B Thermodynamics of computation

"Computers may be thought of as engines for transforming free energy into waste heat and mathematical work." — Bennett (1982)

B.1 Von Neumann-Landauer Principle

B.1.a Information and entropy

We will begin with a quick introduction or review of the entropy concept; we will look at it in more detail soon (Sec. B.2). The *information content* of a signal (message) measures our "surprise," that is, how unlikely it is. $I(s) = -\log_b \mathcal{P}\{s\}$, where $\mathcal{P}\{s\}$ is the probability of signal s. We take logs so that the information content of independent signals is additive. We can use any base, and for $b = 2, e$, and 10, the corresponding units are *bits* (or *shannons*), *nats*, and *dits* (also, *hartleys, bans*). Therefore, if a signal has a 50% probability, then it conveys one bit of information. The *entropy* of a distribution of signals is their average information content:

$$H(S) = \mathcal{E}\{I(s) \mid s \in S\} = \sum_{s \in S} \mathcal{P}\{s\} I(s) = -\sum_{s \in S} \mathcal{P}\{s\} \log \mathcal{P}\{s\}.$$

Or more briefly, $H = -\sum_k p_k \log p_k$. For example, if $\mathcal{P}\{1\} = 1/16$ and $\mathcal{P}\{0\} = 15/16$, we will receive, on the average, $H = 0.3$ bits of information.

According to a well-known story, Shannon was trying to decide what to call this quantity and had considered both "information" and "uncertainty." Because it has the same mathematical form as statistical entropy in physics, von Neumann suggested he call it "entropy," because "nobody knows what entropy really is, so in a debate you will always have the advantage."[10] (This is one version of the quote.) I will call it *information entropy* when I need to distinguish it from the thermodynamical concept.

An important special case is the entropy of a uniform distribution. If there are N signals that are all equally likely, then $H = \log N$. Therefore, if we have eight equally likely possibilities, the entropy is $H = \lg 8 = 3$ bits.[11]

[10] `https://en.wikipedia.org/wiki/History_of_entropy` (accessed 2012-08-24). Ralph Hartley laid the foundations of information theory in 1928, on which Claude Shannon built his information theory in 1948.

[11] I use the notations $\lg x = \log_2 x$ and $\ln x = \log_e x$.

A uniform distribution maximizes the entropy (and minimizes the ability to guess).

In computing, we are often concerned with the *state* of the computation, which is realized by the state of a physical system. Consider a physical system with three *degrees of freedom* (*DoF*), each with 1024 possible values. There are $N = 1024^3 = 2^{30}$ possible states, each describable by three 10-bit integers. If we don't care about the distance between states (i.e., distance on each axis), then states can be specified equally well by six 5-bit numbers or one 30-bit number, etc. (or ten digits, since $30 \log_{10} 2 \approx 9.03$ digits). Any scheme that allows us to identify all 2^{30} states will do. We can say that there are 30 *binary degrees of freedom*.

In computing we often have to deal with things that grow exponentially or are exponentially large (due to combinatorial explosion), such as solution spaces. (For example, NP problems are characterized by the apparent necessity to search a space that grows exponentially with problem size.) In such cases, we are often most concerned with the *exponents* and how they relate. Therefore it is convenient to deal with their logarithms (i.e., with logarithmic quantities). The logarithm represents, in a scale-independent way, the degrees of freedom generating the space.

Different logarithm bases amount to different units of measurement for logarithmic quantities (such as information and entropy). As with other quantities, we can leave the units unspecified, so long as we do so consistently. I will use the notation "$\log x$" for an *unspecific logarithm*, that is, a logarithm with an unspecified base.[12] When I mean a specific base, I will write $\ln x$, $\lg x$, $\log_{10} x$, etc. Logarithms in specific bases can be defined in terms of unspecific logarithms as follows: $\lg x = \log x / \log 2$, $\ln x = \log x / \log e$, etc. (The units can be defined $\text{bit} = \log 2$, $\text{nat} = \log e$, $\text{dit} = \log 10$, etc.)

B.1.b The von Neumann-Landauer bound

Thermodynamic entropy is unknown information residing in the physical state. The macroscopic thermodynamic entropy S is related to microscopic information entropy H by Boltzmann's constant, which expresses the entropy in thermodynamical units (energy over temperature). If H is measured in nats, then $S = k_{\mathrm{B}} H = k_{\mathrm{B}} \ln N$, for N equally likely states. When using

[12] Frank (2005a) provides a formal definition for the *indefinite logarithm*. I am using the idea less formally, an "unspecific logarithm," whose base is not mentioned. This is a compromise between Frank's concept and familiar notation; we'll see how well it works!

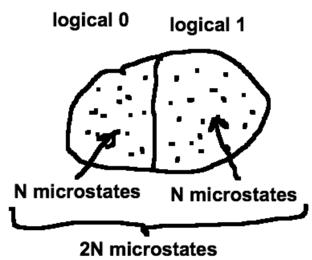

Figure II.3: Physical microstates representing logical states. Setting the binary device decreases the entropy: $\Delta H = \lg N - \lg(2N) = -1$ bit. That is, we have one bit of information about its microstate.

unspecific logarithms, I will drop the "B" subscript: $S = kH = k \log N$. The physical dimensions of entropy are usually expressed as energy over temperature (e.g., joules per kelvin), but the dimensions of temperature are energy per degree of freedom (measured logarithmically), so the fundamental dimension of entropy is degrees of freedom, as we would expect. (There are technical details that I am skipping.)

Consider a macroscopic system composed of many microscopic parts (e.g., a fluid composed of many molecules). In general a very large number of *microstates* (or *microconfigurations*) — such as positions and momentums of molecules — will correspond to a given *macrostate* (or *macroconfiguration*) — such as a combination of pressure and termperature. For example, with $m = 10^{20}$ particles we have $6m$ degrees of freedom, and a $6m$-dimensional *phase space* of its possible microstates.

Now suppose we partition the microstates of a system into two macroscopically distinguishable macrostates, one representing 0 and the other representing 1. For example, whether the electrons are on one plate of a capacitor or the other could determine whether a 0 or 1 bit is stored on it. Next suppose N microconfigurations correspond to each macroconfiguration (Fig. II.3). This could be all the positions, velocities, and spins of the many electrons, which we don't care about and cannot control individually. If we confine the system to one half of its microstate space in order to represent a 0 or a 1, then the entropy (average uncertainty in identifying the microstate) will decrease by one bit. We don't know the exact microstate, but at least

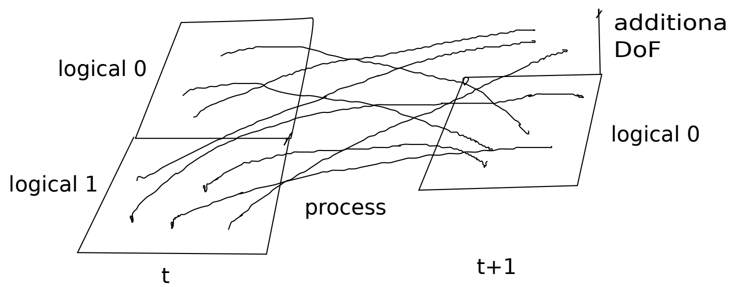

Figure II.4: Thermodynamics of erasing a bit. On the left is the initial state (time t), which may be logical 0 or logical 1; on the right (time $t+1$) the binary device has been set to logical 0. In each case there are N microstates representing each prior state, so a total of $2N$ microstates. However, at time $t+1$ the system must be in one of N posterior microstates. Therefore N of the microstate trajectories must exit the defined region of phase space by expanding into additional, uncontrolled degrees of freedom. Therefore entropy of the environment must increase by at least $\Delta S = k \log(2N) - k \log N = k \log 2$. We lose track of this information because it passes into uncontrolled degrees of freedom.

---

we know which half of the state-space it is in.

In general, in physically realizing a computation we distinguish *information-bearing degrees of freedom* (IBDF), which we control and use for computation, from *non-information-bearing degrees of freedom* (NIBDF), which we do not control and are irrelevant to the computation (Bennett, 2003).

Consider the process of erasing or clearing a bit (i.e., setting it to 0, no matter what its previous state): we are losing one bit of physical information. The physical information still exists, but we have lost track of it; it has been *thermalized*.

Suppose we have N physical microstates per logical macrostate (logical 0 or logical 1). Before the bit is erased it can be in one of $2N$ possible microstates, but there are only N microstates representing its final state. The laws of physics are reversible,[13] so they cannot lose any information. Since physical information can't be destroyed, it must go into NIBDF (e.g., the environment or thermal motion of the atoms) (Fig. II.4). The trajectories

[13]This is true in both classical and quantum physics. In the latter case, we cannot have $2N$ quantum states mapping reversibly into only N quantum states.

have to expand into other degrees of freedom (NIBDF) to maintain the phase space volume.

The information lost, or dissipated into NIBDF (typically as heat), is $\Delta S = k\log(2N) - k\log N = k\log 2$. (In physical units this is $\Delta S = k_\mathrm{B}\ln 2 \approx 10^{-23}\mathrm{J/K}$.) Therefore the increase of energy in the device's environment is $\Delta Q = \Delta S \times T_\mathrm{env} = k_\mathrm{B}T_\mathrm{env}\ln 2 \approx 0.7kT_\mathrm{env}$. At $T_\mathrm{env} = 300\mathrm{K}$, $k_\mathrm{B}T_\mathrm{env}\ln 2 \approx 18$ meV $\approx 3\times 10^{-9}$ pJ $= 3$ zJ. We will see that this is the minimum energy dissipation for any irreversible operation (such as erasing a bit); it is called the *von Neumann–Landauer (VNL) bound* (or sometimes simply the *Landauer bound*). Von Neumann suggested the idea in 1949, but it was published first by Rolf Landauer (IBM) in 1961.[14] Recall that for reliable operation we need minimum logic levels around $40k_\mathrm{B}T_\mathrm{env}$ to $100k_\mathrm{B}T_\mathrm{env}$, which is two orders of magnitude above the von Neumann–Landauer limit of $0.7k_\mathrm{B}T_\mathrm{env}$. "From a technological perspective, energy dissipation per logic operation in present-day silicon-based digital circuits is about a factor of 1,000 greater than the ultimate Landauer limit, but is predicted to quickly attain it within the next couple of decades" (Berut et al., 2012). That is, current circuits have signal levels of about 18 eV, which we may compare to the VNL, 18 meV.

In research reported in 2012 Berut et al. (2012) confirmed experimentally the Landauer Principle and showed that it is the erasure that dissipates energy. They trapped a 2μ silica ball in either of two laser traps, representing logical 0 and logical 1. For storage, the potential barrier was greater than $8k_\mathrm{B}T$, and for erasure, the barrier was lowered to $2.2k_\mathrm{B}T$ by decreasing the power of the lasers and by tilting the device to put it into the logical 0 state (see Fig. II.5). At these small sizes, heat is a stochastic property, so the dissipated heat was computed by averaging the trajectory of the particle over multiple trials:

$$\langle Q\rangle = \left\langle -\int_0^\tau \dot{x}(t)\frac{\partial U(x,t)}{\partial x}\mathrm{d}t\right\rangle_x .$$

(The angle brackets means "average value.") Complete erasure results in the ball being in the logical 0 state; incomplete erasure results in it being in the logical 0 state with probability p. They established a generalized Landauer bound in which the dissipated heat depends on the completeness of erasure:

$$\langle Q\rangle^p_\mathrm{Landauer} = kT[\log 2 + p\log p + (1-p)\log(1-p)] = kT[\log 2 - H(p, 1-p)].$$

[14] See Landauer (1961), reprinted in Leff & Rex (1990) and Leff & Rex (2003), which include a number of other papers analyzing the VNL principle.

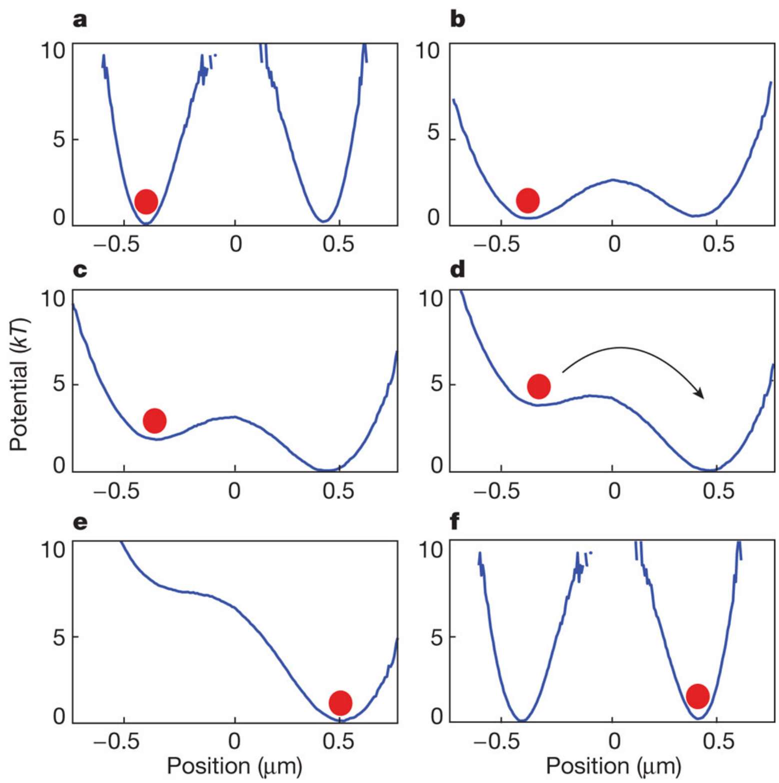

Figure II.5: Erasing a bit by changing potential barrier. (Figure from Berut et al. (2012).)

Therefore, for $p = 1$, heat dissipation is $kT \log 2$, but for $p = 1/2$ (ineffective erasure) no heat needs to be dissipated. Notice how the dissipated energy depends on the entropy $H(p, 1-p)$ of the final macrostate.

B.1.c Irreversible operations

Suppose the phase space is divided into M macrostates of size $N_1, N_2, \ldots, N_M$, where $N = N_1 + N_2 + \cdots + N_M$. Let p_{ij} be the probability the device is in microstate i of macrostate j. The total entropy is

$$S = -k \sum_{ij} p_{ij} \log p_{ij}. \tag{II.3}$$

We can separate this into the *macroscopic entropy* associated with the macrostates (IBDF) and the *microscopic entropy* associated with the microstates (NIBDF). Now let $P_j = \sum_{i=1}^{N_j} p_{ij}$ be the probability of being in macrostate j. Then Eq. II.3 can be rearranged (Exer. II.1):

$$S = -k \sum_j P_j \log P_j - k \sum_j P_j \sum_{i=1}^{N_j} \frac{p_{ij}}{P_j} \log \frac{p_{ij}}{P_j} = S_\mathrm{i} + S_\mathrm{h}. \tag{II.4}$$

The first term is the macrostate entropy (IBDF):

$$S_\mathrm{i} = -k \sum_j P_j \log P_j.$$

The second is the microstate entropy (NIBDF):

$$S_\mathrm{h} = -k \sum_j P_j \sum_{i=1}^{N_j} \frac{p_{ij}}{P_j} \log \frac{p_{ij}}{P_j}.$$

Note that the ratios in the inner summation are essentially conditional probabilities, and that the inner summation is the conditional entropy given that you are in macrostate j.

When we erase a bit, we go from a maximum S_i of 1 bit (if 0 and 1 are equally likely), to 0 bits (since there is no uncertainty). Thus we lose one bit of information, and the macrostate entropy decreases $\Delta S_\mathrm{i} = -k \log 2$. (The actual entropy decrease can be less than 1 bit if the 0 and 1 are not equally likely initial states.) Since according to the Second Law of Thermodynamics

$\Delta S \geq 0$, we have a corresponding minimum increase in microstate entropy, $\Delta S_{\rm h} \geq k \log 2$. Typically this is dissipated as heat, $\Delta Q \geq kT \log 2$. The information becomes inaccessible and unusable.

The standard logic gates (AND, OR, XOR, NAND, etc.) have two input bits and one output bit. Therefore the output will have lower entropy than the input, and so these gates must dissipate at least 1 bit of entropy, $kT \log 2$ energy. Consider, for example, AND. If the four inputs 00, 01, 10, 11, are equally likely, then the input entropy is $H_{\rm i} = 2$ bits. However the output entropy will be $H_{\rm o} = -(1/4)\lg(1/4) - (3/4)\lg(3/4) = 0.811$, so the entropy lost is 1.189 bits. To compute the dissipated energy in Joules, multiply by $\ln 2$ to convert bits to nats (or shannons to hartleys):

$$\Delta Q = T\Delta S \geq -Tk_{\rm B}(H_{\rm o} - H_{\rm i}) \ln 2 \approx 1.189 k_{\rm B} T \ln 2 \approx 0.83 k_{\rm B} T.$$

For each gate, we can express $H_{\rm o}$ in terms of the probabilities of the inputs and compute the decrease from $H_{\rm i}$ (exercise). If the inputs are not equally likely, then the input entropy will be less than 2 bits, but we will still have $H_{\rm i} > H_{\rm o}$ and energy will be dissipated. (Except in a trivial, uninteresting case. What is it?)

More generally, any irreversible operation (non-invertible function) will lose information, which has to be dissipated into the environment. That is, it is irreversible because it loses information, and every time we lose information, we lose energy. If the function is not one-to-one (injective), then at least two inputs map to the same output, and so information about the inputs is lost. For example, changing a bit, that is, overwriting a bit with another bit, is a fundamental irreversible operation, subject to the VNL limit. Therefore, the assignment operation is bound by it. Also, when two control paths join, we forget where we came from, and so again we must dissipate at least a bit's worth of entropy (Bennett, 2003). These considerations suggest that reversible operations might not be subject to the VNL limit, and this is in fact the case, as we will see.

The preceding observations have important connections with the problem of Maxwell's Demon and its resolution. Briefly, the demon has to reset its mechanism after each measurement in preparation for the next measurement, and this dissipates at least $kT \log 2$ energy into the heat bath for each decision that it makes. That is, the demon must "pay" for any information that it acquires. Therefore, the demon cannot do useful work. Further discussion is outside the scope of this book, so if you are interested, please see Leff &

Rex (2003) and Leff & Rex (1990) (which have a large intersection), in which many of the papers on the topic are collected.

B.2 Mechanical and thermal modes

We need to understand in more detail the reason for the increase of entropy and its relation to reversibility and irreversibility.[15] We can classify systems according to their size and completeness of specification:

specification:	complete	incomplete	
size:	~ 1	~ 100	$\sim 10^{23}$
laws:	dynamical	statistical	thermodynamical
reversible:	yes	no	no

Dynamical systems have a relatively small number of particles or degrees of freedom and can be completely specified. For example, we may have 6 degrees of freedom for each particle (x, y, z, p_x, p_y, p_z). We can prepare an *individual* dynamical system in an initial state and expect that it will behave according to the dynamical laws that describe it. Think of billiard balls or pucks on a frictionless surface, or electrons moving through an electric or magnetic field. So far as we know, the laws of physics at this level (either classical or quantum) are reversible.

If there are a large number of particles with many degrees of freedom (several orders of magnitude), then it is impractical to specify the system completely. Moreover, small errors in the initial state will have a larger effect, due to complex interaction of the particles. Also, there are small effects from interaction with the environment. Therefore we must resort to *statistical laws* and we have a *statistical system*. If we can't *manage* all the degrees of freedom, then we average over those that we can't manage. Statistical laws don't tell us how an individual system will behave (there are too many sources of variability), but they tell us how *ensembles* of similar systems (or preparations) behave. We can talk about the average behavior of such systems, but we also have to consider the variance, because unlikely outcomes are not impossible. For example, tossing 10 coins has a probability of 1/1024 of turning up all heads; this is small, but not negligible. Statistical laws are in general irreversible (because there are many ways to get to the same state).

[15] This section is based primarily on Edward Fredkin and Tommaso Toffoli's "Conservative logic" (Fredkin & Toffoli, 1982).

Finally we consider *thermodynamical systems*. Macroscopic systems have a very large number of particles ($\sim 10^{23}$) and a correspondingly large number of degrees of freedom. We call these "Avogadro scale" numbers. It is important to grasp how truly enormous these numbers are; in comparison (Tong, 2012, p. 37): the number of grains of sand on all beaches $\approx 10^{18}$, the number of stars in our galaxy $\approx 10^{11}$, and the number of stars in the visible universe $\approx 10^{22}$, but the number of water molecules in a cup of tea $\approx 10^{23}$. Obviously such systems cannot be completely specified (we cannot describe the initial state and trajectory of every atom). Indeed, because information is physical, it is *physically impossible* to "know" (i.e., to represent physically) the physical state of a macroscopic system (i.e., to use the macrostates of one system to represent the microstates of another macroscopic system).

We can derive statistical laws for thermodynamical systems, but in these cases most macrostates become so improbable that they are virtually impossible (for example, the cream unmixing from your coffee). The central limit theorem shows that the variance decreases with n: By the law of large numbers (specifically, Bernoulli's Theorem), variance in the relative number of successes is $\sigma^2 = p(1-p)/n$, that is, $\sigma^2 = 1/4n$ for $p = 1/2$. Therefore the standard deviation is $\sigma = 1/(2n^{1/2})$. For example, for $n = 10^{22}$, $\sigma = 5 \times 10^{-10}$. Also, 99.99% of the probability density is within $4\sigma = 2 \times 10^{-9}$ (see Thomas, *Intro. Appl. Prob. & Rand. Proc.*, p. 111). The probability of deviating more than ϵ from the mean decreases exponentially with n: $\frac{1}{6}\exp(-\epsilon^2 n/2) + \frac{1}{2}\exp(-2\epsilon^2 n/3)$.

In the thermodynamic limit, the *likely* is *inevitable*, and the *unlikely* is *impossible*. In these cases, *thermodynamical laws* describe the virtually deterministic (but *irreversible*) dynamics of the system.

Sometimes in a macroscopic system we can separate a small number of *mechanical modes* (DoF) from the *thermal modes*. "Mechanical" here includes "electric, magnetic, chemical, etc. degrees of freedom." The mechanical modes are strongly coupled to each other but weakly coupled to the thermal modes. For example, in a rigid body (e.g., a bullet, a billiard ball) the mechanical modes are the positions and momentums of the particles in the body. Thus the mechanical modes can be treated exactly or approximately independently of the thermal modes. In the ideal case the mechanical modes are completely decoupled from the thermal modes, and so the mechanical modes can be treated as an isolated (and reversible) dynamical system. The energy of the mechanical modes (once initialized) is independent of the energy ($\sim kT$) of the thermal modes. The mechanical modes are *conservative*;

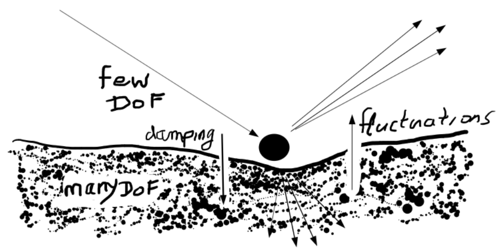

Figure II.6: Complementary relation of damping and fluctuations. The macroscopic ball has few mechanical degrees of (e.g., 6), but during the collision there is an interaction of the enormously many degrees of freedom of the microstates of the ball and those of the wall.

---

they don't dissipate any energy. (This is what we have with *elastic* collisions.) This is the approach of reversible computing.

Suppose we *want* irreversible mechanical modes, e.g., for implementing irreversible logic. The underlying physics is reversible, and so the information lost by the mechanical modes cannot simply disappear; it must be transferred to the thermal modes. This is *damping*: Information in the mechanical modes, where it is accessible and usable, is transferred to the thermal modes, where it is inaccessible and unusable. This is the *thermalization* of information, the transfer of physical information from accessible DoF to inaccessible DoF. But the interaction is bidirectional, so *noise* (uncontrolled DoF) will flow from the thermal modes back to the mechanical modes, making the system nondeterministic.

As Feynman said, "If we know where the *damping* comes from, it turns out that that is also the source of the *fluctuations*" [Feynman, 1963]. Think of a bullet ricocheting off a flexible wall filled with sand. It dissipates energy into the sand and also acquires noise in its trajectory (see Fig. II.6). To avoid nondeterminacy, the information may be encoded redundantly so that the noise can be filtered out. For example, the signal can be encoded in multiple mechanical modes, over which we take a majority vote or an average. Or the signal can be encoded with energy much greater than any one of the thermal modes, $E \gg kT$, to bias the energy flow from mechanical to thermal (preferring dissipation over noise). Free energy must be used to refresh the mechanical modes, and heat must be flushed from the thermal modes. "[I]mperfect knowledge of the dynamical laws leads to uncertainties in the behavior of a system comparable to those arising from imperfect knowledge of its initial conditions... Thus, the same regenerative processes which help overcome thermal noise also permit reliable operation in spite of substantial fabrication tolerances." (Fredkin & Toffoli, 1982)

Damped mechanisms have proved to be very successful, but they are inherently inefficient.

> In a damped circuit, the rate of heat generation is proportional to the number of computing elements, and thus approximately to the useful volume; on the other hand, the rate of heat removal is only proportional to the free *surface* of the circuit. As a consequence, computing circuits using damped mechanisms can grow arbitrarily large in two dimensions only, thus precluding the much tighter packing that would be possible in three dimensions. (Fredkin & Toffoli, 1982)

In an extreme case (force of impact > binding forces), a signal's interacting with the environment might cause it to lose its coherence (the correlation of its constituent DoFs, such as the correlation between the positions and momenta of its particles). The information implicit in the mechanical modes is lost into the thermal modes.