

Physical versus Logical Coupling in Memory Systems

Abstract: A memory system consisting of bistable static dissipationless units such as ferrites, ferroelectrics, or cryotrons is considered. For a given amount of physical material the memory capacity may be increased by using small volumes of the bistable material for each bit. If made sufficiently small, however, the individual bits will become unreliable because of the influence of thermal agitation and quantum-mechanical tunneling processes. Some unreliability can be tolerated, since it can be compensated by redundancy. The optimum size of the individual bit, for maximum information storage, is evaluated. If thermal agitation is the prime source of errors, then the optimum-sized bit involves typically less than 100 of the independent cooperating units (electron spins, dipoles, et cetera) which cause the bistability. The maximization process concerns itself only with the preservation of information and not with possible methods of access to the individual bit. In particular, the maximization process neglects complications in the coding equipment needed to read in and out of memory.

Introduction

The search for continually smaller and faster data processing system components causes us to ask what ultimate limitations the laws of physics impose on the progress in this direction. In this note we are concerned with the increasingly important effects of quantum-mechanical tunneling and thermal agitation on the reliability of a memory, as the physical system storing an individual bit becomes very small. It will be shown that there is a sense in which there is an optimum-sized memory element.

The existence of the optimum can be expected from the following crude line of thought. Assume that we are given a fixed amount, say a pound, of perhaps a ferromagnetic material, which can be used to manufacture bistable units capable of serving as memory elements. How finely shall we divide our pound? We can use the whole pound to make only one bit. The resultant bit, because of its massiveness, will be unlikely to lose its information, either through a tunneling process, or through thermal agitation, and will therefore be a very reliable bit. Nevertheless our whole memory is rather useless since it is only a single bit. At the opposite extreme, we can divide the pound into a great many units, each of which is so small that it loses its information almost immediately. Despite the large number of elements in such a memory, it is again almost useless. There is, therefore, presumably an intermediate degree of subdivision which provides for a maximum memory capacity, if the time

over which we wish to preserve information has been specified.

Our considerations will involve binary symmetric storage elements, i.e., elements having two and only two distinct (but physically equivalent) states. The restriction to binary symmetric elements is convenient, but not essential.

In general a nonvanishing probability exists that a storage element makes a spurious transition from one of its states to the other. Such transitions may be caused by thermal fluctuations, by the imperfect isolation of a storage element from the effect of machine operations supposedly irrelevant to its state, or simply by the fact that neither state is stationary in the quantum-mechanical sense. Such transitions may lead to error in the interpretation of what information has been stored in memory. The probabilities of such errors may be reduced by the introduction of suitable *coupling* between the storage elements.

1. Definition of basic concepts

By *physical coupling* between n storage elements we understand a connection such that a transition may occur if and only if all n elements make the same transition simultaneously. A new storage element is evidently created in this way, having a reduced transition probability.

By *logical coupling* between n elements we understand the introduction of redundancy such that only k ($k < n$) of the group of n elements can be adjusted in

†Deceased. A preliminary and relatively complete version of this manuscript had been prepared by J. A. Swanson, and has been brought into its present form by R. Landauer.

a logically independent fashion. The remaining $(n-k)$ elements have their states uniquely determined, when the k states have been specified, by a definite function of k arguments, chosen so as to make "error correction" possible up to a certain point. Such a coupling will be referred to as (n, k) coupling.

2. Preliminary statement of problem

Both physical and logical coupling reduce error probabilities, and both require extra storage material in order to keep the amount of stored information constant. An obvious question is—which method of coupling uses physical material more efficiently to reduce error? Our purpose is to gain some insight into this problem.

It is evident that a fair appraisal would take into account the limitations involved in the physical realization of logical coupling. However such considerations lie beyond the scope of this discussion. The fact that logical coupling requires complications in the coding and decoding equipment, which is external to the memory being interrogated, will be ignored.

3. Transition probability for one element

The states of an element will be labeled "A" and "B." Suppose at time $t=0$ the state is definitely A. What is the probability as a function of time that the element makes a transition to state B? Clearly, the answer to this question depends on our physical model. It is natural to assume, however, that transitions in both directions proceed as in a Poisson process. The result of this assumption is found most easily by imagining a large number of elements n_0 originally in state A. The number of transitions per unit time to state B is proportional to the number in A, while the number of transitions per unit time to A is proportional to the number in B. The differential equation to which we are led has as solution for the number n_B in B

$$n_B = \frac{n_0}{2} [1 - \exp(-2\gamma t)],$$

where γ is a constant. Thus the probability q of finding an element in state B known to have been in state A at time $t=0$ is

$$q = \frac{1}{2} [1 - \exp(-2\gamma t)].$$

If γt is small, $q = \gamma t$. Thus γ is the *transition probability per unit time*.

4. Behavior of transition probabilities under physical coupling

An Appendix deals with this problem from a physical viewpoint. It is shown there that if one assumes a memory element to be equivalent to a particle in a potential with two minima, then the transition probability due to thermal agitation or quantum-mechanical tunneling of n physically coupled elements is proportional to the expo-

ponential of a negative constant times n . This follows from taking the n coupled element to correspond to a particle n times as massive in a potential n times as strong as a single element. The more intuitive argument given below yields the same type of dependence of the transition probability on n using a completely different sort of assumption as to the exact nature of physical coupling.

We consider n systems coupled physically, each of which, when uncoupled, has transition probability γ per unit time. We seek for the transition probability per unit time, γ_n , of the coupled system.

Let us suppose that a characteristic time τ is required for the constraint to make itself felt. If all n systems were impelled to make a transition during time τ , the transition would occur as in the absence of constraint. The probability for this is $(\gamma\tau)^n$. Thus, if we assume that the transition cannot otherwise occur (similar results are obtained if we require that at least one-half must be impelled to change during time τ)

$$\gamma_n = \tau^{-1} (\gamma\tau)^n.$$

Introducing new quantities

$$v = \tau^{-1}, \mu = -\log_e \gamma\tau, \gamma_n = v e^{-\mu n}.$$

The probability of transition during a time t is (for $\gamma_n t \ll 1$)

$$q_n = v t e^{-\mu n}.$$

A coupling which yields the above formula will be called *uniform physical coupling*. The characteristic of uniform coupling is that the addition of a new element always causes the same fractional reduction in the transitional probability.

While the Appendix points out that "uniform physical coupling" is the obvious way in which matter is coupled against the disturbing influences of thermal agitation and quantum-mechanical tunneling, this does not hold with respect to the undesired cross influences of various operating elements. As far as "cross talk" is concerned, the different portions of the same piece of matter are likely to be subject to related disturbing influences, and hence their coupling does not reduce the error probability as effectively as indicated in this section.

5. Reduction of error probabilities by logical coupling

An (n, k) coupling reduces the number of "messages" written upon a group of n storage elements from 2^n to 2^k . However, the existence of spurious transitions renders the occurrence of all 2^n combinations of states possible. Thus the decoding scheme consists in associating 2^{n-k} configurations with each message. The specification of the message to which a given configuration belongs will be said to determine the *information state* of the n elements. An *error* will be said to occur if the information state changes because of spurious transitions.

Not all possible (n, k) coupling schemes (of which there are many, unless n is very small) are equally effective in reducing error. There is no generally known way, however, to select the best (n, k) couplings without carrying through the tedious process of actually computing and comparing the error probabilities. Since we shall always assume the best logical coupling we are not in a position to demonstrate a function of n and k which shall yield the value of the minimum attainable error probability, but only to assert that such a function is computable. Nevertheless the theorem of Shannon¹ regarding the capacity of a noisy channel applies here and allows us to state that

$$\lim_{n, k \rightarrow \infty} q_{n, k} = 0,$$

as $n \rightarrow \infty$ and k varies in any manner such that

$$\frac{k}{n} \leq C - \delta,$$

where δ is a fixed constant, and

$$C \equiv 1 + q \log_2 q + (1 - q) \log_2 (1 - q).$$

Here q is the transition probability for one memory element, and $q_{n, k}$ is the probability of an error in the information state for the best (n, k) coupling.

6. The fundamental variables

It is convenient to make the number of elementary systems which are physically coupled a continuous variable, and to adopt an attitude toward physical coupling in greater harmony with the facts of natural physical coupling due to co-operative phenomena as it occurs in ferromagnetics, ferroelectrics, cryogenic memory elements, et cetera. Let us imagine a volume V of material as mentioned in the introductory section exhibiting a natural physical coupling. We may physically separate this material into N storage elements of equal volume v . Let us suppose there is a smallest volume v_0 exhibiting uniform coupling. We can adopt v_0 as our unit of volume. The volume V is then a measure of the maximum number N_0 of elements into which we imagine the material may be divided. The smallest element will be called a *minimum cell*.

The transition probability of a minimum cell will be written as

$$q_{\text{min. cell}} = vt e^{-\mu_0}.$$

Thus, in general

$$q = vt e^{-v\mu_0}$$

for an arbitrary cell of volume v , according to our result in the section on physical coupling.

7. The optimum cell

Let us suppose quite generally that we have available a certain physical material which is to store information according to some specified scheme. If information is stored at time $t=0$ and read out at a later time t , we

may regard the entire process as a (delayed) transmission of information, and may define a channel capacity in the sense of Shannon.¹ However we define a capacity per *unit volume* rather than per symbol. If there are any possibilities of variation in the scheme of information storage, we define the *retention capacity* $R(t)$ as the maximum attainable channel capacity per unit volume under variation of the scheme.

We now apply these rather general ideas to the present instance. It follows from information theory that the *information* I stored in N elements cut from a volume V according to the scheme outlined previously is

$$I = N(1 + q \log_2 q + (1 - q) \log_2 (1 - q)).$$

The information per unit volume is

$$\frac{I}{V} = \frac{1}{v} (1 + q \log_2 q + (1 - q) \log_2 (1 - q)).$$

We define $R(t)$ as the maximum of $\frac{I}{V}$ under variation of v . It will be shown that this maximum is achieved for a finite value of v . We shall call a storage element possessing this volume an *optimum cell*.

We shall look for a maximum in I/V under the assumption that the coefficient vt , which multiplies the exponential in the error probability q , is very large compared to unity, i.e., that the storage time involved is tremendous compared to the basic period in which the constraints are felt. Furthermore we shall assume that the coefficient μ is large enough so that volumes small compared to the unit volume will already be exceedingly reliable ($q \ll 1$).

For the maximization process we must consider the derivative

$$\begin{aligned} \frac{d}{dv} (I/V) &= -\frac{1}{v^2} [1 + q \log_2 q + (1 - q) \log_2 (1 - q)] \\ &\quad + \frac{1}{v} \frac{dq}{dv} [\log_2 q - \log_2 (1 - q)], \end{aligned}$$

where $q = \frac{1}{2}(1 - e^{-2\gamma t})$ and $\gamma = vt e^{-\mu v}$. Hence dq/dv ,

which appears in the expression for $\frac{d}{dv} \left(\frac{I}{V} \right)$ is given by

$$\begin{aligned} \frac{dq}{dv} &= te^{-2\gamma t} \frac{d\gamma}{dv} = te^{-2\gamma t} \frac{d}{dv} (vt e^{-\mu v}) \\ &= -vt \mu e^{-2\gamma t} e^{-\mu v}. \end{aligned}$$

We first wish to point out the existence of a maximum. To do this we shall consider the sign of $d(V^{-1}I)/dv$ for very small and very large v .

If v is large enough then q will be very small compared to unity and we can approximate:

$$\begin{aligned} \frac{d}{dv} \left(\frac{I}{V} \right) &= -\frac{1}{v^2} + \frac{1}{v} \frac{dq}{dv} \log_2 q \\ &= -\frac{1}{v^2} + \frac{1}{v} \frac{dq}{dv} \log_2 (vt e^{-\mu v}), \end{aligned}$$

$$= -\frac{1}{v^2} + \frac{1}{v} (vt\mu e^{-2\gamma t} e^{-\mu v}) [\mu v - \log vt].$$

We can simplify this expression by setting $vte^{-\mu v} = \gamma t$ and $e^{-2\gamma t} = 1 - 2q \doteq 1$. This yields

$$\frac{d}{dv} \left(\frac{I}{V} \right) = -\frac{1}{v^2} + \frac{1}{v} [\gamma t] [\mu v - \log vt].$$

The second term on the right-hand side depends exponentially on v through the factor γ , and hence for sufficiently large v the first term will predominate and the derivative will be negative.

Now we must consider the other extreme, where v is sufficiently small so that at the end of the time t the value of q is almost exactly $1/2$. The function

$$1 + q \log_2 q + (1 - q) \log_2 (1 - q),$$

has a minimum at $q = 1/2$. At the minimum the value of the information per bit goes to zero, and increases quadratically as q departs from $1/2$. Let $1/2 - q = \delta$. Then the quantity whose derivative interests us is

$$\frac{I}{V} \sim \delta^2/v.$$

And its derivative is

$$\frac{d}{dv} \left(\frac{I}{V} \right) \sim -\frac{\delta^2}{v^2} + \frac{1}{v} 2\delta \frac{d\delta}{dv}.$$

Using

$$\begin{aligned} \frac{d\delta}{dv} &= \frac{d}{dv} \left(\frac{1}{2} e^{-2\gamma t} \right) = -te^{-2\gamma t} \frac{d\gamma}{dv} \\ &= -te^{-2\gamma t} (-\mu v e^{-\mu v}) = 2\mu\delta vte^{-\mu v}. \end{aligned}$$

we find

$$\frac{d}{dv} \left(\frac{I}{V} \right) = -\frac{\delta^2}{v^2} + \frac{1}{v^2} (4\delta^2 \mu v vte^{-\mu v}).$$

The ratio of the second term to the first term is $4vt\mu v e^{-\mu v}$. It is this ratio which determines the sign of the derivative. It is clear that in the limit $v \rightarrow 0$ this ratio goes to zero and the derivative we are interested in is negative. We are, however, not likely to be interested in such small values of μv , and wish to show that for somewhat larger and more reasonable values of v , which are, however, still associated with $(1/2 - q) \ll 1$, we have

$$4vt\mu v e^{-\mu v} > 1,$$

which indicates that the derivative $d/dv(I/V)$ does have a zero for a value of q not too close to either zero or one-half.

We are interested only in values of v , which make $e^{-\mu v}$ small enough, so that the bistability makes itself felt for at least a few constraint periods τ (or oscillation periods $1/v$). This means μv must be at least of order unity, and hence $\mu v e^{-\mu v}$ must also be of order of unity for this smallest reasonable volume. The information storage periods of interest, however, are ones which

make vt tremendous, and hence for this smallest reasonable volume

$$4vt\mu v e^{-\mu v} \gg 1.$$

This gives the desired sign for the derivative $d/dv(I/V)$. Furthermore since

$$\gamma t = vte^{-\mu v},$$

we also have $\gamma t \gg 1$ at this smallest reasonable volume, and hence $(1/2 - q) \ll 1$, so that this small volume does correspond to almost complete loss of information in the time t .

Note that the volume which we have just called the smallest reasonable volume bears no very obvious relationship to the minimum volume v_0 , which exhibits physical coupling. The quantity v_0 is the smallest volume in which the co-operative effect is sufficiently strong to lock the individual particles, spins, et cetera, together in their bistable behavior.

We have described the two extreme cases, in which q is either very close to zero, or else very close to $1/2$, in some detail, to show that the derivative vanishes in between these ranges, since the actual equation obtained by setting the derivative equal to zero is exceedingly complicated, and we shall not attempt a very accurate solution. At the maximum we have

$$\frac{d}{dv} \left(\frac{I}{V} \right) = 0,$$

which yields

$$\frac{1 + q \log_2 q + (1 - q) \log_2 (1 - q)}{\log_2 q - \log_2 (1 - q)} = v \frac{dq}{dv},$$

or

$$\frac{1 + q \log_2 q + (1 - q) \log_2 (1 - q)}{\log_2 q - \log_2 (1 - q)} = v (te^{-2\gamma t}) (-\mu v e^{-\mu v}),$$

which can also be written:

$$\frac{1 + q \log_2 q + (1 - q) \log_2 (1 - q)}{\log_2 (1 - q) - \log_2 q} = \mu v e^{-\mu v} vte^{-2\gamma t}.$$

We have shown that the solution yields a q which is not very near zero or one-half. Hence the fraction on the left has a magnitude not too far from unity. Similarly the factor $e^{-2\gamma t} = 1 - 2q$ must be smaller than, but comparable to unity. We therefore know that the volume v which solves the above equation cannot be far from the one which solves

$$\mu v e^{-\mu v} vt = 1.$$

Note that vt is very large, hence the exponential must be very small, and therefore very small percentage changes in the volume v can be expected in the transition from the solution of the rigorous equation to the solution of the simplified one.

The simplified equation requires $e^{-\mu v}$ to overcome the large factor vt . Therefore we know that a reasonable

first guess for a solution of the simplified equation is

$$\mu v = \log vt.$$

We can look for a slightly improved solution by letting

$$\mu v = \log vt + \delta,$$

which yields, after taking the logarithm of the simplified equation

$$\log [\log vt + \delta] - \delta = 0.$$

Expanding the left-hand term gives

$$\log \log vt + \frac{\delta}{\log vt} - \delta = 0.$$

The second term is negligible in comparison with the third, leaving us with

$$\mu v = \log vt + \delta = \log vt + \log \log vt.$$

Thus as vt becomes very large, v approaches $\frac{1}{\mu} \log vt$.

We see immediately that longer storage times lead to increasing optimum volumes. The memory capacity per unit volume, neglecting factors which are smaller than unity, but not very small compared to unity will then be

$$\frac{I}{V} = \frac{\mu}{\log vt}.$$

The insensitivity of $\log vt$ to small factors in its argument shows that it will not help a great deal to shorten the storage period by periodically reading out the entire memory content and rewriting it. (This conclusion is, of course, limited to deteriorating influences which can be drastically reduced by small increases in bit size.)

8. Discussion and interpretation

Our procedure for finding the optimum cell has neglected the difficulties associated with logical coupling. A more practical method of finding an optimum would therefore lead to a larger cell. The difficulties in reaching a very small cell for reading and writing purposes have of course also been completely neglected, and are likely to require larger elements than our optimum. The difficulties in reaching a cell are maximized in random access storage, where a separate channel must exist, leading to each bit. The difficulties are less pronounced in storage forms such as tapes and drums, and therefore our criteria are likely to be more applicable to these. The most compact known form of storage is that displayed in genetics, where the information is really kept on a molecular level. The economy of space in genetic storage must be paid for by the slowness of read-out. The read-out proceeds by a chemical reconstruction process whose time duration is essentially a gestation period. But it is presumably at this level of storage efficiency that the limitations due to tunneling and thermal agitation are really met, and it may well be that these effects contribute to mutation rates.

Let us, however, try to get an impression of the significance of our optimum cell, in terms of more conventional storage media, concentrating furthermore on the possibility of information loss by thermal agitation. In the case of thermal activation our criterion requires that $vte^{-U/kT}$ be comparable to unity. Here U is the height of the barrier which must be surmounted by thermal agitation. The frequency v is associated with a molecular vibration, electronic transition, et cetera, and is therefore typically of the order 10^{12} to 10^{16} , although it can be lower. In the ferromagnetic case, v is a resonance frequency determined by the anisotropy field and is at best of the order of a microwave frequency. Furthermore for all systems, magnetic or otherwise, the factor v that multiplies the exponential is not really the natural vibrational frequency, but is in general lower and only becomes comparable to the resonance frequency for a critically damped system. (This point has been discussed in detail by Kramers.²) Therefore we can expect that the factor v of interest typically lies in the range 10^6 to 10^{16} . The storage period of interest, t , is likely to be at least a second and may well be 1000 years for some conceivable applications.³ Hence vt is likely to range between 10^6 and 10^{19} . This requires that the factor $\exp(-U/kT)$ range roughly between 10^{-6} and 10^{-19} . The quantity U/kT then must range between about 14 and 44. Since our calculation is based on an optimum coding, which is not likely to be really available, we should perhaps take larger values of U/kT . In any case, however, U/kT is not likely to exceed 100.

How large is the volume which provides a switching barrier of $100 kT$? This cannot be answered with any real generality, but a rough order-of-magnitude guess can be made. All the co-operative phenomena which give rise to bistability are associated with some critical temperature T_c , at which the co-operative phenomena disappear. The switching barrier per participating unit (electron spin, moving atom, et cetera) is typically of the order kT_c . For n elementary participating particles the barrier is therefore of order nkT_c . Hence n must be large enough to give

$$nkT_c \sim 100 kT.$$

The operating temperature T will be below the critical temperature T_c , but in all known cases T is still of the same order of magnitude as T_c . Hence n is of the order of 100. This is an exceedingly small element. In most co-operative phenomena this would in fact be likely to be less than the critical number required for establishing the co-operative phenomenon involved. Furthermore since a cubic array of 100 units consists of a cube with about 5 units on a side, it can be seen that the difficulties of addressing such an element are extreme.

Despite the remoteness of the results from practical possibilities, they are perhaps more relevant than some more prevalent typical considerations. It has been realized a number of times, that in order to read a signal from memory, against a background of thermal noise, the switching energy should be of order kT , or for high

accuracy, a few times kT .⁴ We see that the criterion for preserving information against thermal fluctuations is a more stringent one than the read-out in the presence of noise.

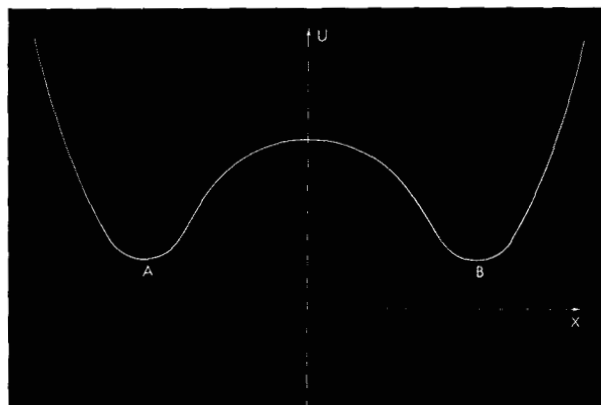
Appendix: Effect of increasing volume on thermal agitation and quantum-mechanical tunneling.

We have in mind a situation somewhat as depicted opposite, which shows the energy (or perhaps free energy) of the switching system as a function of the coordinate x being switched. A and B are the two possible stable states denoting "0" and "1," respectively. There are many possible paths by which the system can pass from A to B , and the one dimensional diagram is merely intended to be symbolic. Our principal assumption is that the energy barrier in this diagram will be proportional to the volume of the switched element, as the latter is changed. This is certainly characteristic of processes such as rotational switching in ferrite films, without domain-wall motion, in which a one-dimensional diagram exactly as shown is really relevant, since the whole volume is locked together in its switching action. The simple energy barrier versus volume proportionality may also, perhaps, hold for more complicated systems in which switching involves nucleation and domain-wall motion. The simple scaling law will obviously break down for switching units which are so small that the cooperative phenomenon which caused the bistability itself disappears—e.g., a minimum number of spins is needed to make a ferromagnet. The range of applicability is therefore somewhat uncertain, but in any case it is clearly the simplest sensible assumption for the dependence of reliability on volume. The probability for transfer from A to B by thermal agitation is then of the form $\nu \exp(-U/kT)$, where U is the barrier height which is taken to be proportional to the volume, and ν is a frequency factor, not likely to be too different from the typical frequency of oscillation associated with well A . The exact value of ν depends on the degree to which vibrations in the system are damped, in a manner which has been discussed by Kramers.² In any case ν is not likely to be appreciably affected by volume increases. The resulting dependence on volume of $\exp(-U/kT)$ is identical with that deduced in the section dealing with physical coupling.

As far as quantum-mechanical tunneling is concerned, we must consider a probability of the form

$$\exp \left[-\frac{1}{\hbar} \int \sqrt{2m(V-E)} dx \right],$$

where the path of integration is taken through that portion of the barrier providing the largest leakage. Strictly speaking, a uniform probability of tunneling, unchanging in time, does not apply to this problem, but rather to the case typified by α -decay, where the particle, after tunneling has no coherent probability of tunneling back again. In the case we are considering, where the barrier separates two identical wells, we find a resonant tunnel-



ing back and forth with a frequency $\Delta E/h$, where ΔE gives the splitting of two states which are similar within each well, but have opposite parity about the barrier center. Therefore if the system is initially in Well A its probability of appearing in Well B rises first as t^2 and is followed by sinusoidal oscillations in the probability of occupation. It seems, however, likely that, if Well B is a many dimensional well, that thermal scattering of the system occurs sufficiently frequently, so that the coherence leading to the resonance is destroyed. That is, the system after arriving in B will go through a considerable history before it arrives once more at the portion of the barrier which permits maximum tunneling. In this period it will have had an opportunity to interact sufficiently with other systems to lose its quantum-mechanical phase memory. Hence tunneling from A to B is just like tunneling from A into free space, except that in the two-well case there is a probability of eventual return tunneling. The probability of return tunneling is, however, a process once again just like the initial tunneling, and therefore independent of phase details.

Even if the system does exhibit the real resonance processes, it seems likely that a simple tunneling rate,

$\nu \exp \left[-\frac{1}{\hbar} \int |p| dx \right]$, will still describe the mean rate of information loss, with fair accuracy. The important point for our considerations, in any case, is that ν is independent of volume, and $|p|$ proportional to it.

Footnotes and references

1. C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication*, University of Illinois Press, Urbana, Illinois (1949).
2. H. A. Kramers, *Physica* 7, 284 (1940).
3. In archival storage, where a great deal of information must be kept for a long time, and where access is infrequent, we are particularly likely to want to push against the maximum limits. Here the insertion of adequate redundancy for information preservation may simply consist in leaving in some of the natural redundancy of material presented in verbal form.
4. This is closely related to discussions by L. Brillouin, in *Science and Information Theory*, Academic Press, New York, 1956. Elegant unpublished considerations along this line have been given by R. L. Garwin and by M. C. Andrews.

Received March 4, 1960