

Chapter II

Physics of Computation

These lecture notes are exclusively for the use of students in Prof. MacLennan's *Unconventional Computation* course. ©2015, B. J. MacLennan, EECS, University of Tennessee, Knoxville. Version of August 20, 2015.

A Energy dissipation

This lecture is based on Michael P. Frank, “Introduction to Reversible Computing: Motivation, Progress, and Challenges” (Frank, 2005b). (Quotations in this section are from this paper unless otherwise specified.)

¶1. Energy efficiency:

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

“where R = performance,

N_{ops} = number of useful operations performed during a job,

t = total elapsed time to perform the job,

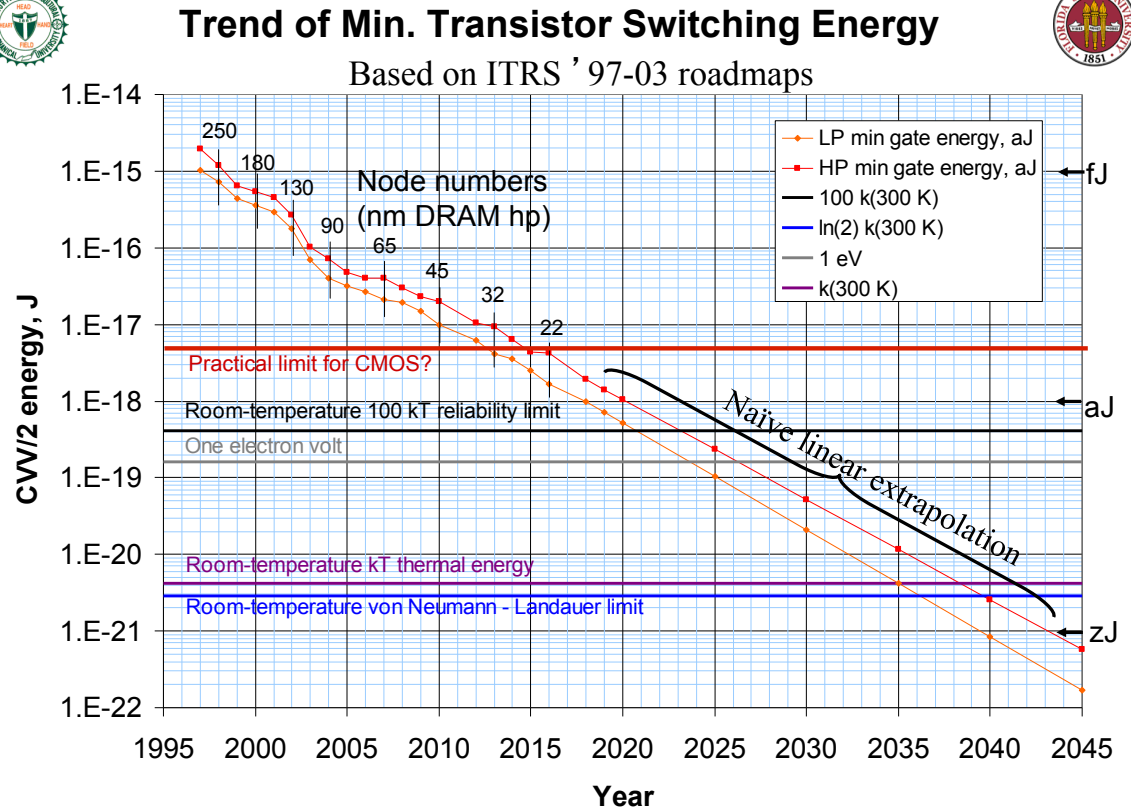
E_{diss} = energy dissipated during the job,

$F_{\text{E}} = N_{\text{ops}}/E_{\text{diss}}$ = 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 (bit ops) has been roughly given by $F_{\text{E}} \approx (1 \text{ op})/(\frac{1}{2}CV^2)$, where C is the typical capacitance of a node in a logic circuit, and V is the typical



2005/05/05

M. Frank, "Introduction to Reversible Computing"

6

Figure II.1: Frank (2005b, slide 9)

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 $E_{\text{sig}} = \frac{1}{2}CV^2$, and this energy gets dissipated whenever the node voltage is changed by the usual irreversible FET-based mechanisms in modern CMOS technology.”

This is the energy to either charge the capacitor or that is dissipated when it’s discharged.



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

¶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.”
The clock rate also goes up with smaller feature sizes.
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 $E_{\text{sig}} = \frac{1}{2}CV^2$ becomes small in comparison with the thermal energy $E_T = k_B T$, (where k_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:** $k_B \approx 8.6 \times 10^{-5} \text{ eV/K} = 1.38 \times 10^{-23} \text{ J/K} \approx 14 \text{ yJ/K}$.

Room temperature $\sim 300\text{K}$,
so $k_B T \approx 26 \text{ meV} \approx 4.14 \times 10^{-21} \text{ J} \approx 4 \text{ zJ}$.

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, $E_{\text{sig}} \gg E_T$ (Fig. II.2). For example, a signal level of”

$$E_{\text{sig}} \gtrsim 100k_B T \approx 2.6 \text{ eV} \approx 400 \text{ zJ}$$

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

A limit of $40k_{\text{B}}T \approx 1$ eV is based on $R = 1/p_{\text{err}}$, formula

$$E_{\text{sig}} \geq -k_{\text{B}}T \ln p_{\text{err}} = k_{\text{B}}T \ln R,$$

and a “decent” $R = 2 \times 10^{17}$.

- ¶9. This implies a maximum $F_{\text{E}} = 1 \text{ op/eV} \approx \frac{1 \text{ OP}}{1.6 \times 10^{-19} \text{ J}} = 6.25 \times 10^{18} \text{ op/J}$.
- ¶10. **Independent of technology:** Note that the preceding conclusions are independent of technology (electronic, optical, carbon nanotube, etc.).
- ¶11. **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.
- ¶12. **Error-correcting codes?:** ECCs don’t help, because we need to consider the *total energy* for encoding a bit.
- ¶13. “It is interesting to note that the energies of the smallest logic signals today [2005] are already only about $10^4 k_{\text{B}}T$. . . , which means there is only about a factor of 100 of further performance improvements remaining, before we begin to lose reliability.
- ¶14. “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 . . .”
- ¶15. **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
 This is also called “The 3 GHz Wall.”
 Note that more ops/sec implied more $E_{\text{diss}}/\text{sec}$.

¹*Spectrum* (Feb. 2011) spectrum.ieee.org/computing/hardware/nextgeneration-supercomputers/0 (accessed Aug. 20, 2012).

- ¶16. Official rates now in 4-5 GHz range, with overclocking up to 8.79 (liq N cooled!).
- ¶17. **Current fastest supercomputer:**² The Tianhe-2 has been clocked at 33.86 petaflops and is expected to run at 54.9 when relocated (and physically rearranged).
 3.12×10^6 cores in 16,000 nodes.
 1.34 PiB (petabytes) of CPU + GPU memory.
 Chips have 16 cores, 1.8 GHz cycle time, achieve 144 gigaflops, consuming 65W.
 It occupies 720m² (7750 ft²). 162 cabinets.³
- ¶18. Processor consumes 17.6 MW.
 Air conditioning consumes 24 MW.
 A total of 41.6MW, which is about enough for a city of 42,000 homes, or 108,000 people.
 This is about the size of Murfreesboro, the sixth largest city in TN.
- ¶19. There are currently plans to scale Tianhe-2 up to 100 petaflops. “Using that same technology to get to exascale would require on the order of 540 megawatts, about the output of a nuclear power plant.”⁴
- ¶20. Scaling up current technology (such as Blue Waters) to 1 exaflop would consume 1.5 GW, more than 0.1% of US power grid.⁵
- ¶21. Some recent supercomputers have had power efficiencies as high as 2 gigaflops/W.⁶
 Tianhe-2: $144 \text{ gflops} / 65\text{W} = 2.22 \text{ gflops/W} = 2.22 \times 10^9 \text{ flop/s/W} = 2.22 \times 10^9 \text{ flop/J} = 2.22 \times 10^{-3} \text{ flop/pJ}$.
 This is about 450 pJ/flop, or $F_E = 2.22 \times 10^{-3} \text{ flop/pJ}$.
 Note that these are flops, not basic logic operations/sec.

²<http://en.wikipedia.org/wiki/Tianhe-2> (accessed Aug. 20, 2013)

³Dongarra, J. “Visit to the National University for Defense Technology Changsha, China.” <http://www.netlib.org/utk/people/JackDongarra/PAPERS/tianhe-2-dongarra-report.pdf> (accessed Aug. 20, 2013).

⁴SOURCE

⁵*Spectrum* (Feb. 2011) spectrum.ieee.org/computing/hardware/nextgeneration-supercomputers/0 (accessed Aug. 20, 2012).

⁶<https://en.wikipedia.org/wiki/Supercomputer> (accessed Aug. 20, 2012).

- ¶22. The most energy-efficient supercomputers are about 4.5 gflops/W. That is, $1/4.5 \times 10^{-9} \text{ J/flop} \approx 220 \times 10^{-12} \text{ J/flop} \approx 220 \text{ pJ/flop}$.
- ¶23. 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.”⁷
(1 pJ $\approx 1.2 \times 10^7$ eV.)
- ¶24. It’s difficult to use more than 5–10% of a supercomputer’s capacity for any extended period; most of the processors are idling.⁸
So with that 3.12×10^6 cores, most of the time three *million* of them are idle.

⁷*Spectrum* (Feb. 2011) spectrum.ieee.org/computing/hardware/nextgeneration-supercomputers/0 (accessed Aug. 20, 2012).

⁸*Spectrum* (Feb. 2011) spectrum.ieee.org/computing/hardware/nextgeneration-supercomputers/0 (accessed 2012-08-20).