PHYSICAL REVIEW LETTERS

VOLUME 46 9 MARCH 1981 NUMBER 10

Energy Cost of Information Transfer

Jacob D. Bekenstein[a]
Princeton University Observatory and Joseph Henry Laboratories, Princeton University, Princeton, New Jersey 08544
(Received 4 December 1980)

From thermodynamic and causality considerations a general upper bound on the rate at which information can be transferred in terms of the message energy is inferred. This bound is consistent with Shannon's bounds for a band-limited channel. It prescribes the minimum energy cost for information transferred over a given time interval. As an application, a fundamental upper bound of 10^{15} operations/sec on the speed of an ideal digital computer is established.

PACS numbers: 02.50.+s, 05.90.+m, 06.50.-x

To transmit information costs energy: Every message is associated with a material object or radiation, and must thus be accompanied by some energy. Can this energy cost be reduced to arbitrarily low values by technical innovation? That the answer may be negative is suggested by Shannon's classic bounds on the rate of information transfer through a communications channel.[1-4] However, these are expressed in terms of the channel's bandwidth and the signal-to-noise ratio. For my particular question and many others of physical interest, a bound expressed directly in terms of the message energy would be preferable. Just such a bound was proposed by Bremermann[5] on the basis of one of Shannon's formulas. The obscurity in which his proposal has remained is perhaps due to the confusion in his arguments: He identifies Shannon's noise energy with the energy uncertainty dictated by $\Delta E\,\Delta t > \hbar$. Yet the fact that a time limited waveform must have a spread in frequency (from which follows $\Delta E\,\Delta t > \hbar$) is actually unrelated to whether a channel is noisy or not in Shannon's sense.[1] Fortunately, it proves possible to derive a bound similar to Bremermann's, but whose range of applicability transcends that of Shannon's bounds, from thermodynamic and causality considerations. I now describe the general idea and implications for the energy cost of information transfer. Further details and examples will be deferred to a later publication.

I start from a recent result: The ratio of entropy to total energy in the rest frame, S/E_0, of a system bound in space is less than $2\pi k\hbar^{-1}c^{-1}R$ with R the effective radius of the system in its rest frame. This result may be inferred[6] from the second law of thermodynamics as generalized to systems involving black holes.[7] In this approach E_0 is interpreted as the precise energy, while for nonspherical systems the meaning of R is obscure. An alternative approach[6] derives the result for systems of massless quantum fields directly from statistical physics. In this second approach E_0 is interpreted as the statistically mean energy, and R as the radius of the sphere which circumscribes the system. I here adopt the latter interpretation, and assume the bound on S/E_0 to be generally valid. I now recall that if a given system's entropy has a maximal value S, then by using every one of its internal states as a symbol, one can code in it information up to an amount $S(k \ln 2)^{-1}$ bits (binary digits).[4] Thus the storage capacity of a system with mean energy E_0 which can just be enclosed in a sphere of radius R is less than $2\pi(\hbar c \ln 2)^{-1}E_0R$ bits, an interesting result in its own right.

A message is in essence a package of matter or radiation (packet of electromagnetic waves in a transmission line, laser pulse in space, burst of relativistic electrons...). First one assumes that it moves slower than light so that it has a rest frame. The receiver which ultimately detects it can recover from it no more than the information I it contains. The bound implies

$$I < 2\pi(\gamma\hbar c \ln 2)^{-1}ER, \tag{1}$$

where γ is the Lorentz factor corresponding to the package-receiver relative motion, and E is the package energy in the *receiver's* Lorentz frame. The rate of information transfer (average rate of information acquisition by the receiver) is just I/τ, where τ is the time that elapses in the receiver's frame between the arrival of the package's front and the instant the last detected signal is conveyed to the receiver's "memory."

Now $\tau \geq \max(T,t)$ where T is the time for the full length of the package to impinge on the receiver, as measured in the latter's frame, and $2t$ is the light travel time across the largest transverse dimension of the system L_t as seen from the receiver [signals cannot be conveyed from every detecting element to the memory—wherever located—in less time than $t = L_t(2c)^{-1}$]. Evidently $T = L_l(v\gamma)^{-1}$ where L_l is the longitudinal dimension (along the line of flight as seen by receiver) of the package in its own rest frame, and v is the package-receiver relative velocity. For given package geometry and receiver disposition, T decreases as v increases, but t remains constant. Hence the smallest possible τ can be attained for $v \to c$ (and for special shapes also for a range of v below c). To relate τ to R without entering into details about the shape, I imagine the parallelepiped with edges of length $A = L_l$, $B = L_t$, and $C = L_t$, which just boxes in the package. The sphere which circumscribes this parallelepiped cannot be smaller than that which circumscribes the package; hence $L_l^2 + 2L_t^2 \geq 4R^2$. Since I am interested in v near c, I may assume that $\tau \geq L_t(2c)^{-1} > L_l(v\gamma)^{-1}$. Combining all these inequalities yields

$$\tau > 2R(v^2\gamma^2 + 8c^2)^{-1/2}. \tag{2}$$

I now take $v \approx c$ to obtain the bound $\tau > 2R(\gamma c)^{-1}$ on the shortest possible τ. Dividing the earlier bound (1) by this last one, I get my basic result

$$\dot{I} < \pi E/\hbar \ln 2, \tag{3}$$

where $\dot{I}$ is the average rate of information transfer in bits per second.

Bound (3) differs from that proposed by Bremermann by a numerical factor as well as in interpretation. Its range of applicability far exceeds that which Bremermann claimed for his because (3) is obtained without recourse to Shannon's bounds. Thus, for example, (3) applies to a hypothetical communication system which employs neutrinos although none of Shannon's results are then relevant because the neutrino field is not itself measurable. Although derived for packages which have a rest frame, bound (3) should also hold for packages traveling exactly at the speed of light. This follows because (3) is good for v arbitrarily near c. Also, assuming that for $v = c$ (3) is not valid leads to contradictions. For example, suppose a message coded in a laser pulse satisfies, in a particular frame, $\dot{I} > \pi E(\hbar \ln 2)^{-1}$ when it propagates in vacuum. If the pulse then *gradually* enters a nonabsorbing medium with index of refraction $1 + \epsilon$ (with $\epsilon \ll 1$) at rest in our chosen frame, its velocity becomes smaller than c, so that (3) becomes applicable. However, since E cannot have increased, this can only happen if I decreased or the minimum τ increased. But if the transition is sufficiently gradual, no information should be lost. Likewise any change in τ from the change in medium must be of $O(\epsilon)$ and can be made arbitrarily small. Hence, our original supposition was wrong, and in fact $\dot{I} \leq \pi E(\hbar \ln 2)^{-1}$ in vacuum. The equality would be attainable only if for $v \neq c$, $\dot{I}$ can approach arbitrarily near $\pi E(\hbar \ln 2)^{-1}$. In fact there is evidence that the thermodynamic bound which I used to obtain (3) cannot be approached closely.[6] Thus (3) must be a strict inequality for all messages, lightlike or not, and in fact it may be possible to obtain a general bound tighter than (3).

Our bound may also be rewritten in the form

$$E/I > \hbar \ln 2/\pi\tau, \tag{4}$$

which expresses the energy cost per bit for a message received and recorded over time τ. In interpreting (3) and (4) one must remember that

E and τ are measured in the frame of the *receiver*. If, as is the case in radio communications, only a fraction of the transmitted energy is received, or if transmitter and receiver are in relative motion, or at different gravitational potentials, one may relate $\dot{I}$ to the *transmitted* energy by scaling τ and E by the appropriate solid-angle and red-shift factors. The moral of (3) and (4) is that rapid reception of information ultimately requires higher energy reception, i.e., higher energy transmission, and/or greater received- to transmitted-energy ratio. In short, fast information is energy expensive.

A well-known rule of thumb in the theory of communications via electromagnetic channels (the Hartley-Nyquist law)[2,8] equates the maximal $\dot{I}$ with the bandwidth Δf (in hertz) of the channel. This rule is consistent with (3). In fact, in a band-limited channel the *mean* photon energy will generally exceed $\frac{1}{2}h\,\Delta f = \pi\hbar\,\Delta f$; since there is at least one photon per message, $E/\hbar > \pi\,\Delta f$. Thus the bound (3) is more generous than the rule, a further indication that it may be possible to improve my bound. The unprecise Hartley-Nyquist rule has been superseded by Shannon's bounds on $\dot{I}$ which, in particular, take account of the role of noise in limiting $\dot{I}$. I shall show in a later publication that (3) is always consistent with Shannon's bound whenever the latter is applicable.

An example which illustrates the power of (4) concerns the maximal conceivable speed of a digital computer. The number of elementary operations that such a machine can perform per second, r, cannot exceed cD^{-1} where D is the characteristic dimension of the "circuits." Presumably a computer must be made of atoms, and so D cannot be reduced arbitrarily. However, it is not clear how many atoms are the minimum for a futuristic computer (10^6 or 10^{15}?) so that this line of reasoning cannot predict a precise maximal r. Here (4) may be used to resolve this difficulty. In performing a simple arithmetical operation, the machine typically transfers information, specifying a couple of numbers over a distance D. Now, to specify a nine-digit decimal number takes 30 bits ($2^{30} \approx 1\times 10^9$), so that the machine transfers about $60r$ bits/sec. According to (4) this costs at least $1.4\times 10^{-26} r$ ergs/operation since the relevant τ here must be shorter than r^{-1} sec if the machine is to carry out one operation on the basis of results from the preceding one. If, as seems probable, the energy accompanying each "message" cannot be recycled, the power dissipated, P_d, must be at least $1.4\times 10^{-26} r^2$ ergs $\sec^{-1}$ (this is unrelated to resistive heating which can be eliminated). Thus for large r the machine has an "overheating" problem.

The biggest allowed r is determined by the largest P_d which can be offset by available cooling mechanisms. For the nearly microscopic machines I have in mind ($r = 3\times 10^{12}$ implies $D < 10^{-2}$ cm) cooling by forcing a fluid through the "circuits" is out of the question. The other options are cooling by heat conduction to the periphery of the circuits and by radiation from the atoms and molecules of the computer. To cool efficiently by conduction one would like to make the machine in the form of a thin plate of thickness d and area of order D^2. The cooling rate (ergs/sec) is then $P_c \approx \kappa D^2 d^{-1}\Delta T$, where ΔT is the typical temperature differential inside the machine and κ the thermal conductivity. Now one knows that $\kappa \approx \frac{1}{3}Cv\lambda$ where v and λ are the rms speed and mean free path of the heat carriers (phonons or electrons) and C is their specific heat per unit volume.[9] Concentrating first on the κ from phonons, one knows that v is about the sound speed and hence $v < 5\times 10^5$ cm $\sec^{-1}$ for the known solids under ordinary conditions. Now λ is set by various scattering mechanisms, but in any case $\lambda < d$ (DeHaas-Biermasz effect).[9] It is well known that each atom in a solid contributes a maximum of $3k$ to the phonon heat capacity. In ordinary solids there are $\sim 10^{23}$ atoms/cm^3. Hence $C < 4.14\times 10^7$ erg $^\circ\mathrm{K}^{-1}$. I recall that $D < cr^{-1}$. Hence $P_c < 6.2\times 10^{35}\xi r^{-2}$ erg $\sec^{-1}$, where I have taken $\Delta T < 100\,^\circ\mathrm{K}$, a reasonable figure; the factor ξ lumps the effects of my various approximations. I expect it to be of order unity. If instead κ is due to electron conductivity, v must be interpreted as the Fermi velocity $(2kT_\mathrm{F}/M)^{1/2}$, where T_F is the Fermi temperature and M the (effective) electron mass. The heat capacity per electron is[9] $\frac{1}{2}\pi^2 kT/T_\mathrm{F}$. As before, $\lambda < d$. The lowest T_F's of known conductors are $2\times 10^4\,^\circ\mathrm{K}$. Taking $M \approx 9.1\times 10^{-28}$ g, $T < 500\,^\circ\mathrm{K}$, and $\Delta T < 100\,^\circ\mathrm{K}$, I find that the previous bound for P_c is still valid if one takes $\xi \approx 6.4$. Setting $P_d = P_c$ gives $r_{\max} < 2.6\times 10^{15}\xi^{1/4}$.

Cooling by radiation will be most efficient in the optical regime where the Einstein coefficients and quanta energies are largest ($\approx 10^8$ $\sec^{-1}$ and 3×10^{-12} erg, respectively) from among the radiative modes, one can hope to couple efficiently to the heat source. With 10^{23} atoms/cm^3 and a volume $\approx D^3$, one gets $P_c < 8.1\times 10^{50}\xi r^{-3}$ erg $\sec^{-1}$, where again ξ is a factor of order unity correcting for my approximations. Comparing with P_d, I get $r_{max} < 2.25 \times 10^{15} \xi^{1/5}$.

Thus whatever the cooling method, a machine operating at $r \gtrsim 10^{15}$ must overheat by hundreds of degrees Kelvin and destroy itself (a conservative time scale is 10^{-6} sec). Therefore, 10^{15} operations/sec is a firm upper bound on the speed of an ideal digital computer. Any realistic machine would fall short of this by orders of magnitude.

I thank Amnon Meisels for long bygone discussions, Gary Gibbons for reawakening my interest in the subject and for calling my attention to Bremermann's work, Professor J. Ostriker and Professor S. B. Treiman for hospitality, and the National Science Foundation for support under Contract No. AST-79-22074.

[a]On sabbatical leave from the Ben Gurion University of the Negev, Beersheva, Israel.

[1]C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication* (Univ. of Illinois Press, Urbana, 1949).

[2]S. Goldman, *Information Theory* (Prentice-Hall, New York, 1953).

[3]D. Middleton, *An Introduction to Statistical Communication Theory* (McGraw-Hill, New York, 1960).

[4]L. Brillouin, *Science and Information Theory* (Academic, New York, 1956).

[5]H. J. Bremermann, in *Proceedings of Fifth Berkeley Symposium on Mathematical Statistics and Probability,* edited by L. M. LeCam and J. Neyman (Univ. of California Press, Berkeley, 1967).

[6]J. D. Bekenstein, Phys. Rev. D 23, 287 (1981).

[7]J. D. Bekenstein, Phys. Rev. D 7, 2333 (1973); S. W. Hawking, Commun. Math. Phys. 43, 199 (1975).

[8]F. J. Dyson, Rev. Mod. Phys. 51, 447 (1979).

[9]C. Kittel, *Introduction to Solid State Physics* (Wiley, New York, 1956), 2nd ed.

Power Spectra of Strange Attractors

B. A. Huberman and Albert B. Zisook[a]
Xerox Palo Alto Research Center, Palo Alto, California 94304

(Received 29 December 1980)

It is shown that, for systems which enter chaos through period doubling bifurcations, the integrated noise power spectrum in the chaotic regime behaves as $N(r) = N_0(r - r_c)^\sigma$, with $\sigma = 1.5247\ldots$. Furthermore, the existence of a new universal constant which describes the scaling behavior of the average bandwidth in the strange attractor is reported. These results are directly applicable to experiments probing the onset of turbulence in physical systems.

PACS numbers: 05.20.Dd, 05.40.+j, 47.25.-c

A number of physical systems, such as stressed fluids, high-temperature plasmas, and Josephson junctions have been observed to undergo a transition into a turbulent regime characterized by broadband noise in the power spectra. A possible explanation for these phenomena is that the phase trajectories for the complete nonlinear many-body problem enter a low-dimensional region of phase space containing a strange attractor. A strange attractor is a region in phase space such that nearby trajectories must enter it but once inside they diverge from each other. Hence we arrive at a description of turbulence involving only very few degrees of freedom. The effectively stochastic motion which these few degrees of freedom undergo gives rise to the observed noise in the power spectra. One common route into this turbulent regime is a universal cascade of period doubling bifurcations which occur as some control parameter is varied.[1-3] This cascade can be easily understood when, through the construction of return maps associated with the Poincaré maps, the dynamical system is mapped onto one-dimensional (1D) recursion relations which possess the same bifurcation structure.[4]

Recently, it has been shown that once in the chaotic regime, the Lyapunov exponent, which measures the rate of divergence of nearby trajectories, behaves very much like the order parameter of a phase transition near the critical point, i.e., it obeys a universal scaling law.[5] This development allows, in principle, for the application of techniques developed in the study of critical phenomena to the onset of turbulence in these nonlinear systems.

Appealing as these ideas might be, they suffer from the fact that one cannot directly measure Lyapunov exponents or discern the topology of at-