Physical Limits of Computing

Article for *Computing in Science and Engineering*

Michael P. Frank

Computational science & engineering and *computer* science & engineering have a natural and long-standing relation.

Scientific and engineering problems tend to place some of the most demanding requirements on computational power, thereby driving the engineering of new bit-device technologies and circuit architectures, as well as the scientific & mathematical study of better algorithms and more sophisticated computing theory. For example, the need for finite-difference artillery ballistics simulations during World War II motivated the ENIAC, and massive calculations in every area of science & engineering motivate the PetaFLOPS-scale supercomputers on today's drawing boards.

Meanwhile, computational methods themselves help us to build more efficient computing systems. Computational modeling and simulation of manufacturing processes, logic device physics, circuits, CPU architectures, communications networks, and distributed systems all further the advancement of computing technology, together achieving ever-higher densities of useful computational work that can be performed using a given quantity of time, material, space, energy, and cost. Furthermore, the global economic growth enabled by scientific & engineering advances across many fields helps make higher total levels of societal expenditures on computing more affordable. The availability of more affordable computing, in turn, enables whole new applications in science, engineering, and other fields, further driving up demand.

Partly as a result of this positive feedback loop between increasing demand and improving technology for computing, computational efficiency has improved steadily and dramatically since the computing's inception. When looking back at the last forty years (and the coming ten or twenty), this empirical trend is most frequently characterized with reference to the famous "Moore's Law" [Moo65,Moo75,Moo95,Moo97], which describes the increasing density of microlithographed transistors in integrated semiconductor circuits. (See figure 1.)

Interestingly, although Moore's Law was originally stated in terms that were specific to semiconductor technology, the trends of increasing computational density inherent in the law appear to hold true even across multiple technologies. One can trace the history of computing technology back through discrete transistors, vacuum tubes, electromechanical relays, and gears, and amazingly, we still see the same exponential curve extending across all these many drastic technological shifts. Furthermore, when looking back far enough, the curve even appears to be super-exponential; the frequency of doubling of computational efficiency appears to *itself* increase over the long term ([Kur99], pp. 20-25).

Naturally, we wonder just how far we can reasonably hope this fortunate trend to take us. Can we continue indefinitely to build ever more and faster computers using our available economic resources, and apply them to solve ever larger and more complex scientific and engineering problems? What are the limits? *Are* there limits? When semiconductor technology reaches its technology-specific limits, can we hope to maintain the curve by jumping to some alternative technology, and then to another one after that one runs out?

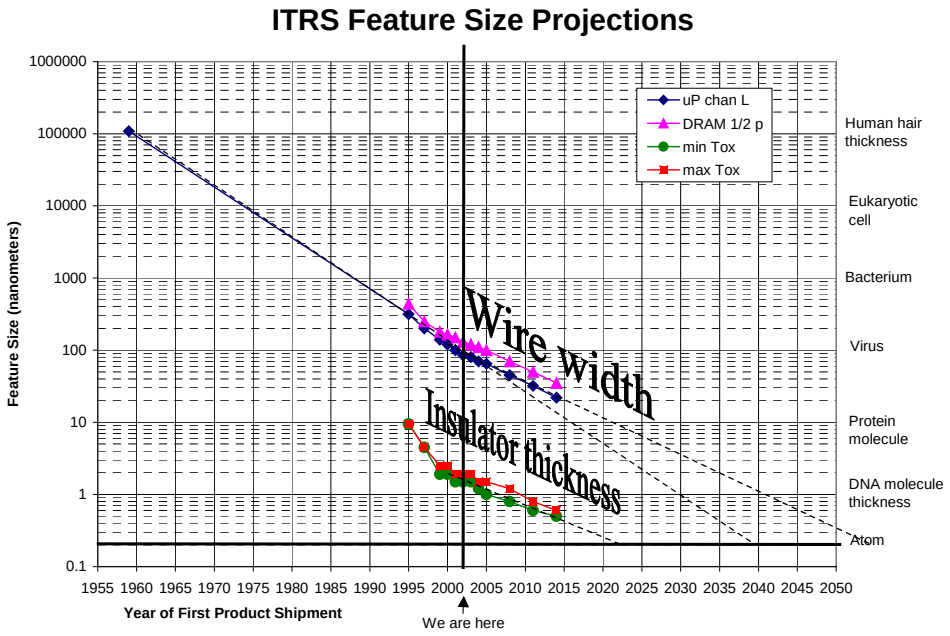

Figure 1. Trends for minimum feature size in semiconductor technology. The data in the middle are taken from the 1999 edition of the International Technology Roadmap for Semiconductors. The point at left represents the first planar transistor, fabricated in 1959. Presently, the industry is actually beating the roadmap targets that were set just a few years ago; ITRS targets have historically always turned out to be conservative, so far. From this data, wire widths, which correspond most directly to overall transistor size, would be expected to reach near-atomic size by 2040 or 2050, unless the trend line levels off before then. It is interesting to note that in some experimental processes, some features such as gate oxide layers are *already* only a few atomic layers thick, and cannot shrink further.

Obviously, it is always a difficult and risky proposition to try to forecast future technological developments. However, 20th-century physics has given forecasters a remarkable gift, in the form of the very sophisticated modern understanding of physics, as embodied in the Standard Model of particle physics. According to all available evidence, this model, together with general relativity, explains the world so successfully that apparently *no experimentally accessible phenomenon* fails to be encompassed within it. That is to say, no definite and persistent inconsistencies between the fundamental theory and empirical observations have been uncovered in physics within the last couple of decades.

And furthermore, in order to probe beyond the range where the theory has already been thoroughly verified, physicists find that they must explore subatomic-particle energies above a trillion electron volts, and length scales far tinier than a proton's radius. The few remaining serious puzzles in physics, such as the masses of particles, the disparity between the strengths of the fundamental forces, and the unification of general relativity and quantum mechanics are of a rather abstract and aesthetic flavor. Their eventual resolution (whatever form it takes) is not currently expected to have any significant applications until one reaches the highly extreme regimes that lie beyond the scope of present physics (although, of course, we cannot assess the applications with certainty until we *have* a final theory).

In other words, we expect that the fundamental principles of modern physics have "legs," that they will last us a while (many decades, at least) as we try to project what will and will not be possible in the comings evolution of computing. By taking our best theories seriously, and exploring the limits of what we can engineer with them, we push against the bounds of what we think we can do. If our present understanding of these limits eventually turns out to be seriously wrong, well, then the act of pushing against the limits is probably the activity that is most likely to lead us to that very discovery. (This methodology is nicely championed by Deutsch [Deu97].)

So, I personally feel that forecasting future limits, even far in advance, is a useful research activity. It gives us a roadmap as to where we may expect to go with future technologies, and helps us know where to look for advances to occur, if we hope to ever circumvent the limits imposed by physics, as it is currently understood.

Interestingly, just by considering fundamental physical principles, and by reasoning in a very abstract and technology-independent way, one can arrive at a number of firm conclusions about upper bounds, at least, on the limits of computing. I have furthermore found that often, an understanding of the general limits can then be applied to improve one's understanding of the limits of specific technologies.

Let us now review what is currently known about the limits of computing in various areas. Throughout this article, I will focus primarily on *fundamental*, technology-independent limits, since it would take too much space to survey the technology-specific limits of all present and proposed future computing technologies.

But first, before we can talk sensibly about information technology in physical terms, we have to define *information* itself, in physical terms.

Information and Physical Entropy: Two kinds of "Infropy"

From a physical perspective, what *is* information? For purposes of discussing the limits of information technology, the relevant definition relates closely to the physical quantity known as *entropy*. As we will see, entropy and information are really but alternate forms of the same fundamental physical entity, which I will dub *infropy* (to coin a word), so as to emphasize the unification of the two concepts, while avoiding the connotations attached to either of the more specific terms. Think of "infropy" as a convenient abbreviation for "information and/or entropy".

The concept of entropy was introduced before it was understood to be an informational quantity. Historically, it was Boltzmann who first identified the *maximum entropy S* of any physical system with the logarithm of its total number of possible, mutually distinguishable states. (This discovery is carved on his tombstone.) I will call this same quantity the *total infropy* of the system, for reasons to soon become clear.

In Boltzmann's day, it was a bold conjecture to presume that the number of states was a finite one that admitted of a logarithm. But today, we know that operationally distinguishable states correspond to orthogonal quantum state-vectors, and the number of these for a given system is well-defined in quantum mechanics, and is finite for finite systems (more on this later).

Now, any logarithm by itself is a pure number, but the logarithm base that one chooses in Boltzmann's relation determines the appropriate *unit* of infropy. Using base 2 gives us the infropy unit of 1 *bit*, while the natural logarithm (base *e*) gives us a unit I like to call the *nat*, which is simply ($\log_2 e$) bits. The *nat* is also more widely known as

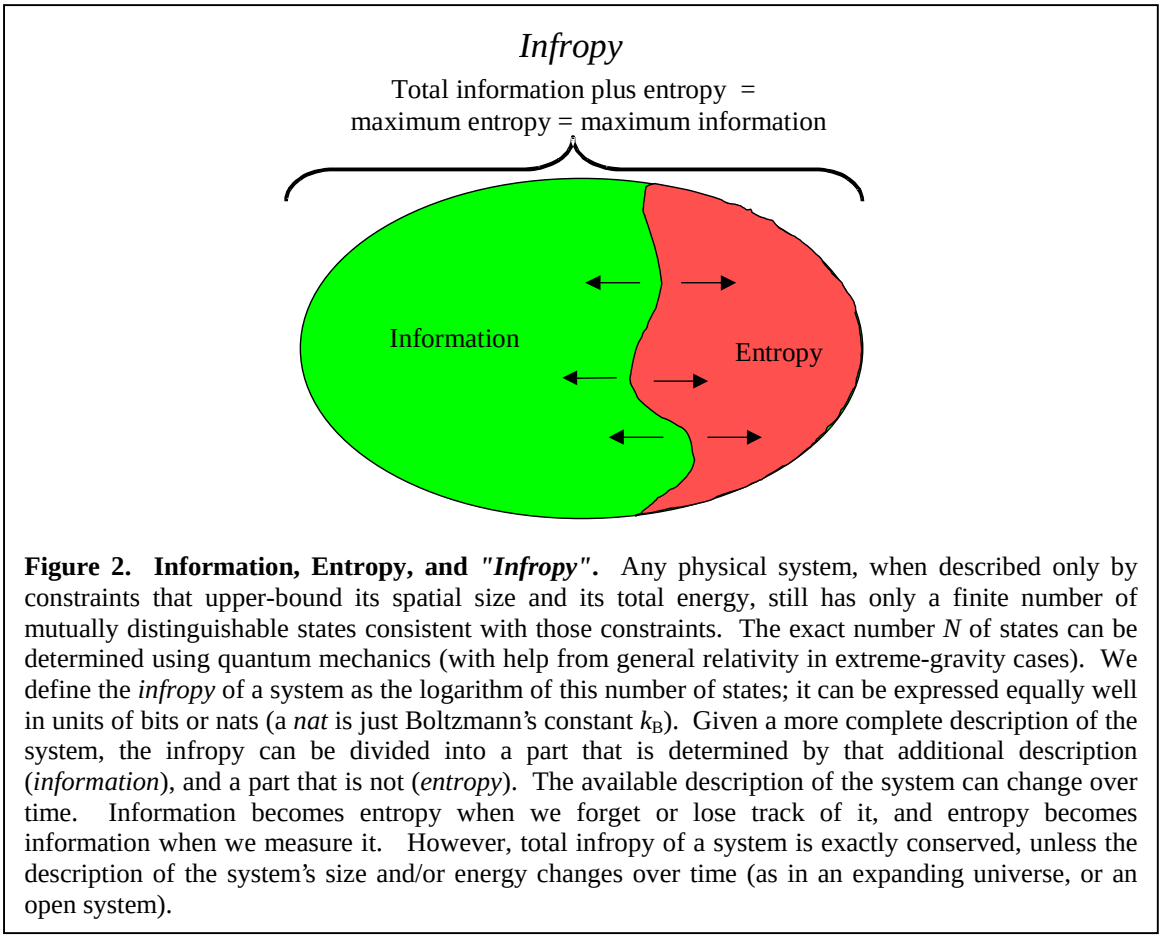

Figure 2. Information, Entropy, and *"Infropy".* Any physical system, when described only by constraints that upper-bound its spatial size and its total energy, still has only a finite number of mutually distinguishable states consistent with those constraints. The exact number N of states can be determined using quantum mechanics (with help from general relativity in extreme-gravity cases). We define the *infropy* of a system as the logarithm of this number of states; it can be expressed equally well in units of bits or nats (a *nat* is just Boltzmann's constant k_B). Given a more complete description of the system, the infropy can be divided into a part that is determined by that additional description (*information*), and a part that is not (*entropy*). The available description of the system can change over time. Information becomes entropy when we forget or lose track of it, and entropy becomes information when we measure it. However, total infropy of a system is exactly conserved, unless the description of the system's size and/or energy changes over time (as in an expanding universe, or an open system).

Boltzmann's constant k_B. Any of these units of infropy can be associated with physical units of energy divided by temperature, because temperature itself can be defined as just a measure of energy required per increment in the log state count, $T = \partial E/\partial S$. For example, a temperature of 1 degree Kelvin can be defined as a requirement of 1.38×10^{-23} Joules of energy to increase the log state count by 1 nat (that is, to multiply the state count by e). A bit, meanwhile, corresponds to a requirement of 9.57×10^{-24} Joules per degree Kelvin, the energy needed to double the state count.

Now, that's infropy, but what distinguishes physical entropy from physical information? The distinction is fundamentally observer-dependent, but in a way that is well-defined, and that coincides for most observers in simple cases.

Information is the infropy in the part of the system that is *known* (by a particular observer), and entropy is the infropy in the part that is *unknown*. The meaning of "known" can be clarified, by saying that a system A (the observer) *knows* the state of system B (the observed system) to the extent that some part of the state of A (some record or memory) is correlated with the state of B, and that the observer can access and interpret the implications of that record regarding the state of B.

To quantify things, the *maximum information* or *maximum entropy* of any system is, as already stated, just the log of its possible number of distinguishable states. If we know nothing about the state, all the system's infropy is entropy, from our point of view. But, as a result of preparing or interacting with a system, we may come to know (or

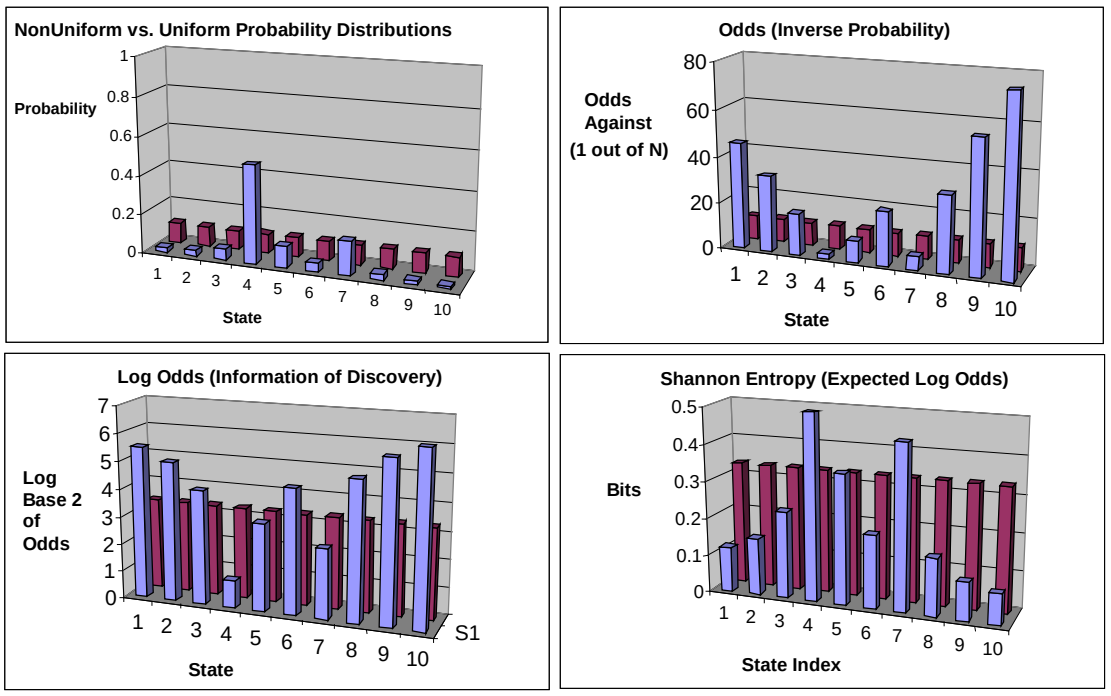

Figure 3. Shannon Entropy. The figure shows an example of Shannon's generalization of Boltzmann entropy for a system having ten distinguishable states. The blue bars correspond to a specific nonuniform probability distribution over states, while the purple bars show the case with a uniform (Boltzmann) distribution. The upper-left chart shows the two probability distributions. Note that in the nonuniform distribution, we have a 50% probability for the state with index 4. The upper-right chart inverts the probability to get the odds against the state; state 4 is found in 1 case out of 2, whereas state 10 (for example) appears in 1 case out of 70. The logarithm of this "number of cases" (lower left) is the information gain if this state were actually encountered; in state 4 we gain 1 bit; in case 10, more than 6 bits (2^6=64). Weighting the information gain by the state probability gives the expected information gain. Because the logarithm function is concave-down, a uniform distribution minimizes the expected log-probability, maximizes its negative (the expected log-odds, or entropy), and minimizes the information (the expected log-probability, minus that of the uniform distribution).

learn) something more about its actual state, besides just that it is one of the N states that were originally considered "possible."

Suppose we learn that the system is in a particular subset of $M<N$ states; only the states in that set are then possible, given our knowledge. Then, the entropy of the system, from our new point of view, is log M, whereas to someone without this knowledge, it is log N. For us, there is (log N) – (log M) = log(N/M) *less* entropy *in* the system. We say we now know log(N/M) more information *about* the system, or in other words that log(N/M) more of its infropy is information (from our point of view). The remaining log M amount of information, *i.e.*, the infropy still unknown in the system, we call *entropy*.

So, you can see that if we know *nothing* about the system, it has entropy log N and we have log N – log N = 0 information about it. If we know the *exact* state of the system, then it has log 1 = 0 entropy, and we have log N – 0 = log N information. Anywhere in between, the system has some intermediate entropy, and we have some intermediate information. Claude Shannon showed how the definition of entropy could be appropriately generalized to situations where our knowledge about the state x is expressed not as a subset of states, but as a probability distribution p_x over states. In that

case the entropy is just $H = \sum_x p_x \log p_x$. The information is $\log N - H$. Note that the Boltzmann definition of entropy is just the special case of Shannon entropy where p_x happens to be a uniform distribution over all N states (see figure 3).

Anyway, regardless of our state of knowledge, note that *the sum of the system's entropy and its information is always conserved*. Information and entropy are just two forms of the same fundamental quantity, somewhat analogously to kinetic and potential energy. Whether a system has information or entropy just depends on whether *our state is correlated* with the system's state, or whether the states are independent. Information is just known entropy. Entropy is just unknown information.

Interestingly, infropy is, apparently, like energy, a *localized* phenomenon. That is, it has a definite location in space, associated with the location of the subsystem whose state is in question. Even information *about* a distant object can be seen as just information in the state of a local object (*e.g.* a memory cell) whose state happens to have become correlated with the state of the distant object through a chain of interactions. Information can be viewed as always flowing locally through space, even in quantum systems [Deu99].

It may seem at first that the relative location and velocity of two isolated, widely-separated objects (a kind of information) provides a counter-example of information that cannot sensibly be localized in either object. But general relativity teaches us that another relationship-determined quantity, namely the energy associated with the objects (which includes their potential energy due to the forces between them, and their relative kinetic energy) *is* actually localized to the space in and around the objects in a very definite way. The precise distribution of this energy has a definite physical effect, manifested in a deformation of the surrounding spacetime fabric. So, the information associated with the relative configuration of two objects must be spread out over the fields that surround the objects, with some definite distribution, perhaps one that is closely related to the energy distribution (though the precise relationship between these distributions is not yet clear to me).

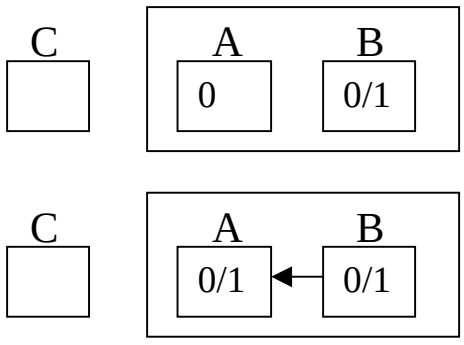

Figure 4. Entropy & measurement. Suppose system B, which contains 1 bit of entropy (two possible states, labeled 0 and 1), is measured by system A (arrow). Now, system A is correlated with B, and B's infropy is now information, from A's point of view. But from the point of view of an outside observer C who does not have access to A's record, the combined system still has 1 bit of entropy, since it could be either in state (A=0,B=0) or state (A=1,B=1). Since physics is reversible, the number of possible states of the whole system can't decrease from the point of view of an outsider who isn't measuring the state. But it can increase, for example, if C loses track of the interactions taking place between A and B, in which case all 4 joint states of the AB system become possibilities from C's point of view.

Further, a system's entropy may be converted to information by measurement, and information may be converted into entropy by forgetting (or erasure of information). But the sum of the two

in a given system is always a constant, unless the maximum number of possible states in the system is itself changing, which may happen if the system's changes in size, or if energy is added or removed. Actually, it turns out that in an expanding universe, the number of states (and thus the total infropy) *is* increasing, but in a small, local system with constant energy and volume, we will see that it is a constant.

To say that entropy may be converted to information through observation may at first sound like a contradiction of the second law of thermodynamics, that entropy always increases. But remember, information is really just another *kind* of entropy, when we say we're "converting" it, we really mean that we're giving it a different label, not that there is less of it. Also, from a point of view external to the whole measurement process, the entropy, even if extracted from the original system through measurement, is still there (and still entropy) from the external point of view. (See figure 4.)

But if entropy can be moved out of a system by measurement, then couldn't you theoretically remove all the entropy from a cylinder of gas (for example) by repeated measurements, freezing it into a known state, and gaining its heat energy as work? Then, couldn't you get more free energy, by allowing the cylinder to be warmed again by its surroundings while expanding against a piston, and repeat the experiment *ad infinitum* as a perpetual motion machine?

This question is exactly the famous Maxwell's Demon "paradox", which only seemed like a paradox (resisting all attempts at resolution) before Charles Bennett of IBM finally resolved it [Ben82,Ben88] with the realization that you have to keep track of where the extracted information goes. Sure, you can take entropy (and energy) out of a system, but you have to put the infropy somewhere, you can't just "disappear" it. Wherever you put it, you will require energy to store it. You'll need *less* energy, if you put the infropy in a lower-temperature system, but the resulting gain of work isn't forbidden by thermodynamics, it's just how any heat engine works!

Now, Boltzmann developed his definition of entropy in the context of classical mechanics by making the seeming *ad hoc* assumption that even the seemingly-continuous states of classical mechanics were somehow discretized into a finite number that admitted a logarithm. However, this notion was later vindicated, when Max Planck and the entire subsequent development of quantum mechanics showed that the world *was* discretized, at least in the relevant respects. The entire classical understanding of the relations between entropy, energy, temperature, *etc.*, remained essentially valid, forming the whole field of quantum statistical mechanics, a cornerstone of modern physics. Only the definition of entropy had to be further generalized, since partially-known states in quantum mechanics are described not by probability distributions, but by a generalization of a probability distribution called a *mixed state* or *density operator*, which can be represented (in finite cases) by *density matrices*. However, entropy can still be defined for these more complex objects in a way that remains perfectly consistent with the more restricted cases addressed by Boltzmann and Shannon (see fig. 5).

The study of the dynamic evolution, under physics, of mixed states leads to a fairly complete understanding of how the irreversible behavior described by the second law arises out of reversible microphysics, and how the quantum world appears classical at large scales [Zur01]. It all comes down to information and entropy. Quantum states, obeying a wave equation, tend to disperse outside of any localized region of state space to which they are initially confined. They evolve deterministically, but when you project

$$\rho = \begin{bmatrix} \rho_{11} & \rho_{12} & \cdots & \rho_{1n} \\ \rho_{21} & \rho_{22} & \cdots & \rho_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{n1} & \rho_{n2} & \cdots & \rho_{nn} \end{bmatrix} \Longrightarrow \rho' = \begin{bmatrix} \rho'_{11} & 0 & \cdots & 0 \\ 0 & \rho'_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \rho'_{nn} \end{bmatrix}$$

von Neumann entropy:

$$\begin{aligned} H(\rho) &= -\mathrm{Tr}\ \rho \ln \rho \\ &= -\mathrm{Tr}\ \rho' \ln \rho' \\ &= -\sum_i \rho'_{ii} \ln \rho'_{ii} \\ &\le -\sum_i \rho_{ii} \ln \rho_{ii} \\ &\le \ln n \end{aligned}$$

Figure 5. Density matrix representation of probabilistic mixtures of quantum states. The rows and columns of ρ are indexed by the system's distinguishable states (in any basis). Each diagonal element ρ_{ii} just gives the probability of basis state i. The off-diagonal elements ρ_{ij}, $i \neq j$ are complex numbers that specify quantum *coherences* between the basis states. Any density matrix ρ has a unique basis such that when ρ is re-expressed in that basis, the resulting matrix ρ' is diagonal, and represents a classical mixture of $\le n$ basis states. The basis-independent *von Neumann entropy* of a mixed state is given by $H = -\mathrm{Tr}\ \rho \ln \rho$ (where ln represents a matrix logarithm, defined as the inverse of matrix exponential, which is defined by a Taylor-series expansion of e^M). This quantity is exactly the same as the Shannon entropy (in nat units) of the probability distribution specified along the diagonal of the *diagonalized* density matrix ρ'. The von Neumann entropy of a (not necessarily diagonal) density matrix ρ is always less than or equal to the Shannon entropy of ρ's own diagonal, which is in turn always less than or equal to the Boltzmann entropy, ln n.

their quantum state down to a classical probability distribution, you see that an initial sharp probability distribution tends to spread out, increasing the Shannon entropy of the state over time. The state, looked at from the right perspective or "basis," is still as definite as it was, but as a matter of practice, we generally lose track of its detailed evolution; so the information the system had (from our point of view) becomes entropy.

Quantum computing, on the other hand, is all about isolating a system and maintaining enough control over its evolution so that we can keep track of its exact quantum state as it deterministically changes; the "infropy" in a quantum computer's state is therefore information, not true entropy.

However, most systems are not so well isolated; they leak state information to the outside world; the environment "measures" their state, as it were. The environment becomes then correlated with the state; the system's state information becomes mixed up with and redundantly spread out over arbitrarily large-scale surrounding systems. This precludes any control over the evolution of that state information, and so we fail to be able to elicit any quantum interference effects, which can only appear in well-defined deterministic situations. The mechanism by which even gradual measurement by the environment eats away at quantum coherences, effectively devolving a pure quantum state into a (higher-entropy) mixed state, and making the large-scale world appear to have an objective classical state, is by now quite well understood [Zur01].

Information Storage Limits

Now that we know what information (or, we should really use our more general term "infropy") physically *is* (more or less), let's talk about some of the limits that can be placed on it, based on known physics.

An arbitrary quantum wavefunction, as an abstract *mathematical* object, requires infinite information to *describe* precisely, because it is selected from among an uncountable set of possible wavefunctions. But remember, the key definition for physical information, ever since Boltzmann, is not the number of *mathematically* distinct states, but rather the number of operationally *distinguishable* states. Quantum mechanics gives

distinguishability a precise meaning: Two states are 100% distinguishable if and only if (considered as complex vectors) they are orthogonal.

A basic result in quantum statistical mechanics is that the total number of orthogonal states for a system consisting of a constant number of non-interacting particles, with relative positions and momenta, is roughly given by the numerical volume of the particles' joint configuration space or *phase space* (whatever its shape), when measured using length and momentum units chosen so that Planck's constant *h* (which has units of length times momentum) is equal to 1. Therefore, so long as the number of particles is finite, and the volume of space occupied by the particles is bounded, and their total energy is bounded, then even though (classically) the number of point particle states is uncountably infinite, and even though the number of possible quantum wavefunctions is also uncountably infinite, the *amount of infropy in the system is finite!*

Now, this model of a constant number of non-interacting particles is a bit unrealistic, since in quantum field theory (the relativistic version of quantum mechanics), particle number is not constant; particles can split (radiation) and merge (absorption). To refine the model one has to talk about possible field states with varying numbers of particles. However, this still turns out not to fundamentally change the conclusion of finite infropy for any system of bounded size and energy. Warren Smith of NEC [Smi95] and Seth Lloyd of MIT [Llo00] have, in independent papers, given an excellent description of the quantitative relationships involved.

Smith [Smi95] argues for an upper bound to entropy *S* per unit volume *V* of $\frac{S}{V} = \left(\frac{q}{2}\right)^{1/4} \frac{16\sqrt{\pi}}{3 \cdot 60^{1/4}} \left(\frac{c}{\hbar} \cdot \frac{M}{V}\right)^{3/4}$, where *q* is the number of distinct particle types (including different quantum states of a given particle type), *c* is the speed of light, $\hbar$ is Planck's constant, and *M* is the total (gravitating) mass-energy of the system. As an example, using only photons with 2 polarization states, a 1 m^3 box containing 1000 kg of light could contain at most 6×10^{34} bits, or 60 kb per cubic Ångstrom (roughly a small atom-sized volume). However, achieving this limit for stable storage is probably unrealistic, since light with this mass density (that of water) would have a temperature of nearly a billion degrees, and exert a pressure on the order of 10^{16} pounds per square inch!

In [Llo00], Lloyd presents a bound nearly identical to Smith's, and derived based on similar arguments. It differs from Smith's only in that it is tighter by the small constant factor of $2\sqrt{2}$. Lloyd presents the example of a 1 kg, 1 liter "ultimate laptop" (at the density of water again) for which, using the same 2-state photon assumption as Smith, the maximum entropy would be 2.13×10^{31} bits, the same entropy density as in Smith's example, less the factor of $2\sqrt{2}$.

One should note that these field-theory based limits do not take into account the effects of gravity and general relativity. Based on very general grounds, Bekenstein [Bek81] has proved a (usually much looser) entropy limit for a system of given size and energy, that holds even when taking general relativity into account: $S < 2\pi ER/\hbar c$, where *E* is total energy and *R* is the system's radius. The only systems known to actually attain this entropy bound are black holes. *The physical system of least radius that can contain a given amount of entropy is a black hole.* (Black hole "radius" has a standard, meaningful definition even in the severely warped spacetime in and around a black hole.) Interestingly, the entropy of a black hole is proportional to its surface area (suitably

defined), not to its volume, as if all the information about the hole's state were stuck at its surface (event horizon). A black hole has exactly 1/4 nat of entropy per square Planck length of surface area (a Planck length is a fundamental unit of length equal to 1.6×10^{-35} m). In other words, the absolute minimum physical size of 1 nat's worth of infropy is a square exactly 2 Planck lengths on a side!

The Bekenstein bound is truly enormous. A hypothetical 1-meter radius (mainframe-sized) machine that achieved this bound would have an average entropy density throughout its volume of 10^{39} bits per cubic Ångstrom, much higher than the limit for the water-density machines described earlier. However, this "machine" would also be a black hole with roughly the mass of the planet Saturn. (I/O difficulties aside, this object might be somewhat difficult to accommodate safely in your average machine room!)

Now of course, both the field-theory and Bekenstein bounds on entropy density are only technology-independent upper bounds. Whether we can come anywhere close to reaching these bounds in any realistic computing technology is another question entirely. Both these bounds require considering states of quantum fields. However, it seems impossible to constrain or control the state of a field in definite ways without a stable surrounding and/or supporting structure. Arbitrary field states are not stable structures. For stability, it seems that one requires long-lived, bound particle states, such as one finds in molecules, atoms and nuclei. This leads us to our next question:

How many bits can you store in an atom?

Nuclei may have an overall spin orientation, which is encoded using a vector space of only dimensionality 2, so the spin only holds 1 bit of infropy. Aside from the spin variability, at normal temperatures, a given nucleus is normally frozen into its quantum ground state. It can only contain additional information if it is excited to higher energy levels. But, excited nuclei are not stable—they are radioactive and decay rapidly, emitting high-energy, damaging particles. Not a very promising storage medium!

Electron configuration is another possibility. Outer-shell electrons may have spin variability, and excited states that, although still unstable, at least do not present a radiation hazard. There may be many ionization states for a given atom that may be reasonably stable in a well-isolated environment. This presents another few bits.

The choice of nucleus species in the atom in question presents another opportunity for variability. However, there are only a few hundred reasonably stable species of nuclei, so at best (even if you have a storage location that can hold any species of atom) this only gives you about an additional 8 bits or so.

An atom in a potential energy well (relative to its neighbors) generally has 6 restricted degrees of freedom, three of position and three of momentum. Each one of these permits exactly one nat's worth of entropy, for vibrational states. So this gives us a few more bits. However, phonons (the quantum particles of mechanical vibration) can easily dissipate out into any mechanical supporting structure, so they do not represent stable storage.

Of course, an arbitrarily large number of bits could be encoded in an atom's position and momentum along *unrestricted* degrees of freedom, *i.e.*, in infinitely large open spaces. However, given bounded spaces and energies, the entropy is still limited by the classical limit of log phase-space-volume (mentioned earlier). Since entropy per atom grows with only *log* volume, entropy *density* (per volume) actually *shrinks* with

increasing volume. So, spreading atoms out, though it increases entropy *per atom* (by some small number of bits) does not increase entropy density. If we are interested in maximizing information *density* using atoms, then we should focus on dense, solid-state materials.

For example, an analysis I performed of the entropy density in pure copper, based on standard CRC tables of empirically-derived thermochemical data, suggests that (at atmospheric pressures) the actual entropy density falls in the rather narrow range of only 0.5 to 1.5 bits per cubic Ångstrom (an Ångstrom is 10^{-10} m, about the radius of a hydrogen atom), over a wide range of temperatures, from room temperature up to just below the metal's boiling point. Entropy densities in a variety of other pure elemental materials are also near this level, though copper had the highest entropy density of the materials I studied. The entropy density would be expected to be somewhat greater for mixtures of elements, but not by much.

One can try to further increase entropy densities by applying high pressures. But if the size of the surrounding structure which is applying the pressure is taken into account, the overall *average* entropy density would probably not increase by very much. An exception might be for the interior of a large gravitationally-bound object, such as an oversize planet. But I don't think we'll be engineering Jupiter into a giant computer anytime *too* soon. Neutron stars are very much denser, but even less feasible.

Based on all this, I would be quite surprised if an infropy density greater than, say, ~10 bits per cubic Ångstrom could be achieved for stable, retrievable storage of digital information anytime within, say, the next 100 years. And, if this limit is *ever* exceeded, I predict that this will require either planetary-scale engineering, or radical and surprising new developments in fundamental physics.

Note, however, that even at an infropy density of only 1 bit/Å^3 (which the Moore's Law trend would have us reach in another 40 years or so), a convenient 1 cm^3 (sugar-cube size) lump of material could theoretically hold 10^{24} bits of information. This quantity (or actually the slightly greater quantity 2^{80} bits) is known, in obscure jargon[1], as 1 *yottabit* or 1 Yb. In more familiar units, it is ~100 billion terabytes, much greater than the total digital storage in the entire world today.

Minimum Energy for Information Storage

One of the most important raw resources involved in computing, besides time and space and manufacturing cost, is energy. When we talk about "using up energy," we really mean converting *free energy* into *heat energy*, since energy itself is conserved. A chunk of energy always carries an associated chunk of infropy. Temperature, as we saw earlier, is just the slope of the energy *vs.* infropy curve for a given (open) system whose total energy and infropy are (for whatever reason) subject to change. Free energy is energy all of whose infropy happens to be information. In other words, the energy is in a a precisely-known state. Heat is energy all of whose infropy happens to be entropy. In other words, its state is completely unknown. Converting free energy to heat energy just means converting the energy's infropy from information to entropy.

Free energy is useful for a couple of reasons. First, by converting it into different forms (all still kinds of energy) we can move things and lift things; do useful work. In

[1] The "yotta" prefix was adopted in 1990 by the *19th Conférence Générale des Poids et Mesures (CGPM), cf.* `http://www.bipm.fr/enus/3_SI/si-prefixes.html`.

particular, it gives us a means to move *entropy* itself out of a system where it is not wanted (perhaps because we want to use the infropy of the system to contain some definite information instead).

Suppose you have a system A containing 1 bit of entropy, and you bring a system B next to it that contains 1 bit of information. Suppose further you arrange to set up an interaction between the two systems that has the effect of swapping their (infropic) contents. Now, the system A contains the information, and system B the entropy. You then move system B away physically, and the unwanted entropy is now out of your hair (though still not destroyed).

Free energy can be seen as a general term for an information-carrying "system B" whose contents can be replaced with entropy, before transmitting the energy out into some external environment.

Suppose we want to reduce the entropy of a target system (clean it up, or cool it down), we can use free energy to do this, by doing a swap: Exchange the entropy of the target with the information in the energy, ending up with a target system with low entropy and lots of information (clean, cool, ordered), and energy with more entropy (hot, disordered). The hot energy moves away. Cleaning up a messy room or cooling down a refrigerator are both everyday examples of this process.

"Erasing" a computer memory is another example of this process. Suppose we want to transfer 1 bit of infropy (whether it's known or not) out of the computer's memory by swapping it onto a chunk of free energy (having at least 1 bit of infropy) which is then carefully transported, and some of it emitted, into an uncontrolled external environment E. If the outside environment E has temperature T, then a minimum amount $T\cdot$(1 bit) of energy, by the very definition of temperature, must be emitted into that environment in order to carry the 1 bit's worth of added infropy into it. Simply re-expressing the logarithm as base e rather than 2 gives $T\cdot$(1 bit) = $T\cdot$(ln 2 nat) = $k_B T$ ln 2 energy, the famous expression of the minimum energy for bit erasure that was first discovered by Landauer [Lan61] (though von Neumann had previously hinted at something similar [von66]).

Note that if the information-carrying energy isn't transported carefully, it is likely to simply be emitted into the device's immediate surroundings, which will probably be fairly warm inside the core of a high-performance computer, and so the free energy requirement is fairly large. However, in the ideal case, the energy would first be transported to a very cool external environment (*e.g.*, into space, ideally to some system in close thermal contact with the cosmic microwave background at 3 K), and so the free energy actually emitted need be much less in that case.

So, what is the minimum energy required to "store a bit of information?" That depends. If all that is needed is to reversibly change the state of a storage location from one definite state to another, then there is no lower limit, apparently [Ben73,Ben82]. However, if the storage location already contains entropy, or just some information that we wish to forget, then we have to move this existing infropy out of the system to the unlimited space in the external world, which costs $k_B T$ ln 2 free energy, where T is the temperature at the point where we thermalize (lose control of) the infropy. We cannot simply wish away the infropy, somehow compressing the system's number of possible states, because physics is reversible, and phase space is incompressible (Liouville's theorem). The true entropy (von Neumann entropy) of a mixed quantum state of an

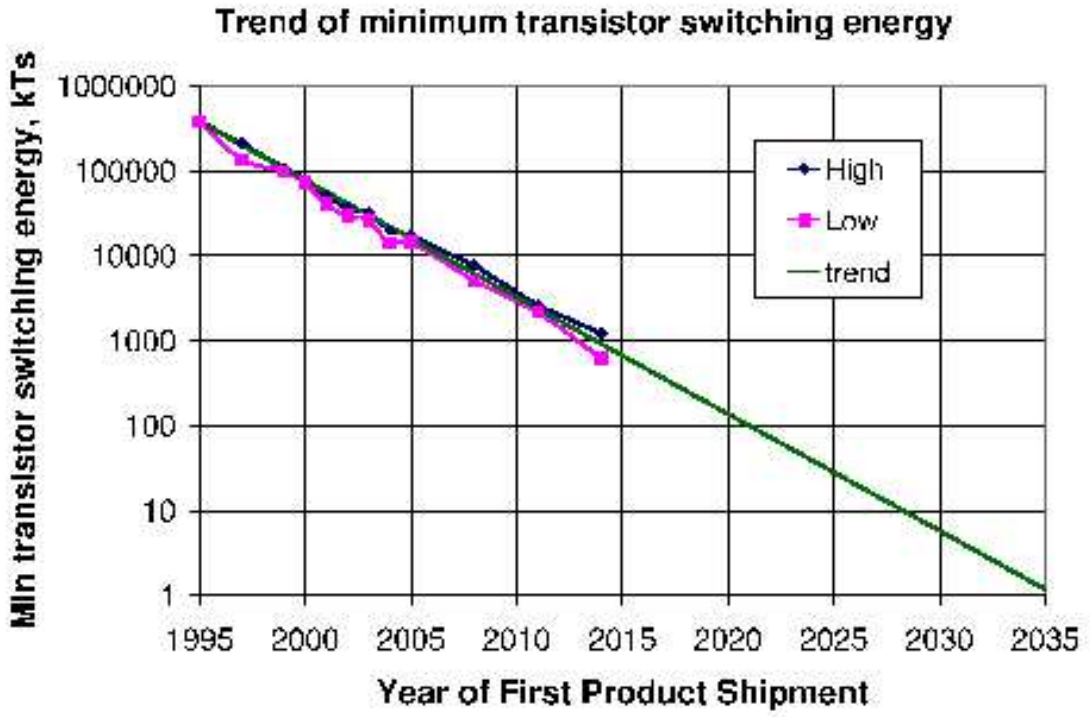

Figure 6: Trendline of minimum 1/2 CV^2 transistor switching energy. Values were calculated using the goals for high-performance and low-power processors in the 1999 International Technology Roadmap for Semiconductors [ITRS99]. Energy is expressed as a multiple of room-temperature kT, which also is the number of nats of infropy associated with that energy. If the trend is followed, thermal noise will begin to become significant in the 2030s, when transistor energies approach small multiples of kT. Shortly after 2035 (if not sooner) this trend will be *forced* to begin leveling off, since a bit of information *requires* at least 0.69 kT to carry it [Mei00]. However, if reversible operations are used, the order-kT bit energies need not be *dissipated* [Lan61], and so the dissipation per reversible bit manipulation might continue decreasing along this curve a while longer.

unobserved, closed system does not decrease as the system evolves unitarily, though it can increase, insofar as we lose track of the system's detailed evolution.

It is interesting to note that current technology is a bit closer to approaching the fundamental limits on energy dissipation for information storage, compared to how far we would have to go to achieve the limits on information density. Current trends would have us reach the limit of ln 2 kT (*i.e.*, 1 bit of physical infropy displaced per bit of digital information irreversibly stored) in only about 35 years (see figure 6). At this time (if not sooner), the performance per unit power of ordinary irreversible computing (which does an irreversible storage operation with every logic-gate operation) will start to level off, at a maximum level of 3.5×10^{22} irreversible bit-operations per second in a 100 W computer that disposes displaced infropy into a room-temperature (300 K) thermal reservoir. This rate is about a million times higher than the maximum rate of bit operations in 30-million-gate, 1 GHz processors in use today. Any possible further improvements in performance-per-power beyond this point would require reversible computing, which we will discuss later.

Communication Limits

Communication is important in computing because it constrains the performance of many parallel algorithms. In his well-known work spawning the field of *information theory*, Claude Shannon derived the maximum information-carrying capacity of a single wave-based communications channel (of given frequency-band width) in the presence of noise. Shannon's limits are widely studied, and are fairly closely approached by coding schemes used in communications today.

However, when considering the ultimate physical limits relevant to computation, we have to go a bit beyond the scope of Shannon's paradigm. We want to know not only the capacity of a *single* channel, but also the maximum bandwidth for communication using *any possible number of channels*, given only area and power constraints.

Interestingly, the limits from the previous section, on information density, directly apply to this. Consider: The difference between information *storage* and information *communication* is, fundamentally, solely a difference in one's frame of reference. Communication is just bit transportation, *i.e.* storage in motion. And storage is just communication across zero distance (but through time).

If one has a limit on information density ρ, and a limit on information propagation velocity v, then this immediately gives a limit of ρv on information *flux density*, that is, bits per unit time per unit area.

Of course, we always *have* a limit on propagation velocity, namely the speed of light c, and so each of the information density limits mentioned earlier directly implies a limit on flux density (though relativistic corrections are needed for speeds approaching c). One can then derive a maximum information bandwidth per unit area (infropy flux), as a function of per-area power density (energy flux).

For example, Smith [Smi95] shows that the maximum entropy flux F_S using photons, given energy flux F_E, is $F_S \leq \frac{4}{3}\sigma_{SB}^{1/4} F_E^{3/4}$, where σ_{SB} is the Stefan-Boltzmann constant $\pi^2 k_B^4/60c^2\hbar^3$. So, for example, a 10-cm square wireless tablet transmitting electromagnetically at a 1 W power level could never communicate at a bit rate of more than 5×10^{20} bits per second (per side), no matter what distribution of frequencies or coding scheme is used, even in the complete absence of noice.

This limit sounds very high at first, but consider that the corresponding bit rate *per square nanometer* is only 50 kbps. For communication among neighboring devices in a computer with nano-scale components, one would like a much higher bandwidth density, perhaps on the order of 10^{11} bps/nm^2, to keep up with the ~100 GHz expected rate of bit-operations in a nanometer-size electronic component. This ~$10^6\times$ higher infropy flux would require a $(10^6)^{4/3} = 10^8\times$ higher power density, that is, on the order of 1 MW/cm^2! Light with this power density would have an effective temperature of about 14,000 K. This is too high to be practical (the computer would melt), so we can rule out light as a practical medium for dense interconnects at the nano-scale, at least until we find a way to build stable structures at such temperatures.

In contrast, notice that if a bit were encoded in more compact particles (atomic or electronic states), rather than electromagnetic field modes (which are too spread-out to be useful at reasonable temperatures), then, given a plausible information density of 1 bit per cubic nanometer, our desired bit-rate could be achieved with a quite reasonable velocity (of atoms or electrons) of only 100 m/s.

Another interesting consideration is the minimum energy *dissipation* (as opposed to energy *transfer*) required for communications. Again, one can look at a communication channel as just a storage element looked at from a different angle, so to speak. If the channel's input bit is in a definite state, then to swap it with the desired information takes no energy [Lan96]. The channel does its thing (ideally, ballistically transporting the energy being communicated, over a definite time span), and the information is then swapped out at the other end—although the receiver needs an empty place to store it. However, if the receiver's storage location is already occupied with a bit that's in the way of your new bit, then you have to pay the energetic price to dispose of the old bit.

Computation Rate Limits

So far we have focused only on limits on information storage and communication. But what about computation itself? What minimum price, in terms of raw physical resources, must we pay for computational operations?

Earlier, we discussed the thermodynamic limit on computational performance of irreversible computations as a function of their power dissipation, due to the need for removal of unwanted (garbage) infropy. However, this limit may not apply to reversible computations. Are there other performance limits that apply to *any* type of computation?

Interestingly, yes: Basic quantum theory can be used to derive a maximum rate at which transitions (such as bit-flips) between distinguishable states can take place [Mar96,Llo00]. This upper bound depends only on the total energy E in the system, and is given by $4E/h$, where $h=2\pi\hbar$ is Planck's constant.

At first, this seems like an absurdly high bound, since the *total* energy presumably includes the rest-mass-energy of the system, which, if the system contains massive particles, is a substantial amount of energy. For example, Lloyd's 1-kg "ultimate laptop" has a mass-energy of 9×10^{16} Joules, and so its maximum rate of operation comes out to be 5×10^{50} state-changes per second!

However, if the system's whole mass-energy is not actively involved in the computation, then apparently it is only that portion of the mass-energy that *is* involved that is relevant in this bound. This gives a much more reasonable level. For example, a hypothetical single-electron device technology in which individual electrons operate at 1V potentials above their ground state could perform state-transitions at a maximum rate of about 1 PHz (10^{15} Hz) per device. Interestingly, as with the speed limit due to energy dissipation, this is only about a factor of a million beyond where we are today.

What about energy limits? We've been talking about Landauer's realization that reversible operations fundamentally need require no energy. It seems that real technologies can approach these predictions, as indicated by Likarev's analysis of the reversible superconducting parametric quantron [Lik82], as well as by the reversible adiabatic CMOS circuits that have been a popular topic of investigation and experimentation (for myself and co-workers, among others) in recent years (*cf.* [You94,Fra98,Vie99,Fra99]). We have designed and built several asymptotically zero-energy (in principle) processors, showing that there is nothing inherently impossible or even especially difficult about building usable computer architectures based on reversible logic.

However, the practicality of these kinds of approaches for breaching sub-*kT* energy levels is not quite firmly established. For example, there is an open research issue of how to provide appropriate synchronization in a reversible processor. Let us explain.

In order to make rapid forward progress through the computation, the machine state needs to evolve nearly ballistically (dominated by forward momentum, rather than by a random walk, along its trajectory in configuration space). A ballistic *asynchronous* (clockless) reversible processor seemingly would be disastrous, since small misalignments in the arrival times of different ballistically-propagating signals would throw off the interactions and lead to chaos [Zur84] (this was a problem with Ed Fredkin's original billiard ball model [Fre82] of ballistic reversible computing). So, apparently, an accurate, synchronous global timing signal needs to be provided in order to keep the logic signals aligned in time. No one has yet proposed a specific mechanism for such a clock generator, accompanied by sufficiently detailed scaling analysis (preferably backed up by a physical simulation) to establish that the entire system can clearly and obviously be scaled to sub-*kT* dissipation levels per logic operation that it drives, while also scaling up cost-effectively to arbitrarily large numbers of processing elements working together. Merkle and Drexler's helical logic proposal [Mer96] is a candidate proof-of-concept which appears to have potential, but its engineering details have probably not been worked out with quite enough thoroughness yet to convince the last skeptics of its feasibility.

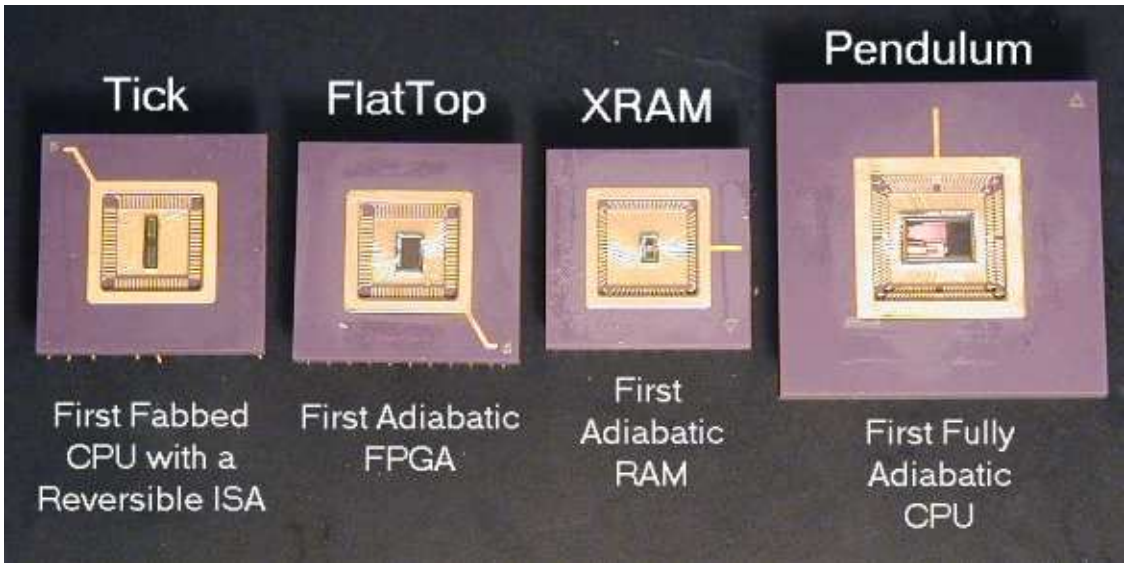

Figure 7. Reversible chips designed at MIT, 1996-99. As graduate students, my co-workers and I designed, outsource-fabricated and tested these four proof-of-concept reversible chips, using the SCRL adiabatic CMOS logic family that was introduced by Knight and Younis in 1993-94 [You94]. Cadence design tools and a 0.5-μm process provided by MOSIS were used. TICK was a benchmark for comparison purposes, an 8-bit, *non*-adiabatic implementation of a reversible instruction set architecture, while PENDULUM was a 12-bit fully-adiabatic implementation with essentially the same ISA, but designed to achieve much lower power [Vie99]. Before PENDULUM, we built the much simpler FLATTOP, a fully-adiabatic programmable array of 400 simple 1-bit processing elements; arrays of these chips could in principle be programmed to simulate arbitrary reversible circuits in a scalable way [Fra98]. XRAM was a small fully-adiabatic static RAM chip.

As an example of the skeptical viewpoint, Smith [Smi99] conjectures a "no free lunch" principle, that any physical mechanism offering sub-*kT* computational operations must necessarily suffer from fatal asymptotic space-time overheads, such as an inability to reuse hardware. Another notable expression of skepticism is that found in Mead and Conway's well-known VLSI textbook [Mea80].

However, despite theese doubts, and despite the lack of a complete, rock-solid proof-of-concept implementation as a counter-example, none of the skeptics have given any rigorous *proof*, or even any very convincing argument (in my opinion), why a hardware-efficient and scalable sub-*kT* logic technology must be *fundamentally* impossible for *any* technology, as opposed to merely being unattainable by a variety of specific mechanisms.

So, to the best of my knowledge, at the moment it is *still* technically an open question whether arbitrarily low-energy computation (with hardware reuse) is truly permitted, or not. Finding the definitive answer to this crucial question (whether it be yes or no) is a key goal of my own research.

Interestingly, even if efficient physically-reversible computing turns out to be permitted in principle, it is probably not quite as good as it sounds at first for *all* applications, because of the *algorithmic* overheads that appear to be associated with reversible computing in general [Fra00]. Furthermore, in any *fixed* device technology, one can show that there will be a maximum degree of reversibility that will be beneficial, so that any given technology's dissipation per operation, even if much less than *kT*, is not *arbitrarily* small.

Cosmological Limits of Computing

So far, everything we've discussed has concerned the limits on computational *efficiency*, namely how many bits we can store and how many ops we can perform *per unit* of some resource such as space, time, matter, or energy. However, in considering the *overall* limits on computation in the very long term, there is also the question of limitations on the resources *themselves* to consider.

Unless one is rather a pessimist, there seems no strong reason why our civilization should not be able to survive, prosper, and continue advancing technologically over the coming centuries. As global wealth increases and technology becomes more affordable, we can expect that eventually a boundary will be crossed where interstellar (perhaps robotic) probes become affordable, though constrained to sub-light speeds. Sober analyses of the subject strongly suggest that there is no fundamental physical reason that a civilization such as ours could not, within a few tens of millions of years, spread throughout the galaxy, then begin to launch probes to other galaxies. If we are truly determined to continue advancing available computing power indefinitely, then perhaps we must plan to eventually re-engineer the stars and galaxies themselves into configurations optimized to provide us maximum resources for computing purposes.

How much of the universe could we eventually grab and control? Estimates vary, and much depends on the expansion rate of the universe and its possible acceleration, issues which are not quite settled yet. But, suppose that we could eventually gain control of a sizable number of superclusters' worth of mass-energy before the rest of the galaxies disappear over the redshift horizon. What could we do with this material?

Obviously, we could do a *lot* of computing. But, perhaps the most interesting question is: Is the total number of computational operations we can do (given a finite supply of mass-energy, and an ever-expanding cosmos) finite or infinite?

One might expect at first that of course it must be finite, due to the finite energy supply, but keep in mind that if reversible computing techniques can be applied to an arbitrarily high degree of perfection, then, in principle, even a finite supply of free energy could suffice to carry out an infinite number of ever-more-nearly-reversible operations.

But, even with an infinite number of operations available, doesn't one eventually run out of new things to compute? If our piece of the universe has a finite infropy, then it has only a finite number of distinguishable states, before we would eventually have to just start repeating states previously encountered (which would not be very interesting). However, recall that all the known fundamental infropy limits increase with increasing spatial extent, even when energy is constant. So, if we can spread our available resources over larger and larger expanses of space, we can (in principle) encode ever more information using a given supply of energy. However, it is not clear whether any kind of meaningful structure could be maintained and contained in a controlled state, as it becomes ever sparser.

In any event, there is starting to be a serious debate among scientists on this important question of cosmological limits [Dys79,Kra00,Dys01], although at the moment it seems unresolved. Apparently, not all of the potentially relevant issues (such as reversible computing) have been thoroughly taken into account yet. Obviously, these cosmic concerns surely seem faint and distant to most researchers today. But, it is still interesting to see that some topics that we can begin exploring today, concerning such esoteric subjects as reversible computing and quantum limits on entropy densities, could conceivably, in a far distant future, allow our descendants to make the difference between our universe being one of eternal life and unbounded variety, or one of death and permanent stagnation. It's the ultimate engineering challenge!

Conclusion

All computer users, including computational scientists & engineers, naturally hope that the trend of increasing affordability of computing power will take us as far as possible. However, our best available knowledge of physics strongly indicates that some ultimate limits *do* exist, and give us, at least, loose upper bounds on what might be achieved.

Interestingly, one of the most imminent of the fundamental limits appears to be the limit on the energy dissipation of irreversible computation, but this particular limit may possibly be circumvented through the use of reversible computing techniques. Although reversible computing has made impressive progress, whether this "fix" can ultimately work out in a scalable and cost-efficient way remains to this day an open question, one that is the subject of active research by myself and others.

In my future work, I am planning to apply methods of computational physics to model and simulate various candidate reversible computing systems, taking all the relevant physical considerations into account, until either a complete and detailed proof-of-concept model of a realistic, cost-efficient, and scalable sub-kT computing system is developed, or it becomes clear how to construct a rigorous and general proof that no mechanism having all the desired properties can physically exist.

In any event, I hope that the present article will help to inspire scientists & engineers in many fields to devote increased attention to finding ways to meet the incredible challenges facing the future of computing, as it approaches the many limits found at the atomic scale. These limits are now close enough to fall within the career horizons of people starting out today: For example, given present rates of improvement, computing will hit the *kT* thermodynamic brick wall before today's 30-year-old Ph.D. graduates will retire.

Given the nearness of some of these limits, basic science and engineering research done today that is likely to impact how closely and how quickly we can approach these ultimate limits has, in business terminology, an enormous net present value for society.

Finally, studies of cosmological limits to computing show us that fundamental discoveries about computing limits that are being made today may potentially even bear on the long-term prospects for life in the universe. Could any conceivable area of scientific inquiry be any grander or more exciting?

References [not yet sorted]

[Moo65] Gordon E. Moore, "Cramming more components onto integrated circuits," *Electronics*, April 19, 1965, pp. 114-117.

[Moo75] G. E. Moore, "Progress in digital integrated electronics," *Technical Digest 1975 International Electron Devices Meeting*, IEEE, 1975, pp. 11-13.

[Moo95] G. E. Moore, "Lithography and the Future of Moore's Law," *Optical/Laser Microlithography VIII: Proceedings of the SPIE*, **2440**, 1995, pp. 2-17.

[Moo97] G. E. Moore, "An Update on Moore's Law," Intel Developer Forum Keynote Speech, Sep. 30, 1997. http://developer.intel.com/pressroom/archive/speeches/gem93097.htm.

[Kur99] Ray Kurzweil, *The Age of Spiritual Machines: When Computers Exceed Human Intelligence*, Penguin Books, 1999.

[Deu99] David Deutsch and Patrick Hayden, "Information Flow in Entangled Quantum Systems," *Proceedings of the Royal Society* **A456**, 2000, pp. 1759-1774. http://arxiv.org/abs/quant-ph/9906007.

[Ben82] Charles H. Bennett, "The Thermodynamics of Computation—a Review," *International Journal of Theoretical Physics* **21**(12):905-940, 1982.

[Ben88] C. H. Bennett, "Notes on the history of reversible computation," *IBM Journal of Research and Development* **32**(1):16-23, Jan. 1988. Reprinted in [Lef90], ch. 4, pp. 281-288.

[Lef90] Harvey S. Leff and Andrew F. Rex, eds., *Maxwell's Demon: Entropy, Information, Computing*, Princeton University Press, 1990.

[Zur01] Wojciech Hubert Zurek, "Decoherence, Einselection, and the Quantum Origins of the Classical," http://arxiv.org/abs/quant-ph/0105127.

[Smi95] Warren D. Smith, "Fundamental physical limits on computation," http://external.nj.nec.com/homepages/wds/fundphys.ps.

[Llo00] Seth Lloyd, "Ultimate physical limits to computation," *Nature* **406**:1047-1054, 31 Aug. 2000.

[Bek81] Jacob D. Bekenstein, "Universal upper bound on the entropy-to-energy ratio for bounded systems," *Physical Review D*, **23**(2):287-298, 15 Jan. 1981.

[Lan61] Rolf Landauer, "Irreversibility and Heat Generation in the Computing Process," *IBM Journal of Research and Development* **5**:183-191, 1961. Reprinted in [Lef90], ch. 4, pp. 188-196.

[Ben73] C. H. Bennett, "Logical Reversibility of Computation," *IBM Journal of Research and Development* **17**(6):525-532, 1973.

[von66] John von Neumann, *Theory of Self-Reproducing Automata*, University of Illinois Press, 1966.

[ITRS99] Semiconductor Industry Association, "International Technology Roadmap for Semiconductors," 1999 Edition, http://public.itrs.net/files/1999_SIA_Roadmap/Home.htm.

[Mei00] James D. Meindl and Jeffrey A. Davis, "*The* Fundamental Limit on Binary Switching Energy for Terascale Integration (TSI)," *IEEE Journal of Solid State Circuits* **35**(10):1515-1516, Oct. 2000.

[Lan96] Rolf Landauer, "Minimal energy requirements in communication," *Science* **272**(5270):1914-1918, Jun. 28, 1996.

[Mar96] Norman Margolus and Lev B. Levitin, "The maximum speed of dynamical evolution," in T. Toffoli *et al.*, eds., *PhysComp96 (Proceedings of the Fourth Workshop of Physics and Computation)*, New England Complex Systems Institute, 1996.

[Lik82] K. K. Likharev, "Classical and Quantum Limitations on Energy Consumption in Computation," *International Journal of Theoretical Physics* **21**(3/4):311-326, 1982.

[You94] Saed G. Younis and Thomas F. Knight, Jr., "Asymptotically zero energy split-level charge recovery logic," *International Workshop on Low Power Design*, 1994, pp. 177-182.

[Fra98] Michael P. Frank *et al.*, "A scalable reversible computer in silicon," in Calude *et al.*, eds., *Unconventional Models of Computation*, Springer, 1998, pp. 183-200.

[Vie99] Carlin J. Vieri, *Reversible Computer Engineering and Architecture*, Ph.D. thesis, MIT, 1999.

[Fra99] Michael P. Frank, *Reversibility for Efficient Computing*, Ph.D. thesis, MIT, 1999. http://www.cise.ufl.edu/~mpf/rc/thesis/phdthesis.html.

[Zur84] W. H. Zurek, "Reversibility and Stability of Information Processing Systems," *Physical Review Letters* **53**(4):391-394, 23 Jul. 1984.

[Fre82] Edward F. Fredkin and T. Toffoli, "Conservative Logic," *International Journal of Theoretical Physics* **21**(3/4):219-253, 1982.

[Mer96] Ralph C. Merkle and K. Eric Drexler, "Helical Logic," *Nanotechnology* **7**(4):325-339, 1996.

[Smi99] Warren D. Smith, "Classical reversible computation with zero Lyapunov exponent," http://www.neci.nec.com/homepages/wds/pu-fred-lyap.ps, Feb. 25, 1999.

[Mea80] Carver Mead and Lynn Conway, *Introduction to VLSI Systems*, Addison-Wesley, 1980.

[Fra00] Michael P. Frank and M. Josephine Ammer, "Relativized Separation of Reversible and Irreversible Space-Time Complexity Classes," submitted to

Information and Computation. http://www.cise.ufl.edu/~mpf/rc/memos/M06_oracle.html.

[Dys79] Freeman J. Dyson, "Time without end: Physics and biology in an open universe," *Reviews of Modern Physics* **51**(3):447-460, Jul. 1979.

[Kra00] Lawrence M. Krauss and Glenn D. Starkman, "Life, The Universe, and Nothing: Life and Death in an Ever-Expanding Universe," *The Astrophysical Journal* **531**(1):22-30, 2000.

[Dys01] Freeman Dyson, "Is Life Analog or Digital?", *Edge* **82**, Mar. 13, 2001, http://www.edge.org/documents/archive/edge82.html.

[Deu97] David Deutsch, *The Fabric of Reality*, Penguin Books, 1997.