Journal of Statistical Physics, Vol. 1, No. 1, 1969

Thermodynamic-Like Transformations in Information Theory

H. Reiss[1]

Received March 7, 1969

This article explores the application of thermodynamic and statistical thermodynamic formalism to information theory problems. In particular, the applicability of the transformation theory of thermodynamics is investigated. After a brief tutorial discussion of thermodynamic and statistical thermodynamic methods and concepts, their information theory analogues are developed. Besides information theory entropy, the information theory counterparts of temperature, chemical potential, Helmholtz free energy, etc., are developed and related to conventional information theory concepts such as channel capacity, matching of source and channel, etc. Information theory theorems are proved via the statistical thermodynamic analogue method; and, finally, several problems are formulated and solved using thermodynamic-like transformations. This article is aimed chiefly at bridging the interface between the two disciplines, and is intended to be provocative. Therefore, no attempt has been made to have it be all inclusive.

KEY WORDS: Information theory; Thermodynamic transformations; Partition function; Statistical thermodynamics; Channel capacity; Matching; Noiseless channel; Compact code.

1. USEFULNESS OF THE INTERDISCIPLINARY APPROACH

The similarity between certain ideas from information theory and those originating in statistical mechanics has been recognized for some time. The most famous of these correspondences involves the information theory *entropy*, which bears a great deal of resemblance to the quantity of the same name employed in statistical thermodynamics. Information theorists have been quick to point out, however, that the similarity is, in many respects, only formal and the *real* tasks of information theory are not measurably lightened by the awareness of this correspondence.

In spite of this disclaimer, several distinguished authors have attempted to

---

[1] Department of Chemistry, University of California, Los Angeles.

combine information theory and physics in various ways. It is convenient to classify these attempts under a number of descriptive headings which serve to characterize unique *modes* of interdisciplinary inquiry.

1.1. Direct Combination

In this mode, which is best typified by the discussions of L. Brillouin,(1) the entropies of information theory and of thermodynamics are considered to be one and the same. Thus in making a *physical* observation the entropy of the universe is always increased, the smallest increase possible (for a 50% chance of having "meter" deflection *not* due to Brownian motion) being $k \ln 2$, where k is Boltzmann's constant. In performing a measurement on a given system, the entropy of *that* system may be decreased since after the measurement it may be in a more well defined state, but the process of measurement is an irreversible one (in the thermodynamic sense) and must be accompanied by an overall increase in the entropy of the universe. By the same line of reasoning, the process of communication in which information is transmitted from a source "system" to a receiving "environment" corresponds to a measurement on the state of the "source" and requires an overall increase in entropy. Energy will be expended in sending the message and the energy–entropy balance can be analyzed in the *physical* thermodynamic sense. Frequently, however, the *information* entropy effects are so small ($k \ln 2$, for example) that although they can be *directly combined* with those due to *physical* entropy, such combination has little value beyond rounding out the internal consistency of the theory.

1.2. Foundations of Statistical Mechanics

In this mode, of which E. T. Jaynes is the chief proponent,(2) the concepts of information theory are substituted for other formulations (for example, the Gibbsian ensemble approach) in order to provide a basis for the discipline of statistical mechanics. The various distributions are exhibited as those which are *least biased*, subject to whatever "information" (constraints) is imposed on the mechanical system. Although this formulation leads to the same relationships as those derived in the conventional approach, it has not been widely adopted, probably because it relinquishes the goal of exhibiting *thermodynamics* as a branch of mechanics; an idea which has consumed chemists and physicists for nearly a century. On the other hand, Jaynes' approach does possess the merit of being self-contained, intuitively pleasing, didactically simple, and, by its very nature, minimally biased.

1.3. Manipulations of Information Theory

A third mode involves the use of the manipulative procedures of statistical thermodynamics as a tool in performing the tasks of information theory itself. Thus the mathematical procedures, the "folklore" of statistical mechanics, are applied wherever possible to lighten the burden of formal manipulation within information theory. In some respects this mode is the inverse of that described above, and it is not surprising that Jaynes has produced what appears to be the only published paper$^{(3)}$ on the subject. This is a very interesting attempt, but for reasons which we shall try to analyze below, the work seems not to have been widely noticed. This is a pity because Jaynes' paper is both first class and provocative. It should be pointed out, however, that the use of this approach does not produce anything which cannot in principle be derived by the standard techniques of information theory. It merely provides an alternative and perhaps, occasionally, more concise method of derivation.

Of the three categories discussed above, the first two represent blendings of physics and information theory in the interest of obtaining additional insight into *physical* phenomena. The last category, however, is conceptual and mathematical. Whether or not entropy, as used in information theory, admits of any physical interpretation is beside the point. Within information theory itself—with Shannon's first theorem,$^{(4)}$ for example—this is the essential view. According to this theorem, the entropy per symbol merely defines the average code-word length for a *compact code.*$^{(5)}$ In Jaynes' paper this point of view is conserved, and any functions or methods reminiscent of statistical thermodynamics are only formally so.

There are, however, several possible advantages (some of which were mentioned by Jaynes) in the interdisciplinary approach. These are:

1. There are mathematical methods, sometimes motivated by physical intuition, which exist in statistical thermodynamics which may be suitable but which have not yet been exploited for information theory.
2. There may be concepts and definitions, useful in statistical mechanics, which are valuable in information theory.
3. By translating information theory into a language familiar to statistical mechanicians, and the reverse, a large group of interested persons may be stimulated to attack problems.
4. The establishment of common languages will permit dialogue which will keep each discipline aware of relevant developments in the other.

In the opinion of the present writer, Jaynes' article did not receive its proper share of attention for the following reasons. In the first place it was brief, only a "note" according to its title. The title itself referred to "unique decipherability" and did not call attention to the use of thermodynamic formalism. Furthermore, it was published in the *Transactions* of the IRE and has probably been exposed to an audience many of whom, in view of their field of specialization, are not deeply familiar with the field of statistical thermodynamics. In spite of this last condition, Jaynes did not provide a tutorial preface on statistical thermodynamics. Finally, Jaynes himself seems not to have pursued the subject or to have publicized his point of view beyond the effort represented by his single article.

Nevertheless, the present writer feels there is considerable merit in Jaynes' original idea. The central purpose of the *Journal of Statistical Physics* is the bridging of interfaces between disciplines, and so it is quite appropriate to attempt to develop Jaynes' theme further, and to provide space for a tutorial preface which will assist the information theorist to gain insight into some of the manipulative philosophy of statistical thermodynamics. Attention is focused on helping the *information theorist* rather than the statistical thermodynamicist since the techniques to be introduced are used for the solution of *information theory* problems, where they may be novel, rather than for statistical mechanics, where they are common. This does not imply that the techniques of *information theory* have nothing to offer statistical mechanicians (or for that matter, physicists in general). They do, and in fact this has been the more usual direction of transfer of methodology, falling under the headings of the first two categories above. In this article we endeavor to reverse the flow.

Before addressing the main task, however, the author feels compelled to make a qualifying statement. This is that he knows very little about information theory and more about statistical thermodynamics. As a result, his judgments concerning what is important in information theory (or for that matter even what is *right*) may leave something to be desired. On the other hand, perhaps this is an undersubscribed method for penetrating an interface—having an expert in one field apply his special skills to the solution of problems in another, rather than the reverse in which an expert in one field attempts to adopt the skills of another for the solution of problems in his own field.

2. THE TRANSFORMATION THEORY OF THERMODYNAMICS

Strictly speaking, Jaynes(3) concentrated on using the *partition function* (or its analogue) of statistical mechanics in the study of an information theory problem which he called "semi-optimal transmission." Although most uses of the partition function imply the *transformation theory* of thermodynamics, Jaynes did not address this body of knowledge explicitly; and in this paper, besides commenting on Jaynes' work, we shall examine how such transformations may be used in information theory. Before doing this, it is useful to undertake a brief review of certain manipulative aspects of thermodynamics.(6)

One of the most powerful features of thermodynamics is the service it provides in making possible the transformation of one set of experimental data on the macroscopic behavior of a system (at equilibrium) into another set. Thus the proper use of thermodynamics may avoid the performance of redundant measurements. For example, if the coefficient of thermal expansion, the isothermal compressibility, the heat capacity at constant volume, and the temperature and volume of a substance have been measured, it is unnecessary to measure the *heat capacity at constant pressure*. It may be calculated from the measured values of the prior properties by means of thermodynamic transformation. This transformation is independent of the microscopic (atomic and molecular) structure of the substance involved.

The transformation is based on the theory of *exact* differentials.(7) Some of the important relations involving exact differentials follow.

If $f(x, y, z)$ is an analytic function of x, y, and z (so that df is an *exact* differential), and we have

$$df = X\,dx + Y\,dy + Z\,dz \tag{1}$$

where X, Y, and Z are also analytic functions of x, y, and z, then it must be true that

$$\begin{aligned} X &= \left(\frac{\partial f}{\partial x}\right)_{y,z} \\ Y &= \left(\frac{\partial f}{\partial y}\right)_{x,z} \\ Z &= \left(\frac{\partial f}{\partial y}\right)_{x,y} \end{aligned} \tag{2}$$

and furthermore

$$\begin{aligned} \left(\frac{\partial X}{\partial y}\right)_{x,z} &= \left(\frac{\partial Y}{\partial x}\right)_{y,z} \\ \left(\frac{\partial Y}{\partial z}\right)_{x,z} &= \left(\frac{\partial Z}{\partial y}\right)_{x,z} \\ \left(\frac{\partial X}{\partial z}\right)_{x,y} &= \left(\frac{\partial Z}{\partial x}\right)_{y,x} \end{aligned} \tag{3}$$

Still another useful relation is the following:

$$\left(\frac{\partial f}{\partial x}\right)_{y,z} = -\left(\frac{\partial f}{\partial y}\right)_{x,z}\left(\frac{\partial y}{\partial x}\right)_{y,z} \tag{4}$$

Several others exist, but it is not necessary to document all of them here.

The essential roles of the *laws* of thermodynamics are to postulate the existence of functions which depend only on the variables which characterize the thermodynamic state and whose differentials with respect to these variables are therefore exact. If in Eqs. (1)–(4) f is such a function and x, y, and z are thermodynamic variables of state, then relations such as Eqs. (2)–(4) represent transformations among thermodynamic variables. Thus the laws of thermodynamics make the transformations possible.

For example, the *zeroth* law of thermodynamics$^{(8)}$ postulates the existence of the *temperature* function of state, and prescribes a method for its measurement. The *first* law$^{(9)}$ postulates the existence of a function of state U called the *internal energy* and prescribes a method for its measurement. The prescription in this case is the following. One encloses the system in an adiabatic shell (one which does not permit the system to exchange energy in the form of *heat* with its surroundings) and causes it to undergo a change of state by performing a certain amount of work $\Delta\omega_a$ on the system. Then the change in U is given by

$$\Delta U = \Delta\omega_a \tag{5}$$

In this way U can be determined to within a constant term for any system. Since only changes in U are important in thermodynamics, it is unnecessary to specify the precise value of this constant.

If the same change of state is conducted without the adiabatic shell, the work may be represented by $\Delta\omega$ rather than $\Delta\omega_a$, and in general it is found that

$$\Delta U \neq \Delta\omega \tag{6}$$

(ΔU having been determined previously by the adiabatic process). The equality can be repaired by adding a compensating term Δq in Eq. (6). Thus

$$\Delta U = \Delta\omega + \Delta q \tag{7}$$

where Δq is referred to as the *heat* absorbed by the system during the change.

The *second* law of thermodynamics(10) postulates the existence of a function of state S called the entropy, and also prescribes a method of measurement. Thus one measures the heat absorbed by the system during a change of state conducted in a thermodynamically *reversible*(6) manner. If this heat is absorbed at one temperature T, then

$$\Delta S = \Delta q/T \tag{8}$$

Again the determination is to within a constant. If the process is not isothermal, then one may consider infinitesimal steps and write

$$dS = Dq/T \tag{9}$$

The total change in entropy is then computed by integration. D is used on the right of Eq. (9) to indicate that Dq, unlike dS, is not an exact differential.

Writing Eq. (7) in differential form and substituting Eq. (9) yields

$$dU = T\,dS + D\omega \tag{10}$$

If, for example, work can be performed on the system only by changing its volume V,

$$D\omega = -p\,dV \tag{11}$$

where p is the *equilibrium* pressure (because, since $T\,dS$ is the *reversible* heat, $D\omega$ must be the *reversible* work) and Eq. (10) becomes

$$dU = T\,dS - p\,dV \tag{12}$$

The power of the thermodynamic method depends upon the number of quantities which can be demonstrated to be state functions. Besides the *primary* functions established by the laws of thermodynamics, it is possible to synthesize *derived* state functions by combining quantities already known to be state functions. Familiar examples of such *derived* functions are the *enthalpy*,

$$H = U + pV \tag{13}$$

the *Helmholtz free energy*,

$$A = U - TS \tag{14}$$

and the *Gibbs free energy,*

$$G = U + pV - TS \tag{15}$$

The application of Eqs. (2) and (3) to (12) yields the interesting transformation

$$\begin{gathered} T = \left(\frac{\partial U}{\partial S}\right)_V, \qquad p = -\left(\frac{\partial U}{\partial V}\right)_S \\ \left(\frac{\partial T}{\partial V}\right)_S = -\left(\frac{\partial p}{\partial S}\right)_V \end{gathered} \tag{16}$$

Similar transformations may be obtained by taking the total differentials of H, A and G, substituting Eq. (12), and, finally, applying Eqs. (2) and (3). Thus

$$dH = dU + p\,dV + V\,dp \tag{17}$$

which, upon substitution of Eq. (12), becomes

$$dH = T\,dS + V\,dp \tag{18}$$

Similarly we find

$$dA = -p\,dV - S\,dT \tag{19}$$

and

$$dG = V\,dp - S\,dT \tag{20}$$

From these relations we develop the analogues of Eq. (16):

$$\begin{gathered} T = \left(\frac{\partial H}{\partial S}\right)_p, \qquad V = \left(\frac{\partial H}{\partial p}\right)_S \\ \left(\frac{\partial T}{\partial p}\right)_S = \left(\frac{\partial V}{\partial S}\right)_p \end{gathered} \tag{21}$$

$$\begin{gathered} p = -\left(\frac{\partial A}{\partial V}\right)_T, \qquad S = -\left(\frac{\partial A}{\partial T}\right)_V \\ \left(\frac{\partial p}{\partial T}\right)_V = \left(\frac{\partial S}{\partial V}\right)_T \end{gathered} \tag{22}$$

$$\begin{gathered} V = \left(\frac{\partial G}{\partial p}\right)_T, \qquad S = -\left(\frac{\partial G}{\partial T}\right)_p \\ \left(\frac{\partial V}{\partial T}\right)_p = -\left(\frac{\partial S}{\partial p}\right)_T \end{gathered} \tag{23}$$

Clearly many other transformations between sets of thermodynamic data are possible. What bearing, however, does this methodology have on information theory? To understand this it is necessary to introduce a few concepts from statistical mechanics and to consider the "molecular theory" analogues of these transformation relations. We accomplish this in the next section.

3. STATISTICAL REPRESENTATION OF THERMODYNAMIC FUNCTIONS

The basic idea of statistical mechanics is to represent thermodynamic quantities as certain averages taken over the molecular or microscopic mechanical states of the system.(11,12) In the final analysis these microscopic states are considered to be nonclassical (quantum) states, and the probability of observing the system to be in the ith state is denoted by P_i. This probability is conditioned by the particular "constraints" imposed on the system. Knowledge of such constraints is in fact equivalent to the possession of "information" on the behavior of the system. They make take the form, for example, of statements concerning its volume or energy and are frequently more subtle than this. In information theory terms, the method of establishing the functional form of P_i is equivalent to maximizing the uncertainty, subject to whatever information (constraints) is already available. In essence, this corresponds to Jaynes' method(2) for identifying the least biased distribution.

Notwithstanding the pragmatic need, at the moment, to determine P_i in this manner, the goal of the physicist still remains; namely, the derivation of P_i from purely mechanical considerations. The Gibbsian ensemble technique to which Jaynes' method (if not his philosophy) is really equivalent does not accomplish this.

In the so-called *canonical* ensemble in which the number of molecules, the temperature, and the volume of the system are constrained to remain constant, it is possible to show(12) that

$$P_i = e^{-E_i/kT}/Q \tag{24}$$

where E_i is the energy of the ith quantum state, k is Boltzmann's constant, T is the temperature; and Q, the *partition* function, is given by

$$Q = \sum_i e^{-E_i/kT} \tag{25}$$

where the sum in Eq. (25) is over all quantum states accessible to the system.

E_i may depend upon the constrained geometric variable V as well as the number of molecules N. In each quantum state the mean rate of transfer of momentum (per unit area) to the walls of the containing vessel represents the "pressure" in that state, and from mechanical considerations is given by

$$p_i = -\frac{\partial E_i}{\partial V} \tag{26}$$

The thermodynamic pressure is then specified by the *ensemble average*

$$p = \sum_i p_i P_i \tag{27}$$

and the termodynamic *internal* energy by the ensemble average

$$U = \sum_i P_i E_i \tag{28}$$

Taking the total differential of U in Eq. (28) yields

$$\begin{aligned} dU &= \sum_i E_i dP_i + \sum_i P_i dE_i \\ &= \sum_i E_i dP_i + \sum_i P_i \frac{\partial E_i}{\partial V} dV \end{aligned} \tag{29}$$

In this equation we have assumed N to remain constant. Substituting from Eqs. (26) and (27) gives

$$dU = \sum_i E_i dP_i - p dV \tag{30}$$

Now we may employ Eq. (24) to write

$$E_i = -kT(\ln P_i + \ln Q) \tag{31}$$

which can be substituted into Eq. (30) to give

$$dU = -kT \sum_i \ln P_i dP_i - p dV \tag{32}$$

where we have used the relation

$$\sum_i dP_i = 0 \tag{33}$$

Actually Eq. (33) shows that

$$d \sum_i P_i \ln P_i = \sum_i \ln P_i dP_i \tag{34}$$

and Eq. (34) may be substituted into (33) to yield

$$dU = Td \left\{ -k \sum_i P_i \ln P_i \right\} - p dV \tag{35}$$

which by comparison with Eq. (12) identifies the entropy as

$$S = -k \sum_i P_i \ln P_i \tag{36}$$

All the thermodynamic functions are now easily derivable. For example, suppose we wish to determine A. Then

$$A = U - TS = \sum_i P_i E_i + kT \sum_i P_i \ln P_i \tag{37}$$

Substituting for E_i from Eq. (31), we obtain

$$A = -kT \sum_i (P_i \ln P_i + P_i \ln Q) + kT \sum_i P_i \ln P_i = -kT \ln Q \tag{38}$$

since

$$\sum_i P_i = 1 \tag{39}$$

Equation (38) is a standard result of statistical thermodynamics, exhibiting the connection between the partition function and the thermodynamic properties in its simplest form.

If we wished to compute the pressure p, we might elect to use Eq. (27) and to perform the average over microscopic states once again. It is much simpler, however, to employ the transformation theory of thermodynamics and to make use of Eq. (22). Thus we may write

$$\begin{aligned} p &= -\left(\frac{\partial A}{\partial V}\right)_T = kT\left(\frac{\partial \ln Q}{\partial V}\right)_T \\ &= kT\left[\frac{\partial}{\partial V}\sum_i e^{-E_i/kT}\right] \\ &= -\frac{1}{Q}\sum_i \frac{\partial E_i}{\partial V} e^{-E_i/kT} \\ &= \sum_i p_i P_i \end{aligned} \tag{40}$$

The last three steps are included to prove the validity of the first two by recovering the proper ensemble average.

What we have managed to do is to find a set of ensemble averages which are supposed to represent thermodynamic quantities and which in fact are *related to each other like thermodynamic quantities.* They therefore transform like thermodynamic quantities so that with a few of them at our disposal it is not necessary to perform averaging processes again in order to obtain the others. We need merely use the transformation theory of thermodynamics.

This is the manipulative feature which we wish to carry over to information theory. Many of the quantities of importance to information theory are carefully chosen averages. We shall show that these are also related to each other like thermodynamic quantities. (Note that this does not necessarily mean that they *are* thermodynamic quantities; most of the time they will not be.) As a result, we may obtain certain important averages from others without having to perform the averaging process anew each time. Instead, we merely invoke the appropriate thermodynamic-like transformation. We address this subject in the following sections.

4. INFORMATION THEORY RELATIONS DERIVED BY THE METHODS OF STATISTICAL THERMODYNAMICS

In this section we derive certain important information theory theorems by the methods of statistical thermodynamics. We shall also define certain information theory equivalents of thermodynamic quantities. Consider then a noiseless channel(13) coupled to a zero-memory source(14) through which we transmit messages, each message requiring a time T for its transmission.

The *capacity* C of the channel measured in *bits* per unit time is defined as

$$C = \lim_{T\to\infty} \frac{1}{\ln 2}\frac{\ln M(T)}{T} \tag{41}$$

where $M(T)$ is the number of distinct messages which can be transmitted in time T. In what follows we shall drop the lim symbol, and it will be understood that T and other quantities related to it in scale will always be very large. (This resembles the custom in statistical mechanics, where, unless otherwise specified, the number of molecules in a thermodynamic system is considered essentially infinite.)

Suppose the words "emitted" by the source are coded in the channel in terms of code symbols for which the average time of transmission per symbol is σ. Then we would expect each message to contain

$$L = T/\sigma \tag{42}$$

symbols. The largest possible number of messages in time T occurs when there are absolutely no constraints, so that the transmission of one symbol in no way limits the symbol which follows it. Then if the alphabet of code symbols contains r such symbols,

$$M(T) = r^L = r^{T/\sigma} \tag{43}$$

from which we compute, using Eq. (41),

$$C^*(r, \sigma) = \frac{\ln r}{\sigma \ln 2} \tag{44}$$

where $C^*(r, \sigma)$ is clearly the maximum capacity possible for given r and σ.

Usually the capacity of the channel will fall far below C^* because there *will* be constraints present (for example, the need to bind the symbols into words) which reduce the possible number of messages. In any event, it is wasteful if the channel is coupled to a source which emits information more rapidly than the channel can handle it, and the problem is to "match" channel and source so that capacity is not *underused* nor information *dissipated* (because it cannot be handled).

We now begin an attack on this problem, using methods which are common to statistical thermodynamics. In order to start with the simplest possible case we consider a situation which will permit us to develop the information theory analogue of the *microcanonical* ensemble,(12) in which a thermodynamic system is constrained to have constant volume, molecular content, and energy. To construct the analogue it is necessary to impose still another constraint on our channel over and above the restriction which binds symbols into words, namely, to limit the number of *words* in each message to be exactly W. This constraint corresponds to what Jaynes(3) has called "semi-optimal" transmission, and determines the mean transmission time per word as

$$t = T/W \tag{45}$$

We assume that the channel possesses a "spectrum" of transmission times for code words such that time t_i is required for the transmission of the ith word. The spectrum of times is conveniently denoted by the *formal* vector $\mathbf{t}$ whose components are the numbers t_i. We wish to answer the following questions.

What assignment of word probabilities P_i for the source will allow us to use the channel at full capacity; and alternatively, given the assignments P_i, what "spectrum" t_i will achieve the same result?

To simplify matters, we shall restrict our considerations to "nondegenerate" word time distributions, i.e., to distributions in which no two words have the same time. No essential complication would result from the inclusion of degeneracy. If, then, messages are considered which contain n_i words of type i, there will be

$$\Omega(\mathbf{n}) = W!/\prod_i n_i! \tag{46}$$

such messages possible, where $\mathbf{n}$ is a *formal* vector (denoting a particular distribution of n_i) whose components are the numbers n_i. The total number of messages is then obtained by summing over $\mathbf{n}$ subject to the obvious constraints

$$\sum_i n_i = W \tag{47}$$

$$\sum_i t_i n_i = T \tag{48}$$

Thus

$$M(T, W) = \sum_{\mathbf{n}} \Omega(\mathbf{n}) \tag{49}$$

The average frequency of appearance of the ith word among these messages is clearly

$$\langle n_i \rangle = \sum_{\mathbf{n}} n_i(\mathbf{n})\, \Omega(\mathbf{n}) / \sum_{\mathbf{n}} \Omega(\mathbf{n}) \tag{50}$$

and the probability of a given word (in the channel) is

$$P_i' = \langle n_i \rangle / W \tag{51}$$

The use of the channel at full capacity implies the possibility of every possible message subject to the imposed constraints and therefore a frequency of appearance of a word given by Eq. (51). Thus the frequency P_i in the source should be

$$P_i = P_i' \tag{52}$$

If M and T are very large, the average value $\langle n_i \rangle$ will be negligibly different from n_i ($\mathbf{n}^*$), where $\mathbf{n}^*$ denotes the distribution which gives the largest value of Ω,(15) and $\ln M$ will be indistinguishable from $\ln \Omega^*$, where Ω^* is the value of Ω that goes with $\mathbf{n}^*$. Thus we find $\langle n_i \rangle = n_i^* = n_i(\mathbf{n}^*)$ by maximizing Ω (or, more conveniently, its logarithm) subject to the constraints imposed by Eqs. (47) and (48). The result proves to be

$$\langle n_i \rangle = n_i^* = WP_i' = WP_i = We^{-\lambda t_i}/q \tag{53}$$

where

$$q = \sum_j e^{-\lambda t_j} \tag{54}$$

and λ is an undetermined multiplier(16) which can, in principle, be determined by substituting Eq. (53) into (48).

The quantity q is the analogue of the *partition* function per molecule[17] (in this case, per word) which occupies such an important position in statistical mechanics. λ is the analogue of a *temperature.*

Thus to use the channel at capacity when the spectrum of word times is fixed, we should select a source whose word probabilities are given by

$$P_i = e^{-\lambda t_i}/q(\lambda) \tag{55}$$

while if we are faced with the more common situation, namely, a source with fixed P_i, we must select word times governed by the relation

$$t_i = -(1/\lambda) \ln P_i q(\lambda) \tag{56}$$

Equations (51) and (52) represent the answers to the two questions posed above.

We are now in a position to define the information theory entropy and to show its relationship to the average word time t, under the condition that we employ the channel at capacity. Under this condition

$$t = \sum_i P_i t_i \tag{57}$$

Assume for the moment that the t_i are fixed and that we accommodate the source to the channel so that only the P_i vary. Then any change in mean word time t can be expressed as

$$dt = \sum_i t_i dP_i \tag{58}$$

If we substitute t_i from Eq. (56), the result is

$$dt = -\frac{1}{\lambda} \sum_i \ln P_i dP_i = \frac{1}{\kappa\lambda} d \left\{ -\kappa \sum_i P_i \ln P_i \right\} \tag{59}$$

where we have used $\sum_i dP_i = 0$, and κ is a constant which we choose as

$$\kappa = 1/\ln 2 \tag{60}$$

so that the quantity in braces will have the dimensions of *bits.* That quantity is in fact the famous information theory entropy per word, and so we write

$$S = -\kappa \sum_i P_i \ln P_i \tag{61}$$

where this S should not be confused with that defined in Eq. (36) of which it is, nevertheless, the analogue. Thus we may write

$$dt = (1/\kappa\lambda)\, dS \tag{62}$$

and defining

$$\tau = 1/\kappa\lambda \tag{63}$$

as the *temperature* of the *source,* we have

$$dt = \tau\, dS \tag{64}$$

and

$$P_i = e^{-t/\kappa\tau}/q(\tau) \tag{65}$$

Equation (65) is the complete analogue of Eq. (24), with κ replacing Boltzmann's constant and t_i taking the place of the energy of the ith quantum state. Equation (64) is the analogue of the thermodynamic relation, Eq. (12), for the case in which V is constant and the thermodynamic system is permitted to exchange energy with its surroundings only through the flow of heat; i.e., it is not allowed to perform work.

In spite of the extensive isomorphism with statistical thermodynamics, the reader is reminded that the correspondence is only formal and that quantities like S, t, and τ are conceptual rather than physical in origin.

The use of an *information theory temperature* τ as a parameter for characterizing a source is a novel and useful concept and may be justifiably regarded as a bonus provided by the interdisciplinary approach. If this source is to be matched with a channel intended for use at full capacity, then the channel will have to contain a code of mean word time t specified by Eq. (57). The mean word time is like the partial molar internal energy(18) (internal energy per molecule) or energy density in a thermodynamic system. The channel may be said to *come to equilibrium with the source* at temperature τ. When this equilibrium is established the mean word time is determined. Source and channel are "matched" when their temperatures are the same.

The source is therefore the analogue of a thermostat in thermodynamics. A system in equilibrium with this thermostat will possess a mean energy density determined by the temperature of the thermostat. The temperature τ of the source also characterizes its degree of randomness. The more equal the probabilities of having the source emit various words, the higher the temperature; i.e., higher temperature corresponds to greater randomness.

Since t is the counterpart of U, the thermodynamic internal energy, and τ is the counterpart of T, we may define the information theory analogue of A, the Helmholtz free energy (per molecule), as

$$\alpha = t - \tau S \tag{66}$$

and by exactly the same process which produced Eq. (38) we now obtain

$$\alpha = -\kappa\tau \ln q(\tau) \tag{67}$$

Reintroducing Eq. (66), we may write this as

$$t - \tau S = -\kappa\tau \ln q(\tau) \tag{68}$$

We shall use this expression later.

Having completed the discussion of this simple example, it is useful to consider a slightly more complicated analogue; the information theory counterpart of the *grand* ensemble.(12) We now no longer require W to be fixed and consider each message of length T to be part of a much longer *grand* message. In fact, the longer message may be regarded as composed of $\mathscr{M}$ messages, each requiring time T for its transmission. Thus the total message will require time $\mathscr{M}T$ for transmission. Although each component message has fixed T, it is "open" insofar as its word content W is concerned. The grand message, however, will have a fixed word content $\mathscr{W}$.

Each component message will still be characterized by a word distribution $\mathbf{n}$ and the number of *possible* messages of type $\mathbf{n}$ will still be as defined in Eq. (49), except that now W will depend on $\mathbf{n}$, so we write

$$\Omega_{\mathbf{n}} = W_{\mathbf{n}}/\pi_i n_i! \tag{69}$$

Suppose we consider a grand message in which the distribution of messages of type $\mathbf{n}$ is characterized by $M_{\mathbf{n}}$, $M_{\mathbf{n}}$ being the number of such component messages in the grand message. Then the number of grand messages going with this distribution is

$$\Gamma_{\mathrm{M}} = \mathscr{M}! \prod_{\mathbf{n}} \frac{\Omega_{\mathbf{n}}^{M_{\mathbf{n}}}}{M_{\mathbf{n}}!} \tag{70}$$

The total number of messages available to the channel is

$$M_G(\mathscr{M}T, \mathscr{W}) = \sum_{\mathbf{M}} \Gamma_{\mathrm{M}} \tag{71}$$

Each term in this sum is subject to the constraints

$$\sum_{\mathbf{n}} M_{\mathbf{n}} = \mathscr{M} \tag{72}$$

$$\sum_{\mathbf{n}} W_{\mathbf{n}} M_{\mathbf{n}} = \mathscr{W} \tag{73}$$

and again we approximate the logarithm of the sum by the logarithm of its maximum term. We thus maximize $\Gamma_{\mathbf{m}}$ subject to Eqs. (72) and (73). The result for $M_{\mathbf{n}}$ is then

$$M_{\mathbf{n}} = \mathscr{M} \Omega_{\mathbf{n}} \gamma^{W_{\mathbf{n}}} / \sum_{\mathbf{m}} \Omega_{\mathbf{m}} \gamma^{W_{\mathbf{m}}} \tag{74}$$

where γ is an undetermined multiplier which can be fixed by substituting Eq. (74) into (73). Obviously the probability $\theta_{\mathbf{n}}$ of the source emitting a message of type $\mathbf{n}$ is

$$\theta_{\mathbf{n}} = M_{\mathbf{n}}/\mathscr{M} = \Omega_{\mathbf{n}} \gamma^{W_{\mathbf{m}}}/Z \tag{75}$$

where

$$Z = \sum_{\mathbf{m}} \Omega_{\mathbf{m}} \gamma^{W_{\mathbf{m}}} \tag{76}$$

is the analogue of the *grand* partition function$^{(12)}$ in statistical thermodynamics.

Note that we are now dealing with a partition function for a whole *message* rather than for a *word*. In fact, if we review the steps leading to Eq. (75) we see that $\Omega_{\mathbf{n}}$ need not be confined to the form offered in Eq. (69); Eq. (75) will still hold. This means that it is possible to deal with non-zero-memory sources in which successive words are correlated. We need only take proper account of these correlations in the evaluation of $\Omega_{\mathbf{n}}$. Nonetheless, in this article attention will be restricted to the zero-memory case and we shall use Eq. (69).

The probability of the ith word in the grand message is clearly

$$P_i = \sum_{\mathbf{n}} \frac{n_i(\mathbf{n})\,\theta_{\mathbf{n}}}{W_{\mathbf{n}}} \tag{77}$$

and for employment of the channel at full capacity we should match to a source which emits words with these probabilities.

The isomorphism with statistical thermodynamics is so complete that a number of important results may now be stated almost without proof, the reader being referred to standard texts on statistical mechanics for satisfaction in this respect. It is convenient to introduce a quantity β defined by

$$\beta = \gamma e^{T/\kappa\tau} \tag{78}$$

In terms of β the grand partition function may be written as

$$Z = \sum_{\mathbf{m}} \Omega_{\mathbf{m}} e^{-TW_{\mathbf{m}}/\kappa\tau} \beta^{W_{\mathbf{m}}} \tag{79}$$

Next we collect all terms with $W_{\mathbf{m}}$ equal to fixed W and write Eq. (79) as

$$Z = \sum_{W} Q_W\, \beta^W \tag{80}$$

where

$$Q_W = e^{-TW/\kappa\tau} \sum_{\mathbf{m}} \Omega_{\mathbf{m}} \tag{81}$$

where now $\mathbf{m}$ refers to the set of formal vectors such that $W_{\mathbf{m}} = W$. Q_W is called the partition function of a *message* in the *canonical* ensemble. If we approximate the sum in Eq. (81) by its maximum term $\Omega_{\mathbf{m}}{}^*$, the result will simply be that which is derived by substituting Eq. (55) into (46), taking $n_i = WP_i$. Thus

$$\Omega_{\mathbf{m}}{}^* = e^{T/\kappa\tau}[q(\tau)]^W \tag{82}$$

and replacing the sum in Eq. (81) by this result gives

$$Q_W = q^W \tag{83}$$

Introducing Eqs. (78) through (82) into (75) gives

$$P_W = Q_W \beta^W / Z \tag{84}$$

where P_W is the probability of a message of word content W. From this it may easily be shown that the average word content $\overline{W}$ of a component message is

$$W = \beta \left(\frac{\partial \ln Z}{\partial \beta}\right)_\tau \tag{85}$$

For $\mathscr{M}$, $\mathscr{W} \to \infty$, it can be demonstrated that $\overline{W}$ is negligibly different from W^*, the value of W going with the maximum term in Z in Eq. (80). That maximum term satisfies the condition

$$\left(\frac{\partial \ln Q_{W^*}}{\partial W^*}\right)_\tau = -\ln \beta \tag{86}$$

or using $W^* = \overline{W}$,

$$\left(\frac{\partial \ln Q_{\overline{W}}}{\partial \overline{W}}\right)_\tau = -\ln \beta \tag{87}$$

Finally, substituting $Q_{\overline{W}} = q^{\overline{W}}$ from Eq. (83), we get

$$\ln \beta = -\ln q = \frac{1}{\kappa\tau} \frac{\partial}{\partial W}\{W\alpha(T)\} \tag{88}$$

where the last step involves the use of Eq. (67). The quantity in braces on the right is just the Helmholtz free energy of a message of word content W and temperature τ, so that the derivative is μ, the *information theory chemical potential*$^{(19)}$ *per word*. Thus

$$\beta = e^{\mu/\kappa\tau} \tag{89}$$

and is the analogue of the statistical thermodynamic *absolute activity*. Equation (85) may in fact be written

$$\overline{W} = \kappa\tau\left(\frac{\partial \ln Z}{\partial \mu}\right)_\tau \tag{90}$$

Thus it is apparent that μ determines the average word content of a component message, just as τ determines the mean time of transmission of a word.

The situation which we have arrived at is the following. We are able to deal with messages emitted by a source of fixed entropy per word (fixed P_i) and to characterize a channel coupled to this source by two parameters τ and μ. These parameters determine the average time of transmission and average word content of a message whose transmission time is T. Alternatively, we could fix the word content and the average word transmission time and employ Eq. (48) to determine τ and Eqs. (88) and (89) to fix μ.

In one case W and t are allowed to fluctuate while μ and τ are the fundamental quantities, and in the other case W and t are fixed and fundamental while μ and τ are derived. The μ and τ which are derived in the last case do give rise, when regarded as fundamental, to average values $\overline{W}$ and $\bar{t}$ which are identical, for $T \to \infty$, with W and t used to derive them in the first place.

In general, the capacity of the channel will depend upon μ and τ, and for a source to be matched to the channel its word probabilities P_i should be specified by Eq. (65). These results are quite general and can be used even in connection with sources having non-zero memories. In such cases, however, Q_W is not simply given by a^W, but must be evaluated taking correlation into account.

The statistical thermodynamic approach can also be used with messages of finite length. For this purpose some of the techniques already in existence for *thermodynamic* systems with "end" effects may be employed. In such systems the value of n_i may fluctuate appreciably from WP_i in a given message. On the other hand, using the grand ensemble approach where W and t are not fixed, it is possible to discuss these fluctuations and to characterize them in terms of τ and μ.

The fixing of both t and W may of course be viewed, alternatively, as the fixing of T and W. This corresponds to channels which, according to Jaynes, are only capable of *semi-optimal* transmission. With *optimal* transmission, the more usual case, only T is fixed.

Under this circumstance we maximize Ω in Eq. (46) without invoking the constraint (47). We will be able to show that $q(\tau) = 1$ and that $\tau = 1/C$. Equation (88) then requires $\beta = 1$ and Eq. (89) shows that $\mu = 0$. Thus for optimal transmission

$$\tau = 1/C, \qquad \mu = 0 \tag{91}$$

Now to the proof.

In Eq. (46) we replace W by $\sum_i n_i$ and take the variation of $\ln \Omega$ with respect to the various n_i, setting it equal to zero. The result is

$$\delta W = \sum_i \ln \left(\frac{\sum n_j}{n_i} \right) \delta n_i = 0 \tag{92}$$

Taking the variation of Eq. (48), with T constant, multiplying by an undetermined parameter λ, and adding to Eq. (92) gives (upon setting each coefficient equal to zero)

$$n_i = \left(\sum_j n_j \right) e^{-\lambda t_i} \tag{93}$$

or

$$P_i = n_i \Big/ \sum_j n_j = e^{-\lambda t_i} \tag{94}$$

Substitution of P_i into Eq. (58) once again shows that $\lambda = 1/\kappa\tau$, so that we have

$$P_i = e^{-t_i/\kappa\tau} \tag{95}$$

and since $\sum_i P_i = 1$, this shows that

$$q(\tau) = \sum_i e^{-t_i/\kappa\tau} = 1 \tag{96}$$

which is part of the desired result.

In a celebrated proof,⁽⁴⁾ Shannon showed that if a channel possesses r code symbols, and there is *no* restriction on W, the capacity C_r in r-ary units (log base r) is the largest real root of

$$\sum_i r^{-C_r t_i} = 1 \tag{97}$$

The typical term in this equation may be expressed in the form

$$r^{-C_r t_i} = [e^{\ln r}]^{-C_r t_i} = e^{-(C_r \ln r) t_i} = e^{-C_{\text{nat}} t_i} \tag{98}$$

where C_{nat} refers to the capacity in natural units. If we use the fact that $1/\kappa = \ln 2$, we can write

$$r^{-C_r t_i} = e^{-C t_i/\kappa} \tag{99}$$

where C is the capacity in *bits* per unit time. Substitution of Eq. (99) in (97) shows that $q(\tau_0) = 1$ if τ_0 is interpreted as $1/C$. This is the conventional information theory approach to "optimal" transmission.

In order to show that the statistical thermodynamic analogue method gives the same result, we must prove, without invoking Eqs. (97) or (98), that $\tau = 1/C$. At the same time, this will identify the τ_0 arbitrarily defined above as $1/C$ with the τ (under optimal transmission) which *we* have introduced as the *information theory* temperature. To accomplish this we substitute Eq. (96) into (68) and obtain (under optimal transmission)

$$S/t = 1/\tau \tag{100}$$

But S/t is simply the capacity at τ_0, i.e., the *entropy per second*; a fact which may be demonstrated by using the maximum terms in Eq. (49) for $M(T, W)$, making use of Eq. (53) with $\lambda = 1/\kappa\tau$, together with Eq. (46), and substituting $M(T, W)$ so obtained into (41). The result is

$$C = (\kappa/t) \ln q + 1/\tau \tag{101}$$

where we have used Eqs. (47), (48), and (60), and the fact that $T = Wt$. We may write Eq. (101) as

$$C = (1/\tau t)(\kappa\tau \ln q + t) \tag{102}$$

Substituting the negative of the left member of Eq. (68) for $\kappa\tau \ln q$ in Eq. (102) yields

$$C = S/t \tag{103}$$

Thus we may identify τ in Eq. (100) with τ_0, the desired result. Whenever the channel is *totally unconstrained* so that it is used at maximum capacity, it follows from Eq. (44) that

$$\tau_0 = \tau^* = \sigma \ln 2/\ln r \tag{104}$$

The maximum capacity of the channel is achieved when the code is free of constraints. Under this circumstance, C^* is given by Eq. (44). At less than maximum capacity, $C < C^*$, and according to Eq. (103),

$$t/S \geqslant \sigma \ln 2/\ln r \qquad \text{or} \qquad t/\sigma S = l/S \geqslant \ln 2/\ln r \tag{105}$$

in which l is the mean length in symbols of the code words.

This is Shannon's first theorem.(4) How closely we achieve the equality depends on how constraint-free we choose the channel code. If the entropy S were measured in r-ary units, the ratio of logarithms on the right would have to be replaced by unity. Equation (103) defines "matching" quite generally as the condition under which the entropy or information per word per unit time delivered by the source (the right number) just equals the channel capacity. To a large degree Shannon's theorem has been established using the statistical thermodynamic analogue method.

Nevertheless, this method is very closely related to those conventional in information theory (even if the language is altered slightly). But even if there is no particular manipulative advantage to deriving Eq. (105) by this method (and under certain circumstances there may be), it has permitted us to define and characterize certain quantities, for example the information temperature and chemical potential, and to round out the analogue.

Now that the analogue is relatively complete, we are in a position to formulate certain questions, the answers to which are conveniently obtained by applying the transformation theory of thermodynamics discussed in Sections 2 and 3. This matter is discussed in the next section.

5. THERMODYNAMIC-LIKE TRANSFORMATIONS

Before proceeding, it is worthwhile to summarize some of the results of the preceding sections. We have considered a channel defined by a code subject to certain constraints. Among these was a fixed "spectrum" of word transmission times t_i. If $\mathbf{t}$ is used to denote the spectrum of word times, the capacity C may be expressed as a function of $\mathbf{t}$ and t, $C = C(\mathbf{t}, t)$. We discovered that if the channel was to be used at capacity it was necessary to match it to a source whose word probabilities were given by

$$P_i = e^{-t_i/\kappa\tau}/q(\tau) \tag{106}$$

where

$$q(\tau) = \sum_j e^{-t_j/\kappa\tau} \qquad \text{and} \qquad \kappa = 1/\ln 2 \tag{107}$$

τ is determined by the requirement $\sum_i P_i t_i = t$. If the word content of a message is *unconstrained*, $q(\tau) = 1$ and τ may be shown to be equal to $1/C$.

If the information theory entropy per word is defined as $S = -\kappa \sum_i P_i \ln P_i$, then it may be shown that the variation in mean word time required of a channel with fixed $\mathbf{t}$ and fixed word content W, if it is to remain matched to a source of entropy S, as S varies, is given by

$$dt = \tau\, dS \tag{108}$$

Alternatively, we can ask for the change in the entropy of the source required for continued matching with a channel whose mean word time changes. Clearly this will be

$$dS = (1/\tau)\, dt \tag{109}$$

As we change the source so that the various P_i change, it will not be possible in general to "match" it to the channel by merely changing τ as long as the spectrum $\mathbf{t}$

remains fixed. However, matching may be accomplished by allowing the t_i to change so as to further accommodate the changing source. It is this subject which we now address, and in connection with which we will have an opportunity to use the transformation theory mentioned in the early sections.

It is easiest to begin by considering a set of parameters, $X_1, X_2, \ldots, X_k$ upon which the various t_i will be made to depend. The exact choice of functional relation $t_i(\mathbf{X})$ (where now $\mathbf{X}$ stands for the set X_1, X_2, etc.) will define a "strategy" for matching under varying conditions. To deal with the simplest case, we will assume that only a single X_i is used so that $\mathbf{X}$ may be replaced by X. No increased difficulty is encountered when several parameters are involved.

Equation (58) must now be replaced by

$$dt = \sum_i t_i dP_i + \sum_i P_i \frac{dt_i}{dX} dX \tag{110}$$

if the variation in t due to a variation in X is to be included. If matching is to be achieved, P_i must still be given by Eq. (65), and Eq. (110) may be expressed in the form

$$dt = \tau\, dS - \pi\, dX \tag{111}$$

where

$$\pi = -\sum_i P_i \frac{dt_i}{dX} \tag{112}$$

The reader is invited to compare Eq. (111) with (12). For convenience we may refer to π as the information theory *pressure.*

We can immediately make use of the transformation theory of thermodynamics to express π in terms of the partition function. Thus taking the total differential of α defined by Eq. (66), we have

$$d\alpha = dt - \tau\, dS - S\, d\tau \tag{113}$$

and substituting Eq. (111) gives

$$d\alpha = -\pi\, dX - S\, d\tau \tag{114}$$

From this we obtain

$$\pi = -\left(\frac{\partial \alpha}{\partial X}\right)_\tau = \kappa\tau \left(\frac{\partial \ln q}{\partial X}\right)_\tau \tag{115}$$

where we have used Eq. (67).

We may also make use of the transformation theory in order to express t directly in terms of q. Thus from Eq. (114) we see that

$$S = -\left(\frac{\partial \alpha}{\partial \tau}\right)_X \tag{116}$$

Substitution in Eq. (66) gives

$$t = \alpha - \tau \left(\frac{\partial \alpha}{\partial \tau}\right)_X = -\tau^2 \left(\frac{\partial (\alpha/\tau)}{\partial \tau}\right)_X \tag{117}$$

and use, once more, of Eq. (67) leads to

$$t = \kappa\tau^2 \left(\frac{\partial \ln q}{\partial \tau}\right)_X \tag{118}$$

S may also be expressed in terms of q by substitution of Eq. (67) into (66):

$$S = \kappa \left[\frac{\partial(\tau \ln q)}{\partial \tau}\right]_X \tag{119}$$

Next we apply the transformation theory in order to solve a practical coding problem. We shall choose a somewhat artificial case as a means of reducing the algebraic manipulations so that the details do not obscure the method, which is, after all, the primary target of our exposition. Thus we will deal with a channel containing a code word spectrum such that

$$t_i(x) = iX + X^2 \tag{120}$$

where i assumes all values from zero to infinity. In connection with this channel we ask the following question: *How does the channel capacity C (under matching) change as the word time spectrum is changed due to changing X, under the condition that the mean word time t remains fixed?*

If we are to answer this question, we shall need the coefficient

$$\Gamma = \left(\frac{\partial C}{\partial X}\right)_t \tag{121}$$

We employ thermodynamic transformation theory in order to express Γ in a more convenient form. We begin by noting, from Eq. (111), that another expression for π, besides that given in Eq. (115), is

$$\pi = -\left(\frac{\partial t}{\partial X}\right)_S = \left(\frac{\partial t}{\partial S}\right)_X \left(\frac{\partial S}{\partial X}\right)_t \tag{122}$$

where Eq. (4) has been used. Referring to Eq. (111), it is seen that the first derivative on the right is τ, so that Eq. (122) can be rearranged to

$$\frac{\pi}{\tau} = \left(\frac{\partial S}{\partial X}\right)_t \tag{123}$$

Now according to Eq. (103), $S = Ct$; and substitution of this relation into Eq. (123) gives

$$\frac{\pi}{\tau} = t\left(\frac{\partial C}{\partial X}\right)_t = t\Gamma \tag{124}$$

or

$$\Gamma = \pi/t\tau$$

By using Eqs. (115) and (118), Γ may be expressed in terms of the partition function. Thus

$$\Gamma = \left(\frac{\partial C}{\partial X}\right)_t = \left(\frac{\partial \ln q}{\partial X}\right)_\tau \bigg/ \tau^2 \left(\frac{\partial \ln q}{\partial \tau}\right)_X \tag{125}$$

In order to calculate Γ, therefore, we must first evaluate $q(\tau, X)$ and then eliminate τ by expressing it in terms of the value of t, maintained fixed in the problem. The evaluation of q is straightforward using Eq. (120). Thus

$$q = \sum_{i=0}^{\infty} \exp\left\{-\frac{iX + X^2}{\kappa\tau}\right\} = e^{-X^2/\kappa\tau} \sum_{i=0}^{\infty} (e^{-X/\kappa\tau})^i = \frac{e^{-X^2/\kappa\tau}}{1 - e^{-X/\kappa\tau}} \tag{126}$$

where we have summed a geometric series. Now

$$t = \kappa\tau^2 \frac{\partial \ln q}{\partial \tau} = \frac{X^2 + X(1 - X)\, e^{-X/\kappa\tau}}{1 - e^{-X/\kappa\tau}} \tag{127}$$

Solving for τ in terms of t, we obtain

$$\frac{1}{\tau} = \frac{\kappa}{X} \ln\left\{1 + \frac{X}{t - X^2}\right\} \tag{128}$$

Equation (120) requires t to be at least as large as X^2 so that the argument of the logarithm must be greater than or equal to unity. Thus, as t increases for fixed X (fixed code word spectrum t_i), τ must increase.

This is consistent with the statistical thermodynamic analogue where internal energy density is increased with increasing temperature, t being the analogue of energy density.

We now use Eq. (115), together with Eq. (126),

$$\pi = \kappa\tau \left(\frac{\partial \ln q}{\partial X}\right)_\tau = -\frac{2X + (1 - 2X)\, e^{-X/\kappa\tau}}{1 - e^{-X/\kappa\tau}} \tag{129}$$

The denominator of Eq. (125) is available in (127), so that now we can use it together with Eq. (129) to obtain

$$\Gamma = -\frac{1}{\tau}\left\{\frac{2X + (1 - 2X)\, e^{-X/\kappa\tau}}{X^2 + X(1 - X)\, e^{-X/\kappa\tau}}\right\} \tag{130}$$

Eliminating τ from this expression, using Eq. (128), finally yields

$$\Gamma = \left(\frac{\partial C}{\partial X}\right)_t = -\kappa\left(\frac{X^2 + t^2}{X^2 t}\right) \ln\left\{1 + \frac{X}{t - X^2}\right\} \tag{131}$$

Thus an increase in X at constant t, which, according to Eq. (120), implies an increase in the length of each word, leads to a decrease in channel capacity. This is not surprising since increasing the various t_i without increasing the average word length will require an increased use of the smaller words and therefore, effectively, a smaller alphabet.

Next we discuss the opposite problem. We are faced with a change in source described (for simplicity) by a single parameter y. Thus we have

$$P_i = P_i(y) \tag{132}$$

In this case S, which is fully specified by the P_i, will be a function of y only, while the t_i, which, under matching, will have to be given by Eq. (106), will depend on both y and τ:

$$t_i = t_i(y, \tau) \tag{133}$$

Equation (110) is then replaced by

$$dt = \sum_i t_i dP_i + \sum_i P_i \left\{ \left(\frac{\partial t_i}{\partial v} \right)_\tau dy + \left(\frac{\partial t_i}{\partial \tau} \right)_y d\tau \right\} \tag{134}$$

Defining

$$\xi = \sum_i P_i \left(\frac{\partial t_i}{\partial y} \right)_\tau \tag{135}$$

and

$$\eta = \sum_i P_i \left(\frac{\partial t_i}{\partial \tau} \right)_y \tag{136}$$

we may write Eq. (134) as

$$\begin{aligned} dt &= \tau\, dS + \xi\, dy + \eta\, d\tau \\ &= \left(\tau \frac{dS}{dy} + \xi \right) dy + \eta\, d\tau \end{aligned} \tag{137}$$

All sorts of transformations are possible using Eq. (137), but if we ask the same question as we did in connection with the previous problem, the procedure is much simpler than in that case. For example, if we ask for the rate of change of capacity with y such that the mean word time t remains fixed, we need the following coefficient:

$$\Gamma' = \left(\frac{\partial c}{\partial y} \right)_t \tag{138}$$

Substituting from Eq. (103) this becomes

$$\Gamma' = \frac{1}{t} \frac{dS}{dy} \tag{139}$$

Since dS/dy is a given quantity, the problem is solved.

6. CONCLUSION

We could carry the process far beyond the point reached in the preceding section but enough has been said to demonstrate the method.

The statistical thermodynamic analogue method undoubtedly has considerable power. It should be especially useful if extended so as to deal with non-zero-memory sources. It may even be useful in connection with noisy channels, but this remains to be demonstrated. One suspects that the analogue of *irreversible* thermodynamics(20) should find application there since one is confronted with problems of entropy *production*, but this also remains to be shown.

REFERENCES

1. L. Brillouin, *Science and Information Theory*, 2nd ed., Academic Press, New York (1963).
2. E. T. Jaynes, *Phys. Rev.* **106**:620 (1957); **108**:171 (1957).
3. E. T. Jaynes, *IRE Trans. Inform. Theory* **5**:98 (1959).
4. N. Abramson, *Information Theory and Coding*, McGraw-Hill, New York (1963), p. 72; C. E. Shannon, *The Mathematical Theory of Communication*, University of Illinois Press (1949).
5. N. Abramson, *op. cit.*, p. 66.
6. H. Reiss, *Methods of Thermodynamics*, Blaisdell, New York (1965).
7. *Ibid.*, Chapter II.
8. *Ibid.*, Chapter III.
9. *Ibid.*, Chapter III.
10. *Ibid.*, Chapter IV.
11. R. C. Tolman, *Principles of Statistical Mechanics*, Oxford University Press (1938), Chapters 2–4.
12. T. L. Hill, *Introduction to Statistical Thermodynamics*, Addison-Wesley, Reading (1960), Chapter 1.
13. N. Abramson, *op. cit.*, p. 111.
14. *Ibid.*, p. 13.
15. T. L. Hill, *op. cit.*, p. 478.
16. H. Reiss, *op. cit.*, Chapter II.
17. T. L. Hill, *op. cit.*, Chapter 3.
18. H. Reiss, *op. cit.*, p. 47.
19. *Ibid.*, Chapters VII and VIII; T. Hill, *op. cit.*, Chapter 1.
20. D. D. Fitts, *Nonequilibrium Thermodynamics*, McGraw-Hill, New York (1962).