Fundamental physical limits on computation

Warren D. Smith *

May 5, 1995

Abstract

Current (May 1995) revision of 1992 report.

We consider limitations on the performance of computers arising from thermodynamics and the laws of physics. We provide upper bounds on three quantities: sustained information flux, information storage density, and sustained computational speed. All of these upper bounds are "tight" in the sense that they could be approached by plausible-sounding physical systems, and they all arise from a single unified point of view. We also make a conjecture about the rate of inevitable decay of stored information. This conjecture may be thought of as a quantitative extension of the second law of thermodynamics. It leads to a bound on the density of stable information. We carefully elucidate the assumptions behind these bounds. We give a list of 4 open problems at the end.

KEYWORDS: thermodynamics, computation, reversible Turing machines, blackbody radiation, decay of information, physics, entropy, information transmission and storage, cooling requirements.

Contents

*NECI, 4 Independence way, Princeton NJ 08540. Email: wds@research.NJ.NEC.COM

Table of physical constants that will be used in this article.

symbol and value	name
$c = 2.998 \times 10^8$ meters/sec	speed of light
$\hbar = 1.055 \times 10^{-34}$ meter^{2}Kg/sec	Planck constant
$h = 2\pi\hbar = 6.626 \times 10^{-34}$ meter^{2}Kg/sec	
$\sigma_{SB} = 5.670 \times 10^{-8}$Kg sec^{-3}Kelvin^{-4}	Stefan-Boltzmann constant, see (13)
$G = 6.673 \times 10^{-11}$ meters^{3}Kg^{-1}sec^{-2}	Gravitational constant
$m_e = 9.110 \times 10^{-31}$ Kg	Mass of electron
$k_B = 1.381 \times 10^{-23}$ meters^{2}Kg Kelvin^{-1}sec^{-2}	Boltzmann constant
$\alpha = \frac{e^2}{4\pi\epsilon_0\hbar c} = \frac{1}{137.0}$	fine structure constant
$r_e = \alpha r_c = 2.818 \times 10^{-15}$ meters	Classical electron radius
$r_c = \frac{\hbar}{m_e c} = 3.862 \times 10^{-13}$ meters	Compton wavelength of electron/2π
$r_b = r_c/\alpha = 5.292 \times 10^{-11}$ meters	Bohr radius
q	Number of "kinds" of elementary particles in universe

1 Ultimate goals

The ultimate, far off, goal of the present research is as follows. Given a problem (such as "factor this integer _____") and given that one has a computer of mass m contained in a sphere of radius R, and supplied with power P. We want to obtain bounds, especially lower bounds, on the time it must take for the computer to solve the problem, in terms of m, R, P, and the problem. [1]

[1] Since quantum mechanics is not deterministic, perhaps it would be better to say "expected waiting time before, with probability $> 2/3$, the computer will report a correct solution to the problem." Also perhaps one may need, or want, to make additional assumptions about the computer, such as "the computer is made of the following 30 chemical elements _____, is surrounded by vacuum, and recieves input and output via 9600 baud ASCII on a RS-232 connector." However, it seems pointless to go into much detail, considering our present state of total incompetence at answering all such questions.

1.1 Flavor

The flavor of the research will be to try to derive lower bounds while making as few physical assumptions as possible. (And explicitly stating what assumptions it is that we *are* making.) We will not be interested in bounds that assume, say, that the computer must be made of CMOS technology, or even made of certain kinds of atoms within a certain temperature range. Such bounds are certainly of interest, but they are obviously subject to the criticism that the bound could be defeated by changing the rules. I am more interested in bounds that depend on truly fundamental physical limits.

Due to the considerable number of errors present in previous papers in this field, I have resolved to be very careful in my arguments. Since nobody has yet found a "grand unified" physical theory that explains everything and which all agree on, it is naturally impossible to "prove" things in physics with the same ease with which one proves things in computer science. Therefore I have restricted my "proofs" to statements of the form "'this list of assumptions implies this claim" and then to discuss the regimes in which said assumptions are (at least empirically) valid. I apologize for the verbosity this has incurred.

1.2 Subgoals

It seems to the present author (or at least, that is the way he has organized this paper) that there are three fundamental tasks going on in every computation:

1. Transmitting information between the parts of the computer.
2. Storage and retrieval of information in memory.
3. Alteration of information, also known as computation, also sometimes this involves destruction of information. For example, computing $C = A + B$ might also be thought of as altering the information in A and B, or throwing some of it away (if C is kept but A and B are not; or if an old value of C is overwritten).

Hence these questions arise:

1. What are limits on information flux?
2. What are limits on memory capacity and access time?
3. What are limits on the permanence of stored information?
4. What are limits on speed of computation?

The present paper will show that very elementary thermodynamics and quantum theory can say something about all of these questions.

1.3 How to read this paper

The reader who knows information theory and statistical thermodynamics can skip §2. The reader interested only in our main results and uninterested in our argumentation, can scan through the paper to find "Questions 1,2, and 4" and "Answers 1A,1B,1C,2A,2B,4A,4B,..,4E" in the text. The conjectures (question 3 and answers 3A, 3B) of §4.7 are also of interest, although I feel they are not quite right, cf. footnote 22.

2 Background

In this section we review a few definitions and standard results in information theory [gall80] [shan49] [slep74] and thermodynamics [reif65] [land80] [lifs80]. See also [jayn57]. This review will in no sense be complete. Also, for a brief discussion of Turing machines, see §5.

2.1 Entropy

In information and probability theory, the "entropy" of a set of w states is merely $\lg w$. In other words, saying which of the w states, the system is actually in, would convey $\lg w$ bits of information. (We use $\lg x = \log_2 x$ and $\ln x = \log_e x$ in this paper.) In thermodynamics, the "entropy" of a physical system, regarded as being in an ensemble of w possible states with equal likelihood, is $k_B \ln w$.

More generally, the information theoretic entropy of w states, the ith occurring with probability p_i, is $s = -\sum_{i=1}^{w} p_i \lg p_i$ bits. (If $p_i = 1/w$, this expression reduces to the previous one, $\lg w$.) In other words, there are

$$\binom{N}{Np_1,\ Np_2, \ldots, Np_w} \tag{1}$$

possible messages, containing a large number N of characters selected from a w-letter alphabet, and such that the total number of occurences of letter i in the message was Np_i. The lg of this number (by Stirling's formula) is asymptotic to $Ns = -N\sum_{i=1}^{w} p_i \lg p_i$, so the information content of a single letter selected according to the probability distribution p is s. The thermodynamic entropy of a physical system is $k_B \ln 2$ times its information theoretic entropy in bits.

According to quantum mechanics (a brief axiomatic description of which is given in the first chapter of [bjor64]), the definition of the "state" of a physical system is precisely an eigenfunction of its multiparticle wave equation. Thus entropy is really a question of counting eigenfunctions.

2.2 Temperature

If S is the entropy of a physical system and E is its energy, (where S is reckoned by assuming that all states of the physical system with energy E are equally likely) then its "temperature" T is given by $1/T = dS/dE$. Here the meaning of the "derivative" can be unclear, since S as we have defined it is a discrete quantity, but in most cases of interest the number of accessible states of a physical system is so large that the discreteness of S does not matter.

Another way around the discreteness problem[2] is to regard the state of a physical system as a single point in a $6N$ (very high) dimensional momentum-position "phase space."[3] One then regards lumps of phase space of hypervolume h^{3N} as a single "state" since close points are "indistinguishable" according to quantum uncertainty principles. Hence entropy is really a log-hypervolume – a continuous quantity.

Vast combinatorial forces cause entropy to "always" increase toward its maximal possible value, explaining why hot things heat up cold things and explaining why "temperature" is a universal and useful concept. In fact, it is worth formalizing this statement:

The second law of thermodynamics: *The entropy of an isolated physical system is nondecreasing, and in "thermodynamic equilibrium" is maximal. Any such system will eventually approach thermodynamic equilibrium.*

This is of course only a statistical law and is subject to violation by statistical fluctuations. All the air molecules in a room could decide to move to the North half of the room. But, all the quantities in physical situations in which one applies statistical mechanics are always so enormous (i.e. the number of air molecules in the room is of order 10^{26}) that such fluctuations are virtually impossible.

2.3 Heat bath vs. maximal entropy probability distributions

A very useful notion in thermodynamic arguments is the notion of a "heat bath," which may be thought of as a very large volume of ideal gas at a constant temperature. A heat bath at temperature T leaking an infinitesimal amount dq of heat, loses thermodynamic entropy dq/T.

We can see immediately that for a small system A in contact with (i.e. rapidly exchanging energy with) a large heat bath B, A will be found in a state with energy E, with probability $\exp(\frac{-E}{k_B T})$. This is because

[2] Perhaps the best way is to use the entropy of a probability distribution, which is continuous

[3] There may be additional discrete coordinates orthogonal to momentum and position, e.g. spin.

B must have an energy E below its usual energy and will thus have an entropy $\Delta S = E/T$, below than its usual entropy. The fraction of the states of B which have this energy, is thus proportional to $\exp(-\Delta S/k_B)$. Thus this "Boltzmann distribution" is an immediate consequence of the notion of "heat bath," conservation of energy, the definition of entropy as a log of a state-count, and the assumption ("fundamental assumption of statistical mechanics") that all states of the combined $(A \cup B)$ system are equally likely.

Jaynes [jayn57] argued that the Boltzmann distribution is the maximum-entropy state-occupation probability distribution with fixed expectation value of energy E. Thus the Boltzmann distribution, according to Jaynes's point of view, was forced without the need to refer to any heat bath. Or if a heat bath was needed, it was merely to serve as a "thermometer" to calibrate one's scale of temperature.

The "equipartition principle," which states that the expected energy in any "degree of freedom" of a system in contact with a temperature-T heat bath, is $k_B T$, is an immediate consequence of Boltzmann's distribution (view any particular "degree of freedom" as a subsystem "A").

2.4 Bose statistics

"Bose statistics" arise from the Boltzmann distribution as follows. We have bins containing indistinguishable particles. The total number of particles in bin i is $n_i \in \{0, 1, 2, \ldots\}$. There is no constraint on the total number of particles. The total energy of the system is $E = \sum_i n_i E_i$. The system is in contact with a temperature-T heat bath. A derivation in [reif65] (§9.3) then shows that the expected number of particles in bin i is

$$\overline{n_i} = \frac{1}{\exp \frac{E_i}{k_B T} - 1}. \tag{2}$$

2.5 Information theory, channel capacity, Shannon's theorem

"Information theory" is concerned with the transmission of information through a "noisy channel" between a "sender" and a "receiver." The information consists a N-character message where each character is selected from an L-letter alphabet. The channel may cause a character i sent by the sender to change into another character j received by the reciever, with some "transition probability" $p(j|i)$. Thus the noisy channel can destroy information. The sender and receiver can try to compensate for this by using error-correcting codes. The "capacity" C of the channel is

$$C = \min_{\substack{0 \le Q_1, \ldots, Q_L \le 1, \\ \sum_{i=1}^{L} Q_i = 1}} \max_{k \in \{1,2,\ldots,L\}} \left(\sum_{j=1}^{L} P(j|k) \lg \frac{P(j|k)}{\sum_{i=1}^{L} Q_i P(j|i)} \right). \tag{3}$$

Note that the determination of C is a minimization of a concave-up function within a convex polytope in an $(L-1)$-dimensional space. Such "convex programming" problems can be solved to high precision in time polynomial in the size of the problem [vaid89].

Shannon's theorem states that *an error-correcting coding scheme (mappings between numbers and N-letter messages) exists, so that the sender can transmit a message representing a number in the set $\{1, \ldots, 2^{(C-\epsilon)N}\}$, for any fixed $\epsilon > 0$, and have confidence approaching 1 (as $N \to \infty$) that the receiver will be able to determine the number. However, if the sender tries to transmit a number picked randomly from a uniform distribution in the set $\{1, \ldots, 2^{(C+\epsilon)N}\}$, for any fixed $\epsilon > 0$, then, no matter what scheme is used, with probability approaching 1 (as $N \to \infty$) the receiver will be unable to determine the number. (Both approaches to probability 1 are exponentially rapid.)*

Similarly to the second law of thermodynamics, Shannon's theorem is only a statistical statement; it is conceivable the channel could suddenly decide to create noise far above or far below its usual level, just when you try to transmit an important message. Again, however, one is usually concerned with values of N so large that this is virtually impossible.

"Shannon's bounds" can enable one to estimate the capacity of a channel without actually computing it. They state that: For any L values $Q_1, Q_2, \ldots, Q_L$ with $0 \le Q_i \le 1$, $\sum_{i=1}^{L} Q_i = 1$,

$$C \le \max_{k \in \{1,2,\ldots,L\}} \left(\sum_{j=1}^{L} P(j|k) \lg \frac{P(j|k)}{\sum_{i=1}^{L} Q_i P(j|i)} \right), \tag{4}$$

and

$$C \ge \sum_{k=1}^{L} Q_k \left(\sum_{j=1}^{L} P(j|k) \lg \frac{P(j|k)}{\sum_{i=1}^{L} Q_i P(j|i)} \right). \tag{5}$$

Both of these bounds are tight if the Q_i are chosen to lie at the optimum of the convex programming problem (3). If the transition probabilties have the property that $\sum_{i=1}^{L} P(j|i) = 1$ for each $j = 1, 2, \ldots L$, then Shannon's lower bound on capacity (with $Q_i = 1/L$) may be written

$$C \ge \lg L - \frac{1}{L} \sum_{k=1}^{L} S_k, \quad S_k = -\sum_{j=1}^{L} P(j|k) \lg P(j|k). \tag{6}$$

This inequality is in fact an equality if it is also the case that all the S_k are equal. In words: "the capacity of a symmetric channel is the entropy of the source ($\lg L$) minus the entropy of the channel."

3 Information flux

Question 1. "What is the maximum amount $\dot{I}$ of sustained information flux (bits transmitted across an area A per unit time) one can possibly achieve using a given amount P of power?"

It will turn out that there is an intimate relationship between fundamental physical limits on heat flux, and fundamental physical limits on information flux. The answer to question 1 which makes this connection most clear is as follows.

Answer 1A: *$\dot{I}$ is bounded by the maximum possible flux of entropy in any physical process that transmits a flux of heat energy, having power P, across the window.*

We will argue that at sufficiently high power densities, it is impossible to transmit much more entropy with a given amount of power (or dually: much more power at a given temperature) than one can transmit by using thermal radiation through a vacuum. The result, which is less general, but usually more useful, than answer 1A, is

Answer 1B: *Assuming that all information must be transmitted using photons, $\dot{I}$ is bounded by the flux of entropy radiated by a blackbody of surface area A radiating power P. In the regime where the Stefan-Boltzmann blackbody law $P = \sigma_{SB} A T^4$ is valid this is*

$$\frac{\dot{I}}{A} \le \frac{4(\sigma_{SB})^{1/4}}{3k_B \ln 2} \left(\frac{P}{A}\right)^{3/4}. \tag{7}$$

Caveat: The formula is only claimed to be true in the 24-order power range 10^3 Kelvin $< T = (\frac{P}{A\sigma_{SB}})^{1/4} <$ 10^9 Kelvin, *see discussion below. Even at significantly higher equivalent blackbody temperatures than* 10^9 *Kelvin, our formula should still be valid except for a possible need to multiply it by a factor of $(q/2)^{1/4}$ where q is an upper bound on the number of types of elementary particles there can be (cf. §4, §4.2) and* $(q/2)^{1/4} < 10$ *is very likely. At* extremely *high power densities where general relativistic effects may ensue, we do not claim validity. If $T < 1000$ Kelvin, answer 1B will still be valid, but we no longer claim that answers 1A and 1B are equivalent; at low T the photons-only restriction actually* is *a restriction.*

For example. Suppose we have a 1 meter^2 window and 1 watt. Then the information flux one can transmit across the window using photons is no more than 2.15×10^{21} bits/second. (But see the remark at the end of §3.2.)

The Stefan-Boltzmann law will be re-derived in §3.3.

Thermodynamic justification for answer 1B.
Assumptions:

1. We assume that we are not allowed to transmit information by any means other than electromagnetic radiation (photons).

2. We assume the second law of thermodynamics (2LT) and the law of conservation of energy (COE).

3. Finally, we assume that once we have used some power, we are not allowed to recycle it. (Alternatively, let P in question and answer 1 denote a bound on the power actually being transmitted through the window, and then no assumptions about recycling need to be made.)[4]

Suppose it were possible to transmit a greater amount of information flux via photons, using less power. Beam in this information using power P through a window of area A into a cavity. The cavity's walls are made of a medium which can interact with and hence thermalize the photons in the cavity, but are assumed to have negligible thermal mass. (If the media are slightly imperfect, a modified form of our argument below will still hold.) After a long time passes, the photons bouncing around the cavity will approach thermal equilibrium. Which will be a blackbody radiation bath having at least as much entropy as the original beam, since according to thermodynamics, entropy always increases during an approach to thermal equilibrium. But then the blackbody bath will also have more energy density than the original beam. Hence, after equilibrium is achieved (or rather, after a "steady state" is reached) more thermal energy will be radiating out the window than we are radiating in, violating COE and solving the world's energy problems. QED.[5]

3.1 More careful discussion.

There are several flaws in the argument above which we will soon discuss more carefully.

The first flaw is the assumption that all information must be transmitted via photons. In fact, we will see that in certain regimes of low power density, it is possible to do better by using phonons. However, we will present parallel arguments of both a theoretical and empirical nature, designed to show that at sufficiently high power densities, the use of non-photon methods will not help.

The second flaw is the fact that the Stefan-Boltzmann $\sigma_{SB}AT^4$ law is not a true description of blackbody radiation. This manifests itself in four ways that the author knows about.

Firstly, we have neglected the presence of blackbody radiation in the form of massless particles besides photons, namely neutrinos and the hypothetical "gravitons." In practice radiation in these forms is negligible and could not be detected, since the coupling constants involved are so weak. Nevertheless, in theory, information could be transmitted in these forms and also would be radiated by an *ideal* blackbody, multiplying the Stefan-Boltzmann $\sigma_{SB}AT^4$ law by a small constant factor.[6]

Secondly, if the temperature T is too large then a blackbody would also emit power in forms other than massless particles, for instance, in the form of spontaneously created electron-positron pairs. The result would be a power output exceeding $\sigma_{SB}AT^4$. We very much doubt that any computer ever built will work at such high power densities, but considering the possibility nevertheless, we argue (cf. §4) that the Stefan-Boltzmann $\sigma_{SB}AT^4$ law should *still* be valid even at these very high temperatures, and even allowing non-photon particles of all kinds, if σ_{SB} is multiplied by a factor $q/2$ where $1 \le q/2 \le 10^4$. (Thus the upper bound of answer 1B would need to be multiplied by a factor $(q/2)^{1/4}$ in the range $[1, 10]$.)

---

[4] This assumption may be dropped. See §3.6.

[5] An alternative but equivalent proof would be to assume the power radiating into the window had to be the same as the power radiating out (by COE), in which case a violation of 2LT would result.

[6] Incidentally, even if a computer designer *were* willing to try to use neutrinos, in some regimes the constant factor information increase still would not be great, since neutrinos are spin $\pm 1/2$ fermions and thus [due to the Pauli exclusion principle] are capable of transmitting less information than photons, which are bosons. Gravitons are bosons [spin 2], but they are even harder to detect than neutrinos, indeed nobody has ever detected one.

Thirdly, if the temperature T and area A are too low, then the typical wavelength of thermal radiation will be larger than the radius of the blackbody and the radiating power will drop below $\sigma_{SB} A T^4$. (Answer 1B will still be a valid *upper bound,* though.) The requisite discussion of finite-size effects is in §3.3.

Fourthly, and finally, at *extremely* large temperatures, general relativistic effects (distortion of the metric of spacetime arising from the gravitating mass of the thermal radiation field) will ensue. These will distort the very notion of the "area" of the window and we refuse to consider such effects here, although see §4.2.

To see why it might be a stupid assumption to require transmission of infomation via photons only, consider a crystal of lithium. There are two stable isotopes of lithium of atomic weights 6 and 7. Thus each position in the crystal lattice can store at least one bit of information (by choice of isotope) and if the crystal is kept sufficiently cold (so that the atoms do not exchange places) and efficient error-correcting codes are used (so that any exchanges will be corrected for) this information will be stored, reliably, essentially permanently. The specific gravity of this crystal will be about 0.50, so that (we conclude) a cylindrical crystal with base of area 1meter^2 and of height 9cm will have mass about 45 Kg and store about 4.1×10^{27} bits. Pushing on this crystal, initially at rest, with 1 watt of continuous power, will accelerate it so that it crosses the 1meter^2 window (having traveled a total distance of 9cm) in about 1 second. Hence, we may successfully transmit information across the window, by throwing memory crystals, at a rate far exceeding 10^{22} bits per second.

The reply to this objection is that we have ignored the power and the time required to assemble the crystal and to read the information out of the crystal. If such reading and writing were required to be done using photons, we would be back to our original assumptions, since surely (?) more information could be transmitted more easily by beaming the photons through the window directly, than by using them wastefully in this indirect manner. We agree that by a scheme alternating long periods of information storage with short bursts of crystal throwing, it is possible to transmit information at a *peak* rate far exceeding our bound, which pertains only to the *sustained* rate of information transmission with a given power input.[7]

A more clear-thinking objection occurs when one realizes that answer 1A is the true answer to question 1, and answer 1B is invalid exactly when it is not equivalent (i.e. when non-photon methods are superior) to answer 1A.

Thermodynamic justification for answer 1A.
Assumptions: same as assumptions 2 and 3 in the justification of answer 1B.
Suppose it were possible to transmit a greater amount of information flux than the bound given in answer 1A, using less power. Transmit this information using power P through a window of area A into a heat bath whose temperature is T. The heat bath will then be gaining entropy at a rate $> P/T$, and after a long time passes and the system reaches steady state, the heat bath will be radiating heat out the window also at the rate P, corresponding to (even with the best heat removal process that can exist) entropy loss of $\leq P/T$. The heat bath is thus continually getting hotter despite the fact that it is radiating power out at the same rate it is radiated in, a contradiction. QED.

So the question now becomes: what are the fundamental physical limits on heat flux, and under what circumstances is blackbody radiation the best way to get rid of heat?

3.2 Fundamental physical limits on heat flux.

According to every high school physics teacher, there are 3 mechanisms for heat flux: conduction, convection (forced or unforced), and radiation. We recommend to the reader the discussions of heat conduction in [berm79] and [AIPH].

We now argue that at high power densities (T large), more heat flux between two heat baths at a given temperature T will be obtained by radiation through a near-vacuum, than would have been obtained by any kind of conduction or convection (driven or not) through any material.

Argument that radiation loses heat more efficiently than other methods when T is large. Any material consists of fundamental particles and conducts heat via vibrational modes (quantum states of

[7] Incidentally, since $x^{3/4}$ and $x^{1/4}$ are a concave-down functions, it is impossible to increase one's average information rate above the bound (7) by unsteady use of power or nonuniform use of area.

nonzero energy) of these particles. In normal metals, the main mechanism is via perturbations in electron momentum. In dielectric crystals such as diamond, it is via phonons (fundamental quantized vibrational modes of the crystal). In gases, it is via motion of the molecules. We claim, and our claim has great experimental and theoretic justification, that in hypothetical materials with infinite mean free path (i.e. of electrons, phonons, or gas molecules, respectively) the heat flux would be greater than in the real materials. Further, we claim that at high T, "ideal" gases (with infinite mean free path) will conduct heat better than any other material. However, even in such a gas, we are limited by the fact that each molecule carries off only about $k_B T$ worth of heat energy, and it only moves only at a speed corresponding to the mass of the molecule and the energy $k_B T$. Meanwhile, photons move at lightspeed and also carry on average about $k_B T$ worth of energy per photon. Further, the flux of gas molecules is proportional to T times the (fixed) gas density, while with photons in a vacuum, the flux is proportional to T^3, which eventually (with T large) becomes higher than in any real material.

Incidentally, it is observed in practice that at temperatures exceeding 1000 Kelvin, "radiative transfer" [berm79] [chan60], a diffusive movement of photons, begins to become more important than ordinary heat conduction through all known materials. (And at 6000 Kelvin, no solids exist, at 7000, no liquids exist, and by 10000, no molecules exist.) It is therefore not suprising that by eliminating the material, causing the photons to move unrestrictedly rather than diffusively, the rate of heat transfer is accelerated.

By use of *forced* convection, one might hope to exceed the heat flux obtainable by pure conduction, but in fact the pumping speed of the coolant over the surface to be cooled, must not exceed thermal velocity or one will find that the surface is being heated, not cooled. One therefore sees that for our hypothetical ideal gas coolant in which incoming molecules are *already* moving, without ever being scattered, at thermal velocities, convection will not help. QED.

This justifies answer 1B, at least for power densities corresponding to temperatures $T > 1000$ Kelvin.[8]

Characteristic energies ($k_B T = 1$ eV corresponds to a temperature $T = 10^4$ Kelvin) and characteristic power fluxes (4 mWatt, which is about a "TTL load," in a 30 gauge wire is 5×10^4 Watt/meter^2, corresponding to 1000 Kelvin) in present day computers seem to be in the range of validity of answer 1B.

However, at *low* power densities (corresponding to T small) the best way to transmit heat through a given-area window between heat baths at temperature T is probably by some combination of conduction or convection through the right kind of material. Therefore, we now discuss the best known ways to conduct heat at low power fluxes.

The best known conductors of heat in temperatures 0-2000 Kelvin are Silver, Copper, and Diamond. The peak thermal conductivity of Silver is 1.93×10^4 Watts/(meter-Kelvin), occuring at $T = 7K$. The peak thermal conductivity of Diamond is about 2700 Watts/(meter-Kelvin), occuring at about $T = 270$ Kelvin.

Schemes based on forced convection of substances supporting a phase transition can lead to even higher heat fluxes. At sufficiently low temperatures, however, I doubt that any convective scheme will be better than conduction through a good quality crystal.

In general if "$X \gg Y$" means "X is a better thermal conductor than Y at the same temperature," then

- solids $\gg$ liquids $\gg$ gases
- perfect crystals $\gg$ poor crystals $\gg$ amorphous solids
- At low temperatures: metals $\gg$ insulators
- normal metals $\gg$ superconductors
- At low T (< 10 Kelvin), thermal conductivity in metals is proportional to T but metals feature a slow decrease in thermal conductivity at high T, like $1/T$
- At low T (< 300 Kelvin), thermal conductivity in dielectrics is proportional to T^3 but dielectrics too feature a slow decrease in thermal conductivity at high T, like $1/T$.

[8]The dividing line of 1000 Kelvin was picked somewhat arbitrarily and should not be taken too seriously.

- "Thermal conductivity" of a vacuum is proportional to T^3 at all T.

All of these empirical facts are easy to understand. Solid crystals are better than amorphous solids, liquids, and gases since the heat is transmitted by phonons (essentially sound waves) which travel in a wavelike manner at the speed of sound, and in a perfect crystal near temperature 0 and with low phonon intensity, phonons never scatter (i.e. have infinite mean free path). Metals are better than insulators at low T since the heat is mainly transmitted by conduction-band electrons, which have enormous zero-point Fermi energies corresponding to a velocity of $\approx 10^6$ meters/sec for Cu at $T = 0$ Kelvin; (and these electrons also have infinite mean free path in a perfect crystal); meanwhile in insulators the heat is conducted by phonons, less of which are available in the phonon gas at low T, and which travel more slowly. In superconductors, there is no Fermi energy to help since the conduction electrons are in Cooper pairs which are "Bosons," hence normal metals are found to be much better heat conductors than superconductors. At high T, metals scatter electrons vigorously, gradually becoming bad heat conductors, but meanwhile more phonons become available in the temperature-T phonon gas in Diamond, so it wins. At $T > 300$ Kelvin phonon Debye cutoff due to the limited number of modes ($3N$) in a Diamond crystal occurs, so the T^3 growth ceases and changes to a gradual $1/T$ decline due to thermal phonon scattering. At $T = 3000$ Kelvin most things have vaporized and the battle is between the gases and the photons in vacuum. Eventually, at high T, the photon density in the vacuum grows so large, and the fact that the mean-free-path in gases is small – forcing heat transfer to be slow and diffusive – becomes so sorrowful, that the photons overwhelm all heat-transfer competition.

All this leads to the conclusion that at sufficiently low power fluxes, corresponding to "temperatures" of 7 Kelvin and 270 Kelvin respectively (assuming one were so foolish as to build a computer operating at such low power fluxes) it should be possible to transmit more information by the use of conduction electrons in silver and phonons in diamond, respectively, than one can transmit by means of electromagnetic radiation with the same amount of power.

It is in fact possible to analyse blackbody *phonon* radiation in a homogenous isotropic medium with speed of sound c_s; the procedure is almost exactly analagous to the analysis of blackbody photon radiation (cf. §3.3), except that phonon instead of photon modes need be used throughout. The result is the "Stefan-Boltzmann law for phonon blackbodies" stating that the acoustic power radiated from a blackbody, per unit area, is

$$\sigma_{SB} \frac{qc^2}{2c_s^2} T^4 \tag{8}$$

where $q = 1$ if only compressional waves are permitted, but $q = 3$ if 2 orthogonal types of shear waves are also permitted. Thus, since $c_s \ll c$, phonon blackbodies emit far *more* power than photon blackbodies at a given temperature. It seems suprising that a slower characteristic speed c_s actually *helps* thermal conduction. The explanation is that more phonon modes exist at a given energy than photon modes. The nth X-photon mode in an $L \times L \times L$ box has energy proportional to $\hbar cL/n$; thus a factor $(c/c_s)^3$ more modes are available at given energy in the phonon gas. This overwhelms the fact that photons move c/c_s times more quickly, and allows far more information flux, or heat flux, in the phonon gas. However, the law (8) only holds at temperatures considerably smaller than the Debye temperature where mode cutoff occurs. These temperatures are on the order of 300 Kelvin for the usual materials. Meanwhile photons do not suffer from a mode-number cutoff and eventually win.

We reiterate that in fact, *photons are not optimal* at low temperatures. At power fluxes equivalent to room temperature, it would seem that the greatest possible information flux would be obtained not by using photons, but rather by transmitting sound waves through a diamond. Indeed, in the numerical example we gave before (1 watt, 1 square meter) the equivalent "temperature" is only 65 Kelvin, so that actually 1A and 1B are in serious disagreement! (We should have used a power density 10^6 times higher to achieve agreement.) By transmitting sound waves through a diamond, assuming $c_s = 18000$ meters/sec and $q = 1$, we could have transmitted information $2^{-1/4}\sqrt{c/c_s} = 109$ times more quickly than the photon bound of answer 1B.

3.3 Blackbodies of finite size

One may ask about diffraction effects. If our photons are being transmitted through a finite window, then diffractive effects will presumably reduce the information content of the photons, an effect we had not considered. (Of course, answer 1B is still valid as an upper bound. More strongly: diffraction is nonexistent if the "window" is purely imaginary.) The best to way answer such objections is to state that, when deriving answer 1B, we should have used the power radiated by a *finite* blackbody instead of $A\sigma_{SB}T^4$.

We now sketch how one might analyse such a blackbody. More precisely, we will analyse a photon gas in a finite container; the relationship between this and the radiation of a finite blackbody is not so clear.[9]

Namely, consider a volume-V cubical cavity with reflective walls, filled with a gas of photons in thermal equilibrium with a heat bath at temperature T. We wish to compute the energy density in this finite cavity in order to compare it with the blackbody energy density in an infinite cavity, which is the usual sort of approximation. The answer is that the photons must be selected from a discrete set of Fourier modes instead of a continuous set, namely the permissible wavevectors lie on a lattice Z^3 rather than in R^3. Specifically, a photon is specified by 3 nonnegative integers n, m, p and a Boolean variable b ([mors53] 13.3.45 page 1849) and its energy is $\pi\hbar c\sqrt{n^2+m^2+p^2}/V^{1/3}$.

Bose-Einstein statistics then applies[10] to see that the energy density is

$$\frac{\pi\hbar c}{V^{4/3}}\sum_{m,n,p,b}\frac{\sqrt{n^2+m^2+p^2}}{\exp(\frac{\pi\hbar c}{k_B V^{1/3}T}\sqrt{n^2+m^2+p^2})\ -\ 1} \tag{10}$$

where the sum is over all integer lattice points (m,n,p) and the Boolean variable b, and where we require that if $b=0$ then $m>0$ and $n,p\ge 0$ and $n+p\neq 0$; and if $b=1$ then $m\ge 0$, $n>0$, $p>0$.

This sum arises instead of the usual octant integral

$$\frac{\pi\hbar c}{4V^{4/3}}\int_0^\infty \frac{x\cdot 4\pi x^2\,dx}{\exp(\frac{\pi\hbar c}{k_B V^{1/3}T}x)-1} = \frac{k_B^4T^4}{\pi^2\hbar^3c^3}F_4 \tag{11}$$

where

$$F_a=\int_0^\infty \frac{y^{a-1}\ dy}{e^y-1}=\zeta(a)\ \Gamma(a). \tag{12}$$

We have $F_4=\pi^4/15$ so that the energy density of the blackbody radiation inside the cavity as $V\to\infty$ tends to

$$\frac{4}{c}\sigma_{SB}T^4,\ \text{ where }\ \sigma_{SB}=\frac{\pi^2}{60}\frac{k_B^4}{c^2\hbar^3}. \tag{13}$$

From $1/T=dS/dE$ we deduce that the entropy density is $16\sigma_{SB}T^3/(3c)$. The flux of energy out of a surface, for instance out of a small aperture in the cavity, is then $\sigma_{SB}T^4$ since the photons are traveling at speed c and the remaining factor of 4 is chopped by a factor of 2 due to the fact that we only count flux *out* of the surface, and by another factor of 2 due to isotropy – the average height of a unit hemisphere is $1/2$. The corresponding flux of entropy is $4\sigma_{SB}T^3/3$. Viewing the aperture of area A as a "blackbody" of surface area A, we have recovered the Stefan-Boltzmann law.

It should be clear that the relative difference between the sum (10) and the integral (11) is relatively arbitrarily small if the dimensionless quantity $\frac{\pi\hbar c}{k_B V^{1/3}T}$ is small. Hence, for a window (of rotund shape – not long and thin – so that our inside-a-cube viewpoint has *some* relevance; we will take $A^3\approx V^2$) diffraction

[9] For a much more extensive discussion of thermal radiation in finite containers, see [balt72].

[10] Alternatively, to get a firm, but weak, lower bound on the entropy, we may proceed as follows. Only allow at most one photon to occupy a mode. Then a set of N modes exists whose total energy is $\le E$, where

$$N-O(N^{2/3})=12^{-1/4}\pi^{-1/2}(\hbar c)^{-3/4}V^{1/4}E^{3/4}. \tag{9}$$

The entropy within the cavity is then at least $Nk_B\ln 2$, and the information content is at least N bits.

effects will be negligible, and answer 1B will be valid, if

$$\frac{\pi\hbar c(\sigma_{SB})^{1/4}}{k_B(AP)^{1/4}} \ll 1 \tag{14}$$

(Observe that the left hand side is a ratio of two lengths.)

In our previous example with $P = 1$ Watt and $A = 1$ square meter, this ratio is 1.11×10^{-4} so the approximation should indeed be valid. From (14) we see that diffraction effects will only insignificantly lower the information flux ceiling in the range

$$AP \gg 1.5 \times 10^{-16} \text{ WattM}^2. \tag{15}$$

On the other hand, it would seem that in the range

$$AP \ll 1.5 \times 10^{-16} \text{ WattM}^2, \tag{16}$$

where diffraction effects *dominate*, the sum (10) will be well approximated by its 3 equal biggest terms, i.e. the photon gas will be dominated by the 3 equal [and minimal] energy modes, and indeed there will usually be nothing in these modes either! Note: The entropy in the box is nonzero, even when the energy is too small for even a single photon. This is an artifact of Bose statistics, which may be thought of [for Jaynesians who wish to avoid the use of heat baths] as assuming that we have an ensemble of a large number of boxes; just because the *average* energy of a box is too low for a photon, does not mean that none of the boxes contain photons. Thus, thinking of an ensemble of $B \to \infty$ boxes, where BE is sufficient energy for a pB photons, $0 < p \ll 1$, the information content of a single box is $-p\lg(3p) - (1-p)\lg(1-p) \approx p\lg(3p/e)$, where $e \approx 2.71828$, bits, i.e.

$$\frac{V^{1/3}E}{\pi\sqrt{2}\,\hbar c}\lg\left(\frac{3\pi\sqrt{2}\,\hbar c}{e\,V^{1/3}E}\right) \tag{17}$$

bits, leading to a typical flux of information (across surfaces of area $A = V^{2/3}$ within the box) of

$$\frac{A^{1/2}P}{\pi\sqrt{2}\,\hbar c}\lg\left(\frac{3\pi\sqrt{2}\,\hbar c^2}{4e\,A^2 P}\right), \tag{18}$$

where $E = 4VP/c$. This suggests the use of this expression as a replacement for (7) in the low power limit.

In the usual previous work in information theory, where information transmission over "realistic, band-limited" analog channels was considered [shan59] [slep61] [land62], it was found that the amount of information which could be transmitted grew like the logarithm of the available power as power became large, whereas we see from answer 1B that in fact, far more information flux is possible – growing like $P^{3/4}$. It appears that previous notions of bandlimiting are irrelevant when wavelengths far smaller than the dimensions of the channel between the two communicators are readily accessible.

Probably a better way to treat *radiation* by blackbodies of finite size is to consider a temperature-T thermal radiation field in a very large cavity escaping slowly from an aperture of (comparatively small) area A in its wall. The only problem then is to compute the transmission coefficient of the aperture as a function of ingoing and outgoing wavevector. If the aperture is such that a fraction $\frac{A}{A+\lambda^2}$ of the incident thermal power with wavelength λ emerges from the aperture (cf. [born80]), then in the low-T limit, the radiated power will behave proportionally to

$$\sigma_{SB}\frac{(k_B)^2 A^2}{\hbar^2 c^2}T^6. \tag{19}$$

This suggests that in the low power limit, the bound of answer 1B should be replaced by

$$\frac{I}{A} \le \text{const.} \times \frac{(\sigma_{SB})^{1/6}A^{1/6}}{(k_B)^{2/3}\hbar^{1/3}c^{1/3}}\left(\frac{P}{A}\right)^{5/6}. \tag{20}$$

instead of by (18).

3.4 Answer 1B is a "tight" bound

Having been convinced that 1B is a valid upper bound on information flux, one may ask whether it is in any sense tight, i.e. whether it is actually physically possible to transmit and receive information at a rate anywhere near this upper bound.

The answer in some sense is "yes," although this sense may not be quite the one the reader wanted. Namely: if the entropy of the corresponding black body radiation is S, then there really are about $\exp(S/k_B)$ possible configurations of photons in spacetime passing through the window, and each configuration is in principle distinguishable from each other, i.e. uncertainty principles do not forbid such distinguishment. Hence that much information really can be made to come through the window.[11]

One may ask whether it is possible to *detect* all of this information. The answer is again yes.

Imagine an a extremely large hemisphere centered at and lying to one side of, the window. The inside surface of this hemisphere is coated with spectroscopically analysing, time-of-arrival recording, polarization recording, photon detectors, all cooled to near absolute zero. The hemisphere is so large and the detectors so small that typically only one photon arrives at any detector at a time. It seems that such a collection of detectors, en masse, could in principle detect at least a constant fraction of the information coming through the window.

Finally, one may ask whether it is possible to *transmit* this much information. Yes, in the sense that a thermal blackbody is generating that much (albeit random) information, and in theory one could have a non-random blackbody (i.e. whose state is known) doing the job instead.

3.5 Transmission of information in the presence of thermal noise

One may also ask "what about noise?" Noise has been ignored above, which is perfectly acceptable if we only are seeking an upper bound, but hardly if we are considering how tight it is. We now introduce noise by postulating that there is an underlying blackbody radiation bath, of temperature T_0, filling the vacuum we are transmitting our photons through. This is an unavoidable consequence of the fact that our computer has some finite temperature T_0. If $T_0 > 0$, then this noise will corrupt our signal. The question is then what fraction of the information remains. Assuming we encode our information as radiation-field states by the use of optimal-rate error correcting codes, the answer is "the channel capacity of free space."

Shannon's bounds (5) tell us that this capacity is bounded below by

$$\frac{4}{3\ln 2}\left[\frac{(\sigma_{SB})^{1/4}}{k_B}\,(\frac{P}{A})^{3/4}-\frac{\sigma_{SB}T_0^3}{k_B}\right] \tag{21}$$

Thus the presence of noise hardly affects the transmission capacity in the high P limit where the $P^{3/4}$ law is valid.

3.6 "Recycling" transmitted power?

In answers 1A and 1B we had to assume that "recycling" transmitted power was not allowed. In this section we will show (by use of later results; you'll have to read answer 2A before you can comprehend our argument) that this assumption was not really necessary.

So, we ask: Is it possible to transmit information using power P, then receive the information, then "recycle" the power, using it to transmit more information, perhaps, thus beating the $\dot{I} < O(1 + P^{3/4})$ bound of answer 1B?

The answer is no. (You can recycle power *if* you were transmitting more than was required. But you'll never beat 1B or 1A.) If you could, then since it is impossible to get rid of entropy (2nd law of thermodynamics) you would have to accomplish this by taking the information contained in your power P,

[11] Thus one could determine the state and look it up in a giant table of all possible states of the radiation field. "Aha," one would say, "this is state number 5743 in my table of the $10^{10^{10}}$ possible states! Therefore, the fellow on the other side of the window was trying to transmit the binary number 1011001101111."

and compressing it into a smaller amount of power $P' < P$, thus having power $P - P'$ left over to use. However, it is *impossible* to compress information into a smaller amount of mass-energy (by answer 2A) per volume, than is in a time-t chunk of the power-P beam. Thus you would only be able to accomplish said compression (i.e. into a smaller amount of energy) by increasing its volume. But the volume of the computer is bounded, so that approach is unsustainable. So, you have to radiate it away into infinite space. That works, but you can't radiate any faster than the information is already coming in, per unit area (cf. §3.2), so the best you can do is to get rid of it all.

Thus it is thermodynamically impossible to recycle the power. QED.

According to R. Landauer [land91], "...in communication, as in other areas, inevitable minimal dissipation arises only when information is discarded." At first the reader may think that the argument above refutes this statement, but actually it supports it. That is because it is essential in the argument above that the information be "received." (Meaning: stored, used, or discarded. Different view of the same thing: to recycle some power to carry new information, you have to erase the information already inside of that power.) Consider a beam of information-carrying light traveling (thanks to a few mirrors) in a cyclic path many millions of times. The total energy used is miniscule, but the total information being transmitted through a window (admittedly, the same information, many times, although one could interpose some reversible optical transformation by the use of funny-shaped mirrors) is very large. In this counterexample the power *is* being recycled, but the information is never being "received," so the anti-recycling argument still stands.

4 Memory Capacity

The second fundamental computational quantity of interest is memory.

Question 2. "How much information I can be stored in a memory whose gravitating mass M is confined within a region of volume V?"

Notice that by speaking of "gravitating mass" we have tacitly allowed the interconversion of mass and energy inside the memory, and avoided having to worry about where "zero" lies on our energy scale. Also, note that any memory which can always be accessed in a time τ must occupy a volume V with $V \leq 4\pi(c\tau)^3/3$ (assuming that space is Euclidean), so that the volume constraint could equally well have been replaced by an access time constraint.

An answer is as follows. Assume that there are at most a finite number q of kinds of elementary particles in the universe. Throughout this paper we adopt the unconventional semantic convention that we regard particles with different quantum numbers (e.g. different "spin," "charge," or "strangeness") as actually being different "kinds" of particles. We must assume that q is finite, otherwise an infinite amount of information could be stored as the identity of a single particle. (Although the cornucopia of known elementary particles changes seemingly all the time, it still seems likely that $q \leq 2 \times 10^4$. Certainly $q \geq 10$.)[12] Then:

Answer 2A. *Let I be the number of bits in a memory of volume V and gravitating mass M. Then I is bounded by $(q/2)^{1/4}$ times the entropy of a volume-V box full of blackbody radiation, whose temperature T is selected so that the radiation has total mass-energy M. Thus in the regime of validity of the Stefan-Boltzmann law, the maximum information density I/V, in bits per unit volume, which can be stored in a memory of mass M and volume V is*

$$\frac{I}{V} \leq (q/2)^{1/4} \frac{16}{3\ln 2} \frac{(\sigma_{SB})^{1/4} c^{5/4}}{k_B} \left(\frac{M}{V}\right)^{3/4} = (q/2)^{1/4} \frac{16\sqrt{\pi}}{3 \cdot 60^{1/4} \ln 2} \left(\frac{cM}{\hbar V}\right)^{3/4}. \tag{22}$$

Note that if there were only two kinds of particles in the universe ($q = 2$) namely two polarizations of photon, then this bound would be precisely the bound arising from blackbody radiation. In that case, as a

[12] If there were more than 1000 kinds of massless (or very small mass) bosons, then the 3 Kelvin thermal background of these particles, which presumably would exist due to the big bang, would have more mass-energy than the current estimate (cf. section 4.2) of the mass of the universe. The fact that the consequent gravitational effects have not been observed constitutes convincing experimental evidence that there are < 1000 kinds of massless bosons. There are also good reasons to believe (see §4.2) that there are $< 10^4$ kinds of elementary particles, period.

numerical example, consider a 1 cubic meter region of space containing 1000 Kg of mass-energy. With $q = 2$, the bound tells us that no more than 6.0×10^{34} bits can be stored in this space. If $q = 2 \times 10^4$, this should be multiplied by 10.

Either way, this far exceeds the number of bits which can be stored (as before) in an equivalent size Lithium crystal with 1 bit per atom in the crystal; $< 10^{29}$ bits may be stored with Lithium.

Justification of answer 2A.
Assumptions:

1. We assume De Broglie's relation $h = p\lambda$ between the momentum p and wavelength λ of particles.
2. We assume the validity of the formula from special relativity
$$E = \sqrt{m_0^2 c^4 + p^2 c^2} \tag{23}$$
giving the mass-energy (here expressed as an energy E) of an object of rest mass m_0 whose momentum is p.
3. We assume the "ideal gas conjecture:" the entropy of a set of N particles with given masses, fixed total kinetic energy, and constrained within a region of fixed volume, cannot be made larger than the entropy which would arise from the assumption that the particles were noninteracting. (This assumption makes sense theoretically, since any interaction among particles tends to decrease the accessible volume of phase space; and also experimentally, since everything seems to behave like an ideal gas at its highest entropy state. Discussion in §4.1.)
4. Gravitational effects distorting the metric of space, are negligible.
5. Finally, we've assumed, as we just explained, that q is an upper bound on the number of kinds of elementary particles there are, including accounting for all the quantum numbers intrinsic to the particle.

Assume the memory is contained in a $L \times L \times L$ cubical box with reflecting walls, $L^3 = V$. (The cubicity of the box and the boundary conditions are pretty irrelevant, we merely are making this assumption for simplicity.)

Then the quantum state of any set of particles/waves inside the box is completely specified by the occupation number of each mode. A "mode" is specified by 4 integers n_1, n_2, and n_3 (n_i gives the number of wavelengths in direction i) and $n_4 = 1, 2, \ldots, q$ telling what sort of particle this is. The energy of each mode is
$$E(n_1, n_2, n_3, n_4) = c\sqrt{m_0^2 c^2 + h^2(n_1^2 + n_2^2 + n_3^2)/(2L)^2}. \tag{24}$$

Since the energy of a mode is a monotonically increasing function of the rest mass m_0 of the particle in that mode, we may without loss of generality (since we are only seeking an *upper bound* on entropy at fixed energy) assume the rest masses of all the elementary particles in the universe, are 0. Since any set of occupation numbers achievable by fermions are also achievable by bosons, we may also assume, again without loss of generality for the purpose of deriving an upper bound, that all the elementary particles in the universe are bosons. In that case, we have precisely q copies of the standard picture for (polarization-free) blackbody radiation inside the box. This gives answer 2A. The justification of the *formula* of answer 2A arises from Bose statistics as usual (assuming the energy density is high enough, and the dimensions of the box large enough, that finite-size effects may be neglected; see (3.3) for the energy density of blackbody radiation and nearby discussion for the entropy density).

Our earlier remark that "the boundary conditions and the cubicity of the box are irrelevant" is correct in the high-energy limit, since the only thing that matters is the number of eigenmodes of the Laplacian inside the box, which asymptotically depends only on the volume of the box and not on its shape [kac66]. It may not be correct at low energies or for boxes some of whose dimensions are very small. QED.

4.1 The "Ideal gas conjecture" – discussion

I was planning on proving this conjecture in this section, but haven't managed it yet... so here are some remarks, anyway.

Here is a precise mathematical formulation of this conjecture. (We formulate it nonrelativistically, and on the surface of a sphere, for simplicity.)

Ideal gas conjecture: Let S be the surface of a sphere in some Euclidean space of fixed dimensionality $d+1$. For $x \in S$, let $V(x)$ be any piecewise differentiable real valued "potential function" such that the Schrödinger operator $H = -\nabla^2 + V$ which operates on complex differentiable functions $\Psi : S \mapsto \mathbf{C}$ has a nowhere dense spectrum with a finite greatest lower bound. Indeed, let its eigenvalues be $\lambda_0 \le \lambda_1 \le \lambda_2 \le ....$ Let

$$f_V(N) = \lambda_N - \lambda_0, \quad N = 0, 1, 2, ... \tag{25}$$

We conjecture that $f_V(N) \gtrsim f_0(N)$.

A stronger version of this conjecture would replace the "$\gtrsim$" (asymptotic inequality) by "$\ge$" (strict inequality).

In other words, the constant potential conjecturally has the most spectral lines, and funny structure in the potential tends to decrease the volume of phase space that is accessible and hence decrease the number of available states.

Partial and suggestive results:

1. An old conjecture of Rayleigh that, of all d-dimensional "drums" of a given d-volume, the d-ball has the "gravest tone," has been proven [boss86] and even extended [boss88]. This suggests that funny structure is unhelpful.

2. The asymptotic behavior of λ_N for Schrödinger's equation on any compact region with any given *bounded* potential function (whose mean value, wlog, is 0), depends *only* on the d-volume of that region, e.g. [kac66]; but (as may be shown using the Schrödinger variational principle with Ψ = constant as a test function) the ground state eigenvalue is lower. Thus our conjecture is asymptotically true for bounded potentials, and indeed the strict form of it is true for all except perhaps a finite number of λ_n.

Presumably the same is true for unbounded potential functions, so long as their singular behavior is not too severe. (How severe? When $d = 1$, I suspect the answer is that it will suffice if the singularities are nowhere dense and $O(x^{-2})$ in character.)

3. I haven't seen these references yet, but they look interesting:
C.Bandle: Isoperimetric inequalities and applications, Pitman 1980
Fefferman and Phong: Bull. AMS 9.2 (1983) 129-206
Kirsch and Simon: Commun.Math.Phys. 97,3 (1985) 453-460
B. Simon: J. Funct. Anal. 53,1 (1983) 84-98
Glaser et al.: Commun.Math.Phys. 59,2 (1978) 197-212
A. Martin: Commun.Math.Phys. 129,2 (1990) 161-168.

4. The case $d = 1$ may be solved exactly for a piecewise constant potential (although we have to solve a system of linear equations to find the coefficients of the exponential waves in each piece, and finding the eigenvalue then requires the solution of a transcendental equation). This allows some strengthening of remark 2 in this case.

5. One can also gain insight about the $d = 1$ case in a different way. We may write the Schrödinger operator H as a matrix of countably infinite order by using the complex exponential waves as a basis. For a nonconstant potential V (wlog with zero mean) this matrix be Hermitian, will have $1, 4, 9, 16, ...$ on the diagonal, and each super and subdiagonal will have constant entries (Toeplitz).

4.2 Discussion. No limit to density? Black hole thermodynamics.

Again we have found, in answer 2A, a 3/4 power law.

Based on what we've learned so far, it would appear that there is no real limit on the density of a memory, other than those set by limitations on the stability of matter at extremely high densities, which would cause too-dense memories to collapse into black holes, and (cf. §4.7) concerns about the permanence of the stored information. Memories which store a bit in molecular dimensions are possible (the DNA inside our cells), and indeed bits may be stored in atomic dimensions (the Lithium crystal example; also storing information in the vibrational modes of a perfect crystal theoretically could achieve subatomic-size bits). It would seem, though, that much more is possible. Thus in the numerical example above we saw that photon modes in a box could store information $> 6 \times 10^5$ times more densely than Lithium, using the same density of mass-energy. By going to higher mass-energy densities, more should be possible: Memories (or indeed, whole computers) located on neutron stars could be a factor of $\approx 10^4$ times faster and smaller (in terms of linear dimensions) than functionally equivalent molecular-size computers (although they would require a correspondingly larger energy consumption per bit operation), and these in turn could be considerably faster and smaller than today's computers, whose gates are much larger than atomic dimensions.

So it would seem that to understand the *real* limitations on computer memory, one would need to understand the fundamental limitations on the stability of matter and the strongest forces and smallest particles there are [lieb90].

We have made no attempt to consider general relativistic effects which should become important at very high densities. However, some attempt *has* been made by J. Bekenstein [beke73] [beke74] [beke81] and S. Hawking [hawk74], which we will now survey.

First, one should realize that the presence of "black holes" in general relativity seems to contradict the second law of thermodynamics, since, to get rid of entropy, one may merely toss it into a black hole. Due to "no hair" theorems [hawk79] [beke72a] [beke72b] [teit72a] [teit72b] [teit72c] [hart71] [hart72] [robi75] black holes are completely characterized, as far as an external observer is concerned, by exactly three parameters: their mass, charge, and angular momentum. The information-bearing degrees of freedom in matter or radiation tossed into the hole are thus eviscerated. According to classical general relativity, this evisceration is permanent, since nothing can escape from a black hole. To make a heretical machine capable of generating useful work from thermal motion, one may connect the "hot side" of a heat engine to a heat bath and the "cold side" to a radiator radiating thermal radiation into a black hole.

The work of Bekenstein and Hawking, combining quantum mechanics and general relativity, has begun to resolve this paradox.

Hawking [hawk74] [hawk75] [hawk76] ascribed a "temperature" T to black holes of mass M (the formula below assumes a nonrotating uncharged hole)

$$T = \frac{\hbar c^6}{8\pi M k_B G} \tag{26}$$

(More generally, for a rotating charged black hole, the temperature is proportional to the "surface gravity" of the hole.) Hawking found that black holes would actually radiate thermal radiation in all forms (photons, but also neutrinos, gravitons, lepton pairs, etc.) at their characteristic temperature, causing black holes to gradually lose mass-energy. This radiation could be thought of as mass-energy quantum mechanically "tunnelling" out of the black hole (a potential barrier insurmountable without quantum effects) in order to raise its entropy. A black hole would thus have a finite lifetime, which Hawking estimated as 10^{71} seconds for a hole the mass of the sun. Other remarkable consequences of the existence of Hawking's radiation include: the nonconservation of Baryon and Lepton number, and the impossibility of the "oscillating universe."

Bekenstein [beke73] [beke74] had earlier ascribed an "entropy" to black holes and argued it had to be proportional to the surface area of the hole. A putative value for the precise constant of proportionality was derived by Bekenstein, but it was incorrect. From Hawking's temperature formula, the entropy of an uncharged nonrotating hole is

$$S = \frac{4\pi k_B G}{\hbar c} M^2 \tag{27}$$

More generally, for a rotating charged black hole, the entropy is proportional to the surface area A of the hole's event horizon:

$$S = \frac{k_B c^3}{4\hbar G} A \tag{28}$$

and A is determined by the charge, mass, and angular momentum of the hole.

Finally, the *Generalized second law of thermodynamics* formulated and conjectured by Bekenstein, and proven[13] to follow from the ordinary second law and the laws of Hawking radiation by Hawking [hawk76], states that *The total entropy of the universe outside of black holes, plus the entropy of all the holes, is nondecreasing.*

Again this is only a statistical statement, just as is the usual second law of thermodynamics. All these developments since became a small industry among theoretical physicists [beke75] [gibb75] [gibb76] [hawk77] [hawk79] [hart76] but are now "well-accepted and uncontroversial[14]."

Finally, Bekenstein [beke81] argued that these statements about black holes in fact lead to a remarkable upper bound on the entropy S of *any* physical system of gravitating mass M which may be enclosed by a surface of area A:

$$\frac{S}{M} < \frac{\sqrt{\pi} k_B c}{\hbar} \sqrt{A} \tag{29}$$

Bekenstein's derivation of this bound is as follows. From classical general relativity [beke74] (see also [beke73] [bard73] [hawk72]) it is found that by dropping a physical system of mass M and enclosable by a surface of area A into a very much larger Schwarzschild black hole, starting with the body very near to, and stationary with respect to, the hole, the surface area of said hole will increase by an amount

$$\frac{4\sqrt{\pi}G}{c^2} M \sqrt{A}. \tag{30}$$

Then due to the generalized second law of thermodynamics and the expression (28) for the entropy of a black hole in terms of its surface area, the upper bound (29) must hold, otherwise a contradiction (decrease in generalized entropy) would occur during this process.

Bekenstein, in deriving this bound as sketched above, made four assumptions (aside from assuming the validity of the generalized second law and of classical general relativity) none of which were ever stated nor justified.[15] First, he assumed that the entropy of any gravitational or other radiation generated during the infall into the hole, and escaping from it, would be negligible. Second, he did not actually derive the expression (30) for the area increase during his infall process. He in fact derived the expression $\frac{8\pi G}{c^2} MR$ where R was the distance from the center of mass of the body (initially entirely outside the hole) to the black hole's event horizon. Then, without justification, he assumed that $A \geq 4\pi R^2$. This is true in Euclidean geometry, if we assume the body is initially oriented so that its center of mass is as close to the hole as possible. But it is not entirely clear that it is true in the highly distorted nonEuclidean metric near the hole[16]. Finally, Bekenstein's argument only proves that the entropy of a system *initially touching a black hole* obeys the bound (29). It does not prove that the entropy of *any* system obeys this bound. To prove this, one must also show that there is some way of lowering the system until it is stationary and touching the hole with its center of mass as close to the hole as possible, such that during this lowering process: (1) the area of the hole does not increase, (2) the area A of a surface enclosing the body does not increase, (3) the mass-energy M of the body does not increase, (4) the entropy of whoever and whatever is supervising the lowering process, does not increase.

It is quite possible that these assumptions underlying Bekenstein's bound could be justified, I am merely saying that neither he nor his critics [unru82] [unwi82] [page82] [beke82] [beke83] [beke88] have ever done so. In the rest of this section, we will pretend that this has been done and we will explore the consequences of Bekenstein's bound, assuming its validity. Translating Bekenstein's bound into information-theoretic entropy, we see that,

[13] Actually, Hawking allows the possibility that the right hand side of (28) should have some positive constant added on to it, but says it is "plausible" that the constant should be zero.

[14] Bryce De Witt. It should perhaps be recalled that nobody has ever observed, with certainty, either Hawking radiation or a black hole.

[15] His fourth assumption that the ground state mass-energy density of any object was always nonnegative now seems to have been justified [witt81] [choq84].

[16] Actually, I have to check; it seems to me possible that for very large holes, the geometry near the event horizon is in fact Euclidean. If so, that is one less objection.

Answer 2B. *If A is the area of a surface enclosing a computer memory (of volume V and gravitating mass M) storing I bits of information, then*

$$I < \frac{\sqrt{\pi}c}{\hbar \ln 2} M \sqrt{A}. \tag{31}$$

The area A grows proportionally to M^2 when the system collapses to a (nonrotating uncharged) black hole [cf. (27-28)], since we certainly require the "enclosing" surface to include the gravitational radii of the mass inside. However, in the usual circumstance where gravitational effects are negligible and Euclidean geometry applies, A does not depend on M and may be taken proportional to $V^{2/3}$ as usual.

An elegant feature of answer 2B, is that it needs to make *no* assumptions about the nature of the fundamental particles of the universe. (In answer 2A we had to assume there were only a finite number q of possible types of particles, although we didn't need to assume anything else about them.) Still, answer 2B will be a weaker bound than answer 2A in almost every conceivable application. For example, in the previous numerical example, (1000 Kg inside 1 cubic meter) Bekenstein's bound (assuming the 1-cubic meter region is a sphere) gives $I < 1.6 \times 10^{46}$ bits while the bound of answer 2A was $I < 6.0 \times 10^{34}$ bits.

Indeed, as Bekenstein [beke82] argued, his bound implies an upper limit on the number q of kinds of elementary particles there are in the universe. As usual for us, we may let the "kind" of a particle include all the measurable quantum number information about a particle, not just its name. Then (to correct and simplify Bekenstein's argument), Bekenstein assumed that any particle of mass-energy M could be confined inside a radius-R spherical region whose circumference was at most K times greater than its Compton wavelength:[17] $R = K\hbar/(Mc)$ (where K is some dimensionless real number of order 1), without much increase in its mass due to the energy of confinement, whence it follows that

$$q \le \lfloor \exp(2\pi K) \rfloor. \tag{32}$$

In particular, with $K = 1$ we get $q \le 535$ and with $K = 2$ we get $q \le 286751$.

As another numerical example, we will now compute an upper bound on the amount of information one could possibly store in the entire universe. Assume that the universe is contained in a sphere of radius 10^{10} parsecs$=10^{26}$ meters. Observations of galactic rotation speeds have shown that only 10-20% of the mass in the universe is "visible," (i.e. stars); the rest is mass of an unknown form ("missing mass"). Anyway, present opinion is that the mass density of the universe, including both visible and invisible mass, is at most 5×10^{-28}Kg/meter^3. If we compute area and total mass using the usual Euclidean formulas for the volume and surface of a sphere, we estimate that the total mass of the universe is $M = 2.1 \times 10^{51}$ Kg and the total surface area of an enclosing sphere is 1.3×10^{53} meters^2. Answer 2B then allows us to conclude that the total amount of information that could be stored in an object the size and mass of the universe, is bounded by 5×10^{120} bits.[18]

Meanwhile, a complete solution of the oriental game of "Go" (recording, for each of about 3^{361} positions, the best score that could be forced when starting play from that position) would require considerably more space than this to write down.

4.3 Answers 2A and 2B are "tight."

Answer 2B is tight in the sense that it is actually achieved by the "entropy of a black hole." On the other hand, it is impossible to extract information at will from a black hole, so answer 2B is hardly tight in the

[17] This argument of Bekenstein's would fail if there were an infinite sequence of particles $p_1, p_2, \ldots, p_n, \ldots$ such that the "absolute constant" K actually depended on what particle it was: $K = K_n$, and such that $K_n \ge (\ln n)/(2\pi)$. Another logical flaw is that we are both pretending that there is only one spherical region, which could have contained any of q particles, and also that there is a different sized box for each particle mass. This flaw is repairable if one assumes that the lighter particles can be confined inside the smallest box by the simple expedient of increasing their energies.

[18] From answer 2A with $(q/2) = 1$ one finds 5×10^{90} bits, which is probably a lot closer to the correct answer. However, I felt obligated to use Bekenstein's weaker bound since the universe probably contains black holes.

sense that we know how, in principle, to store and retrieve an amount of information comparable to the upper bound, in a time comparable to the speed-of-light delay.

We are now going to claim that answer 2A in fact *is* tight in this stronger sense.

Namely, we may store information as the occupation numbers of the photon modes of a cubical box of volume V. Since these modes are time-invariant[19] the information will persist forever. According to our statistical analysis in the "justification of answer 2A" given previously, the information really is there (i.e. there really are that many distinguishable states of the radiation field in the box), up to a factor of $q^{1/4}$, which we ignore here. Now, to retrieve the information in the box, remove the box. All the photons will then escape in a time of order $V^{1/3}/c$ and their number, polarizations, and wavelength may be measured by an array of detectors on a very large sphere surrounding the box, while only losing a constant fraction of the information, as discussed in §3.4.

We will now claim that, in principle, all this measuring could be done in an amount of time comparable to the speed-of-light delay $V^{1/3}/c$. Certainly, in the scheme we have suggested, all the photons hit the sphere during an interval of time only $\approx V^{1/3}/c$ wide, so all the measuring takes place in a short time interval. But if the sphere were very large, the delay between opening the box and doing the measurements could be large. To make a claim that this delay may be reduced also to $\approx V^{1/3}/c$, we need a definition of what it means to "measure" something. Hopefully the scheme we will now suggest will be included in the reader's own mental definition of this term.

If we surround the box with an emulsion of a silver halide, we could in principle make a hologram which would record the photons that hit it. We could imagine, instead of silver halide, some substance with light-amplitude recording properties made out of nuclear matter or particles with higher mass than electrons, the point being that higher-density forms of matter have arbitarily small "granularity" (since the granularity [atomic size] of matter ultimately arises from $\Delta x \Delta p > \hbar/2$). By control of its microscopic structure, we could in principle imagine such a substance capable of also recording light polarization as well as intensity, thus going beyond silver halides. If the granularity were reduced sufficiently that it were comparable to the smallest wavelength of the photons in the box, then by Nyquist's theorem the "sampling" one would do in time $\approx V^{1/3}/c$ would be sufficient to reconstruct the mode occupation numbers. We claim that inefficiency in the "photographic emulsion" should reduce the information by only a constant factor.

This argument might still be objected to, in the event that the granularity of the emulsion had to be so fine, that sufficiently massive particles might not exist with which to construct it.[20] In that case, we point out that it has been argued [zure91b] that any interaction between any two entities may be regarded as a "measurement." By forcibly compressing a large volume of our "emulsion" into a small volume, we could force its "grain size" to be arbitrarily small, although at the cost of delocalizing its particles (due to the zero-point motion arising from their confinement). The compressed and delocalized emulsion would still "measure" at least a constant fraction of the information in the photons, in this weaker sense of the word "measure."

Finally, we point out that this photon-mode memory we have just constructed allows the possibility of alteration or measurement of a single bit, without affecting the rest of the memory. To do this, we must alter our storage scheme slightly. We now store bit i by either occupying the ith photon mode by a *single* photon, or not; we no longer allow the possibility of occupation of a mode by more than one photon. It turns out that this "Fermi" restiction only reduces the information storage capacity by a constant factor. (From (9) the factor is < 16.17. Actually, the true factor is $(8/7)^{1/4}$.).

[19] Assuming quantum mechanics and electromagnetism are *linear*. In fact, one of the significant developments in physics during the last 40 years has been the realization, due to the success of Feynman and Schwinger's "quantum electrodynamics," that these things can be nonlinear. Nonlinear effects can cause photon-photon scattering (due to the creation of "virtual" electron-positron pairs) and can prevent the possibility of true time-invariance. See §4.7. For now, the reader will have to be satisfied that the information in our memory will persist, while not forever, at least for some nonzero amount of time. Although nonlinear effects are generally thought of as small at "everyday" energies, they can still have profound effects [zure91b] [hawk74].

[20] And even if they did, the emulsion might still need to be so massive that it would collapse under its own gravity. We will continue to assume such gravitational effects are negligible, though. This is permissible since the mass of the emulsion will be comparable to the mass of the photons in the memory in the extreme relativistic limit we are talking about here, and we'd previously assumed the photons didn't have that much total mass.

Since all of the modes are orthogonal functions, in principle one could measure the ith mode by taking the inner product of all the modes with the complex conjugate of the ith mode. Physically, to take such an "inner product," one could place an "antenna" inside the memory-box constructed to have electric field sensitivity orthogonal to all the modes except for the desired one. Again, the antenna would have to be made out of material whose granularity was much smaller than the minimum wavelength of the photons, but, ignoring this granularity issue and also the issue of how to introduce the antenna into the box, in principle there seems to be nothing stopping us from measuring a single bit in a time comparable to the speed-of-light delay. Similarly by stimulating the antenna, we could introduce a photon in a desired mode, into the box, thus writing a bit into the memory.

4.4 Information storage in the presence of noise.

In the presence of thermal noise, the storage capacity of our photon-mode memory will only be reduced by at most an *additive* constant, similarly to §3.5. In the high energy-density limit, at finite temperature, the information storage capacity will thus be negligibly reduced by thermal noise.

4.5 Cubbyholes

Here is a quick way to gain intuition about why the $I/V \propto (M/V)^{3/4}$ behavior of answer 2A arose. Imagine our volume V were divided somehow into I "cubbyholes" – one per bit stored – and each cubbyhole either contained a photon (whose wavelength would have to be of order $(V/I)^{1/3}$ and hence whose energy would have to be of order $hc(V/I)^{-1/3}$), or not. Then the total mass-energy in the memory would be $M \approx hc^{-1} I^{4/3} V^{-1/3}$, which is the right result, albeit with a somewhat non-optimal value for the constant.

4.6 Memory add-ons

You might imagine (or hope) that, although any volume-V memory containing I bits must have mass-energy M where $M \propto I^{4/3} V^{-1/3}$, that this is just "manufacturing cost" and not so relevant to the actual energy consumption you'd need to store I bits. Not so.

Suppose we have cleverly pre-manufactured a memory with all sorts of expensive (in terms of mass-energy) stuff inside it. For example, perhaps we've manufactured cubbyholes as in §4.5. We want now to be able to store I bits in the memory for less energy cost than answer 2A would have allowed. If we then retrieve the I bits we get back to the memory's original state – prepared to accept deposits of I other bits.

The reason this is not going to work is that the pre-manufactured stuff is really information. (If nothing else, it is recording what it is.) Call this information I_0. When we add I bits, we are up to $I_0 + I$ bits stored in the memory – and there is no way to have less while still remembering the way the memory used to be, plus its contents – then upon regurgitating them we are back to I_0. Hence, the energy required to get from I_0 to $I_0 + I$ is proportional to $(I + I_0)^{4/3} - I_0^{4/3}$, which is greater than $I^{4/3}$, the energy cost bound without any pre-manufacturing.

4.7 Fundamental law of information destruction?

Question 3. "How quickly must stored information decay?"

In practice, unfortunately, our memory will not remain in a time-invariant quantum state forever. The states will mix due to thermal noise and quantum fluctuations. Although it has been suggested [benn88] that in classical physics, computers could be built using arbitrarily rigid mechanical parts, thus preventing errors due to thermal noise, in fact special relativity tells us that perfectly rigid bodies cannot exist. By the same token, two quantum systems cannot be perfectly isolated from one another since one cannot build insurmountable potential barriers. Further, even supposedly noninteracting particles, such as photons in orthogonal modes, in fact usually will interact, due to nonlinear effects. We therefore suspect that any information stored in a memory must eventually decay and that any computer must eventually develop a hardware bug.

Several intriguing problems now arise in physics, computer science, and mathematical biology: From the laws of physics: how quickly "must" information decay and hardware bugs arise? In computer science and biology: Can self-repairing computers exist? For instance in some reasonable model such as 3D cellular automata, in which parts are failing every so often (for example, at each time step, each cell of the automaton gets randomized with some small probability p), can one still assume that the algorithm will still be carried out successfully (with some probability bounded away from zero)? Can self-repairing organisms survive for an exponentially long expected lifetime? Can a reproducing organism exist such that its line of descendants will continue forever with probability bounded above 0 in some (nonmutated? or at least still reproducing) form? See [fede89] [pipp88] [burk70] [pipp92] [vNeu56] [dobr77] [dobr77b] [haje91] [pipp90] [gacs83] [gacs86].

Returning to physics, we speculate that there should be a quantitative version of the second law of thermodynamics, saying not only that entropy increases, but in fact giving a lower bound on the rate of such increase. In terms of our information-theoretic point of view, we are suggesting that there is a **fundamental law (or laws) of information decay** saying: "if you have I bits of information stored in a volume V and a mass M, then it will inevitably decay at at rate of at least $f(M, V, I)$ randomized bits/second." We will now attempt to shed a small amount of light on the possible form of such a law by analysing information decay in the particular memory model of §4.3 in which information was stored as occupation numbers of photon modes in a box.

Firstly, we will note that our idealization of "boxes" with perfectly hard, rigid, reflecting walls cannot actually exist. (A rigid object would have a speed of sound greater than the speed of light – impossible, according to special relativity.) Nevertheless, suitable "boxes" might exist, with some suitably weak sense of the word "box." For instance [whee57], one can imagine "geons" which are blobs of radiation confined by their own gravity. A nickel atom may be thought of as a "box" containing 28 electrons. An (essentially) perfectly reflecting box for low energy microwave photons may be made out of superconductor. All one really needs is some system exhibiting time invariant solutions. So, in the below we will assume that information decay due to imperfections in the box can be made negligible.

Secondly, it would seem unavoidable that thermal noise will leak in from the outside universe, despite all possible attempts to screen the box. (Screening out neutrinos and gravitons could prove difficult!) Still, this screening could in principle be made arbitrarily good if sufficiently thick shielding and sufficiently intense refrigeration were used. So, in the below we will assume that information decay due to thermal leakage can be made negligible.

The third cause of information decay in our photon-mode memory, which we do *not* know any way to prevent, and which we are therefore going to analyse, is mode-mixing due to the nonlinear phenomenon (brought to us courtesy of quantum electrodynamics) of photon-photon scattering.

According to formulas in [jauc76] (see also [itzy80] [karp50] [karp51] [achi37] [brei34]) the first order quantum electrodynamics analysis of photon-photon scattering shows that two photons colliding head-on on a line L_1, each having energy $\omega m_e c^2$, will scatter off one another, becoming transmogrified into two other photons of the same energy, but having a *different* direction-line L_2, with a probabilistic cross section given by (74-75). Scattering at any angle $\angle L_1 L_2$ is possible. Once a scattering event has occurred, there is no way to reconstruct the original memory from the present one since any of a large number of possible predecessor states could have resulted in the present state via various possible scatterings. Scatterings are completely random, nondeterministic events.

It is an easily verified fact (see the appendix) from special relativity that for any two photons, moving on lines such that they will collide at an angle θ, and having energies E_1 and E_2, there is a reference frame, called their "center of mass frame," in which both photons are colliding head-on (what would be $\theta = \pi$) and they each have exactly the same energy

$$E = \sqrt{E_1 E_2} \sin \frac{\theta}{2}. \tag{33}$$

Thus the cross-section formula (74-75) in fact suffices to compute the scattering probability of any two photons. (And the beam fluxes differ in other reference frames, see the appendix.)

The conservation of momentum condition is presumably trivial for photons which are *modes* in a box (since a mode has zero average momentum), as opposed to wave packets in free space. The wavevectors $\vec{k_i}$, (in the reference frame of the box), are integer 3-vectors for modes in the box. The conservation of energy condition

$$|\vec{k_1}| + |\vec{k_2}| = |\vec{k_3}| + |\vec{k_4}| \tag{34}$$

actually cannot be met (except for a constant number of trivial symmetries) for most choices of $(\vec{k_1}, \vec{k_2})$ while maintaining integrality of $(\vec{k_3}, \vec{k_4})$! Thus by sticking to a set of modes such that all mode-pair energies are distinct, one could theoretically entirely prevent first order photon-photon scattering with little loss in information storage density. However (recalling that rigid walls are impossible), if we assume that the box's walls are sufficiently "soft" that the exact integrality requirement in (34) can be relaxed, even very slightly, then potentially, almost any two modes can scatter in a large number of possible ways.

Let us now analyse the frequency of photon-photon scattering events in a "non-relativistic" ($k_B T \ll m_e c^2$) thermal photon gas at temperature T. This scattering frequency will be (cf. [reic80])

$$\frac{1}{2} n^2 \overline{\sigma} c, \tag{35}$$

events per unit volume per unit time, where n is the number density of photons in the gas and $\overline{\sigma}$ is the average scattering cross section, averaged over all pairs of photons (i.e. over directions and energies of each) in the gas. The overline denotes thermal averaging.

Refer back to (11) and (12). The number density of photons is

$$n = \frac{F_3}{\pi^2} \frac{k_B^3 T^3}{\hbar^3 c^3} = \frac{2\zeta(3)}{\pi^2} \frac{k_B^3 T^3}{\hbar^3 c^3}. \tag{36}$$

By (33) and the "NR" cross section (74) we find that

$$\overline{\sigma} = \frac{973}{10125} \frac{\alpha^2}{\pi^2} (r_e)^2 \times \overline{\omega^3}^2 \times A \tag{37}$$

where the angular average A is

$$A = \frac{1}{2} \int_0^\pi (\sin\frac{\theta}{2})^{6+2} \sin\theta \, d\theta \; = \; \frac{1}{5} \tag{38}$$

and the energy average $\overline{\omega^3}$ is given by

$$\overline{\omega^3} = \frac{F_6}{F_3} (\frac{K_B T}{m_e c^2})^3 = \frac{4}{63\zeta(3)} (\frac{K_B T}{m_e c^2})^3. \tag{39}$$

So, finally, combining all our results gives for the frequency f/V of scattering events per unit volume

$$\frac{f}{V} = \frac{4448\alpha^2}{28704375\pi^6} \frac{r_e^2 k_B^{12} T^{12}}{\hbar^6 c^{17} m_e^6} \qquad \text{(NR)}. \tag{40}$$

Now by using (13), $E = Mc^2$, and answer 2A with $q = 2$, we may express the "temperature" T of our photon-mode memory in terms of I/V:

$$\frac{k_B T}{m_e c^2} \; = \; (\frac{45\sqrt{2}\ln 2}{16\pi^2})^{1/3} r_c \, (\frac{I}{V})^{1/3}, \tag{41}$$

which leads to (43) below, under the assumption that each scattering event destroys exactly 1 bit of information. (The "true" destruction of information arising from a scattering is the entropy of the scattering probability distribution, which we shall not calculate. Certainly, this is at least 1 and no more than $O(\log I)$ bits.)

Next, let's consider the photon gas in the opposite, extreme-relativistic, limit. We see that the two-photon scattering cross section (75) is much smaller than the two photon pair-production cross section (77). Therefore, we need only concern ourselves with the latter process[21].

A derivation is given in the appendix of the pair production rate per unit volume in temperature T blackbody radiation, $k_B T \gg m_e c^2$.

$$\frac{f}{V} = \frac{r_e^2 m_e^2 k_B^4}{72 c \hbar^6} T^4 [O(1) + \ln \frac{k_B T}{m_e c^2}] \qquad \text{(ER)}. \tag{42}$$

Re-expressing this in terms of I/V instead of T by use of (41) gives (44) below.

We have at last:

Conjectural Answer 3A. *If the information content of a photon-mode gas is I bits and its volume is V, then the rate f/V of scattering events (which is a lower bound on the rate $\frac{-d}{dt}(\frac{I}{V})$ of information destruction) per unit volume is*

$$\frac{f}{V} = \frac{139(\ln 2)^4 \alpha^4}{3584 \pi^{14}} \frac{r_c^{14} m_e^6 c^7}{\hbar^6} (\frac{I}{V})^4, \quad \text{(NR)}. \tag{43}$$

$$\frac{f}{V} \approx \frac{45^{1/3} 5 (\ln 2)^{4/3} \alpha^2}{2^{2/3} 128 \pi^{8/3}} \frac{m_e^6 r_c^6}{\hbar^6} (\frac{I}{V})^{4/3} [O(1) + \ln(r_c [\frac{I}{V}]^{1/3})] \quad \text{(ER)} \tag{44}$$

Here "NR" means the expression is valid in the low-energy limit $(I/V)^{1/3} r_c \ll 1$ and "ER" means the expression is valid in the high-energy limit $(I/V)^{1/3} r_c \gg 1$. The conjectural part is that in the ER limit, any *type of "maxent" memory must have an information decay rate bounded below by the decay rate of a photon-mode memory.*

These, then, are conjectured asymptotic forms for a putative fundamental law of information destruction, and indeed information in the photon-mode memory of §4.3 actually will decay at approximately these rates, assuming the memory is isolated from external thermal noise and imperfections in the box do not matter.[22]

Observe that at low information densities I/V, the rate of destruction becomes asymptotically negligible, and presumably with self-repair mechanisms using error-correcting codes, it would be possible to maintain purity for exponentially long (in I) times.

At high information densities I/V, the rate of destruction becomes large, eventually so large that almost all of the stored information will be destroyed in an amount of time comparable to the speed-of-light delay in accessing the memory. Now, conceivably, some self-repair scheme would still be possible even if the information decayed in times less than the speed-of-light transit time for the entire memory (i.e. divide the memory into submemories), but it would *surely* be impossible for a self-repair scheme to work if all the information decayed in a time comparable to the light transit time over a length the size of one bit ($\ell_{\text{bit}} \approx (V/I)^{1/3}$):

$$\frac{d}{dt} \frac{I}{V} \approx -(\frac{I}{V})^{4/3} c. \tag{45}$$

This leads to a conservative limit on stable information *density* of any memory, which would be about $I/V < 5 \times 10^{27} (r_c)^{-3}$??? by using (43), except for the fact that (43) is only valid in the NR limit, which this certainly isn't; so by using (44) one obtains

$$\frac{I}{V} < (r_c)^{-3} K \tag{46}$$

where K is, according to a strict interpretation of (44), $K = \exp(5 \times 10^6 6)$ but actually due to the fact that the "$O(1)$" term in (44), and hence the constant factor K in (46) are not known, and the fact that unknown

[21] Actually, we really want the cross section for *all* processes, including multiple pair production, multiple photon production, etc. Pair production seems to be the lion's share at accessible energies, but we are not sure this is really valid in asymptopia.

[22] Here by "maxent" memories, we mean those that are storing nearly the maximum amount of information possible in that amount of mass-energy with its volume. Since by increasing the energy of each photon in a photon memory by a factor of κ, one can make a heavier memory which decays $\approx \kappa^2$ times more slowly, since ER photon scattering cross sections decrease like energy^{-2}, we see that if we were allowed to prodigiously "waste" mass-energy, we could actually store information more stably than the conjectured bound.

quantities appear in an exponential, this estimate is surely way off. For example if we insist that the critical length be the diameter of 1000 bits instead of 1, we would a much different answer. The important thing to remember here is that

Conjectural Answer 3B. *There is a dimensionless constant K such that, for any maxent memory with pretensions to permanence and storing I bits in a volume V, (46) holds.*

Recalling that $r_c = \alpha r_b$ and $\alpha \approx 1/137$, this can be colloquially rephrased as *"You can't store more than $K \cdot 137^3$ bits in an atom."*

This may be profound. Why is it that atoms have their usual sizes? Why is it that the planet Earth is the size that it is? Well, atoms and the Earth are more or less permanent entities, so they must be being stored stably. One may argue that any attempt to stably store the information represented by "the Earth" in something significantly smaller (e.g. ≈ 137 times smaller) could not succeed (unless one resorted to non-maxent schemes such as in footnote 22, which are so far below maximal entropy that they "must" not be able to occur in nature).

Multiplying by c would yield an *upper limit on achievable information flux* of

$$Kc(r_c)^{-3} = 5.20 \times 10^{45} \cdot K \frac{\text{bits}}{\text{meter}^2\ \text{sec}}. \tag{47}$$

It is also possible to obtain limits on the information contained in a memory by arguing that too much information would require too much mass, which would cause the memory to collapse into a black hole. Unlike the argument above, though, this method will *not* lead to bounds on information (or mass) density per se. The argument follows. If a mass M may be enclosed by a surface of area A and

$$\frac{M^2}{A} > \frac{c^4}{16\pi G^2}, \tag{48}$$

then the mass is (assuming it is nonrotating and uncharged) a black hole. Now by combining answer 2B with (48), we see that to prevent collapse we must have

$$\frac{I}{A} < \frac{c^3}{4G\hbar \ln 2}, \tag{49}$$

and by combining the bound of answer 2A with (48), we see that (assuming that the Euclidean sphere formula $A^3 = 36\pi V^2$ is at least approximately true)

$$\frac{I^2}{V} < \approx \frac{16\sqrt{6\pi}c^{9/2}\sqrt{q/2}}{9\sqrt{15}(\ln 2)^2 G^{3/2}\hbar^{3/2}} \tag{50}$$

Either one of these collapse-preventing conditions would still permit unboundedly large information density I/V in sufficiently small regions[23], and thus are weaker than the conjectural answer 3B.

However, because of their nonlocality, these conditions can actually be stronger than answer 3B in the limit when we are storing a very large amount of information.

Plausibility argument for 'conjectural answers 3A,3B'. All this can hardly be said to be "rigorous" even in our present very particular photon-mode memory model. The implied extrapolation to all memories, no matter how constructed, seems outrageous. Nevertheless, there is *some* reason for suspecting that it might be justified in the high density limit.

1. By previous arguments (answer 2A) our photon-mode memory model is the way to store the most possible information density (up to a small constant factor of $\leq (q/2)^{1/4}$) with the least possible mass density.

2. Indeed in the limit of high-density maxent information storage, it seems that one would be *forced* to use something like the photon mode scheme.

[23] But they would not permit *one bit* to be stored in an arbitrarily small volume!

3. The known cross-section formulas given in the appendix indicate that in both the non-relativistic and the extreme-relativistic limits, photon-photon cross sections are unboundedly smaller than the cross sections you would be forced to consider when designing a memory containing electrons and/or positrons. In other words, to avoid scattering, you are best off using photons only.

4. In the extreme-relativistic limit, any photon competitors will be traveling at nearly the same speed as photons, hence will scatter more often.

5. It also appears (preliminarily; see the appendix) that a memory based on neutrinos would also be corrupted much more quickly, asymptotically, than a memory based on photons.

All this combined makes it plausible that information must decay at at least the rates given in conjectural answer 3A. "QED."

As a numerical example, we compute the maximum amount of stable information that could possibly be stored in a sphere of volume 1 cubic meter (and surface area 4.836 meter^2). The maximum amount of mass that could be placed in such a region without causing it to be a black hole, is about 4×10^{24} Kg. Then from (49), we find that no more than about 7×10^{69} bits could be stored in this region. This figure is reduced to 2×10^{52} bits if we use (50) instead with $q = 2$. Meanwhile (46) yields a much stronger (?) upper bound of $K \cdot 1.7 \times 10^{37}$ bits. (And the smaller the memory, the better (46) will perform relative to the black hole bounds.)

Next, returning to a previous example of a photon-mode memory with weight 1000 Kg, volume of 1 cubic meter, and storing 6.0×10^{34} bits, we ask: how quickly must the information in this memory be decaying? We see that $(I/V)^{1/3} r_c = 0.15 \ll 1$ so that we are in the "low-energy" limit of our conjectured answer. Then (43) leads to a decay rate of 5.35×10^{30} bits/second. To get a decay rate of < 1 bit/second, we would need to store $< 1.3 \times 10^{27}$ bits in a cubic meter, to make the decay rate of < 1 bit/year, we would need to store $< 1.8 \times 10^{25}$ bits in a cubic meter, and to make it < 1 bit in 10^{10} years ($\approx$ the age of the universe), we would need to store $< 5.7 \times 10^{22}$ bits.

Let us now compare this with the decay rate, due to diffusion, of information stored in a lithium crystal as before. Using an atomic weight of 6.5 and a density of 500 Kg/meter^3, we find that there are 4.6×10^{28} Lithium atoms in a cubic meter. According to experimental data [lodd70], the self-diffusion coefficient D_T of solid Lithium in the range 35-178 Celsius is

$$D_T \approx 1.2 \times 10^{-5} \exp\left(\frac{-6345 \text{ Kelvin}}{T}\right) \frac{\text{meter}^2}{\text{second}}. \tag{51}$$

Pretending Lithium is simple cubic with lattice constant of 2.79×10^{-10} meters (which is compatible with the density) and diffusion only occurrs via interchanges of adjacent atoms, this corresponds to a rate of

$$7.1 \times 10^{42} \exp\left(\frac{-6345 \text{ Kelvin}}{T}\right) \tag{52}$$

atom-exchanges per second per meter^3. Then extrapolating to $T = 100$ Kelvin, and assuming an atom-exchange is the same as destroying 1 bit, the lithium memory would be losing 2.0×10^{15} bits per second. However, it seems that (52) allows the possibility of making the diffusion rate arbitrarily small by lowering T. In fact [reif65], the "Fermi temperature" of lithium atoms at this density is $T_F \approx 4.58$ Kelvin, and the Fermi temperature of the conduction band electrons (assuming one conduction electron per atom) is $T_F \approx 54600$ Kelvin. The Fermi temperature is a measure of zero-point motion, and presumably it is impossible to lower the effective diffusion temperature below T_F. It is very unclear to me just what the lowest achievable effective diffusion temperature is (does the 54600 affect the 4.58?) Anyway, assuming 10 Kelvin is possible would yield a bit-destruction rate of 2.0×10^{-233} bits per second, *far* below a photon memory of comparable capacity.

The reason that the photon memory was beaten by lithium, despite all the good reasons in the "plausibility argument" above showing that this was impossible, is due to the sole advantage rest mass has over photons – it stands still. Thus although photons *do* have a much smaller cross section than Lithium atoms, they still will scatter much more often because they are moving at the speed of light.

In the high-density ER limit, any mass would also have to have high kinetic energy and hence *would* be moving around near the speed of light, so that this advantage of rest-mass over photons, would dissapear. This is the reason that conjectural answer 3A was conjectured to hold for *all* memories, only in the high density limit.

The Lithium example we have just given, indicates that very large memories are possible which, for practical purposes, will never decay. Nevertheless, the following open question remains of some theoretical interest: can memories of volume V and capacity I bits exist, whose characteristic decay time τ is "exponentially" long, i.e. there exist positive constants A, B such that, for all sufficiently small bit densities I/V,

$$\tau > \exp(A[\frac{V}{I}]^B) \quad ? \tag{53}$$

5 Computation.

Question 4. "How quickly can we execute fundamental computational operations? How much energy does this require?"

Before answering this question, we will require some definitions. A "Turing machine" consists of a small finite set of states, and a tape and certain finite set of transition rules, which are tuples $(R, m, W)_{i,j}$ which indicate: if R is the symbol written in the current position of the tape and you are in state i, then move m places along the tape (m is 0, +1, or -1), write W in the new position on the tape, and change your state to j. (W can be a special "null" symbol, in which case the new position on the tape is *not* overwritten.) These rules include a possible transition to a special "calculation done" state and also there is a special "start of calculation" state. A Turing machine is "universal" if it is capable of computing any function computable by any Turing machine, assuming a suitable "simulation program" has been pre-written on some finite portion of the tape beforehand.

It is possible that for each transition rule that exists, one may write a unique reverse transition rule $(W, -m, R)_{j,i}$. It is also possible that no such unique reverse transition rules can be written, due to many-to-one behavior of the forward transitions. For instance if forward transitions $(A, 0, B)_{0,1}$, $(C, 0, B)_{0,1}$, $(D, 0, B)_{0,1}$ exist, then in an attempt to reverse the transition while only using the local information [state, current tape symbol] accessible to the Turing machine, there would be at least 3 possibilities, none of which could be ruled out. Also, the overwriting of the symbols on the tape in transitions such as $(A, 1, B)_{0,1}$ leads to at least L-to-1 behavior where we are using an L-symbol alphabet, since there is no way to know what was written before the "B" overwrote it. If one can write unique reverse transition rules corresponding to every forward transition, then the Turing machine will be said to be "reversible." Otherwise, it has irreversible, or information-destroying, transitions.

It is not hard to construct universal Turing machines [mins67].

Answers 4A, 4B, 4C below were previously observed by some combination of R. Landauer and C. Bennett [land61] [benn88] [benn82] [zure86] (arguably it may be traced by to J. von Neumann) and are merely repeated here. Throughout this section, if the reader prefers some other suitable model of computation (i.e. 3D cellular automata) to "Turing machines" he may generally just change the words "Turing machine" to his favorite machine and everything we say will still hold. In no irrevocable way have we used any particularly special properties of a Turing machine as opposed to other kinds of computational models.

Answer 4A. *If a computer immersed in a heat bath of temperature T erases a bit of information, then it must expend an amount $k_B T \ln 2$ of energy. More generally if a computer makes a n-to-1 information-destroying ($n > 1$) transition, it must expend an amount $k_B T \ln n$ of energy.*

Justification of answer 4A. The information-destroying transition reduced the entropy of the computer by $k_B \ln n$. Thus by the second law of thermodynamics the entropy of the heat bath must have increased by at least this amount. But that requires an energy expenditure of $k_B T \ln n$. QED.

Caveat: there are some subtleties involved in the application of answer 4A. If a $10 \to 1$ transition occurs, but the Turing machine is so constructed that 5 of these 10 states can never arise, then the necessary energy expenditure presumably could be made less than $k_B T \ln 10$. Answer 4A is assuming that all of the n states could have actually been responsible, equiprobably, as far as somebody trying to reverse the algorithm is concerned. Those denying this point of view, and who would not mind allowing omnipotent reversing Turing machines which, when they worry about how to reverse, unrealistically know much more than just their *local* state, will prefer

Answer 4A (more careful). *If a Turing machine immersed in a heat bath of temperature T could, at some point in time, have arisen from n possible start states (here by "state" we also include all the information written on the tape) then its "negentropy" will be said to be $\lg n$. In any transition increasing the negentropy by δ, an amount $k_B T \delta \ln 2$ of energy must be expended.*

Thus in order to avoid the expenditure of energy, it would seem all one needs to do is to avoid making irreversible transitions. To see that reversible computation can simulate any algorithm, merely store the old state (or, just the part of the state that is going to change) before executing any irreversible step of the algorithm, on a stack. The algorithm running in reverse pops the stack. Indeed, it is possible to construct universal Turing machines that are reversible [benn73].[24]

Answer 4B. *It is possible to compute any result computable by a reversible universal Turing machine in N steps, in an expected time equivalent to N^2 steps, by use of a physical system (designed to simulate a reversible universal Turing machine) and with the expenditure of zero energy.*

Justification of answer 4B. A *Thermodynamically biased Universal Reversible Turing Machine* (TURTM) will make a forward transition with probability p and will make the reverse transition with probability $1-p$.

One can build a physical system implementing a set of reversible transition rules (and in which the energy barriers are huge for illegal transitions) and by immersing it in a heat bath, it will make transitions with $p = 1/2$, randomly walking forward or backward in "time" in the N-state-chain region between the "start" and "done" states (the walk-boundary conditions are reflecting). All the energy is extracted from and returned to the heat bath so that no energy is expended, and after cN^2 random transitions have occurred, the probability becomes arbitrarily large (if c is an arbitrarily large constant) that the "calculation done" state has been reached at least once. QED.

We now observe that by the expenditure of energy $E > 0$ at each forward step the computer may *bias* the random walk so that

$$\frac{p}{1-p} = \exp(\frac{2E}{k_B T}). \tag{54}$$

In this case, it will become exponentially likely (for c large) that the calculation has been completed after t steps where

$$(2p-1)t > N + c\sqrt{(1-p)pt} \tag{55}$$

Incidentally, it is impossible to expend energy less than E on some steps and more than E on others in some clever way designed to still use average energy E per step but to increase the rate of forward progress. This is because the expected forward progress per step is

$$2p - 1 = \tanh(\frac{2E}{k_B T}). \tag{56}$$

which is a *concave-down* function of E, $E \geq 0$.

Thus we obtain

Answer 4C. *There is a tradeoff between energy-usage and speed. Expending energy of order $k_B T N$ will assure completion of an N-step TURTM calculation in order N basic time units (unless one is exponentially unlucky). On the other hand, one may expend a factor of N^β less energy, $0 \leq \beta \leq 1$, in which case the calculation will almost certainly take about N^β times longer.*

For most interesting computations it would seem N is large and the use of $\beta > 0$, incurring an asymptotically unbounded slowdown factor, would be *unacceptable.* Such computers will be called "irreversible." We observe that refrigerating an irreversible computer to a temperature T lower than the temperature T_{universe} of the surrounding universe, will not work to reduce the energy consumption of order $k_B T$ per step, since

[24] According to Bennett (private communication) this was apparently first realized by Lecerf in the 1960's.

even with a perfect Carnot refrigerator, the energy needed to export this amount of heat to the outside universe, will be $\geq k_B T_{\text{universe}}$.

One may ask just what a "basic time unit" is. It would seem that the expenditure of $\approx k_B T$ energy, (required each basic time unit, at least if $\beta = 0$) would have to take time

$$\geq \frac{\hbar}{k_B T} \tag{57}$$

due to the energy-time uncertainty principle or Bekenstein's answer 1C. Perhaps by expending more energy than $k_B T$ per step, one could go faster than this, though.

Also, if the computer has physical size L, then some internal communications will take time at least L/c.

The above statements lead to obvious power-speed-size-temperature tradeoffs, but as far as any of these tradeoffs are concerned, it seems that by increasing the power consumption P of the computer, one can increase the computational speed linearly with P.

A less obvious bound, which shows that this will not work, is below.

5.1 Limits on the speed of computation.

Answer 4D. *Let an irreversible computer be enclosed in a surface of area A and have no interaction with the outside universe except for heat lost and power consumed and the inputting of the problem and the outputting of the answer. Let its power consumption be P and let its sustained speed (in irreversible bit operations per second) be s. Then*

$$s \leq \frac{(\sigma_{SB})^{1/4}}{k_B \ln 2} A^{1/4} P^{3/4}. \tag{58}$$

This is yet another "$P^{3/4}$ law." It neatly encapsulates the idea that the need to get rid of the heat limits the performance of computers. Thus a chip with 1 square centimeter of surface area and consuming 100 Watts cannot perform a sustained rate of more than 1.61×10^{22} irreversible bit operations per second.

Justification for answer 4D.
Assumptions:

1. We assume we are in the high-power limit in which the best way to dump heat is via thermal radiation (or alternatively, assume that the only allowed cooling method is thermal radiation), and in which the Stefan-Boltzmann law is valid.

2. We ignore the possible use of non-photon thermal radiation for cooling; if this is permitted, the right hand side of (58) would have to be multiplied by $(q/2)^{1/4}$.

3. Gravitational effects are negligible.

The computer is generating heat at a rate P, which is $\geq k_B T s \ln 2$, where T is the temperature of the computer. Now this heat must also be being radiated out of the computer, which occurs at a rate $A\sigma_{SB}(T^4 - T^4_{\text{universe}})$.

The rate of power input will, in steady state operation, be equal to the rate of heat output. Upon solving for T in terms of P (and taking $T_{\text{universe}} \geq 0$) the rate inequality yields the desired bound on s. (Note that this bound now does not involve T and only depends on P and A.) QED.

Of course, similarly to the discussion in §3.1 of answer 1A vs. answer 1B, the "true" version of answer 4D should be based on the true capacity of the computer's cooling system, which could be greater than a Stefan-Boltzmann blackbody at low heat fluxes. Therefore:

Answer 4E. *Suppose that the cooling system can export heat $H(P_c, T)$ to the outside universe, where T is the temperature of the computer, P_c is the power consumed by the cooling system, and H is some entirely general function (monotonic increasing in its arguments) which is not necessarily the blackbody function. Then upon solving*

$$P - P_c = H(P_c, T) - P_c \geq s k_B T \ln 2 \tag{59}$$

for the (uniquely determined) unknowns T and P_c subject to $0 \leq P_c < P$ and $T > 0$ and maximizing s, we will have an upper bound on computational speed depending on the power supply and the cooling capacity.

5.2 Tightness.

We claim that answers 4A, 4B, 4C, 4D, and 4E are in fact tight bounds in the sense that realistic physical models exist, which will approach the bounds.

For 4A-4C this should be obvious by constructing reversible Turing machines.

To construct a computing device approaching the bounds of 4D and 4E is slightly trickier. The crux is to define "computing device." If we regard the destruction of information as computing (and indeed, this is arguably all we need to do, since answers 4D and 4E were phrased in terms of "irreversible bit operations") then we can simply take information in the form of boxes of photons (cf. §4.3) and destroy it by opening the boxes, thus converting the information to heat. The heat will radiate away at precisely the same rate as the information is destroyed, leading to precise tightness of the bound in 4D. If the reader is unsatisfied that this constitutes "computing," we will proclaim that the DNA copying machine inside bacteria in fact is a thermodynamic reversible (although highly biased) computing device, and indeed many biochemical reactions may be regarded as computing. Indeed Bennett [benn85] has proposed a hypothetical "enzymatic" TURTM which would work exceedingly well, provided one knew how to manufacture it. The "power supply" for such a computer is the reactants, especially adenosine triphosphate ("ATP"). A setup in which reactants are continually fed in, possibly along with varying concentrations of regulating chemicals (which could be telling the machine: start copying DNA; start destroying DNA, or whatever), and products are drained out (along with heat) will keep working if and only if bound 4E is obeyed.

Besides enzymes, we mention that Likharev [likh77] [likh82] [likh85] has proposed a way to accomplish reversible computation using electrical circuits involving Josephson junctions. It seems likely that Likharev's devices could actually be built.

6 Remarks about previous work.

Various calculations concerning information flux and "photon channels" have been made previously [pier81] [levi66] [bowe67] [mand59] [karp70] [vour90] [wall83] [mcel81] [lebe63] [gord61] [beke81b] [pend83] [gabo61] [ster60], but all of them, at least according to our present perspective, were incorrect; none of them found the $P^{3/4}$ law of answer 1B or understood its simple thermodynamic genesis.

Some previous authors have made some statements which may appear to contradict ours, so it is worth a short discussion. It was pointed out [land82] that the $\Delta E \Delta t > \hbar/2$ uncertainty principle need not slow down a computer whose energy is not carefully kept track of. It still seems valid to claim that any computational step which involves a dissipation of ΔE worth of energy, must use time $\geq \hbar/(2\Delta E)$. Available Josephson junction switching technology actually *has* switching times Δt approaching $\hbar/(2\Delta E)$, where ΔE is a relevant superconducting bandgap energy [rugg90] [likh90] and Josephson logic circuits exist which manipulate single flux quanta.

H. Heffner [heff62] found a lower bound

$$T \geq \frac{hf}{k_B \ln[(2G-1)/(G-1)]} \tag{60}$$

on the "noise temperature" T of any linear amplifier with gain $G \geq 1$. His derivation assumes that the amplifier inputs and outputs photons with frequency f (outputting a factor of G more photons than it inputs). This bound arises from the $\Delta n \Delta \phi \geq \frac{1}{2}$ (n is the number of photons, ϕ is their average phase-angle) uncertainty principle.

A general discussion of "uncertainty principles" may be found in [robe29] and the article by Wigner in [whee83]. For a different view of the same subject, see [slep61] [land62].

Paul Benioff [beni82a] [beni82b], and later R.P. Feynman [feyn85], observed (essentially) that one could in principle construct a quantum-mechanical reversible Turing machine, whose Turing transition matrix was implemented by a unitary state-transition matrix with 0-1 coefficients. Such a machine could perform N Turing steps, in time proportional to N, while using zero energy. Unfortunately these machines, which R. Landauer [land91] has called "ballistic," only work at zero temperature and zero entropy, in a non-thermodynamic frictionless noiseless universe. (One can also make ballistic Turing machines in Newtonian physics [marg84] [fred82]. See [land86] for a critical discussion of related devices. Certain present-day optical computing devices (e.g. optical Fourier transform) in fact *are* ballistic reversible computers, but imperfections make it impractical to put very many such devices in series.)

David Deutsch [deut85], also working in a zero-temperature hypothetical world, went beyond Feynman and Turing by explicitly allowing certain unitary state-transition matrices whose entries were *not* necessarily 0's and 1's only. At any time Deutsch's machine could be in a linear superposition (with complex coefficients whose squared norms sum to 1) of a potentially vast number of states of the corresponding Turing machine. The coefficients in the linear superposition could be (to some extent) controlled "in software" via extensions of the basic state-transitions of the Turing machine to include certain unitary linear transformations. Deutsch argued that his machine was more powerful than a Turing machine and truly obeyed a (modified) Church's thesis: it could simulate any finite physical system. He further argued that somehow the usual interpretations of quantum mechanics were placed under an "intolerable strain" by the presence of his machine and that Everett's "many worlds" interpretation was the only one which was not. Further, he actually proposed an experiment to prove that the many-worlds interpretation was correct and the Copenhagen interpretation was not.

I am afraid I must differ drastically with Deutsch. First, his machine is *not* more powerful than a Turing machine, since a Turing machine clearly can simulate, to arbitrary accuracy, a (probabilistic ensemble of) Deutsch machines. Thus any function computable by a Deutsch machine is computable by a Turing machine. (Further, a Turing machine with a hardware random number generator can pick out one element of this ensemble, too.) However it still *may* be true that simulating N Deutsch-machine steps could take an exponential (κ^N) number of Turing machine steps, with some definitions of the word "simulate." Peter Shor [shor94] has shown that that is true, assuming integer factoring is superpolynomially hard. Deutsch has also claimed that the use of certain quantum mechanical devices [benn92a] [benn92b] [benn92c] [benn93] gives capabilities to Turing machines with random number generators, which they would not otherwise have. Specifically, two Turing machines A and B equipped with this device may communicate in such a way that eavesdropping is statistically impossible, in the sense that any eavesdropper must destroy the message before it reaches B. Although Deutsch has ascribed great significance to this, I do not.

Secondly, Deutsch's ideas about many-worlds are ridiculously baroque and seem to place a much more "intolerable strain" on this author than do simpler ways to understand the Copenhagen interpretation [zure91b].

Most importantly, the nonlinear "decoherence" effects explained in [zure91b] would in fact cause all Deutsch-machines that you could actually build, to fail to perform as advertised. This invalidates Deutsch's proposed many-worlds-experiment and makes Shor's algorithm [shor94] of much less interest.

We will not provide a detailed argument for this; we will simply mention Zurek's [zure91b] estimate of the decoherence time of any quantum complex superposition having spatial extent Δx and mass m

$$\tau_{\text{decoh}} = \tau_{\text{relax}} \frac{\hbar^2}{2mk_B T(\Delta x)^2} \tag{61}$$

where T is the temperature of the surrounding vacuum and τ_{relax} is the classical "relaxation time" of the system (assumed imperfectly isolated from its environment, with which it communicated through the "warm vacuum."). Since the smallest useful computer components I can conceive of anybody using are molecules of small protein mass, and we need to get 1000 components correlated to do anything interesting, let us take $m = 10^6$AMU. Then this is

$$\tau_{\text{decoh}} = \tau_{\text{relax}} \frac{2.4 \times 10^{-7}}{T(\Delta x)^2} \tag{62}$$

where T is in μKelvin and Δx in μmeters. Even if our systems are so well isolated from the environment that they feature $Q \approx 10^8$, comparable to the highest-Q systems ever made, and even if the entire computer fits in a 1 micron sphere and is refrigerated to 1 μKelvin... we conclude that it will only be able to execute 24 cycles before it decoheres. This is nowhere near enough to factor any nontrivial integer, and certainly all that quantum computers could ever gain for us compared to conventional computers (even with ridiculously more optimism) would be a constant factor.

Bekenstein [beke81b] started with his bound (cf. answer 2B) $I < 2\pi ER/(\hbar c \ln 2)$ on the number of bits of information that could be contained in a sphere of radius R containing gravitating energy E, and then assumed that it would take at least time $2R/c$ to transmit this information anywhere. He concluded that

Answer 1C: *If Bekenstein's bound (answer 2B) is valid, then the rate r of information transmission, via the transmission of a message encoded inside energy (or equivalent mass) E, is bounded by*

$$r \leq \frac{\pi E}{\hbar \ln 2} \tag{63}$$

bits per unit time. In other words, the mass-energy transmitted per bit transmitted, assuming this bit is transmitted and received in a time interval t, must be at least

$$\text{Energy per duration} - t - \text{bit} \quad \geq \quad \frac{\hbar \ln 2}{\pi t}. \tag{64}$$

This is an admirable bound and in some sense turns the energy-time uncertainty principle into a useful statement about information transmission. Unfortunately we cannot support the ways in which Bekenstein then went on to use this bound to "conclude" that no digital computer could process more than "a firm upper bound" of 10^{17} bits per second. Among the flaws in those arguments, we mention the assumptions (all five of which were used, although only the third and fifth were stated) that all the energy, including the ground-state rest mass-energy, of transmitted messages must be dissipated, that energy has to be dissipated in order to measure it (not so – you could weigh it!), that no recycling[25] of transmitted mass-energy can occur, and that the computer "must" operate below a temperature of about 500 Kelvin and "must" not have liquid cooling. In fact, the author suspects that digital processing elements located on neutron stars *could* exceed 10^{17} bits/second. Deutch [deut82] pointed out some of these flaws but for some unknown reason "concluded" that "there is no fundamental limit on the speed of information processors."

7 Sorting N numbers, each b bits long

As an example of what our bounds can say about a concrete problem, we consider the problem of sorting N numbers, each b bits long, where $b \geq \lg N$. The input is thus $I = Nb$ bits.

Let our computer be a sphere of radius R, surface area $A = 4\pi R^2$, and volume $V = 4\pi R^3/3$, supplied with power P. Let T be the temperature of the surrounding heat bath.

The input is I bits of information, so with power P, our information flux bound shows that it will take time at least

$$t_{\text{input}} \geq I P^{-3/4} A^{-1/4} c_1, \quad c_1 = 4.6 \times 10^{-22} \text{ in SI units} \tag{65}$$

just to read it. (This is what VLSI theorists [ullm84] would call a "Vt^6 bound.")

If we assume that the computer must be capable of storing all the data, this will require energy at least

$$E_{\text{storage}} \geq I^{4/3} V^{-1/3} c_2, \quad c_2 = 3.8 \times 10^{-27} \text{ in SI units} \tag{66}$$

(Since $Pt \geq E_{\text{storage}}$, VLSI theorists would call this a "Vt^3 bound.")

Both these bounds would still permit the computer to sort arbitrarily quickly with a sufficiently large computer (even with finite power) or to use an arbitrarily small energy consumption, but those are ruled out by the next two bounds.

[25] Actually, this claim may be justified via the same sort of argument as we used in §3.6 but using answers 1C and 2B in place of 1B and 2A.

Since the entropy reduction in the sorting process is $\Delta S = \lg N! \approx N \lg N$ bits, we conclude that energy

$$E_{\Delta\text{ent}} \geq k_B T \ln N!. \tag{67}$$

must be expended.

Lightspeed delay shows that the time expenditure must be at least

$$t_{\text{light}} \geq 2R/c, \tag{68}$$

otherwise some part of the computer would remain unused, which would be pointless.

By considering the optimal choice of the computer size R (not too small, so that the storage energy or input time will not be too large, but also not so large that the lightspeed delay will dominate) we may obtain a lower bound on sorting time if the power supply is fixed:

$$t \geq I^{2/3} P^{-1/2} c_3, \quad c_3 = 2.4 \times 10^{-18} \text{ in SI units} \tag{69}$$

Although this is a better bound when $P \to \infty$ than (67), it would still allow solving the sorting problem in an arbitrarily small amount of time by using a tiny computer with a huge power supply. The only way to stop that is to call upon our *conjectural* bounds which provide an absolute limit $\rho_{\max} \approx (r_c)^{-3} = 1.736 \times 10^{37}\text{bits/meter}^3$ on the density of information one can stably store in 3-space. One then is forced to have $V \geq \rho_{\max} I$ so that $R \geq I^{1/3} 10^{-13}\text{meter}$, so that

$$t_{\text{light}} \geq I^{1/3} c_4, \quad c_4 = 10^{-21}\text{sec}. \tag{70}$$

Actually, one could also stop it by using the bound (50) which prevents too much information I from being stored in too small a sphere because it would force gravitational collapse. This is rather weak, and one hates to have to rely upon gravity, but since it is better than relying upon a shaky conjecture, here is the result:

$$t_{\text{light}} \geq I^{2/3} c_5, \quad c_5 = 10^{-35}\text{sec}. \tag{71}$$

This is actually stronger than (70) when $I \to \infty$, but in practice would not be stronger until a truly enormous amount of information was being sorted.

8 Open questions.

1. Corruption of stored information; self-repairing organisms. See the 2nd paragraph of §4.7.

2. Algorithmic efficiency. All of our results about information storage and transmission have ignored the *computational burden* of encoding and decoding information, with the aid of suitable error correcting codes to defeat thermal noise, required. These were purely information-theoretic existence arguments. What happens if one tries to fix this flaw? See [spie95].

3. A new complexity theory? In §3 we considered bounds on information flux in which one party was interested in transmitting information, through a window of fixed area, to another party. What if we have N independent entities, each enclosed in disjoint spheres of given sizes, and each of these parties may wish to communicate with any of the other parties, and collaborate in doing computations?

We have proven (answer 2A) that any volume V with gravitating mass M inside can only contain information I with $I/V = O((M/V)^{3/4})$. If N entities live in the volume V and communicate at speeds $< c$, then the total amount of information *in flight* (over paths whose total length is L_{tot}) at any time is

$$I_{\text{in flight}} = L_{\text{tot}} \cdot (\text{bit rate})/c \leq I \leq V \cdot O([\frac{M}{V}]^{3/4}). \tag{72}$$

Thus if M/V is bounded and bitrate is constant in all communications, then a bound of Ozatkas and Goodman's [ozat90] form[26] now holds:

$$L_{\text{tot}} = O(V) \tag{73}$$

in other words, a computer has to use up at least the same amount of volume as would be occupied by *wires* of constant cross sectional area and length L_{tot}. This is getting perilously close to the well-studied "VLSI theory" model of computation [leng90] [yao81] [ullm84] [leig83] [thom80] [leis83], in which many lower bounds have been proved. The VLSI model in these references isn't quite "right:"

- It ignores cooling problems
- It ignores length-dependent propagation delay (but see [chaz85])
- It ignores the possibility of "cheating" by using more power and more bitrate when you want to transmit more information than usual, or communicating by radio...

but, it's close enough to make me suspect that a usable complexity theory could come of this.

Can we make a model of computation resembling VLSI theory, but based now not on just some model, but actually on the Laws of Physics, in which we can prove tradeoffs between power usage, cooling needs, mass, volume, N, and runtime for various computational tasks? We have begun this task in §7.

4. Ideal gas conjecture. We assumed in answer 2A that N noninteracting particles in a fixed volume had at least as much entropy as if some interaction Hamiltonian were turned on. This may be formulated as a purely mathematical conjecture about the asymptotic number of eigenmodes of the appropriate wave equation, see §4.1. Prove it.

9 Acknowledgments

I had helpful conversations with Eric B. Baum, William Bialek, Richard A. Linke, and Tineke Thio.

10 Appendix: Collected scattering cross sections.

All the cross sections we give are averaged over directions and polarizations.

The dimensionless quantity ω denotes the ratio of the energy of the incoming particle to the rest mass $m_e c^2$ of an electron. The dimensionless quantity β denotes the ratio of the velocity of the incoming particle to the speed of light. "ER" denotes the extreme relativistic ($\omega \gg 1$, $\beta \to 1-$) and "NR" the nonrelativistic ($\omega \to 1+$, $\beta \to 0$) limit.

The second of the two particles mentioned is always at rest, except in the photon-photon cases, when we are in the center-of-mass frame of the two photons, the polarizations are assumed to be random, and ω refers to the energy of either of the incoming photons.

Photon-Photon scattering [itzy80][achi37] [itzy80] [jauc76][karp51] [eule36] [ward50]:

$$\frac{973}{10125}\frac{\alpha^2}{\pi^2}(r_e)^2\omega^6 \quad \text{(NR)} \tag{74}$$

$$\text{const.} \times \alpha^2 (r_e)^2 \omega^{-2} \quad \text{(ER)} \tag{75}$$

Photon-Photon pair production [jauc76][brei34] [karp51]:

$$0 \quad \text{(NR)} \tag{76}$$

$$\pi(r_e)^2 \operatorname{arccosh}(\omega)\omega^{-2} \quad \text{(ER)} \tag{77}$$

---

[26] Ozatkas and Goodman prove a very nice statement, but their subsequent claim that "therefore... all the results of VLSI area theory apply also to optics" is overconfident. O&G's "L_{tot}" refers to the total path length actually followed by the light. Wires cannot necessarily follow the paths followed by light (i.e. if light self-intersects).

For instance in 2D VLSI, laying out the complete graph K_N takes area $\theta(N^4)$, whereas O&G's 2D bound, if all the N nodes are located in a square sub-region of area N, gives order $N^{2.5}$. Hence, optics, even at a single wavelength, could still be better than wires, as far as O&G's bound is concerned, despite their claims.

(Maybe this counterexample can be defeated, and maybe O&G's statement is true, but I'm just saying: they never proved it.)

Photon-Electron pair production [beth34] [heit54][haug75] [whee39] [jauc76]:

$$0 \quad \text{(NR)} \tag{78}$$

$$\alpha(r_e)^2 \left[\frac{28}{9}\ln(2\omega) - \frac{218}{27}\right] \quad \text{(ER)} \tag{79}$$

Photon-Electron scattering [bjor64][klei29] [itzy80][thom33]:

$$\frac{8\pi}{3}(r_e)^2 \quad \text{(NR) Thompson} \tag{80}$$

$$\pi(r_e)^2 \left[\ln(2\omega) + \frac{1}{2} + O(\frac{\ln\omega}{\omega})\right]\omega^{-1} \quad \text{(ER) Compton, Klein Nishina} \tag{81}$$

Electron-Positron scattering [bjor64][itzy80][bhab35]:

$$\text{const.} \times (r_e)^2\beta^{-4} \quad \text{(NR)} \tag{82}$$

$$\text{const.} \times (r_e)^2\omega^{-2} \quad \text{(ER)} \tag{83}$$

Electron-Electron (or positron-positron) scattering [itzy80][bjor64][moll32] [mott29]:

$$\text{const.} \times (r_e)^2\beta^{-4} \quad \text{(NR) Mott} \tag{84}$$

$$\text{const.} \times (r_e)^2\omega^{-2} \quad \text{(ER) Moller} \tag{85}$$

Electron-Electron (or positron-positron) collision with pair-production [heit54]:

$$0 \quad \text{(NR)} \tag{86}$$

$$\frac{28}{27\pi}\alpha^2(r_e)^2\ln(\omega)^3 \quad \text{(ER)} \tag{87}$$

According to Heitler [heit54], the pair production cross section in the ER limit "increases at least as $(\ln\omega)^2$ even for a screened field."

Electron-Electron (or positron-positron) collision with bremstrahlung [jauc76][whee39]:

$$\text{small} \quad \text{(NR)} \tag{88}$$

$$4\alpha(r_e)^2\ln(2\omega) \quad \text{(ER)} \tag{89}$$

Electron-Positron annihilation [dira30] [bjor64][itzy80][jauc76]:

$$\pi(r_e)^2\beta^{-1} \quad \text{(NR)} \tag{90}$$

$$\pi(r_e)^2\left[\ln(2\omega) - 1 + O(\frac{\ln\omega}{\omega})\right]\omega^{-1} \quad \text{(ER)} \tag{91}$$

These expressions are all well-confirmed by experiment except for (74-75) and (77) which are virtually impossible to confirm *directly* by experiment, due to the high fluxes of gamma rays that would be required. All of these expressions arise from quantum electrodynamics. Quantum electrodynamics has so far proven to be "the most accurate physical theory ever devised," and "essentially correct up to the highest energies known [10^6 MeV]... there is no visible limit (as yet) to the validity..." [heit54] but the fact that almost every calculation in it involves a diverging asymptotic series still leaves grounds for concern.

Neutrino-electron scattering [bail82][quig83]involves a cross section (according to the Salam-Weinberg unification of the weak force into quantum electrodynamics) which behaves proportionally to the energy of the neutrino. (The same remark is true for electronic or muonic neutrinos or antineutrinos scattering off quarks, electrons, or positrons.) These predictions are compatible with all the observations that have been made in the range 0-200 GeV [bari81] [bail82] [alle86]. Thus [alle86] found

$$\frac{\sigma(\nu_e e \text{ scattering})}{E_{\text{neutrino}}} \approx 9.8 \times 10^{-45}\frac{\text{cm}^2}{\text{MeV}}. \tag{92}$$

Neutrino-neutrino scattering should presumably (?) also have a cross section proportional to energy, but it has never been observed, nor am I aware of any theoretical analysis.

Heitler [heit54] has argued that cross section formulas such as (92) (88) and (86) which tend to infinity in the high energy limit, are "impossible" (despite their agreement with experiment, so far) and must be an artifact of their derivation via "first order" quantum electrodynamics. Some other authors have expressed no such qualms. (The bottom line is that nobody knows what happens at the extremely high energies currently innaccessible to experiment.) In any case, we presume that the true cross sections are still rather large.

To conclude this section we give a derivation of the pair production rate in a blackbody radiation bath.

Breit and Wheeler [brei34] have (their eq. 21.3 p.1090; this formula was derived independently by [karp51]) that the pair creation cross section for head on collision of *randomly polarized* light quanta in the COM frame, each quanta having energy $E = m_e c^2 C$ where $C = \cosh\theta$, $S = \sinh\theta$, is

$$\sigma(E) = 2\pi r_e^2[2\theta C^{-2} - SC^{-3} + 2\theta C^{-4} - SC^{-5} - \theta C^{-6}] \tag{93}$$

(Note $S = \sqrt{C^2-1}$ and $\theta = \text{arccosh}C$. The terms are given in decreasing order of importance when C is large. Of course we only consider the case $C > 1$ since otherwise $\sigma = 0$.) Here σ is defined so that if we have two head on beams of photons, one with N_1 and one with N_2 photons per unit cross sectional area per second, then the number of pairs produced per second in a common volume V of the beam will be $V\sigma N_1 N_2/(2c)$.

If we have photons of energies E_1 and E_2 in beams of fluxes N_1 and N_2 photons per second per unit area colliding at angle ϕ (head on would be $\phi=\pi$), then in their center of mass frame the collisions are head on, the photon energies are both $E=\sqrt{E_1E_2}\sin\frac{\phi}{2}$, and the fluxes are N_1E_1/E and N_2E_2/E. The pair production rate (events per unit volume per unit time) is the same in all reference frames, since it is a Lorentz invariant density.

Thus the cross section for two photons traveling at angle ϕ and having energies E_1, E_2 is

$$\sigma(\sqrt{E_1E_2}\sin\frac{\phi}{2})\ \sin^2\frac{\phi}{2}. \tag{94}$$

Let $F_T(E)\mathrm{d}E$ be the number flux per solid angle of thermal photons at temperature T in the energy range $[E, E+\mathrm{d}E]$ – namely

$$F_T(E)=\frac{2}{c^2h^3}\frac{E^2}{\exp(\frac{E}{k_BT})-1}. \tag{95}$$

Then the total pair production rate (pairs per unit volume per unit time) in thermal radiation at temperature T is then – according to Breit and Wheeler's meaning of "cross section,"

$$\frac{2\pi}{c}\int_0^\infty\int_0^\infty\int_0^\pi F_T(E_1)F_T(E_2)\sigma(\sqrt{E_1E_2}\sin\frac{\phi}{2})\frac{\sin\phi}{2}\sin^2\frac{\phi}{2}\,\mathrm{d}\phi\mathrm{d}E_1\mathrm{d}E_2. \tag{96}$$

Letting $Z=k_BT/(m_ec^2)$, and assuming $Z\gg 1$ so that the "$O(1)$" term in $\sigma(E)=4\pi r_e^2C^{-2}[O(1)+\ln C]$ is relatively negligible, this is

$$\frac{16\pi^2r_e^2m_e^6c^7}{h^6}Z^4\int_0^\infty\int_0^\infty\int_0^\pi(O(1+\ln x+\ln y)+\ln Z)\frac{x}{e^x-1}\frac{y}{e^y-1}\sin\phi\,\mathrm{d}\phi\mathrm{d}x\mathrm{d}y. \tag{97}$$

Finally by using $\int_0^\infty\frac{x^{a-1}\,\mathrm{d}x}{e^x-1}=\zeta(a)\Gamma(a)$ so that $\int_0^\infty\frac{x\,\mathrm{d}x}{e^x-1}=\pi^2/6$ and $\int_0^\infty\frac{x^{a-1}\ln x\,\mathrm{d}x}{e^x-1}=\frac{\mathrm{d}}{\mathrm{d}a}(\zeta(a)\Gamma(a)$, and $\int_0^\pi\sin\phi\mathrm{d}\phi=2$ we see that this is

$$\text{pair production rate}=\frac{r_e^2m_e^2k_B^4}{72c\hbar^6}T^4[O(1)+\ln\frac{k_BT}{m_ec^2}],\qquad\frac{k_BT}{m_ec^2}\gg 1. \tag{98}$$

(For small T, the behavior is more like an Arhenius cutoff times power corrections.)

I have tried to compile a bibliography of sufficient size to serve as an aid to future work in this field. Some of the items in it are not cited.

References

[achi37] A. Achieser. *Physik Z. Sowjetunion*, 11:263–, 1937.

[bail82] David Bailin. *Weak interactions.* Adam Hilger Ltd., second edition, 1982.

[balt72] H.P. Baltes and F.K. Kneubühl. Thermal radiation in finite cavities. *Helvetica Physica Acta*, 45:481–529, 1972.

[bara87] Richard Barakat and John Reif. Lower bounds on the computational efficiency of optical computing systems. *Applied Optics*, 26(6):1015–1018, 1987.

[bard73] J.H. Bardeen, B. Carter, and S. W. Hawking. The four laws of black hole mechanics. *Commun. Math. Phys.*, 31:161–170, 1973.

[bari81] Barry C. Barish. Neutrino physics. In *Gauge theories and experiments at high energies*, pages 303–336. Scottish universities summer school in physics, 1981.

[baum70] L.E. Baum, T. Petrie, G. Soules, and N.Weiss. A maximization technique occuring in the statistical analysis of probabilistic functions of markov chains. *Ann. Math. Statist.*, 41:164–171, 1970.

[beke72a] Jacob D. Bekenstein. Nonexistence of baryon number of static black holes, i. *Physical Review D*, 5:1239–1246, 1972.

[beke72b] Jacob D. Bekenstein. Nonexistence of baryon number of static black holes, ii. *Physical Review D*, 5:2403–2412, 1972.

[beke73] Jacob D. Bekenstein. Black holes and entropy. *Physical Review D*, 7(8):2333–2346, April 1973.

[beke74] Jacob D. Bekenstein. Generalized second law of thermodynamics in black hole physics. *Physical Review D*, 9(12):3292–3300, June 1974.

[beke75] Jacob D. Bekenstein. Statistical black hole thermodynamics. *Physical Review D*, 12:3077–, 1975.

[beke81b] Jacob D. Bekenstein. Energy cost of information transfer. *Physical Review Letters*, 46:623–626, 1981.

[beke81] Jacob D. Bekenstein. Universal upper bound on entropy-to-energy ratio for bounded systems. *Physical Review D*, 23(2):287–298, January 1981.

[beke82] Jacob D. Bekenstein. Black holes and everyday physics. *General relativity and gravitation*, 14(4):355–359, 1982.

[beke82b] Jacob D. Bekenstein. Specific entropy and the sign of the energy. *Physical Review D*, 26(4):950–953, 1982.

[beke83] Jacob D. Bekenstein. Entropy bounds and the second law for black holes. *Physical Review D*, 27(10):2262–2270, 1983.

[beke83b] Jacob D. Bekenstein. Entropy content and information flow in systems with limited energy. *Physical Review D*, 30(8):1669–1679, 1984.

[beke88] Jacob D. Bekenstein. Communication and energy. *Physical Review A*, 37(9):3437–3449, 1988.

[bell64] J.S. Bell. On the einstein-podolsky-rosen paradox. *Physics*, 1:195, 1964.

[beni82b] P. Benioff. *J. Stat. Phys.*, 29:515–545, 1982.

[beni82a] Paul Benioff. Quantum mechanical models of turing machines that dissipate no energy. *Phys. Rev. Lett.*, 48:1581–1585, 1982.

[benn73] C.H. Bennett. The logical reversibility of computation. *IBM J. Res. Devel.*, 17:525–532, 1973.

[benn82] C.H. Bennett. The thermodynamics of computation – a review. *Int. J. Theor. Phys.*, 21(3/4):905–940, December 1982.

[benn87] C.H. Bennett. Demons, engines, and the second law. *Scient. American*, pages 108–116, November 1987.

[benn88] C.H. Bennett. Notes on the history of reversible computation. *IBM J. Res. Devel.*, 32(1):16–23, 1988.

[benn92a] C.H. Bennett, Francois Bessette, Gilles Brassard, Louis Salvail, and John Smolin. Experimental quantum cryptography. *J. Cryptology*, 5(3), 1992.

[benn85] C.H. Bennett and R. Landauer. The fundamental physical limits on computation. *Scientific American*, pages 48–71, July 1985.

[benn92c] Charles H. Bennett. Quantum cryptography using any two nonorthogonal states. *Phys. Rev. Lett.*, May 1992.

[benn93] Charles H. Bennett, Gilles Brassard, Claude Crepeau, and Marie-Helene Skubiszewska. Practical quantum oblivious transfer. In *Advances in Cryptology–Crypto-91 Proceedings*. Springer, 1993.

[benn92b] Charles H. Bennett, Gilles Brassard, and N. David Mermin. Quantum cryptography without bell's theorem. *Phys. Rev. Lett.*, 68:557–559, 1992.

[berm79] R. Berman. *Heat conduction in solids*. Oxford University Press, 1976.

[beth34] H. Bethe and W. Heitler. On the stopping of fast particles and the creation of positive electrons. *Proc. Royal Soc. London*, 146:83–112, 1934.

[bhab35] H.J. Bhabha. The scattering of positrons by electrons with exchange. *Proc. Royal Soc. London A*, 154:195–206, 1935.

[bjor64] James D. Bjorcken and Sidney D. Drell. *Relativistic quantum mechanics*. McGraw-Hill, 1964.

[bjor65] James D. Bjorcken and Sidney D. Drell. *Relativistic quantum fields*. McGraw-Hill, 1965.

[born80] Max Born and Emil Wolf. *Principles of Optics*. Pergamon, sixth corrected edition, 1980.

[boss86] Marie H. Bossel. C.R.Acad.Sci. Paris Ser I Math 302,1 (1986) 47-50.

[boss88] Marie H. Bossel. Z. Angew. Math. Phys. 39,5 (1988) 733-742.

[bowe67] Julius I. Bowen. On the capacity of a noiseless photon channel. *IEEE Trans. On Information Theory*, IT-13(2):230–236, April 1967.

[boyd61] G.D. Boyd and J.P. Gordon. Confocal multimode resonator for millimeter through optical wavelength masers. *Bell Syst. Tech. J.*, 40:489–508, March 1961.

[brei34] G. Breit and J.A. Wheeler. Collision of two light quanta. *Physical Review*, 46:1087–1091, December 1934.

[burk70] Arthur W. Burks, editor. *Essays on cellular automata*. University of Illinois Press, 1970.

[chan60] S. Chandrasekhar. *Radiative transfer*. Dover, 1960.

[chaz85] B.M. Chazelle and L. Monier. A model of computation for VLSI with related complexity results. *Journal of the ACM*, 32,3:573–588, 1985.

[choq84] Yvonne Choquet-Brouhat. Positive energy theorems. In B.S. DeWitt and R. Stora, editors, *Relativity, groups, and topology (Les Houches session XL)*, volume 2, pages 739–785. North Holland, 1984.

[chur90] Patricia Churchland. *The neurophilosphy of Mind*. MIT Press, 1990.

[demp77] A.P. Dempster, N.M. Laird, and D.B. Rubin. Maximum likelihood from incomplete data using the em algorithm. *J. Royal Stat. Soc. B*, 39:1–38, 1977.

[deut82] David Deutsch. Is there a fundamental limit on the rate at which information can be processed? *Physical Review Letters*, 48(4):286–288, 1982.

[deut85] David Deutsch. Quantum theory, the church-turing principle and the universal quantum computer. *Proc. Royal Soc. London A*, 400:97–117, 1985.

[deut89] David Deutsch. Quantum computational networks. *Proc. Royal Soc. London A*, 425:73–90, 1989.

[dira30] P.A.M. Dirac. On the annihilation of positrons and electrons. *Proc. Cambridge Philos. Soc.*, 26:361–375, 1930.

[dobr77b] R.L. Dobrushin and S.I. Ortyukov. Lower bound for the redundancy of self-correcting arrangements of unreliable functional elements. *Prob. Inform. Trans.*, 13:203–218, 1977.

[dobr77] R.L. Dobrushin and S.I. Ortyukov. Upper bound for the redundancy of self-correcting arrangements of unreliable functional elements. *Prob. Info. Transmission*, 13:203–218, 1977.

[alle86] R.C. Allen et al. A study of $\nu_e e^-$ elastic scattering. In *Proc. 12th int'l. conf. on Neutrino physics and astrophysics*, pages 392–399. World Scientific, 1986.

[eule36] H. Euler. *Ann. Phys.*, 26:398–, 1936.

[fabe78] Donald S. Faber and Henri Kron, editors. *Neurobiology of the Mauthner Cell*. Raven Press, 1978.

[fede89] Tomas Feder. Reliable computation by networks in the presence of noise. *IEEE Trans. Info. Theory*, 35(3):569–571, May 1989.

[feyn82] R.P. Feynman. Simulating physics with computers. *Int. J. Theor. Phys.*, 21(6/7):467, 1982.

[feyn85] R.P. Feynman. Quantum mechanical computers [reprinted in found. phys. 16,6 (1986) 507-532]. *Optics News*, 11, Feb 1985.

[fred82] E. Fredkin and T. Toffoli. Conservative logic. *Int. J. Theor. Phys.*, 21(3/4):219–253, April 1982.

[gabo50] D. Gabor. Communication theory and physics. *Philosophical Magazine*, 41:1161–1187, November 1950.

[gabo61] D. Gabor. Light and information. *Prog. Optics*, 1:109–, 1961.

[gacs83] P. Gacs. Reliable computation with cellular automata. *J. Computer Systems Sci.*, 32(32):15–78, 1986.

[gacs86] P. Gacs and J. Reif. A simple 3d real-time reliable cellular array. *J. Computer Systems Sci.*, 36:125–147, 1986.

[gall80] R.G. Gallager. *Information Theory and Reliable Communication*. Wiley, 2 edition, 1980.

[gibb75] G.W. Gibbons. Vacuum polarization and the spontaneous loss of charge by black holes. *Commun. Math. Phys.*, 44:245–264, 1975.

[gibb76] G.W. Gibbons and M.J. Perry. Black holes in thermal equilibrium. *Phys. Review Letters*, 36:985–, 1976.

[glim87] J. Glimm and A. Jaffe. *Quantum physics: a functional integral point of view*. Springer, 1987.

[gord61] J.P. Gordon. Information capacity of a communications channel in the presence of quantum effects. In J.R. Singer, editor, *Advances in quantum electronics*, pages 509–519. Columbia Univ. Press, New York, 1961.

[gord62] J.P. Gordon. Quantum effects in communications systems. *Proc. IRE*, 50:1898–1908, September 1962.

[AIPH] Dwight E. Gray, editor. *AIP handbook of physics*. McGraw-Hill, third edition, 1972.

[grei87] P.T. Greiling. High-speed digital ic performance outlook. *IEEE Trans. Microwave Theory Tech.*, 35:245–259, 1987.

[haje91] Bruce Hajek and Timothy Weller. On the maximum tolerable noise for reliable computation by formulas. *IEEE Trans. Info. Theory*, 37(2):388–391, March 1991.

[hart71] J.B. Hartle. Long-range neutrino forces exerted by kerr black holes. *Physical Review D*, 3:2938–2940, 1971.

[hart72] J.B. Hartle. Can a schwarschild black hole exert long range neutrino forces? In J. Klauder, editor, *Magic without magic, John Archibald Wheeler*, pages 259–276. W.H. Freeman, 1972.

[hart76] J.B. Hartle and S.W. Hawking. Path-integral derivation of black hole radiance. *Physical Review D*, 13:2188–2203, 1976.

[haug75] E. Haug. Bremstrahlung and pair production in the field of a free electron. *Zeit. f Naturf.*, 30a:1099–1113, 1975.

[hawk74] Stephen W. Hawking. Black hole explosions? *Nature*, 248:30–31, March 1974.

[hawk75] Stephen W. Hawking. Particle creation by black holes. *Commun. Math. Phys.*, 43:199–220, 1975.

[hawk76] Stephen W. Hawking. Black holes and thermodynamics. *Physical Review D*, 13:191–197, 1976.

[hawk77] Stephen W. Hawking. Zeta function regularization of path integrals in curved spacetime. *Commun. Math. Phys.*, 56:133–148, 1977.

[hawk72] Stephen W. Hawking and J.B. Hartle. Energy and angular momentum flow into a black hole. *Commun. Math. Phys.*, 27:283–290, 1972.

[heff62] H. Heffner. The fundamental limit of linear amplifiers. *Proc. IRE*, 50:1604–1608, 1962.

[heit54] W. Heitler. *The quantum theory of radiation*. Clarendon Press, Oxford, 3rd edition, 1954.

[hirs64] Hirschfelder, Curtiss, and Bird. *Molecular theory of gases and liquids*. Wiley, 1964.

[itzy80] Claude Itzykson and Jean-Bernard Zuber. *Quantum field theory.* McGraw-Hill, 1980.

[jauc76] J.M. Jauch and F. Rohrlich. *The theory of photons and electrons.* Springer, 2nd expanded edition, 1976.

[whee39] J.A.Wheeler and W.E. Lamb. Influence of atomic electrons on radiation and pair production. *Physical Review*, 55:858–862, 1939.

[jayn57] E.T. Jaynes. Information theory and statistical mechanics. *Physical Review*, 106(4):620–630, May 1957.

[kac66] Marc Kac. Can you hear the shape of a drum. *American Mathematical Monthly*, 73 part II:1–23, 1966.

[karp70] S. Karp, E.L. O'Neill, and R.M. Gagliardi. Communication theory for the free-space optical channel. *Proc. IEEE*, 58:1611–1626, October 1970.

[karp50] R. Karplus and M. Neuman. Nonlinear interactions between electromagnetic fields. *Physical Review*, 80:380–, 1950.

[karp51] R. Karplus and M. Neuman. The scattering of light by light. *Physical Review*, 83:776–784, 1951.

[klei29] O. Klein and Y. Nishina. Über die streuung von strahlung durch freie elektronen nach der neuen relativistische quantumelektrodynamik von Dirac. *Zeit. Physik*, 52:853–868, 1929.

[land62] H.J. Landau and H.O. Pollak. Prolate spheroidal wave functions, fourier analysis, and uncertainty, ii,iii. *Bell Syst. Tech. J.*, 40:65–84, 1295–1336, 1961.

[land80] L.D. Landau and E.M. Lifshitz. *Statistical Physics part 1.* Pergamon, third edition, 1980.

[land61] R. Landauer. Irreversibility and heat generation in the computing process. *IBM J. Res. and Dev.*, 5(3):183–191, July 1961.

[land82] R. Landauer. Uncertainty principle and minimal energy dissipation in the computer. *Int. J. Theor. Phys.*, 21(3/4):283–297, 1982.

[land85] R. Landauer. Fundamental physical limitations on the computing process. *Annals NY Acad. Sci.*, 426:161–170, 1985.

[land86] R. Landauer. Computation and physics: Wheeler's meaning circuit. *Found. Phys.*, 16:551–564, 1986.

[land89] R. Landauer. Computation, measurement, communication, and energy dissipation. In S. Haykin, editor, *Selected topics in signal processing*, pages 18–. Prentice-Hall, 1989.

[land91] R. Landauer. Information is physical. *Physics Today*, pages 23–29, May 1991.

[lebe63] D.S. Lebedev and L.B. Levitin. The maximum amount of information transmissible by an electromagnetic field. *Soviet Phys. Doklady*, 8(4):377–379, October 1963.

[leff90] H.S. Leff and A.F. Rex, editors. *Maxwell's demon – entropy, information, computing.* Princeton University Press, 1990.

[leig83] F.T. Leighton. *Complexity issues in VLSI.* MIT Press, 1983.

[leig86] F.T. Leighton and A.L. Rosenberg. Three-dimensional circuit layouts. *SIAM J. Comput.*, 15(3):793–813, 1986.

[leis83] C.E. Leiserson. *Area-efficient VLSI computation.* MIT Press, 1983.

[leng90] Thomas Lengauer. VLSI theory. In J. van Leeuwen, editor, *Handbook of theoretical computer science volume A*, pages 836–868. MIT Press, 1990.

[levi66] L.B. Levitin. Photon channels with small occupation numbers. *Problems of Information Transmission*, 2:48–56, 1966.

[lieb90] Elliott Lieb. The stability of matter. *Bull. AMS*, 22(1):1–49, January 1990.

[lifs80] E.M. Lifshitz and L.P. Pitaevskii. *Statistical Physics part 2.* Pergamon, 1980.

[likh77] K.K. Likharev. Dynamics of some single flux quantum devices: 1. parametric quantron. *IEEE Trans. Magnetics*, 13:242–244, 1977.

[likh82] K.K. Likharev. Classical and quantum limitations on energy consumption in computation. *Int. J. Theor. Phys.*, 21:311–326, 1982.

[likh86] K.K. Likharev. *Dynamics of Josephson Junctions and circuits.* Gordon and Breach, 1986.

[likh85] K.K. Likharev, S.V. Rylov, and V.S. Semenov. Reversible conveyor computation in an array of parametric quantrons. *IEEE Trans. Magnetics*, 21:947–, 1985.

[likh90] K.K. Likharev, V. Semenov, and A.B. Zorin. New possibilities for superconductor devices. In Ruggiero and Rudman, editors, *Superconducting devices.* Academic press, 1990.

[lodd70] A. Lodding, J.N. Mundy, and A. Ott. Isotope inter-diffusion and self diffusion in solid lithium metal. *Physica Status Solidi*, 38:552–569, 1970.

[loud83] Rodney Loudon. *The quantum theory of light.* Oxford University Press, second edition, 1983.

[mand59] L. Mandel. Fluctuation of photon beams; the distribution of the photoelectrons. *Proc. Phys. Soc.*, 74:233–243, August 1959.

[mand76] L. Mandel. Ideal light source for an optical communication channel. *J. Opt. Soc. Amer.*, 66:968–970, September 1976.

[marg84] Norman Margolus. Physics-like models of computation. *Physica*, 10D:81–95, 1984.

[mcel81] Robert J. McEliece. Practical codes for photon communication. *IEEE Trans. Info. Theory*, IT-27(4):393–398, July 1981.

[mins67] Marvin Minsky. *Computation: finite and infinite machines.* Prentice-Hall, 1967.

[misn73] C.W. Misner, K.S. Thorne, and J.A. Wheeler. *Gravitation.* W.H. Freeman, 1973.

[moll32] C. Moller. *Ann. d. Phys.*, 14:531–, 1932.

[mors53] Philip M. Morse and Herman Feshbach. *Methods of theoretical physics.* McGraw Hill, 1953.

[mott29] N.F. Mott. The scattering of fast electrons by atomic nuclei. *Proc. Royal Soc. London A*, 124:425–442, 1929.

[ozat90] H.M. Ozatkas and J.W. Goodman. Lower bound for the communication volume required for an optically interconnected array of points. *J. Optical Soc. America*, 7(11):20–26, November 1990.

[page82] Don N. Page. Comment on a universal upper bound on the entropy to energy ratio for bounded systems. *Physical Review D*, 26(4):947–949, 1982.

[paxi90] G. Paxinos, editor. *The human nervous system.* Academic Press, July 1990.

[pend83] J.B. Pendry. Quantum limits to the flow of information and entropy. *J. Physics A*, 16(10):2161–2171, 1983.

[pier81] John R. Pierce, Edward C. Posner, and Eugene R. Rodemich. The capacity of the photon counting channel. *IEEE Trans. Info. Theory*, IT-27(1):61–77, January 1981.

[pipp90] N. Pippenger. Developments in the synthesis of reliable organisms from unreliable components. In J. Glimm, J. Impagliazzo, and I. Singer, editors, *The legacy of John von Neumann*, volume 50 of *Proc. symp. pure math.* AMS, 1990.

[pipp85] Nicholas Pippenger. On networks of noisy gates. *Proc. IEEE Symp. Foundations of Computer Sci.*, 26:30–38, 1985.

[pipp88] Nicholas Pippenger. Reliable computation by formulas in the presence of noise. *IEEE Trans. Info. Theory*, IT-34(2):194–197, March 1988.

[pipp92] Nicholas Pippenger. Symmetry in self-correcting cellular automata. manuscript. *Tech Report, CS dept., University of British Columbia, Vancouver*, 1992.

[quig83] Chris Quigg. *Gauge theories of the strong, weak, and electromagnetic interactions.* Benjamin Cummings, 1983.

[redn84] R.A. Redner and H.F. Walker. Mixture densities, maximum likelihood, and the em algorithm. *SIAM Review*, 26:195–239, 1984.

[reic80] L.E. Reichl. *A modern course on statistical physics.* University of West Texas Press, Austin, 1980.

[reif65] John Reif. *Fundamentals of Statistical and Thermal Physics.* McGraw-Hill, 1965.

[robe29] H.P. Robertson. The uncertainty principle. *Physical Review*, 34:163–164, 1929.

[robi75] D.C. Robinson. Uniqueness of the kerr black hole. *Physical Review Lett.*, 34:905–906, 1975.

[rugg90] S.B. Ruggiero and D.A. Rudman, editors. *Superconducting devices.* Academic, 1990.

[russ78] R.M. Russell. The cray-1 computer system. *Communications of the ACM*, 21(1):63–72, January 1978.

[schi68] Leonard I. Schiff. *Quantum Mechanics.* McGraw-Hill, 1968.

[shan59] C.E. Shannon. Probability of error for optimal codes in a gaussian channel. *Bell Syst. Tech. J.*, 38(3):611–656, May 1959.

[shan49] Claude E. Shannon and W. Weaver. *The mathematical theory of communication.* Univ. Illinois Press, Urbana, 1949.

[shor94] Peter W. Shor. Algorithms for Quantum Computation: Discrete Log and Factoring, *Proc. 35th IEEE Symp. Foundations of Computer Sci.*, 124-134, Nov 1994.

[slep61] D. Slepian and H.O. Pollak. Prolate spheroidal wave functions, fourier analysis, and uncertainty, i. *Bell Syst. Tech. J.*, 40:43–63, 1961.

[slep74] David Slepian, editor. *Key papers in the development of information theory.* IEEE Press, 1974.

[spie95] Daniel A. Spielman. Linear-time encodable and decodable error-correcting codes, *Proc. 27th ACM Symp. Theory of Computing*, May 1995.

[ster60] T.E. Stern. Some quantum effects on information channels. *IRE Trans. Inform. Theory*, IT-6:435–440, September 1960.

[hawk79] S.W.Hawking and W. Israel, editors. *General Relativity, an Einstein centenary survey.* Cambridge University Press, 1979.

[sze81] Simon M. Sze. *Physics of semiconductor devices.* Wiley, 2 edition, 1981.

[szil29] L. Szilard. On the decrease of entropy in a thermodynamic system by the intervention of intelligent beings. *Zeitschr. Phys.*, 53:840–856, 1929.

[teit72a] C. Teitelboim. Nonmeasurability of the baryon number of a black hole. *Nuovo Cimento Letters*, II(3):326–328, 1972.

[teit72b] C. Teitelboim. Nonmeasurability of the lepton number of a black hole. *Nuovo Cimento Letters*, II(3):397–400, 1972.

[teit72c] C. Teitelboim. Nonmeasurability of the quantum numbers of a black hole. *Physical Review D*, 5:2941–2954, 1972.

[thom80] Clark D. Thompson. *A complexity theory for VLSI.* CMU Computer Science Dept PhD Thesis, 1980.

[thom33] J.J. Thompson. *Conduction of electricity through gases.* Cambridge University Prees, 1933.

[ullm84] J.D. Ullman. *Computational aspects of VLSI.* Computer Science Press, Rockville MD, 1984.

[unru82] W.D. Unruh and R.M. Wald. Entropy bounds, acceleration, radiation, and the second law. *Physical Review D*, 25:942–943, 1982.

[unwi82] S.D. Unwin. Possible violations of the entropy-to-energy ratio bound. *Physical Review D*, 26(4):944–946, 1982.

[vaid89] P.M. Vaidya. A new algorithm for minimizing convex functions over convex sets. In *30th IEEE Symp. on Foundations of Computer Science*, pages 338–343. IEEE, 1989.

[vita88] Paul M.B. Vitanyi. Locality, communication, and interconnect length in multicomputers. *SIAM J. Computing*, 17(4):659–673, August 1988.

[vNeu55] J. von Neumann. *The mathematical foundations of quantum mechanics.* Princeton University Press, 1955.

[neum58] J. von Neumann. *The computer and the brain.* Yale Univ. Press, 1958. (book published despite being incomplete when vN died.).

[vNeu66] J. von Neumann. *Theory of self-reproducing automata (companion volume to burk70; published despite being incomplete at time of vN's death).* University of Illinois Press, 1966.

[vNeu56] John von Neumann. Probabilistic logics and the synthesis of reliable organisms from unreliable components. In C.E. Shannon and J. McCarthy, editors, *Automata Studies*, pages 43–98. Princeton Univ. Press, Princeton NJ, 1956.

[vour90] A. Vourdas. Information in optical binary communications. *IEEE Trans. Info. Theory*, 36(1):208–211, January 1990.

[wall83] D.F. Walls. Squeezed states of light. *Nature*, 306:141–146, November 1983.

[ward50] J. Ward. The scattering of light by light. *Physical Review*, 77:293, 1950.

[whee57] J.A. Wheeler and E.A. Power. Thermal geons. *Reviews in Modern Physics*, 29:480–, 1957.

[whee83] J.A. Wheeler and W.H. Zurek, editors. *Quantum Theory and Measurement.* Princeton Univ. Press, 1983.

[witt81] Edward Witten. A new proof of the positive energy theorem. *Commun. Math. Phys.*, 80:381–402, 1981.

[yao81] Andrew C. Yao. The entropic limitations on VLSI computations. *Proc. ACM Symp. Theor. Computer Science*, 13:308–311, 1981.

[zure89] W. H. Zurek. Thermodynamic cost of computation; algorithmic complexity and the information metric. *Nature*, 34:119–124, 1989.

[zure84a] W.H. Zurek. Reversibility and stability of information processing systems. *Phys. Rev. Lett.*, 53:391, 1984.

[zure86] Wojciech H. Zurek. *Annals NY Acad. Sci.*, 480:89–97, 1986.

[zure91] Wojciech H. Zurek, editor. *Complexity, Entropy, and the physics of information*, volume VIII of *Advanced book program; Santa Fe Institute studies in the sciences of complexity.* Addison-Wesley, 1991.

[zure91b] Wojciech H. Zurek. Decoherence and the transition from quantum to classical. *Physics Today*, pages 36–44, October 1991.