

УСПЕХИ ФИЗИЧЕСКИХ НАУК

536 : 53.08

**ТЕРМОДИНАМИЧЕСКИЕ МОДЕЛИ
ИНФОРМАЦИОННЫХ ПРОЦЕССОВ***Р. П. Поплавский**Светлой памяти дорогих родителей посвящая*

СОДЕРЖАНИЕ

1. Основная проблема термодинамики информационных процессов	465
а) Введение (465). б) История вопроса (466). в) Процесс измерения и его информационные характеристики (467).	
2. Термодинамические пределы точности физического измерения	472
а) Дефект энтропии при измерении (472). б) Нижняя граница необратимости процесса физического измерения (475). в) Энергетическая цена точности и количества информации (481).	
3. Термодинамические характеристики процесса передачи информации . . .	484
а) Энергетический смысл кодирования (484). б) Натуральное представление числа (485). в) Позиционные способы кодирования (489).	
4. Термодинамические модели процесса обработки информации	491
а) Критерии сложности обработки информации (491). б) Энергетический критерий сложности процесса обработки информации (492). в) Об эффективных методах вычислений (496).	
5. Заключение	498
Цитированная литература	499

**1. ОСНОВНАЯ ПРОБЛЕМА ТЕРМОДИНАМИКИ
ИНФОРМАЦИОННЫХ ПРОЦЕССОВ**

а) В в е д е н и е. Основная проблема термодинамики информационных процессов состоит в установлении связей (предельных соотношений) между важнейшими термодинамическими характеристиками (энергия, энтропия) и информационными (точность, количество информации). Применительно к первичному информационному процессу — измерению — можно выделить три стороны проблемы:

— связано ли получение информации с уменьшением статистической энтропии системы и какова эта связь;

— каково компенсирующее увеличение энтропии в системе за счет диссипации энергии;

— за счет чего удастся после измерения обеспечить дополнительное уменьшение энтропии в системе посредством управляющего воздействия.

Несмотря на очевидное значение этой проблемы, литература, посвященная ей, немногочисленна и весьма противоречива. Поэтому круг обсуждаемых в обзоре вопросов ограничен именно этой проблемой и притом применительно к классическим (не квантовым) объектам. Последнее ограничение связано с тем, что, как будет видно из дальнейшего, для решения сформулированной выше основной проблемы термодинамики информационных процессов необходимо и в первом приближении доста-

точно ограничиться исследованием классической части измерительного прибора.

С другой стороны, сделана попытка с единой точки зрения рассмотреть все основные информационные процессы: получение информации (измерение), передачу информации и обработку информации (вычислительный процесс).

б) И с т о р и я в о п р о с а. После введения в термодинамику Клаузиусом понятия энтропии как функции состояния Больцман придал энтропии статистический (информационный) смысл. За несколько лет до этого (см., например, ¹) Максвелл сформулировал парадокс с демоном, сортирующим молекулы по скоростям и «нарушающим» вследствие этого второе начало термодинамики. Именно при исследовании парадокса с «демоном Максвелла» Сцилард в 1929 г. впервые ² указал на физическую связь информации и энтропии.

Математическое определение количества информации, введенное в 1928 г. Хартли ³, а затем обобщенное в 1948 г. в работах Винера ⁴ и, в особенности, Шеннона ⁵, с точностью до постоянного множителя совпадает со статистическим определением энтропии по Больцману. Это формальное обстоятельство послужило основанием для введения Шенноном понятия «энтропии источника сообщений». Физический смысл совпадения формул для количества информации и энтропии Шенноном не обсуждался. На полезность исследования содержательного смысла аналогии между этими понятиями указывали Винер ⁶ и Колмогоров ⁷.

Однако теория информации, начиная с основополагающих работ Шеннона ⁵, развивалась преимущественно как математическая дисциплина ^{31,32} применительно к процессу передачи информации. Колмогоровым ⁷ дано строгое изложение основных проблем этой теории и рассмотрены границы ее применимости.

Основные работы 1944—1951 гг., анализирующие связь между термодинамикой и информацией, относились к исследованию проблемы «демона Максвелла» с различных точек зрения и развивали упомянутую основополагающую работу Сциларда ². Это работы Габора, Демерса, Джекобсона и, наконец, Бриллюэна, подробно изложенные в книге Бриллюэна ⁸. Было показано, что основное в парадоксе Максвелла не то, что «демон» — мыслящее существо (его можно заменить автоматом), а то, что для целенаправленного управления, приводящего к уменьшению энтропии (упорядочению системы либо по температуре, либо по давлению — «демон давления» ⁸), используется информация. Получение информации неизбежно связано с рассеянием энергии, приводящим к росту энтропии, превышающему ее убыль вследствие упорядочения молекул. Представляется, что приведенные в книге Бриллюэна ⁸ результаты дают исчерпывающее разрешение парадокса с «демоном Максвелла». Трудно согласиться с Шамбадалем ⁹, который считает «изгнание демона» кажущимся. Его доводы основаны на различной размерности энтропии и количества информации, а также на том, что в ряде упомянутых работ рассматривался одномолекулярный газ. Известно, что безразмерной энтропией (при измерении температуры в энергетических единицах) часто пользуются *) в статистической физике ¹⁰, а для одномолекулярного газа в термостате термодинамика еще справедлива (см. ¹¹, стр. 294). Однако

*) Всюду в настоящей работе также используется безразмерная энтропия и применяется ее обозначение H . Именно такое обозначение принято в теории информации, начиная с работ Шеннона. Впрочем, в физической литературе это обозначение встречается существенно раньше: у Больцмана и Гиббса.

можно согласиться⁹ с тем, что парадокс Максвелла разрешен лишь качественно: вопрос о количественных связях между информационными и энергетическими характеристиками в упомянутых работах нельзя считать решенным.

Поэтому естественна высказанная в 1955 г. Харкевичем¹² неудовлетворенность в связи с отсутствием в теории информации зависимостей типа законов сохранения. Этим же можно объяснить известные высказывания в 1950 г. Габора¹³ и в 1951 г. фон Неймана¹⁴ о необходимости развития физического (термодинамического) направления в теории информации.

Применение термодинамики к теории информации нашло последовательное развитие лишь в 1956 г. в известной книге Бриллюэна⁸. Здесь поставлен и рассмотрен целый ряд вопросов, далеко выходящих за пределы проблемы «демона Максвелла». В частности, подробно проанализирована связь между энергетическими и информационными характеристиками на целом ряде мысленных экспериментов по измерению. Сформулирован в общем виде «негэнтропийный принцип информации», который устанавливает нижнюю границу энергетической цены измерения. Однако ни из рассмотренных примеров, ни из общих соображений не выяснено, достижима ли эта нижняя граница и, если нет, то нельзя ли оценку сделать более точной. Таким образом, и на основании книги Бриллюэна⁸ предельные соотношения между информационными и энергетическими характеристиками нельзя считать окончательно установленными. Поэтому значение этого вопроса подчеркивается и в литературе шестидесятых годов (см., например,¹⁵).

В следующей книге Л. Бриллюэн¹⁶ развивает свои концепции в несколько ином направлении: обсуждаются проблемы информационного содержания физических законов и теорий, значение ошибок измерений в познавательном процессе и пр.

В последние годы появились работы, использующие методы теории информации для анализа технических информационных систем^{17, 18, 37} и рассматривающие формальные аналогии между термодинамическими и информационными понятиями^{19, 20}. Работы этих направлений, естественно, не ставили задачу дать ответ на вопрос о физических связях между энергетическими и информационными характеристиками.

Дальнейшее развитие физической теории информации пошло, в основном, по пути исследования ограничений, накладываемых квантовой природой материи на процесс передачи информации. Эта сторона физики информационных процессов имеет большое теоретическое и прикладное значение и представлена обширной литературой.

Однако лишь немногие из работ этого направления имеют отношение к поставленному выше вопросу. Результаты этих работ, в частности^{21, 22}, а также непосредственно относящихся к теме обзора работ, в частности²³⁻²⁵, будут подробно обсуждаться ниже.

в) П р о ц е с с и з м е р е н и я и е г о } и н ф о р м а ц и о н н ы е
х а р а к т е р и с т и к и. Для установления предельных физических соотношений между энергетическими и информационными характеристиками представляется естественным начать с анализа первичного информационного процесса — физического измерения.

Под измерением обычно понимается любое однозначное преобразование измеряемой физической величины в некоторую другую физическую величину, называемую регистрирующим параметром. Однако для термодинамического анализа измерения это определение требуется сузить: необходимо представить измерение как термодинамический процесс перехода

системы из одного равновесного состояния в другое. Будем предполагать, что:

1) измеряемая физическая величина является внутренним параметром исследуемой системы (ИС);

2) в качестве регистрирующего параметра выбрана физическая величина, характеризующая состояние измерительного прибора (ИП) и потому допускающая ее использование в качестве управляющего параметра;

3) в процессе взаимодействия ИС и ИП (т. е. в процессе измерения) устанавливается новое стационарное значение регистрирующего параметра, однозначно связанное с измеряемой величиной.

Согласно этому определению, для выполнения измерения только тогда достаточно ограничиться одним этапом (и соответственно, одним актом элементарного взаимодействия ИС с ИП), когда:

а) измеряемая величина удовлетворяет условию 1);

б) она может быть непосредственно преобразована в регистрирующий параметр, удовлетворяющий 2), 3).

В общем случае это не так, и для измерения может потребоваться длинная цепь последовательных преобразований. Представляется, что для выяснения общих закономерностей достаточно рассмотреть цепочку из четырех этапов:

$$l \rightarrow \lambda \rightarrow F \rightarrow y \rightarrow x; \quad (1.1)$$

здесь l — измеряемый параметр, который может быть как классическим, так и квантовым, — далеко не всегда удовлетворяет условию 1) и поэтому должен быть преобразован к физической величине, удовлетворяющей условию 1). На пути этого преобразования, реализуемого датчиком, часто (явно или неявно) присутствует промежуточный этап, на котором l оставляет «след» λ в некотором закодированном виде. Поясним сказанное на примерах возможных пар $l - \lambda$:

частота обнаруживаемого сигнала — положение (или номер) перестраиваемого резонатора;

дальность до объекта — временной интервал, отмеченный зондирующим и отраженным импульсом;

разность фаз или скорости двух электромагнитных сигналов — интерференционная картина в месте их встречи;

импульс микро- или макрочастиц — мгновенное значение отклонения пробного тела и т. д.

Видно, что «след» λ представляет собой позиционный код измеряемой величины и явно присутствует при таких измерениях, когда предполагается наличие наблюдателя. Поэтому часто λ принимается в качестве окончательного результата измерения, как это следует из приведенных примеров. Если же наличие наблюдателя не предполагается, то этап $l \rightarrow \lambda$ может присутствовать неявно, так как одновременно с перестройкой прибора по параметру l в датчике происходит преобразование его к F . Параметр F имеет характер обобщенной силы (в частности, электродвижущей), действующей на ИП. Таким образом, ИС (вместе с датчиком) имеет некоторый запас свободной энергии, обеспечивающей взаимодействие ИС и ИП.

Действие обобщенной силы F на ИП, приводящее к соответствующему изменению координаты x , не всегда может реализоваться непосредственно. Например, если F — измеряемое напряжение в исследуемой электрической цепи, а x — угол поворота стрелки с пружиной, то y — установившееся в процессе измерения значение тока, или заряда, или напряжения в цепи прибора, непосредственно действующее на стрелку. Параметр y и устройство, реализующее этап $F \rightarrow y$ (1.1), будем называть согласо-

щими. Этот этап введен в рассмотрение для того, чтобы оценить его влияние на термодинамические характеристики процесса измерения. Из дальнейшего будет видно, что предельные соотношения определяются непосредственным преобразованием $F \rightarrow x$, а любой дополнительный этап только ухудшает оценки.

Из сказанного видно, что для термодинамических характеристик измерения по 1)–3) существенна только классическая часть ИП. При этом введенные выше определения конкретизируются следующим образом.

Всякое физическое измерение связано с взаимодействием двух систем — исследуемой и измерительного прибора — таким, что происходит обмен энергией между ИС и ИП, вследствие чего изменяется значение параметра x регистрирующего устройства измерительного прибора. Диссипативные процессы, связанные с энергетическим обменом между ИС и ИП, являются единственной возможной причиной необратимости физического измерения.

Предполагается, что ИС и ИП помещены в термостат с температурой T . Тепловые флуктуации параметров F, y, x являются единственной неустранимой причиной погрешности измерения. Измерение построено на том, что имеется однозначное соответствие между средними (по реализациям флуктуаций) значениями $\bar{F}, \bar{y}, \bar{x}$, и по значению \bar{x} судят об измеряемой величине F . Предполагается линейная связь между F, y, x .

Подобная модель измерения рассмотрена в ^{23, 26, 27} и будет подробно описана в гл. 2.

Переходим к определению основных информационных характеристик измерения.

Пусть истинное значение измеряемой скалярной величины l , измеренное значение l' , их разность

$$l - l' = \Delta l \quad (1.2)$$

— погрешность данного измерения.

Измерение в области, или оценка случайной величины ²⁸, характеризуется тем, что заранее известна область (отрезок) L длины l_m

$$l_m = l_{\max} - l_{\min}, \quad l \in L \Leftrightarrow l_{\min} \leq l \leq l_{\max} \quad (1.3)$$

и задано априорное распределение $P(l)$ в L .

Погрешность Δl данной реализации измерения, естественно, не может служить его критерием качества: для этого следует выбрать некоторую величину, усредненную по совместному распределению $P(l', l) = P(l)P(l'/l)$. Физическое измерение характеризуется обычно свойством однородности, т. е. тем, что $P(l'/l) = P(|l - l'|) = P(|\Delta l|)$ и не зависит от l .

Усреднение по всем флуктуациям (т. е. по $P(|\Delta l|)$) будем обозначать чертой сверху, а усреднение по априорному распределению — треугольными скобками $\langle \dots \rangle$.

Тогда средний квадрат флуктуаций равен

$$\overline{\langle \Delta l^2 \rangle} = \overline{\Delta l^2}. \quad (1.4)$$

Кроме того, измерение часто характеризуют ^{8, 29} интервалом Δl_p , имеющим смысл минимального разрешаемого интервала или цены деления. Следует подчеркнуть, что Δl_p , в отличие от $\sqrt{\overline{\Delta l^2}}$, является неоднозначной характеристикой, так как

$$P(|\Delta l| > \frac{\Delta l_p}{2}) = w > 0, \quad (1.5)$$

т. е. имеется ненулевая вероятность w ошибки (превышения интервала Δl_p). В связи с этим возникает вопрос о связи погрешности Δl_p с надежностью ²⁹ эксперимента.

Измерение по схеме (1.4) характеризуется линейным преобразованием

$$\frac{x}{x_m} = \frac{l - l_0}{l_m} \quad (1.6)$$

отрезка L (1.3) в отрезок X длины x_m . При этом, как известно ³⁰, распределения l и x принадлежат одному и тому же типу и различаются лишь параметрами расположения: масштабным l_m/x_m и центрирующим l_0 .

Будем в дальнейшем рассматривать лишь симметричные относительно $\langle l \rangle$ распределения $P(l)$ и, соответственно, случай симметричного измерения:

$$l_0 = \langle l \rangle, \quad -\frac{x_m}{2} \leq x \leq \frac{x_m}{2} \Leftrightarrow x \in X_+ \quad (1.7)$$

и асимметричного измерения:

$$l_0 = l_{\min}, \quad 0 \leq x \leq x_m \Leftrightarrow x \in X_- \quad (1.8)$$

Одновременно рассматривая оба случая (1.7) и (1.8), будем писать: $x \in X$, имея в виду либо X_+ , либо X_- . Преобразование (1.6) связано в общем случае с тем, что в датчике используется некоторая эталонная обобщенная сила $F_{\text{эт}}$, которая совместно с $F(l)$ дает

$$\begin{aligned} \langle F_{\text{эт}} + F(l) \rangle &= 0 \quad \text{при} \quad (1.7), \\ F_{\text{эт}} + F(l_{\min}) &= 0 \quad \text{при} \quad (1.8). \end{aligned} \quad (1.9)$$

Обратим внимание на то, что совершенствование экспериментов обычно связывается с увеличением их чувствительности, т. е. различимости малых разностей $l_1 - l_2$. Естественно поэтому, что в этих случаях основное внимание уделяется снижению абсолютной погрешности $\sqrt{\Delta l^2}$ (1.4). При этом, когда говорят о повышении точности, то имеется в виду уменьшение погрешности, хотя формально размерная величина $(1/\sqrt{\Delta l^2})$ вряд ли может быть названа абсолютной точностью.

В противоположность этому, относительная точность может быть формально определена как обратная относительной погрешности величина. Как будет видно из дальнейшего, именно эта безразмерная величина играет определяющую роль в термодинамике информационных процессов.

Определим относительную точность $1/\sigma_x$ однократного измерения

$$\frac{1}{\sigma_x} = \frac{x}{\sqrt{\Delta x^2}} = \frac{l - l_0}{\frac{x_m}{l_m} \sqrt{\Delta l^2}} \quad (1.10)$$

Эта величина имеет смысл либо как характеристика одной реализации измерения в области, либо при оценке постоянной величины ²⁸. В последнем случае, говоря о точности величины l , обычно предполагают, что в (1.10) и (1.8) $l_0 = l_{\min} = 0$. В первом же случае для характеристики всего процесса измерения в области определим среднюю относительную точность $1/\sigma$ как отношение априорной среднеквадратической погрешности к апостериорной:

$$\frac{1}{\sigma} = \sqrt{\frac{\langle (x - \langle x \rangle)^2 \rangle}{\langle \Delta x^2 \rangle}} \equiv \sqrt{\frac{\langle \delta x^2 \rangle}{\langle \Delta x^2 \rangle}} \quad (1.11)$$

Часто вводят относительную (или, согласно ³⁷, приведенную) погрешность $\sqrt{\Delta \xi^2}$ через безразмерную переменную

$$\xi = \frac{x}{x_m} = \frac{l-l_0}{l_m}, \quad (1.12)$$

и тогда

$$\frac{1}{\sigma^2} = \frac{\langle \delta \xi^2 \rangle}{\langle \Delta \xi^2 \rangle} \equiv \frac{\langle \delta \xi^2 \rangle}{\sigma_0^2}. \quad (1.13)$$

Введем еще относительный (приведенный) разрешаемый интервал

$$\varepsilon = \Delta \xi_p = \frac{\Delta x_p}{x_m}, \quad P\left(\left|\frac{x-x'}{x_m}\right| > \frac{\varepsilon}{2}\right) = w > 0. \quad (1.14)$$

Будем называть $1/\varepsilon$ разрешающей способностью, $1/w$ — надежностью (следуя работе ⁸), $1/\sigma$ — точностью измерения.

Остается определить среднее количество информации $I(x, x')$, характеризующее процесс измерения в области.

Имеем, согласно работам ^{7, 31},

$$I(x, x') = H(x) - MH(x/x'), \quad (1.15)$$

где $H(x)$ — начальная энтропия, выражающая неопределенность априорного распределения $P(x)$, $H(x/x')$ — условная энтропия, характеризующая апостериорную неопределенность x при измеренном значении x' (т. е. неопределенность распределения $P(x/x')$), $MH(x/x')$ — усредненное по $P(x')$ значение условной энтропии $H(x/x')$.

Учитывая, в соответствии с ³², что максимальной энтропией на отрезке обладает равномерное распределение, а максимальную энтропию при заданной дисперсии имеет распределение Гаусса, будем в дальнейшем полагать для соответствующих плотностей:

$$p(x) = \text{const} = \frac{1}{x_m}, \quad x \in X, \quad (1.16)$$

$$p(x/x') = p(|x - x'|) = \frac{1}{\sqrt{2\pi}\sigma_0} e^{-\Delta x^2/2\sigma_0^2} \quad (-\infty \leq x - x' \leq \infty). \quad (1.17)$$

Здесь, в соответствии с формулами (1.13), (1.15), (1.16), (1.7), (1.8), имеем

$$\langle \Delta x^2 \rangle = \frac{1}{12} x_m^2, \quad \sigma_0^2 = \frac{1}{12} \sigma^2. \quad (1.18)$$

Если при $\sigma_0^2 \ll 1$ (точные измерения) пренебречь отличием $p(x')$ от $p(x)$ и вычислить (1.15) по (1.16), (1.17), получим:

$$I(x, x') = \frac{1}{2} \ln \frac{1}{\sigma^2} + \frac{1}{2} \ln \left(\frac{6}{\pi e} \right) = \frac{1}{2} \ln \frac{1}{\sigma^2} - 0,17 \approx \ln \frac{1}{\sigma}, \quad \sigma^2 \ll 1. \quad (1.19)$$

Таким образом, при высокой точности $1/\sigma$ (формула (1.11)) измерения количество информации определяется ее логарифмом *) ($\ln 1/\sigma$) достаточно точно (было взято отношение дисперсий столь разных распределений, как равномерного и нормального).

*) Для частного случая, когда и априорное, и апостериорное распределения — нормальные, зависимости типа (1.19) и последующей — (1.22) приведены в работе ³⁷. Следует подчеркнуть, что до последнего времени встречаются определения точности по Бриллюэну ^{8, 16} (в наших обозначениях (1.14)) как $1/\varepsilon$, а количества информации $I = \ln(1/\varepsilon)$. Понятно, что вследствие условия $w > 0$ (1.14) такое определение неоднозначно и поэтому неверно. Ниже (гл. 3) этот вопрос будет рассмотрен подробнее в связи с энергетическими оценками.

Если вычислить $p(x')$ на основании выражений (1.16), (1.17), получим, что внутри интервала $P(x')$ не отличается от $P(x)$, а лишь на краях отрезка $(x_b \equiv x_{\max}, x_{\min})$ при $|x - x_b| \approx \sigma_0 x_m$ различие $p(x')$ от $p(x)$ становится заметным. При вычислении $I(x, x')$ это эквивалентно эффективному уменьшению

$$\langle \Delta \xi^2 \rangle \equiv \sigma_0^2 \approx \overline{\Delta \xi^2} (1 - \sigma_0^2), \quad (1.20)$$

что в формуле (1.19) несущественно (при $\langle \delta \xi^2 \rangle \gg \overline{\Delta \xi^2}$, $\sigma^2 \ll 1$), однако должно быть учтено *) при вычислении $I(x, x')$ для грубых измерений. В этом случае, вспоминая свойство симметрии $I(x, x')$:

$$I(x, x') = I(x', x) = H(x') - MH(x'/x), \quad (1.21)$$

отметим, что условная энтропия строго определяется нормальным распределением, так как усреднение в (1.21) по $P(x)$, а $H(x') > H(x)$ вследствие упомянутого краевого эффекта.

Расчеты показывают, что краевой эффект может быть достаточно точно учтен, если представить себе, что эффективное увеличение $H(x')$ эквивалентно тому, что начальная дисперсия равна сумме априорной $\langle \delta x^2 \rangle$ и апостериорной $\overline{\Delta x^2}$. Тогда из предыдущей формулы имеем равенство

$$I(x, x') \approx \frac{1}{2} \ln \left(\frac{\langle \delta x^2 \rangle + \overline{\Delta x^2}}{\overline{\Delta x^2}} \right) = \frac{1}{2} \ln \left(\frac{1}{\sigma^2} + 1 \right), \quad (1.22)$$

которое, в отличие от (1.19), справедливо и при $\sigma^2 \gg 1$.

Видно, что для очень грубых измерений, когда $1/\sigma^2 \ll 1$,

$$I(x, x') \approx \frac{1}{2\sigma^2} \quad \text{при } \frac{1}{\sigma^2} \rightarrow 0. \quad (1.23)$$

Приведем в заключение выражения для $\langle \delta x^2 \rangle$ (1.11), (1.13), которые в случае равномерного распределения (1.16) упрощаются. Для симметричного (1.7) измерения

$$\langle \delta x^2 \rangle = \langle x^2 \rangle = \frac{1}{12} x_m^2, \quad (1.24)$$

для асимметричного (1.8)

$$\langle \delta x^2 \rangle = \frac{1}{4} \langle x^2 \rangle = \frac{1}{12} x_m^2. \quad (1.25)$$

2. ТЕРМОДИНАМИЧЕСКИЕ ПРЕДЕЛЫ ТОЧНОСТИ ФИЗИЧЕСКОГО ИЗМЕРЕНИЯ

а) Дефект энтропии при измерении. Приступая к анализу характеристик процесса измерения, выделим главные вопросы сформулированной выше основной проблемы термодинамики информационных процессов (см. раздел а) гл. 1).

*) Можно (1.20) получить формально, если и при определении погрешности $p(x, x')$ учесть априорные данные:

$$p(x, x') = \begin{cases} |x - x'| & \text{при } x' \in X, \\ x_{\max} - x & \text{при } x' > x_{\max}, \\ x - x_{\min} & \text{при } x' < x_{\min}. \end{cases}$$

В этом случае, естественно, равенства типа (1.4) становятся неверными.

I) Отражает ли совпадение формул для количества информации I и изменения энтропии ΔH физическое содержание процесса или же это просто формальная аналогия? Точнее: связано ли получение информации с уменьшением статистической энтропии (дефектом $\Delta H_1 \leq 0$ энтропии ²¹) системы и какова эта связь?

II) Каково компенсирующее увеличение энтропии $\Delta H_2 \geq 0$ в системе, т. е. какова минимальная диссипация энергии в процессе измерения, как она связана с информационными характеристиками, т. е. какова энергетическая цена информации? Иначе: обратим или необратим процесс классического измерения и какова нижняя граница необратимости его?

III) За счет чего удается при получении информации произвести некоторое упорядочение в системе (типа того, которое производит «демон Максвелла», а в общем случае, реализовать управление системой за счет информации), т. е. обеспечить дополнительное уменьшение энтропии ΔH_3 ($\Delta H_3 \leq 0$)? В этом разделе рассмотрим ответы на вопрос 1).

До самого последнего времени встречаются высказывания (см., например, ^{9, 33}), в которых авторы настаивают на существенно различной природе физической и информационной (в соответствии с ³³, «энергетической» H_z и «кибернетической» H_k) энтропии. В работе ⁹, как уже упоминалось выше, основные доводы основаны на различии в их размерности. В работе ³³, кроме того, подчеркивается, что H_z определена только для равновесного состояния замкнутой термодинамической системы, тогда как H_k — для стационарного случайного процесса выдачи сообщений источникам информации и; таким образом, может быть отнесена к любой математической модели источника сообщений, не связанной с состоянием физической системы. С этим, конечно, нельзя не согласиться, однако понятно, что сопоставление H_k и H_z (вернее, I и ΔH_1 по (I)) имеет смысл лишь при анализе изменения состояния реальной физической системы в процессе получения информации *) (аналогично тому, как это обсуждается в работе ³³ применительно к процессу управления для других составляющих: ΔH_2 (II) и ΔH_3 (III)).

Большинство авторов отождествляют физическую и информационную энтропию, подобно Бриллюэну ^{8, 16}, ограничиваясь эвристическими соображениями (см., например, ³⁴⁻³⁶), которые не могут служить доказательством этой точки зрения. «Положительная» (согласно ²¹) сторона сформулированного Бриллюэном ^{8, 16} неэнтропийного принципа информации как раз и состоит в утверждении об эквивалентности получения информации (связанной *) уменьшению физической энтропии системы.

При этом просто предполагается, что после получения информации в системе уменьшается число возможных микросостояний в $1/\epsilon$ (см. (1.14)) раз и постулируется, что $I = |\Delta H_1| = \ln(1/\epsilon)$. Выше уже упоминалось (см. сноску на стр. 471) о том, что определения относительной точности и количества информации в работе ⁸ нестрогие. Здесь следует подчеркнуть, однако, что именно Бриллюэн впервые обратил внимание на важность относительной (а не абсолютной) точности для установления количественных связей между термодинамическими и информационными характеристиками.

В работе ²¹ вопрос I) исследован более основательно. Авторы рассматривают процесс передачи информации и связывают ее получение с отклонением системы от состояния равновесия. Предполагается, что каждый i -й сигнал переводит систему в i -е макросостояние. Вводится

*) Бриллюэн ⁸ вводит поэтому различие между свободной и связанной информацией. Последняя возникает тогда, когда возможные случаи (события) могут быть представлены как микросостояния физической системы.

понятие дефекта энтропии ΔH_1 , который показывает, насколько состояние системы при действии определенного сигнала в среднем дальше от состояния термодинамического равновесия, чем состояние, возникающее при действии случайно выбранного сигнала. Утверждается, что количество информации I , полученное физической системой, равно ее дефекту энтропии $|\Delta H_1|$, однако здесь же отмечается, что этот принцип должен рассматриваться как эвристический, так как он строго не доказан. Далее утверждается ²¹, что всегда справедливо неравенство $I \leq |\Delta H_1|$ и что это является следствием второго начала термодинамики. Последняя часть этого утверждения ошибочна, так как второе начало может быть применено лишь при оценках компенсирующего возрастания энтропии ΔH_2 (II) в другой части системы, т. е. при использовании «отрицательной» (согласно ²¹) стороны негэнтропийного принципа Бриллюэна, согласно которому $\Delta H_2 \geq |\Delta H_1|$. Это утверждение Бриллюэна может считаться следствием второго начала термодинамики.

Представляется, что ответ на вопрос I) следует из анализа равновесной термодинамической модели, рассмотренной в работе ²³. Согласно ²³, ²⁶, ²⁷ и данному в разделе в) гл. 1 определению, процесс измерения сводится к взаимодействию двух тел, первое из которых Y включает в себя исследуемую систему (вместе с согласующим устройством), второе X является регистрирующим устройством (для конкретности — механическим) измерительного прибора. Предполагается, что до измерения ($t < 0$) тела Y и X не взаимодействуют и находятся в начальном состоянии равновесия: $y(t \leq 0) = x(t \leq 0) = 0$. Равновесными моделями будем называть такие, в которых в процессе измерения ($t \geq 0$) устанавливается новое равновесное состояние теперь уже связанной системы $Y + X$, характеризующее новыми равновесными значениями параметров y_∞ , x_∞ , пропорциональными измеряемой величине F .

Рассмотрим вначале, в соответствии с ²³, модели первого класса, в которых точность измерения однозначно определяется переданной от тела Y телу X энергией U .

Переходной процесс в регистрирующем устройстве (теле X) описывается уравнением

$$\mu \ddot{x} + \eta \dot{x} + \beta x = \kappa y \equiv f(t), \quad x(0) = \dot{x}(0) = 0. \quad (2.1)$$

Переходной процесс в теле Y (ИС с согласующим устройством) описывается либо уравнением

$$M \ddot{y} + N \dot{y} = F - \kappa \dot{x}, \quad y(0) = 0 \quad (2.2)$$

(если y — обобщенная скорость), либо уравнением

$$N \ddot{y} + B \dot{y} = F + \kappa \dot{x}, \quad y(0) = 0 \quad (2.3)$$

(если y — обобщенная координата). Здесь введены обобщенные коэффициенты *): инерции μ , M ; сопротивления η , N ; упругости (жесткости) $\beta = \frac{\partial^2 U}{\partial x^2}$, B ; κ — коэффициент связи подсистем Y и X .

Видно, что в соответствии с формулой (1.10)

$$\sigma_x^2 = \frac{\overline{\Delta x^2}}{x_\infty^2} = \frac{kT}{\beta x_\infty^2} = \frac{kT}{2U}. \quad (2.4)$$

*) В частном случае, когда Y — электрическая цепь, система (2.1), (2.2) описывает переходной процесс в гальванометре (см. ²⁷), система (2.1), (2.3) — в электрометре.

Рассматривая измерение в области (симметричное, в соответствии с (1.7)), получим из (1.11), (1.24)

$$\frac{1}{\sigma^2} = \frac{\langle x^2 \rangle}{\Delta x^2} = \frac{2 \langle U \rangle}{kT}, \quad (2.5)$$

где $\langle U \rangle$ — среднее по $P(x)$ значение переданной от тела Y телу X энергии.

В равновесной модели легко непосредственно подсчитать дефект энтропии ΔH_1 . Действительно, до измерения

$$H_0 = H(Y) + H(X) = H(\alpha) + H(x) + H(y), \quad (2.6)$$

где $H(\alpha)$ — энтропия, определяемая всеми микроскопическими и макроскопическими параметрами, кроме y в теле Y и x в теле X . После измерения между двумя (ранее не зависимыми) параметрами (y и x) устанавливается связь, что приводит к уменьшению энтропии теперь уже связанных подсистем. Имеем

$$H_\infty = H(Y + X) = H(\alpha) + H(x) + MH(y/x). \quad (2.7)$$

Дефект энтропии равен

$$-\Delta H_1 = H_0 - H_\infty = H(y) - MH(y/x) = I(x, y). \quad (2.8)$$

Но неопределенность в y при известном x связана лишь с погрешностью $x - x' = \Delta x$ измерения. Вспоминая (1.15) и (1.22), получим, учитывая (2.5):

$$I(x, y) = I(x, x') = \frac{1}{2} \ln \left(\frac{1}{\sigma^2} + 1 \right) = \frac{1}{2} \ln \left(\frac{2 \langle U \rangle}{kT} + 1 \right). \quad (2.9)$$

Видно, что степень связи между подсистемами Y и X характеризуется количеством полученной информации, равным уменьшению энтропии, и основная задача анализа модели состоит в определении минимально допустимой диссипации энергии в процессе измерения, т. е. в ответе на вопрос II).

Однако и до проведения этого анализа высказываемая часто точка зрения о том, что при неограниченном времени измерения можно получить любую относительную точность, затратив сколь угодно мало энергии, легко опровергается приведенными выше оценками.

Казалось бы, эта точка зрения вполне соответствует классическому представлению об обратимых процессах. Однако рассматривая процесс получения информации, нельзя пренебрегать, как это обычно делается, малым *), но в данном случае определяющим эффектом взаимодействия двух систем, приводящим к тому, что энтропия связанной системы, вообще говоря, меньше суммы энтропий не взаимодействующих ранее систем. А если это так, то, согласно второму закону термодинамики, в процессе взаимодействия обязательно должна происходить компенсирующая это уменьшение энтропии диссипация энергии, т. е. скорость взаимодействия должна быть отличной от нуля.

В следующем разделе будут получены оценки этой скорости и даны объяснения физического смысла необратимости процесса измерения.

б) Нижняя граница необратимости процесса физического измерения. Исследования, проведенные Дж. фон Нейманом¹¹, а также другими авторами³⁸⁻⁴¹, доказывают

*) Заметим, что этот эффект $|\Delta H_1|$ мал лишь по сравнению с энтропией всей системы (H_0 или H_∞ в (2.6), (2.7)), однако для $H_0 - H(\alpha)$, т. е. для энтропии, определяемой параметрами y и x , эффект $|\Delta H_1|$ весьма заметен. Это существенно, так как обмен энергией U происходит именно по этим степеням свободы (y и x).

необратимость процесса измерения, связанную лишь с квантовой природой исследуемого объекта.

В последнее время появились работы, исследующие квантовомеханическую необратимость процесса измерения с точки зрения понятий теории информации^{22, 42, 43}: в терминах²¹, обсуждавшихся выше, показывается, что $I < |\Delta H_1|$, если хотя бы два оператора, соответствующие измеряемым параметрам, не коммутируют. Заметим в этой связи, что квантовомеханические ограничения на чувствительность (абсолютную погрешность Δl) имеют место лишь при одновременном измерении по крайней мере двух некоммутирующих переменных.

Ограничения на точность измерения одной величины возникают лишь в классической макроскопической части ИП и связаны с тепловыми флуктуациями. В работах^{26, 27} приведены соответствующие оценки дисперсии абсолютной погрешности Δx^2 . Выше, в соответствии с²³, были приведены оценки (2.4), (2.5) относительной дисперсии, связывающие ее (и относительную точность $1/\sigma$) с энергией U взаимодействия исследуемой системы и измерительного прибора, а также с дефектом энтропии (2.8), (2.9). Для ответа на вопрос II) существенно, однако, связать точность не с переданной U , а с диссипированной Q в процессе измерений энергией.

Известно, что диссипированная энергия Q в процессе (2.1) передачи энергии U телу X существенно определяется временем взаимодействия. Если сила нарастает быстро (т. е. время взаимодействия ограничено постоянной времени ИП), то произведенная работа вдвое превышает U : кинетическая энергия, равная потенциальной, диссипируется в системе — передается термостату. С другой стороны, известно¹⁰, что в обратимом процессе минимальная работа $R_{\min} \rightarrow U$ (т. е. $Q \rightarrow 0$). Причем дело здесь не в том, что коэффициент трения пренебрежимо мал (как это иногда объясняется): если пренебречь силами трения, то новое равновесное состояние связанной системы никогда не установится. Приближение к обратимому процессу может быть обеспечено, независимо от величины коэффициента трения η (с изменением η меняется лишь масштаб времени, так как изменяется постоянная времени ИП), — путем замедления процесса взаимодействия (медленным нарастанием силы). Объяснение этому хорошо известно: переданная энергия U определяется перемещением $x = vt$ (v — средняя скорость, t — время взаимодействия), а диссипированная энергия $Q \sim v^2 t$. Поэтому, сколь угодно замедляя переходной процесс, т. е. увеличивая t и пропорционально уменьшая v , можно, при той же U , обеспечить $Q \rightarrow 0$.

Отметим, что обычно не исследуется, что значит замедлить переходной процесс: предполагается, что экспериментатор медленно вводит силу (двигает поршень, выводит сопротивление в электрической цепи и пр.). Но оценить, как зависит совершенная экспериментатором работа от замедления, не представляется возможным.

Поэтому в рассматриваемой модели роль экспериментатора сводится к минимуму: он лишь выводит подсистемы из состояния, когда они не взаимодействуют, в состояние взаимодействия, когда начинается переходной процесс в связанной системе ИС + ИП. Это действие экспериментатора (выдвижение заслонки, замыкание ключа в электрической цепи и пр.) — однократное и не связано ни с передаваемой энергией, ни с энергией, запасенной в заслонке или разомкнутом ключе. Работа, совершаемая при этом экспериментатором, должна лишь превышать (вообще говоря, во много раз) kT — тот энергетический барьер, который не позволяет подсистемам спонтанно прийти в состояние взаимодействия. Далее в моделях рассматривается сам механизм замедления, реализуемый не экспериментатором, а некоторым регулятором, который (в отличие от экспери-

ментатора) может быть исследован физическими методами. При этом оказывается, что так как сам регулятор участвует в тепловом движении, то реализация замедления связана с некоторой дополнительной диссипацией энергии, обычно не учитываемой.

Возвращаясь к модели (2.1), (2.2) или (2.1), (2.3), видим, что если $f(t) = \kappa y(t)$ нарастает быстро ($\tau_y \ll \tau_x$, τ_y и τ_x — постоянные времени тел Y и X соответственно), то диссипативные потери в теле X : $Q_x = U$. Видно также, что согласующее устройство может служить демпфером, замедляющим нарастание силы в (2.1). При этом, однако, хотя и можно добиться $Q_x \rightarrow 0$, диссипативные потери в теле Y : $Q_y > U$. Этот результат следует из того, что коэффициент связи κ и пропорциональные ему безразмерные коэффициенты $\delta = \kappa^2/\eta H$ в системе (2.1), (2.2) и $\Delta = \kappa^2/\beta V$ в системе (2.1), (2.3) — ограничены сверху. Уменьшение δ и Δ относительно верхней границы возможно, но связано с ростом Q_y . Энергетически оптимальными являются такие значения, при которых $\sigma_y^2 = \sigma_x^2$ (см. (2.4)). Тогда

$$\begin{aligned} Q_\Sigma &= Q_y + Q_x = 2Q_x = 2U, \\ \sigma^2 &= \sigma_y^2 + \sigma_x^2 = 2\sigma_x^2, \end{aligned} \quad (2.10)$$

и достигается нижняя граница оценки для $\sigma^2 Q_\Sigma$ в случае двух степеней свободы (см. 23)

$$\sigma^2 Q_\Sigma \geq 2kT. \quad (2.11)$$

Эта оценка, вследствие (2.10), в 4 раза хуже предельной для одной степени свободы, когда согласующий параметр y отсутствует, а сила F действует непосредственно на регистрирующее устройство.

Тогда, в соответствии с (2.4) и $Q_x = U$, имеем

$$\sigma^2 Q_\Sigma \geq \frac{1}{2} kT. \quad (2.12)$$

Обе оценки относятся к необратимой реализации процесса измерения: либо когда время взаимодействия ограничено постоянной времени τ_x — (2.12), либо когда замедляется нарастание силы в X , но за счет этого увеличиваются потери в Y (2.11).

Таким образом, для уменьшения диссипации при измерении недостаточно замедлить нарастание силы в X : необходимо также замедлить нарастание силы в Y . В рассматриваемой модели это можно сделать лишь путем введения третьего тела Z — регулятора, который медленно меняет параметры H в (2.2) или V в (2.3). Но так как регулятор также участвует в тепловом движении, то оказывается, что при заданной U (т. е. точности $1/\sigma$) уменьшение Q_Σ не беспредельно *).

Пусть регулятор обеспечивает оптимальный закон нарастания силы $f(t)$ в (2.4):

$$f(t) = \begin{cases} \beta x(t) + \varepsilon & \text{при } 0 \leq t \leq t_0, \\ f_\infty = \beta x_\infty = \text{const} & \text{при } t \geq t_0. \end{cases} \quad (2.13)$$

Тогда диссипация Q_1 , связанная с уменьшением скорости, уменьшается в $n = t_0/4\tau_x$ раз (необратимая реализация процесса соответствует средней скорости $x_\infty/4\tau_x$). С другой стороны, вследствие (2.13), уравнение (2.1)

*) Можно показать, что аналогичные приводимым ниже результаты имеют место, если регулятор, вместо параметров H или V , медленно меняет параметр β в (2.1).

превращается в $m\ddot{x} + \eta\dot{x} = \varepsilon$, когда следует учитывать броуновский уход *)

$$\overline{\Delta x^2}(t_0) = \frac{2kT}{\eta} t_0. \quad (2.14)$$

При $t = t_0$, когда $f(t) = f_\infty$, тело X попадает в потенциальную яму, причем его энергия тем больше превышает равновесное значение $\beta \overline{\Delta x_0^2}/2 = kT/2$, чем больше t_0 . Эта энергия отдается термостату, что соответствует второму члену диссипативных потерь Q_2 . Из условия

$$Q_\Sigma = Q_1 + Q_2 = 2U \frac{4\tau_x}{t_0} + 2kT \frac{t_0}{4\tau_x} \rightarrow \min \quad (2.15)$$

получим

$$Q_{\Sigma \min} = 4 \sqrt{U kT}, \quad \left(\frac{t_0}{4\tau_x} \right)_{\text{opt}} \equiv n_{\text{opt}} = \sqrt{\frac{U}{kT}} = \frac{1}{\sqrt{2} \sigma_x} = \frac{1}{\sigma}, \quad (2.16)$$

откуда сразу следует предельная оценка (см. ²³) для «обратимой» реализации **) процесса измерения:

$$\sigma Q_\Sigma \geq 4kT. \quad (2.17)$$

Для модели с одной степенью свободы, аналогично предыдущему, оценка улучшается:

$$\sigma Q_\Sigma \geq 2kT, \quad (2.18)$$

однако в данном случае не в 4 раза, как (2.12), (2.11), а лишь вдвое.

Тот же результат получается при дискретной реализации замедления, когда регулятор фиксирует каждое i -е промежуточное равновесное состояние ($i = 1, 2, \dots, n$), причем в оптимальном случае $\Delta x_i = \Delta x_j = x_\infty/n$. Так как фиксация каждого равновесного состояния требует энергетической связи регулятора с системой, превышающей kT , то каждый i -й переход связан с диссипацией $Q_{i \min} \geq kT$.

Имеем в этом случае

$$Q_1 = \sum_{i=1}^n \Delta Q_{1i}, \quad Q_2 = \sum_{i=1}^n \Delta Q_{2i}, \quad U = \sum_{i=1}^n \Delta U_i, \quad (2.19)$$

$$\Delta Q_{1i} = \Delta Q_{1j} \equiv \Delta Q_1 = \frac{U}{n^2}.$$

С другой стороны, учитывая связь регулятора с системой, имеем:

$$\Delta Q_{2i} = a^2 kT, \quad a \geq 1. \quad (2.20)$$

Запишем аналогично (2.15):

$$Q_\Sigma = na^2 kT + \frac{U}{n} \rightarrow \min. \quad (2.21)$$

откуда, полагая $a^2 = 1$ (слабые ограничения), сразу получим (2.18).

Для того чтобы получить оценки роста энтропии ΔH_2 , необходимо от однократного измерения перейти к измерению в области. Легко видеть, что оценки (2.11) и (2.12) для необратимой реализации процесса справедли-

*) Т. е. с уменьшением разности сил ε увеличивается t_0 , а с ним и недетерминированное (броуновское) рассогласование сил.

**) Кавычки указывают на то, что оптимальная с точки зрения связи между σ и Q_Σ реализация процесса измерения, как видно из (2.16), соответствует конечной скорости процесса — порядка тепловой. Не только увеличение, но и уменьшение скорости относительно «естественной» (тепловой) связано с ростом диссипативных потерь.

ллы и для средних значений $\langle Q_\Sigma \rangle$ при симметричном (1.7) измерении (вследствие (1.24), аналогично (2.4), (2.5)). Для асимметричного измерения (1.8) оценки, вследствие (1.25), ухудшаются в 4 раза.

При оптимальной реализации процесса, однако, в оценках (2.17), (2.18) коэффициент, строго говоря, несколько уменьшается (в $\sqrt{3}/2$ раза), вследствие того, что $1/\sigma \sim \sqrt{\langle x^2 \rangle}$, а $\langle Q \rangle \sim \langle \sqrt{x^2} \rangle$. По этой же причине для асимметричного измерения оценки (2.17), (2.18) ухудшаются не в 4, а в 2 раза.

В дальнейшем мы будем пренебрегать коэффициентом $\sqrt{3}/2$ в (2.17), (2.18) и применять эти оценки, как и (2.11), (2.12), также и к случаю симметричного измерения в области *), когда $Q_\Sigma = \langle Q_\Sigma \rangle$ (скобки будем опускать).

Из формул (2.12), (2.18) получаем, что минимальный рост энтропии ΔH_2 за счет диссипации полностью определяется точностью измерения и способом его реализации (необратимым или предельно «обратимым»):

$$\Delta H_2^{(н)} \geq \frac{Q_\Sigma}{kT} = \frac{1}{2\sigma^2}, \quad \Delta H_2^{(о)} \geq \frac{2}{\sigma} = 2 \sqrt{2\Delta H_2^{(н)}}. \quad (2.22)$$

Согласно предыдущему (см. (2.9)),

$$|\Delta H_1| = I = \frac{1}{2} \ln \left(\frac{1}{\sigma^2} + 1 \right) = \begin{cases} \ln(1/\sigma), & \sigma^2 \ll 1, \\ 1/2\sigma^2, & \sigma^2 \gg 1. \end{cases} \quad (2.23)$$

Выше упоминалось, что по Бриллюэну ^{8, 16}

$$\Delta H_2 \geq |\Delta H_1|, \quad (2.24)$$

причем знак равенства определяет нижнюю границу необратимости процесса измерения. Из (2.22), (2.23) видно, что эта нижняя граница сильно занижена и может достигаться лишь в пределе при $I \rightarrow 0$. Впрочем, даже это не совсем точно, так как формулы (2.22) справедливы, вообще говоря, лишь при $\sigma^2 \leq 1$. В противном случае (т. е. при $U < kT$ (2.4)), в соответствии с (2.20), $Q_\Sigma > kT > U$, т. е. $\Delta H_2 > 1$ всегда. В связи с этим заметим, что возможность достижения нижней границы (2.24) на рассмотренных в работе ⁸ примерах не доказана. Если же она все же достижима при $\sigma^2 < 1$, то из (2.23), (2.5) получается крайне неестественная связь оптимальной скорости переходного процесса с параметрами измерения и ИП (см. в связи с этим вторую сноску на стр. 478).

Главное же принципиальное возражение против того, что при $I > 0$ может осуществляться равенство в (2.24), возникает при ответе на сформулированный выше вопрос III). Еще с рассмотрения парадокса Максвелла известно, что полученная информация не полностью характеризует достигнутый результат: она может быть использована в дальнейшем для внесения упорядочения в систему (для управления), т. е. для уменьшения энтропии $\Delta H_3 < 0$.

Именно низкое значение информационного коэффициента полезного действия

$$r_{\text{инф max}} = \frac{I}{\Delta H_{2 \text{ min}}} = \frac{\sigma}{2} \ln \frac{1}{\sigma} = \frac{1}{2} I e^{-I}, \quad \sigma^2 \ll 1, \quad (2.25)$$

*) В связи с тем, что соотношения (2.11), (2.12), (2.17), (2.18), а также последующее (2.22) справедливы как для измерения в области (оценки случайной величины), так и для однократного измерения (оценки постоянного параметра), представляется естественным распространить формулу (2.23) и понятие количества информации (определенное лишь для случайной величины) и на однократное измерение.

получающееся из соотношений (2.23), (2.22), обеспечивает возможность эффективного управления без нарушения второго начала термодинамики:

$$\Delta H_2 - |\Delta H_1| = \Delta H_2 (1 - r_{\text{инф max}}) \geq |\Delta H_3| > 0. \quad (2.26)$$

В отличие от информационного, механический КПД $r_{\text{мех}}$ полностью характеризует достигнутый результат

$$r_{\text{мех max}} = \frac{U}{U + Q_{\Sigma \text{ min}}} \approx 1 - 2\sqrt{\frac{kT}{2U}}, \quad U \gg kT, \quad (2.27)$$

и тем меньше отличается от единицы, чем больше полезный эффект (переданная энергия U).

Таким образом, анализ замедления переходного процесса, реализуемого идеальным (однако участвующим в тепловом движении) регулятором, приводит к естественным ограничениям (2.16), (2.22) в приближении к обратимому процессу.

Все эти результаты относились к моделям первого класса, когда переданная энергия U полностью определяла информационные характеристики процесса измерения (см. (2.4), (2.16), (2.23)).

Можно рассмотреть модели второго класса, когда оценка (2.12), связывающая Q_{Σ} с σ , получается непосредственно и не может быть улучшена, но зато при этом значение переданной энергии U (уже не определяющей σ) может существенно превышать ограничения (2.16), (2.27).

Пусть измеряемая величина F , в отличие от рассмотренной модели, — не сосредоточенный параметр, а медленно убывает с расстоянием z . Тогда роль регулятора Z играет «тележка», медленно приближающаяся тело X к ИС.

В этом случае оптимальный закон нарастания силы:

$$f(t) = bx(t) + \epsilon, \quad \beta - b = \beta_{\text{эф}} \ll \beta, \quad (2.28)$$

и в уравнении (2.1), вместо β , будет $\beta_{\text{эф}}$. Тогда диссипация уменьшается в $\beta/\beta_{\text{эф}}$ раз, но средний квадрат погрешности увеличивается во столько же раз, так что оценка (2.12) не улучшается. При этом, однако, связь между диссипированной энергией Q_{Σ} и переданной U может быть лучше, чем по (2.16), но между точностью и U — существенно хуже, чем в (2.4).

Такого рода взаимодействие характерно тем, что на пробное тело действует малая разность двух больших сил, созданных соответствующими полями. Поэтому имеется весьма слабая связь с параметром x (точность $1/\sigma$, характеризующая эту связь, мала), и возникает «сильное» взаимодействие в равновесии: незначительная флуктуация суммарной энергии приводит к большим флуктуациям параметра x , а значит, к большим перекачкам энергии от одного поля к другому. Как известно⁴⁴, нормальность макроскопических переменных характеризуется высокой стабильностью их ($1/\sigma \gg 1$). В данном случае энергия системы — нормальная переменная, однако энергия подсистем — нет, ибо их тепловые флуктуации велики, а, соответственно, информация, извлекаемая пробным телом (с регистрирующим параметром x), мала.

Таким образом, лишь при такой передаче энергии U , при которой из системы извлекается информация, характеризуемая точностью по (2.4), имеет место ограничение (2.16) в приближении к обратимому процессу.

Остановимся теперь на выборе параметров регистрирующего устройства.

Из предыдущего видно, что при заданном значении непосредственно измеряемой величины $x_{\infty} = f_{\infty}$ увеличение точности (т. е. U) может быть достигнуто путем уменьшения коэффициента упругости β . Если, кроме энергетических ограничений, учесть также ограничения на время

измерения, которые во всех случаях (как в необратимом, так и оптимальном) определяются временем релаксации τ_x , легко получить следующие рекомендации по синтезу регистрирующего устройства. Известно, что характер переходного режима для уравнения (2.1) полностью определяется параметром

$$\gamma = \frac{\eta}{2\sqrt{\mu\beta}}, \quad (2.29)$$

обратным добротности. Легко показать, что в граничном ($\gamma = 1$) между колебательным ($\gamma < 1$) и апериодическим ($\gamma > 1$) режимами достигается минимум времени релаксации τ_x (а с ним — и времени измерения) — при заданном значении β (т. е. заданной точности). Имеем

$$\tau_x = \frac{2\mu}{\eta} = \frac{\eta}{2\beta} \quad \text{при} \quad \gamma = 1. \quad (2.30)$$

Отсюда видно, почему точные измерения характеризуются малыми значениями β и μ .

в) Энергетическая цена точности и количества информации. Из полученных выше оценок (2.12), (2.18), (2.22), (2.23) для процесса измерения немедленно следуют предельные оценки энергетической цены точности \hat{e}_σ и единицы количества информации \hat{e}_I . Имеем из (2.18)

$$\hat{e}_\sigma = \frac{Q_{\Sigma \min}}{1/\sigma} = 2kT, \quad (2.31)$$

т. е. при оптимальном замедлении процесса цена точности постоянная.

Однако при необратимой реализации процесса энергетическая цена точности

$$\hat{e}_\sigma^{(n)} = \frac{1}{2} kT \left(\frac{1}{\sigma} \right) \quad (2.32)$$

растет пропорционально точности.

Из (2.23) и предыдущих формул имеем

$$\hat{e}_I = \frac{Q_{\Sigma \min}}{I} = 2kT \frac{1}{\sigma \ln(1/\sigma)} = 2kT \frac{e^I}{I}, \quad \sigma^2 \ll 1, \quad (2.33)$$

$$\hat{e}_I^{(n)} = \frac{kT}{\sigma^2 \ln[(1/\sigma^2) + 1]} = \begin{cases} (kT/2I) e^{2I}, & \sigma^2 \ll 1, \\ kT, & \sigma^2 \gg 1, \end{cases} \quad (2.34)$$

т. е. для $\sigma^2 \ll 1$ не только при необратимой, но и при оптимальной реализации процесса цена единицы количества информации I растет экспоненциально (относительно I). Лишь при $I \rightarrow 0$ ($\sigma^2 \gg 1$) эта цена приближается к kT — нижней оценке Бриллюэна⁸, полученной также в²¹.

Одна из причин получения заниженных оценок для \hat{e}_I состоит в том, что ни в⁸, ни в²¹ не вводится явно ограничение на энергию (либо наоборот, на количество информации или точность^{*}), в связи с чем эти оценки относятся к малоинтересному предельному случаю $I \rightarrow 0$.

Вторая причина имеет другое происхождение: очень часто оцениваются энергетические затраты лишь на первом этапе измерения (см. схему (1.1)), в особенности, если этот этап — обнаружение^{**}).

^{*}) Соответствующие естественные условные экстремальные задачи обсуждаются в следующей главе.

^{**}) Практически это иногда вполне оправдано: например, при радиолокационном измерении дальности рассчитывается лишь энергия сигнала, необходимая для надежного обнаружения (это соответствует этапу $I \rightarrow \lambda$ в (1.1)), и никого не интересует энергия, затрачиваемая в развертывающем устройстве или в счетчике. Вместе с тем

Рассмотрим кратко предельные энергетические соотношения при обнаружении в зависимости от точности. Статистическая теория обнаружения развивалась примерно в те же годы, что и теория информации, и представлена обширной литературой. Здесь следует выделить основополагающие работы Котельникова^{45, 46} и более поздние⁴⁷⁻⁴⁹. Не останавливаясь на многочисленных результатах этой теории, отметим основной важный для обсуждаемого вопроса факт: пороговая энергия определяется прежде всего энергией шума kT , а уровень порога можно найти из требований к вероятности обнаружения $(1 - w_+)$ и ложных тревог w_- . Известно, что при любых соотношениях между длительностью $\hat{\tau}$ импульса сигнала и постоянной времени τ релаксации приемника энергия сигнала для надежного обнаружения $E_c > kT$.

Если речь идет об обнаружении постоянной силы длительности $\hat{\tau}$ (видеоимпульса) или импульса синусоидальных колебаний известной частоты ν и длительности $\hat{\tau}$, то согласованный приемник, как известно⁵⁰, характеризуется временем релаксации $\tau \approx \hat{\tau}$: в этом случае оптимально используется полоса $\Delta\nu$ приемника, и основная часть энергии сигнала передается приемнику и диссипируется в нем.

Часто используется сигнал с большим числом $n \gg 1$ степеней свободы (широкополосный, либо повторяющиеся импульсы, либо длительно — по сравнению с τ — действующая постоянная сила). Тогда энергия сигнала на одну степень свободы может быть меньше kT (мощность шума превышает мощность сигнала). В этих случаях, как известно⁵⁰, используется метод накопления (для синусоидального сигнала — когерентного накопления). При этом энергетическое отношение сигнал/шум на выходе приемника возрастает в $n = \hat{\tau}/\tau \gg 1$ раз, однако суммарная энергия сигнала $E_c^{(n)}$ в \sqrt{n} раз выше, чем энергия $E_c^{(1)}$ сигнала с одной степенью свободы, обеспечивающего те же w_+ и w_- .

В последнее время приобрел большое значение противоположный случай соотношения времени $\tau/\hat{\tau} = m \gg 1$. В работе Брагинского⁵¹ показано, что повышение чувствительности механического приемника — осциллятора, описываемого (2.1), — неизбежно связано с увеличением времени релаксации τ и соответствующим сужением полосы приемника. В работе⁵¹ предложена оптимальная стратегия принятия решения, которая, несмотря на столь значительное рассогласование приемника относительно сигнала, обеспечивает обнаружение почти предельного порогового сигнала. При этом передаваемая приемнику (или извлекаемая из него) порция энергии U может быть в m раз меньше kT . Следует заметить, однако, что пороговая энергия сигнала E_c при этом не уменьшается: лишь $1/m$ -я часть этой энергии передается приемнику (вследствие того, что приемник имеет в m раз более узкую полосу, чем спектр сигнала), а основная часть энергии сигнала рассеивается, не попадая в приемник, и поглощается в термостате. Именно поэтому применение описанного эффекта кажущегося снижения порога обнаружения для передачи информации или активной радиолокации, естественно, нецелесообразно, так как этот способ приводит не к уменьшению энергии передатчика, а лишь к неполному использованию ее в приемнике.

Однако предложенная в работе⁵¹ и развитая в работе⁵² методика имеет большое значение в физических экспериментах по обнаружению тонких эффектов (поиск кварков, гравитационных волн и др. — см. ⁵¹).

именно на этом этапе ($F \rightarrow x$ в (1.1)) формируется физическая величина, пропорциональная дальности, и к нему относятся полученные выше оценки. Из (2.32) видно, что для обычно реализуемой точности ($\sim 10^3$) затраты на втором этапе пренебрежимо малы по сравнению с энергией сигнала.

В связи с тем, что передаваемая приемнику энергия U может быть в m раз меньше kT , квантовые ограничения начинают сказываться существенно раньше (при $U = kT$ $m \approx h\nu$). Эти вопросы подробно рассмотрены в ⁵².

Остановимся кратко на последовательных этапах измерения, описанных в ^{51, 52}, и сопоставим их с этапами по схеме (1.1). Измеряемая сила вызывает малые нестационарные изменения амплитуды колебаний механического осциллятора (2.1) — этап $(l \rightarrow \lambda)$ в (1.1). Далее при помощи емкостного датчика механические колебания преобразуются в электрические, что соответствует этапу $\lambda \rightarrow F$ в (1.1). В работе ⁵² показано, что и этот этап целесообразно осуществлять в нестационарном режиме: тогда можно реализовать минимальный пороговый сигнал (при высоком начальном уровне энергии осциллятора). Следующий этап представляет особый интерес: показано ⁵², что измерение энергии электрического осциллятора до и после действия силы может быть выполнено практически без возмущения. Взаимодействие электронного пучка с резонатором может происходить почти полностью упруго — без передачи энергии. Между взаимодействиями в фазе и противофазе часть электронов попадает в регистрирующее устройство, что соответствует последним этапам в (1.1): $F \rightarrow y \rightarrow x$; здесь y — ток, а x — значение регистрирующего параметра. Последний этап, как всегда, квазистационарный и связан с передачей и рассеянием энергии, превышающей kT .

Таким образом, во всех случаях обнаружения рассеянная в приемнике и окружающей среде энергия превышает kT .

Рассматривая точность измерения при обнаружении, следует различать две задачи.

С одной стороны, можно при обнаружении сохранить амплитудную информацию и таким образом не просто обнаруживать, но и измерять силу. При этом в первых двух случаях, когда $\hat{\tau} \gg \tau$, требование 3) раздела в) гл. 1 об установлении стационарного состояния выполняется, и полученные выше энергетические оценки справедливы для точности измерения амплитуды. В третьем случае, как видно из только что сказанного, стационарным является лишь последний этап в (1.1), для которого полученные оценки также справедливы.

С другой стороны, что наиболее характерно для обнаружения, интерес представляет точность измерения не амплитуды, а некоторого параметра l , по которому перестраивается обнаружитель. Такое обнаружение также является первым этапом ($l \rightarrow \lambda$) измерения по схеме (1.1), причем λ — либо мгновенное положение приемника на некоторой шкале, либо номер обнаружителя с положительным ответом (при параллельном многоканальном приеме).

Для того чтобы обеспечить точность $1/\sigma_0$ (см. (1.10), (1.13), (1.14)) порядка разрешающей способности $1/\varepsilon$, необходимо, чтобы вероятность ошибки

$$w \approx \frac{1}{\varepsilon} w_- \approx \varepsilon, \quad w_- \approx \varepsilon^2, \quad (2.35)$$

где w_- — вероятность ложного выброса в одном акте обнаружения (разрешаемом элементе). Отсюда легко найти энергию E_0 порогового уровня для обнаружения сигнала

$$E_0 \approx kT \ln \frac{1}{w_-} = 2kT \ln \frac{1}{\varepsilon}$$

и примерное значение энергии сигнала

$$E_c \approx 2E_0 = 4kT \ln \frac{1}{\varepsilon} \approx 4kTI. \quad (2.36)$$

Видно, таким образом, что при обнаружении (т. е. на первом этапе измерения) достигается по порядку величины минимальная цена единицы информации; однако эта цена относится лишь к энергии сигнала и не учитывает энергетических затрат на втором этапе, соответствующем измерению времени или расстояния. Хотя практически эта (вторая) часть энергетических затрат бывает незначительна (см. сноску на стр. 481), принципиально именно она определяет энергетическую цену точности и количества информации (2.31)–(2.34).

В заключение этой главы остановимся на причинах того, что термодинамические ограничения не сказывались практически на совершенствовании физических экспериментов.

В физических экспериментах стремятся, как правило, добиться повышения чувствительности^{51–55}, т. е. снижения абсолютной среднеквадратической погрешности $\sqrt{\Delta l^2}$ (1.4) и, соответственно, уменьшения разрешаемого интервала Δl_p (1.5). При этом, однако, относительную точность эксперимента искусственно уменьшают, прибегая к преобразованию (1.6) за счет использования эталонов (1.9).

Сказанное относится к большинству — даже наиболее тонких и точных — экспериментов: к классическим опытам Майкельсона и экспериментам по использованию эффекта Мёссбауэра⁵³; к опытам по обнаружению гравитационных волн⁵¹ или вообще «решающим экспериментам»⁵⁴ и др.

Во всех этих экспериментах, несмотря на высокую чувствительность, точность $1/\sigma$ не превышала 10^3 . При этом, даже согласно худшей оценке (2.32), при $T = 300^\circ\text{K}$ в одном акте измерения необходимо рассеивать лишь 10^{-7} эрг. Если, однако, согласно⁵⁶, в обычных экспериментах высокая относительная точность и не очень нужна, то фундаментальные физические постоянные необходимо знать с высокой точностью. Тем не менее все фундаментальные постоянные определены с относительной точностью, не превышающей 10^6 – 10^7 (см., например, ^{56, 57}). Один акт измерения с такой точностью требует рассеять ~ 1 эрг, что также меньше реальных технических ограничений.

Термодинамические ограничения приобретают не только принципиальное, но и практическое значение в современных системах сбора, хранения, передачи и обработки информации. Во-первых, эти системы характеризуются передачей и обработкой больших массивов информации с высоким быстродействием. Во-вторых, передача информации связана с большими потерями на затухание сигналов. В-третьих, современные вычислительные машины оперируют с числами весьма высокой точности ($\sim 10^9$ – 10^{13}).

Остановимся поэтому на термодинамических особенностях процессов передачи и обработки информации, вытекающих из ограничений на точность измерения.

3. ТЕРМОДИНАМИЧЕСКИЕ ХАРАКТЕРИСТИКИ ПРОЦЕССА ПЕРЕДАЧИ ИНФОРМАЦИИ

а) Энергетический смысл кодирования. Для выяснения термодинамических особенностей процесса достаточно ограничиться анализом передачи одномерных некоррелированных сообщений (чисел) $x \in X$. Именно этот случай непосредственно связан с рассмотренным выше процессом измерения и позволяет получить энергетические оценки точности различного представления чисел, необходимые в дальнейшем для исследования процесса обработки информации.

Пусть измеренное значение скалярной величины в виде числа необходимо передать по линии связи. Аналогично предыдущему, потребуем, чтобы на приемном конце интересующая нас величина также была представлена в виде одной скалярной характеристики. Очевидно, что получение сообщений в виде скалярной величины с необходимой точностью требует не меньших энергетических затрат, чем измерение. В отличие от измерения, однако, здесь существенно, какая часть энергии расходуется исследуемой системой (в данном случае — энергия пришедшего сигнала), какая — измерительным прибором (в данном случае — приемником, реализующим декодирование). Если единственной характеристикой сигнала, несущей информацию о числе, является его амплитуда (энергия), то все оценки, полученные при анализе измерения, непосредственно относятся к энергии сигнала, поглощаемой в приемнике^{24, 25}, и никакие дополнительные затраты в приемнике не требуются. Такой способ кодирования будем называть аналоговым, или натуральным, представлением числа. Этот способ передачи соответствует дистанционному измерению, крайне невыгодному в энергетическом отношении: вследствие затухания сигнала энергетические затраты в передатчике оказываются весьма большими. Поэтому целесообразно перераспределить энергию, необходимую для обеспечения заданной точности передачи, таким образом, чтобы максимально возможно уменьшить энергию принятого (значит, и переданного) сигнала, соответственно увеличив энергетические затраты при декодировании на приемном конце. Такое перераспределение энергии и осуществляется различными способами кодирования, когда информация о числе содержится не только в амплитуде сигнала, но и в его временной позиции (позиционные коды) или в несущей частоте.

Эти (и другие) способы кодирования основаны на том, что вместо одной используется множество степеней свободы: скалярной величине приводится в однозначное соответствие вектор, каждая из составляющих которого требует существенно меньшей точности представления, чем исходный скаляр. Выше было показано, что при высокой точности энергетические затраты растут экспоненциально с увеличением количества информации о данной скалярной величине. С другой стороны, понятно, что при росте количества информации не за счет точности, а вследствие увеличения числа составляющих, энергетические затраты пропорциональны увеличению количества информации (т. е. числу составляющих). Именно в этом заключается физический смысл любых способов кодирования, в том числе оптимального — по Шеннону.

б) **Натуральное представление числа.** Прежде чем оценивать более экономные (за счет кодирования) способы, рассмотрим вначале энергетическую цену точности натурального представления числа.

Выше была получена цена точности для однократного измерения, а затем эти результаты путем простых рассуждений были распространены на измерение в области. Строго говоря, однако, постановки соответствующих экстремальных задач различаются. Исследуем этот вопрос для процесса передачи.

Следуя работам^{5, 7}, будем рассматривать передачу сообщений по схеме

$$x \rightarrow y \rightarrow y' \rightarrow x', \quad (3.1)$$

где x и x' — соответственно исходное и принятое сообщения, y и y' — сигналы на входе и выходе передающей линии.

В отличие от традиционных^{5, 7, 31}, рассматриваемые ниже постановки экстремальных задач упрощаются тем, что относятся лишь к некоррели-

рованным сообщениям (числам), однако имеют важную отличительную особенность, связанную с целью исследований. Если удовлетворение «требований к качеству (точности)» передачи⁷ обычно ограничивается пропускной способностью канала, то ниже в качестве ограничивающей величины принимается энергия^{*}). Поэтому допустимы лишь такие ограничения на качество передачи, которые не требуют неограниченно большой энергии. Так, например, используемое в обычных математических постановках ограничение

$$P \left[\frac{|x - x'|}{x_m} \leq \frac{\varepsilon}{2} \right] = 1, \quad (3.2)$$

недопустимо.

Такое требование к точности (3.2) исходит из классической теории приближения функции, где x — точное задание функции, x' — приближенное. В этом случае, когда x, x' — абстрактные математические объекты, условие (3.2) вполне корректно. Но когда в переходе $x \rightarrow x'$ участвуют физические сигналы ($y \rightarrow y'$), условие (3.2) не может быть достигнуто, если энергия ограничена.

Для любой физической закодированной величины предположение о нулевой вероятности w ошибки, превышающей ε (если ε меньше априорного интервала), неизбежно связано с предположением о сколь угодно большой энергии сигнала (при ненулевой температуре T). Это следует из классической статистической теории флуктуаций^{10, 26, 27}.

Таким образом, с учетом ограниченности энергии следовало бы, аналогично (1.5), (1.14), условие (3.2) переписать так:

$$P \left[\frac{|x - x'|}{x_m} \leq \frac{\varepsilon}{2} \right] = 1 - w, \quad w > 0. \quad (3.3)$$

Однако такое задание требований к передаче, в отличие от (3.2), не характеризует ее качество и не позволяет сравнивать различные зависимости $y(x)$: для различных $y(x)$ те же значения ε, w могут приводить к существенно различной точности. Сказанное особенно наглядно проявляется, в частности, при сравнении натурального представления числа в аналоговых системах с разрядным представлением — в цифровых.

Поэтому в дальнейшем рассматриваются такие требования к качеству передачи, которые выражаются через одну скалярную характеристику — среднюю относительную точность $1/\sigma$.

Как и при измерении, аналогично (1.8), (1.16), полагаем, что числа x распределены равномерно на отрезке

$$p(x) = \frac{1}{x_m}, \quad 0 \leq x \leq x_m \Leftrightarrow x \in X. \quad (3.4)$$

Рассмотрим наиболее естественный случай натурального представления числа — амплитудный линейный:

$$y = Bx. \quad (3.5)$$

Предполагая, что в приемнике происходят те же процессы, которые рассмотрены при анализе измерения, т. е. что положение x регистрирующего элемента полностью определяется переданной этому элементу энергией $U(x) = \beta x^2/2$ (в данном случае частью энергии сигнала), и прене-

^{*}) Существенно, что в этой постановке полученные результаты всегда приводят к конечному значению пропускной способности. С другой стороны, некоторые из результатов в обычной постановке в действительности справедливы лишь при неограниченно большой энергии. Заметим, что вводимые в обычных постановках ограничения по пропускной способности, связанные с ограничением средней мощности сигнала, не эквивалентны ограничениям на среднюю энергию.

брегая шумами в канале, зададим условное распределение $P(x'/x)$ его плотностью

$$p(x'/x) = p(|x - x'|) = \sqrt{\frac{\beta}{2\pi kT}} \exp\left[-\frac{\beta}{kT} \frac{(x - x')^2}{2}\right]. \quad (3.6)$$

Так как энергетические затраты определяются относительной погрешностью σ (или σ_0 , в соответствии с (1.13), (1.12), (1.18)), перепишем (3.6), учитывая, что $U_{\max} = \beta x_m^2/2$, а $p(x) = 1/x_m$ (3.4):

$$p(x'/x) = p(x) \sqrt{\frac{U_{\max}}{\pi kT}} \exp\left[-\frac{U_{\max}}{kT} \frac{(x - x')^2}{x_m^2}\right]. \quad (3.7)$$

Физический смысл коэффициента B в соотношении (3.5) становится ясным, если считать, что энергия сигнала

$$\tilde{E}_c = y^2 = (1 + \alpha) \frac{U_{\max}}{kT} \frac{x^2}{x_m^2}, \quad (3.8)$$

где α определяет долю диссипированной энергии Q относительно переданной U .

Из предыдущего известно (2.16)–(2.18), что

$$\alpha = \begin{cases} \alpha_n = 1, \\ \alpha_0 = 2\sqrt{\frac{kT}{U}}, \end{cases} \quad (3.9)$$

где индекс «н» соответствует необратимой, а «0» — «обратимой», т. е. оптимальной, согласно разделу 6) гл. 2, реализации процесса декодирования в приемнике. Из (3.8) видно, что, в отличие от измерения, где переход к оптимальному режиму дает значительный (пропорциональный точности $1/\sigma$ — см. (2.18), (2.11)) выигрыш в энергетике, в случае передачи это не так (энергия сигнала (3.8), которая определяет энергию передатчика, изменится менее чем вдвое). Это и понятно, если учесть, что независимо от дальнейшего использования переданной энергии U (полезного или рассеяния в приемнике), на передающем конце все равно следует обеспечить после затухания $E_c > U$.

Будем измерять энергию в относительных единицах, обозначив

$$\frac{U_{\max}}{kT} \equiv A, \quad \frac{\langle U \rangle}{kT} \equiv S, \quad B = \sqrt{(1 + \alpha)A} \frac{1}{x_m}. \quad (3.10)$$

Из (3.4), вспоминая формулы (1.8), (1.25), получим

$$S = \frac{A}{3}. \quad (3.11)$$

При передаче чисел x , заданных в области (3.4), производится дискретизация (квантование) их и соответствующих им амплитуд сигналов y (3.5), (3.10), согласно (1.14): $\Delta x_p/x_m = \varepsilon$. Любой выбор ε в соответствии с (3.3) приводит к ненулевой вероятности ошибки w . Обе эти величины по-разному влияют на точность $1/\sigma$ и на выбор средней энергии S (3.11) для передачи однократного сообщения. Для определения суммарной энергии S_Σ на передачу всех сообщений x_i ($i = 1, 2, \dots, 1/\varepsilon$) необходимо среднюю энергию S умножить на количество отсчетов $1/\varepsilon$:

$$S_\Sigma = S \frac{1}{\varepsilon}. \quad (3.12)$$

Запишем постановки экстремальных задач ²⁴:

$$\sigma(\varepsilon, w) \rightarrow \min, \quad S(\varepsilon, w) \leq C_1 = \text{const}, \quad (3.13)$$

$$\sigma(\varepsilon, w) \rightarrow \min, \quad S_\Sigma(\varepsilon, w) \leq C_2 = \text{const}, \quad (3.14)$$

каждая из которых соответствует энергетической оптимизации параметров передачи — либо для однократного сообщения (3.13), либо во всей области (3.14).

Решение задачи (3.13) показывает, что σ_{\min} (и соответственно I_{\max}) достигается при условиях

$$\varepsilon \rightarrow 0, \quad w \rightarrow 1$$

так, что

$$\frac{1-w}{\varepsilon} = \sqrt{\frac{3}{\pi} S}, \quad (3.15)$$

т. е., как и следовало ожидать, для однократной передачи (или измерения) дискретизации не требуется ($\varepsilon \rightarrow 0$).

Имеем в этом случае (учитывая (3.4), (1.8), (1.25))

$$\sigma^2 = \frac{2}{S}, \quad I \approx \frac{1}{2} \ln S, \quad S \gg 1. \quad (3.16)$$

Видно, что и для непрерывного сообщения ($\varepsilon \rightarrow 0$) количество информации — конечное и определяется отпущенной средней энергией. Заметим, что при использовании симметричного кодирования (т. е. если аналогично (1.7) в (3.4) положить $-x_m/2 \leq x \leq x_m/2$) энергетическая цена точности уменьшается в 4 раза:

$$\sigma^2 = \frac{1}{2S}, \quad -\frac{x_m}{2} \leq x \leq \frac{x_m}{2}. \quad (3.17)$$

Решение задачи (3.14) для симметричной области (3.17) дает энергетически оптимальные соотношения между количеством отсчетов $1/\varepsilon$ (или разрешающей способностью для измерения), точностью $1/\sigma$ и надежностью $1/w$:

$$w = 0,4; \quad \varepsilon^2 = \frac{0,6}{(S_\Sigma)^{2/3}} = \frac{0,47}{S}, \quad (3.18)$$

$$\sigma^2 = 3\sigma_0^2 = \varepsilon^2 = \frac{5}{8S}.$$

Сравнивая значения минимальных погрешностей (σ , σ_0) с соответствующими значениями (3.17) (где также, $\sigma^2 = 3\sigma_0^2$), полученными из задачи (3.13), видим, что ошибки дискретизации приводят к увеличению энергетической цены точности в 5/4 раза. При симметричном кодировании (3.4), в соответствии с (1.18),

$$\sigma^2 = 12\sigma_0^2 = 4\varepsilon^2 = \frac{5}{2S}, \quad (3.19)$$

т. е. проигрыш по сравнению с (3.16) также в 5/4 раза.

Обратим внимание на то, что при энергетически оптимальной дискретизации (3.18) и симметричном кодировании: $\varepsilon = \sigma$, т. е. в соответствии с (3.16) и (1.19),

$$I \approx \ln \frac{1}{\sigma} = \ln \frac{1}{\varepsilon}, \quad \varepsilon^2 \ll 1. \quad (3.20)$$

Это значит, что можно говорить об определении энергетически оптимальной ε -энтропии при решении экстремальной задачи (3.14). Условия (3.18) определяют единственное значение ε , когда при заданном S_Σ энтропия $H_\varepsilon = \ln 1/\varepsilon$ (несмотря на то, что $w > 0$ в (3.3)).

Из формул (3.20), (3.16) видно, что количество информации логарифмически зависит от энергии (при $S \gg 1$), и лишь в предельном случае, когда $S \ll 1$, приходим к линейной зависимости: $I \approx S$.

Более точное (по сравнению с (3.16)) выражение, справедливое для любых S , имеет, в соответствии с (2.23), (1.22), вид

$$I = \frac{1}{2} \ln (2S + 1). \quad (3.21)$$

Сравним это выражение с формулой Шеннона⁵ для предельной средней за $\Delta\tau$ пропускной способности C (в полосе частот $\Delta\nu$):

$$C = \Delta\nu \ln \left(\frac{P_c}{P_{ш}} + 1 \right). \quad (3.22)$$

Так как мощность шума $P_{ш} = kT\Delta\nu$, мощность сигнала $P_c = E_c/\Delta\tau$, а $I = C\Delta\tau$, получаем:

$$I = \Delta\tau\Delta\nu \ln \left(\frac{E_c}{kT\Delta\nu\Delta\tau} + 1 \right). \quad (3.23)$$

В соответствии с^{5, 8, 12}, полагая, что для количества информации на одну степень свободы (что выражается формулой (3.21)) следует положить $\Delta\nu\Delta\tau = 1/2$, приходим, вспоминая (3.10), (3.9), (3.8), к формуле (3.21), если в (3.9) $\alpha = \alpha_0 = 0$ (1). Если $\alpha = \alpha_n = 1$, то $2S = \tilde{E}_c$ (3.8), и количество информации уменьшается (крайне незначительно: меньше чем на единицу). Применение оптимального замедления для передачи нецелесообразно не только вследствие малости положительного эффекта, но и потому, что требует использования многих степеней свободы (накопления) для передачи одного отсчета. Но при этом более выгодно применение кодирования, при котором скалярной величине ставится в соответствие сигнал со многими степенями свободы (вектор).

в) **Позиционные способы кодирования.** Выше было показано, что ограниченность энергетических ресурсов приводит к необходимости дискретизации непрерывных сообщений. Количество отсчетов $1/\varepsilon$ определяет при этом максимальное количество разрешимых значений передаваемого скалярного сообщения (числа). Переход от обычных ограничений по мощности к ограничениям на энергию сигнала показывает, что широкополосные сигналы более экономны энергетически (а не по мощности) лишь вследствие использования многих степеней свободы, однако только до определенных пределов. При обсуждении обнаружения было выяснено, что всегда для надежного обнаружения энергия сигнала $E_c > kT$, и различные методы накопления не меняют этого ограничения. Поэтому увеличение числа степеней свободы M сигнала больше $1/\varepsilon$ бессмысленно.

Наиболее экономный с точки зрения энергии передатчика способ кодирования числа — одноимпульсный позиционный, когда

$$M = M_{\max} = 1/\varepsilon. \quad (3.24)$$

При этом используется частный случай векторного представления — единичным ортом, когда все составляющие, кроме одной, равны нулю. Вследствие (3.24) этот способ кодирования наименее экономен с точки зрения полосы (или длительности), т. е. требует существенного увеличения пропускной способности канала по сравнению с натуральным представлением числа.

Другой способ позиционного кодирования, связанный с разрядным представлением числа, наиболее распространен. Этот способ несколько менее экономен по энергетике, но более экономен по числу степеней свободы:

$$M_a = (\ln a)^{-1} \ln \frac{1}{\varepsilon},$$

где a — основание позиционной системы счисления. Из предыдущего известно, что энергетически выгодней взять минимально возможное значение $a_{\min} = 2$. Тогда

$$M_2 = (\ln 2)^{-1} \ln \frac{1}{\varepsilon}, \quad (3.25)$$

и энергия сигнала, как и при (3.24), определяется требованиями, предъявляемыми к его надежному обнаружению. Вследствие нерегулярного (аномального ¹²) закона распределения ошибок (по сравнению с натуральным амплитудным кодированием (3.7)), связи между σ_0 , ε и w существенно отличаются от (3.18).

Пренебрегая коэффициентами порядка единицы, сразу найдем:

$$w \cong \varepsilon \cong \sigma_0. \quad (3.26)$$

Отсюда легко найти требования к вероятности ложных тревог w_- -ложного выброса и в любой позиции j , отличной от i -й (переданного сообщения).

Для однопозиционного кодирования, вследствие (3.24), (3.26), аналогично (2.35), получим

$$w = M_{\max} w_-, \quad w_- = \varepsilon^2. \quad (3.27)$$

Для двоичного разрядного представления (3.25) имеем:

$$w = M_2 w_-, \quad w_- = \frac{\varepsilon \ln 2}{\ln (1/\varepsilon)}. \quad (3.28)$$

Отсюда легко найти пороговую энергию для обоих случаев:

$$E_{\text{пор}} \approx kT \ln \frac{1}{w_-}, \quad (3.29)$$

и, положив $w_- = w_+$ (вероятность необнаружения), получим для энергии сигнала, представляющего число однопозиционным кодом (3.27):

$$E_c^{(0)} \approx 2E_{\text{пор}}^{(0)} = 4kT \ln \frac{1}{\varepsilon} \approx 4kTI. \quad (3.30)$$

Для двоичного разрядного представления сигналы (единицы) могут быть во всех M_2 (3.25) позициях. Считая, что в среднем в числе содержится половина единиц, получим:

$$E_p = \frac{1}{2} M_2 E_c^{(p)} = M_2 E_{\text{пор}}^{(p)} = kT \ln \frac{1}{\varepsilon} (\ln 2)^{-1} \left(\ln \frac{1}{\varepsilon} + \ln \ln \frac{1}{\varepsilon} + \ln \frac{1}{\ln 2} \right),$$

$$E_p \approx kT \left(\ln \frac{1}{\varepsilon} \right)^2 = kT (I)^2. \quad (3.31)$$

Видно, что при однопозиционном кодировании (3.30) энергетическая цена единицы информации

$$\hat{e}_0 = \frac{E_c^{(0)}}{I} \approx 4kT, \quad (3.32)$$

как и для обнаружения (2.36), по порядку величины совпадает с минимальной оценкой по ^{8, 21}. Однако, как уже упоминалось в разделах а) гл. 3 и б) гл. 2, после обнаружения при декодировании сигнала производится измерение с точностью $\sim 1/\varepsilon$, когда справедливы все энергетические оценки разделов б) гл. 2 и б) гл. 3. Существенно при этом, что энергия, затрачиваемая при декодировании на приемном конце, не влияет на выбор энергии сигнала (которая затухает в канале передачи).

Энергетическая цена единицы информации при разрядном представлении

$$e_D = \frac{E_D}{I} \approx kT \ln \frac{1}{\epsilon} \approx kTI \quad (3.33)$$

растет с увеличением I (количества разрядов $\sim \ln 1/\epsilon$). Как и при позиционном кодировании, для перехода к скалярной величине, т. е. при декодировании (дешифрации) здесь снова справедливы энергетические оценки разделов б) гл. 2 и б) гл. 3.

4. ТЕРМОДИНАМИЧЕСКИЕ МОДЕЛИ ПРОЦЕССА ОБРАБОТКИ ИНФОРМАЦИИ

а) К р и т е р и и с л о ж н о с т и о б р а б о т к и и н ф о р м а ц и и. В последнее десятилетие получила значительное развитие математическая теория сложности, отдельные направления которой достаточно полно освещены в монографической и обзорной литературе⁵⁸⁻⁶⁴. В работе⁶⁵ рассмотрены прикладные аспекты теории сложности с точки зрения задач теории управления. В связи с этим предложен⁶⁵ единый подход к различным направлениям теории сложности и проведена классификация этих направлений. Остановимся на двух основных классах критериев сложности^{65, 63}, имеющих непосредственное отношение к обработке информации.

Будем под обработкой информации понимать однозначное отображение Γ множества элементов $x \in X$ в множество элементов $\Gamma x = f \in X_f$, когда оба множества X , X_f являются подмножествами некоторого множества чисел. Неформально — в духе предыдущих разделов — будем представлять обработку информации как косвенное измерение. Пусть необходимо определить значение некоторой физической величины f , измерить которую непосредственно невозможно, но каким-то образом известно однозначное соответствие между ней и другой физической величиной x , доступной для непосредственного измерения. Тогда измерение x и дальнейшее преобразование соответствующей значению x физической величины в значение функции $f(x)$, представленное в таком же виде, и составляет косвенное измерение (обработку информации).

Следовательно, необходимо, во-первых, задать каким-то образом отображение Γ , а во-вторых, реализовать его, т. е. совершить соответствующее преобразование.

Первый класс критериев сложности^{65, 63} применительно к обработке информации характеризует сложность задания отображения Γ . Это может быть сложность текста, алгоритма (количество знаков) или схемы (количество элементов), в которых описаны правила перехода от объектов $x \in X$ к объектам $f \in X_f$. В частности, это может быть длина двоичного слова (программы), по которой некоторый автомат (машина) восстанавливает слово f из слова x . В работах Маркова⁶⁶ и Колмогорова^{67, 68} развита общая теория критериев сложности этого класса — применительно к любым вычислимым функциям. Существенно, что в⁶⁷ доказана инвариантность (с точностью до аддитивной константы) этих критериев сложности по отношению к автомату, осуществляющему дешифровку программы.

Заметим, что описанные критерии характеризуют сложность алгоритма (его записи), которая, вообще говоря, не связана со сложностью его реализации (т. е., например, длительностью вычислений). Эта сторона сложности обработки информации, т. е. трудоемкость вычислений по данному алгоритму, относится ко второму классу критериев^{63, 65} и характеризует сложность самого процесса реализации обработки. Здесь

вводятся сигнализирующие функции⁶³: временные или емкостные, характеризующие соответственно ресурсы времени (например, количество элементарных операций) или памяти, необходимые для реализации данного отображения Г. Существенно, что эти оценки зависят как от алгоритма, так и, вообще говоря, от машины (математической модели автомата), на которой реализуется алгоритм. Удовлетворительной машиннонезависимой теории критериев этого класса пока не существует^{63, 65}.

Вместе с тем было бы желательно оценивать вычислительную сложность (сложность процесса обработки информации) инвариантно по отношению к алгоритмам и машине, исходя лишь из самых общих закономерностей. Представляется, что для определения такой инвариантной характеристики сложности процесса обработки необходимо от математических моделей перейти к физическим. В предыдущих главах показано, что сложность процессов измерения и передачи информации однозначно характеризуется произведенной работой и рассеянной при этом энергией. Было бы заманчиво представить процесс обработки информации также как термодинамический процесс и характеризовать его сложность рассеянной энергией*). Изложим некоторые соображения на этот счет, кратко намеченные в^{24, 25}.

б) Энергетический критерий сложности процесса обработки информации. Существенное отличие математических моделей обработки информации от физических состоит в том, что первые оперируют с самого начала только с дискретными объектами (чаще всего, двоичными словами), тогда как вторые рассматривают макроскопические физические параметры, имеющие характер непрерывных величин. С другой стороны, физические модели явно вводят ограничения на энергетические ресурсы, лишь вследствие чего, согласно разделу б) гл. 3 возникает необходимость дискретизации (квантования) исследуемых объектов. Математические модели, напротив, исходят из гипотезы потенциальной осуществимости⁶⁴, которая, согласно⁶⁹, состоит в «отвлечении от реальных границ наших конструктивных возможностей, обусловленных ограниченностью нашей жизни в пространстве и во времени». Таким образом, практическая осуществимость⁶⁴ как раз и определяется физическими ограничениями, причем если речь идет о термодинамических ограничениях, то определяемые ими практические ограничения в значительно меньшей степени зависят от уровня развития техники, чем, например, ограничения по быстрдействию.

Остановимся теперь на основном отличии процессов измерения и передачи информации от процесса ее обработки. Если для первых энергетический критерий сложности определялся лишь одним критерием качества этих процессов — точностью, то для обработки это не так. Обработка информации (как, впрочем, и любая обработка — например, материалов) не может характеризоваться одной скалярной величиной, в отличие от передачи информации (или транспортировки материалов). Здесь существенно меняется качество (ценность) информации (или другого обрабатываемого продукта). Если нужно получить деталь не только заданного веса, но и определенной формы, то сложность такой детали определяется не только весом исходной заготовки, но и процессом ее обработки. Аналогично, сложность функции характеризуется не только ее точностью, но и видом и, соответственно, определяется не только точностью исходной информации, но и процессом ее обработки. Естественно, сложность обра-

*) Заметим в связи с этим, что даже в формальных математических постановках используется⁶³ термин «вычислительная работа» — конечно, не в термодинамическом смысле.

ботки детали зависит от технологии, так что в качестве безусловной характеристики должна быть принята минимальная (по всем мыслимым технологиям) сложность. Аналогично, сложность процесса обработки информации следует характеризовать минимальными (по всем мыслимым моделям обработки) энергетическими затратами. Представляется, что можно приблизиться к таким оценкам, рассмотрев лишь две термодинамические модели обработки информации²⁴.

Возвращаясь к трактовке процесса обработки информации как косвенного измерения*), представим его по схеме

$$x \rightarrow y \xrightarrow{\Gamma} z \rightarrow f. \quad (4.1)$$

Здесь x — измеренная физическая величина (или число, ей соответствующее), y — представляющий ее сигнал, z — в том же виде (что и y) представленный сигнал, соответствующий значению функции — числу $f(x)$. В (4.1) обозначения соответствуют (3.1) (а не (1.1)); поэтому, подобно (3.5), (3.10), y и z характеризуют амплитуду сигналов, а y^2 и z^2 — их энергию (см. (3.8), (3.9)).

В связи с недостаточностью скалярной характеристики для оценки качества обработки информации и в соответствии с общепринятым способом векторного представления функций с интегрируемым квадратом, будем представлять их в виде точек (векторов) в гильбертовом пространстве L_2 . Пространство L_2 следует выбрать потому, что, как было показано выше, средние энергетические затраты пропорциональны **)

$$\langle f^2 \rangle = \int_X p(x) f^2(x) dx = \|f\|^2. \quad (4.2)$$

Здесь $\|f\|$ — норма функции в L_2 , расстояние $\rho(f, g)$ между функциями f и g определяется так:

$$\rho^2(f, g) = \|f - g\|^2 = \int_X p(x) [f(x) - g(x)]^2 dx. \quad (4.3)$$

Как и ранее, будем считать, что требования к точности реализации $f(x)$ заданы. Из требования к точности $1/\sigma_{0f}$ (см. (1.13), (1.18), (1.24), (1.25)) легко найти требования к точности представления исходной переменной (непосредственно измеряемой величины x):

$$\sigma_{0f}^2 = \sigma_{0x}^2 \int_X p(x) \left(\frac{df}{dx} \right)^2 dx = \sigma_{0x}^2 \left\| \frac{df}{dx} \right\|^2. \quad (4.4)$$

Видно, что, вообще говоря, при $p(x) = \text{const}$

$$\frac{1}{\sigma_{0x}} \geq \frac{1}{\sigma_{0f}}, \quad (4.5)$$

что вполне соответствует интуитивному представлению о том, что исходного продукта должно быть количественно не меньше конечного (в данном случае — по количеству информации $I \sim \ln(1/\sigma)$). Неравенство (4.5) тем сильнее (см. (4.4)), чем больше осциллирует $f(x)$.

*) Аналогичная данной в разделе а) гл. 4 трактовка вычислительного процесса содержится в⁷⁰, где рассматриваются математические задачи теории (косвенных) измерений, связанные с приближенным решением некоторых интегральных уравнений на вычислительных машинах.

**) См. (3.4) и (3.8) — (3.11) для энергии сигналов. Рассеянная энергия Q , в соответствии с (2.5), (2.11), (2.12), также пропорциональна квадрату точности при необратимой (оперативной) реализации процессов. Переход к оценкам Q при оптимальном замедлении процессов легко осуществляется по формулам (2.16) — (2.18), (2.22), что мы, как правило, делать не будем.

Из (4.2) и предыдущего видно, что норма функции определяется только требованиями к точности $1/\sigma_{of}$ (или $1/\sigma_f$) — пропорциональна ей. Потребуем, чтобы кодирование исходной информации (3.5) было таким же, как и искомой, т. е. в (4.1)

$$z = Bf. \quad (4.6)$$

Это означает, что, вследствие (4.4), (4.5), исходная переменная должна быть пронормирована так, чтобы обеспечить более высокую точность при том же энергетическом коэффициенте A (3.10). Это сводится к увеличению диапазона x_m изменения x , т. е. для исходного сигнала следует положить

$$y = Bu, \quad u = cx, \quad c > 1, \quad (4.7)$$

$$0 \leq x \leq x_m, \quad 0 \leq u \leq u_m = cx_m.$$

Простейший способ реализации обработки информации по схеме (4.1) состоит в измерении x с необходимой по (4.4) точностью и в последующем функциональном преобразовании измеренной величины в искомую $f(x)$. Если этот второй этап реализуется функциональным преобразователем (например, потенциометром), то он полностью соответствует второму измерению. Этот этап может также соответствовать передаче информации (с последующим декодированием), если сигнал $y = Bcx$ использовать для запроса значений f из таблицы.

В последнем случае особенно наглядно представляется необходимость дискретизации x (и y) и квантования f (и z). Согласно (3.18), (3.19), могут быть выбраны энергетически оптимальные значения

$$\varepsilon_x = \sqrt{3} \sigma_{0x}, \quad \varepsilon_f = \sqrt{3} \sigma_{0f} \geq \varepsilon_x. \quad (4.8)$$

Таким образом, из этой модели обработки информации следует, что энергетическая сложность ее

$$S_x^f = S_f + S_x,$$

$$S_x^f \approx C \|f\|^2 \left(1 + \left\| \frac{df}{dx} \right\|^2\right), \quad (4.9)$$

где коэффициент пропорциональности C определяется на основании предыдущих глав. Легко видеть, что (4.9) не всегда определяет минимально необходимые энергетические затраты: так, если даже $f(x)$ весьма близка к x ($f \rightarrow x$, $df/dx \rightarrow 1$), рассмотренная модель требует удвоенных, по сравнению с непосредственным измерением, затрат. Близость функций (4.7), (4.6) характеризуется (4.3) расстоянием $\rho(f, u)$:

$$\rho^2(f, u) = \int_x p(x) [f(x) - cx]^2 dx. \quad (4.10)$$

Обозначив

$$f(x) - cx \equiv \varphi(x), \quad (4.11)$$

получим

$$\rho^2(f, u) = \|\varphi(x)\|^2 \quad (4.12)$$

и перейдем к описанию второй модели²⁴. Кажется очевидным, что при $\|\varphi\| \ll \|f\|$ целесообразно в устройстве обработки информации (УОИ) использовать энергию каждого i -го входного сигнала y_i , несколько изменив ее значение, т. е. реализовать энергетические скачки:

$$\zeta_i^2 = B^2 \varphi_i^2 = B^2 (f_i - cx_i)^2. \quad (4.13)$$

Легко понять, однако, что этой энергии недостаточно для обработки: требуется еще знать величину скачка φ_i , соответствующего каждому x_i на

описанная модель счета дает выражение для энергии, которую необходимо передать от одного тела другому. А эта энергия, как следует из предыдущего, полностью определяет значение диссипированной энергии (в необратимом случае просто равна ей). Таким образом, формула (4.18) представляет значение среднего энергетического скачка для заданной обработки. Среднее значение роста энтропии в системе пропорционально квадрату расстояния $\rho^2(f, u)$ (в необратимом случае) или расстоянию $\rho(f, u)$ (при оптимально замедленной реализации счета).

Возвращаясь к определению энергетической сложности $S_{x,f}$ обработки как косвенного измерения (4.15), (4.14), (4.9), замечаем, что всегда

$$S_{x,f} = S_f \frac{1}{r}, \quad r < 1, \quad (4.19)$$

где S_f — энергетическая сложность непосредственного измерения. Таким образом, коэффициент $1/r$ характеризует относительную сложность обработки: увеличение сложности воспроизведения $f(x)$ по косвенной информации об x , доступной для непосредственного измерения.

Можно попутно ввести также понятие «ценность исходной информации»²⁴, которую естественно характеризовать коэффициентом $r < 1$, обратным коэффициенту относительной сложности обработки *).

Как следует из рассмотренного выше, сложность вычислений тем больше (и ценность исходной информации тем меньше), чем больше расстояние $\rho(x, f)$ между векторами исходной и искомой функции.

Такое рассмотрение сложности вычислений и ценности исходной информации позволяет, как нам кажется, рассеять недоразумения, возникающие при обсуждении (в терминах скалярного представления информации) полезного эффекта работы вычислительной машины (см., например,⁸, стр. 349) и связанного с ним вопроса о сравнении информации, содержащейся в уравнении, с информацией, полученной при его решении. Вычислительная машина в процессе решения задачи, конечно, не увеличивает количества информации, а уменьшает его, одновременно увеличивая ценность информации.

в) О б э ф ф е к т и в н ы х м е т о д а х в ы ч и с л е н и й. В практических вычислениях функциональные преобразователи (или табличный способ реализации функций), соответствующие первой модели обработки раздела б) гл. 4, применяются достаточно редко. Кроме неуниверсальности подобных устройств, здесь существенно использование амплитудного кодирования, когда энергетическая сложность пропорциональна квадрату точности (в реальных случаях ограничения на время обработки). При $T = 300^\circ \text{K}$ число с точностью $10^9 \approx 2^{30}$ требует энергии $10^5 \text{ эрг} = 10^{-2} \text{ Дж}$. Именно поэтому для точных вычислений аналоговые способы не используются. Выше уже упоминалось, что столь (и более) высокая точность натурального представления чисел не только не требуется и практически не реализуема из энергетических соображений, но и ничему физически не соответствует, так как даже точность фундаментальных постоянных значительно меньше (см. также другие принципиальные соображения по этому поводу⁷⁴).

Требования к высокой точности возникают в многомерных задачах, когда число различных исходных ситуаций экспоненциально растет

*) Понятие «ценность информации», вне связи с вычислительным процессом, рассматривалось в^{33, 71-73} и др. работах, где это понятие предполагало более общий случай использования полученной информации: для достижения некоторой цели (управления). В рассматриваемом здесь случае цель состоит лишь в вычислении заданной функции.

с увеличением числа n измерений исходной области $X^{(n)}$ определения функции. В этом случае использование первой модели обработки затруднительно. Действительно, для этого требуется размерность n области $X^{(n)}$ входных переменных привести в соответствие с размерностью μ памяти. При этом энергетические затраты $S^{(n)} \sim nq^{n/\mu}$, где $q = 1/\epsilon$ — число различных ситуаций по каждой из переменных. Так как $\mu \leq 3$ (практически даже $\mu \leq 2$), то при больших n рост $S^{(n)}$ значителен.

Во второй модели обработка информации реализация энергетических скачков при переходе $y \rightarrow z$ (4.1) не обязательно связана с запросом из таблицы значений φ_i . Собственно счет характеризуется обычно последовательностью некоторых преобразований аргумента x и соответствующего ему сигнала y (например, умножений и сложений при реализации полинома). В этом случае переход от аргумента к функции происходит не по прямой, соединяющей соответствующие векторы $\{x_i\}$ и $\{f_i\}$, а по некоторой ломаной, которая тем ближе к прямой, чем шире набор элементарных операций. Тем не менее, этот переход по ломаной может оказаться энергетически выгодным, так как не требует обращения к таблице. В частности, в этом случае можно использовать (как это и делается в ЦВМ) двоичное разрядное представление чисел, которое энергетически существенно дешевле натурального *) (см. (3.34), (3.33)).

При разрядном кодировании энергия каждого данного числа в среднем одинакова (в отличие от амплитудного, где энергия пропорциональна квадрату значения x) и зависит лишь от разрядности (n). Поэтому все элементарные операции связаны с примерно одинаковыми затратами энергии (порядка n единичных скачков (3.33)), а энергетическая сложность вычислений естественным образом достаточно хорошо характеризуется количеством элементарных операций. При этом расстояние $\rho(f, u)$ по-прежнему определяет сложность, однако вследствие (3.31) энергетические затраты пропорциональны $(\ln \rho)^2$, а не ρ^2 , как при амплитудном кодировании.

Заметим, что определения и модели обработки информации, рассмотренные выше применительно к реализации функции $f(x)$, могут быть легко распространены на решение некоторой задачи (в частности, понятия сложности решения, ценности исходной информации и др.). Эффективность того или иного метода решения данной задачи — при заданных требованиях к точности решения — достаточно хорошо характеризуется числом элементарных операций. Вспомним, что эффективность второй модели обработки по сравнению с общим табличным способом (первая модель) определялась удачным выбором нулевого приближения к искомой функции. В данном случае в качестве нулевого приближения использовалась сама независимая переменная, а целесообразность такого способа определялась близостью по (4.10), которая соответствовала изменению требований к относительной точности для второго этапа (запроса из таблиц значений φ_i вместо f_i).

В связи с только что отмеченными обстоятельствами выскажем несколько качественных соображений о конструктивных методах в вычислительной математике. Применение хороших нулевых приближений особенно эффективно в многомерных задачах (как это, в частности, следует из рассмотренных выше моделей). Поэтому существенно иметь некоторые общие подходы к отысканию таких нулевых приближений. Если рассматривать некоторый класс задач, то для «почти всех» задач этого класса единственно возможным является некоторый общий и, как

*) Если при счете с разрядными числами приходится пользоваться таблицей, то требуется переход к натуральному числу (например, в дешифраторе), что влечет за собой увеличение энергетических затрат.

правило, крайне громоздкий по вычислениям метод. В частности, например, для решения граничных задач, описываемых уравнениями эллиптического типа, таким методом служит метод сеток.

С другой стороны, как правило, в каждом классе задач существуют «особые точки» во всей области этого класса, для которых известны (или могут быть предложены) частные, но весьма эффективные методы их решения. В частности, для упомянутого класса граничных задач предложен простой по вычислениям метод для областей, границы которых описываются уравнениями второго порядка (см. ^{75, 76}).

Если использовать решения, полученные для частных задач, в качестве нулевых приближений (главной части ⁷⁵) для решения более широкого класса задач, в каком-то смысле «близких» к этим частным задачам, то естественно ожидать существенной экономии в вычислительной работе. Так, например, для граничных задач ^{75, 76} даны критерии близости. Если при переходе далее от нулевого приближения к последующим использовать, например, тот же общий метод сеток, то окажется, что шаг сетки может быть выбран существенно большим. В работе ⁷⁷ показано, что эффективность такого рода нулевых приближений связана с существенным понижением требований к относительной точности остаточной части решения.

5. ЗАКЛЮЧЕНИЕ

Целый ряд физических задач требует проведения такой вычислительной работы, которую не может обеспечить современная вычислительная техника ⁷⁸. Аналогичные трудности возникают и в задачах планирования, управления и проектирования (см. ^{79, 80}). Интересно иметь оценки, принципиально ограничивающие возможности вычислительной техники. При этом желательно, чтобы такие оценки не были слишком завышены, подобно приводимым в литературе ⁸¹⁻⁸³ и характеризующимся количеством единиц обрабатываемой информации $\sim 10^{100}$. Эта цифра получена либо из подсчета числа нуклонов во Вселенной и количества событий на атомном уровне с момента образования земной коры ⁸¹, либо из оценки энергии Вселенной, эквивалентной ее массе и используемой для записи информации ^{82, 83}.

Из термодинамических соображений выше была определена энергетическая цена точности различного вида представления чисел, на основании чего можно получить более реальные оценки предельных возможностей вычислительной техники.

Требования к высокой точности реализуются при разрядном (в частности, двоичном) представлении чисел, когда энергия пропорциональна квадрату логарифма точности. Здесь высокая точность связана с большой глубиной вычислений и вызывает повышенные требования к надежности ^{14, 84-86}. В этом случае для тридцатиразрядного двоичного числа, если учесть требования к надежности, энергетическая цена двоичной единицы при $T = 300^\circ\text{K}$ имеет порядок 10^{-11} эрг. Анализ данных по логическим схемам современных технических устройств (электронных ⁸⁷ и оптических ⁸⁸) показывает, что в широком диапазоне частот (от 10^4 до 10^9 гц) энергия одного акта переключения примерно одинакова и имеет порядок 10^{-3} эрг. Таким образом, предельное значение на 8 порядков отличается от реально достигнутого. Считая, что это расхождение принципиально преодолимо, получим следующую оценку.

Если всю энергию, производимую на Земле в течение года ($\sim 10^{14}$ кет-час), затратить на обработку информации, то ее хватит на проведение 10^{36} ($300/T$) элементарных операций (типа сложения тридцати-

входе. В общем случае можно запросить значения φ_i из таблицы, используя часть энергии входного сигнала для запроса памяти (как и в первой модели), а другую часть — просуммировать с энергией скачка ξ_i^2 (4.13). Видно, что в такой модели (второй) обеспечивается непрерывный переход от обработки информации к измерению при $\|\varphi\| = \rho(f, u) \rightarrow 0$, а ее энергетическая сложность

$$S_x^{\varphi} = S_{\varphi} + S_x, \quad (4.14)$$

$$S_x^{\varphi} \approx C \|f\|^2 \left(1 + \frac{\|\varphi\|^2 + \|d\varphi/dx\|^2}{\|f\|^2} \right).$$

Общее определение энергетической сложности $S_{x,f}$ реализации функции $f(x)$, в соответствии с предыдущим и (4.13), (4.14):

$$S_{x,f} = \min \{S_x^f, S_x^{\varphi}\}. \quad (4.15)$$

То, что сложность перехода $y \rightarrow z$ в (4.1) действительно характеризуется расстоянием $\rho(f, u)$ (4.10)–(4.12), а ρ^2 пропорционально среднему энергетическому скачку (4.13), следует из модели счета ²⁴, составляющей центральную часть второй модели обработки.

Пусть УОИ заменено «черным ящиком», на вход которого поступает сигнал $y = Bcx$ (4.7). Если «черный ящик» помещен в термостат, то на сигнал накладываются аддитивные шумы со средней энергией kT , определяемой температурой T термостата. При $T > 0$ есть ненулевая вероятность того, что произвольное значение входа перейдет в любое другое значение, в том числе и в заданное обработкой $z = Bf$ (4.6). Чем выше эта вероятность, тем, по предположению, проще обработка.

Запишем выражение для плотности вероятности $w(f/u)$ перехода входных значений в выходные, исходя из гауссова закона флуктуаций (3.6), (3.7) и (4.13):

$$w(f/u) = \sqrt{\frac{B^2}{\pi}} e^{-B^2(f-cx)^2} = \sqrt{\frac{B^2}{\pi}} e^{-B^2\varphi^2(x)}$$

и перейдем, учитывая необходимость дискретизации значений входа и выхода, а также конечную точность воспроизведения функции, от плотности к конечным значениям вероятностей перехода точки cx_i в ε_f -отрезок:

$$W_i(f/u_i) = \frac{B}{\sqrt{\pi}} e^{-B^2\varphi_i^2(x)} \varepsilon_f. \quad (4.16)$$

Запишем далее среднее (по $P(x)$) значение логарифма этой вероятности:

$$\langle -\ln W(f/u) \rangle = \sum_{i=1}^{1/\varepsilon} p_i \left[B^2\varphi_i^2(x) + \ln \frac{\sqrt{\pi}}{B\varepsilon_f} \right], \quad (4.17)$$

где, в соответствии с (3.10), (3.18), $B\varepsilon = \text{const}$ (порядка единиц). Взяв отношение (4.17) к логарифму вероятности спонтанного появления единицы из нуля (т. е. исключая точность воспроизведения — коэффициент B) и устремляя $T \rightarrow 0$, получим:

$$\lim_{T \rightarrow 0} \frac{\langle -\ln W(f/u) \rangle}{\langle -\ln W(1/0) \rangle} = \rho^2(f, u) = \|\varphi(x)\|^2. \quad (4.18)$$

Известно ¹⁰, что показатель в экспоненте, описывающей вероятность соответствующей флуктуации, равен той минимальной работе, которую необходимо произвести над системой для перевода ее из исходного равновесного состояния в другое, определяемое флуктуацией. Таким образом,

разрядных двоичных чисел). Подчеркнем, что, несмотря на то, что приведенная оценка учитывает лишь тепловые шумы, она справедлива во всех случаях. Действительно, если при $kT \gg \hbar\nu$ оценка является точной, то при необходимости учета квантовых шумов она может быть принята в качестве верхней (по количеству операций). Эта оценка уже не столь сильно завышена, как приведенные выше, и показывает, что энергетические ресурсы реально и достаточно сильно ограничивают возможности ЦВМ современной структуры.

Чтобы конкретизировать это утверждение, сошлемся на пример ⁷⁸. Для квантовомеханического расчета молекулы метана требуется провести вычисления по методу сеток в 10^{42} точках. Если считать, что в каждой точке следует выполнить всего 10 элементарных операций, и предположить, что все вычисления производятся при сверхнизкой температуре ($T = 3 \cdot 10^{-3}$ °K), то и при этом расчет молекулы метана потребует израсходовать энергию, производимую на Земле примерно за столетие.

Аналогичные требования возникают в задаче расчета плазменных ловушек ⁷⁸ и в многочисленных задачах управления и планирования ^{79, 80}. Поэтому, наряду с увеличением производительности ЦВМ и наращиванием их парка, актуальным является поиск других (не арифметических) способов решения сложных задач.

Автор выражает благодарность В. Б. Брагинскому и Л. А. Ривлину за полезные обсуждения, способствовавшие улучшению настоящей статьи.

ЦИТИРОВАННАЯ ЛИТЕРАТУРА

1. Я. М. Гельфанд, История и методология термодинамики и статистической физики, М., «Высшая школа», 1969.
2. L. Szilard, Zs. Phys. 5, 840 (1929).
3. Р. Хартли, сборник «Теория информации и ее приложения», М., Физматгиз, 1959, стр. 5.
4. Н. Винер, Кибернетика, М., «Сов. радио», 1968.
5. К. Шеннон, Работы по теории информации и кибернетике, М., ИЛ, 1958.
6. Н. Винер, Кибернетика и общество, М., ИЛ, 1958.
7. А. Н. Колмогоров, сборник «Сессия АН СССР по научным проблемам автоматизации производства. Пленарные заседания», М., Изд-во АН СССР, 1957, стр. 66.
8. Л. Бриллюэн, Наука и теория информации, М., Физматгиз, 1960.
9. П. Шамбадал, Развитие и приложение понятия энтропии, М., «Наука», 1967.
10. Л. Д. Ландау, Е. М. Лифшиц, Статистическая физика, М., «Наука», 1964.
11. Дж. фон Нейман, Математические основы квантовой механики, М., «Наука», 1964.
12. А. А. Харкевич, Очерки общей теории связи, М., Гостехиздат, 1955.
13. D. Gabor, Phil. Mag. 41, 1161 (1950).
14. Дж. фон Нейман, Теория самовоспроизводящихся автоматов, М., «Мир», 1971.
15. Ю. А. Шрейдер, сборник «Информация и кибернетика», М., «Сов. радио», 1967, стр. 15.
16. Л. Бриллюэн, Научная неопределенность и информация, М., «Мир», 1966.
17. П. В. Новицкий, Основы информационной теории измерительных устройств, М., «Энергия», 1968.
18. В. И. Рабинович, М. П. Цапенко, Информационные характеристики средств измерения и контроля, М., «Энергия», 1968.
19. В. Ю. Урбах, сборник «Проблемы кибернетики», вып. 10, М., Физматгиз, 1963, стр. 99.
20. H. Reiss, J. Statist. Phys. 1, 107 (1969).
21. Д. С. Лебедев, Л. Б. Левитин, сборник «Теория передачи информации», М., «Наука», 1964, стр. 5.
22. Л. Б. Левитин, сборник «IV конференция по теории передачи и кодирования информации. Секция II», Москва — Ташкент, АН СССР, 1969, стр. 111.
23. Р. П. Поплавский, ДАН СССР 202, 562 (1972).
24. Р. П. Поплавский, цит. в ²² сборник, секция I, стр. 209.

25. Р. П. Поплавский, сборник «V Всесоюзный симпозиум по кибернетике», Тбилиси, АН СССР, 1970, стр. 45.
26. Я. И. Френкель, Статистическая физика, М.—Л., Изд-во АН СССР, 1948.
27. М. А. Леонтович, Статистическая физика, М.—Л., Гостехиздат, 1944.
28. Г. Крамер, Математические методы статистики, М., ИЛ, 1948.
29. М. И. Корнфельд, УФН 85, 533 (1965).
30. В. Феллер, Введение в теорию вероятностей и ее приложения, т. 2, М., «Мир», 1967.
31. Р. Фано, Передача информации, М., «Мир», 1965.
32. С. Голдман, Теория информации, М., ИЛ, 1957.
33. И. А. Полетаев, сборник «Исследования по кибернетике», М., «Сов. радио», 1970, стр. 228.
34. R. C. Raupond, Am. Sci. 38, 273 (1950).
35. D. A. Bell, *ibid.* 40, 682 (1952).
36. М. М. Волькенштейн, Молекулы и жизнь, М., «Наука», 1965.
37. Г. И. Кавалеров, С. М. Мандельштам, Введение в информационную теорию измерений, М., «Энергия», 1974.
38. Л. И. Мандельштам, Полн. собр. трудов, т. V, М.—Л., Изд-во АН СССР, 1950.
39. Д. Бом, Квантовая теория, М., Физматгиз, 1961.
40. Д. М. Блохинцев, Принципиальные вопросы квантовой механики, М., «Наука», 1966.
41. Д. И. Блохинцев, УФН 95, 75 (1968).
42. А. С. Холево, Проблемы передачи информации 9 (2), 31 (1973).
43. А. С. Холево, *ibid.*, № 3, 3.
44. Г. Уленбек, УФН 103, 275 (1971).
45. В. А. Котельников, Проблемы помехоустойчивости радиосвязи («Радиотехнический сборник»), М., Госэнергоиздат, 1947.
46. В. А. Котельников, Теория потенциальной помехоустойчивости, М., Госэнергоиздат, 1956.
47. Ф. Вудворт, И. Дэвис, сборник «Теория передачи электрических сигналов при наличии помех», М., ИЛ, 1953, стр. 239.
48. В. Петерсон, Т. Бердсал, В. Фокс, сборник «Теория информации и ее приложения», М., Физматгиз, 1959, стр. 210.
49. Д. Ван-Митер, Д. Мидлтон, сборник «Прием сигналов при наличии шума», М., ИЛ, 1960, стр. 57.
50. А. А. Харкевич, Борьба с помехами, М., «Наука», 1965; С. Е. Фалькевич, Оценка параметров сигнала, М., «Сов. радио», 1970.
51. В. Б. Брагинский, Физические эксперименты с пробными телами, М., «Наука», 1970.
52. В. Б. Брагинский, Ю. И. Воронцов, УФН 114, 41 (1974).
53. Р. Мёссбауэр, УФН 72, 658 (1960); Р. Паунд, *ibid.*, стр. 673.
54. Дж. Тригг, Решающие эксперименты в современной физике, М., «Мир», 1974.
55. Д. Вик, УФН, 101, 303 (1970).
56. Б. Тейлор, Д. Лангенберг, У. Паркер, УФН 105, 575 (1971).
57. Основные формулы физики, под ред. Д. Мензела, тт. 1—2, М., ИЛ, 1957.
58. С. В. Яблонский, сборник «Информационные материалы», вып. 5 (42), М., Изд. Научного совета АН СССР по комплексной проблеме «Кибернетика», 1970, стр. 5.
59. О. Б. Лупанов, *ibid.*, стр. 16.
60. Б. А. Трахтенброт, *ibid.*, стр. 61.
61. А. К. Звонкин, Л. А. Левин, УМН 25 (6), 85 (1970).
62. А. Г. Витушкин, Оценка сложности задачи табулирования, М., Физматгиз, 1959.
63. Б. А. Трахтенброт, Сложность алгоритмов и вычислений, Новосибирск, Изд-во НГУ, 1967.
64. Б. А. Трахтенброт, Алгоритмы и вычислительные автоматы, М., «Сов. радио», 1974.
65. Д. Б. Юдин, А. П. Горяшко, Изв. АН СССР (Техническая кибернетика), № 3, 34 (1974).
66. А. А. Марков, ДАН СССР 157, 262 (1964).
67. А. Н. Колмогоров, Проблемы передачи информации 1 (1), 3 (1965).
68. А. Н. Колмогоров, *ibid.* 5 (3), 3 (1969).
69. А. А. Марков, Тр. МИАН СССР, 42 (1954).
70. В. Я. Арсенин, Тр. Ленингр. отделения МИАН СССР 83, 33 (1973).
71. А. А. Харкевич, цит. в¹⁹ сборник, вып. 4, 1960, стр. 53.
72. М. М. Бонгард, *ibid.*, вып. 9, 1963, стр. 71.
73. Р. Л. Стратонович, Изв. АН СССР (Техническая кибернетика), № 5, 3 (1965).
74. П. К. Рашевский, УМН 28 (4), 156 (1973).

75. Р. П. Поплавский, сборник «Вычислительная математика», вып. 6, М., Изд-во АН СССР, 1960, стр. 63.
 76. Р. П. Поплавский, ПММ 24, 387 (1960).
 77. Р. П. Поплавский, сборник «Исследования по современным проблемам конструктивной теории функций», М., Физматгиз, 1961, стр. 358.
 78. А. Шлютер, УФН 106, 119 (1972).
 79. Д. Б. Юдин, Автоматика и телемеханика 5, 47 (1972).
 80. Д. Б. Юдин, Изв. вузов (Радиофизика) 15, 957 (1972).
 81. У. Эшбп, сборник «Общая теория систем», М., «Мир», 1966, стр. 171.
 82. Н. J. Времерманн, Proc. of the Conference on Self-Organizing Systems, Washington, Spartan, 1962, p. 93.
 83. Р. Ландауэр, УФН 106, 125 (1972).
 84. Дж. фон Нейман, сборник «Автоматы», М., ИЛ, 1956, стр. 68.
 85. Дж. фон Нейман, в книге Тьюринга «Может ли машина мыслить», М., Физматгиз, 1960, стр. 59.
 86. Дж. фон Нейман, Кибернетический сборник, вып. 1, М., ИЛ, 1960, стр. 11.
 87. Р. Дайли, Г. Кунцлеман, Заруб. радиоэлектрон, вып. 6, 39 (1971).
 88. Н. Г. Басов, В. В. Никитин, А. С. Семенов, УФН 97, 561 (1969).
-