Chapter II

Physics of Computation

A Reversible computation

These lectures are based on Michael P. Frank, "Introduction to Reversible Computing: Motivation, Progress, and Challenges" (CF05, May 46, 2005, Ischia, Italy). (Quotations in this section are from this paper unless otherwise specified.)

A.1 Problem of energy dissipation

¶1. **Energy efficiency:**

$$R = \frac{N_{\text{ops}}}{t} = \frac{N_{\text{ops}}}{E_{\text{diss}}} \times \frac{E_{\text{diss}}}{t} = F_{\text{E}} \times P_{\text{diss}} \tag{II.1}$$

"where R = performance,
N_{ops} = number of useful operations performed during a job,
t = total elapsed time to perform the job,
E_{diss} = energy dissipated during the job,
$F_{\text{E}} = N_{\text{ops}}/E_{\text{diss}}$ = energy efficiency,
$P_{\text{diss}} = E_{\text{diss}}/t$ = average power dissipation during the job."
The key parameter is F_{E}.

¶2. **Energy efficiency of FET:** "Energy efficiency for the lowest-level ops has been roughly given by $F_{\rm E} \approx (1 \text{ op})/(\frac{1}{2}CV^2)$, where C is the typical capacitance of a node in a logic circuit, and V is the typical voltage swing between logic levels."
(The charge stored in a capacitor is $Q = CV$ and the energy stored in it is $\frac{1}{2}CV^2$.)

¶3. "This is because voltage-coded logic signals have an energy of $E_{\rm sig} = \frac{1}{2}CV^2$, and this energy gets dissipated whenever the node voltage is changed by the usual irreversible FET-based mechanisms in modern CMOS technology."

¶4. Moore's law is a result of "an exponential decline in C over this same period [1985–2005] (in proportion to shrinking transistor lengths), together with an additional factor of $\sim 25\times$ coming from a reduction of the typical logic voltage V from 5V (TTL) to around 1V today."
See Fig. II.1.

¶5. Neither the transistor lengths nor the voltage can be reduced much more.

¶6. **Thermal noise:** "[A]s soon as the signal energy $E_{\rm sig} = \frac{1}{2}CV^2$ becomes small in comparison with the thermal energy $E_T = k_{\rm B}T$, (where $k_{\rm B}$ is Boltzmann's constant and T is the temperature), digital devices can no longer function reliably, due to problems with thermal noise."

¶7. **Room-temperature thermal energy:** $k_{\rm B} \approx 8.6 \times 10^{-5}$ eV/°K $= 1.38 \times 10^{-11}$ pJ/°K.
I'll write k when it's clear from context.
Room temperature ~ 300°K,
so $k_{\rm B}T \approx 26$ meV $\approx 4.14 \times 10^{-9}$ pJ $= 4.14$ zeptojoules.
This is *room-temperature thermal energy.*

¶8. **Reliable signal processing:** "For a reasonable level of reliability, the signal energy should actually be much larger than the thermal energy, $E_{\rm sig} \gg E_T$ (Fig. II.2). For example, a signal level of

$$E_{\rm sig} \gtrsim 100 k_{\rm B}T \approx 2.6 \text{ eV}[= 4.14 \times 10^{-7} \text{ pJ}]$$

(at room temperature) gives a decently low error probability of around $e^{-100} = 3.72 \times 10^{-44}$."

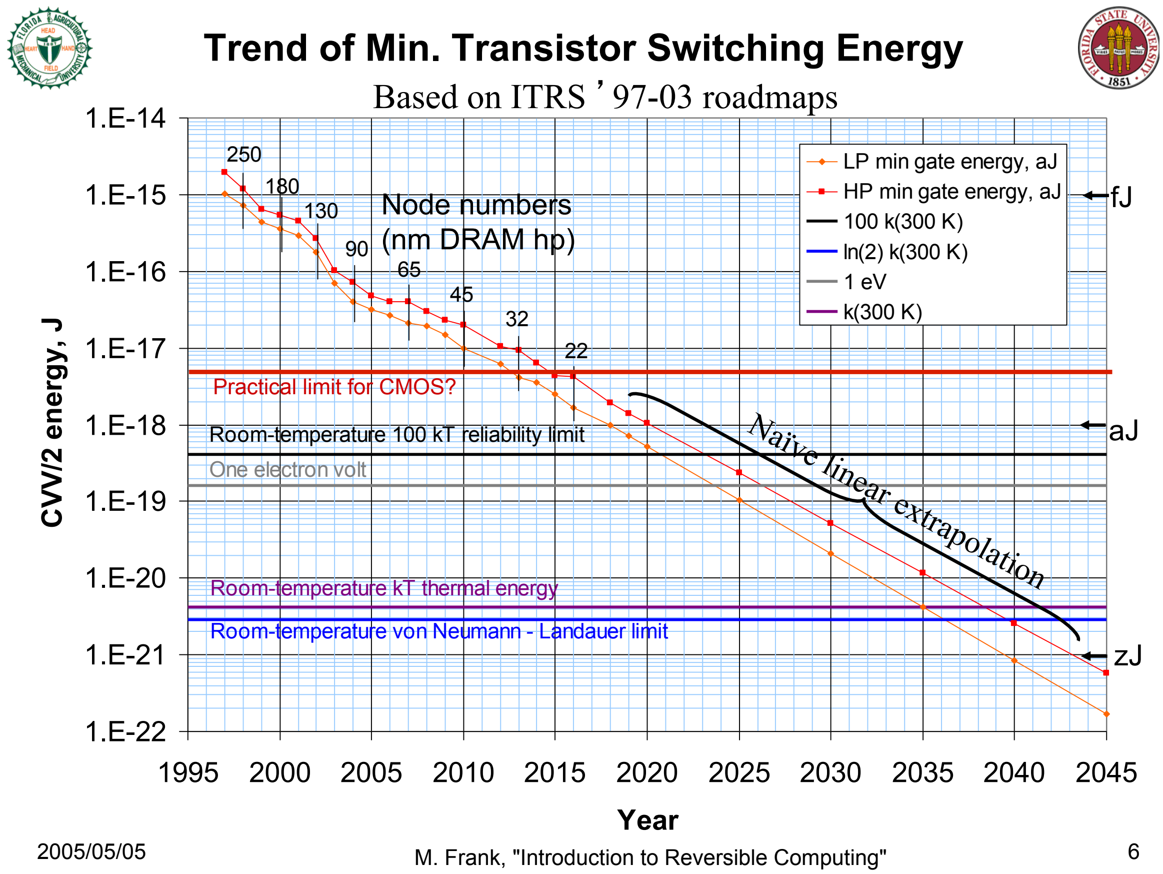

Figure II.1: [slide from Frank, RC 05 presentation]

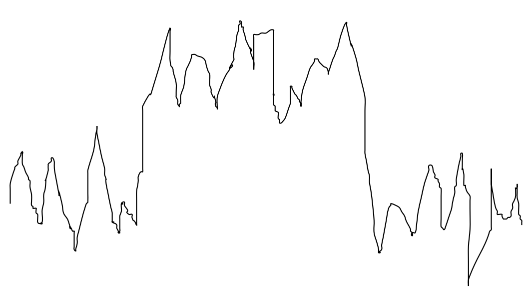

Figure II.2: Depiction of 0-1-0 pulses in the presence of high thermal noise.

A limit of $40k_{\mathrm{B}}T \approx 1$ eV is based on $R = 1/p_{\mathrm{err}}$, formula $E_{\mathrm{sig}} \geq k_{\mathrm{B}}T \ln R$, and a "decent" $R = 2 \times 10^{17}$.

¶9. **Lower operating temperature?:** Operating at a lower temperature does not help much, since the effective T has to reflect the environment into which the energy is eventually dissipated.

¶10. **Error-correcting codes?:** ECCs don't help, because we need to consider the *total energy* for encoding a bit.

¶11. "It is interesting to note that the energies of the smallest logic signals today are already only about $10^4 k_{\mathrm{B}}T$ ..., which means there is only about a factor of 100 of further performance improvements remaining, before we begin to lose reliability."

¶12. "A factor of 100 means only around 10 years remain of further performance improvements, given the historical performance doubling period of about 1.5 years. Thus, by about 2015, the performance of conventional computing will stop improving, at least at the device level ..."

¶13. **Power wall:** "About five years ago [2006], however, the top speed for most microprocessors peaked when their clocks hit about 3 gigahertz. The problem is not that the individual transistors themselves can't be pushed to run faster; they can. But doing so for the many millions of them found on a typical microprocessor would require that chip to dissipate impractical amounts of heat. Computer engineers call this the power wall."[1]

¶14. **Current fastest supercomputer:**[2] The IBM Sequoia runs at 16.325 petaflops (soon the be increased to 20).
1,572,864 processors. 1.6 petabytes of RAM.
Considered very efficient because consumes only 7.89 MW [enough for a city of about 8000 homes].

¶15. Scaling up current technology (such as Blue Waters) to 1 exaflop would consume 1.5 GW, more that 0.1% of US power grid.[3]

[1] *Spectrum* (Feb. 2011) `spectrum.ieee.org/computing/hardware/nextgeneration-supercomputers/0` (accessed 2012-08-20).

[2] *Computer* 44, 8 (Aug. 2012), p. 21.

[3] *Spectrum* (Feb. 2011) `spectrum.ieee.org/computing/hardware/nextgeneration-supercomputers/0` (accessed 2012-08-20).

¶16. Some recent supercomputers have had power efficiencies as high as 2 gigaflops/W.[4]
This is about 500 pJ/flop, or $F_{\mathrm{E}} = 2 \times 10^{-3}$ flop/pJ.
Note that these are flops, not basic logic operations/sec.

¶17. It might be possible to get it down to 5 to 10 pJ/flop, but "the energy to perform an arithmetic operation is trivial in comparison with the energy needed to shuffle the data around, from one chip to another, from one board to another, and even from rack to rack."[5]
(5 pJ $\approx 3 \times 10^7$ eV.)

¶18. It's difficult to use more than 5–10% of a supercomputer's capacity for any extended period; most of the processors are idling.[6]
So with more that 1.5×10^6 cores, most of the time more than a *million* of them are idle.

A.2 Reversible computing as solution

A.2.a Possible solution

¶1. Notice that the key quantity F_{E} in Eqn. II.1 depends on the energy *dissipated* as heat.

¶2. The $100 k_{\mathrm{B}} T$ limit depends on the energy in the signal (necessary to resist thermal fluctuation causing a bit flip).

¶3. There is nothing to say that information processing has to dissipate energy; an arbitrarily large amount of it can be recovered for future operations.

"Arbitrary" in the sense that there is no inherent physical lower bound on the energy that must be dissipated.

¶4. It becomes a matter of precise energy *management*, moving it around in different patterns, with as little dissipation as possible.

---

[4] `https://en.wikipedia.org/wiki/Supercomputer` (accessed 2012-08-20).

[5] *Spectrum* (Feb. 2011) `spectrum.ieee.org/computing/hardware/nextgeneration-supercomputers/0` (accessed 2012-08-20).

[6] *Spectrum* (Feb. 2011) `spectrum.ieee.org/computing/hardware/nextgeneration-supercomputers/0` (accessed 2012-08-20).

¶5. Indeed, E_{sig} can be increased to improve reliability, provided we minimize dissipation of energy.

¶6. This can be accomplished by making the computation *logically reversible* (i.e., each successor state has only one predecessor state).

A.2.b Reversible physics

¶1. All fundamental physical theories are Hamiltonian dynamical systems.

¶2. All such systems are time-reversible. That is, if $\psi(t)$ is a solution, then so is $\psi(-t)$.

¶3. In general, *physics is reversible.*

¶4. Physical information cannot be lost, be we can lose track of it. This is entropy: "unknown information residing in the physical state."
Note how this is fundamentally a matter of *information* and *knowledge.*
What is irreversible is the *information loss.*

A.2.c Von Neumann-Landaur Principle

¶1. **Entropy:** A quick introduction/review of the entropy concept. We will look at it in more detail soon (Sec. B.1).

¶2. **Information content:** The information content of a signal (message) measures our "surprise," i.e., how unlikely it is.
$I(s) = -\log_b \mathcal{P}\{s\}$, where $\mathcal{P}\{s\}$ is the probability of s.
We take logs so that the information content of independent signals is additive.
We can use any base, with corresponding units *bits*, *nats*, and *dits* (also, hartleys, bans) for $b = 2, e, 10$.

¶3. **1 bit:** Therefore, if a signal has a 50% probability, then it conveys one bit of information.

¶4. **Entropy 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, $H = -\sum_k p_k \log p_k$.

¶5. **Shannon's entropy:** 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."[7]
(This is one version of the quote.)

¶6. **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.

¶7. **Macrostates and microstates:** 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.

¶8. **Thermodynamic entropy:** 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).
$S = k_\mathrm{B} H$.
(There are technical details that I am skipping.)

¶9. **Microstates representing a bit:** Suppose we partition the microstates of a system into two macrostates, one representing 0 and the other representing 1.
Suppose N microconfigurations correspond to each macroconfiguration (Fig. II.3).
If we confine the system to one half of its microstate space, to reresent a 0 or a 1, then the entropy (average uncertainty in identifying the microstate) will decrease by one bit.

[7] `https://en.wikipedia.org/wiki/History_of_entropy` (accessed 2012-08-24).

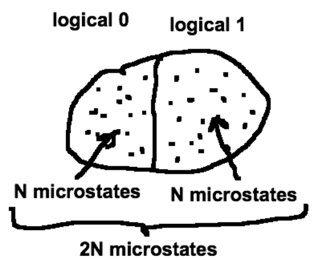

Figure II.3: Physical microstates representing logical states. Setting the bit changes the entropy by $\Delta H = \lg N - \lg(2N) = -1$ bit. That is, we have one bit of information about its microstate.

We don't know the exact microstate, but at least we know which half of the statespace it is in.

¶10. **Overwriting a bit:** Consider the erasing or overwriting of a bit whose state was originally another known bit.

¶11. We are losing one bit of physical information. The physical information still exists, but we have lost track of it.

Suppose we have N physical microstates per logical macrostate (0 or 1). Therefore, there are N states in the bit we are writing and N in the bit to be overwritten. But there can be only N is the rewritten bit, so N must be dissipated into the environment. $\Delta S = k \ln(2N) - k \ln N = k \ln 2 = 1$ bit dissipated. (Fig. II.3)

¶12. The increase of entropy is $\Delta S = k \ln 2$, so the increase of energy in the heat reservoir is $\Delta S \times T_{\text{env}} = kT_{\text{env}} \ln 2 \approx 0.7kT_{\text{env}}$. (Fig. II.4)
$kT_{\text{env}} \ln 2 \approx 18$ meV $\approx 3 \times 10^{-9}$pJ.

¶13. **von Neumann – Landauer bound:** This is the *von Neumann – Landauer (VNL) bound*. VN suggested the idea in 1949, but it was published first by Rolf Landauer (IBM) in 1961.

¶14. "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

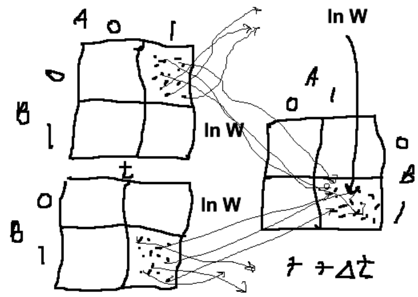

Figure II.4: Bit $A = 1$ is copied over bit B (two cases: $B = 0$ and $B = 1$). In each case there are $W = N^2$ micro states representing each prior state, so a total of $2W$ microstates. However, at time $t + \Delta t$ the two-bit system must be on one of W posterior microstates. Therefore W of the trajectories have exited the $A = B = 1$ region of phase space, and so they are no longer logically meaningful. $\Delta S = k \ln(2W) - k \ln W = k \ln 2 = 1$ bit dissipated.

quickly attain it within the next couple of decades." [EVLP]
That is, current circuits are about 18 eV.

¶15. **Experimental confirmation:** In March 2012 the Landauer bound was experimentally confirmed [EVPL].

A.2.d Reversible logic

¶1. To avoid dissipation, don't erase information. The problem is to keep track of information that would otherwise be dissipated.

¶2. This is accomplished by making computation *logically reversible.*

(It is already *physically* reversible.)

¶3. The information is rearranged and recombined *in place.* (We will see lots of examples of how to do this.)

A.2.e Progress

¶1. In 1973, Charles Bennett (IBM) first showed how any computation could be embedded in an equivalent reversible computation. Rather than discarding information, it keeps it around so it can later "decompute" it. This was *logical* reversibility; he did not deal with the problem of *physical* reversibility.

¶2. **Brownian Computers:** Or "Brownian motion machines." This was an attempt to suggest a possible physical implementation of reversible computation.
"the mean free path of the system's trajectory was much shorter than the distance between neighboring computational states" (see also [B82]).

¶3. Therefore: "In absence of any energy input, the system progressed essentially via a random walk, taking an expected time of $\Theta(n^2)$ to advance n steps."

¶4. A small energy input biases the process in the forward direction, so that it precedes linearly, but still very slowly.

¶5. Compare "DNA polymerization, which (under normal conditions, such as during cell division) proceeds at a rate on the order of only 1,000 nucleotides per second, with a dissipation of $\sim 40 k_{\mathrm{B}} T$ per step."
This is about 1 eV (see §8 below).
Note that DNA replication includes error-correcting operations.

¶6. **Digression: DNA data storage:** A team at Harvard has just reported converting a 53,000 word book to DNA and then reading it out by DNA sequencing.[8]
This is the densest consolidation of data in any medium (including flash memory, but also experimental media, such as quantum holography).
The book included 11 jpegs and a javascript program, for a total of 5.27 Mbits.
The decoded version had only 10 incorrect bits.
The book could be reproduced by DNA replication.
(The cost of DNA synthesis has been decreasing about 5× per year, and the cost of sequencing by about 12× per year.)

¶7. **Energy coefficient:** Since "asymptotically reversible processes (including the DNA example) proceed forward at an adjustable speed,

[8] `http://spectrum.ieee.org/biomedical/imaging/reading-and-writing-a-book-with-dna/` (accessed 2012-08-24). See also *Science* (Aug. 16, 2012), `DOI: 10.1126/science.1226355` (accessed 2012-08-24).

proportional to the energy dissipated per step," define an *energy coefficient*:

$$c_{\mathrm{E}} \stackrel{\mathrm{def}}{=} E_{\mathrm{diss}}/f_{\mathrm{op}},$$

"where E_{diss} is the energy dissipated per operation, and f_{op} is the frequency of operations."

¶8. "In Bennett's original DNA process, the energy coefficient comes out to about $c_{\mathrm{E}} = 1\mathrm{eV/kHz}$."
That is, for DNA, $c_{\mathrm{E}} \approx 40kT/\mathrm{kHz} = 40 \times 26$ meV/kHz ≈ 1 eV/kHz.

¶9. But it would be desirable to operate at GHz frequencies and energy dissipation below $k_{\mathrm{B}}T$.
Recall that at room temp. $k_{\mathrm{B}}T \approx 26$ meV (Sec. A.1 §6, p. 8).
So we need energy coefficients much lower than DNA.
This is an issue, of course, for molecular computation.

¶10. **Information Mechanics group:** In 1970s, Ed Fredkin, Tommaso Toffoli, et al. at MIT.

¶11. **Ballistic computing:** F & T described computation with idealized, perfectly elastic balls reflecting off barriers. Minimum dissipation, propelled by (conserved) momentum. Unrealistic. Later we will look at it briefly.

¶12. They suggested a more realistic implementation involving "charge packets bouncing around along inductive paths between capacitors."

¶13. Richard Feynman (CalTech) had been interacting with IM group, and developed "a full quantum model of a serial reversible computer" (Quant. mech. comps., *Found. Phys.*, 16, 6 (1986), 507–531).

¶14. **Adiabatic circuit:** Since 1980s there has been work in *adiabatic circuits*, esp. in 1990s.
An *adiabatic process* takes place without input or dissipation of energy.
Adiabatic circuits minimize energy use by obeying certain circuit design rules. "[A]rbitrary, pipelined, sequential logic could be implemented in a fully-reversible fashion, limited only by the energy coefficients and leakage currents of the underlying transistors."

¶15. As of 2004, est. $c_{\mathrm{E}} = 3$ meV/kHz, about 250× less than DNA.

¶16. "It is difficult to tell for certain, but a wide variety of post-transistor device technologies have been proposed ... that have energy coefficients ranging from 10^5 to 10^{12} times lower than present-day CMOS! This translates to logic circuits that could run at GHz to THz frequencies, with dissipation per op that is still less (in some cases orders of magnitude less) than the VNL bound of $k_{\mathrm{B}}T \ln 2$ ... that applies to all irreversible logic technologies. Some of these new device ideas have even been prototyped in laboratory experiments [2001]."

¶17. "Fully-reversible processor architectures [1998] and instruction sets [1999] have been designed and implemented in silicon."

¶18. But this is more the topic of a CpE course...