около 2%.

В момент t мы наблюдаем больше чем q квантов в резонаторе. Если это обусловлено тепловыми флуктуациями, то средняя избыточная энергия получена от окружающего термостата T, и согласно (14.5) средняя энергия в этом случае будет равна

$$\bar{E}_q = \bar{n}_q h\nu = qh\nu + \bar{n}h\nu. \qquad (14.14)$$

Если q квантов от постороннего источника действительно поглощаются, то избыточная энергия $qh\nu$ получается от этого источника и добавляется к нормальному среднему значению $\bar{n}h\nu$, что снова дает (14.14). В дальнейшем избыточная энергия с той или иной скоростью рассеивается вследствие трения, вязкости или омических потерь и поглощается окружающим термостатом при температуре T.

Если (14.14) есть результат флуктуаций, то это просто означает, что термостат получает обратно часть энергии, которую он первоначально отдал. Если (14.14) есть результат действительного поглощения энергии от внешнего источника, то в конечном счете q квантов поглощаются термостатом и получается увеличение энтропии

$$\Delta S = \frac{qh\nu}{T} = qkx = \begin{cases} k \ln 2, & \text{если } q = m, \\ 4k & \text{по Айзингу.} \end{cases} \tag{14.15}$$

Это увеличение энтропии окружающего термостата есть цена, уплачиваемая за наблюдение. Наименьшее значение этого увеличения $k \ln 2 \approx 0{,}7\,k$ получается, когда мы допускаем вероятность ошибки, равную 50%.

Очень интересно, что, если выбрать предел $m < \bar{n}$ (см. (14.6) и (14.8)), то потребуется, тем не менее, поглощение некоторого количества энергии для того, чтобы вернуть резонатор в нормальное состояние. Это объясняется тем, что предел m может быть ниже среднего $\bar{n}$, но среднее $\bar{n}_m$ для квантов выше m больше, чем $\bar{n}$ (см. (14.12)).

Полученный результат является общим и применим как к низким, так и к высоким частотам, хотя численный пример этого раздела относился к низким частотам.

3. Высокочастотный резонатор и цена наблюдения

Рассмотрим высокочастотный резонатор, для которого $h\nu$ может быть порядка kT или больше kT. Величина x уже не мала, и условие (14.6) для медианы обычно не дает целого числа. Мы должны поэтому взять ближайшее большее целое:

$$q \geqslant m. \tag{14.16}$$

Надежность наблюдения при этом выше 50% и ΔS больше, чем $k \ln 2$ (см. (14.15)), если весь избыток энергии рассеивается и передается термостату при температуре T.

Возьмем теперь другой случай, который соответствует проблеме демона Максвелла. Молекулы находятся в замкнутом объеме при температуре T, и для того, чтобы увидеть молекулы, мы применяем дополнительный источник света, испускающий кванты $h\nu$. Молекула может быть наблюдаема, если она рассеивает по меньшей мере один квант $h\nu$, который затем поглощается фотоэлементом или глазом наблюдателя. Этот квант $h\nu$ должен быть выше уровня излучения черного тела для того, чтобы можно было различить его на общем фоне.

Ранее предполагалось (см. (13.4) и (13.38)), что эти требования приводят к условию

$$h\nu \geqslant kT.$$

Покажем теперь, что мы можем снизить предел до

$$h\nu \geqslant kT \ln 2. \tag{14.17}$$

Для этого заметим прежде всего, что наименьшее количество энергии, которое может быть поглощено, — это один квант.

Далее, если мы хотим, чтобы надежность наблюдения одного (или более) кванта была не ниже 50%, мы должны взять $m=1$. Совместно с (14.6) и (14.7) это приводит к соотношениям

$$m=\bar{n}=1 \quad \text{и} \quad x=\frac{h\nu}{kT}=\ln 2. \tag{14.18}$$

Среднее и медиана совпадают и соответствуют поглощению ровно одного кванта.

Основное состояние (без квантов) имеет вероятность $1/2$, равную общей вероятности всех состояний с 1, 2, 3 или более квантов. Если мы наблюдаем один (или более) квант в резонаторе, то вероятность того, что он может быть обусловлен действительным поглощением, составляет 50%, и вероятность того, что он обусловлен флуктуациями, также равна 50%. Когда этот квант рассеивается и снова поглощается окружающей средой при температуре T, мы получаем увеличение энтропии среды на

$$\Delta S = k \ln 2. \tag{14.19}$$

Таким образом, предел k, которым мы пользовались в разделе 3 главы 13, должен быть заменен на $k \ln 2$, что несколько повышает эффективность действия демона.

Если в замкнутом объеме при температуре T мы рассматриваем резонатор с более высокой частотой

$$h\nu > kT \ln 2, \quad \text{или} \quad x > \ln 2,$$

то получаем $m < 1$, и мы должны согласно (14.16) взять $q = 1$. Это дает надежность выше 50% и ΔS больше, чем $k \ln 2$. Во всех этих примерах наблюдение дает нам в точности одну двоичную единицу информации, ответ «да» или «нет» на определенный вопрос:

Была ли энергия поглощена данным резонатором?

Имеется ли молекула газа в данном месте?

Одна двоичная единица информации соответствует $k \ln 2$ и должна быть оплачена негэнтропией большей, чем $k \ln 2$.

Нужно отметить, что информация действительна лишь на короткое время: затухание резонатора вскоре рассеивает избыточную энергию или молекула смещается из наблюдаемого положения. Этот закон убывания соответствует общему смыслу второго начала.

Предшествующее рассуждение ясно показывает, что второе начало справедливо лишь в среднем (см. п. В раздела 1). Второе начало всегда ограничено возможностью непредсказуемых флуктуаций. Может случиться, что отдельное наблюдение сможет быть сделано за исключительно малую цену, но мы не в состоянии предвидеть, когда и как это может случиться. Только средние значения могут уверенно предсказываться.

Количество информации, соответствующее значительному числу двоичных единиц, дает чрезвычайно малое изменение энтропии системы благодаря множителю 10^{-16} в (14.1).

Обсуждаемые нами вопросы не имеют большого значения для термодинамики, но связь между энтропией и информацией имеет фундаментальное значение для последовательного построения теории информации. Кроме того, мы рассмотрим специальные условия, когда энтропийная цена наблюдения может оказаться много выше предела (14.1).

4. Эксперименты, требующие многих одновременных наблюдений при низких частотах

В разделах 2 и 3 мы обсуждали очень простой пример, в котором эксперимент был основан на наблюдении одного-единственного осциллятора, и мы нашли предел $\Delta S = k \ln 2$ увеличения энтропии, когда вероятность правильного наблюдения считается равной $1/2$. Этот предел был получен для низких частот. Если включить и случай высокой частоты (раздел 3), то результаты могут быть представлены следующим образом: увеличение энтропии при отдельном наблюдении равно

$$\Delta S \geqslant A_1 k, \text{ где } A_1 = \begin{cases} \ln 2 \text{ — низкие частоты } h\nu \leqslant kT, \\ \dfrac{h\nu}{kT} \text{ — высокие частоты } h\nu > kT. \end{cases} \tag{14.20}$$

На практике в излучении черного тела высокие частоты не возбуждаются и одного кванта $h\nu$ достаточно для надежного наблюдения.

Нужно отметить, что в рассуждениях предыдущих разделов предполагалось, что наблюдение состоит в отсчете значения энергии в резонаторе выше E_l. Считается, что отрицательный отсчет, т. е. наблюдение низкой энергии резонатора, не дает какой-либо информации в силу соображений, приведенных в разделе 6 главы 13 в связи с обсуждением тепловой машины Силарда. Основное соображение, на основании которого мы ограничиваемся только информацией, полученной из положительных результатов (т. е. наблюдений высокой энергии в обнаруживающем приборе), состоит в том, что неясно, как определить надежность отрицательного отсчета.

Рассмотрим теперь более сложную задачу, в которой эксперимент требует наблюдений над n осцилляторами. В некоторых случаях наблюдения с помощью n приемников делаются одновременно, но при иных обстоятельствах мы можем делать наблюдения последовательно. Для наблюдения может требоваться, чтобы только один резонатор обладал в данный момент энергией выше определенного предела E_l, однако в иных экспериментальных устройствах наблюдение может быть основано на высоких значениях энергии в некотором числе m резонаторов:

$$1 \leqslant m \leqslant n.$$

Предполагается, что все резонаторы поддерживаются в термостате при постоянной температуре T; мы рассмотрим сначала случай низких частот ($h\nu \ll kT$). Задача состоит в определении предела E_l, который дает общую вероятность правильного наблюдения, равную 1/2, и вероятность, равную 1/2, для ложного наблюдения, обусловленного флуктуациями.

Возьмем произвольный предел E_l и рассмотрим случай одного из n осцилляторов. Из (14.5а) имеем следующие вероятности:

$$P_1 = e^{-E_l/kT} \tag{14.21a}$$

— вероятность того, что тепловые флуктуации дают данному резонатору энергию больше E_l, тогда как

$$P_2 = 1 - e^{-E_l/kT} \tag{14.21b}$$

есть вероятность того, что тепловые флуктуации дают резонатору энергию ниже E_l. Для случая, когда наблюдение ведется над n резонаторами, нетрудно вычислить следующую таблицу вероятностей, зависящих от тепловых флуктуаций.

Таблица 14.1

$P_{n,\,n} = P_1^n$	вероятность того, что все n резонаторов имеют энергию $\geqslant E_l$.
$P_{n,\,n-1} = nP_1^{n-1}P_2$	вероятность того, что какой-либо резонатор имеет энергию $< E_l$, тогда как остальные $(n-1)$ резонаторов имеют энергию $\geqslant E_l$.
$P_{n,\,n-2} = \frac{n(n-1)}{2} P_1^{n-2}P_2^2$	вероятность того, что какие-либо два резонатора имеют энергию $< E_l$, тогда как остальные $(n-2)$ резонатора имеют энергию $\geqslant E_l$.
$P_{n,\,n-m} = \frac{n!}{m!\,(n-m)!} P_1^{n-m} P_2^m$	вероятность того, что какие-либо m резонаторов имеют энергию $< E_l$, тогда как остальные $(n-m)$ резонаторов имеют энергию $\geqslant E_l$.
$P_{n,\,0} = P_2^n$	вероятность того, что все n резонаторов имеют энергию $< E_l$.

Сумма вероятностей для этих различных случаев должна, очевидно, равняться единице:

$$\sum_{m=0}^{n} P_{n,\,n-m} = P_1^n + \ldots + \frac{n!}{m!\,(n-m)!} P_1^{n-m} P_2^m + \ldots + P_2^n =$$
$$= (P_1 + P_2)^n = 1^n = 1.$$

Уточним теперь экспериментальные условия наблюдения. Положим прежде всего, что наблюдение дает положительный результат, когда какой-либо из n осцилляторов обладает энергией большей, чем E_l. Вероятность того, что это происходит не благодаря флуктуациям, равна P_2^n, а вероятность того, что один осциллятор (или более) поднят на уровень выше E_l тепловыми флуктуациями, составляет $1 - P_2^n$. Таким образом, вероятность правильного наблюдения есть P_2^n, так как наблюдаемая энергия выше E_l должна в этом случае быть результатом поглощения энергии внешнего источника. Вероятность ложного наблюдения, являющегося результатом флуктуаций, есть $1 - P_2^n$. Требуя, чтобы вероятность правильного наблюдения равнялась $1/2$, получаем условие

$$P_2^n = (1 - e^{-A_n})^n = \frac{1}{2}, \quad \text{где} \quad E_l = A_n kT, \tag{14.22}$$

или

$$e^{-A_n} = 1 - \left(\frac{1}{2}\right)^{1/n}. \tag{14.22a}$$

При $n = 1$ мы получаем наш прежний результат (14.15):

$$E_{l1} = kT \ln 2, \quad A_1 = \ln 2.$$

При увеличении n E_l должно возрастать, так как с увеличением n правая часть (14.22a) убывает. Следовательно, и e^{-A_n} убывает. Логарифмируя (14.22), получим асимптотическое выражение для A_n:

$$n \ln(1 - e^{-A_n}) \approx -ne^{-A_n} \approx -\ln 2,$$

или

$$A_n \approx \ln n - \ln(\ln 2) \approx \ln n + 0{,}3667. \tag{14.23}$$

Значения коэффициента A_n даны в следующей таблице:

$n =$	1	2	4	100	10 000
$A_n \approx$	0,69	1,23	1,84	4,95	9,58

(14.24)

В нашем эксперименте возбужденный резонатор получил свою избыточную энергию от внешнего источника, если наблюдение правильно, или, если наблюдение ложно, резонатор получил эту энергию от термостата. В любом случае эта избыточная энергия будет впоследствии поглощена термостатом, что приведет в случае правильного наблюдения к увеличению энтропии на

$$\Delta S = \frac{E_l}{T} = A_n k.$$

Это рассуждение в точности совпадает с предыдущим (см. (14.14) и (14.15)).

Нужно подчеркнуть, что в этой новой задаче вычисленное значение ΔS представляет лишь нижний предел. Действительное увеличение может быть больше, чем kA_n. Мы наблюдаем поглощение энергии от внешнего источника только тогда, когда оно достаточно для превышения резонатором уровня E_l. Но многие другие резонаторы могут поглотить некоторую энергию от внешнего источника в меньших количествах, которые не поднимут энергию резонаторов выше E_l и ускользнут, таким образом, от нашего внимания. Эта поглощенная энергия также перейдет в термостат и даст дополнительное неизвестное увеличение энтропии. Итак, окончательно мы приходим к условию

$$\Delta S \geqslant A_n k. \tag{14.25}$$

Так как A_n есть возрастающая функция n, то нижний предел увеличения энтропии возрастает с числом необходимых для эксперимента резонаторов.

Выражения (14.22), (14.22а), (14.23) и (14.24) для A_n были получены в предположении, что только один из резонаторов дает положительный отсчет. Эти выражения, однако, действительны в случае, когда мы получаем m $(1 \leqslant m \leqslant n)$ положительных отсчетов, если мы потребуем, чтобы вероятность того, что все они истинны, равнялась $1/2$. Это, как и ранее, означает, что, поскольку мы ограничиваемся флуктуациями, мы должны приравнять $1/2$ вероятность того, что все n резонаторов имеют энергию ниже E_l. Таким образом, все предыдущие выражения для A_n сохраняют силу. Нужно, однако, видоизменить выражение (14.25) для увеличения энтропии, так как теперь термостатом должна быть поглощена энергия mE_l. Итак,

$$\Delta S = \frac{mE_l}{kT} = mA_n k,$$

или, если учесть неопределенность отрицательных отсчетов.

$$\Delta S \geqslant mA_n k. \tag{14.26}$$

В этом рассуждении использована лишь небольшая часть таблицы 14.1. В других задачах, например, когда лишь некоторая доля положительных отсчетов считается истинной, понадобятся другие из приведенных вероятностей.

Иные соотношения получаются в случае высоких частот. Одного-единственного кванта $h\nu$ на каждый резонатор достаточно для надежного наблюдения, что и представляет, очевидно, необходимый минимум энергии. Эти частоты практически не встречаются в излучении черного тела. Поэтому мы берем

$$E_l = h\nu = AkT, \quad A = \frac{h\nu}{kT}.$$

Суммируя результаты, мы видим, что эксперимент, требующий n одновременных наблюдений над n резонаторами, находящимися в термостате, соответствует увеличению энтропии на

$$\Delta S = Ak \tag{14.27}$$

для каждого резонатора, энергия которого превосходит E_l, причем

$$A = \begin{cases} A_n \text{ согласно (14.23) и (14.24) для низких частот,} \\ \dfrac{h\nu}{kT} \text{ для высоких частот } h\nu \gg kT. \end{cases} \tag{14.28}$$

Промежуточный случай $h\nu \approx kT$ должен исследоваться отдельно для каждой частной задачи.

5. Проблемы, требующие высокой надежности

Наши предыдущие рассуждения основывались на допущении, что мы можем удовлетвориться в наших экспериментах вероятностью ошибок, составляющей 50%. Этот очень низкий предел был выбран для того, чтобы получить наименьшее значение увеличения энтропии, связанного со всяким измерением. Представляет интерес распространить исследование на проблемы, требующие большей надежности, и выяснить, как возрастает энтропийная цена измерения.

Подчеркнем здесь различие между *надежностью* и *точностью*. Точность (относящиеся к ней вопросы подробно обсуждаются в главе 15) может быть определена следующим образом: мы измеряем величину x в диапазоне l с возможной погрешностью Δx; мы полагаем, что величина

$$\mathscr{A} = \frac{l}{\Delta x} \tag{14.29a}$$

определяет точность измерения. Если, например, мы пользуемся метром, подразделенным на миллиметры, то $\mathscr{A} = 1000$. Для мерки длиной в один ярд[1]) с делениями через $^1/_4$ дюйма $\mathscr{A} = 144$.

Надежность соответствует другой задаче: если дана экспериментальная методика с определенными требованиями в отношении точности, то всегда существует возможность, что тепловые флуктуации дадут ложные отсчеты и неправильные показания. Это будет происходить с определенной вероятностью.

Рассмотрим теперь задачи, в которых вероятность ошибки вследствие теплового движения будет уменьшена с принятого нами ранее значения $^1/_2$ до

$$P = \frac{1}{r}, \quad r > 2. \tag{14.29b}$$

Условие Айзинга (см. (14.3), (14.4), (14.13) и (14.15)) соответствует значению $r = 54$.

Начнем с эксперимента с низкочастотным осциллятором, в котором делается один-единственный отсчет. Рассуждение раздела 2 здесь непосредственно применимо. Мы должны выбрать предел q, дающий

$$P_{n \geq q} = e^{-qx} = \frac{1}{r}$$

согласно (14.5a), и мы получаем:

$$\ln r = qx = \frac{qh\nu}{kT}. \tag{14.30}$$

[1]) Ярд = 3 футам = 36 дюймам.

Предел энергии резонатора есть $qh\nu$. После того как наблюдение сделано, эта дополнительная энергия рассеивается в затухании резонатора (см. (14.14)), в результате чего энтропия термостата возрастает (см. (14.15)) на

$$\Delta S = \frac{qh\nu}{T} = k \ln r. \tag{14.31}$$

Энтропийная цена измерения возрастает с увеличением надежности r по логарифмическому закону.

Рассмотрим теперь эксперимент, в котором делается много наблюдений. Это расширит результаты раздела 4. Наше предыдущее равенство (14.22) заменяется теперь на

$$P = (1 - e^{-A_n})^n = 1 - \frac{1}{r}, \quad E_l = A_n kT, \tag{14.32}$$

где P — вероятность того, что все положительные отсчеты, взятые по m из n резонаторов, являются истинными, и E_l — предельное значение энергии каждого из n резонаторов. Предыдущий результат (14.31) соответствует $n = 1$, и

$$A_1 = \ln r. \tag{14.33}$$

Для больших значений r и n можно получить асимптотическое выражение. Логарифмируя (14.32), имеем:

$$n \ln (1 - e^{-A_n}) \approx - ne^{-A_n}, \quad \ln \left(1 - \frac{1}{r}\right) \approx - \frac{1}{r},$$

откуда

$$ne^{-A_n} \approx \frac{1}{r},$$

или

$$A_n \approx \ln (rn), \quad rn \gg 1. \tag{14.34}$$

Коэффициент A_n для низких частот может изменяться практически от 0,7 до нескольких сотен. Эта формула может давать хорошее приближение даже при малых n, поскольку она дает точное значение при $n = 1$. Энергия, рассеиваемая в конечном счете в m резонаторах, равна $mA_n kT$, а полная энтропийная цена составляет:

$$\Delta S \geqslant kmA_n = km \ln (rn). \tag{14.35}$$

В случае высоких частот дело обстоит иначе. Для наблюдения достаточно одного кванта, так как высокие частоты практически отсутствуют в излучении черного тела:

$$\left.\begin{array}{c} E_t = h\nu = AkT, \quad A = \frac{h\nu}{kT}, \quad h\nu \gg kT, \\ \Delta S \geqslant mkA. \end{array}\right\} \qquad (14.36)$$

Для коэффициента A имеем:

$$A = \begin{cases} A_n \text{ согласно (14.34) для низких частот } h\nu \ll kT, \\ \frac{h\nu}{kT} \text{ для высоких частот } h\nu \gg kT. \end{cases} \qquad (14.37)$$

Переход от низких частот к высоким более труден и зависит от частных особенностей рассматриваемой задачи.

Во всех примерах, обсужденных до сих пор в этой главе, мы предполагали, что наблюдение производится над некоторым числом n осцилляторов, поддерживаемых при температуре T в термостате. В некоторый момент мы наблюдаем энергии этих осцилляторов, а затем равновесие температур между осцилляторами и термостатом постепенно восстанавливается. Некоторое количество энергии перераспределяется, и энтропия термостата возрастает. Такие условия являются простейшими для обсуждения, однако случай высоких частот требует более детального исследования.

6. Более подробное обсуждение экспериментов с высокими частотами

Условия, принятые в предыдущих разделах, дали удовлетворительную теорию для случая низких частот, но некоторые эксперименты требуют в действительности применения высоких частот, и возникает вопрос, действительно ли вся требуемая для эксперимента энергия должна всегда в конце концов рассеяться. Равенство (14.36), конечно, определяет полное количество энергии, необходимое для наблюдения, но вместо применения в качестве приемника осциллятора, поддерживаемого при температуре T, мы можем представить себе другое устройство, например, фотоэлемент при температуре T. Используемое в эксперименте излучение может попадать на кусок металла и выбивать из него электрон с кинетической энергией

$$E_{k_0} = h\nu - W_0, \tag{14.38}$$

где W_0 означает работу вылета электрона из катода. Электрон может быть затем замедлен электрическим полем, и может отдать в результате этого энергию W_1, соответствующую работе электрического поля. Наконец, электрон попадает на флуоресцирующий экран при температуре T и поглощается. В момент, когда происходит поглощение, электрон имеет низкую кинетическую энергию

$$E_{k_1} = h\nu - W_0 - W_1, \tag{14.39}$$

которая, превращаясь в тепло, дает увеличение энтропии экрана на

$$\Delta S = \frac{(h\nu - W_0 - W_1)}{T}. \tag{14.40}$$

Если замедляющее поле слабо, ΔS лишь немного меньше, чем $h\nu/T$. Если же поле слишком сильно, то электрон не может достичь экрана и наблюдение становится невозможным. Эксперимент может дать положительный результат только при условии, что электрон, достигающий экрана, имеет энергию большую, чем тепловые флуктуации в экране; это возвращает нас к задаче предыдущего раздела.

Необходимо подчеркнуть различие между следующими двумя величинами:

$$\Delta E = nh\nu \tag{14.41}$$

— наименьшей полной энергией, гребуемой для эксперимента с n приемными ячейками, и

$$\Delta S = knA_n \tag{14.42}$$

— наименьшим увеличением энтропии при наблюдении с надежностью r над n приемниками. Здесь A_n — коэффициент, ранее определенный равенствами (14.23), (14.24) и (14.34). Для высоких частот имеем:

$$\Delta Q = T\Delta S < \Delta E. \tag{14.43}$$

Лишь часть полной энергии ΔE превращается в тепло ΔQ;

разность может быть восстановлена в форме электрической или механической работы ΔW:

$$\Delta W = \Delta E - \Delta Q. \tag{14.44}$$

В большинстве экспериментальных устройств мы не заботимся о том, действительно ли энергия ΔE превращается в тепло или восстанавливается; однако предыдущее рассуждение важно в теоретическом отношении.

Итак, применение высоких частот для получения данного количества информации оказывается более дорогим (с точки зрения затраты негэнтропии), чем применение низких частот, хотя бы часть энергии, связанной с высокими частотами, и была восстановлена в форме работы с помощью какого-либо специального устройства. Однако всегда необходимо удержать достаточно энергии для воздействия на низкочастотный обнаруживающий прибор, так что предыдущие результаты, относящиеся к низкочастотному резонатору, устанавливают наименьшую цену наблюдения. Введение специального устройства для уменьшения потери энергии в случае высоких частот можно представить просто как преобразование ситуации, в которой требуются высокие частоты, к такой ситуации, где можно обойтись низкими частотами.

Здесь интересно также упомянуть о фотографическом методе. Когда квант $h\nu$ попадает в фотографическую пластинку, происходит ионизация атома и испускается электрон с кинетической энергией E_{k_0}, выражаемой равенством (14.38), где W_0 теперь означает порог ионизации. Электрон, пройдя некоторый путь, останавливается в эмульсии и теряет свою энергию E_{k_0}, которая частично затрачивается на совершение работы, а частично превращается в тепло в результате многих столкновений. Здесь W_0 много больше, чем kT (так как ионизация производится не тепловыми флуктуациями), и тепло, выделенное в эмульсии, будет больше kT (в противном случае электрон не смог бы пройти достаточно далеко и был бы отброшен тепловым движением). Эти условия существенно не отличаются от только что рассмотренных и приводят к сходным заключениям. Процесс проявления пластинки не требует обсуждения. Он представляет собой просто усиление (и с точки зрения энтропии очень дорогое), преобразовывающее микроскопический эффект в макроскопический, могущий наблюдаться непосредственно.

7. Пример, показывающий наименьшую негэнтропию, необходимую для наблюдения

Обсудим упрощенный пример, на котором действительно возможно показать существование наименьшей негэнтропии, требуемой для наблюдения. Рассмотрим снова задачу раздела 4 главы 13 об определении местоположения частицы при помощи луча света. Возьмем последовательность световых импульсов длительностью τ_1 с интервалом τ_2 между

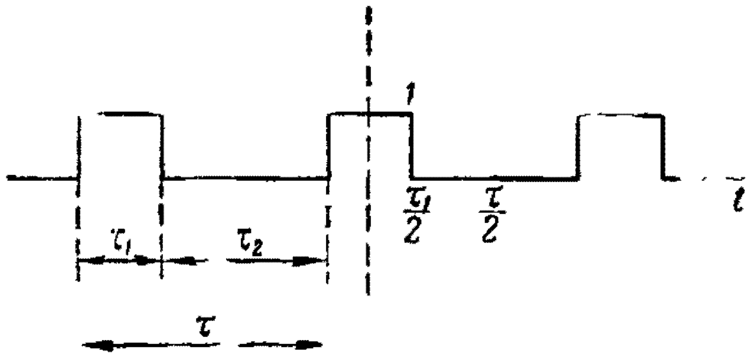

Рис. 14.1. Световые импульсы длительностью τ_1, повторяемые с интервалом τ. Свет выключен на протяжении интервала $\tau_2 = \tau - \tau_1$.

импульсами (рис. 14.1). Полный период есть $\tau = \tau_1 + \tau_2$, и функция, представляющая последовательность импульсов, запишется в виде

$$f(t) = \begin{cases} 1 & |t| < \frac{\tau_1}{2}, \\ 0 & \frac{\tau_1}{2} < |t| < \frac{\tau}{2}. \end{cases} \tag{14.45}$$

Такая последовательность импульсов требует полосы частот, определяемой в основном длительностью короткого импульса. Если $\tau_1 < \tau_2$, то имеем согласно (13.32)

$$0 < \nu < \nu_M, \quad \nu_M \approx \frac{1}{2\tau_1}. \tag{14.46}$$

Вместо импульсов $f(t)$ рассмотрим обращенную последовательность $F(t)$ с тем же периодом τ:

$$F(t) = \begin{cases} 0 & |t| < \frac{\tau_1}{2}, \\ 1 & \frac{\tau_1}{2} < |t| < \frac{\tau}{2}, \end{cases} \tag{14.47}$$

$$F(t) = 1 - f(t) \tag{14.48}$$

(рис. 14.2). Обе функции f и F имеют, очевидно, одинаковый спектр, хотя одна из них представляет короткие импульсы с длинными интервалами, а вторая — длинные импульсы с короткими интервалами. Наивысшая частота ν_M в обоих случаях зависит от более короткого интервала τ_1 или τ_2.

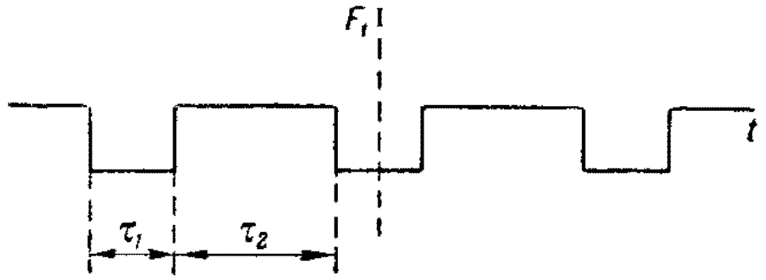

Рис. 14.2. Световые импульсы длительностью τ_2, повторяемые с интервалом τ_1.

Этот график является обращением рис. 14.1.

Эта наивысшая частота имеет минимум, когда f и F совпадают, т. е. при

$$\tau_1 = \tau_2 = \frac{\tau}{2} \quad \nu_M = \frac{1}{\tau}. \tag{14.49}$$

Этот случай соответствует, очевидно, наименьшему количеству энергии, требуемому для наблюдения, и, следовательно, наименьшему увеличению энтропии. Рассмотрим ситуацию более внимательно.

Сначала мы можем легко вычислить информацию. Мы применяем последовательность импульсов $f(t)$ (рис. 14.1) и наблюдаем рассеянный молекулой свет. Это значит, что молекула оказывается в луче света на протяжении одного из импульсов τ_1. Полное априорное число P_0 возможностей пропорционально τ, а после получения информации число возможностей P_1 пропорционально $\tau_1 < \tau$. Пользуясь формулой (1.6), находим информацию

$$I = k \ln\left(\frac{P_0}{P_1}\right) = k \ln\left(\frac{\tau}{\tau_1}\right). \tag{14.50}$$

Чем меньше длительность импульса τ_1, тем выше точность и тем больше информация.

Для того чтобы подсчитать используемую при наблюдении энтропию, мы исследуем сначала спектр нашей функции (14.45)

$$f(t)=\frac{\tau_1}{\tau}+\sum_{n=1}^{\infty}a_n\cos n\varphi,$$

$$\varphi=\frac{2\pi t}{\tau},\quad a_n=\frac{2}{\pi n}\sin\frac{\pi n\tau_1}{\tau}=\frac{2}{\pi n}(-1)^{n+1}\sin\frac{\pi n\tau_2}{\tau}. \qquad (14.51)$$

Симметрия относительно τ_1 и τ_2 очевидна и соответствует (14.48). Амплитуды a_n имеют наибольшее значение для малых n и убывают постепенно до очень малых значений. Положим, что $\tau_1<\tau_2$, тогда первая равная нулю амплитуда получается при

$$n_0=\frac{\tau}{\tau_1},\quad \nu_0=\frac{n_0}{\tau}=\frac{1}{\tau_1},\quad (\tau_1<\tau_2)\ ^{1)}.$$

Теперь мы должны вспомнить о существовании *шума*. Очень малые амплитуды в спектре (14.51) будут совершенно стерты вследствие искажения сигнала шумом. Предположим, что мы можем устранить эти высокие частоты и сохранить только конечный спектр, простирающийся до частоты $\frac{1}{2}\nu_0$ (половина частоты, при которой амплитуда первый раз обращается в нуль). Итак, примем:

$$\left.\begin{aligned}&\text{A:}\quad \nu_{\max}=\frac{1}{2\tau_1},\quad n_{\max}=\frac{\tau}{2\tau_1},\quad (\tau_1<\tau_2),\\&\text{B:}\quad \nu_{\max}=\frac{1}{2\tau_2},\quad n_{\max}=\frac{\tau}{2\tau_2},\quad (\tau_2<\tau_1).\end{aligned}\right\} \qquad (14.52)$$

Случай $\tau_1=\tau_2$ требует особого внимания. Условие (14.52) дает для этого случая $n_{\max}=1$, что означает, что мы применяем приближение

$$f=\frac{1}{2}+\frac{2}{\pi}\cos\frac{2\pi t}{\tau},\quad \tau_1=\tau_2=\frac{1}{2}\tau$$

с одним членом вместо прямоугольной ломаной (рис. 14.3).

Для того чтобы преодолеть тепловой шум, каждая из сохраняемых составляющих должна иметь энергию порядка kT.

---

[1]) Если не накладывать условия $\tau_1<\tau_2$, то в общем виде можно записать: $n_0=\frac{\tau}{\tau_1}m$, где m — наименьшее целое, дающее целое значение n_0. (*Прим. перев.*)

В общем, полная энергия конечного спектра равна

$$E = n_{\max} \alpha kT + \beta \frac{kT}{2}, \tag{14.53}$$

где численные множители α, β имеют порядок единицы. Последний член выражает энергию, которую должна иметь постоянная составляющая, чтобы превзойти $\frac{1}{2} kT$. Эта энергия

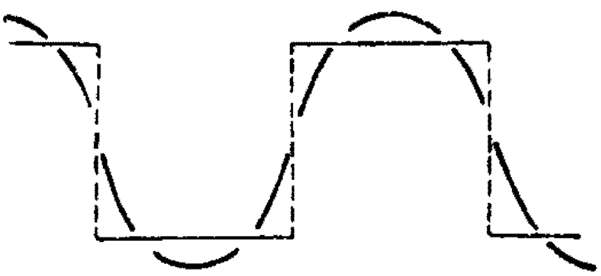

Рис. 14.3. Приближение для системы импульсов при $\tau_1 = \tau_2 = \frac{\tau}{2}$ двумя первыми членами ряда Фурье:

$$f = \frac{1}{2} + \frac{2}{\pi} \cos \frac{2\pi t}{\tau}.$$

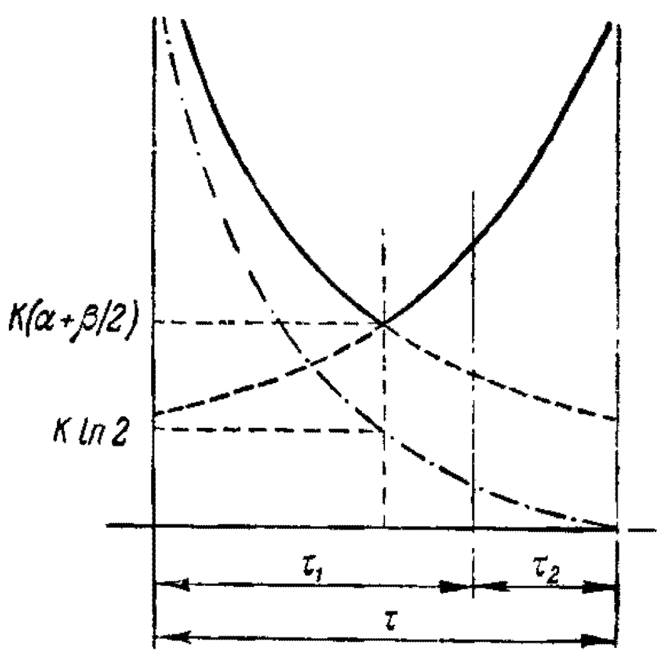

Рис. 14.4. Информация и связанное с ней увеличение энтропии в функции $\tau_1 = \tau - \tau_2$ при $0 < \tau_1 < \tau$

— · — · — информация, — энтропия, — — — — — неиспользованная доля $k\,(\alpha n_{\max} + \beta/2)$.

рассеивается и поглощается (например, в фотоэлементе) за время наблюдения, что дает увеличение энтропии

$$\Delta S = k\left(\alpha n_{\max} + \frac{1}{2}\beta\right), \tag{14.54}$$

изменяющееся в соответствии с (14.52). Кривая имеет минимум при $\tau_1 = \tau_2 = \frac{1}{2}\tau$, и общий ее вид показан на рис. 14.4, где нанесена также кривая I (см. (14.50)).

Этот пример снова доказывает, что увеличение энтропии всегда больше полученной информации и что количество энтропии, требуемой для наблюдения, не может быть ниже определенного предела, равного в нашем примере $k\left(\alpha + \frac{1}{2}\beta\right)$.

ГЛАВА 15

НАБЛЮДЕНИЕ И ИНФОРМАЦИЯ

1. Экспериментальные ошибки и информация

Мы уже подчеркивали тот факт, что количество информации, получаемое из эксперимента, тесно связано с точностью экспериментальной установки. Мы начали (глава 12) с определения информации

$$I_1 = k \ln\left(\frac{P_0}{P_1}\right), \tag{15.1}$$

где P_0 — число равновероятных возможностей до наблюдения, P_1 — число равновероятных возможностей после наблюдения, I_1 — информация, получаемая в результате наблюдения.

Мы пользовались этим определением в главах 13 и 14 при самых различных условиях эксперимента. Теперь мы обсудим общую проблему точности, экспериментальных ошибок и информации. Начнем с простого и определенного примера. Мы наблюдаем положение точки x на отрезке длиной L с погрешностью Δx, как показано на рис. 15.1. Введем определения:

$$\left.\begin{array}{ll}\text{абсолютная погрешность} & \varepsilon_a = \Delta x,\\[6pt] \text{относительная погрешность} & \varepsilon_r = \dfrac{\Delta x}{x},\\[6pt] \text{сравнительная (или пропорцио-} & \\ \text{нальная) погрешность} & \varepsilon_c = \dfrac{\Delta x}{L}.\end{array}\right\} \tag{15.2}$$

Точность $\mathscr{A}$ эксперимента была определена равенством (14.29) как величина, обратная ε_c:

$$\mathscr{A} = \frac{1}{\varepsilon_c} = \frac{L}{\Delta x}. \tag{15.2a}$$

Это определение удовлетворяет тому условию, что большая точность соответствует меньшей сравнительной погрешности. Абсолютная и относительная погрешности хорошо знакомы физику. Что касается сравнительной ошибки ε_c, то о ней, по-видимому, упоминается редко. Тем не менее именно эта величина имеет значение в связи с теорией информации.

Количество информации, получаемое при наблюдении, согласно (15.1) равно

$$\Delta I = k \ln\left(\frac{L}{\Delta x}\right) = -k \ln \varepsilon_c = k \ln \mathscr{A}. \qquad (15.3)$$

Оно зависит от отношения неопределенности Δx (которая остается и после того, как наблюдение сделано) к исходной

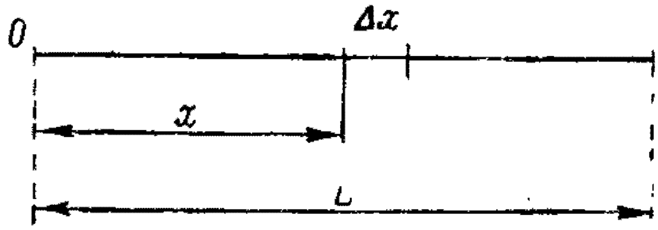

Рис. 15.1. Местоположение частицы определено с погрешностью Δx в точке x в пределах возможного интервала положений L.

неопределенности L, соответствующей всему диапазону наблюдения. В случае оптического прибора мы имели бы дело с отношением разрешающей способности к полной апертуре.

В главе 14 (раздел 7) уже применялись сходные определения в связи с вопросом об измерении времени. Рассматривая случай короткого импульса длительностью τ, повторяемого с интервалами t, мы нашли, что определяющей величиной является отношение τ/t, так как она фигурирует как в выражении для информации, так и в формуле для энтропийной цены. Обсудим теперь сходные задачи в пространстве и во времени и выясним, как измерить положение (координаты) x, y, z частицы в момент t с погрешностями Δx, Δy, Δz и Δt. Определение требует введения величин a, b, c и θ, покрывающих диапазон возможных значений x, y, z и t. Если диапазон наблюдения не установлен, то информацию, получаемую при измерении, нельзя определить, а энтропийная цена становится бесконечно большой.

Напомним различие между *точностью* и *надежностью*. В главе 14 мы определили надежность путем рассмотрения возможности неправильных отсчетов, обусловленных тепловыми флуктуациями в измерительном устройстве. Простой рисунок покажет различие между этими двумя терминами. На рис. 15.2 мы полагаем, что применяются распространяющиеся вдоль линии импульсы и что эти импульсы наблюдаются в подходящей приемной системе. Полная длительность последовательности импульсов обозначена через θ, и кратчайший импульс, который приемная система способна измерить, соответствует Δt. Точность есть $\theta/\Delta t$.

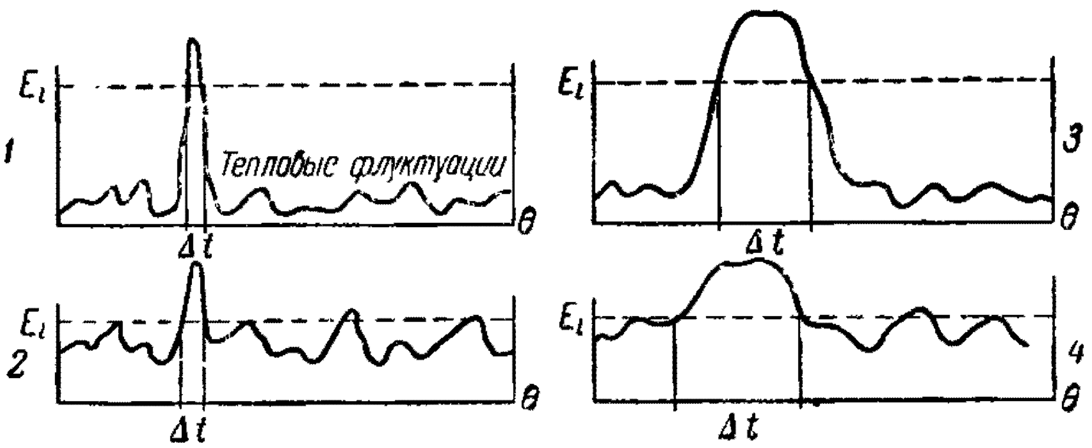

Рис. 15.2. Примеры импульсов с тепловым шумом, дающих четыре комбинации высокой и низкой точности с высокой и низкой надежностью:

1 — высокая точность, высокая надежность; 2 — высокая точность, низкая надежность; 3 — низкая точность, высокая надежность; 4 — низкая точность, низкая надежность.

При построении приемника мы выбрали некоторый предел E_l интенсивности, которую приемник записывает. Импульсы с интенсивностью меньше E_l не записываются, тогда как импульсы с интенсивностью выше E_l записываются. Если E_l много выше, чем средние тепловые флуктуации, то надежность велика, так как маловероятно, чтобы флуктуации достигли предела. Низкий предел E_l, слишком близкий к интенсивности тепловых флуктуаций, дает малую надежность. Общее положение поясняется рис. 15.2.

2. Измерения длины с низкой точностью

Начнем с очень простой задачи: измерение длины при помощи метровой мерки. Мерка подразделена на мелкие части Δx (1 *мм*, например), и мы прикладываем мерку к измеряемому объекту так, чтобы их начала совпадали. Для того чтобы видеть мерку и объект, необходим свет, и мы должны смотреть на каждое подразделение мерки, пока не найдем то деление, которое содержит конец объекта. Для точного обсуждения мы абстрагируем процедуру следующим образом. мы рассматриваем частицу, находящуюся где-то внутри интервала от 0 до L на оси x, и хотим определить ее положение с точностью Δx. Для этого мы подразделяем сегмент L на n интервалов длиной Δx каждый:

$$n = \frac{L}{\Delta x} = \frac{1}{\varepsilon_c} = \mathscr{A}. \tag{15.4}$$

Теперь задача состоит в том, чтобы найти интервал, в котором находится частица. Мы перемещаем луч света (например, системой зеркал) через каждый из n интервалов. Каждый интервал снабжен резонатором, который будет воспринимать свет, рассеянный частицей, если она находится в данном интервале. Каждый из резонаторов поддерживается при температуре T, и мы пока предполагаем, что применяется свет низкой частоты ($h\nu \ll kT$). Измерение может производиться одним из двух способов.

Случай I. Мы наблюдаем первый резонатор, а затем все последующие по очереди, пока не отметим, скажем в i-м интервале ($i \leqslant n$), положительное отклонение, которое может быть истинным или ложным.

Случай II. Мы наблюдаем все n резонаторов и отмечаем, скажем, j положительных отклонений ($1 \leqslant j \leqslant n$), из которых одно соответствует наличию частицы, а остальные являются ложными.

Подсчитаем получаемую информацию и связанное с ней увеличение энтропии, чтобы показать, что в обоих случаях

$$\Delta(S - I) \geqslant 0.$$

Для этого воспользуемся результатами главы 14, относящимися к наблюдениям, требующим применения многих резонаторов. Как и в главе 14, мы не будем придавать значения отрицательным отсчетам. Считается, что только положительные отсчеты (т. е. наблюдение высокой энергии в резонаторе) дают информацию. И так же, как и в главе 14, вычисленное увеличение энтропии будет представлять собой нижнюю грань.

Случай I. В этом случае частица наблюдается в интервале i, причем $i \leqslant n$. Таким образом, частица локализована в некотором положении из числа априори возможных i положений, и, следовательно, полученная информация есть

$$\Delta I = k \ln i. \tag{15.5}$$

Было сделано наблюдение над i резонаторами, из числа которых один резонатор поглотил энергию, рассеиваемую затем в термостате, так что связанное с этим увеличение энтропии равно согласно (14.26)

$$\Delta S \geqslant kA_i. \tag{15.5a}$$

Если i велико, то, используя выражения (15.5), (15.5a) и (14.23) для A_i, имеем:

$$\Delta (S - I) \geqslant k [\ln i - \ln (\ln 2)] - k \ln i = \\ = -k \ln (\ln 2) > 0. \tag{15.6}$$

Конечно, i может быть невелико: мы можем обнаружить частицу в первом или втором интервале. В этом случае мы не можем пользоваться асимптотическим выражением (14.23) для A_i, хотя неравенство (15.6) сохраняет силу.

В таблице 15.1 приведены данные для нескольких малых значений i.

Таблица 15.1

	i			
	1	2	3	4
A_i	0,69	1,23	1,58	1,84
$\frac{\Delta I}{k}$	0,00	0,69	1,10	1,39
$\frac{\Delta (S-I)}{k}$	0,69	0,54	0,48	0,45

Случай II. В этом случае мы имеем j положительных отсчетов, так что число возможных положений частицы сократилось от исходного n до j. Полученная информация равна

$$\Delta I = k \ln \left(\frac{n}{j}\right). \tag{15.7}$$

В этом случае j резонаторов поглотили энергию, которая затем должна быть рассеяна. Это дает увеличение энтропии на

$$\Delta S \geqslant kjA_n, \tag{15.7a}$$

так как наблюдались все n резонаторов. Мы снова воспользовались равенством (14.26). Полагая, что n достаточно велико для того, чтобы можно было воспользоваться асимптотическим выражением (14.23) для A_n, мы получаем с помощью (15.7) и (15.7a):

$$\Delta(S - I) \geqslant k\left\{j[\ln n - \ln(\ln 2)] - \ln\binom{n}{j}\right\} = \\ = k[\ln j + (j-1)\ln n - j\ln(\ln 2)] > 0. \tag{15.8}$$

Это равенство совпадает с (15.6) при $j = 1$, т. е. если только один резонатор дает положительный отсчет. При $j > 1$ случай II дает меньшую информацию и большее увеличение энтропии.

В последующих разделах мы воспользуемся результатами, полученными для случая II при $j = 1$. В этом случае получается наименьшая разность (из числа рассмотренных примеров) между увеличением энтропии и информацией; в то же время в этом случае мы можем считать, что n велико и асимптотическое выражение для A_n применимо.

Нужно отметить, что другие возможные устройства дают всегда большее увеличение энтропии для равного количества информации. Так, например, можно устроить резонаторы так, чтобы они получали свет непосредственно; при этом положительный отсчет указывает на отсутствие частицы. В этом случае ищется отрицательный отсчет, указывающий, что частица преграждает прохождение света через данный интервал. Легко убедиться, что такой эксперимент дает много большее увеличение энтропии, так как большая часть резонаторов поглощает свет.

3. Измерения длины с высокой точностью

В предыдущем рассуждении предполагалась низкая точность, когда погрешность Δx измерения длины не очень мала, так что может применяться свет низкой частоты. Для получения очень малой погрешности Δx необходимо применять свет высокой частоты с малой длиной волны λ с тем, чтобы луч освещал одновременно только один интервал. Сосредоточение света на одном-единственном интервале Δx практически возможно только при условии, что длина волны λ меньше $2\Delta x$:

$$\lambda \leqslant 2\Delta x. \tag{15.9}$$

На рис. 15.3 показаны схематически три различных осветительных устройства. В первом из них (A) используется волновод, толщина которого d равна Δx. Этот волновод имеет

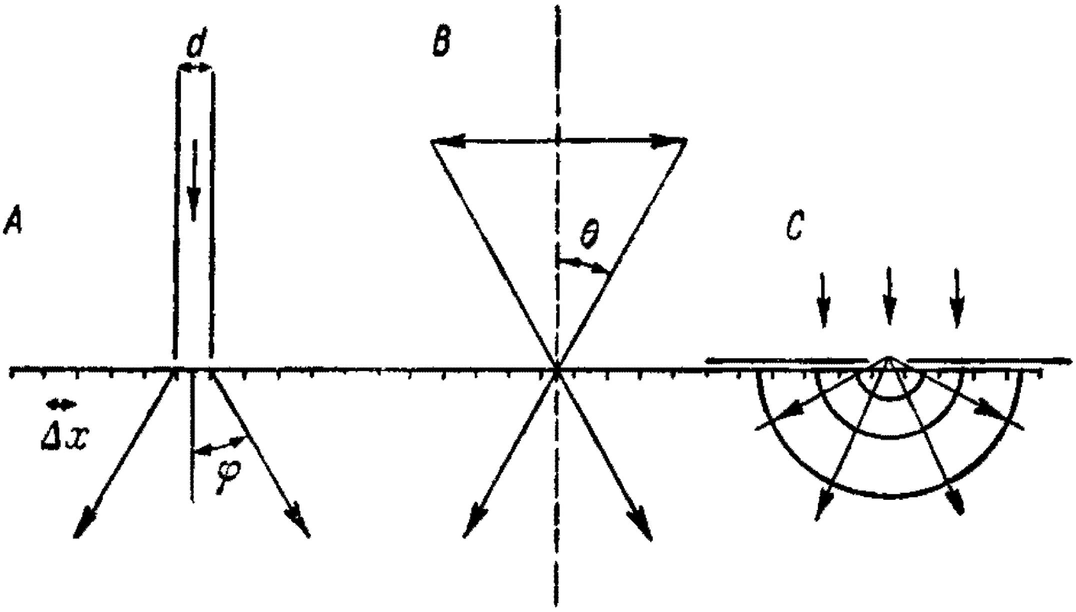

Рис. 15.3. Три устройства для сосредоточения света на интервале Δx:

A. Волновод толщины $d = \Delta x$; B. Линза с апертурой θ; C. Экран с апертурой ширины Δx.

нижнюю граничную (критическую) частоту, соответствующую $\lambda/2 = d$, и может передавать частоты только выше граничной. Это и есть в точности условие (15.9). По выходе из волновода свет расходится под углом φ, тем меньшим, чем выше частота. Освещение может быть резко ограничено на интервале Δx. В схеме B применена линза с апертурой θ. Ширина фокального пятна согласно известной формуле для разрешающей способности равна

$$\Delta x = \frac{\lambda}{2\sin\theta}.$$

Это снова приводит к условию (15.9). В схеме C делается попытка применения большой длины волны λ и плоского экрана с отверстием размера Δx, но вследствие дифракции по другую сторону экрана свет рассеивается на соседние интервалы, и сосредоточение света на одном-единственном интервале Δx этим методом недостижимо.

Таким образом, для резкого сосредоточения нужно применять для наблюдения частоту, удовлетворяющую условию:

$$\nu = \frac{c}{\lambda} \geqslant \frac{c}{2\Delta x}. \tag{15.10}$$

Определим теперь характеристическую длину, относящуюся к температуре T, при которой поддерживаются резонаторы, примененные для наблюдения,

$$\frac{h\nu}{kT} = \frac{hc}{2kT\Delta x} = \frac{\Delta x_T}{2\Delta x}, \tag{15.11}$$

где

$$\Delta x_T = \frac{hc}{kT} \approx \frac{1{,}44}{T}. \tag{15.12}$$

Здесь взято наименьшее значение ν, удовлетворяющее условию (15.10).

Напомним, что условия раздела 2 требуют низкой частоты, так что

$$\frac{h\nu}{kT} \ll 1,$$

и, следовательно,

$$\Delta x \gg \frac{\Delta x_T}{2}. \tag{15.13}$$

Это означает, что точность может быть низкой, поскольку длина интервала Δx ограничена со стороны малых значений.

С другой стороны, для получения высокой точности требуется

$$\Delta x \ll \frac{\Delta x_T}{2}, \tag{15.13a}$$

для чего нужно (см. (15.11)), чтобы было

$$\frac{h\nu}{kT} \gg 1. \tag{15.14}$$

Для обсуждения случая высоких частот нам потребуется, однако, более точное утверждение по поводу величины $h\nu$. На основании предыдущего мы можем принять, что при высоких частотах

$$h\nu > E_l = kTA_n, \tag{15.14a}$$

где E_l — предельная энергия осцилляторов в низкочастотном случае. Если считать n большим (т. е. применять много резонаторов), то это условие принимает вид

$$h\nu > kT\,[\ln n - \ln(\ln 2)\,]. \tag{15.14b}$$

В этом высокочастотном случае мы можем подсчитать $\Delta(S-I)$ для условий, описанных в конце раздела 2 (наблюдаются n резонаторов, только один отсчет положителен). Так как поглощение одного-единственного кванта достаточно для одного положительного отсчета, то имеем:

$$\left.\begin{aligned} &\Delta S = \frac{h\nu}{T} > k\,[\ln n - \ln(\ln 2)\,], \\ &\Delta I = k\ln n, \\ &\Delta(S-I) > k\ln(\ln 2) > 0, \end{aligned}\right\} \tag{15.15}$$

как и в низкочастотном случае. В действительности в рассматриваемом случае увеличение энтропии может быть много больше, так как ν может быть очень велико. Не следует, однако, забывать рассуждения раздела 6 главы 14 по поводу возможности уменьшения доли рассеянной энергии $h\nu$. Мы не можем получить результаты лучшие, чем в низкочастотном случае, так как рано или поздно наблюдение должно быть сделано при помощи низкочастотного устройства, если не высокочастотного. Обобщенный принцип Карно, таким образом, удовлетворяется, и мы обсудим позднее роль, которую играет необходимая для эксперимента большая энергия. Сходное положение имеется и в других примерах.

Остающийся случай $h\nu \approx kT$ требует специального обсуждения в каждой отдельной задаче. Если, например A_n настолько мало, что равенство (15.14a) применимо, а (15.14) — не применимо, то мы имеем ситуацию, требующую подробного исследования, основанного на точных выражениях различных средних величин, необходимых для получения значения A_n.

4. Эффективность наблюдения

Эффективность $\mathscr{E}$ экспериментального метода наблюдения может быть определена как отношение полученной информации ΔI к негэнтропийной цене $|\Delta N|$, равной увеличению энтропии ΔS, сопровождающему наблюдение:

$$\mathscr{E}=\frac{\Delta I}{|\Delta N|}=\frac{\Delta I}{\Delta S}, \quad \Delta N=-\Delta S. \tag{15.16}$$

При низкой точности мы получаем условия раздела 2, а в случае II (с одним только положительным отсчетом) имеем для больших n и при надежности, равной 2,

$$\mathscr{E}=\frac{\Delta I}{\Delta S}=\frac{\ln n}{A_n}\approx\frac{1}{1-\frac{\ln(\ln 2)}{\ln n}}\approx 1+\frac{\ln(\ln 2)}{\ln n}. \tag{15.17}$$

Чем выше точность ($\mathscr{A}=n$ велико), тем выше эффективность $\mathscr{E}$ наблюдения; эффективность стремится к единице с увеличением n.

Формула (15.17) применима в случае низкой надежности, когда вероятность того, что наблюдаемый отсчет обусловлен тепловыми флуктуациями, равна $1/2$. Для задач, требующих высокой надежности, мы можем воспользоваться формулами раздела 5 главы 14.

Мы нашли (см. (14.34)), что если как r, так и n велики, то

$$A_n=\ln(rn), \quad r\gg 1, \quad n\gg 1,$$

что дает увеличение энтропии

$$\Delta S=k\ln(rn), \tag{15.18}$$

и эффективность

$$\mathscr{E}_r=\frac{\Delta I}{\Delta S}\approx\frac{\ln n}{\ln(rn)}=\frac{1}{1+\frac{\ln r}{\ln n}}. \tag{15.19}$$

Эффективность будет зависеть от соотношения n и r.

Мы отмечали, что высокая точность требует применения высокой частоты. Если, однако, применяются соответствующие устройства для уменьшения рассеяния энергии, то вышеприведенные соотношения, относящиеся к низкочастотному случаю, сохраняют силу и для высокочастотного случая. Без таких устройств равенство (15.19) видоизменяется, и мы имеем:

$$\mathscr{E}_r = \frac{\Delta I}{\Delta S} \approx \frac{\ln n}{h\nu}, \tag{15.19a}$$

где $h\nu > E_l \approx \ln(rn)$, так что эффективность получается меньше.

5. Измерение расстояния с помощью интерферометра

Метод измерения длины или положения, рассмотренный в предыдущих разделах, дает сравнительно малое увеличение энтропии; иными словами, эффективность измерения (по крайней мере при низкой надежности) близка к единице, и эффективность возрастает с повышением точности.

Этот метод высокоэффективен также и при большей надежности, причем точное значение эффективности зависит согласно (15.19) как от надежности, так и от точности.

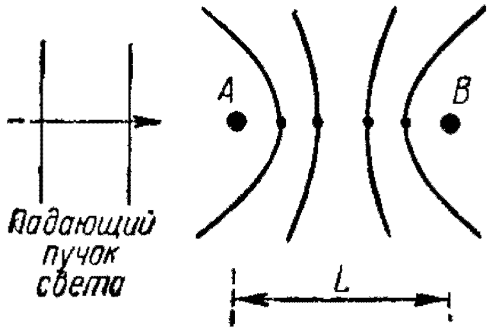

Рис. 15.4. Интерференционное окаймление, производимое рассеянием света двумя частицами A и B, отстоящими друг от друга на L.

Интерферометрические измерения оказываются очень дорогостоящими (с точки зрения увеличения энтропии), и их эффективность — малой даже при низкой надежности. Как мы увидим, основной причиной такого положения является необходимость записи многих положительных отсчетов.

Представим себе, что для измерения расстояния L между точками A и B мы помещаем в A и B материальные частицы, освещаем их и наблюдаем интерференционную картину, возникающую вследствие рассеяния света частицами. Число n интерференционных полос между A и B дает число полуволн на расстоянии L между ними (рис. 15.4):

$$\frac{n\lambda}{2} = L. \tag{15.20}$$

Для определенности рассмотрим действительный интерферометр, например, прибор Фабри — Перо, применяемый для измерения длины эталонного метра. Сводя метод к его существенным чертам, мы можем представить себе две параллельные отражающие пластины A и B и систему стоячих волн между ними, как показано на рис. 15.5. Для нахождения расстояния между A и B нужно сосчитать число плоскостей максимальной интенсивности между A и B. Это можно сделать при помощи фотоэлемента P, передвигаемого от A к B. Соотношение (15.20) дает частоту света

$$\nu = \frac{nc}{2L}. \qquad (15.21)$$

Если для надлежащего наблюдения положения плоскости наибольшей интенсивности требуется q квантов $h\nu$, то потребуется общее количество лучистой энергии

$$\Delta E = nqh\nu = \frac{n^2qhc}{2L}. \qquad (15.22)$$

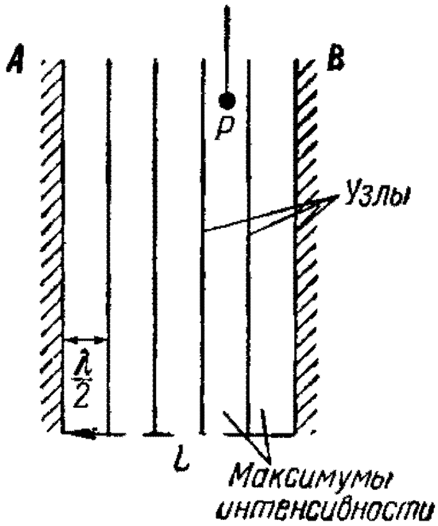

Рис. 15.5. Стоячие волны между двумя параллельными отражающими пластинами A и B, отстоящими друг от друга на L.

Фотоэлемент P применяется для нахождения плоскостей наибольшей интенсивности.

При низких частотах число q будет велико, но при высоких частотах можно обойтись одним квантом на полосу ($q = 1$). Исследуем этот пункт более тщательно.

Если вся энергия ΔE поглощается и превращается в тепло в термостате, то соответствующее увеличение энтропии за время наблюдения равно

$$\Delta S = \frac{\Delta E}{T} = \frac{n^2qhc}{2LT} = \frac{kn^2q\Delta x_T}{2L}. \qquad (15.23)$$

Здесь мы снова вводим характеристическое расстояние Δx_T для температуры T, определенное равенством (15.12). Положим, что мы используем наименьшее число q квантов, допустимое в данных экспериментальных условиях (к этому вопросу мы еще вернемся). Мы измеряем, таким образом, длину L с погрешностью, не превосходящей $\lambda/2$:

$$\Delta L \approx \frac{\lambda}{2} = \frac{L}{n}. \qquad (15.24)$$

Следовательно, если условия (15.23) выполнены, то

$$\Delta S = \frac{knq\Delta x_T}{2\Delta L}. \tag{15.25}$$

Обсудим теперь вопрос о числе q квантов, требуемом для наблюдения положения полосы.

А. Большое расстояние, низкая точность

Эти условия соответствуют малым квантам и низким частотам, наблюдаемым на резонаторах в термостате при температуре T:

$$h\nu \ll kT,$$

$$\frac{h\nu}{kT} = \frac{nhc}{2LkT} = \frac{\Delta x_T}{2\Delta L} \ll 1. \tag{15.26}$$

Мы работаем при условиях, установленных в разделе 4 главы 14, где в общем виде обсуждалось экспериментальное устройство с применением большого числа n одновременных наблюдений. Для случая надежности, равной 2 (вероятность ошибки вследствие флуктуаций равна 1/2), мы получили следующий результат:

$$E_l = qh\nu = A_n kT,$$

где $A_n > 0{,}7$. Это дает:

$$q = \frac{A_n kT}{h\nu} = \frac{2A_n \Delta L}{\Delta x_T} \gg 1.$$

Коэффициент A_n достигает нескольких единиц. Повышение надежности ведет к увеличению коэффициента A_n. При помощи (15.25) мы можем подсчитать увеличение энтропии в случае A, так как в этом случае вся энергия ΔE, несомненно, рассеивается при наблюдении:

$$\Delta S = knA_n. \tag{15.27}$$

Мы можем полагать, что это наблюдение, скажем, измеряет положение B по отношению к A, так что, как и в предыдущем, полученная информация равна

$$\Delta I = k \ln \frac{L}{\Delta L} = k \ln n, \tag{15.28}$$

а эффективность наблюдения для больших n

$$\mathscr{E}=\frac{\Delta I}{\Delta S}=\frac{\ln n}{nA_n}\approx\frac{\ln n}{n\left[\ln n-\ln(\ln 2)\right]}\approx$$
$$\approx\frac{1}{n}\left[1+\frac{\ln(\ln 2)}{\ln n}\right]\leqslant 1. \qquad (15.29)$$

Эта величина значительно меньше, чем в предыдущих случаях, где требовался только один положительный отсчет. Если требуется более высокая надежность, т. е. меньшая вероятность ложных наблюдений, обусловленных тепловыми флуктуациями, то нужно воспользоваться для A_n формулами раздела 5 главы 14. Увеличение энтропии ΔS будет в n раз больше, чем в соответствующем случае раздела 3, и во столько же раз уменьшится эффективность.

Какова бы ни была надежность, A_n всегда больше, чем $\ln n$, и негэнтропийный принцип информации приводит к нашему обобщенному принципу Карно

$$\Delta(S-I)=k(nA_n-\ln n)\geqslant 0. \qquad (15.30)$$

B. Малые расстояния, или высокая точность и низкая температура

В этих условиях нужны высокие частоты

$$h\nu\gg kT, \qquad \frac{h\nu}{kT}=\frac{\Delta x_T}{2\Delta L}\gg 1. \qquad (15.31)$$

Теперь мы работаем далеко за пределами спектра излучения черного тела при температуре T. Мы можем применять малую интенсивность в один квант на каждую полосу, и, таким образом, общее количество энергии, необходимое для измерения, составляет:

$$q=1, \quad \Delta E=nh\nu=\frac{nhc}{2\Delta L}, \quad \Delta L=\frac{L}{n}. \qquad (15.32)$$

Мы пришли к ситуации, описанной в разделе 6 главы 14, и мы должны делать различие между требуемой полной энергией и той ее долей, которая должна в конце концов рассеяться в форме тепла ΔQ. Как было пояснено, мы можем воспользоваться фотоэлементом с замедляющим полем, по крайней мере теоретически (практически это было бы очень неэффективным устройством!). Это позволило бы нам уменьшить рассеиваемую энергию до той же самой величины, что и в случае низких частот, так что формулы (15.27) — (15.30) могли бы по-прежнему применяться.

Остается, однако, в силе тот факт, что для эксперимента требуется большая энергия ΔE, как показывает равенство (15.32), которое мы можем переписать в следующем виде:

$$\Delta E \Delta L = \frac{nhc}{2}, \quad n = \frac{L}{\Delta L} = \mathscr{A}, \tag{15.33}$$

где $\mathscr{A}$ — точность. Мы обсудим позднее физический смысл этого соотношения, показывающего, что требуемая энергия ΔE может беспредельно возрастать с уменьшением ΔL.

6. Другая схема для измерения расстояния

Предшествующее рассуждение выявляет большие трудности, возникающие при измерении интерферометрическим методом весьма малых расстояний, и показывает, что цена таких экспериментов очень высока. Ввиду важности этого результата полезно показать, что другие методы обходятся по меньшей мере так же дорого. Можно, например, предположить, что трудности, встреченные в предыдущем разделе, были обусловлены тем, что мы применили длины волн значительно меньшие, чем измеряемые расстояния. Это привело к чрезвычайно коротким волнам в задаче измерения очень малых расстояний. Можно выбрать другую методику, основанную на применении волн значительно более длинных, чем измеряемое расстояние L. Итак, рассмотрим следующие условия:

$$h\nu \ll kT, \quad \lambda \gg \Delta x_T \tag{15.34}$$

в соответствии с (15.24) и определением Δx_T (15.12). Мы имеем также:

$$\lambda \gg L. \tag{15.35}$$

Эта ситуация представлена схематически на рис. 15.6.

Мы будем в этом случае наблюдать интенсивность света, рассеянного частицами A и B как вперед, так и назад. Волны, рассеянные частицами A и B вперед, совпадают по фазе. Пусть α есть амплитуда волны, рассеянной каждой из частиц. Полная интенсивность волн, рассеянных вперед, есть

$$I_f = 4\alpha^2, \tag{15.36}$$

тогда как волны, рассеянные назад, сдвинуты по фазе на угол φ и дают результирующую интенсивность

$$I_b = 4\alpha^2 \cos^2\left(\frac{\varphi}{2}\right) \approx 4\alpha^2\left(1 - \frac{\varphi^2}{4}\right), \tag{15.37}$$

где

$$\varphi = \frac{4\pi L}{\lambda} \tag{15.38}$$

— малая величина на основании (15.35), так что разложение $\cos^2\left(\frac{\varphi}{2}\right)$ в (15.37) применимо. Разность i между I_f и I_b дает меру L:

$$i = I_f - I_b = \frac{I_f\varphi^2}{4} = I_f\, 4\pi^2\left(\frac{L}{\lambda}\right)^2. \tag{15.39}$$

Для измерения I_f и I_b нужно поглотить рассеянный свет в каком-либо фотоэлементе или в настроенных приемниках.

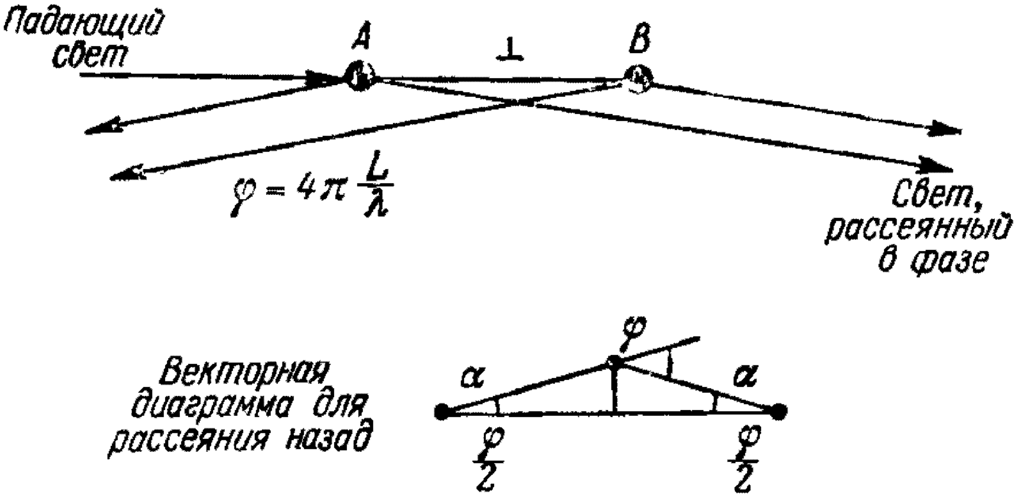

Рис. 15.6. Свет низкой частоты, рассеянный вперед частицами A и B, сохраняет фазу. Свет, рассеянный назад, отличается по фазе на $\varphi = 4\pi L/\lambda$. Показана также векторная диаграмма для рассеивания назад. Амплитуда равна α.

Таким образом, рассеивается энергия $I_f + I_b$, и эта сумма равна приблизительно $2I_f$. Соответствующее увеличение энтропии согласно (15.39) равно

$$\Delta S = \frac{2I_f}{T} = \frac{i\left(\frac{\lambda}{L}\right)^2}{2\pi^2 T}. \tag{15.40}$$

Тепловое движение в приемниках имеет порядок kT (согласно (15.34)), и поэтому как I_f, так и I_b измеряются с погрешностью kT, так что

$$\Delta i = 2kT \tag{15.41}$$

есть погрешность измерения разности i. Мы должны применять большие интенсивности для получения удовлетворительной точности. Если положить

$$i = \beta kT, \quad \beta \gg 1, \tag{15.41a}$$

и

$$I_f = fkT,$$

то из (15.39) и (15.41a) получаем:

$$i = \beta kT = fkT 4\pi^2 \left(\frac{L}{\lambda}\right)^2,$$

откуда с помощью (15.38)

$$\frac{\beta}{f} = 4\pi^2 \left(\frac{L}{\lambda}\right)^2 \ll 1, \quad \text{или} \quad f \gg \beta \gg 1.$$

Теперь из (15.41) и (15.41a) находим:

$$\frac{\Delta i}{i} = \frac{2}{\beta}, \tag{15.42}$$

тогда как из (15.39) имеем:

$$\Delta i = \Delta I_f 4\pi^2 \left(\frac{L}{\lambda}\right)^2 + I_f 4\pi^2 \left(\frac{L}{\lambda^2}\right) 2\,\Delta L,$$

или из (15.42)

$$\frac{\Delta i}{i} = \frac{\Delta I_f}{I_f} + \frac{2\,\Delta L}{L} = \left(\frac{1}{f}\right) + \frac{2\,\Delta L}{L} = \frac{2}{\beta}.$$

Совместно с условием $f \gg \beta$ это дает:

$$\frac{\Delta L}{L} \approx \frac{1}{\beta}. \tag{15.42a}$$

Коэффициент β получает смысл точности $\mathscr{A}$ при измерении L, так что выражение (15.40) для увеличения энтропии ΔS_i

с использованием (15.41a) принимает вид

$$\Delta S_l = k\left(\frac{e\mathcal{A}}{2\pi^2}\right)\left(\frac{\lambda}{L}\right)^2 \qquad (15.43)$$

для длинных волн $\lambda \gg L$. Мы должны сравнить этот результат с формулами, полученными в разделе 5. Мы рассмотрим снова два случая: A и B.

А. Короткие волны, большие расстояния (см. (15.26) и (15.27))

Имеем условие

$$L \gg \Delta x_T. \qquad (15.44)$$

Энтропийная цена для коротких волн при надежности эксперимента r была

$$\Delta S_{S,\,A} = knA_n, \qquad (15.45)$$

где $n = e\mathcal{A}$ и $A_n \approx \ln(rn)$. В конце концов коротковолновый метод дает:

$$\Delta S_{S,\,A} = ke\mathcal{A}\ln(re\mathcal{A}), \qquad (15.46)$$

т. е. $ke\mathcal{A}$, умноженное на численный коэффициент, значение которого составляет несколько единиц, или, самое большее, несколько десятков.

Применяя длинные волны, мы имеем ΔS_l, выраженное формулой (15.43). Чтобы удовлетворить условию (15.35), берем

$$\lambda = \gamma L, \quad \gamma \gg 1.$$

Достаточно значение γ порядка нескольких десятков, и условие (15.34) автоматически выполняется с учетом (15.44). Окончательно

$$\Delta S_l = \frac{ke\mathcal{A}\gamma^2}{2\pi^2}, \qquad (15.47)$$

где численный множитель $\gamma^2/2\pi^2$ имеет величину в несколько десятков. Итак, в первом случае (А) оба метода приводят к величинам одного порядка для негэнтропийной цены измерения.

В. Короткие волны, малые расстояния (см. (15.31), (15.32) и (15.34))

Положим, что

$$L \ll \Delta x_T. \tag{15.48}$$

При коротких волнах необходимая энергия согласно (15.32) была

$$\Delta E = \frac{n^2 hc}{2L} = kTn^2 \frac{\Delta x_T}{2L}.$$

Лишь некоторая доля этой энергии должна рассеяться.

Если, однако, взять наиболее неблагоприятный случай (вся энергия рассеивается), то увеличение энтропии составляет:

$$\Delta S_{S,\,B} \geqslant \frac{kn^2 \Delta x_T}{2L} = \frac{k\mathscr{A}^2 \Delta x_T}{2L}. \tag{15.49}$$

Пользуясь формулой (15.43) для случая длинных волн, мы снова возвращаемся к условиям (15.34) и (15.35). Теперь более стеснительным условием является (15.34), и мы полагаем:

$$\gamma' h\nu = kT, \quad \text{где} \quad \gamma' \gg 1,$$

что дает:

$$\lambda = \gamma' \Delta x_T \quad \text{и} \quad \frac{\lambda}{L} = \gamma' \frac{\Delta x_T}{L}.$$

Здесь коэффициент γ' величиной в несколько единиц достаточно велик, чтобы сделать тепловую энергию практически равной kT, как это требуется соотношением (15.41). Таким образом,

$$\Delta S_l = k\mathscr{A} \frac{\gamma'^2}{2\pi^2} \left(\frac{\Delta x_T}{L}\right)^2. \tag{15.50}$$

Множитель $(\gamma'/\pi)^2$ имеет порядок нескольких единиц. Для малых расстояний L энтропийная цена возрастает в случае длинных волн как L^{-2} (см. (15.50)), тогда как в случае коротких волн — лишь как L^{-1} (см. (15.49)).

Таким образом, для достаточно малых расстояний метод с применением коротких волн дает меньшую энтропийную цену, нежели метод с применением длинных волн; действительно, для чрезвычайно малых расстояний длинные волны обходятся очень дорого. Процедура раздела 5 более выгодна, чем рассматриваемый здесь метод.

Рассуждение еще не совсем закончено. Увеличение энтропии зависит от точности $\mathscr{A}$: оно растет как $\mathscr{A}^2$ при коротких волнах и как $\mathscr{A}$ — при длинных.

Случай умеренно малых расстояний потребовал бы более пристального изучения, но рассмотрение этого случая едва ли дало бы что-либо практически важное.

7. Измерение промежутков времени

Обсудив с некоторыми подробностями измерение длины, обратимся к другой основной проблеме: к измерению времени. В большинстве часов механизм передвигает стрелки по циферблату. Это попросту сводит измерение промежутка времени к измерению угла или длины. В этом случае наблюдение выполняется, как описано в предыдущих разделах.

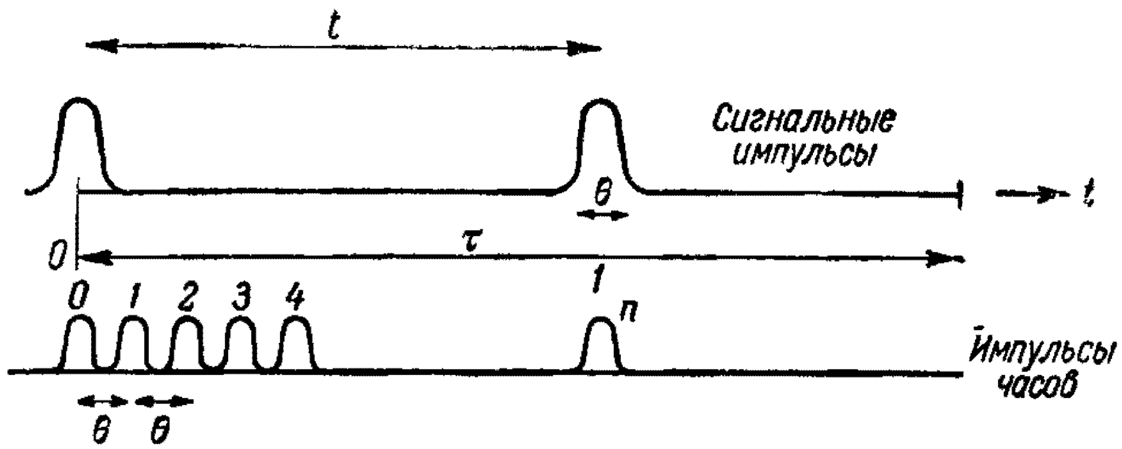

Рис. 15.7. Два сигнала длительностью θ посылаются со сдвигом на t в пределах возможного интервала τ. Электронные часы дают короткие импульсы с интервалом θ, приводящие в действие приемники сигналов, положение которых во времени мы желаем определить.

Эта процедура, однако, не очень точна, и мы рассмотрим другую методику, совпадающую по существу с той, которая применяется в электронных хронирующих устройствах, позволяющих достигнуть очень высокой точности. Временный интервал t определяется как расстояние во времени между двумя импульсами, например электрическими импульсами, посылаемыми по линии в приемник. Метод измерения поясняется схематически рис. 15.7. Первый импульс совпадает с импульсом, приходящим от электронного хронирующего устройства. Это устройство (электронные часы) дает короткие импульсы с интервалами θ; длительность импульсов также равна θ. Каждый импульс, поступающий от часов, включает некоторый приемник. Приемники за номерами 0, 1, 2, 3, ..., n подключаются по очереди к линии, по которой должны поступить сигнальные импульсы. Если за время включения приемника сигнала нет, то ничего не происходит. Если сигнал есть, приемник приходит в действие. Мы замечаем, например, что пришли в действие приемники за номером 0 и за номером n, и мы заключаем, что длина временного интервала есть

$$t = n\theta. \tag{15.51}$$

Переключающий механизм в принципе не требует какой-либо энергии. Его действие может быть основано на изменении напряжения на кристаллическом диоде, через который ток не проходит, пока не появляется сигнальный импульс. Можно также воспользоваться изменением напряжения на сетке электронной лампы. Когда приходит сигнальный импульс, он должен быть записан, а это означает рассеяние энергии.

Можно представить себе применение настроенного приемника, но обычно требуется усилитель. Он является вспомогательным устройством для увеличения мощности, и мы не будем вводить в рассмотрение энергию, потребляемую самой усилительной системой.

В такого рода задаче для определения количества информации и энтропийной цены мы должны установить полную длительность эксперимента τ. Воспользуемся определениями, сходными с введенными в разделе 1. Рассмотрим сравнительную погрешность ε_c и точность $\mathscr{A}$:

$$\varepsilon_c = \frac{\Delta t}{\tau} = \frac{\theta}{\tau} = \frac{1}{\mathscr{A}}. \tag{15.52}$$

В общем, ситуация в точности совпадает с той, которую мы обсуждали ранее в связи с измерениями длины.

Полученная информация согласно (15.3) равна

$$\Delta I = k \ln \mathscr{A}, \tag{15.53}$$

тогда как энтропийная цена составляет:

$$\Delta S \geqslant 2kA_{n+1}, \tag{15.54}$$

так как мы произвели наблюдение над $(n+1)$ резонаторами с положительными отсчетами по двум из них. Мы можем воспользоваться значением A_{n+1} по формуле (14.23) в случае надежности, равной 2 (вероятность ошибки 1/2), или взять

$$A_{n+1} = \ln [r(n+1)],$$

если требуется высокая надежность (глава 14, раздел 5). Обобщенный принцип Карно соблюдается:

$$\Delta (S - I) \geqslant 0. \tag{15.55}$$

В проблеме измерения времени в силу соотношения неопределенности всегда сохраняются условия, соответствующие случаю A раздела 5. Импульсы длительности θ требуют полосы шириной

$$\Delta\nu = 2\nu_M = \frac{1}{\theta}.$$

Покажем, что это условие приводит к низким частотам:

$$h\nu_M < kT, \quad \text{откуда} \quad \theta = \frac{1}{2\nu_M} > \frac{h}{2kT}, \tag{15.56}$$

или

$$\theta > \frac{\Delta x_T}{2c} = \frac{1{,}44 \cdot 10^{-10}}{6T} \text{ секунд}, \tag{15.57}$$

где T — абсолютная температура, Δx_T — характеристическая длина для температуры T, определяемая равенством (15.12). Условие (15.56) на практике всегда выполняется. Это условие можно рассмотреть с точки зрения соотношения неопределенности. Известно, что время t и энергия E представляют собой пару сопряженных переменных. Поэтому, если мы измеряем t с точностью Δt, мы обязательно вводим неизвестное возмущение ΔE в энергию E:

$$\Delta E \Delta t \approx h. \tag{15.58}$$

Эту неопределенность ΔE нужно сравнить с нормальными флуктуациями ε, обусловленными тепловым движением при температуре T:

$$\overline{\Delta_1 E} \approx \sqrt{\overline{\varepsilon^2}} = \sqrt{n}\, kT, \tag{15.59}$$

где n представляет число *активных* степеней свободы системы, а именно, число колебаний низкой частоты ($h\nu \leqslant kT$). Высокочастотные колебания практически не возбуждаются и оказывают малое влияние на флуктуации энергии. Итак, нужно сравнить ΔE и $\Delta_1 E$.

Случай A

Если $\Delta E \leqslant \Delta_1 E$, то

$$\Delta t \gg \frac{h}{\sqrt{n}\,kT}. \tag{15.60}$$

Это есть случай, рассматриваемый в данном разделе. Возмущением, обусловленным измерением времени, можно пренебречь. Условия (15.57) и (15.60) различаются только множителем $\sqrt{n}$.

Случай B

Если $\Delta E \geqslant \Delta_1 E$, то

$$\Delta t \leqslant \frac{h}{\sqrt{n}\,kT}. \tag{15.61}$$

Эти условия приводят к значительному возмущению системы. Слишком большая точность в определении времени могла бы даже разрушить наблюдаемую систему, так как потребовались бы высокие частоты, дающие большие кванты $h\nu$, порядка $\Delta_1 E$ или больше.

8. Наблюдения под микроскопом

Обсужденные в предыдущих разделах проблемы измерений расстояния и времени в своих применениях покрывают почти все области физического эксперимента. Так, например, измерение электрического тока производится по показанию вольтметра, что сводится к определению положения стрелки на шкале, т. е. к измерению длины. Измерение длины, измерение интервалов времени и счет — вот основные методы эксперимента, и наши рассуждения показывают применение негэнтропийного принципа информации к этим проблемам.

Вопрос имеет, однако, настолько большое значение, что мы исследуем еще несколько дополнительных примеров. Рассмотрим прежде всего вопрос о том, каким образом определяется положение частицы при помощи микроскопа, концентрирующего свет в фокусе F, как показано на рис. 15.8. Для того чтобы получить четкое определение фокуса, требуется определенная полоса частот и очень большое число фотонов, падающих на линзу. Задача становится определенной только тогда, когда мы ограничиваем поле наблюдения, как показано на рис. 15.8. Свет проходит через волновод, который может иметь, например, прямоугольное сечение со сторонами a и b вдоль осей x и y соответственно. Волновод определяет дискретную совокупность собственных волн[1]), наложение которых образует фокус в фокальной плоскости, координата которой $z = 0$.

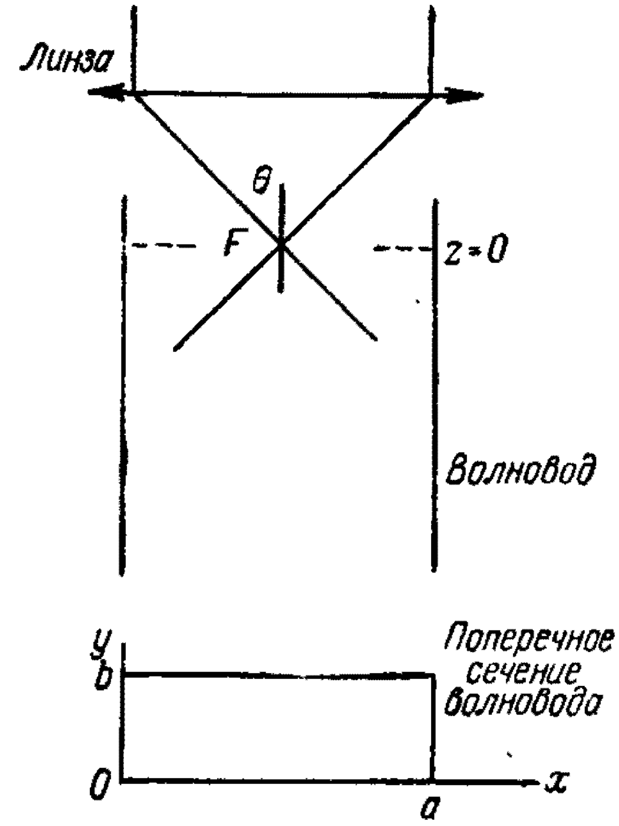

Рис. 15.8. Волновод ограничивает пределы наблюдений линзой микроскопа с апертурой θ. Фокус F линзы находится на уровне $z = 0$ прямоугольного волновода.

Выберем волны H (поперечно-электрические) с вертикально направленным магнитным полем, нормальная производная которого на границе должна обращаться в нуль:

$$H_z = \sum_{l,\, m} A_{lm} \cos \frac{\pi l x}{a} \cos \frac{\pi m y}{b} \cos \omega \left(t - \frac{z}{V_{lm}}\right), \quad (15.62)$$

где l и m — числа узлов в направлениях x и y, V_{lm} — фазовая скорость данной волны в направлении z. Целые числа l и m пробегают все значения от 0 до ∞, но комбинация 0, 0 исключена. Волновое уравнение

$$\nabla^2 H_z - \frac{1}{c^2} \frac{\partial^2 H_z}{\partial t^2} = 0 \quad (15.63)$$

---

[1]) См., например, C. G. Montgomery, R. H. Dicke, E. M. Purcell, Principles of Microwave Circuits, MIT Radiation Lab. series, vol. 8, pp. 34, 37, 55, McGraw-Hill, NY, 1948.

дает:

$$\omega^2\left(\frac{1}{c^2}-\frac{1}{V_{lm}^2}\right)=\left(\frac{\pi l}{a}\right)^2+\left(\frac{\pi m}{b}\right)^2. \tag{15.64}$$

Фазовая скорость V_{lm} больше, чем c, тогда как групповая скорость U_{lm} меньше, чем c. Обе скорости стремятся в пределе к c с повышением частоты. Фазовая скорость V_{bm} становится бесконечно большой (а U_{lm} приближается к нулю) при нижней граничной частоте

$$\frac{1}{c^2}\,\omega_{lm}^2=\left(\frac{\pi l}{a}\right)^2+\left(\frac{\pi m}{b}\right)^2, \tag{15.65}$$

и так как

$$\frac{\omega_{lm}}{c}=\frac{2\pi}{\lambda_{lm}},$$

то соответствующая длина волны выражается соотношением

$$\frac{4}{\lambda_{lm}^2}=\left(\frac{l}{a}\right)^2+\left(\frac{m}{b}\right)^2.$$

Данная l, m-волна может быть создана плоской волной, вступающей в волновод наклонно к оси и отражающейся попеременно от боковых стенок[1]). Обозначим через φ_x, φ_y и θ углы, образованные падающим лучом с осями x, y и z соответственно. Мы имеем:

$$\left.\begin{aligned}
&\cos\varphi_x=\pm\frac{l\lambda}{2a},\quad \cos\varphi_y=\pm\frac{m\lambda}{2b},\\
&\text{откуда}\\
&V_{lm}=\frac{c}{\cos\theta}>c,\\
&\frac{\omega^2}{c^2}\sin^2\theta=\left(\frac{\pi l}{a}\right)^2+\left(\frac{\pi m}{b}\right)^2,
\end{aligned}\right\} \tag{15.66}$$

или

$$\frac{4}{\lambda^2}\sin^2\theta=\left(\frac{l}{a}\right)^2+\left(\frac{m}{b}\right)^2.$$

[1]) Эта концепция предложена автором в 1936 г. Она основана на разложении поля в волноводе на плоские волны. Последующее рассмотрение отдельной плоской составляющей несколько формально в случае радиоволновода, но может иметь прямой физический смысл в оптическом случае. (*Прим. перев.*)

Граничные условия соответствуют $\theta = \pi/2$. В случае рис. 15.9, очевидно, высшие l, m-составляющие возбуждаются посредством наклонных лучей в сходящемся пучке. Подходящим наложением членов (15.62) можно образовать фокус F. Если частота света ω задана, то (15.65) ограничивает значения индексов l, m, так как всегда должно быть $\omega_{lm} < \omega$. Если задан апертурный угол θ_0, то (15.66) дает более тесное ограничение значений l, m. Высокая точность в определении фокуса потребует наличия в сумме членов с большими значениями l, m, а это в свою очередь потребует высокой частоты ω и большой апертуры θ_0.

Мы имеем, однако, дело с иной задачей. Мы будем удовлетворены ограниченной точностью в определении фокуса и хотим связать эту точность с информацией, получаемой при наблюдении. Возьмем в сумме (15.62) конечное число членов

$$0 \leqslant l \leqslant l_M, \quad 0 < m \leqslant m_M \tag{15.67}$$

и выберем частоту выше пределов, установленных равенством (15.66). Условия (15.67) дают фокальное пятно конечных размеров

$$\Delta x \geqslant \frac{a}{l_M}, \quad \Delta y \geqslant \frac{b}{m_M}. \tag{15.68}$$

Мы не можем разделить переменные z и t, так как волны свободно распространяются в направлении z и предполагается, что волновод либо бесконечен, либо нагружен на конце так, чтобы избежать отражений. Можно лишь применить короткие импульсы длительности τ, как мы делали в главе 13, и для того, чтобы получить конечное число степеней свободы, нужно периодически повторять импульсы на промежутке t. Это дает для каждой волны число степеней свободы (см. (13.36)):

$$n \geqslant \frac{t}{2\tau}. \tag{15.69}$$

В общем, пучок имеет $l_M m_M n$ степеней свободы частоты ω.

Положим, что при температуре T частота низка: $h\nu < kT$. Каждая степень свободы имеет энергию kT, и, следовательно, полная тепловая энергия будет равна

$$E_T = l_M m_M n kT. \tag{15.70}$$

Если мы хотим наблюдать фокус, различимый на тепловом фоне, мы должны применить энергию E, по меньшей мере равную AE_T, где A — коэффициент, определенный в разделах 4 и 5 главы 14.

Когда наблюдение сделано, лучистая энергия рассеивается рассматриваемой частицей и поглощается какими-либо фотоэлементами. Это дает увеличение энтропии

$$\Delta S = \frac{E}{T} \geqslant l_M m_M n k A, \tag{15.71}$$

или, с помощью (15.68) и (15.69),

$$\Delta S \geqslant \frac{AkF}{2}, \quad \text{где} \quad A > 1 \quad \text{и} \quad F = \frac{a}{\Delta x} \frac{b}{\Delta y} \frac{t}{\tau}. \tag{15.72}$$

Полученная информация согласно (15.1) и (15.3) равна

$$\Delta I = k \ln F \tag{15.73}$$

— величине, всегда меньшей, чем ΔS. Негэнтропийный принцип информации удовлетворяется, так как

$$\Delta (S - I) \geqslant k \left[\left(\frac{AF}{2} \right) - \ln F \right] > 0 \tag{15.74}$$

— неравенство, очень сходное с полученным Гэйбором при обсуждении задачи Силарда (см. (13.55)). Выражение (15.74) всегда положительно. Оно имеет при $F = 2/A$ минимум, равный $k \left[1 - \ln \left(\frac{2}{A} \right) \right] > 0$, так как A больше единицы.

Когда поле наблюдения очень велико и точность высока, величина F становится очень большой, и энтропийная цена наблюдения ΔS много больше полученной информации ΔI. Это замечание может иметь важное значение в некоторых физических задачах. Наблюдение фокуса без ограничения времени и поля сделало бы энтропийную цену бесконечно большой, что, однако, не соответствует никакой физической проблеме [1]).

[1]) Этот вопрос исследован G. Toraldo di Francia, J. Opt. Soc. Amer. 45, 497—501 (1955).

9. Рассуждение о фокусе в волноводе

Вернемся снова к условиям (15.67) и исследуем более тщательно условия вблизи фокуса в волноводе. Мы должны помнить, что стенки трубы действуют как совершенные зеркала, беспредельно отражающие сечение трубы вверх и вниз, вправо и влево, как показано на рис. 15.9. Фокус $F_{0,0}$ внутри трубы имеет свои изображения в $F_{-1,0}$, $F_{0,-1}$, $F_{-1,-1}$, в соседних прямоугольниках, и так без конца. Как можно

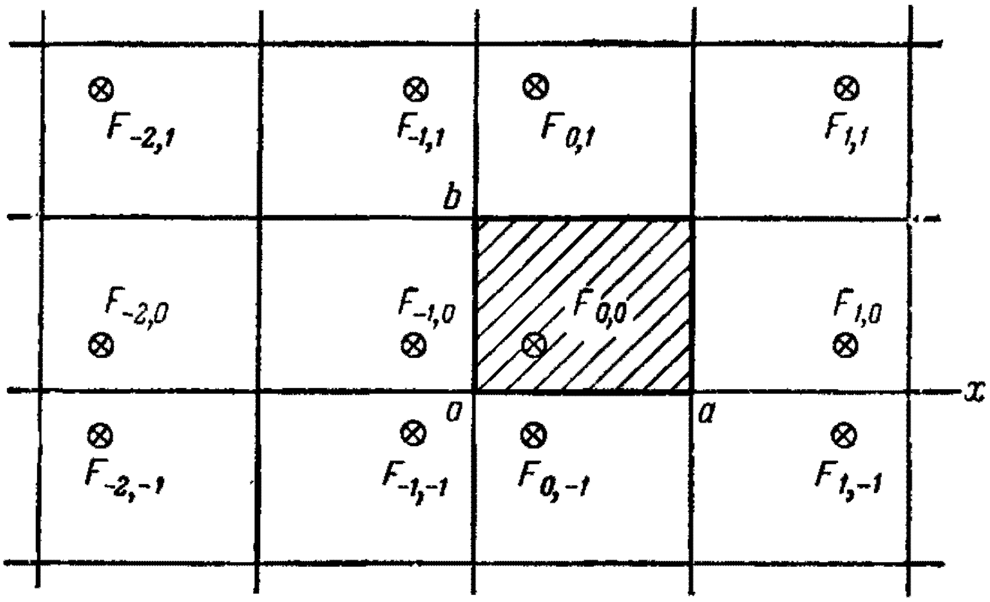

Рис. 15.9. Поперечное сечение волновода показано заштрихованной площадкой; фокус линзы находится в F_{00}. Так как стенки волновода являются совершенными отражателями, то получаются изображения фокуса, как показано на рисунке.

представить такую ситуацию при помощи системы волн типа (15.62)? Заметим прежде всего, что фокальная плоскость $z=0$ характеризуется тем, что все составляющие разложения (15.62) совпадают по фазе. Для другого значения z эти же составляющие имеют уже разные фазы вследствие того, что их скорости V_{lm} различны, и пучок расходится.

Рассматривая только фокальную плоскость $z=0$, имеем:

$$H_{z=0}=\sum_{l,m} A_{lm}\cos l\varphi \cos m\psi \cos \omega t, \qquad (15.75)$$

где

$$\varphi=\frac{\pi x}{a}, \quad \psi=\frac{\pi y}{b}.$$

Мы можем разбить двойной ряд Фурье на два отдельных ряда, один по x, другой по y, и получим:

$$\left.\begin{aligned} &H_{z=0} = X(x)\,Y(y)\cos\omega t,\\ &X(x) = \sum_l X_l \cos l\varphi, \quad 1 \leqslant l \leqslant l_M,\\ &Y(y) = \sum_m Y_m \cos m\psi, \quad 1 \leqslant m \leqslant m_M,\\ &A_{lm} = X_l Y_m. \end{aligned}\right\} \tag{15.76}$$

Мы исключаем $l=0$ и $m=0$, чтобы избавиться от запрещенного члена 0, 0.

Задача симметрична относительно x и y. Рассмотрим только координату x. Мы желаем получить фокус $F_{0,0}$ и все его изображения. Это означает, что мы должны иметь острые максимумы функции $X(x)$ в точках

$$x = x_0 + 2pa \quad \text{и} \quad x = -x_0 + 2pa, \tag{15.77}$$

где p — целое, положительное или отрицательное.

Мы построим ряд с конечным числом членов, представляющий такого рода ситуацию, но предварительно нам потребуются некоторые математические соотношения. Начнем с тождества Лагранжа, записав его в форме, несколько отличной от (13.35):

$$\cos\varphi + \cos 2\varphi + \ldots + \cos l_M\varphi = \sum_1^{l_M} \cos l\varphi =$$

$$= \frac{1}{2}\left[\frac{\sin\left(l_M + \frac{1}{2}\right)\varphi}{\sin\frac{1}{2}\varphi} - 1\right] = \frac{\sin\frac{l_M}{2}\varphi\cos\frac{l_M+1}{2}\varphi}{\sin\frac{\varphi}{2}}. \tag{15.78}$$

Этот результат получается путем простых тригонометрических преобразований, основанных на соотношениях

$$\left(l_M + \frac{1}{2}\right)\varphi = \frac{l_M}{2}\varphi + \frac{l_M+1}{2}\varphi,$$

$$\frac{1}{2}\varphi = -\frac{l_M}{2}\varphi + \frac{l_M+1}{2}\varphi.$$

Возьмем теперь

$$1 \leqslant l \leqslant l_M, \quad X_l = \cos l\varphi_0, \quad \varphi_0 = \frac{\pi x_0}{a}, \tag{15.79}$$

что дает:

$$X(x) = \sum_0^{l_M} \cos l\varphi \cos l\varphi_0 = \frac{1}{2} \sum_1^{l_M} [\cos l(\varphi - \varphi_0) + \\ + \cos l(\varphi + \varphi_0)],$$

или с применением (15.78)

$$X(x) = \frac{\sin \frac{l_M}{2}(\varphi - \varphi_0) \cos \frac{l_M + 1}{2}(\varphi - \varphi_0)}{2 \sin \frac{\varphi - \varphi_0}{2}} + \\ + \frac{\sin \frac{l_M}{2}(\varphi + \varphi_0) \cos \frac{l_M + 1}{2}(\varphi + \varphi_0)}{2 \sin \frac{\varphi + \varphi_0}{2}}. \tag{15.80}$$

Первый член дает острые максимумы при

$$\varphi = \varphi_0 + 2p\pi,$$

когда знаменатель обращается в нуль. Второй член дает максимумы при

$$\varphi = -\varphi_0 + 2p\pi,$$

и это дает полную систему фокуса и его изображений, определяемых равенством (15.77). Система вырождается при $\varphi_0 = 0$, т. е. когда фокус находится в одном из углов сечения волновода.

Последовательность коэффициентов (15.79) дает в точности требуемые условия. Рассмотрим первоначальный фокус и его окрестность

$$x = x_0 + \delta x, \quad \varphi = \varphi_0 + \delta\varphi, \quad \delta x = \frac{a\delta\varphi}{\pi}. \tag{15.81}$$

Вторым членом (15.80) можно пренебречь. Первый член имеет в фокусе ($\delta\varphi = 0$) максимум, равный $l_M/2$, и первые нули при

$$\delta\varphi = \pm \frac{\pi}{l_M + 1},$$

когда косинусный член знаменателя первого члена обращается в нуль. Ширина максимума может быть практически определена половиной расстояния между первыми нулями справа и слева:

$$\left.\begin{aligned} \Delta\varphi &= \frac{2\delta\varphi}{2} = \frac{\pi}{l_M+1}, \\ \Delta x &= \frac{a}{l_M+1} \approx \frac{a}{l_M}. \end{aligned}\right\} \tag{15.82}$$

Когда l_M велико, это условие сводится к прежнему соотношению (15.67).

Аналогичные соображения применимы к координате y. Легко убедиться, что эти условия в точности соответствуют обычному определению разрешающей способности. Рассмотрим квадратную апертуру и возьмем одинаковые пределы вдоль x и y:

$$a = b = r, \quad \Delta x = \Delta y, \quad l_M = m_M = M.$$

Радиус фокального пятна равен

$$\Delta r = \sqrt{(\Delta x)^2 + (\Delta y)^2} = r\sqrt{\frac{1}{l_M^2} + \frac{1}{m_M^2}} = \frac{\sqrt{2}\, r}{M}.$$

Предельный угол θ_0 получается из (15.66):

$$\frac{4\nu^2}{c^2}\sin^2\theta_0 = \left(\frac{l_M}{a}\right)^2 + \left(\frac{m_M}{b}\right)^2 = 2\left(\frac{M}{r}\right)^2, \tag{15.83}$$

так что

$$\Delta r = \frac{c}{\nu\sin\theta_0} = \frac{\lambda}{\sin\theta_0}, \tag{15.84}$$

что и представляет собой обычную формулу для разрешающей способности, как функции длины волны λ и апертурного угла θ_0.

10. Примеры и обсуждение

Читатель может поинтересоваться, почему соотношения (15.67) и (15.69) совпадают, за исключением множителя 1/2 в (15.69). Причина состоит в том, что нам нужна система стоячих волн вдоль x и y. Стоячие волны образуются из двух одинаковых бегущих волн, распространяющихся в противоположных направлениях. Эти две встречные волны возникают вследствие отражения от стенок волновода. В направлении z мы имеем только одну бегущую волну, отраженной волны нет. Этим и объясняется множитель $1/2$ в выражении (15.69) для числа n степеней свободы. В начале раздела 9 уже было пояснено, что все составляющие волны в точности совпадают по фазе в фокальной плоскости $z=0$ и быстро расходятся по фазе для других значений z из-за разных скоростей распространения V_{lm}. Мы могли бы с равным успехом рассматривать образование фокуса сходящимся пучком с максимальным углом θ_0, определяемым равенством (15.83). Пучок расходится по другую сторону фокуса, причем θ_0 есть угол расхождения.

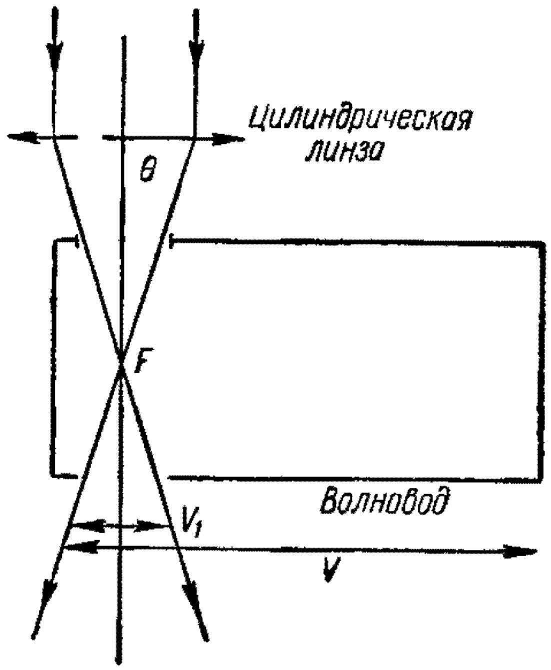

Рис. 15.10. Слабо расходящийся луч, полученный при помощи цилиндрической линзы и пригодный для применения в машине Силарда.

Вместо острого фокуса нам может понадобиться длинный иглообразный пучок. Его можно получить, применяя высокую частоту и малый угол θ. Формулы (15.66) или (15.83) показывают возможность подобрать эти параметры так, чтобы целые индексы l и m оставались большими и чтобы пучок имел малое поперечное сечение.

Возможно также получить с известным приближением плоский пучок заданной толщины. Это можно сделать с помощью цилиндрической линзы при малых углах θ и длинных щелей, ограничивающих пучок в направлении x (но не y), как показано на рис. 15.10. Разложение волны (15.76) дает для фокальной плоскости

$$H_{z=0} = X(x) \cos \omega t$$

вне зависимости от y. Получаемый пучок выражается как

$$H = \sum_{l=1}^{l_M} X_l \cos l\varphi \cos \omega \left[t - \left(\frac{z}{V_l} \right) \right] \qquad (15.85)$$

при $m=0$. Высокая частота ω много выше граничной частоты

$$\omega \gg \omega_{l_M} = \frac{\pi c l_M}{a} \tag{15.86}$$

даст пучок с очень малым расхождением. Эти условия соответствуют задаче о машине Силарда, обсужденной в разделе 6 главы 13. Рис. 15.10, повернутый на 90°, соответствует ситуации рис. 13.5. В этом случае мы имеем $\Delta y = b$, и выражение (15.72) для F сводится к

$$F = \frac{a}{\Delta x}\,\frac{t}{\tau} = \left(\frac{V}{V_1}\right)^2. \tag{15.87}$$

Это есть в точности значение F из (13.55), и совмещение (15.74) с (15.87) идентично результату главы 13, полученному другим методом.

Равенства (15.72) и (15.74) содержат численный множитель A, значения которого были обсуждены в разделе 5 главы 14. При низких частотах этот множитель имеет значение A_n, могущее достигать нескольких единиц, но в некоторых экспериментальных условиях могут потребоваться высокие частоты, при которых A значительно больше:

$$A = \begin{cases} A_n & \text{при} \quad h\nu \ll kT, \\ \dfrac{h\nu}{kT} & \text{при} \quad h\nu \gg kT. \end{cases} \tag{15.88}$$

Первый случай относится к классическим наблюдениям. Второй случай соответствует наблюдению, возмущающему наблюдаемую систему и требующему учета квантовых условий.

11. Заключение

Во всех примерах, обсужденных в этой и предыдущих главах, мы нашли, что информация, полученная из эксперимента, связана с точностью логарифмической зависимостью

$$\Delta I = k \ln \mathscr{A} \tag{15.89}$$

и что сопутствующее увеличение энтропии удовлетворяет неравенству

$$\Delta S \geqslant \Delta I. \tag{15.90}$$

Эти примеры послужили, таким образом, подтверждением обобщенного принципа Карно, который мы назвали негэнтропийным принципом информации:

$$\Delta (S - I) \geqslant 0. \tag{15.91}$$

Мы исследовали эффективность наблюдения, которая в силу негэнтропийного принципа удовлетворяет условию

$$\mathscr{E} = \frac{\Delta I}{\Delta S} \leqslant 1. \tag{15.92}$$

В случае прямого измерения длины (см. разделы 2 и 3 этой главы) эффективность очень близка к единице и негэнтропийная цена наблюдения мала. Во всех других рассмотренных случаях, однако, эффективность много меньше единицы, и увеличение энтропии возрастает с повышением точности по меньшей мере линейно:

$$\Delta S = a\mathscr{A} + b. \tag{15.93}$$

Постоянная b обычно мала. Коэффициент a в некоторых случаях есть постоянная, а в других является логарифмической функцией $\mathscr{A}$. Как интерферометрический метод измерения расстояний, так и измерение положения при помощи микроскопа приводят к увеличению энтропии, выражаемому равенством (15.93). Очевидно, в этих случаях эффективность быстро убывает с повышением точности.

Мы видели, что эффективность, вообще говоря, выше при низких частотах. Исключение составляют случаи, когда из-за очень высоких требований к точности, оказывают влияние квантовые условия. При таких обстоятельствах предпочтительно применение высоких частот, даже если эффективность остается очень низкой. Однако в этом случае мы можем повысить эффективность при помощи подходящих устройств, уменьшающих рассеяние энергии. По существу, такая процедура сводит высокочастотную задачу к низкочастотной, и соответствующие результаты для низких частот могут, таким образом, применяться во всех случаях.

Следует, быть может, заметить, что прямой метод измерения расстояния, дающий очень большую эффективность, нелегко применить в лаборатории для измерения малых расстояний из-за трудностей в построении соответствующей аппаратуры. Очень малые расстояния всегда измеряются интерферометрическим методом, и, таким образом, практически мы имеем дело с малыми эффективностями, полученными в большинстве рассмотренных примеров.

ГЛАВА 16

ТЕОРИЯ ИНФОРМАЦИИ, ПРИНЦИП НЕОПРЕДЕЛЕННОСТИ И ФИЗИЧЕСКИЕ ПРЕДЕЛЫ НАБЛЮДАЕМОСТИ

1. Общие замечания

Мы обсудили целый ряд экспериментов и получили общий результат: негэнтропийный принцип информации, согласно которому информация, полученная из физического наблюдения, должна быть оплачена увеличением энтропии в лаборатории. В среднем это увеличение энтропии больше, чем полученная информация, если обе величины выражаются в одних и тех же единицах. Эти условия представляют собой новое ограничение возможностей наблюдения, и мы сравним их теперь с известным соотношением неопределенности.

На протяжении последних пятидесяти лет велась обширная дискуссия по поводу взаимоотношения классической физики и квантовой физики. Точка зрения Н. Бора, выраженная в его принципе *дополнительности,* и его толкование соотношения неопределенности остаются, по-видимому, здравым и логичным подходом к проблеме. Недавняя дискуссия [1]) ясно показала важность этих вопросов.

Физика начинается исторически и фактически на классическом уровне с опытов человеческого масштаба и основных определений (длина, время, масса, общие законы), основанных на этих опытах. От этой прочной основы физик переходит к исследованию проблем атома и открывает квантовые условия

---

[1]) W. Elsasser, Phys. Rev. **52,** 987 (1937); Phil. Sci. **18,** 300 (1951); в книге «Louis de Broglie Physicien et Penseur», p. 87, Albin Michel, Paris, 1953; N. Bohr, L. Rosenfeld, Phys, Rev. **78,** 794 (1950); L. Rosenfeld в книге «Louis de Broglie Physicien et Penseur», p. 43, Albin Michel, Paris, 1953; D. Bohm, Quantum Theory, Prentice-Hall, NY, 1951.

со всеми ограничениями, вытекающими из принципа неопределенности, и квантовую механику. Для совместимости необходимо, чтобы эти новые законы сводились в пределе при больших квантовых числах к классическим законам. Н. Бор неоднократно подчеркивал тот факт, что измерительный прибор всегда описывается в терминах классической физики, но он ввел также *принцип соответствия*, согласно которому квантовые законы должны в пределе переходить в классические законы, когда постоянная Планка h может считаться пренебрежимо малой.

Бор говорит, что «квантовая теория предполагает классические концепции при описании этого уровня».

Как можно было бы определить классическую физику? Эльзассер характеризует этот этап возможностью невозмущающих экспериментов, в которых между наблюдателем и наблюдаемой системой может быть сделано ясное различие. Типичным примером является оптическое наблюдение механической системы, например, падающего тела. Эксперимент не возмущает систему заметным образом и может быть повторен много раз с совпадающими результатами. Он обладает фундаментальным свойством воспроизводимости. Нелегко установить точные условия таких невозмущающих наблюдений. Несколько положений можно, однако, подчеркнуть: наблюдаемая система должна содержать много частиц и большое число квантов. Точность эксперимента должна быть не слишком высока. Очевидно условие, что опыт должен производиться вдали от границы соотношения неопределенности:

$$\Delta p \cdot \Delta q \gg h \quad \text{и} \quad \Delta E \cdot \Delta t \gg h. \tag{16.1}$$

Погрешности Δp, Δq, ΔE и Δt должны быть не слишком малы для обеспечения возможности одновременного определения p, q, E и t.

Мы можем также ввести другое условие, вытекающее из приведенных в настоящей книге рассуждений. Энтропийная цена ΔS наблюдения должна быть пренебрежимо малой по сравнению с полной энтропией S_0 наблюдаемой системы:

$$\Delta S \ll S_0,$$

и, следовательно,

$$\Delta I \leqslant \Delta S \ll S_0, \tag{16.2}$$

так как получаемая информация ΔI всегда меньше энтропийной цены ΔS. Эти условия несомненно необходимы, но могут быть недостаточными. Общая схема такова: сначала система и измерительное устройство не связаны между собой. На время наблюдения устанавливается связь с некоторой степенью взаимодействия. После наблюдения система и измерительное устройство не связаны. Во время наблюдения значения p, q, E и S системы претерпевают изменения, и эти неизвестные изменения совершаются на протяжении промежутка времени Δt, на котором длится эксперимент. Наблюдение может считаться *невозмущающим*, если эти неизвестные изменения чрезвычайно малы по сравнению с точностью эксперимента. Условия (16.1) и (16.2), таким образом, необходимы. Рассмотрим, например, случай (16.2). В результате эксперимента полная энтропия наблюдаемой системы и измерительного устройства увеличивается на ΔS. Если системы после эксперимента не связаны, то часть этого приращения ΔS будет относиться к наблюдаемой системе, а остальное — к измерительным приборам, но отчетливо различить эти две составляющие, вообще говоря, невозможно. Поэтому мы должны потребовать, чтобы полное увеличение ΔS было всегда очень мало по сравнению с энтропией S_0 наблюдаемой системы.

Эльзассер рассматривает в качестве примера задачу о молекулах газа. В принципе возможно измерить очень точно положения всех молекул, но для такого наблюдения потребовался бы луч света очень высокой частоты. Взаимодействие этого света с молекулами сообщило бы им очень большие импульсы и существенно изменило бы всю систему. Это, очевидно, не является невозмущающим наблюдением и противоречит условиям (16.1) и (16.2). С другой стороны, измерения объема, давления и плотности могут выполняться с соблюдением условий (16.1) и (16.2), если только точность не слишком высока.

2. Наблюдение есть необратимый процесс

Во избежание недоразумений и для разъяснения положения нужно уточнить, что увеличение энтропии ΔS, необходимое для наблюдения, не имеет никакого отношения к изменениям энтропии, происходящим в самой системе, пока никакое измерение не производится.

Рассмотрим, например, систему, в которой происходит адиабатическое преобразование[1]). Энтропия внутри системы остается неизменной, но адиабатическое преобразование проходит незамеченным и ненаблюдаемым, если система не связана с устройством, позволяющим осуществить наблюдение.

Мы предполагаем, что это адиабатическое преобразование вносит изменения в физические переменные (p, v, T и т. д.); пока мы эти переменные не измеряем, у нас нет возможности узнать, что происходит. Для того чтобы наблюдать p, v, T и т. д., мы должны соединить систему с какими-либо физическими приборами, а этим мы вносим некоторое возмущение. Наблюдение само по себе является необратимым процессом, увеличивающим энтропию лаборатории. Нам всегда приходится отсчитывать положение стрелки на шкале (измерение расстояний) и отмечать промежутки времени. Эти основные наблю-

---

[1]) Адиабатическое преобразование описывается как последовательность равновесных состояний. Если система находится внутри цилиндра длиной L с движущимся поршнем, то перемещение поршня должно быть очень мало на протяжении времени δt, достаточно большого, чтобы возмущение могло распространиться много раз туда и обратно на расстоянии L. Это требует выполнения условия $\frac{\delta x}{\delta t} = v \ll c$, где c — скорость возмущения. Для излучения c — это скорость света; для газа c — тепловая скорость молекул; для твердого или жидкого тела c есть скорость звука.

Адиабатические преобразования квантованных систем рассматривались многими авторами. Бом (D. Bohm, Quantum Theory, гл. 20, Prentice Hall, NY, 1951) устанавливает условие

$$\frac{h}{(\Delta E)^2}\frac{\partial E}{\partial t} \ll 1,$$

где E — энергия, ΔE — разность энергии до следующего квантового состояния. Рассмотрим стоячие волны на длине L с N узлами, так что $L = \frac{N\lambda}{2} = \frac{Nc}{(2\nu)}$ и, следовательно, $E = nh\nu = \frac{hcnN}{(2L)}$ и $\Delta E = h\nu = hcN/(2L)$, если в нашей стоячей волне имеется n квантов $h\nu$. Тогда

$$\frac{\partial E}{\partial t} = -\frac{hcnNv}{2L^2}, \qquad \text{где} \qquad v = \frac{\partial L}{\partial t},$$

и условие Бома для адиабатического преобразования имеет вид

$$\frac{2nv}{Nc} \ll 1$$

— соотношение, сходное с нашим предыдущим неравенством $v \ll c$.

дения обсуждались в главе 15, и всякое физическое наблюдение может быть к ним сведено.

В этом примере мы ясно видим различие между изменениями энтропии, происходящими в изолированной (и, следовательно, ненаблюдаемой) системе, и увеличением энтропии ΔS, требуемым для наблюдения и обусловленным связью между наблюдаемой системой и измерительными приборами.

Наблюдение есть, по существу, необратимый процесс. Исходя из чисто термодинамической точки зрения, мы показали, что никакое наблюдение не может быть сделано без соответственного увеличения энтропии, либо самой физической системы, либо оборудования, применяемого в эксперименте и связанного с системой на протяжении наблюдения:

$$\Delta S > 0. \tag{16.3}$$

Если мы примем более общую точку зрения, описанную в предыдущих главах, то мы включим информацию I в энтропию и получим условие

$$\left.\begin{aligned} &\Delta (S - I) \geqslant 0, \\ &\Delta S \geqslant \Delta I > 0. \end{aligned}\right\} \tag{16.4}$$

В обоих случаях увеличение энтропии неизбежно, если только производится наблюдение, а так как увеличение энтропии означает необратимость, то необратимость оказывается свойством наблюдения.

Сходные замечания были ранее сделаны многими авторами, в особенности фон Нейманом, при обсуждении некоторых парадоксальных проблем квантовой механики.

Нельзя говорить о том, что система находится в некотором состоянии, не измеряя какой-либо характеризующей состояние величины. Процесс измерения необратим и определяет, хотя бы отчасти, будущее поведение системы.

3. Общие ограничения точности физических измерений

Мы встретились с трудностями в измерении расстояний с очень высокой точностью, в особенности, если речь идет об очень малых расстояниях. Характеристическая длина Δx_T, соответствующая данной температуре T и определенная равенством (15.12) как

$$\Delta x_T = \frac{hc}{kT} \approx \frac{1{,}44}{T}\, см, \tag{16.5}$$

где T измерено в градусах Кельвина, является, по существу, границей между большими и малыми расстояниями. Возрастающая трудность измерения все меньших и меньших расстояний указывает на то, что основные допущения евклидова пространства (и времени) должны рассматриваться лишь как идеализация, которая не может и не должна считаться действительной в пределе при чрезвычайно малых расстояниях. Математическое определение *бесконечно малых* расстояний соответствует невозможным физическим условиям. Мы вернемся к этому вопросу в разделе 4.

Связь между квантовой механикой и теорией информации легко усмотреть из различных рассмотренных нами задач. Квантовые ограничения необходимы в обсуждении, так как мы живем в квантовом мире. Если же, однако, эксперимент производится на классическом уровне, то мы замечаем, что постоянная Планка h не появляется в окончательных результатах, которые содержат только постоянную Больцмана k. Мы видим, таким образом, что негэнтропийный принцип информации является действительно новым принципом и не может быть сведен к квантовым соотношениям и соотношениям неопределенности. Постоянная Планка h не может быть устранена из окончательных результатов, если выполняемый эксперимент достигает квантового уровня, и обе теории смыкаются на этом уровне более сложным образом.

Это может быть иллюстрировано задачами, рассмотренными в разделах 3 и 5 главы 15. Большие расстояния (случай A) соответствуют классической задаче. Измерение очень малых расстояний достигает квантового уровня. При измерении чрезвычайно малых расстояний при очень низких температурах играют роль ограничения как со стороны квантовых условий, так и со стороны теории информации.

Во всем нашем общем рассуждении мы стремились пока свести проблемы к их основным элементам, и мы систематически игнорировали деградацию энергии при всякого рода усилении. Усилители всегда необходимы в экспериментальных устройствах. Они выполняют в основном следующее назначение: преобразовать квантовый эффект путем увеличения его амплитуды в такой эффект, который может быть зарегистрирован прибором, действующим на классическом уровне. Роль усилителя была очень ясно определена Эльзассером. Несколько примеров поясняют общее положение: фотографическая пластинка может реагировать на один-единственный квант, а процесс проявления есть усиление, дающее видимую черную точку (классический эффект). Фотоэлемент может зарегистрировать одиночный квант, будучи соединен с электронным усилителем. Счетчик Гейгера может обнаружить одиночную заряженную частицу, а усиление происходит здесь вследствие ионного заряда. Камера Вильсона дает аналогичный пример.

До тех пор пока усилитель не содержит цепи обратной связи, он не изменяет наблюдение, и его можно игнорировать, хотя, конечно, он дает добавку к увеличению энтропии, связанному с наблюдением. Усилители с обратной связью составляют очень важную проблему, и их роль была подчеркнута Винером в его теории кибернетики.

Как отмечалось выше, мы имеем классические условия при измерении в случае A (см. разделы 3 и 5 главы 15), когда ΔL больше, чем Δx_T. Исключительные обстоятельства возникли в случае B при

$$\Delta L < \Delta x_T, \tag{16.6}$$

что соответствует высокой точности, очень низкой температуре и чрезвычайно малым расстояниям. При таких условиях мы должны делать различие между энергией ΔE, требуемой для наблюдения, и той ее долей, которая должна перейти в тепло и вызвать увеличение энтропии. Во многих практических случаях вся энергия ΔE рассеивается и переходит в тепло, но мы показали в главах 14 и 15, что большей части этого рассеяния можно избежать и что увеличение энтропии ΔS может быть (по крайней мере теоретически) сведено к значению, получаемому при классических наблюдениях и лишь немного превышающему количество полученной информации ΔI.

В главе 15 мы исследовали проблему измерения длины и нашли, что наиболее экономичным является метод, обсужденный в разделах 2 и 3. Эффективность была дана формулой (15.17):

$$\mathscr{E} = \frac{\Delta I}{\Delta S} = 1 + \frac{\ln(\ln 2)}{\ln n} \tag{16.7}$$

При точности, равной, скажем, 100:

$$\frac{L}{\Delta L}=\mathscr{A}=n=100 \tag{16.8}$$

(16.7) дает:

$$\mathscr{E}=1+\frac{\ln(\ln 2)}{\ln 100}=1-\frac{0{,}3667}{\ln 100}\approx 1-0{,}08=0{,}92. \tag{16.9}$$

Интерферометрические измерения, обсужденные в разделах 5 и 6, и микроскопические наблюдения, обсужденные в разделе 8, дают значительно меньшую эффективность.

Классические условия (случай A) соответствуют случаю, когда энергия, требуемая для наблюдения, полностью превращается в тепло, и необходимая энергия, таким образом, непосредственно выражается соотношением

$$\Delta E=T\,\Delta S. \tag{16.10}$$

Это есть положение, когда погрешность ΔL в измерении длины не слишком мала и остается больше Δx_T.

В случае B, однако, мы желаем уменьшить погрешность ΔL и достичь условия (16.6), и ситуация совершенно изменяется. Можно поддержать низкое значение теоретического минимума увеличения энтропии ΔS, но требуемая для эксперимента энергия ΔE становится все больше:

$$\Delta E>T\,\Delta S. \tag{16.11}$$

Этот пункт требует более подробного рассмотрения, к которому мы и переходим.

4. Пределы евклидовой геометрии

Рассмотрим фундаментальные трудности, связанные с измерением чрезвычайно малых расстояний, и общие ограничения, которые эти трудности накладывают на физические определения. Для измерения длины L с погрешностью ΔL необходимо применить излучение с длиной волны

$$\lambda\leqslant 2\Delta L. \tag{16.12}$$

Мы тщательно обсудили это условие в главе 15 и сравнили различные методы измерения; было найдено, что в отдельных случаях могут применяться длинные волны, но что их энергетическая или негэнтропийная цена много выше, чем для коротких волн. Малая длина волны λ (см. (16.12)) соответствует большому импульсу

$$p=\frac{h}{\lambda}\geqslant\frac{h}{2\Delta L}, \qquad (16.13)$$

и это излучение будет взаимодействовать с наблюдаемой системой и сообщать ей изменение импульса ΔP, которое может лежать в пределах от $-p$ до $+p$, и, следовательно,

$$\Delta P=2p\geqslant\frac{h}{\Delta L}, \qquad (16.14)$$

что представляет собой обычное соотношение неопределенности.

Новое обстоятельство, которое следует подчеркнуть, состоит в том, что большой импульс p всегда связан с большой энергией, и в соответствии с этим измерение длины с очень малой погрешностью ΔL требует всегда большого расхода энергии.

Предположим сначала, что мы применяем электромагнитное излучение. Случай измерений другого рода рассматривается в следующем разделе. Если мы применяем свет, то длина волны λ соответствует кванту $h\nu$:

$$h\nu=\frac{hc}{\lambda}\geqslant\frac{hc}{2\Delta L}. \qquad (16.15)$$

Наименьшее количество энергии ΔE, с помощью которого мы можем выполнить эксперимент, равно $h\nu$, и, следовательно,

$$\Delta E=h\nu, \quad \Delta E\,\Delta L\geqslant\frac{hc}{2}. \qquad (16.16)$$

Не существует точной границы для малых расстояний или длин, которые могут быть измерены, но количество энергии, требуемой для эксперимента, растет как $(\Delta L)^{-1}$ и быстро становится огромным препятствием для измерений.

Эксперименты, выполняемые в условиях случая B, разумеется, сильно возмущают наблюдаемую систему. Должны применяться очень высокие частоты, а это вызывает большие импульсы отдачи, образование пар и т. д.

Эти замечания ясно показывают, что действительная стоимость оборудования и аппаратуры ядерных исследований находится в соответствии с высокой энергетической ценой. Наконец, стоимость эксплуатации большого циклотрона или синхроциклотрона пропорциональна большой энергии, требуемой для каждого наблюдения и рассеиваемой в мишени[1]).

Математики могут определять бесконечно малые расстояния, но это есть абстракция, соответствующая невозможным физическим условиям. Рассмотрим, например, некоторую лабораторию с определенным ограниченным запасом энергии. Если вся эта энергия необходима для измерения некоторого расстояния, то это расстояние есть наименьшее расстояние, которое может быть измерено в данной лаборатории.

Часто высказывалось мнение, что можно было бы избежать многих трудностей квантовой теории путем введения своего рода наименьшей длины. Фиксированная наименьшая длина едва ли могла бы быть обоснована в свете предыдущих замечаний. Мы не решились распространить пример лаборатории с ограниченным запасом энергии на всю вселенную, так как дать точные определения ее протяженности и вместимости нельзя. Тем не менее мы можем сказать, что измерения все более коротких расстояний представляют неизменно возрастающие трудности, и возрастающая цена наблюдения таких расстояний может являться новым фактором, который нужно ввести в теорию.

Равенство (16.16) очень похоже на (15.33):

$$\Delta E\,\Delta L \approx \frac{nhc}{2}, \tag{16.17}$$

и весьма сходно с соотношением неопределенности (16.14). Однако условия (16.16) имеют значение, совершенно отличное от соотношения неопределенности. В соотношении неопределенности ΔE означает погрешность в определении E. В (16.16), с другой стороны, ΔE есть количество энергии, необходимое для эксперимента и превращаемое, хотя бы частично, в тепло на протяжении процесса измерения L с приближением ΔL. Оба соотношения имеют, таким образом, только формальное сходство; их физический смысл совершенно различен.

---

[1]) L. Brillouin, J. Appl. Phys. **24**, 1152 (1953); **25**, 887 (1954).

5. Возможность использования тяжелых частиц вместо фотонов

Мы обсудили различные экспериментальные устройства в предположении, что для освещения применяется луч света. Нужно рассмотреть возможность применения других волн, таких, как волны, связанные с материальными частицами массы m, вместо фотонов, имеющих нулевую массу.

Начнем опять с равенств (16.12) и (16.13), приложимых к любому виду излучения, и рассмотрим применение частиц с массой покоя m_0. Мы должны подсчитать кинетическую энергию такой частицы. Частица имеет импульс p и кинетическую энергию E_{kin}, равные

$$p=\frac{h}{\lambda}=\frac{m_0 v}{\sqrt{1-\beta^2}}, \quad \text{где} \quad \beta=\frac{v}{c}, \tag{16.18}$$

$$E_{\text{kin}}=m_0c^2\left(\frac{1}{\sqrt{1-\beta^2}}-1\right)=c\sqrt{m_0^2c^2+p^2}-m_0c^2. \tag{16.19}$$

В предельных случаях можно воспользоваться приближением: для фотонов или частиц с малой массой покоя имеем

$$\left.\begin{aligned} m_0c &\ll p=\frac{h}{\lambda}, \\ E_{\text{kin}} &\approx pc-m_0c^2, \end{aligned}\right\} \tag{16.20}$$

тогда как для тяжелых частиц

$$\left.\begin{aligned} m_0c &\gg p=\frac{h}{\lambda}, \\ E_{\text{kin}} &\approx \frac{p^2}{2m_0}. \end{aligned}\right\} \tag{16.21}$$

При заданном p и переменной массе покоя m_0 кривая зависимости E_{kin} от m_0 имеет вид, показанный на рис. 16.1.

На первый взгляд применение тяжелых частиц представляет интерес, так как уменьшает количество энергии E_{kin}, требуемое для эксперимента. Это, однако, невозможно, если измеряются очень малые расстояния d. Положим, например, согласно (16.14),

$$\Delta L=d=10^{-13}, \quad p\approx 3\cdot 10^{-14}.$$

Мы должны сравнить p с m_0c для того, чтобы определить, какое из приближенных соотношений (16.20) или (16.21) применимо:

электрон: $m_0 \approx 10^{-27}$, $m_0 c \approx 3 \cdot 10^{-17}$, случай (16.20), протон: $m_0 \approx 2 \cdot 10^{-24}$, $m_0 c \approx 6 \cdot 10^{-14}$, промежуточный случай, тяжелое ядро: $m_0 \approx \frac{5}{3} \cdot 10^{-22}$, $m_0 c \approx 5 \cdot 10^{-12}$, случай (16.21).

Применение очень тяжелых ядер может уменьшить требуемую энергию, но эти ядра имеют большие поперечные сечения и непригодны для измерения длин, меньших их диаметров.

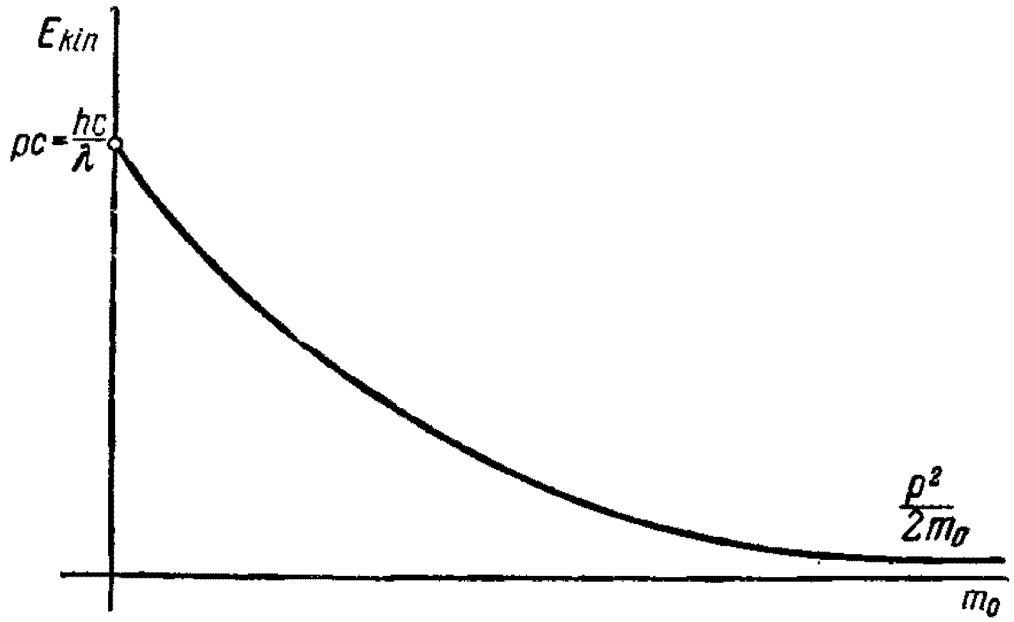

Рис. 16.1. Кинетическая энергия как функция m_0 при данном и фиксированном импульсе p.

Кривая приближается к $p^2/2m_0$ асимптотически при $m_0 c \gg p$.

В общем фотоны, мезоны, электроны и нуклоны являются единственными частицами, которые практически могут применяться в освещающем луче, и они дают сходные результаты, когда расстояние d чрезвычайно мало. Трудности, обсужденные в разделах 3 и 4, являются действительно фундаментальными и не могут быть устранены.

6. Соотношения неопределенности в экспериментах с микроскопом

Методы теории информации были применены в разделах 8 и 9 главы 15 при обсуждении наблюдений под микроскопом. Это типичная задача, часто приводимая в связи с соотношением неопределенности. В рассуждении обычно предполагается, что для получения информации о положении частицы используется один фотон. Это, очевидно, слишком упрощенное предположение. Единственный фотон, наблюдаемый в точке I рис. 16.2, может прийти откуда угодно и не обязательно указывает наличие рассеивающей частицы в M. Рассуждение главы 15 ясно показывает, что для четкого определения фокального изображения в I нужно большое число фотонов. При одном-единственном фотоне получается соотношение неопределенности

$$\Delta x \cdot \Delta p_x \approx h, \qquad (16.22)$$

тогда как при использовании многих фотонов неопределенность становится много больше:

$$\Delta x \cdot \Delta p_x \approx Ah, \qquad (16.23)$$

где численный множитель A может быть очень большим. Это типично для экспериментального устройства, в котором положение x измеряется при помощи некоторой оптической системы без применения материальных экранов, диафрагм или затворов. Рассмотрим этот вопрос подробнее.

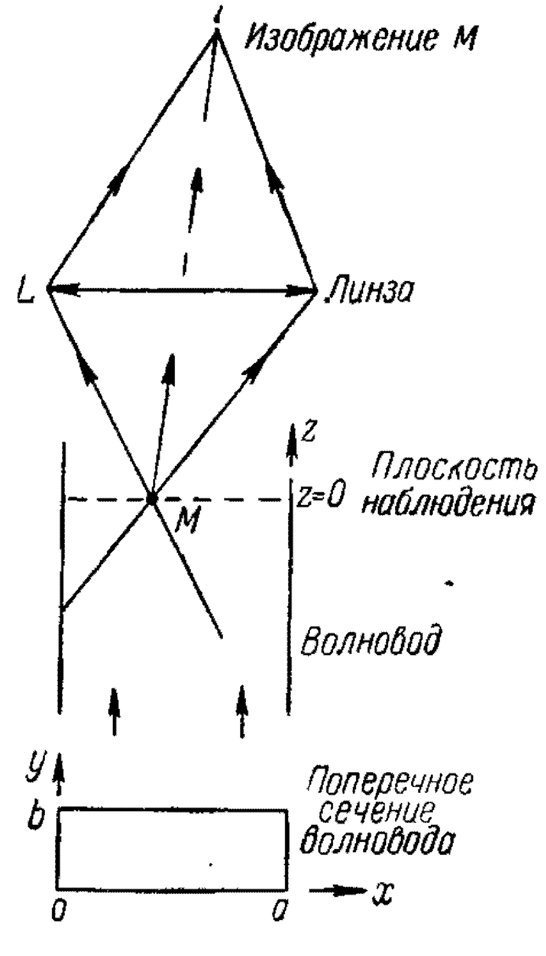

Рис. 16.2. Условия наблюдения, когда волновод с линзой применяется для обнаружения частицы в фокусе M.

Условия наблюдения показаны схематически на рис. 16.2. Мы хотим измерить x на интервале a с погрешностью Δx, но мы не интересуемся значениями y или b, или Δy. В выражении (15.62) для собственных волн мы ограничиваемся множителем $\cos(\pi l x/a)$ и значениями l и игнорируем множитель $\cos(\pi m y/b)$ и значение m всякий раз, когда это возможно. Мы можем, например, выбрать $b > a$, беря основную волну $l = 0$ и $m = 1$, так как эти значения дают низшую граничную частоту (см. (15.65)). Эта волна будет представлять луч, проходящий сквозь волновод и падающий на частицу M:

$$\text{падающий луч } l = 0, \quad m = 1. \qquad (16.24)$$

Этот падающий луч представляется наложением двух плоских волн, для которых согласно (15.66)

$$\cos \varphi_y = \pm \frac{\lambda}{2b}.$$

Мы можем также считать, что m принимает как положительные, так и отрицательные значения, так что $m = \pm 1$. Углы φ_x и φ_y были определены в разделе 8 главы 15, и мы имеем, вообще,

$$\cos \varphi_x = \frac{l\lambda}{2a}, \quad \cos \varphi_y = \frac{m\lambda}{2b}$$

и будем рассматривать l и m как положительные или отрицательные целые числа.

Свет, рассеянный частицей M, может содержать собственные волны со всеми возможными значениями l и m. Мы можем устранить большую часть составляющих, применяя цилиндрическую линзу длиной b в L. Устройство исключает все члены, кроме $m = 0$. Таким образом, можно описать процесс рассеяния как переход

$$\left.\begin{array}{l} \text{для } l\text{: от } 0 \text{ до } l, \\ \text{для } m\text{: от } 1 \text{ до } 0. \end{array}\right\} \qquad (16.25)$$

Однако для того чтобы учесть положительные и отрицательные знаки $\cos \varphi_x$ и $\cos \varphi_y$, мы должны рассматривать переходы

$$\left.\begin{array}{l} \text{для } l\text{: от } 0 \text{ до } \pm l, \\ \text{для } m\text{: от } \pm 1 \text{ до } 0. \end{array}\right\} \qquad (16.25a)$$

Изменение δl за время рассеяния фотона дает частице M импульс δp_x:

$$\delta p_x = \frac{h}{\lambda} \delta (\cos \varphi_x) = \frac{hl}{2a}. \qquad (16.26)$$

Знак $\pm$ содержится теперь в l, и значение l может пробегать интервал от $-l_M$ до -1 или от $+1$ до $+l_M$. Значение $l = 0$ исключено, так как l и m не могут оба равняться нулю. Условие (15.69) требует использования n членов в каждой волне для того, чтобы t было определено с погрешностью $\tau = \Delta t$. Так как импульсы δp_x могут быть как положительными, так и отрицательными, мы составляем сумму квадратов, обозначив ее $(\Delta p_x)^2$:

$$(\Delta p_x)^2 = \sum_n \sum_l (\delta p_x)^2 = \\ = n\left(\frac{h}{2a}\right)^2 4\,(1 + 4 + 9 + \ldots + l_M^2), \tag{16.27}$$

или

$$\Delta p_x = \frac{h}{a}\left|\frac{n}{3}\, l_M\,(l_M + 1)\left(l_M + \frac{1}{2}\right)\right|^{\frac{1}{2}}.$$

Множитель 4 в (16.27) появляется потому, что имеются четыре случая (16.25а), соответствующих комбинациям $\pm l$ и ± 1. Пользуясь теперь значением l_M из (15.68), получаем:

$$\Delta p \cdot \Delta x = h\left|\frac{\tau}{6\Delta t}\left(1 + \frac{\Delta x}{a}\right)\left(\frac{a}{\Delta x} + \frac{1}{2}\right)\right|^{\frac{1}{2}}. \tag{16.28}$$

Это есть уточненное соотношение неопределенности, учитывающее определение времени и пределы изменения переменной x в эксперименте с микроскопом.

Значение τ может изменяться в диапазоне от $2\Delta t$ до бесконечности. Значение $\tau = \Delta t$ не должно применяться, так как оно не дает информации. Параметр a может изменяться от Δx до бесконечности. Таблица 16.1 дает результаты вычисления для нескольких простых случаев.

Таблица 16.1

$\frac{\tau}{\Delta t}$	$\frac{a}{\Delta x}$	$\left\|\frac{\tau}{6\Delta t}\left(1 + \frac{\Delta x}{a}\right)\left(\frac{a}{\Delta x} + \frac{1}{2}\right)\right\|^{\frac{1}{2}}$
2	1	1
2	2	$\frac{1}{2}\sqrt{5} = 1{,}12$
2	3	$\frac{1}{3}\sqrt{14} = 1{,}25$
3	1	$\sqrt{1{,}5} = 1{,}22$
3	2	$\frac{1}{2}\sqrt{7{,}5} = 1{,}37$
большие значения		$\left(\frac{a\tau}{6\Delta t \cdot \Delta x}\right)^{\frac{1}{2}}$

Если точность очень низка и если определение происходит в основном за счет экранов и препятствий ($a=\Delta x$), то имеет место обычная неопределенность. Нас могут, однако, интересовать случаи, когда определение времени и измерение расстояния выполняются первоначально при помощи оптической системы, что ведет к гораздо большей неопределенности. Возьмем, к примеру,

$$\frac{\tau}{\Delta t}=6000, \quad \frac{a}{\Delta x}=1000: \quad \Delta x \Delta p_x = h\,10^3. \qquad (16.29)$$

Мы приходим, таким образом, к заключению, что неопределенность может стать очень большой при измерении положения *свободной* частицы, т. е. частицы, могущей свободно перемещаться на большое расстояние на протяжении долгого времени без возмущений со стороны материальных экранов, затворов или других механических устройств. Такие механические устройства, очевидно, уже не могут применяться, когда речь идет о малых расстояниях, меньше 1Å, и наше рассуждение имеет большое значение для этого диапазона.

7. Измерение импульса

В эксперименте с микроскопом мы предполагали, что нам известна начальная скорость частицы, и мы пытались измерить ее положение. Результаты были не очень благоприятны. Испробуем обратную процедуру. Начнем с измерения положения x в момент t с большой точностью. Это дает частице большой неизвестный импульс (см. (16.28)), но это не играет роли, потому что мы не знаем начальной скорости. Затем мы можем измерить скорость v_x, а следовательно, и p_x при помощи эффекта Допплера[1]). Обычное рассуждение, дающее

$$\Delta x\, \Delta p \approx h, \qquad (16.30)$$

основано на предположении, что одного фотона $h\nu'$ достаточно для хорошего измерения частоты ν' в эффекте Допплера при условии, что может применяться длинный волно-

---

[1]) См., например, D. Bohm, Quantum Theory, стр. 105, Prentice-Hall, NY, 1951.

вой пакет общей длительностью θ:

$$\Delta\nu'\,\theta \approx 1. \tag{16.31}$$

Исследуем теперь вопрос более тщательно, с использованием замечаний, основанных на теории информации. Мы намерены показать, что одного фотона недостаточно, а также обсудим связь числа фотонов, необходимого для эксперимента, с точными спецификациями экспериментальной процедуры. Это приведет нас к видоизмененному соотношению неопределенности

$$\Delta x\,\Delta p_x \approx nh, \quad n \geqslant 1, \tag{16.32}$$

где n зависит от условий наблюдения. Этот практический предел может быть снижен до прежнего значения (16.30), если условия выбраны так, чтобы n приближалось к единице.

Обсуждение информации показало, что нужно всегда определять диапазон изменения и погрешность измеряемых величин. В данной задаче мы берем:

$$\begin{array}{l} \text{время наблюдения } t\text{: погрешность } \Delta t=\theta\text{, диапазон } \tau, \\ \text{скорость частицы } v_x\text{: погрешность } \Delta v_x\text{, диапазон } u. \end{array} \tag{16.33}$$

В прежних примерах мы определяли диапазон τ во времени, предполагая, что вспышки длительностью θ повторяются с интервалом τ. Такие световые сигналы могут быть получены путем наложения различных частот:

$$n = \frac{\tau}{2\Delta t} = \frac{\tau}{2\theta}. \tag{16.34}$$

Эти частоты занимают полосу $\delta\nu = 1/\theta$. Мы определяем также в (16.33) диапазон скоростей, с которыми хотим работать. Будем конструировать систему, способную измерять скорости от v_0 до v_0+u с погрешностью Δv. Это значит, что необходимо применить m резонаторов, настроенных на m различных допплеровых частот ν', причем каждый резонатор способен принимать полосу частот $\Delta\nu'$, соответствующую Δv;

$$m = \frac{u}{\Delta v}, \quad \Delta\nu' = \nu\frac{\Delta v}{c}, \tag{16.35}$$

где ν — средняя исходная частота. Измерение возможно при условии, что исходная полоса $\delta\nu$ в (16.34) меньше, чем $\Delta\nu'$, введенная равенством (16.35). Вопрос об эксперименте, требующем m одновременных наблюдений, обсуждался в главе 14, и было показано, что нужно выбрать для каждого резонатора энергетический предел E_l, равный

$$E_l = A_m kT. \tag{16.36}$$

Если в резонаторе наблюдается энергия $E < E_l$, то резонатор считается невозбужденным, тогда как $E > E_l$ означает, что резонатор возбужден. В случае, когда для эксперимента принята высокая надежность r, имеем (см. главу 14):

$$A_m \approx \ln(rm) \quad \text{при} \quad r \gg 1 \quad \text{и} \quad m \gg 1. \tag{16.37}$$

Этим условиям можно удовлетворить, используя только один квант $h\nu'$, если

$$h\nu' > A_m kT.$$

Энергия E_l будет соответствовать одному кванту на каждой частоте ν', если взять

$$h\nu' = E_l \approx A_m kT, \tag{16.38}$$

и мы имеем n различных частот ν', покрывающих полосу $\delta\nu$ (см. (16.34)), для того, чтобы определить момент измерения времени. Рассеянная энергия равна $nh\nu'$, а увеличение энтропии

$$\Delta S = \frac{nh\nu'}{T} \approx n\, A_m k = kn \ln(rm). \tag{16.39}$$

Полученная информация равна

$$\Delta I = k \ln\left(\frac{\Delta t}{\tau}\frac{\Delta v}{u}\right) = k \ln\frac{nm}{2}, \tag{16.40}$$

и обобщенное второе начало удовлетворяется:

$$\Delta(S - I) = k\,[(n-1)\ln m + n \ln r - \ln(\ln 2)] > 0. \tag{16.41}$$

Численный множитель n может быть мал, но всегда больше, чем 2. Мы предположили высокую надежность и высокую точность ($r \gg 1$, $m \gg 1$), и поэтому ΔS много больше, чем ΔI.

Так как мы применяем n квантов, то получаем отдачу в n раз больше, чем в обычном случае, и соотношение неопределенности выражается формулой (16.32).

Возьмем, к примеру,

$$r = e^{10}, \quad m = e^{10}, \; A_m = 20,$$

где значение A_m вычисляется по формуле (16.37). На основании (16.38) эти значения дают:

$$h\nu' \approx 20kT.$$

Если эксперимент производится при комнатной температуре, то это условие дает частоты вблизи видимого диапазона. Теория информации и квантовые условия не вступают в противоречие; напротив, они дополняют друг друга.

В приложениях к теории поля мы будем пользоваться равенством (16.30), представляющим нижний предел (16.32).

8. Неопределенность в измерении поля

Соотношение неопределенности в теории поля обсуждалось Бором и Розенфельдом[1]) применительно к электромагнитным полям. Применяется зонд, имеющий объем V и заряд ε, в течение времени τ. Импульс p_x измеряется в момент t и принимает другое значение p_x' в момент $t+\tau$. Среднее по V поле $\bar{E}_x$ за время τ равно

$$\bar{E}_x = \frac{p'_x - p_x}{\varepsilon\tau}. \tag{16.42}$$

Положение x зонда измеряется с погрешностью Δx, что дает погрешность Δp_x в определении p_x' и p_x. Мы утверждаем, что в этом случае можно пользоваться соотношением (16.30). Мы измеряем сначала x, затем p_x в момент t (см. раздел 6). Затем мы измеряем p'_x в момент $t+\tau$ и, наконец, x'. Последнее измерение сообщает частице большой неизвестный импульс, но это уже не имеет значения.

---

[1]) N. Bohr, L. Rosenfeld, Kgl. Danske Videnskal. Selskab. Medd. **12**, paper 8 (1953): В. Гайтлер, Квантовая теория излучения, гл. 2, ИЛ, 1956.

Погрешность в определении p_x дает погрешность в измерении поля:

$$\Delta\overline{E}_x \approx \frac{h}{\varepsilon\tau\Delta x} \approx \frac{h}{e\tau Z\Delta x}, \tag{16.43}$$

где $\varepsilon = Ze$, e — заряд электрона. Наблюдения под микроскопом позволяют нам измерить поле по его действию на свободную частицу при отсутствии физических препятствий, возмущающих поле. В этой формуле мы можем взять за Δx диаметр частицы, а $Z\Delta x$ есть характеристика частицы, используемой в качестве зонда. Она может изменяться от 10^{-13} для электрона, до 10^{-10} для тяжелого ядра ($Z = 96$, $\Delta x \approx \approx 10^{-12}$). Для того чтобы определить промежуток времени τ, нам нужен луч света с частотами, простирающимися от 0 до ν_M:

$$\nu_M = \frac{1}{2\tau}, \tag{16.44}$$

откуда

$$\varepsilon\,\Delta\overline{E}_x\,\Delta x \approx 2h\nu_M. \tag{16.45}$$

Эту формулу легко истолковать, так как $\varepsilon\overline{E}_x\Delta x$ есть работа, совершенная полем $\overline{E}_x$ на расстоянии Δx. Эта работа не может быть измерена с большой точностью, так как имеется излучение и поглощение квантов $h\nu_M$ светового луча, применяемого для определения промежутка времени τ.

В общем наше рассуждение не изменяет основной формулы (16.43) Бора и Розенфельда, но подчеркивает некоторые трудности, связанные с соотношением неопределенности. Обычное соотношение неопределенности соответствует абсолютному теоретическому пределу, но этот предел не может быть достигнут во многих экспериментальных устройствах, в которых, вообще говоря, получается значительно большая неопределенность. Это справедливо, в особенности, когда измеряется положение почти свободной частицы.

ГЛАВА 17

НЕГЭНТРОПИЙНЫЙ ПРИНЦИП ИНФОРМАЦИИ В ПРИМЕНЕНИИ К СВЯЗИ

1. Анализ сигналов с конечной шириной полосы

В главе 8 мы обсуждали общую проблему анализа сигналов, и мы смогли показать (см. раздел 6), что система сигналов, представляющих сообщение общей длительности τ и с конечной шириной полосы

$$0 \leqslant |\nu| \leqslant \nu_M, \tag{17.1}$$

имеет лишь конечное число степеней свободы

$$N = 2\nu_M\tau + 1. \tag{17.2}$$

Если полная длительность τ велика, то можно заменить (17.2) приближенным выражением

$$N \approx 2\nu_M\tau. \tag{17.3}$$

Этот очень важный результат был получен двумя методами: при помощи разложения Фурье периодически повторяемого сообщения (см. раздел 6 главы 8) и методом отсчетов для одиночного сообщения (см. раздел 7). В методе Фурье степени свободы соответствуют амплитудам первых N гармоник ряда Фурье. В методе отсчетов функция $f(t)$, представляющая сообщение, измеряется в N равноотстоящих моментах с постоянным интервалом между ними:

$$\theta = \frac{\tau}{N} \approx \frac{1}{2\nu_M}, \tag{17.4}$$

и было показано, как восстановить исходную функцию по этим отсчетам (см. (8.70)):

$$f_m = f(m\theta),$$
$$f(t) = \sum_m f_m \frac{\sin \pi (2\nu_M t - m)}{\pi (2\nu_M t - m)}. \qquad (17.5)$$

Мы обсуждали также (см. раздел 8) другое представление, основанное на *информационных ячейках* Гэйбора:

$$\Delta\nu \, \Delta t = 1, \qquad (17.6)$$

каждая из которых содержит две составляющие — симметричную и антисимметричную. Всего имеется

$$n = \nu_M \tau \qquad (17.7)$$

таких ячеек, что дает:

$$N = 2n = 2\nu_M \tau$$

различных составляющих, которые нужно задать, чтобы определить сообщение.

2. Сигналы и тепловой шум: представление в многомерном пространстве

Анализ главы 8 приложим к любым сообщениям и сигналам. Каким образом применить его к тепловому шуму с конечной шириной полосы ν_M, наблюдаемому на протяжении времени τ? В каждой ячейке мы можем определить среднюю интенсивность, но фаза случайна.

Таким образом, мы имеем n составляющих, каждая из которых представляет собой гармонический осциллятор со средней энергией kT. Мы предполагаем низкие частоты и игнорируем квантовые условия. Мы придем к такому же заключению, если рассмотрим разложение Фурье для всей системы с его N составляющими, и будем считать среднюю кинетическую энергию равной $kT/2$ для каждого синусного члена и для каждого косинусного члена. Можно также начать с шенноновых N отсчетов и присвоить каждому из них среднюю кинетическую энергию $kT/2$. В общем, энергия шума в среднем равна

$$\overline{E}_n = \frac{NkT}{2} = \nu_M \tau kT, \qquad (17.8)$$

а мощность шума

$$P_n = \frac{\overline{E}_n}{\tau} = \nu_M kT. \qquad (17.8a)$$

Это соответствует сигналам, распространяющимся вдоль линии в одном направлении: например, сигнал генерируется на входе слева и распространяется направо. Результат совпадает с равенством (11.7), полученным при обсуждении той же самой задачи по Найквисту.

Тепловой шум определяется своими средней мощностью и флуктуациями (см. главы 9, 10 и 11). Если бы мощность шума была в точности постоянна, она легко могла бы быть устранена и скомпенсирована, но возмущения возникают в действительности из-за непредсказуемых флуктуаций, которые всегда имеют тот же порядок величины, что и средняя мощность.

Рассмотрим сигнал (или сообщение), заданный в виде некоторой произвольной функции с максимальной частотой ν_M и длительностью τ. Функция имеет N степеней свободы $X_1, X_2, \ldots, X_m, \ldots, X_N$, как определено в предыдущем разделе. Она может быть представлена точкой с координатами $X_1, X_2, \ldots, X_m, \ldots, X_N$ в N-мерном пространстве S_N[1]). Каждая совокупность ячеек будет соответствовать некоторой координатной системе в S_N, и изменение координат в многомерном пространстве означает изменение системы ячеек. Для представления данной функции мы пользуемся ортогональными функциями, и полная энергия будет пропорциональна сумме квадратов координат. При надлежащем выборе единиц можем записать:

$$E = \sum_{m=1}^{m=N} X_m^2 = R_N^2 \qquad (17.9)$$

без отдельного численного множителя. Если энергия задана, точка должна лежать на гиперсфере радиуса R_N. Объем такой сферы, как известно, при больших N равен

$$V_N = C_N R_N^N, \quad \text{где } C_N \approx \left(\frac{2\pi e}{N}\right)^{\frac{N}{2}}. \qquad (17.10)$$

[1]) Этот метод подробно обсужден Шенноном, см. PIRE 37, 10 (1949).

Постоянная C_N зависит от числа измерений N. Интересным свойством пространства S_N с большим числом измерений N является то, что объем V_N в основном сосредоточен вблизи поверхности. Это можно доказать, показав, что малое приращение радиуса ΔR_N дает очень большое приращение объема:

$$\frac{\Delta V_N}{V_N} = N \frac{\Delta R_N}{R_N}. \tag{17.10a}$$

Когда N превосходит $R_N/(\Delta R_N)$, это выражение может стать больше единицы! Итак, мы будем иметь в виду, что практически весь объем гиперсферы сосредоточен в тонком слое около поверхности.

3. Пропускная способность канала с шумом

Данный невозмущенный сигнал представляется некоторой точкой M в многомерном пространстве S_N. При наличии шума точка будет смещена на некоторое расстояние в пределах сферы с центром в M. Радиус этой сферы равен $\overline{E}_n^{1/2}$, где $\overline{E}_n$ — средняя энергия шума. Вследствие неопределенности шума мы не знаем, где окажется новая точка M. Поэтому мы не должны применять другой сигнал, представляемый точкой, расположенной слишком близко к точке M. Задача состоит в том, чтобы подсчитать точно, сколько точек может быть надежно отличено друг от друга, или, другими словами, сколько различных сообщений может быть распознано при наличии шума.

Начнем с простого рассуждения, которое позднее мы рассмотрим более тщательно. Полная энергия есть сумма энергии E сигнала и средней энергии $\overline{E}_n$ шума. Согласно (17.9) эта сумма определяет гиперсферу радиуса

$$R' = (E + \overline{E}_n)^{\frac{1}{2}} \tag{17.11}$$

для сигнала вместе с шумом. Сигнал является центром малой сферы радиуса

$$r = \overline{E}_n^{\frac{1}{2}}, \tag{17.12}$$

и мы должны принять, что шум, добавленный к данному сигналу, затемняет сферу радиуса r вокруг точки M, изображающей собственно сигнал. Ставится вопрос: сколько таких малых сфер радиуса r может содержаться в большой сфере радиуса R'? Разумеется, верхним пределом этого числа является отношение объемов

$$G = \frac{R'^N}{r^N} = \left(\frac{E + \bar{E}_n}{\bar{E}_n}\right)^{\frac{N}{2}} = \left(1 + \frac{P}{P_n}\right)^{\frac{N}{2}}, \tag{17.13}$$

где P и P_n — средние мощности сигнала и шума соответственно [1]).

Мы определяем в (4.3) пропускную способность как предел

$$C = K \lim_{\tau \to \infty} \frac{\ln G(\tau)}{\tau}, \tag{17.14}$$

где $G(\tau)$ есть полное число различимых сообщений длительностью τ, которые можно передать по каналу.

Сравнивая (17.13) и (17.4), получаем:

$$C = K \lim_{\tau \to \infty} \frac{N}{2\tau} \cdot \ln\left(1 + \frac{P}{P_n}\right).$$

Но N дано равенством (17.3), откуда

$$C = K\nu_M \ln\left(1 + \frac{P}{P_n}\right), \tag{17.15}$$

---

[1]) На недостаточность этого рассуждения, целиком заимствованного у Шеннона, уже указывалось в нашей литературе. Рассуждение содержит два дефекта. Во-первых, беря отношение объемов, мы принимаем коэффициент заполнения пространства сферами равным единице. Между тем известно, что при наиплотнейшей укладке сфер коэффициент заполнения ими пространства убывает с увеличением числа измерений и стремится к нулю при $N \to \infty$. Во-вторых, для возможности надежного различения сигналов вовсе не требуется, чтобы сферы малого радиуса r не пересекались; напротив, можно показать, что в пределе при $N \to \infty$ возможность получения сколь угодно малой вероятности ошибки сохраняется вплоть до полного покрытия пространства сферами (т. е. когда каждая сфера всеми своими частями лежит внутри других или, иными словами, когда пространство заполнено сферами без пустых промежутков). Оба эти соображения, конечно, существенно меняют геометрическую формулировку задачи. Однако связанные с ними поправки действуют в противоположные стороны и взаимно компенсируются. Решив корректно поставленную геометрическую задачу, мы придем к формуле Шеннона, правильность которой сомнению не подвергается. (*Прим. перев.*)

где K — постоянная, зависящая от выбора единиц, ν_M — верхняя граничная частота полосы, P — средняя мощность сигнала, P_n — средняя мощность шума. Равенство (17.15) известно как формула Хартли — Таллера — Шеннона[1]).

4. Обсуждение формулы Таллера — Шеннона

Столь значительный результат требует более подробного обсуждения для того, чтобы обосновать сделанные приближения и доказать, что предел (17.13) действительно может быть достигнут. Прежде всего мы должны быть уверены в том, что учли влияние флуктуаций и непредсказуемых изменений шума. Все это содержится в общей статистической формуле (9.32):

$$p(E,\ T) = Be^{-\beta E}, \quad \beta = \frac{1}{kT}, \tag{17.16}$$

где $p(T, E)$ — вероятность микросостояния («комплексии» или квантового состояния) с энергией E при температуре T, а B — коэффициент, определяемый тем условием, что сумма вероятностей должна равняться единице. Это соотношение может непосредственно применяться к шуму. Пусть x_1, x_2, ..., x_m, ..., x_n — переменные, представляющие только шум. Тогда (17.9) дает для полной энергии шума

$$E_n = \sum_m x_m^2 = r^2. \tag{17.17}$$

С другой стороны, между r и $r + dr$ содержится объем dV (см. (17.10)), и число квантованных микросостояний пропорционально этому объему. Таким образом, вероятность того, что точка лежит между r и $r + dr$, равна

$$p(r,\ T)\,dr = B'e^{-\beta r^2} r^{N-1}\,dr, \tag{17.18}$$

так как

$$dV = NC_N r^{N-1}\,dr.$$

B' — новая нормирующая постоянная. Новая вероятность $p(r,\ T)$ имеет очень острый максимум при некотором зна-

[1]) Превосходное изложение вопроса дано в статье M. Leifer, W. F. Schreiber, Advances in Electronics 3, 306 (1951).

чении r, когда

$$\left.\begin{aligned} \frac{\partial}{\partial r}\left(r^{N-1}e^{-\beta r^2}\right) = (N-1-2\beta r^2)\,r^{N-2}e^{-\beta r^2} = 0, \\ r_{mp}^2 = \frac{N-1}{2\beta} = \frac{(N-1)\,kT}{2}. \end{aligned}\right\} \quad (17.19)$$

Легко показать, что с увеличением N этот максимум становится все более острым. Равенство (17.19) дает наиболее вероятное значение r^2, и, следовательно, наиболее вероятное значение энергии шума равно

$$E_{mp} = r_{mp}^2 = \frac{(N-1)\,kT}{2} \approx \frac{NkT}{2} = \bar{E}_n, \quad N \gg 1. \quad (17.20)$$

Это — типичный пример совпадения наивероятнейшего значения E_{mp} и среднего значения $\bar{E}_n$, когда число переменных N становится чрезвычайно большим. Мы уже имели пример этого явления в главе 4 (см. разделы 7 и 8).

В нашей задаче сообщение представлено точкой с координатами X_m в пространстве N измерений, а сигнал с шумом дает точку с координатами

$$X'_m = X_m + x_m,$$

где x_m — случайно. Предыдущее рассуждение показывает, что эта новая точка с координатами X'_m практически должна лежать на поверхности сферы радиуса r_{mp} с центром в исходной точке M. Это в точности те условия, которые были приняты при выводе формулы Таллера — Шеннона (см. (17.12) и (17.13)).

Эти условия применимы только при низких частотах, когда $h\nu$ много меньше kT и квантовые состояния могут считаться бесконечно близкими. При этих обстоятельствах шум может быть определен как гауссов шум. Распределение по какой-либо переменной x_m не зависит от других переменных и следует закону $e^{-\beta x_m^2}$ согласно (17.16) и (17.17). Гауссов шум неприменим в квантованной системе, если только кванты $h\nu$ не могут считаться бесконечно малыми. Рис. 17.1 дает геометрическое представление ситуации с учетом соотношения (17.10), согласно которому точки сферы должны практически лежать на ее поверхности. Точка M, представляющая исходное сообщение с энергией E, находится на сфере радиуса $E^{1/2}$. Сообщение плюс шум представлено точкой M' на сфере радиуса $(E+\bar{E}_n)^{1/2}$; расстояние MM' равно $r=\bar{E}_n^{1/2}$ (см. (17.12) и (17.19)). M_1 представляет возможное положение точки другого сообщения, которая могла бы быть спутана с M, поскольку эта точка находится на таком же расстоянии r от M'. Таким образом, MM_1 есть наименьшее возможное расстояние между различимыми точками сообщений. Рис. 17.1 дает плоское сечение многомерного пространства, определяемое тремя точками M, M' и M_1. Расстояние $MM_1=2H$ принимает предельное значение, когда MM' перпендикулярно к OM, а $M'M_1$ — перпендикулярно к OM_1. Это значит, что площадь A треугольника OMM' равна

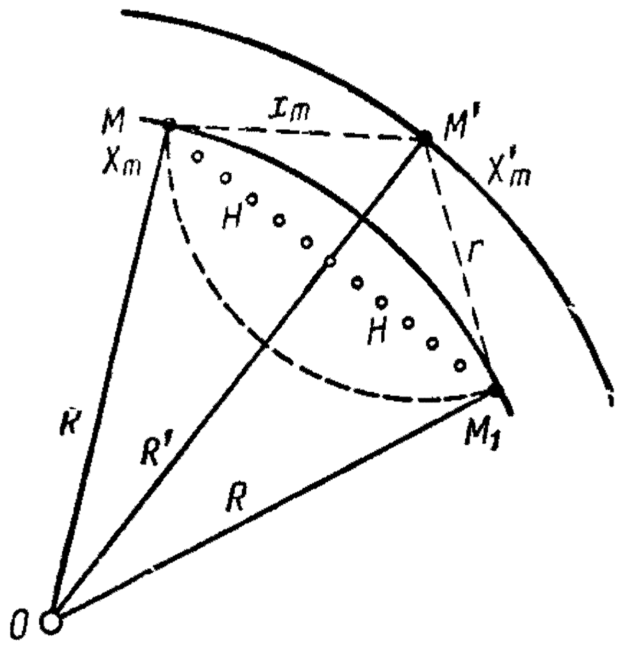

Рис. 17.1. Сечение N-мерной гиперсферы плоскостью, показывающее две «ближайшие» точки сообщений и влияние шума.

$$A_{OMM'}=\frac{H(\overline{OM'})}{2}=\frac{(\overline{MM'})(\overline{OM})}{2}, \qquad (17.21)$$

откуда

$$HR'=Rr \text{ и } H=\sqrt{\frac{E\bar{E}_n}{E+\bar{E}_n}}<r=\sqrt{\bar{E}_n}, \qquad (17.22)$$

Расстояние $2H$ между точками M и M_1 действительно меньше, чем расстояние $2r$, где r — радиус шумовой сферы, окружающей M.

Мы должны теперь выяснить, как расположить точки сообщений M в сфере R объема

$$V=C_NR^N=C_NE^{\frac{N}{2}}, \qquad (17.23)$$

с тем, чтобы обеспечить передачу со сколь угодно малой вероятностью ошибки. Пусть

$$\eta = e^{-\alpha} \leqslant 1 \tag{17.24}$$

есть допустимая вероятность ошибки. Распределим точки сообщений по объему V с постоянной средней плотностью. Один возможный способ получения постоянной плотности состоит в случайном распределении по объему, как предложено Шенноном. Пусть G_1 есть полное число точек сообщений. Каждая точка сообщения, попадающая внутрь сферы радиуса H, окружающей точку M, может быть спутана с M, что и дает ошибку. Объем этой сферы равен

$$u = C_N H^N. \tag{17.25}$$

Точки сообщений, попадающие в остающийся объем $V - u$, не могут дать ошибки. Вероятность ошибки, очевидно, пропорциональна G_1 и u/V, и мы требуем:

$$\mathscr{E} = \frac{G_1 u}{V} \leqslant \eta. \tag{17.26}$$

Это означает, что мы можем взять число точек

$$G_1 \leqslant \frac{\eta V}{u} = \eta \left(\frac{R}{\bar{H}}\right)^N = \eta \left(\frac{E + \bar{E}_n}{\bar{E}_n}\right)^{\frac{N}{2}} \tag{17.27}$$

согласно (17.12), (17.13) и (17.15). Соотношение (17.27) можно переписать в виде

$$G_1 \leqslant \eta G, \tag{17.28}$$

где G — величина, определенная равенством (17.13). Наибольшее число G_1 соответствует знаку равенства в (17.28) и дает пропускную способность канала:

$$C_1 = K \lim_{\tau \to \infty} \frac{\ln G_1}{\tau} = K \lim_{\tau \to \infty} \left(\frac{\ln G}{\tau} + \frac{\ln \eta}{\tau}\right),$$

или согласно (17.24)

$$C_1 = C - \lim_{\tau \to \infty} \frac{\alpha}{\tau}. \tag{17.29}$$

Пропускная способность канала может быть сколь угодно приближена к теоретическому пределу C, определенному ранее (см. (17.15)). Для этого нужно только распределить точки сообщений по сфере радиуса R с постоянной средней плотностью.

5. Практический пример

Предыдущее рассуждение основывалось на рассмотрении целых длинных сообщений общей длительностью τ. Посмотрим теперь, как такое сообщение может быть построено из некоторых стандартных сигналов. Начнем с системы ячеек раздела 1. Годится любая система ячеек, например та, которая соответствует ряду Фурье, или та, которая соответствует методу отсчетов Шеннона. Мы имеем (см. (17.7) и (17.8)):

$$n = \nu_M \tau \tag{17.30}$$

ячеек, каждая из которых содержит энергию шума

$$\overline{e}_n = kT. \tag{17.31}$$

Полная энергия шума равна в среднем

$$\overline{E}_n = n\overline{e}_n = nkT.$$

Выберем уровень энергии E для данного сигнала в некоторой ячейке. На сигнал накладывается шум, что добавляет к энергии сигнала E среднюю энергию $\overline{e}_n$ с флуктуациями того же порядка величины, так что практически покрывается интервал энергий $2\overline{e}_n$. В соответствии с этим довольно грубым рассуждением мы должны попробовать выбрать в каждой ячейке систему энергетических уровней, отстоящих друг от друга на $2\overline{e}_n$, и использовать эти уровни в качестве элементарных сигналов. Сами сигналы обладают энергиями вплоть до некоторого наибольшего значения

$$e_s = 0, \quad 2\overline{e}_n, \ 4\overline{e}_n, \ \ldots, \ 2(m-1)\overline{e}_n, \tag{17.32}$$

а полные энергии сигнала вместе с шумом составляют:

$$e_t = e_s + \overline{e}_n = \overline{e}_n, \ 3\overline{e}_n, \ 5\overline{e}_n, \ \ldots, \ (2m-1)\overline{e}_n.$$

Мы имеем, таким образом, m различимых сигналов на каждую ячейку. Целое сообщение длительностью τ получается путем задания энергий сигнала в n ячейках.

Полная энергия сообщения в среднем равна

$$\left.\begin{aligned} \overline{E}_s = n\overline{e}_s = n(m-1)\overline{e}_n, \\ \overline{E}_s + \overline{E}_n = nm\overline{e}_n. \end{aligned}\right\} \tag{17.33}$$

Эти определения согласуются с предыдущими. Если длительность сообщения велика, то E_s мало отличается от своего среднего значения и все сообщения будут иметь практически энергию, выраженную соотношением (17.33). Это снова случай совпадения наивероятнейшего и среднего значений.

Число получаемых таким путем различимых сообщений равно

$$G' = m^n,$$

и пропускная способность канала

$$C' = K \lim_{\tau\to\infty} \frac{\ln G'}{\tau} = K\nu_M \ln m. \tag{17.34}$$

Но согласно (17.33)

$$m = 1 + \frac{\overline{e}_s}{\overline{e}_n} = 1 + \frac{\overline{E}_s}{\overline{E}_n} = 1 + \frac{\overline{P}_s}{P_n}, \tag{17.35}$$

где $\overline{P}_s = \frac{P_M}{2}$, P_M — наибольшая мощность сигнала.

Итак, пропускная способность C' равна предельной пропускной способности, выражаемой формулой Таллера — Шеннона (17.15). Эта модель будет использована для сравнения в последующем рассуждении.

Это рассуждение оправдывает выбор интервала $2\overline{e}_n$, введенного в (17.32), поскольку он приводит к формуле Шеннона. Более тесное расположение дало бы, очевидно, перекрытие энергетических интервалов, соответствующих двум соседним уровням сигнала. Более просторное расположение могло бы дать бóльшую надежность передачи, но не дало бы полного использования пропускной способности канала.

Несколько дополнительных замечаний полезны для того, чтобы показать связь с анализом предыдущего раздела. Покажем сначала, что равноотстоящие уровни энергии (см. (17.32)) соответствуют равномерной плотности точек сигналов в пространстве $N=2n$ измерений. Рассмотрим некоторую, скажем i-ю, ячейку, имеющую две степени свободы, т. е. одну синусную составляющую и одну косинусную, которые мы обозначим X_i и Y_i. Эти $2n$ координат X_i, Y_i образуют N-мерное пространство. Предположим, что точки сигналов имеют в плоскости X_i, Y_i равномерную плотность $\rho(X_i, Y_i)$. В среднем в малой области $dX_i\,dY_i$ число точек будет:

$$dG=\rho(X_i, Y_i)\,dX_i\,dY_i.$$

Возьмем теперь кольцевую область между r_i и r_i+dr_i, определив r_i условием

$$r_i^2=X_i^2+Y_i^2=E_i, \tag{17.36}$$

где E_i — соответствующая энергия. Среднее число точек сигналов в этой области равно

$$dG=\rho(X_i, Y_i)\,2\pi r_i\,dr_i=\pi\rho(X_i, Y_i)\,dE_i,$$

и мы получаем по шкале энергий равномерную плотность для i-й ячейки

$$\rho'(E_i)=\pi\rho(X_i, Y_i). \tag{17.37}$$

Это соответствует предположению (17.32).

В (17.32) мы применили равноотстоящие и равновероятные энергетические уровни вплоть до некоторого наибольшего. Мы можем ввести иное допущение и рассматривать энергетические уровни с априорной вероятностью

$$\pi_i=Be^{-\gamma E_i} \tag{17.38}$$

для энергии E_i в i-й ячейке. Это дает вероятность

$$\pi_i=Be^{-\gamma r_i^2}=Be^{-\gamma(X_i^2+Y_i^2)} \tag{17.39}$$

в N-мерном пространстве. Задача очень сходна с ранее рассмотренной (см. (17.16) и (17.20)). Новая постоянная γ просто заменяет прежнюю β, и мы рассматриваем энергию сигнала вместо энергии шума. Все точки сигналов практически лежат на сфере

$$R^2=\frac{N-1}{2\gamma}=\bar{E}, \tag{17.40}$$

и $\bar{E}$ есть средняя энергия сигналов.

6. Негэнтропийный принцип в применении к каналу с шумом

В предыдущих главах мы подчеркивали тот принцип, что получение любой информации о физической системе оплачивается соответствующим количеством отрицательной энтропии, отбираемой у системы, и обратно, что всякая информация, запасенная или переданная физической системой, соответствует увеличению негэнтропии системы.

Принцип может быть проверен на обсуждаемых примерах. Мы обсудим посвященную этому вопросу работу Рэймонда [1]) и покажем, как она должна быть исправлена в отношении ошибочного численного множителя 1/4.

Кабель, по которому мы передаем сообщение, имеет первоначально температуру T. При нормальных условиях он имеет энтропию S_0 и уровень шума, соответствующий средней мощности P_n в полосе шириной ν_M и для определенного направления распространения (см. (11.17) и (17.8а)):

$$P_n = kT\nu_M. \tag{17.41}$$

Здесь Рэймонд вводит неправильный множитель 4 в свое равенство (4), что приводит к множителю 1/4 в его равенстве (5).

При передаче сообщения мы используем на протяжении некоторого времени τ среднюю мощность P, но не тепло. Энтропия кабеля вначале не меняется, но если мы не приняли сообщения и оставили его в кабеле, то оно будет распространяться туда и обратно и постепенно затухать вследствие эффекта Джоуля в сопротивлениях, пока вся энергия $P\tau$ не превратится в тепло. Это конечное состояние соответствует энтропии S большей, чем исходная энтропия S_0. Негэнтропия ΔN, введенная вначале в кабель в процессе передачи сигнала, представляется разностью

$$\Delta N = S - S_0. \tag{17.42}$$

Эту величину мы и должны вычислить.

---

[1]) R. C. Raymond, Am. Scientist **38**, 275 (1950); см. также C. A. Desoer, Mass. Inst. Technol. Research Lab. Electronics, Quart. Progr. Rept. **46** (1951); J. H. Felker, PIRE **40**, 728 (1952); D. A. Bell, Am. Scientist **40**, 682 (1952); L. Brillouin, J. Appl. Phys. **25**, 595 (1954).

Предположим сначала очень сильную связь между сигналами и физической системой (кабелем). Это есть предположение Рэймонда, «что теплоемкость замкнутой системы настолько велика, что изменением температуры при действии устройства связи можно пренебречь». Это ведет к окончательной энтропии

$$S = S_0 + \frac{P\tau}{T}, \tag{17.43}$$

откуда с помощью (17.41)

$$\Delta N = \frac{P\tau}{T} = \frac{k\nu_M P\tau}{P_n}. \tag{17.44}$$

Негэнтропия ΔN представляет информацию, текущую со скоростью $\Delta N/\tau$ в секунду, и, следовательно, пропускная способность (в секунду) равна

$$C_1 = \frac{\Delta N}{\tau} = \frac{k\nu_M P}{P_n}. \tag{17.45}$$

Сравнивая этот результат с формулой Шеннона (17.15):

$$C_{Sh} = k\nu_M \ln\left(1 + \frac{P}{P_n}\right), \tag{17.46}$$

мы видим, что обе формулы согласуются только при очень малых отношениях P/P_n, когда можно ограничиться первым членом в разложении логарифма.

Формула Шеннона была получена в специальных предположениях, не соответствующих только что рассмотренной модели. Шеннон имел в виду кабель без потерь, без омического сопротивления. Это означает отсутствие связи между электрическими токами, текущими по кабелю, и самим физическим кабелем. Предполагается, что источник шума находится вне кабеля и действует независимо. Здесь нет прямого механизма, приводящего в конце концов к увеличению энтропии, которое мы обнаружили в нашем предыдущем примере. Если мы примем представление об отсутствии связи между сигналами и кабелем, то мы должны предположить какой-либо иной механизм рассеяния. Увеличение энтропии может происходить вследствие искажения сигнала или из-за нерегулярного изменения интервалов между импульсами тока и интенсивности импульсов.

Начальное состояние есть организованная система импульсов, несущих желаемую информацию, а конечное состояние есть дезорганизованная беспорядочная последовательность импульсов, расположенных случайно и не несущих разборчивой информации. Система имеет n колебательных степеней свободы (ячейка раздела 1), и случайное распределение энергии E_r соответствует температуре T (см. (17.8))

$$E_r = nkT. \tag{17.47}$$

Если теперь предположить, что энергия E_r постепенно нарастает ступенями dE_r, то увеличение энтропии есть

$$S - S_0 = \int_{E_1}^{E_2} \frac{dE_r}{T} = nk \int_{E_1}^{E_2} \frac{dE_r}{E_r} = nk \ln \frac{E_2}{E_1}. \tag{17.48}$$

Начальное состояние E_1 соответствует только мощности P_n шума длительностью τ. Конечный уровень энергии E_2 получается после того, как вся добавочная энергия $P\tau$ распределена случайным образом между n осцилляторами и добавлена к энергии шума

$$E_1 = P_n \tau, \quad E_2 = (P_n + P)\tau. \tag{17.49}$$

Отсюда имеем для негэнтропии

$$\Delta N = S - S_0 = nk \ln \frac{P_n + P}{P_n}.$$

Это дает пропускную способность канала (на единицу времени)

$$C_2 = \frac{\Delta N}{\tau} = k\frac{n}{\tau} \ln\left(1 + \frac{P}{P_n}\right) = k\nu_M \ln\left(1 + \frac{P}{P_n}\right), \tag{17.50}$$

что совпадает с формулой Шеннона (17.46). Различие между нашим предыдущим результатом (17.45) и новым (17.50) обусловлено тем, что первоначальная модель поддерживалась при постоянной температуре T, тогда как в новой модели наблюдается повышение температуры, пропорциональное увеличению случайной энергии E_r, причем равенство (17.47) все время соблюдается. Это повышение температуры уменьшает разность энтропий (17.48) и приводит к меньшей пропускной способности канала.

Логарифм появляется здесь таким же образом, как известный член $\ln V$ в энтропии идеального газа (см. (12.12)). Ситуация в этой задаче очень сходна с той, которую мы ранее обнаружили в разделе 7 главы 13 (см. (13.55)), когда нашли, что информация I растет как логарифм некоторой величины F, тогда как изменение энтропии ΔS есть линейная функция F. В проблеме связи мы имеем сходные формулы: информация I (или пропускная способность канала C, формулы (17.46) и (17.50)) есть логарифм величины

$$F = 1 + \frac{P}{P_n},$$

тогда как изменение энтропии (см. (17.44)) пропорционально $F - 1$ в модели Рэймонда, представляющей реальную физическую систему с рассеянием.

В заключение нужно подчеркнуть физическое значение этого рассуждения: формула Шеннона действительно дает предел, соответствующий оптимальным условиям с наибольшей эффективностью преобразования негэнтропии в информацию. Это следует из того факта, что вычисления, основанные на информации (раздел 3) и на негэнтропии (см. (17.50)), дают совпадающие результаты при условии, что все величины измеряются в одной системе единиц.

Если информация измеряется в двоичных единицах, а энтропия — в термодинамических единицах, то каждая двоичная единица соответствует $k \ln 2 \approx 10^{-16}$ термодинамических единиц. Это снова доказывает (см. главу 14), что одна двоичная единица информации никогда не может быть получена за негэнтропийную цену меньшую, чем $k \ln 2$.

7. Видоизмененная формула Гэйбора и роль биений

Гэйбор[1]) после обсуждения системы ячеек, описанной в главе 8, обстоятельно исследовал вопрос о том, какие уровни энергии могут быть использованы в каждой ячейке для представления различимых сигналов. Рассуждение Гэйбора основано на квантовой теории, но во всех практических приложениях кванты $h\nu$ настолько малы, что может применяться классическая теория. Сделав это, можно убедиться, что новые

[1]) D. Gabor, Phil. Mag. [7] 41, 1161 (1950).

члены в формуле Гэйбора соответствуют биениям между сигналом и шумом. Эти биения увеличивают флуктуации энергии. Если наложить на сигнал с амплитудой A_1 шум с амплитудой A_2 и переменной фазой ψ, то получим:

$$I = A_1 \cos \omega t + A_2 \cos(\omega t + \psi), \tag{17.51}$$

$$I^2 = A_1^2 \cos^2 \omega t + A_2^2 \cos^2(\omega t + \psi) + 2A_1A_2 \cos \omega t \cos(\omega t + \psi).$$

Полагая, что A_2 и ψ меняются медленно, усредним это соотношение по нескольким периодам:

$$\overline{I^2} = \frac{A_1^2}{2} + \frac{A_2^2}{2} + A_1A_2 \cos \psi. \tag{17.52}$$

За такой сравнительно малый промежуток времени мощность равна

$$\overline{P} = P_1 + \overline{P}_2 + 2(P_1\overline{P}_2)^{\frac{1}{2}} \cos \psi. \tag{17.53}$$

Среднее по большому промежутку времени будет содержать только первые два члена, так как $\cos\psi$ в среднем равен нулю:

$$\overline{\overline{P}} = P_1 + \overline{P}_2. \tag{17.54}$$

Последний член в (17.53) представляет биения между сигналом и шумом. Они оказывают существенное влияние на флуктуации мощности Δp:

$$\overline{P} = \overline{\overline{P}} + \Delta p, \tag{17.55}$$

$$\overline{\Delta p^2} = \overline{P_2^2} + 4 P_1\overline{P}_2 \overline{\cos^2\psi} = (\overline{P}_2)^2 + 2 P_1 \overline{P}_2,$$

так как для теплового шума

$$(\overline{P}_2)^2 = \overline{P_2^2} \text{ и } \overline{\cos^2\psi} = \frac{1}{2}.$$

Из (17.55) видно, что флуктуации возрастают с увеличением мощности сигнала P_1.

Эти результаты можно использовать для вычисления числа различимых уровней энергии. Для согласования с формулой Шеннона (см. (17.32)) выберем ступени $2(\overline{\Delta p^2})^{\frac{1}{2}}$. Это вдвое больше ступеней, первоначально предложенных Гэйбором.

Пусть P_M есть наибольшая мощность, применяемая при передаче сигналов. Число ступеней (включая $P=0$) есть

$$m_G = 1 + \int_0^{P_M} \frac{dP}{2\sqrt{\Delta p^2}} = 1 + \frac{1}{2}\int_0^{P_m} \frac{dP}{\sqrt{P_2^2 + 2PP_2}} =$$
$$= \frac{1}{2}\sqrt{1 + 2\frac{P_M}{P_2}} + \frac{1}{2}. \quad (17.56)$$

Для малых отношений P_M/P_2 число m_G сводится к $\left(1 + \frac{1}{2}\frac{P_M}{P_2}\right)$, что совпадает с (17.35). Для больших отношений сигнал/шум значение m_G растет с увеличением P_M медленнее, чем в (17.35), что соответствует формуле Шеннона.

В соответствии с экспериментальными условиями наблюдения нужно выяснить, имеют ли значение дополнительные флуктуации, обусловленные биениями. Это обсуждение покажет, пользоваться ли формулой Шеннона или формулой Гэйбора. Формула Гэйбора (17.56) приводит к следующему выражению для пропускной способности канала:

$$C_G = K\nu_M \ln m_G = K\nu_M \ln \frac{1 + \sqrt{1 + \frac{2P_M}{P_2}}}{2}, \quad (17.57)$$

где P_M — наибольшая мощность сигналов. Это соотношение получено из (17.34) с гэйборовым значением m_G. Если время τ велико, то среднее $\overline{\cos\psi}$ обращается в нуль и биениями можно пренебречь, что и оправдывает применение формулы Шеннона в качестве предельной.

ГЛАВА 18

ПИСЬМО, ПЕЧАТЬ И ЧТЕНИЕ[1])

1. Передача информации; живая информация

В предыдущей главе мы обсуждали проблему связи, применяя обычный подход, разработанный Шенноном, Винером и группой из лабораторий компании Белл. Проблема сведена к простейшему случаю связи между лицами посредством телеграфа или того или иного телефонного устройства. Это вполне определенная и очень важная задача, решение которой может быть распространено на большое число сходных ситуаций, например, когда одно лицо непосредственно говорит с другим.

Теория, однако, может не ограничиваться этим специальным случаем и может быть распространена на значительно более общие проблемы при условии, что специальное внимание обращено на условия действия. Мы намерены в данной главе обсудить эти общие проблемы и показать, что негэнтропийный принцип информации позволяет нам исследовать все возможные обстоятельства связи и сделать некоторые важные заключения об основных законах.

Первая интересная проблема — это вещание. Здесь мы имеем одно лицо, сообщающееся со многими.

A. Диктор говорит перед микрофоном, соединенным N кабелями с N различными приемниками.

B. Одна телеграмма посылается N различным станциям.

C. Оратор выступает перед слушателями в аудитории.

D. Вещательная программа передается по радио.

---

[1]) Эти проблемы обсуждались автором на семинаре в Харварде (октябрь 1951), в лекции в университете Брауна (февраль 1952) и в докладе Американскому физическому обществу (май 1952). См. также L. Brillouin, J. Appl. Phys. 25, 595 (1954).

В случаях A и B мы, очевидно, имеем примеры прежней связи между двумя лицами при N параллельно работающих каналах. Передатчик должен развивать мощность, необходимую для всех каналов. Если все каналы одинаковы, то мы должны просто обеспечить мощность в N раз большую, чем для одного канала. Случаи C и D менее благоприятны из-за того, что большая доля мощности не используется: звуковые волны поглощаются в стенах, мебели и других звукопоглощающих материалах в случае C; радиоволны рассеиваются и теряются в пространстве в случае D. За исключением этих очень больших дополнительных потерь два последних случая аналогичны случаям A и B. В общем мы можем свести все это к одиночным каналам, работающим параллельно.

Эти проблемы имеют следующие общие черты:

1. Мощность и негэнтропия распространяются вместе с информацией.

2. Приемник поглощает энергию и негэнтропию тем же действием, которым осуществляется прием информации. Мы характеризуем эту ситуацию выражением *живая информация*. Информация передается с энергией, необходимой для ее обнаружения.

3. В процессе передачи некоторая энергия деградирует, некоторая негэнтропия теряется, и информация теряется посредством того же механизма.

4. Во всех примерах приложим обобщенный принцип Карно, устанавливающий, что информация должна быть оплачена количеством негэнтропии, большим или равным полученной информации.

2. Проблема чтения и письма

Мы встречаемся с совершенно иной ситуацией, когда обращаемся к другому виду передачи информации: к письму и чтению. Очевидно, что процесс письма сопровождается некоторой затратой энергии, но мы не находим никаких следов этой энергии на исписанной странице. Здесь не осталось также видимой негэнтропии; теперь информация совершенно разобщена с этими другими элементами. Это не отдельный пример, так как сходное положение мы имеем во всех случаях записи информации: печатные материалы, граммофонные пластинки, магнитная лента, перфолента и перфокарты, чертежи и т. д.

Во всех этих примерах мы имеем запасенную информацию, не связанную ни с энергией, ни с негэнтропией. Более того, мы можем извлекать информацию, не разрушая ее. Представляется возможным неограниченное размножение информации.

В качестве примера предположим, что автор пишет книгу, и книга публикуется. Мы не рассматриваем процесса мышления и работу автора. Это есть человеческий элемент, который мы тщательно избегали включать в наши рассуждения (см. главы 1 и 20). Не составляет разницы, будем ли мы рассматривать книгу, написанную автором, или просто собрание числовых данных вроде биржевых цен, метеорологических наблюдений и т. п. Если рукопись имеет N букв, то она содержит некоторое количество информации. Наше определение не идет далее. Предположим теперь, что напечатано 1000 экземпляров книги и что каждый экземпляр прочитали 100 человек. Общее количество информации возросло в 10^5 раз, тогда как соответствующее увеличение энтропии представляется несравнимым. Конечно, увеличение информации снижается такими причинами, как опечатки, ошибки передачи, недостаток понимания и т. д. Однако коэффициент умножения все же может быть практически очень велик. Для того чтобы показать, что и в этих случаях приложим негэнтропийный принцип, необходимо изучить вопрос более внимательно.

3. Мертвая информация и как ее оживить

Мы характеризуем приведенную выше ситуацию как запасенную *мертвую информацию*, не связанную с энергией или негэнтропией. Для того чтобы прочесть информацию, мы должны снова снабдить ее энергией. Дополнительный источник энергии совершенно необходим для чтения, и этот источник доставляет негэнтропию, которая превращается в новую информацию.

Для того чтобы читать книгу, нам нужен источник света. Граммофон не работает без мотора, вращающего пластинку. Слепой, читающий книгу, напечатанную рельефным шрифтом, должен двигать свой палец по бумаге и развивать достаточное давление, чтобы отличить печать от других неровностей страницы; это движение и давление требуют работы. Нам нужен свет и фотоэлемент, чтобы прочесть перфоленту, либо мы применяем батарею и щетки, пропускающие ток через отверстия перфокарты. Можно показать, что во всех этих примерах количество негэнтропии, взятой от внешнего источника, больше, чем полученная информация. Случай чтения перфоленты является типичным примером, для которого мы можем

Информация в дв. ед.	0	1	0	0	1	1	1	0	1	0	0	1
Лента с отверстиями	0		0	0				0		0	0	

Рис. 18.1. Диаграмма, показывающая ленту с отверстиями, пробитыми на месте нулей в сообщении, закодированном по двоичной системе.

подсчитать как энтропию, так и информацию и проверить соотношение между ними.

Положим, что информация записана на металлической перфоленте в виде двоичных цифр. Как показано на рис. 18.1, нули представлены отверстиями, а единицы — неперфорированными отрезками ленты. Каждый отрезок ленты содержит одну двоичную единицу информации, или

$$\Delta I = k \ln 2 \qquad (18.1)$$

на отрезок в термодинамических единицах, где k — постоянная Больцмана. Читающее устройство показано на рис. 18.2. Луч света от источника попадает на ленту. Если имеется отверстие, луч падает на фотоэлемент P_1. Если отверстия нет, луч отражается в фотоэлемент P_2. Мы могли бы устранить

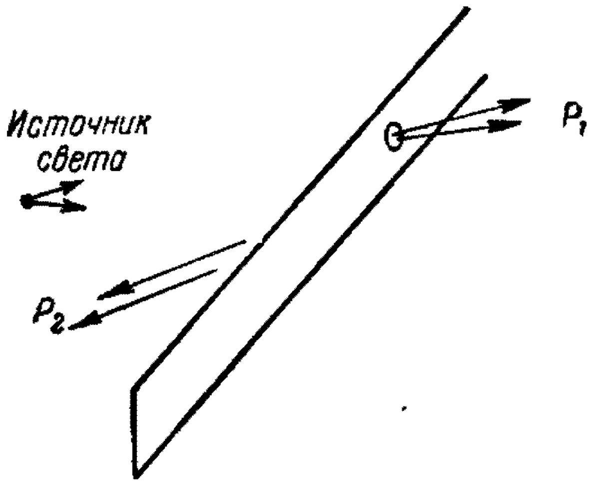

Рис. 18.2. Читающее устройство для ленты рис. 18.1.

Фотоэлемент справа P_1 получает свет через пробитые отверстия, тогда как фотоэлемент слева P_2 получает свет, отраженный от непробитых позиций.

второй фотоэлемент P_2 и считать, что мы имеем единицу всякий раз, когда P_1 бездействует. Нужно, однако, напомнить, что мы всегда требовали, чтобы отрицательному отсчету не придавалось значения. Итак, каждый отрезок ленты потребует по меньшей мере одного кванта $h\nu$, и эта энергия будет поглощена либо в P_1, либо в P_2. Для того чтобы быть различимым на фоне теплового излучения, квант $h\nu$ должен быть достаточно большим, так что

$$h\nu \geqslant kT \ln 2, \tag{18.2}$$

как было показано в разделе 3 главы 14 (см. (14.17)). При поглощении кванта происходит потеря негэнтропии (увеличение ΔS энтропии)

$$\Delta S = \frac{h\nu}{T} \geqslant k \ln 2, \tag{18.3}$$

и мы имеем соотношение

$$\Delta (S - I) \geqslant 0, \tag{18.4}$$

показывающее, что обобщенный принцип Карно (глава 12) удовлетворяется. В этом простом случае мы показали, что информация получается за счет негэнтропии светового луча, и мы можем теоретически достичь предела, соответствующего эффективности, равной единице. Вопрос ускользал от внимания вследствие очень малой величины требуемой негэнтропии

$$k \ln 2 \approx 10^{-16}\ \mathrm{CGS}. \tag{18.5}$$

Вышеописанный способ пробивки не очень надежен, так как единицы фактически не записываются. Гладкий кусок ленты без отверстий будет прочитан как последовательность единиц,

Информация	0	1	0	0	1	1	1	0	1	0	0	1
Лента	0		0	0				0		0	0	
		0			0	0	0		0			0

Рис. 18.3. Более совершенный метод пробивки ленты с применением отверстий на двух уровнях, один из которых предназначен для нулей, а другой — для единиц.

что не верно. Более надежен метод, применяющий два разных уровня для отверстий, как показано на рис. 18.3, и два фотоэлемента, принимающих свет, который может пройти через то или другое отверстие. Такое устройство по-прежнему требует поглощения одного кванта на двоичную единицу, так что соотношения (18.1) — (18.4) сохраняют силу.

Вместо света мы могли бы воспользоваться системой щеток и батареей для пропускания тока через отверстия. Ток отсчитывается при помощи амперметра или регистрируется посредством реле. Согласно рассуждению раздела 2 главы 14 предел составляет и в этом случае $k \ln 2$ на один отсчет. В механическом устройстве типа граммофонной пластинки небольшие положительные или отрицательные импульсы могут быть обнаружены лишь при условии, что механическая система дает энергию больше, чем $kT \ln 2$ на импульс, что соответствует увеличению энтропии на $k \ln 2$. Аналогичные соображения относятся к импульсам, записанным на магнитной ленте. Во всяком случае негэнтропия, потерянная в процессе чтения, больше, или в крайнем случае равна получаемой информации, и обобщенный принцип Карно (см. (18.4)) всегда соблюдается.

Сходные соображения относятся к вопросу об информации, записанной в чертежах. Каждый раз, когда нам нужна информация, мы должны читать чертеж. Процесс чтения поглощает негэнтропию от источника света, и часть этой негэнтропии преобразуется в требуемую информацию, скажем, для построения изделия по чертежу. Мы еще вернемся к этим вопросам.

4. Письмо и печать

Рассмотрим теперь другой вопрос: сколько негэнтропии затрачивается в процессе письма? Эта негэнтропия теряется, но мы получаем новую копию информации, так что печатающий механизм дает еще один пример преобразования негэнтропии в информацию.

Фотографическая печать есть простейший для обсуждения пример: мы кладем перфоленту на фотобумагу и освещаем каждое отверстие. Каждый раз для отпечатания положения отверстия нужен по меньшей мере один световой квант $h\nu$. Ситуация сходна с описанной в разделе 3 и ведет к аналогичным заключениям: отпечатание новой копии стоит ровно столько, сколько прочтение. Новая копия информации оплачивается негэнтропией.

Во всех этих рассуждениях мы игнорировали типичные необратимые процессы: проявление фотобумаги, усиление электрических импульсов и т. п. Все эти технические усовершенствования стоят нам большого количества негэнтропии, и общая цена процесса чтения или печати много выше вычисленного нами предела.

Существенно, что вся информация оплачивается негэнтропией. Количество требуемой негэнтропии конечно, но настолько мало (10^{-16} на дв. ед.), что до сих пор полностью игнорировалось. Связь между энтропией и информацией логически необходима. Мы были бы не в состоянии обсуждать проблемы этой главы без учета этой связи. Очень малое значение (10^{-16}) негэнтропии, требуемой для получения одной двоичной единицы информации, играет однако очень важную роль в современной жизни и делает возможным сообщать информацию за пренебрежимо малую цену.

Отметим теперь закон распада: любая система, в которой записана информация, будет с течением времени вырождаться. Кроме того, и сама информация может постепенно терять свою ценность. По обеим причинам записанная информация следует общему закону распада, согласующемуся с обобщенным принципом Карно. Этот распад, однако, совершается гораздо медленнее для мертвой (записанной) информации, нежели для живой информации.

5. Обсуждение специального примера

Письмо и чтение настолько широко применяются в повседневной жизни, что мы не замечаем, как часто мы к ним прибегаем. Рассмотрим, например, процедуру посылки телеграммы. Производятся следующие действия:

A. Отправитель пишет телеграмму в отделении связи.

B. Телеграфист читает телеграмму и переводит ее в импульсы тока в передатчике.

C. Импульсы тока принимаются приемником, и аппарат печатает сообщение.

D. Телеграфист читает адрес и посылает копию телеграммы по этому адресу.

E. Адресат читает телеграмму.

Мы опустили еще два действия, потому что в них участвует мысль, а наша теория не способна трактовать такие процессы:

A_0. Отправитель составляет свою телеграмму так, чтобы сообщить некоторую информацию адресату.

F. Адресат понимает значение телеграммы.

Ограничивая процедуру этапами A, B, C, D и E, мы замечаем, что в этих этапах участвуют следующие процессы:

A. Письмо.

B. Чтение, кодирование, передача.

C. Декодирование, печатание.

D. E. Чтение.

Каждая операция письма, чтения или печати имеет свою негэнтропийную цену в соответствии с рассуждениями разделов 3 и 4. Операции кодирования и декодирования обычно требуют применения запоминающих устройств, из которых извлекаются кодовые числа. Этот процесс также имеет некоторую негэнтропийную цену. Операция передачи и приема, обсужденная Шенноном, была исследована в главе 17. Копии телеграммы попадут в руки различных заинтересованных людей, и эти копии соответствуют мертвой информации, которая может быть снова и снова прочитана при условии вложения соответствующей энергии.

Упомянутые выше запоминающие устройства будут рассмотрены полнее в главе 19, посвященной вычислительным машинам, в которых операции чтения и письма механизированы и всякий раз используют некоторое количество негэнтропии.

6. Новая информация и избыточность

Мы показали в разделе 3, что процесс чтения требует увеличения энтропии ΔS, которое больше или по крайней мере равно количеству информации ΔI, получаемому читающим. Поставим теперь следующий вопрос: должна ли эта информация рассматриваться как новая информация или она является лишь избыточностью, не несущей новой информации, поскольку она всего лишь повторяет информацию, содержащуюся в написанном документе? Этот вопрос стоит на грани философии, но мы попытаемся дать на него практический ответ. Когда некто читает книгу впервые, он, конечно, получает информацию, и это может считаться новой информацией для читателя. Если он снова читает книгу второй и третий раз, то получает только избыточность, а не новую информацию[1]).

---

[1]) С этим трудво согласиться. Если бы это было так, то мы не перечитывали бы книги по нескольку раз, а между тем мы постоянно это делаем. Дело в том, что, во-первых, при первом чтении, как правило, извлекается не вся содержащаяся в книге информация. Во-вторых, с течением времени происходит распад запасенной в мозгу информации (она, попросту говоря, забывается), и эта утечка требует пополнения. Таким образом, избыточной естественно было бы считать такую информацию, которая повторяет информацию, уже имеющуюся в запасенной форме в нашем мозгу в данный момент. (*Прим. перев.*)

С абсолютной точки зрения мы можем представить себе идеального ученого, знающего все, что известно человечеству. Для этого идеального философа каждая новая книга содержит информацию. Однако другие читатели получают только избыточную информацию, так как она уже известна идеальному ученому. Это весьма нереалистическая точка зрения.

Для того чтобы сделать различие между возможными точками зрения на информацию и избыточность, связанные с чтением и печатанием копий, мы введем следующие определения [1]):

Абсолютная информация: любая информация, доступная какому-либо человеку на Земле, причем эта информация считается только один раз, каково бы ни было число людей, знающих ее. Так, сверхсекретный документ имеет столько же информации, сколько номер газеты, содержащий то же количество букв.

Распределенная информация: произведение количества абсолютной информации на число людей, располагающих этой информацией. Второе определение представляется более реалистичным, чем, скорее философская, *абсолютная информация*, и эта точка зрения принята в данной главе.

Имеются, однако, проблемы, в которых можно предпочесть определение абсолютной информации, и различие между обоими видами информации следует иметь в виду.

---

[1]) L. Brillouin, Am. Scientist 38, 594 (1950).

ГЛАВА 19

ПРОБЛЕМА ВЫЧИСЛЕНИЯ

1. Вычислительные машины

Мы рассмотрим теперь вычислительные машины и их действие на информацию. Часто говорят, что вычислительная машина производит новую информацию. Это в действительности не так. Машины могут обрабатывать информацию: они берут сырой материал и выдают законченный продукт, но общее количество информации не возрастает. При идеальных обстоятельствах информация может оставаться постоянной на протяжении вычисления, но в обычных условиях имеются некоторые потери, и конечная информация меньше начальной.

Рассмотрим вопрос более внимательно. Процесс распадается на две стадии: программирование и собственно вычисление. Программирование выполняется учеными, которые намерены использовать машину. Они обсуждают задачу, выбирают математические законы, которые должны быть применены, приготавливают программу арифметических операций, которые должны быть выполнены в определенной последовательности. Все эти операции записываются в таком виде, чтобы машина могла их прочитать: на перфоленте, на магнитной ленте и т. п. Эта часть работы требует от ученых мышления и знания. Это может быть названо информацией в общем смысле, но это является типичным случаем, когда наши определения информации неприложимы. Мы неоднократно подчеркивали, что наше статистическое определение информации не включает какого-либо человеческого рассуждения и что мы не можем, по крайней мере в настоящее время, обсуждать проблему мышления.

Когда программа работы приготовлена в виде перфоленты, мы можем сосчитать число двоичных единиц на ней и найти количество информации в этой окончательной программе. Едва ли это имело бы большое значение.

Следующий шаг состоит в рассмотрении машины с ее специальной программой и в изучении того, как она в действительности работает. Мы вводим определенные цифровые данные на вход. Машина обрабатывает эти данные, следуя заданной программе, вычисляет окончательные данные и выдает их на выход. Если машина не делает ошибок в вычислениях, то она действует в точности как кодирующая машина. Она использует информацию, содержащуюся во входных данных, и переводит ее в выходные данные, которые в лучшем случае могут содержать всю входную информацию, но не более.

Машину можно сравнить с каналом передачи, включая кодирование и декодирование. При идеальных условиях сообщение передается правильно, и информационное содержание остается постоянным. Ошибки и приближения при вычислении ведут к потере информации. Работу машины лучше всего сравнить с работой переводчика: переводчику даны словарь и правила грамматики, что позволяет ему перевести на английский язык статью, первоначально написанную по-японски. Словарь и грамматика соответствуют программе работы. Японская статья — это входные данные, а английский перевод — выходные данные. Общая информация, имевшаяся на входе, может быть (в лучшем случае) обнаружена на выходе.

Сходным образом машине, вычисляющей, например, таблицы стрельбы, дана программа, учитывающая законы движения снаряда и атмосферные условия. Входные данные представлены начальными положением и скоростью. Выходные данные — точка попадания и конечная скорость. Входные данные логически содержат решение, но мы не в состоянии непосредственно прочесть его. Это есть кодированное сообщение, требующее декодирования, и это, в сущности, означает то же, что и *перевод*. Если действие машины совершенно, то она должна быть обратима: если дано попадание, то она может вычислять в обратном порядке и найти начальные условия. Аналогично переводчик может сделать перевод обратно на японский язык и получить оригинал, или его эквивалент. Идеальные условия работы соответствуют обратимости, а следовательно, сохранению информации.

Обратимость в целом может быть получена путем последовательных шагов, которые могут включать и необратимые операции, как-то: запись данных в запоминающем устройстве с последующим прочтением их. Каждая из этих операций имеет определенную негэнтропийную цену (см. разделы 3 и 4 главы 18). Работа машины по определенной программе действительно обойдется в некоторое количество негэнтропии (энергия деградирует). Это имеет значение с технической точки зрения, но не изменяет наших заключений, и можно себе представить, что эта цена работы будет снижена в будущих лучших машинах.

Процитируем здесь очень интересное рассуждение, написанное более ста лет назад Эдгаром По. Этот автор рассматривает различные автоматы, и, в частности, вычислительную машину Бэббеджа. Нижеследующая цитата очень точно выражает нашу теперешнюю точку зрения:

«Но если эти машины остроумны, то что же мы должны думать о вычислительной машине Бэббеджа? Что мы должны думать о машине из дерева и металла, которая может не только вычислять астрономические и навигационные таблицы любой заданной протяженности, но и сделать точность своих действий математически достоверной благодаря своей способности исправлять свои возможные ошибки? Что мы должны думать о машине, которая может не только выполнять все это, но и печатать свои сложные результаты, когда они получены, без малейшего вмешательства интеллекта человека? Будет, возможно, сказано в ответ, что машина, какую мы описали, всецело выше, в сравнении с игроком в шахматы Мельцеля. Никоим образом — она всецело ниже его, если мы допустим (чего никогда не следует допускать), что игрок в шахматы есть *чистая машина* и выполняет свои действия без всякого непосредственного воздействия человека. Арифметические и алгебраические вычисления по своей природе фиксированы и определены. Если имеются определенные *данные*, то определенные результаты последуют необходимо и неизбежно. Эти результаты ничем не определяются и ничем не обусловливаются, кроме как исходными данными. И подлежащий решению вопрос продвигается к своему окончательному определению последовательными непогрешимыми шагами, не подверженными никаким изменениям. Если так, мы можем без труда представить себе *возможность* построения такого механизма, который после запуска его в соответствии с *данными* подлежащего решению вопроса продолжает свои движения регулярно, последовательно и неуклонно в направлении требуемого решения, так как эти движения, хотя бы и сложные, никогда нельзя представить себе иначе, как конечными и определенными» [1]), [2]).

2. Вычислительная машина как математический элемент

Интересно исследовать роль вычислительной машины, когда она работает с входными данными, полученными из эксперимента. Эти данные являются результатом наблюдений над некоторой физической системой, и серия наблюдений дает обычно непрерывную функцию $f(t)$. Так как всякая физическая система имеет конечную ширину полосы, то это относится также и к $f(t)$.

---

[1]) Edgar A. Poe, Maelzel's Chess Player, 1836.

[2]) Общие положения этого раздела наводят на следующие размышления. Итак, машина не создает информации; иначе говоря, никакая последовательность математических и логических действий не приводит к появлению новой информации (мы оставляем в стороне интереснейший вопрос о том, какова же действительная природа какого-либо *источника* информации). Следовательно, *решение поставленной задачи не дает нам новой информации.* Новая информация не содержится также ни в таблицах, ни в справочниках и т. п. Это уже может показаться странным, хотя несомненным образом вытекает из принятых определений. Но в таком случае ощущается настоятельная необходимость введения еще какой-то новой категории, — может быть, не информации, — которая выражала бы эффект, т. е. полезный результат работы вычислительной машины. Большая вычислительная машина стоит дорого и затрачивает значительную мощность, превращая ее в тепло. (Известно, что только проблема охлаждения машины представляет немалые технические трудности.) Что же мы получаем от машины? Мы получаем от нее результат *переработки* информации по заданной программе. Этот результат имеет определенную ценность, и мы могли бы ввести количественную меру еще не названной величины, характеризующей полезный эффект машины. Эта количественная мера, очевидно, связана со сложностью обработки и могла бы быть непосредственно выражена числом элементарных арифметических действий в двоичной системе счисления. (Заметим, что *практической* характеристикой машины является именно число элементарных операций в секунду. Мы обсуждаем, однако, принципиальную сторону дела.)

До введения определенной величины, характеризующей полезный результат работы вычислительной машины, невозможна никакая разумная оценка эффективности машины. (*Прим. перев.*)

Цифровая вычислительная машина работает с дискретными сигналами. Операция отсчитывания в точности соответствует задаче, обсужденной в разделах 6 и 7 главы 8. Если мы хотим использовать всю информацию, содержащуюся в непрерывной функции, то мы должны взять отсчеты с интервалом

$$\theta = \frac{1}{2\nu_M}, \quad f_m = f(m\theta). \tag{19.1}$$

Исходная функция f практически определена только на конечном промежутке времени τ, но для получения точной математической задачи мы должны установить поведение функции от $-\infty$ до $+\infty$, как мы сделали в разделе 6 главы 8. Предположим снова, что мы имеем периодическую функцию с периодом τ, удовлетворяющую условию (8.52). На вход машины поступают N отсчетов f_m, и если τ велико, то мы имеем:

$$N = 2\nu_M\tau \tag{19.2}$$

(см. (8.51)). По этим отсчетам мы можем найти коэффициенты Фурье, и наоборот (см. 8.63) и (8.63a)):

$$A_n = \frac{1}{N}\sum_{m=1}^{N} f_m e^{-inm\omega_0\theta}, \quad \omega_0 = \frac{2\pi}{\tau}, \tag{19.3}$$

$$f_m = \sum_{n=-n_M}^{+n_M} A_n e^{inm\omega_0\theta}. \tag{19.4}$$

Машина обрабатывает отсчеты f_m и дает выходные значения

$$\psi_\mu = F_\mu(f_1 \ldots f_m \ldots f_N). \tag{19.5}$$

Вообще говоря, функции F_μ, представляющие программу вычисления, могут быть очень сложными. Если, однако, мы ограничим свое рассуждение *линейной математической задачей*, то операция будет представлена матрицей $(M_{\mu m})$

$$\psi_\mu = \sum_{m=1}^{N} M_{\mu m} f_m, \tag{19.6}$$

и эта матрица выражает программу вычисления. Мы имеем N различных значений m на входе, но мы можем иметь иное число N_1 значений μ на выходе. Возможно также, что имеется некоторая избыточность как во входных, так и в выходных данных. Если мы не хотим изменять ширину полосы системы в целом, то наши N_1 выходных данных потребуют иной длительности τ_1 для непрерывной функции на выходе

$$\Psi(t)=\sum_{\mu=1}^{N_1}\psi_\mu g(t-\mu\theta),\quad \tau_1=N_1\theta, \tag{19.7}$$

которая восстанавливается по правилам главы 8 (см. (8.60)) и может быть снова разложена в ряд Фурье (как в (8.58))

$$\Psi(t)=\sum_{\beta=-\beta_M}^{+\beta_M}\alpha_\beta e^{i\beta\omega_1 t},\quad \beta_M\nu_1=\nu_M,\quad \nu_1=\frac{1}{\tau_1},\quad \omega_1=\frac{2\pi}{\tau_1}. \tag{19.8}$$

Соотношение между коэффициентами Фурье α_β и отсчетами выходной функции ψ_μ в точности такое же, как в (19.3)

$$\alpha_\beta=\frac{1}{N_1}\sum_{\mu=1}^{N_1}\psi_\mu e^{-i\beta\mu\omega_1\theta}. \tag{19.9}$$

Из (19.4)—(19.9) мы можем непосредственно получить соотношение между входным спектром A_n и выходным спектром α_β, предполагая линейное вычисление, представляемое матрицей $M_{\mu m}$:

$$\alpha_\beta=\frac{1}{N_1}\sum_{\mu=1}^{N_1}\sum_{m=1}^{N}\sum_{n=-n_M}^{n_M}M_{\mu m}A_n e^{i(nm\omega_0-\beta\mu\omega_1)\theta}. \tag{19.10}$$

Если мы положим, что число N_1 выходных данных равно числу N входных данных, то формула упрощается. Матрица M есть квадратная матрица и

$$N_1=N,\quad \tau_1=\tau,\quad \omega_1=\omega_0,\quad n_M=\beta_M,$$

$$\alpha_\beta=\frac{1}{N}\sum_{\mu=1}^{N}\sum_{m=1}^{N}\sum_{n=-n_M}^{n_M}M_{\mu m}A_n e^{i(nm-\beta\mu)\omega_0\theta}, \tag{19.11}$$

где $-n_M \leqslant \beta \leqslant n_M$. Величина $(nm - \beta\mu)$ есть положительное или отрицательное целое число, а коэффициенты α_β выходного спектра являются рациональными функциями аргумента

$$z = e^{-i\omega_0\theta} = e^{-\frac{i\pi}{\tau\nu M}}. \tag{19.12}$$

Эти формулы полностью подкрепляют точку зрения, развитую в предыдущем разделе: машина обрабатывает информацию, но не производит никакой новой информации. Операция может выполняться в обратном порядке, если матрица несингулярна и может быть обращена. Это значит, что обратная матрица J может быть найдена

$$(M)(J) = (J)(M) = 1. \tag{19.13}$$

Обозначим через $J_{m'\mu'}$ элементы этой обратной матрицы, представляющей программу обратного действия машины.

Условия (19.13) означают:

$$\left.\begin{aligned} &\sum_{m=1}^{N} M_{\mu m} J_{m\mu'} = \delta_{\mu\mu'} = \begin{cases} 1, & \mu = \mu', \\ 0, & \mu \neq \mu', \end{cases} \\ &\sum_{\mu=1}^{N} J_{m'\mu} M_{\mu m} = \delta_{mm'}, \end{aligned}\right\} \tag{19.14}$$

и обратное действие машины представляется соотношением, сходным с (19.11):

$$A_{n'} = \frac{1}{N} \sum_{m'=1}^{N} \sum_{\mu'=1}^{N} \sum_{\beta'=-n_M}^{n_M} J_{m'\mu'} \alpha_{\beta'} e^{i(\beta'\mu' - n'm')\omega_0\theta}. \tag{19.15}$$

Доказательство нетрудно и использует (19.14) совместно с (8.63b):

$$\frac{1}{N} \sum_{m=1}^{N} e^{ipm\omega_0\theta} = \delta_{p0} \quad \begin{cases} 1, & p = 0, \\ 0, & p \neq 0, \text{ целое.} \end{cases} \tag{19.16}$$

Подставляя (19.11) в (19.15), получаем (при $\beta' = \beta$):

$$A_{n'} = \frac{1}{N^2} \sum_{m'\mu'\beta'\mu m n} J_{m'\mu'} M_{\mu m} A_n e^{i(nm - n'm')\omega_0\theta} e^{i\beta'(\mu' - \mu)\omega_0\theta}. \tag{19.17}$$

Суммирование по β' в последнем экспоненциальном множителе и использование (19.16) дает $\delta_{\mu'-\mu, 0}$, откуда $\mu' = \mu$. Затем мы вычисляем сумму

$$\sum_{\mu} J_{m'\mu} M_{\mu m}$$

и получаем $m' = m$ в соответствии с (19.14). Последний шаг состоит в суммировании $e^{i(n-n')m\omega_0\theta}$ по m, что дает $n = n'$ согласно (19.16), так что окончательно мы получаем:

$$A_n = A_{n'}$$

что и требовалось доказать.

Мы получили полное представление машины, работающей по линейной обратимой программе, причем информация не теряется, но и не создается заново. Небольшая потеря информации будет обусловлена округлением ошибок вычисления. Вычисление по линейной программе, представляемой прямоугольной матрицей M, будет терять информацию, если $N_1 < N$, вводить избыточность, если $N_1 > N$. Нелинейные программы (см. (19.5)) исследовать труднее.

Мы обсудим в следующем разделе несколько иную задачу, соответствующую возможному использованию вычислительной машины.

3. Вычислительная машина как элемент схемы; отсчитывание и восстановление

Вычислительная машина может быть введена в состав системы управления, что приводит к многим важным применениям. Эта проблема подробно исследована Линвиллом[1]) и Зальцером[2]). Здесь данные получаются из наблюдений над некоторым механизмом, а результаты вычислений используются для управления работой этого механизма. Вычислительная машина является частью системы управления и должна рассматриваться как элемент схемы. Система управления в целом работает с непрерывными сигналами с конечной

---

[1]) W. K. Linvill, диссертация MIT, 1950; PIRE **40**, 230 (1952).

[2]) J. M. Salzer, диссертация MIT, 1951; PIRE **40**, 231 (1952); **41**, 901 (1953); **42**, 457 (1954); доклад в группе электроники IRE, май 1952; см. также Hughes Aircraft Co. Memo № 338, 1954.

шириной полосы, а вычислительная машина применяет дискретные значения. Отличие от проблемы, обсужденной в предыдущем разделе, состоит в том, что мы полагали ранее, что располагаем для вычисления полной совокупностью данных, соответствующих законченному эксперименту длительностью τ, и что мы обсуждаем результаты на досуге, после того как эксперимент завершен.

В системе управления мы должны производить вычисления немедленно, используя все данные, полученные вплоть до текущего момента t, для того, чтобы быть в состоянии управлять работой механизма как можно скорее. В такой задаче полная длительность эксперимента заранее не определена. Метод предыдущего раздела должен быть заменен шенноновым методом отсчетов (см. раздел 7 главы 8).

Для того чтобы использовать всю информацию, содержащуюся в проходящей полосе, без введения ненужной избыточности, мы опять берем отсчеты с интервалом θ (см. (19.1)). Исходная функция может быть восстановлена согласно (8.69) с помощью импульсных функций:

$$g_s(t) = \frac{\sin 2\pi\nu_M t}{2\pi\nu_M t} = \frac{1}{2\nu_M} \int\limits_{\nu=-\nu_M}^{\nu_M} e^{i2\pi\nu t}\, d\nu \qquad (19.18)$$

и

$$f(t) = \sum_m f_m g_s(t - m\theta) = \int\limits_{\nu=-\nu_M}^{\nu_M} A(\nu)\, e^{i2\pi\nu t}\, d\nu. \qquad (19.19)$$

Импульсные функции g_s простираются на некоторое расстояние как вперед, так и назад. Это значит, что функция $f(t)$ в момент t не может быть найдена, если мы располагаем только прошлыми значениями ($m < 0$).

Линвилл применяет другую импульсную функцию $g_L(t)$, представляющую очень острый импульс, показанный на рис. 19.1. Мы можем определить такой острый импульс формулой, сходной с (19.18):

$$g_L(t) = \frac{\sin 2\pi\nu_1 t}{2\pi\nu_1 t} = \frac{1}{2\nu_1} \int\limits_{-\nu_1}^{\nu_2} e^{i2\pi\nu' t} d\nu', \qquad (19.20)$$

где $\nu_1 \gg \nu_M$. Полученная таким образом функция обозначена $\tilde{f}(t)$:

$$\tilde{f}(t) = \sum_m f_m g_L(t - m\theta) = \int_{-\nu_1}^{\nu_1} \tilde{A}(\nu') e^{i2\pi\nu' t} d\nu'. \qquad (19.21)$$

Из (19.18) и (19.19) находим:

$$A(\nu) = \frac{1}{2\nu_M} \sum_m f_m e^{-i2\pi\nu m\theta}, \quad -\nu_M \leqslant \nu \leqslant \nu_M, \qquad (19.22)$$

тогда как (19.20) и (19.21) дают:

$$\tilde{A}(\nu') = \frac{1}{2\nu_1} \sum_m f_m e^{-i2\pi\nu' m\theta}, \quad -\nu_1 \leqslant \nu' \leqslant \nu_1. \qquad (19.23)$$

Сравнивая (19.22) и (19.23), мы замечаем, что спектр $\tilde{A}(\nu')$ совпадает с $A(\nu)$ в области низких частот и воспроизводит

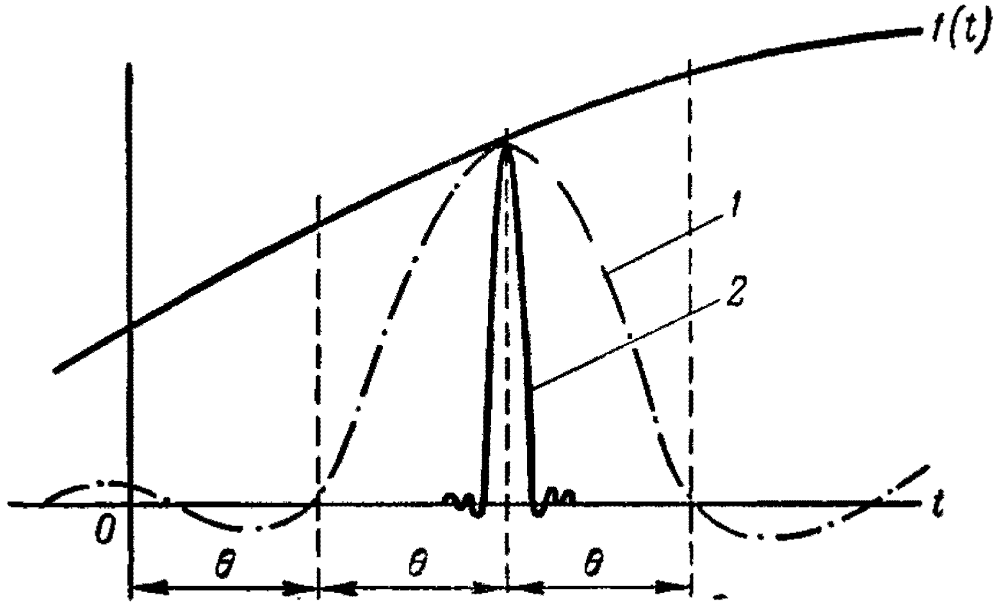

Рис. 19.1. Импульсные функции Шеннона (*1*) и Линвилла (*2*) для восстановления функции $f(t)$.

этот спектр периодически в области более высоких частот:

$$-\nu_M \leqslant \nu' \leqslant \nu_M; \quad \tilde{A}(\nu') = \frac{\nu_M}{\nu_1} A(\nu), \qquad (19.24)$$

и если мы возьмем, например, $\nu_M \leqslant \nu \leqslant 3\nu_M$ и положим $\nu' = \nu + 2\nu_M$, то

$$A(\nu') = \frac{1}{2\nu_1} \sum_m f_m e^{i2\pi\nu m\theta - i2\pi m} = \frac{\nu_M}{\nu_1} A(\nu), \qquad (19.25)$$

так как $2\nu_M\theta = 1$. Вообще, если

$$(2k-1)\nu_M \leqslant \nu \leqslant (2k+1)\nu_M,$$

то мы берем

$$\nu' = \nu + 2k\nu_M$$

и получаем:

$$\tilde{A}(\nu') = \frac{\nu_M}{\nu_1} A(\nu). \tag{19.26}$$

Подытоживая эти результаты, мы имеем:

$$\frac{\nu_1}{\nu_M} \tilde{A}(\nu') = \sum_{k=-\infty}^{\infty} A(\nu' - 2k\nu_M). \tag{19.27}$$

Метод отсчетов Линвилла дает функцию $\tilde{f}(t)$, которая легко может быть вычислена по прошлым значениям, так как ее импульсы четко локализованы. Она отличается от исходной функции добавочными спектрами высшего порядка. Процесс восстановления сводится к простой фильтрации, сохраняющей только полосу $|\nu| \leqslant \nu_M$. Эта операция дает снова исходную функцию $f(t)$ с точностью до постоянного множителя ν_M/ν_1.

4. Вычисление по отсчетам в момент t

Вычисление ведется по данным, известным в момент t. Входными данными являются:

$$f_m = f(m\theta), \quad m \leqslant m_t, \quad m_t\theta = t. \tag{19.28}$$

Выходные данные, которые можно получить в этот момент,

$$\psi_\mu = \psi(\mu\theta), \quad \mu \leqslant m_t - 1. \tag{19.29}$$

Если вычисление производится по *линейной программе*, то мы можем записать соотношение (Зальцер):

$$\psi_{m_t} + \sum_{k=1}^{k_1} b_k(m_t)\psi_{m_t-k} = \sum_{h=0}^{h_1} a_h(m_t) f_{m_t-h}, \tag{19.30}$$

где используются k_1 ранее вычисленных значений и h_1+1 последних входных данных. Коэффициенты b_k, a_h могут зависеть от времени (через m_t). Важным случаем является ***линейная, не зависящая от времени*** программа вычисления с постоянными коэффициентами b_k и a_h. Заменяя выходные значения ψ_{m_t-k} их линейными выражениями через входные значения, можно записать:

$$\psi_{m_t}=\sum_{h=0}^{h_2} c_h f_{m_t-h}. \tag{19.31}$$

Число членов суммы h_2 велико и может даже стать бесконечным, образуя бесконечный ряд. Эта формула позволяет сделать сравнение с результатами раздела 2.

Возьмем

$$m_t=\mu, \quad m_t-h=m, \quad h=\mu-m, \tag{19.32}$$

$$\psi_\mu=\sum_{m=\mu}^{\mu-h_2} c_{\mu-m} f_m. \tag{19.33}$$

Сравнивая с (19.6), получаем матрицу преобразования

$$M_{\mu m}=\begin{cases} 0, & m>\mu, \\ c_{\mu-m}, & \mu-h_2\leqslant m\leqslant\mu, \\ 0, & m<\mu-h_2. \end{cases} \tag{19.34}$$

Эта матрица представлена таблицей для $h_2=3$:

	0	1	2	3	4	5	6	7 $\to\mu$
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)

Элементы матрицы ниже диагонали равны нулю. Все строки имеют одинаковые постоянные коэффициенты, просто смещенные на μ позиций вправо. Эта специальная матрица дает более простые результаты, чем более общая матрица раздела 2.

5. Коэффициент передачи вычислительной машины

Исследуем соотношение между входным и выходным спектрами. Представим функции f и ψ интегралами Фурье, как в (19.12), и получим:

$$f_m = \int_{-\nu_M}^{\nu_M} A(\nu)\, e^{i2\pi\nu t}\, d\nu, \quad t = m\theta, \tag{19.36}$$

$$\psi_m = \int_{-\nu_M}^{\nu_M} \alpha(\nu)\, e^{i2\pi\nu t}\, d\nu. \tag{19.37}$$

Подставим эти выражения в (19.30) и получим интегралы с $e^{2\pi i\nu t}\, d\nu$ в обеих частях. Левая часть имеет множитель

$$\alpha(\nu) + \sum_{k=1}^{k_1} b_k \alpha(\nu)\, e^{-i2\pi\nu k\theta}, \tag{19.38}$$

тогда как в правой части имеем:

$$\sum_{h=0}^{h_1} a_h A(\nu)\, e^{-i2\pi\nu h\theta}. \tag{19.38a}$$

Так как обе функции в (19.30) равны, то равны и их преобразования Фурье:

$$\alpha(\nu) + \sum_k b_k \alpha(\nu)\, e^{-i2\pi\nu k\theta} = \sum_h a_h A(\nu)\, e^{-i2\pi\nu h\theta}. \tag{19.39}$$

Из этого соотношения находим коэффициент передачи

$$W(\nu)=\frac{\alpha(\nu)}{A(\nu)}=\frac{\sum\limits_{h=0}^{h_1} a_h e^{-i2\pi\nu h\theta}}{1+\sum\limits_{k=1}^{k_1} b_k e^{-i2\pi\nu k\theta}}=$$

$$=\frac{a_0+a_1 z+a_2 z^2+\ldots+a_{h_1} z^{h_1}}{1+b_1 z+b_2 z^2+\ldots+b_{k_1} z^{k_1}}, \qquad (19.40)$$

дающий отношение выходного спектра к входному для рассматриваемой вычислительной машины. Коэффициент передачи выражается рациональной функцией аргумента

$$z=e^{-i2\pi\nu\theta}, \qquad (19.41)$$

где θ — интервал отсчетов. Весьма примечательно, что такого рода вычисление, как представленное выражением (19.30), может быть исследовано этим методом и дает на каждой частоте выходной спектр $\alpha(\nu)$, зависящий только от входного спектра $A(\nu)$ на той же частоте. Коэффициент передачи (19.40) был получен Зальцером и позволил ему рассмотреть целый ряд задач, в которых вычислительные машины и обычные схемы были объединены в сложную систему управления.

Мы могли бы с равным успехом начать с (19.31), где предполагалось, что решение предыдущей системы получено для того, чтобы исключить ψ_{m-k}. Это соотношение дает, как и прежде,

$$W(\nu)=\frac{\alpha(\nu)}{A(\nu)}=\sum_{h=0}^{h_2} c_h e^{-i2\pi\nu h\theta} \qquad (19.42)$$

— многочлен или бесконечный ряд по z.

Все это рассуждение полностью согласуется с точкой зрения на роль вычислительной машины, развитой в предыдущих разделах: вычислительная машина представляет собой автоматическое устройство для обработки информации, точно так же как станок обрабатывает куски металла. Новая информация не создается, и на выходе вычислительной машины получается в точности то же количество информации, какое имелось на входе (в идеальном случае без потерь), только другим образом кодированное.

Мы основывали наше рассуждение на методе Фурье, тогда как Линвилл и Зальцер применяли преобразование Лапласа. Оба метода эквивалентны, и формулы Зальцера могут быть получены простой заменой $i\omega$ на s:

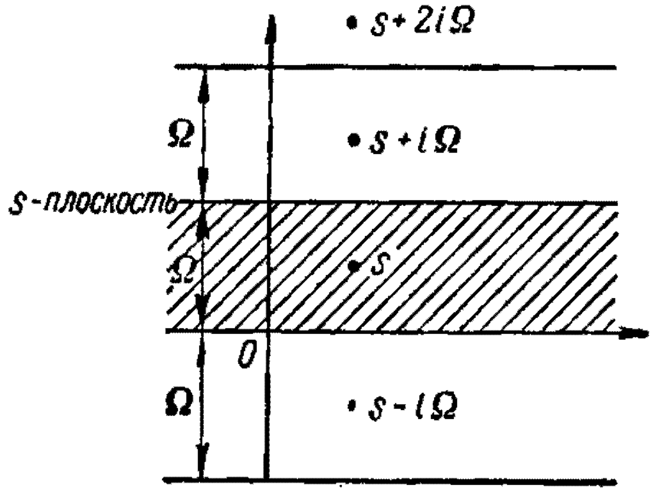

Рис. 19.2. Горизонтальные полоски на комплексной плоскости s, в которых значения $\tilde{A}(s)$ повторяются.

$$\left.\begin{aligned} s &= i\omega, \quad \omega = -is, \\ z &= e^{-i\omega\theta} = e^{-s\theta}. \end{aligned}\right\} \qquad (19.43)$$

Для общности рекомендуется рассматривать s как комплексную величину, но θ (интервал отсчетов) — всегда вещественная величина. Мы нашли, что дискретные данные $\tilde{f}(t)$ имеют периодический спектр $\tilde{A}(\nu)$ (см. раздел 3):

$$\left.\begin{aligned} &\tilde{A}(\omega + n\Omega) = \tilde{A}(\omega) \quad (n \text{ — целое}) \\ \text{и}\quad & \\ &\Omega = 4\pi\nu_M = \frac{2\pi}{\theta} \text{ — вещественно.} \end{aligned}\right\} \qquad (19.44)$$

Это свойство легко перевести на переменную s, что означает:

$$\tilde{A}(s - in\Omega) = \tilde{A}\left(s - \frac{2\pi i n}{\theta}\right) = \tilde{A}(s). \qquad (19.45)$$

В плоскости комплексной переменной s функция $\tilde{A}$ принимает одинаковые значения в следующих друг за другом полосах, как показано на рис. 19.2. Если функция известна в пределах одной из этих полос (например, в заштрихованной полосе), то она известна на всей плоскости, так как периодически повторяется на других полосах.

6. Схемы, содержащие вычислительную машину; проблема устойчивости

Предыдущее исследование показало, что вычислительная машина, работающая по заданной программе, может рассматриваться как элемент схемы с определенным коэффициентом передачи $W(\nu)$, зависящим от частоты ν. Когда вычислительная машина включена последовательно с обычной схемой, то остальная часть схемы характеризуется другим коэффициентом передачи $W_1(\nu)$, и система в целом имеет коэффициент передачи

$$W_{\text{total}}(\nu) = W(\nu)\, W_1(\nu). \tag{19.46}$$

Это замечание позволяет обсуждать задачу о схеме с вычислительной машиной, следуя хорошо известным методам теории цепей. Характерно, что коэффициент передачи $W(\nu)$ вычислительной машины есть рациональная функция переменной z (см. 19.41)), тогда как обычная схема имеет коэффициент передачи, рациональный относительно ω.

Одним из важных вопросов исследования является проблема *устойчивости*. Вычислительная машина может умножить функцию $f(t)$ на множитель C, действуя, таким образом, как усилитель. Программа общего вида, рассмотренная Зальцером (раздел 4, равенство (19.30)), использует некоторое число прежних выходных значений для вычисления выходного значения в момент t. Это значит, что программа требует системы обратной связи. Если машина работает в качестве *усилителя с обратной связью*, то мы должны, разумеется, обсудить проблему устойчивости, так как усилитель с обратной связью может поддерживать незатухающие колебания, если обратная связь достаточно сильна.

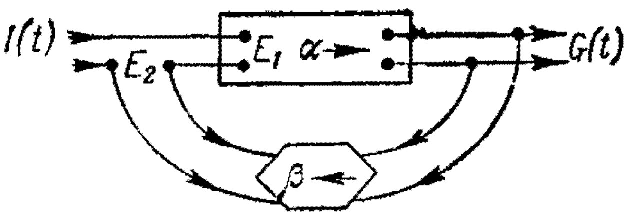

Рис. 19.3. Усилитель α с обратной связью β.

Поскольку этот вопрос имеет фундаментальное значение, займемся им подробно и рассмотрим обычный усилитель α с обратной связью β, как показано на рис. 19.3. Напряжение E_1 на входе дает отклик $G(t)$ на выходе:

$$G = \alpha E_1, \tag{19.47}$$

где α — комплексная величина. Доля β этого отклика подается обратно и добавляется к первоначальному входному воздействию $I(t)$, так что

$$\left.\begin{aligned} E_2 &= \beta G = \alpha\beta E_1, \ \beta \text{ комплексная}, \\ E_1 &= I + E_2 = I + \alpha\beta E_1, \end{aligned}\right\} \tag{19.48}$$

откуда

$$E_1 = \frac{I}{1-\alpha\beta}, \quad G = \frac{I\alpha}{1-\alpha\beta}. \tag{19.49}$$

Это дает коэффициент передачи усилителя с обратной связью

$$W_a = \frac{G}{I} = \frac{\alpha}{1 - \alpha\beta}. \tag{19.50}$$

Это выражение очень сходно с коэффициентом передачи (19.40), полученным для вычислительной машины, если положить:

$$\alpha(\nu) = \sum a_h z^h, \quad z = e^{-2\pi i\nu\theta}, \quad -\alpha\beta = \sum b_k z^k. \tag{19.51}$$

Функция вида (19.50) имеет некоторое число полюсов при

$$\alpha\beta = 1 \tag{19.52}$$

— условие, которое может выполняться при определенных частотах. При этих частотах мы можем получить конечный отклик G при нулевом воздействии I, так как W бесконечно: система самовозбуждается. Положение полюсов на комплексной плоскости $s = \sigma + i\omega$ имеет первостепенное значение и будет обсуждено в следующем разделе.

В итоге мы можем сказать о ситуации следующее:

A. Вычислительная машина аналогична *активной цепи.*

B. Устойчивость программы вычисления может быть исследована теми же методами, что и устойчивость активных цепей.

C. Можно исследовать амплитудные и фазовые характеристики данной программы, а также траектории, импульсную реакцию, диаграмму Найквиста и т. п.

D. Ширина полосы вычислительной машины непосредственно определяется интервалом отсчетов θ (см. (19.1)):

$$\text{наивысшая частота: } \nu_M = \frac{1}{2\theta}. \tag{19.53}$$

Зальцер показал на многочисленных примерах эффективность этого метода рассуждения. Он пояснил также, как согласовать схему с вычислительной машиной, или наоборот. Встречаемые здесь проблемы согласования сходны с такими же проблемами теории цепей.

7. Обсуждение устойчивости программы

Вычислительная машина, работающая по заданной программе, характеризуется коэффициентом передачи $W(s)$, и отклик $\alpha(s)$ пропорционален воздействию $A(s)$, умноженному на W (см. 19.40)):

$$\alpha(s) = A(s)\, W(s). \tag{19.54}$$

Пока $|W|$ остается конечным, отклик конечен и $|W|$ выражает получаемое усиление. Если $|W|$ становится бесконечно большим, мы можем получить отклик без всякого воздействия. Это соответствует полюсам функции $W(s)$. Пусть

$$s = \sigma + i\omega \tag{19.55}$$

будет полюсом. На выходе имеем:

$$\alpha(s)\, e^{st} = \alpha(s)\, e^{\sigma t + i\omega t} \tag{19.56}$$

Если α положительно или нуль, колебания со временем нарастают или сохраняют неизменную амплитуду. Когда σ отрицательно, колебания затухают. Они могут возбудиться в переходном режиме, но в конце концов исчезнут.

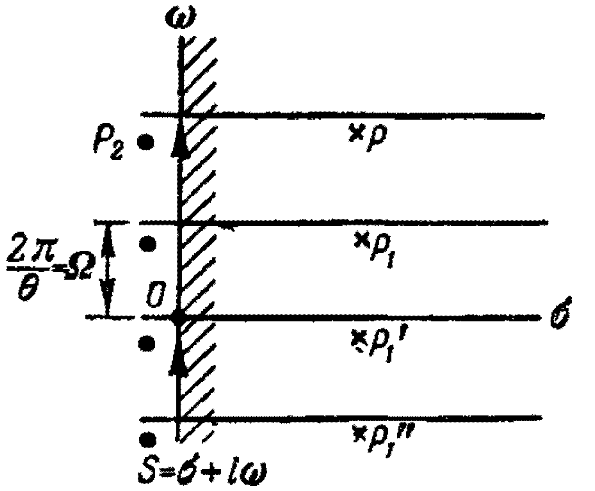

Рис. 19.4. На комплексной плоскости s полюс P_1 справа от мнимой оси означает неустойчивость, тогда как полюс P_2 слева не дает неустойчивости. Показано повторение в последовательных полосах; устойчивые полюсы обозначены точками, неустойчивые — крестиками.

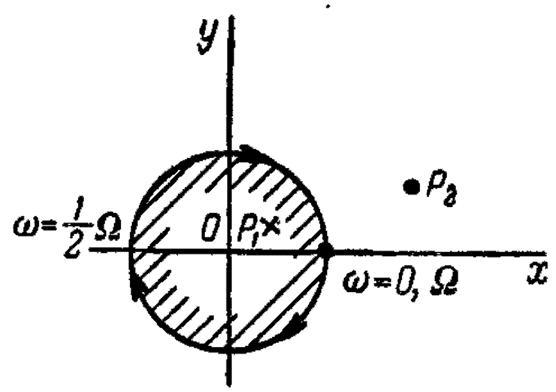

Рис. 19.5. Конформное преобразование $z = e^{-s\theta}$ комплексной плоскости s на комплексную плоскость z.

Мнимая ось ω плоскости s преобразована в единичную окружность плоскости z. Полюсы, соответствующие P_1 (неустойчивые) рис. 19.4, попадают внутрь круга; полюсы, соответствующие P_2 (устойчивые), — вовне.

Таким образом, неустойчивость соответствует полюсам коэффициента передачи с положительной вещественной частью. Эти полюсы находятся справа от мнимой оси переменной s. Вследствие периодичности данный полюс P_1 встречается в каждой полосе высотой Ω. Полюс P_2 слева от мнимой оси не дает неустойчивости. Рис. 19.4 иллюстрирует общее положение: полосы с отрицательным σ (устойчивые) отмечены точками, тогда как полюсы с неотрицательным σ (неустойчивые) отмечены крестиками. Рекуррентность полюса P_1 в последовательных полосах высотой Ω показана, как P_1', P_1'', ...

Вместо плоскости s можем рассматривать комплексную плоскость z, показанную на рис. 19.5. Формула

$$z = x + iy = e^{-s\theta} \tag{19.57}$$

представляет конформное преобразование плоскости s на плоскость z. Начало 0 в плоскости z означает $z = 0$ и соответствует $s = +\infty$. Окружность радиуса 1 в плоскости z соответствует мнимой оси плоскости s

$$s = i\omega, \quad z = e^{-i\omega\theta},$$

а внутренность единичного круга z представляет правую полуплоскость s. Полюс P_1 (и его изображения P_1', P_1'', ...) в правой полуплоскости s соответствует полюсу p_1 внутри единичного круга на плоскости z. Полюс P_2 в левой полуплоскости s представляется точкой p_2, лежащей вне единичного круга на плоскости z.

Задача определения местонахождения полюсов коэффициента передачи может теперь быть поставлена более точно. Коэффициент передачи W из (19.40) может быть записан следующим образом:

$$W = \frac{P(s)}{Q(s)} = \frac{\sum a_h z^h}{1 + \sum b_k z^k}. \tag{19.58}$$

После того как общие корни P и Q исключены, полюсы соответствуют остающимся корням Q:

$$Q = 1 + \sum b_k z^k = 0. \tag{19.59}$$

Мы можем теперь рассмотреть другое конформное преобразование, перейдя от комплексной плоскости z к комплексной плоскости Q, как показано на рис. 19.6. Начало 0 плоскости Q соответствует всем корням функции Q (см. (19.59)); следовательно, оно представляет все полюсы коэффициента передачи W. Точка $Q = 1$ на плоскости Q отображает точку $z = 0$ — начало плоскости z. Единичная окружность плоскости z преобразуется в кривую (траекторию) L на плоскости Q, и внутренность единичного круга преобразуется во внутренность траектории L. Если полюс P_1 расположен внутри единичного круга плоскости z, то начало плоскости Q лежит в области, ограниченной траекторией L, как показано на рис. 19.6, *b*. Такое положение выражает неустойчивость, На рис. 19.6, *a* показано начало, лежащее вне траектории L, что соответствует устойчивости.

Зальцер рассмотрел различные примеры и показал большую практическую ценность метода. Очень часто грубый набросок траектории L достаточен, чтобы определить устойчивость. По тому, насколько траектория обходит начало, можно судить

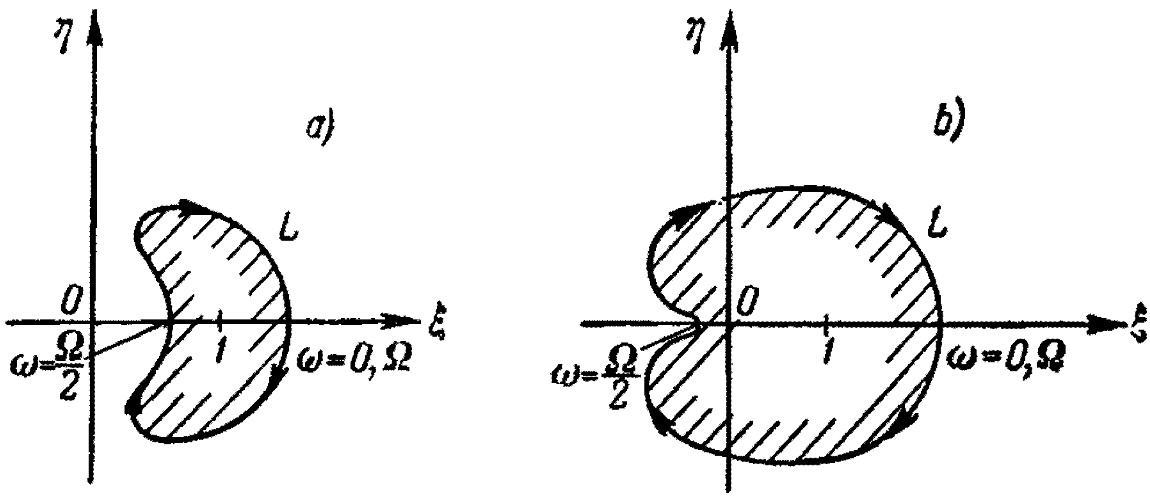

Рис. 19.6. Комплексная Q-плоскость есть конформное преобразование

$$Q=\xi+i\eta=1+\sum_k b_k z^k.$$

a) пример, в котором начало координат плоскости Q лежит вне траектории L. Это — устойчивый случай; *b*) пример, в котором начало координат плоскости Q лежит внутри траектории L. Это — неустойчивый случай.

об относительной устойчивости. Частота наибольшего приближения к началу очень часто является резонансной частотой программы вычисления.

Коэффициент передачи программы вычисления всегда выражается функцией типа (19.40) или (19.58). Когда коэффициент передачи для операции известен, выполнимость операции вычислительной машиной доказана, если коэффициент передачи может быть записан в виде дроби (19.40) или (19.58). Это является главным и, разумеется, очень важным результатом рассуждений Зальцера.

8. Несколько примеров

Приведем несколько примеров коэффициентов передачи, соответствующих некоторым типичным вычислительным задачам.

А. Вычислительная машина как интегратор

В этом случае отклик ψ должен представлять определенный интеграл от входного воздействия $f(t)$, так что можем записать:

$$\psi_m = \psi_{m-1} + \int_{(m-1)\theta}^{m\theta} f(t)\,dt = \psi_{m-n} + \int_{(m-n)\theta}^{m\theta} f(t)\,dt = \int_0^{m\theta} f(t)\,dt. \qquad (19.60)$$

Последние выражения получаются итерацией равенства между первыми двумя. Коэффициент передачи выражается формулой (19.40) или (19.58)

$$W = \frac{\Sigma\, a_n z^n}{1 + \Sigma\, b_n z^n}, \qquad (19.61)$$

и задача состоит в нахождении a_n и b_n. Для этого воспользуемся приближенным методом вычисления интеграла в (19.60) совместно с (19.30):

$$\psi_m + \sum_n b_n \psi_{m-n} = \sum_n a_n f_{m-n}, \qquad (19.62)$$

где $f_k = f(k\theta)$. Мы дадим результаты вычисления интеграла при помощи ступенчатой функции, по правилу трапеций и по *правилу* $^1/_3$ и *правилу* $^3/_8$ Симпсона.

Ступенчатая функция. Мы находим приближенное значение интеграла в (19.60), заменяя $f(t)$ ступенчатой функцией и вычисляя площади получаемых прямоугольников шириной θ и с высотами $f_1, f_2, \ldots, f_m$:

$$\psi_m - \psi_{m-1} = \int_{(m-1)\theta}^{m\theta} f(t)\,dt = f_m\theta.$$

Сравнение с (19.62) дает:

$$b_1 = -1, \quad a_0 = \theta, \quad b_j = a_k = 0 \quad \text{при} \quad j > 1, \quad k > 0,$$

и, следовательно, на основании (19.61) коэффициент передачи равен

$$W_0 = \frac{\theta}{1-z}. \tag{19.63}$$

Правило трапеций. Применение этого правила при $n=1$ дает:

$$\psi_m - \psi_{m-1} = \int\limits_{(m-1)\theta}^{m\theta} f(t)\,dt = \frac{\theta}{2}(f_m + f_{m-1}),$$

откуда с помощью (19.62)

$$b_1 = -1, \quad a_0 = a_1 = \frac{\theta}{2},$$

$$b_j = a_k = 0 \quad \text{при} \quad j > 1,\ k > 1.$$

Таким образом, коэффициент передачи есть

$$W_1 = \frac{\theta}{2}\,\frac{1+z}{1-z}. \tag{19.64}$$

Правило $^1/_3$ Симпсона. Мы должны взять $n=2$ в (19.60):

$$\psi_m - \psi_{m-2} = \int\limits_{(m-2)\theta}^{m\theta} f(t)\,dt = \frac{\theta}{3}(f_{m-2} + 4f_{m-1} + f_m),$$

откуда

$$b_2 = -1, \quad a_0 = a_2 = \frac{\theta}{3},\ a_1 = \frac{4\theta}{3},$$

$$b_j = a_k = 0 \quad \text{при} \quad j \neq 2, \quad k > 2,$$

и коэффициент передачи равен

$$W_2 = \frac{\theta}{3}\,\frac{1+4z+z^2}{1-z^2}. \tag{19.65}$$

Правило $^3/_8$ Симпсона. Аналогично, беря $n=3$, получаем:

$$W_3 = \frac{3\theta}{8}\,\frac{1+3z+3z^2+z^3}{1-z^3}. \tag{19.66}$$

В. Вычислительная машина как дифференциатор

Теперь отклик ψ должен представлять производную входного воздействия f. Так как дифференцирование есть операция, обратная интегрированию, коэффициент передачи идеального дифференциатора должен быть обратным по отношению

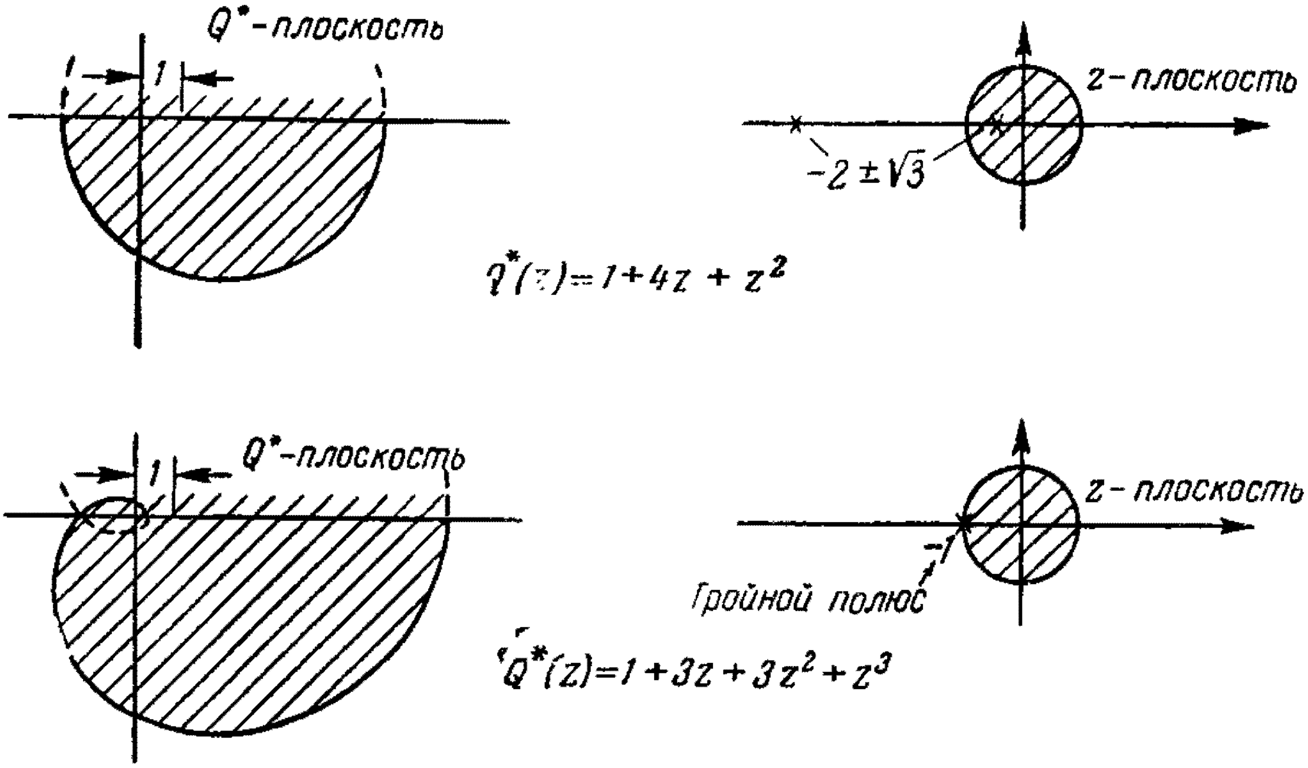

Рис. 19.7. Кривые, показывающие неустойчивость программы дифференцирования, соответствующей W_2^{-1} и W_3^{-1} как в плоскости Q^*, так и в плоскости z.

Q^* означает знаменатель функции передачи (числитель соответствующего W). Отметим, что обе траектории охватывают начало координат плоскости Q^*.

к коэффициенту передачи идеального интегратора. Это значит, что W_0^{-1}, W_1^{-1}, W_2^{-1} и W_3^{-1} могут годиться в качестве программ дифференцирования. Оказывается, однако, что только W_0^{-1} и W_1^{-1} устойчивы, тогда как W_2^{-1} и W_3^{-1} неустойчивы. Положение показано на рис. 19.7, взятом из работы Зальцера. Изображены полюсы на плоскости z и форма траектории на плоскости Q^*.

С. Восстановление (desampling)

Процесс восстановления может быть выполнен путем замены импульсов Линвилла снова импульсами Шеннона или очень тщательной отфильтровкой всех частот $\nu > \nu_M$. Первый способ не вызывает ни фазовых сдвигов, ни изменений интенсивности при $\nu < \nu_M$ и, таким образом, дает коэффициент передачи, равный 1. Второй способ требует нулевой интенсивности при $\nu > \nu_M$ и, следовательно, дает $W = 0$ для этого диапазона частот. Это едва ли осуществимо.

Большинство фильтров обеспечивает в полосе прозрачности хорошую передачу по интенсивности, но дает фазовый сдвиг. Это не имеет значения для телефонной передачи, так как ухо малочувствительно к фазе. Однако при восстановлении функции фаза существенна, и поэтому построение фильтров, предназначенных для восстановления, отличается от построения обычных фильтров.

D. Временные задержки

Если мы применяем импульсы вида $(\sin x)/x$ в предположении, что функция постоянна при больших x, мы должны вычислить:

$$\sum_{\substack{\text{по большим} \\ \text{значениям } x}} \frac{\sin x}{x} \approx \int_{\text{большое } x}^{\infty} \frac{\sin x}{x}\, dx \approx \frac{\cos x}{x}.$$

Это выражение сходится очень медленно. Если мы желаем иметь ошибку, не превосходящую 10^{-n} в окончательном результате, то нам потребуется $x > 10^n = N$ членов. Это большое число членов вызывает задержку, так как функция в момент t будет с требуемой точностью известна лишь на N членов позднее, что означает задержку на $N\theta$.

ГЛАВА 20

ИНФОРМАЦИЯ, ОРГАНИЗАЦИЯ И ДРУГИЕ ПРОБЛЕМЫ

1. Информация и организация

Мы смогли исследовать информационное содержание записанного предложения, совокупности чисел и многих других случаев, к которым наше статистическое определение информации может непосредственно применяться. Мы можем сходным образом подсчитать информацию, содержащуюся в чертеже, при условии, что чертеж может быть представлен конечным числом символов. Рассмотрим, например, схему некоторого электрического устройства. Мы можем занумеровать все имеющиеся на схеме зажимы от 0 до N и описать схему, указав номера соединенных между собой зажимов: скобка (1023—0216) будет означать, что мы должны соединить зажим 1023 с зажимом 0216. Для конечного числа зажимов число P возможных соединений конечно; данная схема выбирает из этого числа одну частную систему. Схема содержит, следовательно, информацию

$$I = K \ln P \tag{20.1}$$

в соответствии с определениями главы 1. Еще раз подчеркнем точный смысл такого определения: оно не делает различия между схемой действующего устройства и другой схемой, которая дала бы никчемную структуру с короткими замыканиями и другими недостатками. Информация, которую мы подсчитываем, это как раз та информация, которая дается человеку, выполняющему монтаж схемы, и она ничего не говорит о том, что устройство должно делать.

Вопросы о том, почему выбрана данная схема и как она работает, не рассматриваются. Это составляло задачу ученого или инженера, который проектировал устройство. Это потребовало мышления и научного обсуждения и, может быть, других человеческих процессов, которые все находятся вне рамок наших теперешних методов исследования.

Рассмотрим пример: электрическое устройство содержит n элементов, каждый из которых имеет m зажимов, и q зажимов для внешних соединений. Всего имеется

$$N = nm + q \tag{20.2}$$

зажимов, каждый из которых может быть соединен с любым другим, что дает:

$$P = N^N$$

различных возможностей. Информация, содержащаяся в одной схеме, равна

$$I = K \ln P = KN \ln N. \tag{20.3}$$

Беря $K = \log_2 e \approx 1{,}4$, получим информацию в двоичных единицах

$$I = N \log_2 N.$$

Возьмем разумные по порядку величины

$$n = 999, \quad m = 10, \quad q = 10,$$

что дает:

$$N = 10\,000 \approx 2^{13{,}3}; \quad I = 133\,000 \text{ дв. ед.}$$

Это — большая цифра в двоичных единицах, но если перевести в единицы энтропии, то мы должны умножить ее на 10^{-16} (этот множитель представляет приближенно отношение единиц) и получим негэнтропию

$$-S = 1{,}33 \cdot 10^{-11},$$

т. е. пренебрежимо малую величину.

Машина, построенная по нашим спецификациям, содержит такую структурную негэнтропию, которая представляет информацию или организацию машины. Отрицательным членом можно, однако, полностью пренебречь в сравнении с полной энтропией машины. Этот пример выявляет некоторые характерные особенности всей теории.

Негэнтропийный принцип информации требует, чтобы всякая информация была связана с соответствующим отрицательным членом в выражении для физической энтропии системы. Эта связь существенна для логической полноты, что было показано рассуждениями глав 8 — 18.

На практике этими отрицательными членами можно в большинстве случаев пренебречь вследствие малости множителя 10^{-16}, вводимого при переходе от двоичных единиц к термодинамическим единицам[1]). Частные организации могут содержать очень большое количество информации в двоичных единицах. Рассмотрим, например, телефонную сеть масштаба, сравнимого с американской системой. Число абонентов имеет порядок десятков миллионов, но будем щедры и положим, что число абонентов составляет сто миллионов:

$$N = 10^8.$$

Прямое применение соотношения типа (20.3) показывает, что в некоторый данный момент количество информации, содержащееся во всей системе, по порядку величины равно

$$I = N \log_2 N \approx 40 \cdot 10^8 = 4 \cdot 10^9 \text{ дв. ед.},$$

что остается все же очень малой величиной ($4 \cdot 10^{-7}$), будучи выражено в единицах энтропии.

Трудно представить себе какое-либо техническое устройство, содержащее много большее количество информации, чем в предыдущем примере, но, обращаясь к живым организмам, мы встречаемся с величинами совершенно другого порядка. Число клеток может быть порядка 10^{12}, и трудно оценить число взаимосвязанных химических молекул в живом организме. Информация, содержащаяся в живом организме, может достигать значительных величин по шкале энтропий. Но это — пока неисследованная область.

---

[1]) Эта аргументация, неоднократно повторяемая в книге, представляется методически неудачной. Та или иная величина должна считаться большой или малой в зависимости не от того, в каких единицах она измеряется, а от того, с чем она сравнивается. (Иначе говоря, понятия «большой» и «малый» обычно применяются в физике к безразмерным отношениям.) По существу, речь идет все время о том, что относительное изменение физической энтропии (связанное с информацией) очень мало. Это положение остается в силе вне зависимости от выбора единиц. (*Прим. перев.*)

2. Информация, содержащаяся в физическом законе

Мы испытали на многих примерах полезность нашего статистического определения *количества информации*. Теперь мы попытаемся распространить это определение на более общие проблемы и предложить некоторые возможные новые области применения.

В главах 14 — 16 мы исследовали информацию, получаемую при физическом эксперименте, и нашли, что количество информации не может быть определено, пока полностью не установлены пределы наблюдения (field of observation) и точность эксперимента. Мы встретимся со сходным положением в вопросе об информации, содержащейся в физическом законе. Для того чтобы дать точный ответ на такой вопрос, нужно прежде всего точно поставить вопрос и указать условия, в которых закон был получен, и как он будет применяться.

Рассмотрим, например, уравнение идеального газа

$$f(p,\ v,\ T) = \frac{pv}{T} = R = \text{const}. \qquad (20.4)$$

Мы должны установить область применения, указать, к какому газу уравнение применяется, каков диапазон изменения переменных p, v и T и какую точность мы ожидаем получить.

Физический закон есть всего лишь итог экспериментальных результатов, и справедливость закона не может обсуждаться, если не указаны условия эксперимента. Имея в виду вышеприведенный пример, мы можем обсудить проблему в общем виде. Положим, что мы наблюдали некоторую физическую систему и измерили некоторые переменные $x_1, x_2, \ldots, x_n$. Для каждой переменной x_i мы указываем диапазон изменения X_i и наименьший интервал наблюдения α_i:

переменная x_i, диапазон X_i, погрешность наблюдения

$$\delta x_i = \alpha_i,$$

полное число экспериментальных точек на X_i равно X_i/α_i. После обсуждения результатов мы заключаем, что некоторое выражение

$$f(x_1,\ x_2,\ \ldots,\ x_n) = C = \text{const} \qquad (20.6)$$

правильно отображает экспериментальные факты, с погрешностью δf, никогда не превосходящей некоторого числа φ.

Но это еще не полная постановка задачи. Мы должны добавить еще одну величину, а именно, полный диапазон F изменения функции f. Этот полный априорный диапазон изменения зависит от предыдущих знаний, которые мы имели до начала наших экспериментов. Если мы ничего не знаем о системе, то F определяется через пределы наших измерительных устройств. Таким образом, функция f окончательно определяется на полном интервале F со ступенями φ и, следовательно, с числом возможных значений F/φ, которые все равновероятны до того, как мы начали наши наблюдения. Итак, мы имеем:

$$N = \frac{X_1 X_2 \ldots X_n}{\alpha_1 \alpha_2 \ldots \alpha_n} \tag{20.7}$$

различных точек наблюдения в многомерном пространстве $x_1, \ldots, x_n$, и для каждой точки мы можем ожидать одного из F/φ значений f. Число возможностей, таким образом, составляет:

$$\left.\begin{aligned} &P_0 = \left(\frac{F}{\varphi}\right)^N \quad \text{до эксперимента,} \\ &P_1 = 1^N = 1 \quad \text{после эксперимента,} \end{aligned}\right\} \tag{20.8}$$

и информация определяется по формулам (1.5) и (1.6)

$$I = K \ln \frac{P_0}{P_1} = KN \ln \frac{F}{\varphi} = KN(\ln F - \ln \varphi). \tag{20.9}$$

Область изменения переменных, как предполагается в (20.5), представляет собой параллелепипед. Это ограничение не является необходимым. Мы можем иметь в многомерном пространстве $x_1, \ldots, x_n$ объем U любой формы, и мы обозначим через α малый объем, соответствующий погрешности одного наблюдения. Это дает вместо (20.7)

$$N = \frac{U}{\alpha}, \tag{20.10}$$

и информация, содержащаяся в физическом законе, равна

$$I = \frac{KU}{\alpha}(\ln F - \ln \varphi). \tag{20.11}$$

Эта формула удовлетворяет следующим общим требованиям: чем точнее данные (т. е. чем меньше α и φ) или чем шире диапазон наблюдений U, тем больше информация.

Применение логарифма, как указано в главе 1, необходимо для обеспечения аддитивности независимых наблюдений.

Мы можем теперь обсудить значение величины F, сравнивая информацию, получаемую последовательными шагами с возрастающей точностью. Предположим, что первая серия наблюдений дает функцию f_1 с погрешностью φ_1, тогда как вторая более точная серия экспериментов дает функцию f_2 с меньшей погрешностью φ_2. Таково, например, положение, когда закон Бойля (20.4) заменяется формулой Ван-дер-Ваальса. Равенство (20.11) дает соответствующее увеличение информации:

$$\Delta I = I_2 - J_1 = \frac{KU}{\alpha}(\ln \varphi_1 - \ln \varphi_2). \qquad (20.12)$$

Эта формула совпадает с (20.11), за исключением того, что F и φ заменены на φ_1 и φ_2. Величина F исключена. Она фактически играет роль аддитивной постоянной, сходной с неизвестной аддитивной постоянной во многих формулах для энтропии.

Этот пример показывает, что наше статистическое определение информации может применяться в весьма разнообразных примерах и приводит к интересным результатам.

Мак-Кэй [1]) обсуждал проблемы научной информации в двух недавних работах, в которых он пытался доказать, что различные определения приложимы в различных областях. Он называет *селективной информацией* величину, которую мы определили как *информацию*, и вводит термины *структурная информация* и *метрическая информация*. Эти последние термины применяются в связи с проблемами точности экспериментальных устройств. Наши исследования и рассуждения стремятся показать, что все эти проблемы могут и должны обсуждаться при помощи одних и тех же определений и формул. Если бы это было не так, мы, без сомнения, натолкнулись бы на несогласованность, так как области

---

[1]) D. M. MacKay, Phil. Mag. [7] 41, 289 (1950); Symposium on Information Theory, London, Trans. IRE (PGIT), p. 60 (1953).

применения различных формул перекрываются, и мы имели бы три различные и не связанные между собой теории информации вместо одной. Мы определенно ощущаем, что необходима только одна теория и что она может покрыть адекватным образом все поле исследований.

3. Информация, содержащаяся в числовой таблице

Мы встречаемся с интересным примером в вопросе об информации, содержащейся в числовой таблице. Этот вопрос часто ставился, и предлагалось сравнение между:

А. Таблицей случайных чисел.

В. Таблицей, представляющей данную функцию.

Общий ответ прост: наша таблица содержит N позиций, соответствующих N значениям переменной x. Для каждого значения x таблица дает значение другой переменной $y(x)$, представленное числом из n десятичных цифр или b двоичных цифр. Числа n и b связаны в среднем соотношением

$$b \approx 3{,}3n. \tag{20.13}$$

В общем таблица содержит

$$I = Nb \tag{20.14}$$

двоичных единиц информации, каков бы ни был закон, связывающий y и x. В случае А, для таблицы случайных чисел, совокупность чисел y должна удовлетворять некоторым средним соотношениям, не должна обнаруживать в среднем корреляции, и т. д., но как только таблица дана, в ней нет уже ничего случайного. Она представляет собой такую же таблицу, как и всякая другая, и можно сказать, что она представляет определенную функцию $y(x)$, которая, вообще говоря, не непрерывна и не имеет производной, но тем не менее является функцией дискретной переменной x.

Если таблица представляет определенную функцию $y(x)$, мы можем различать следующие случаи:

B_1. Если принять, что мы ничего не знаем о функции наперед, то числовая таблица по-прежнему содержит информацию Nb, как в (20.14).

B_2. Мы можем иметь некоторые предварительные сведения о функции. Мы можем, например, располагать доказательствами того, что функция имеет первую, вторую, ... n-ю производные, за исключением некоторых точек на интервале (некоторых значений x). Эта предварительная информация I_0 должна быть вычтена из информации, даваемой формулой (20.14), и информация, содержащаяся в таблице, равна

$$I = Nb - I_0. \tag{20.15}$$

Пусть, например, Δx означает разность между двумя соседними значениями x в таблице, и предположим, что нам известно, что приращение $|\Delta y|$, соответствующее приращению Δx, не может превосходить b' (в двоичных единицах). Тогда имеем:

$$-2^{b'} \leqslant \Delta y \leqslant +2^{b'}. \tag{20.16}$$

Для первого значения функции при $x = 0$ имеем 2^b возможных значений y. Для следующего значения диапазон сокращается до

$$2 \cdot 2^{b'} = 2^{b'+1}.$$

Общее число возможностей есть

$$P = 2^b (2^{b'+1})^{N-1},$$

и информация в двоичных единицах равна

$$I = \log_2 P = b + (N-1)(b'+1) \tag{20.17}$$

— величина меньшая, чем Nb, так как $b'+1$ меньше b.

B_3. Мы можем, наконец, предположить, что нам в точности известно наперед, что представляет собой табулированная функция; в этом случае таблица вовсе не дает нам новой информации.

Обычно мы имеем случай B_2, когда мы располагаем лишь частичным знанием табулированной функции и нуждаемся в таблице для более подробной информации. Наши определения показывают, что наибольшую информацию дает таблица случайных чисел, о которых наперед ничего не известно.

4. Общие замечания

Теория информации возникла из проблем связи, и большая часть её современных приложений относится к этой области. Мы старались показать, что теория может быть также очень полезна в чистой науке, и в особенности в физике. Аналогия между информацией и энтропией была отмечена

Шенноном и восходит фактически к старой работе Силарда, выполненной в то время, когда практическая ценность теории еще не была осознана.

Информация и физическая энтропия имеют одинаковую природу. Энтропия есть мера недостатка детальной информации о физической системе. Чем больше информация, тем меньше энтропия. Информация представляет собой отрицательное слагаемое в энтропии системы, и мы установили негэнтропийный принцип информации.

Всякий раз, когда мы делаем наблюдение над физической системой, мы должны иметь разного рода источники негэнтропии. Мы используем эту негэнтропию и увеличиваем энтропию лаборатории, содержащей наблюдаемую систему и измерительные приборы. В результате мы получаем некоторое количество информации о системе. Увеличение энтропии, однако, всегда больше полученной информации.

Этот результат представляет собой обобщение принципа Карно, и мы проверили его справедливость на многочисленных примерах. Мы показали, таким образом, что теория информации не может быть построена как самостоятельная сущность. Связь с термодинамикой настолько тесна, что для обеспечения последовательности и полноты необходима физическая теория информации. Доказательство этого положения и является целью этой книги.

Практически теория даст для большинства задач чрезвычайно малые изменения энтропии физической системы. Информация измеряется обычно в двоичных единицах, а энтропия — в термодинамических единицах. Отношение этих единиц равно приблизительно 10^{-16}, а это значит, что очень большое число двоичных единиц (скажем, 10^6) соответствует пренебрежимо малому (10^{-10}) изменению энтропии. Однако учет связи между обеими величинами необходим для логической полноты. Многие из задач, обсужденных в главах 16 и 17, были камнями преткновения. Они не могли бы быть решены без помощи нашей *физической теории*, и для их толкования необходима эта новая форма теории.

Очень малое значение негэнтропии, соответствующее очень большим количествам информации, является основной причиной того, что передача информации обычно обходится недорого. Письмо, печать и электрическая связь стоят очень немного в единицах энтропии. Соответственно низка и их денежная стоимость. Современная жизнь основана на этих фактах, и она была бы совершенно иной в мире, в котором негэнтропия информации имела бы большую величину.

То обстоятельство, что в наших рассуждениях мы имели дело с такими ничтожными количествами, не должно нас обескураживать. Вначале казалось, что теория относительности дает лишь очень малые поправки к классической механике. Новые приложения к ядерной физике показывают теперь фундаментальное значение соотношения между массой и энергией. Можно надеяться, что и связь между энтропией и информацией выступит рано или поздно на первый план и мы откроем, где эта связь получает свое полное использование. Проблема измерения чрезвычайно малых расстояний, рассмотренная в главе 16, дает нам первый пример; другие могут вскоре последовать.

5. Примеры проблем, выходящих за рамки данной теории

Существующая теория полностью игнорирует ценность передаваемой и обрабатываемой информации. Это обстоятельство настойчиво подчеркивалось на протяжении всей книги. Многие другие авторы недооценивали, по-видимому, важность этого ограничения, что повело ко многим недоразумениям по поводу возможностей существующей теории. Только игнорируя человеческую ценность информации мы смогли построить научную теорию информации, основанную на статистике, и была показана полезность этой теории. Существует, однако, много проблем, которые нельзя разрешить методами этой теории. Цель этого раздела состоит в том, чтобы вкратце поставить некоторые из этих вопросов и показать, что все они сводятся к одной и той же общей проблеме: как ввести в теорию элемент ценности информации.

Надобность в этом новом элементе возникает всякий раз, когда информация рассматривается как основание для предсказания и для практического применения. Сперва подытожим кратко существующее положение: положим, например, что посылаемая телеграмма закодирована двоичным кодом (0 или 1) и содержит n двоичных цифр. Если кодирование не имеет избыточности, то телеграмма содержит всего n двоичных единиц информации. Вопрос о том, имеет телеграмма значение для получателя или нет, не рассматривается.

На научной конференции в Милане в апреле 1954 г. д-р Куфиньяль (Париж) поставил следующий вопрос: после того как телеграмма послана, автор добавляет одну двоичную цифру со следующим значением:

0: все неверно; не обращайте на телеграмму внимания,

1: телеграмма правильная; можете ею воспользоваться.

С точки зрения информации это означает добавление одной двоичной единицы, так что всего получается $(n+1)$ единица вместо n. Практически же положение иное:

0 («нет») — разрушает всю ценность содержащейся в телеграмме информации,

1 («да») — попросту добавляет избыточность.

Мы можем, следовательно, сказать, что «нет» (0) несет отрицательную информацию, равную по абсолютной величине полной информации в телеграмме. С точки зрения существующей теории информация всегда положительна и никогда не является отрицательной. Здесь ясно обнаруживается различие между статистической мерой информации и практической ценностью этой информации для получателя [1]).

---

[1]) Эти рассуждения вызывают ряд замечаний. Во-первых, отрицательная информация может появиться и в рамках существующей теории. Поясним это примером. Пусть получена телеграмма: «Разыщите квитанцию; она заложена в одну из книг, находящихся в столовой», а вслед за тем другая телеграмма: «А может быть, и в кабинете». Вторая телеграмма увеличивает число априори возможных исходов, т. е. увеличивает неопределенность исходной ситуации. В силу основного определения (см. (1.5)) это означает, что вторая телеграмма несет отрицательную информацию. Во-вторых, следует, быть может, пояснить, что *отрицание сообщения* дает, вообще говоря, информацию, не равную нулю, так как это отрицание уменьшает число возможных исходов. Куфиньяль имеет в виду, конечно, другое; он говорит не об отрицании сообщения, а об отрицании того, что сообщение содержало какую-либо информацию. Таким образом, сообщение вместе с дополнительным отрицанием в смысле Куфиньяля действительно дает информацию, равную нулю, так как исходная неопределенность ситуации полностью восстанавливается. Мы можем с полным основанием говорить, что происходит аннигиляция информации в смысле взаимной компенсации положительной и отрицательной информации. Но спрашивается, какой информации следует приписать отрицательный знак? Нам представляется естественным считать отрицательной первоначальную *ложную* информацию и положительной ту информацию, которая содержится в последующем отрицании, аннулирующем ложную информацию и восстанавливающем исходное положение.

Наконец, можно было бы указать на возможность некоторого обобщения, которое позволяет, в известной степени, ввести рассматри-

Коды с самопроверкой и самоисправлением, рассмотренные в главе 6, соответствуют практическим задачам, очень сходным с рассмотренной выше. К телеграмме, содержащей n двоичных знаков, мы добавляем 0 или 1 так, чтобы число единиц было всегда четно. Приемник проверяет общее число единиц. Если оно нечетно, то может быть 1, 3, 5, ... ошибок. Если оно четно, то либо ошибок нет, либо имеется 2, 4, 6, ... ошибок. Первый случай, очевидно, соответствует сигналу «нет» в задаче Куфиньяля.

Дополнительные сигналы для проверки или исправления не добавляют никакой информации. По существующей теории эти сигналы просто добавляют избыточность. Здесь речь идет, однако, о практической ценности информации.

Вместо того чтобы разделять утверждения на истинные и ложные («да» или «нет», см. выше), мы можем условиться указывать вероятность p того, что сообщение является истинным. Выберем шкалу с 8 ($=2^3$) значениями, и пусть число m ($0 \leqslant m \leqslant 8$) соответствует $p = m/8$, что означает, что утверждение в среднем истинно в m случаях из 8 и ложно в $(8-m)$ случаях из 8. С информационной точки зрения это означает добавление к телеграмме трех двоичных единиц информации, что дает в общем $(n+3)$ дв. ед. вместо n. Поскольку рассматривается ценность информации для получателя, разумно считать, что ценность выражается произведением pn. Во многих случаях, однако, информация вовсе не представляет интереса для получателя, если вероятность ошибки превосходит некоторый установленный минимум, например 50%; в этом случае практическая ценность будет равна нулю для вероятностей 0, 1/8, 2/8, 3/8 и принимать некоторые возрастающие значения, когда вероятности составляют 4/8, 5/8, 6/8, 7/8, 1.

ваемую ситуацию в рамки существующей теории. Речь идет, в сущности, о том, что отправитель может ошибаться и посылать как верные, так и неверные сообщения. Но если так, то можно *включить отправителя в систему связи* и рассматривать его как звено этой системы, в котором с определенной вероятностью правильные сообщения могут заменяться неправильными. Роль непогрешимого отправителя остается при этом за той объективной действительностью, наблюдая которую действует фактически составитель сообщения, будь то человек или машина. При такой постановке вопроса все определения и понятия существующей теории применимы и сохраняют свой первоначальный смысл. (*Прим. перев.*)

Вопрос о надежности физического наблюдения обсуждался в главе 14. Всякое физическое измерение содержит возможность ошибок. Надежность r экспериментального метода была определена через вероятность $1/r$ ошибки, обусловленной тепловыми флуктуациями, и вероятность $\left(1-\frac{1}{r}\right)$ правильного наблюдения. Эта задача, очень сходная с предыдущей, была дополнена предположением, что наименьшее рассматриваемое значение есть $r=2$, что дает вероятность правильного наблюдения, равную 0,5.

Вычислительные машины рассматривались в главе 19, и было показано, что вычисление очень сходно с переводом, кодированием и декодированием. В лучшем случае вычисление сохраняет всю информацию, содержащуюся во входных (исходных) данных, но обычно некоторое количество информации теряется за счет приближений, вводимых вычислительной процедурой. С информационной точки зрения вычисление не добавляет никакой новой информации, а лишь повторяет уже имеющуюся информацию на другом языке и, возможно, с некоторыми потерями. Но вычисление, как и декодирование, и перевод, конечно, увеличивают практическую ценность информации и ее значение для получателя. Вопрос об информации, содержащейся в программе вычисления, является другим аспектом проблемы, до сих пор не исследованным, хотя он и был поставлен на миланской конференции.

Весьма важным для практического применения является вопрос о циркуляции информации в каналах с обратной связью, или без нее. Положительная обратная связь обладает свойствами, существенно отличными от отрицательной. По нашему определению информация есть положительная величина, которая никогда не может стать отрицательной. Следовательно, в каналах с обратной связью мы снова имеем дело не с самой информацией, а с ее ценностью. Вопрос о том, как определить ценность в такого рода задачах, остается открытым. По-видимому, единственное заключение, которое мы пока что можем сделать (и которое не имеет большого практического значения), состоит в следующем: рассмотрим рис. 20.1, на котором входной канал C_1 дает в среднем C_1 дв. ед. в секунду (пропускная способность C_1). Эта информация проходит по C_2, и некоторая доля α $(0<\alpha<1)$

возвращается обратно по цепи обратной связи. Канал C_2 должен быть в состоянии передать

$$C_1(1+\alpha+\alpha^2+\ldots)=\frac{C_1}{1-\alpha} \text{ дв. ед. в сек.},$$

так что его пропускная способность есть

$$C_2=\frac{C_1}{1-\alpha},$$

и выходной канал имеет пропускную способность

$$C_3=C_2-\alpha C_2=C_1.$$

Это, однако, не затрагивает хорошо известных проблем цепей с обратной связью. К задаче такого рода могли бы быть применены методы Линвилла и Зальцера, рассмотренные в главе 19.

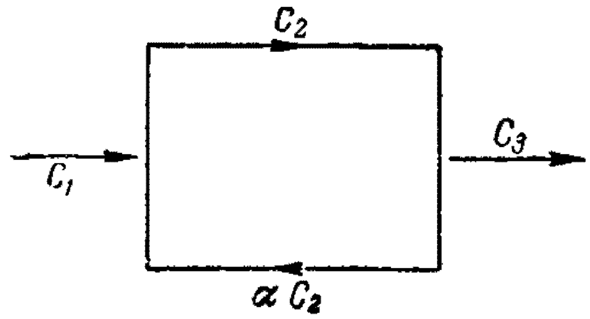

Рис. 20.1. Схематическое представление пропускных способностей в цепи с обратной связью.

Большое число задач показывает важность обсуждения практической ценности информации для получателя. В то время как мера информации, как она определена в настоящее время, является *абсолютной*, ценность информации, конечно, *относительна* для некоторого получателя. Предполагается, что эти вопросы можно обсуждать с эмпирической точки зрения, пытаясь в каждом случае найти разумное определение ценности, а затем сравнивая различные задачи, с целью обнаружить общие свойства различных определений и выяснить, может ли быть развита теория, применимая к различным частным случаям. Такое обсуждение, вероятно, сделало бы теорию информации значительно более ценной для практических применений и могло бы открыть новые важные направления исследований.

Во всяком случае, одно непосредственно очевидно: любой критерий *ценности* ведет к оценке получаемой информации. Это равносильно отбору информации, имеющей определенную меру качества. Некоторая часть информации может быть признана важной и сохранена со всей ее ценностью, тогда как другая часть может считаться не имеющей ценности и быть отброшена. Эта задача очень сходна с процессом фильтрации. Фильтр является хорошо известным в общей теории элементом и обладает двумя важными свойствами: он необратим и он уменьшает общее количество информации. Мы можем установить следующее положение: относительная ценность информации для некоторого данного получателя меньше или равна абсолютной информации. Это уже является отправной точкой, действительной в общем случае, и задача состоит в установлении общего определения различных методов фильтрации, применяемых для оценки относительной ценности информации.

6. Проблемы семантической информации

Развитая в этой книге теория основана исключительно на статистических данных. Согласно нашим определениям, предложения, встречающиеся очень редко, несут большое количество информации. Проблема ценности, кратко обсужденная в предыдущем разделе, не рассматривается в нашей теории. Мы полностью игнорировали также и другую проблему: проблему смысла. Вопрос о том, имеет ли сообщение смысл для отправителя или получателя, мы считали совершенно не относящимся к делу. Некоторые исследования семантических проблем выполнили недавно Вилль[1]) во Франции и Карнап и Бар-Хиллель[2]) в США.

Работа Карнапа и Бар-Хиллеля основана на методах и символике символической логики. Мы дадим здесь краткий очерк их исследования, которое ограничено (отчасти ради простоты, а отчасти вследствие ограниченности методов индуктивной логики, на которой основывается большая часть рассуждений) предложениями на весьма простом языке. Они изучают несколько возможных определений меры информации, содержащейся в данном предложении, и анализируют преимущества и недостатки каждого из них. Их анализ исключает из рассмотрения вопросы оценки и интерпретации: они

---

[1]) J. Ville, 18-th Intern. Congr. of Philosophy of Sci., Paris (1949); опубл. в Actualités Sci. et Ind. **1145**, 101—114, Hermann, Paris (1951).

[2]) R. Carnap, Y. Bar-Hillel, An Outline of a Theory of Semantic Information, MIT Research Lab. Electronics, Tech. Rept. № 247 (1952); Brit. J. Phil Sci. **4**, 147—157 (1953).

игнорируют, будет ли сообщение истолковано получателем в соответствии с намерениями отправителя.

По всем приведенным определениям такое предложение, как $17 \times 19 = 323$, не имеет информации. Основание для этого состоит в предположении, что получатель сообщения располагает весьма возможными логическими выводами из структуры данной языковой системы, так что такого рода предложение является избыточным. Считается, что информацию содержат только такие предложения, хотя бы часть которых не определяется структурой языка.

Карнап и Бар-Хиллель берут за основу весьма простой язык, состоящий из конечного числа n существительных и π прилагательных (называемых в их работе соответственно *индивидуалами* и *предикатами*) и одного-единственного глагола ***имеет*** или ***есть***. Так, если a есть существительное, а P — прилагательное, то Pa читается как «a имеет свойство P» или, проще, «a есть P».

В языке имеются следующие связи, которые могут быть использованы для построения более длинных предложений:

$\sim$ не: отрицание; $(\sim Pa)$ означает: «a не есть P»;

$\vee$ или: дизъюнкция; $(Pa \vee Qb)$ означает «a есть P или b есть Q»;

$\cdot$ и: конъюнкция; $(Pa \,.\, Qb)$ означает: «a есть P и b есть Q»;

$\supset$ если... то: импликация;

$\equiv$ если и только если: равнозначность[1]).

Выбор π прилагательных в языке несколько ограничен. Прилагательные никак не должны перекрываться по смыслу; кроме того, одно прилагательное не должно ни в какой мере обусловливать применение или неприменение другого прилагательного. Словарь языка дополняется введением для каждого из π прилагательных P противоположного ему, или его отрицания, которое может быть обозначено $\sim P$, или новым символом, или зачастую и тем и другим.

Предложения в этом языке разделяются на логически ложные (противоречивые, как, например, $Pa \cdot \sim Pa$), фактуальные (логически неопределенные, как, например, Pa) и логически истинные (тавтологические, как, например, $Pa \vee \sim Pa$).

---

[1]) В нашей литературе встречаются и другие обозначения, как-то: не: $\neg$; и: $\wedge$; или: $\vee$; если и только если: $\sim$. (*Прим. перев.*)

Один очевидный метод определения информационного содержания предложения в этом языке состоит в следующем: имеется конечное число предложений, которые можно построить, и между этими предложениями существуют определенные логические соотношения. Информация некоторого частного предложения i может быть определена как подходящим образом выбранная функция числа предложений, обусловленных i. Существуют по меньшей мере два пути к построению этого определения; мы отсылаем читателя за подробностями к книге Карнапа[1]), так как эти определения находятся с нашей точки зрения в весьма отдаленной связи.

Работа MIT[2]) содержит определения, более тесно связанные с теми, которыми мы пользовались. Эти определения основаны на особом виде фактуального предложения, а именно, на *описании состояния*, обозначаемом через Z. Описание состояния есть предложение, состоящее из конъюнкции πn простых (*атомных*) предложений и содержащее каждое из n существительных в соединении с каждым из прилагательных или с его отрицанием (но не с тем и другим одновременно). Имеется πn возможных соединений существительных с прилагательными, и так как каждое соединение может осуществляться двояким образом (т. е. либо с прилагательным, либо с его отрицанием), то всего имеется $2^{\pi n}$ возможных описаний состояния. Описание состояния полностью определяет возможное состояние рассматриваемой системы, и в определения информации должны войти априорные вероятности описаний состояния. В структуре языка не содержится априори никакого основания для присвоения определенных вероятностей описаниям состояния, хотя некоторые свойства приемлемой совокупности вероятностей до некоторой степени ограничивают выбор. В задачах, обсуждавшихся в этой книге, свойства физической системы, которую мы желаем описать, указывают выбор вероятностей. Так, например, при обсуждении температур мы можем обозначить те, которые выше 40° F, как «высокие», а те, которые ниже 40° F, — как «низкие».

---

[1]) Carnap, Logical Foundations of Probability, U. of. Chicago Press, Chicago, Ill., 1950.

[2]) R. Carnap, Y. Bar-Hillel, An Outline of a Theory of Semantic Information, MIT Research Lab. Electronics, Tech. Rept. № 247 (1952); Brit. J. Phil. Sci. 4, 147—157 (1953).

Очевидно, что вероятности «высоких» и «низких» температур на Северном полюсе и на экваторе будут различны.

Для предложения i (которое, вообще, не является описанием состояния) область (range) i, обозначаемая $R(i)$, есть совокупность описаний состояния, в которых i сохраняет силу. На этом этапе могут быть введены определения информации, основанные на теории совокупности точек. Нас, однако, больше интересуют количественные определения, т. е. определения, дающие числовую меру информации. С этой целью мы вводим функцию меры, обладающую следующими свойствами:

A. Для каждого описания состояния Z существует мера $m(Z)$ такая, что

$$0 \leqslant m(Z) \leqslant 1.$$

Эта величина может толковаться как априорная вероятность описания состояния Z.

B. Сумма $m(Z)$, взятая по всем Z, равна единице.

C. Для всякого ложного предложения f, $m(f)=0$.

D. Для всякого неложного предложения i мы определяем $m(i)$ как сумму $m(Z)$ по всем Z, содержащимся в области $R(i)$.

E. Все существительные трактуются наравне, т. е. $m(i)$ инвариантно при перестановке существительных, встречающихся в предложении i.

F. Все прилагательные трактуются наравне, т. е. $m(i)$ инвариантно при перестановке прилагательных, встречающихся в i.

G. Любое прилагательное может быть заменено его отрицанием без изменения значения $m(i)$.

H. Если два предложения i и j не имеют общих прилагательных, то они считаются индуктивно-независимыми и

$$m(i \cdot j) = m(i) \times m(j).$$

I. Число существительных, не входящих в предложение i, не должно влиять на значение $m(i)$.

Такая функция меры называется ***собственной m-функцией***, и Карнап полагает, что ее можно рассматривать как вероятность, которую он называет: ***абсолютная логическая вероятность***. Условия E, F и G вводят связи между вероятностями различных описаний состояния, но не определяют их единственным образом. Свойства $m(i)$ могут быть использованы для вывода многих теорем и соотношений, необходимых для детальной разработки теории. Из числа этих соотношений мы воспользуемся следующим:

$$m(i)+m(\sim i)=1.$$

Так как $R(i)$ и $R(\sim i)$ не содержат общих описаний состояния (описание состояния не может быть связано как с i, так и с его отрицанием) и каждое описание состояния находится либо в $R(i)$, либо в $R(\sim i)$ (каждое описание состояния связано либо с i, либо с его отрицанием), то

$$m(i)+m(\sim i)=\sum_{\text{по всем } Z} m(Z)=1.$$

Простейшее определение информации через функцию меры названо Карнапом и Бар-Хиллелем *мерой содержания* (cont) и выражается как

$$\operatorname{cont}(i)=m(\sim i)=1-m(i).$$

Мера содержания обладает многими интересными свойствами и находится в качественном согласии с определениями, использованными в этой книге, в следующем смысле. Если $m(i)$ мало, то $\operatorname{cont}(i)$ относительно велико. Но малое $m(i)$ означает, что принимаются сравнительно маловероятные описания состояния, так что $\operatorname{cont}(i)$ есть мера редкости. Другое определение, *мера информации*, ближе к нашим определениям. Она выражается как

$$\inf(i)=-\log_2 m(i)=\log_2\left(\frac{1}{1-\operatorname{cont}(i)}\right).$$

Так же как и для $\operatorname{cont}(i)$, имеется качественное согласие этого определения с нашими. Далее, можно показать, что условие H, которому удовлетворяет функция $m(i)$, приводит к аддитивности меры информации для конъюнкции индуктивно-независимых предложений. Рассмотрение нижеследующего специального случая сделает более ясной тесную связь между обоими определениями. Выберем функцию меры вида

$$m(Z)=\frac{1}{N}$$

для всех Z, где N — общее число описаний состояния. Мы предполагаем, таким образом, что все описания состояния равновероятны. Это предположение совместимо с условиями, наложенными на функцию меры. Если обозначать через N_r число описаний состояния в области $R(i)$, то из определения имеем:

$$\inf(i) = -\log_2 \frac{N_r}{N} = \log_2 \frac{N}{N_r}.$$

Это равенство совпадает с тем, которое применялось на протяжении книги, если мы отождествим описания состояния с равновероятными исходами в некоторой данной ситуации. Вначале имеются N возможных исходов (N описаний состояния). Знание предложения i уменьшает возможное число до N_r, так что информация, даваемая предложением i, согласно нашим прежним определениям, равна

$$I = \log_2 \frac{N}{N_r} \text{ дв. ед.}$$

Для описания состояния Z $R(Z)$ содержит только Z, и

$$I = \log_2 N \text{ дв. ед.}$$

Карнап и Бар-Хиллель дают дальнейшую детальную разработку этих семантических определений и обсуждают приложения к различным вопросам заключения и дедукции.

В этом кратком очерке мы старались лишь показать связь между этими проблемами и теми, которым посвящена эта книга.

ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ