Chapter 8

Future reversible device technologies

In the previous chapter, we reviewed adiabatic circuits, the reversible computing technology that is most similar to the irreversible semiconductor technology that is the basis of essentially all present-day computing.

In this chapter, we look a bit farther afield, and review a number of proposals that have been made for computing technologies that might supersede traditional CMOS, once the limits of MOSFET technology are reached, and manufacturing processes develop to the point where constructing machines based on these alternative device technologies is feasible and economical. A number of the technologies we describe are (time-proportionately) reversible, or at least have reversible variants. We will describe the important parameters of these technologies that impact the scaling issues we discussed in chapter 6. We also will review several proposals that have been made for advanced cooling technologies.

We then calculate, using the formulas developed in §6.2.2.1, p. 128, how large a reversible machine would have to be, in the various proposed reversible technologies and under the various proposed cooling systems, in order for it to be faster per unit area than a machine built using various irreversible technologies. Most of these calculations were previously reported in §7 of our journal article [70].

8.1 Cooling technologies

Ordinary CPU chips in most present-day computers rarely dissipate more than 100 W of heat from a square centimeter of chip surface using normal passive cooling mechanisms, such as conduction through a ceramic package, and natural or forced convection through air. The chip surface is normally at least at room temperature

Cooling technology	Max entropy flux F_S in bits/s cm^2
Digital optic fiber	10^{13}
Typical passive emission	3.5×10^{22}
Drexler's fractal plumbing	3.8×10^{24}
Slow atomic ballistic	10^{26}
Fast atomic ballistic	3×10^{33}
Quantum maximum	5×10^{40}

Table 8.1: Estimates of the maximum entropy flux per unit area achievable with various existing and hypthetical cooling technologies. These are all rather rough estimates, and the last limit is especially arbitrary, since it is technically only valid for black holes having an arbitrarily-chosen 1 Å radius.

(300 K), so the entropy flux attained by these mechanisms is no larger than $F = 100\text{ W}/(k_B(300\text{ K})\ln(2)/\text{bit}) = 3.5\times10^{22}$ b/s-cm^2.

David Tuckerman [166, 167] has created and tested advanced semiconductor cooling systems which use forced convection of liquid coolant through micron-scale channels etched into the back of a silicon wafer. He has experimentally verified cooling rates on the order of $\sim$ 1000 W from a square centimeter-size chip, and has projected that higher rates are possible.

Drexler (1992, [51]), §11.5.3, p. 332, has designed a nanotechnological cooling system using a fractal plumbing network that ought to be able to remove at least 10kW/cm^2 of heat at 273 K from a flat slab of material up to 1 cm thick. This corresponds to an entropy removal rate of 3.8×10^{24} bit/s cm^2.

This figure corresponds roughly to the heat flux in the cooling systems of current-day nuclear reactors, which transport megawatts of heat through massive pipes on the order of a square meter in cross-section. (According to an acquaintance in the nuclear engineering department.)

If entropy were to be encoded in some material at the atomic scale at a density ρ_S of no more than 1 bit per cubic Ångstrom (roughly the volume of a small atom), and the material moves nearly ballistically through the computer at a speed of $v = 1$ m/s, the maximum entropy flux $F_S = \rho_S v$ is 10^{26} bit/s cm^2. (Allowing most of the machine's volume to be occupied by the cooling material.)

If the material instead moves at a tenth of the speed of light (a very fast speed that is still easy to analyze since relativistic effects are small), then the maximum flux is 3×10^{33} bit/s cm^2.

For materials around the density of water, 1 g/cm^3, Bekenstein's fundamental quantum-mechanical/general-relativistic bound on entropy (see §2.2.1, p. 33 and [15]) implies that even if all the material's mass-energy could be used for storing information, no more than about 1.7×10^6 bits can exist in an region 1 Å across. At a tenth the speed of light this gives an entropy flux of 5×10^{40} bit/s cm^2.

If entropy is removed digitally through 1 mm wide 100 GHz optical fiber available today, the maximum flux is only about 10^{13} bits/s cm^2. The maximum entropy flux that can be achieved using electromagnetic radiation is the blackbody flux, as we described with eq. 2.16 (§2.3, p. 39). We should note that the entropy density $S/\mathcal{V}$ in a thermal photon gas scales in proportion to T^3 ([88], p. 571), so achieving unboundedly high entropy densities using a stream of photons would require unboundedly high temperatures, which we may reasonably disallow.

Further, we should remember that the limit on entropy density given by Bekenstein's bound actually increases as information is encoded across regions of smaller and smaller diameters. If some technology can achieve Bekenstein's limit, then it may change the entire form of the appropriate scaling analysis. However, Bekenstein's bound may not actually be achievable, and in any case it seemingly only applies in the high-gravity realm that we have decided to avoid. So for now, we will stick with our general assumption that for any particular technology, entropy density is finite.

Table 8.1 summarizes the above figures.

Now, let us examine how these different flux limits affect the maximum possible rate of computing per unit area, under various computing technologies.

8.2 Irreversible device technologies

Based on the switching energy issues we discussed in §7.1.1 (p. 148), and typical parameters of modern VLSI fabrication processes, we estimate that the best present-day CMOS irreversible device technologies still generate at least $\sim 10^6$ bits of entropy per device-switching operation. This number will decrease somewhat over successive VLSI technology generations, as power supply voltages and circuit node capacitances decrease. However, as we saw in §7.1.2.1 (p. 155), supply voltages cannot decrease too much because of difficulties in setting device thresholds accurately. Moreover, in order to cope with thermal noise, total entropy generation per operation in irreversible CMOS circuits cannot decrease below a reliability-dependent number of nats per operation.

One very interesting alternative semiconductor logic technology is the "rapid single flux quantum" (RSFQ) superconducting logic family being developed by K. Likharev's research group at SUNY, and colleagues [109, 110, 195, 50]. This technology may be able to dissipate as little as 1 aJ (10^{-18} J) of energy per irreversible bit-operation at a temperature of 5 K, which corresponds to an entropy generation of only 21 kilobits.

Irreversible device technology	Entropy generated per bit erased	Operations per second per cm^2 surface in each cooling technology		
		Typical passive	Fractal Plumbing	Slow atomic ballistic
Modern CMOS	10^6	3.5×10^{16}	3.8×10^{18}	10^{20}
Likharev RSFQ	2.1×10^4	1.7×10^{18}	1.8×10^{20}	4.8×10^{21}
Best possible	1	3.5×10^{22}	3.8×10^{24}	10^{26}

Table 8.2: Maximum rate $\mathcal{R}_{\mathrm{A}}$ of irreversible operations per unit area achievable with various irreversible device technologies and cooling technologies from table 8.1.

Finally, we would like to consider a "best possible" irreversible technology that produces only 1 bit of physical entropy for each bit of information that is discarded. Merkle and Drexler (1996, [126]) argue that their proposed "helical logic" electronic logic technology could perform irreversible bit erasure with an energy dissipation approaching $k_{\mathrm{B}}T\ln 2$, which would create just 1 bit of entropy. Drexler's nanomechanical "rod logic" is also estimated to be capable of performing close to this limit as well (Drexler 1992 [51], §12.4.3d, p. 359). We expect that in general, as computational devices approach the nanoscale, a wide variety of different device designs will be found that are capable of asymptotically approaching the minimum entropy generation of 1 bit of entropy per bit of logical information that is irreversibly erased.

In table 8.2 we combine these entropy generation figures with the entropy flux rates from the previous section to calculate a maximum rate of irreversible bit operations per second, per unit of enclosing surface area, for various combinations of irreversible device technologies and cooling technologies. Note that these limits apply no matter how much extra hardware one packs in along the third dimension! As we saw in §6.2.2.1, ultimately, all irreversible technologies are limited to a fixed processing rate per unit of outer surface area, such as the limits given here.

Now, let us examine some reversible technologies and estimate the scales above which they exceed these irreversible rates of performance per unit area.

8.3 Reversible technologies

We now examine the entropy coefficients of a variety of reversible device technologies. Recall that the entropy coefficient of a technology expresses the amount of entropy generated per device operation, per unit of frequency at which the device is operated.

Based on the SCRL adiabatic circuit technology described in the previous chapter, we calculated the entropy coefficient for typical reversible logic gates fabricated in the fairly recent 0.5 μm VLSI process (HP14) that we used for FlatTop, when operating at room temerature. We obtained a value of about 43 bits/kHz. In an estimated "best available" process with around 10 kΩ transistor on-resistance, 1 V power supply, and 60 fF node capacitance, we estimate a somewhat lower value of $\sim$ 6 bits/kHz.

SCRL's entropy coefficient might be even better in an implementation based on low-resistance micro-electro-mechanical switches, as was suggested by Younis ([191], §2.7.3, p. 34). However, based on calculations I did using figures obtained from the MEMS (micro-electro-mechanical systems) community, although some of the best available MEMS switches apparently might offer an entropy coefficient as low as $\sim$ 0.003 bits/kHz, the size of these switches (on the order of 100 microns) is large enough that they do not end up outperforming MOSFETs in terms of reversible cost-efficiency. In other words, although the individual switches can run faster for a given dissipation per operation, a machine of a given speed per unit area must be larger.

Merkle [123] analyzed the energy dissipation of the reversible transfer of a packet of 100 electrons through a minimal quantum FET, and found it to be around $3{\times}10^{-21} J$ at a rate of 1 GHz. The corresponding entropy coefficient at room temerature is about 1.2 bits/GHz.

Drexler's rod logic, operated reversibly, would dissipate about $2{\times}10^{-21}$ J per operation at a speed of 10 GHz ([51], p. 354). Its entropy coefficient at room temperature thus comes out to 0.070 bits/GHz.

The "parametric quantron" superconducting reversible device of Likharev [108] dissipates about 10^{-24} J during a 1 ns operation at around 4 K ([108] p. 322); its entropy coefficient thus comes out at about 0.026 bits/GHz.

Finally, Merkle and Drexler's proposed helical logic [126] was analyzed by them to dissipate around 10^{-27} J at 10 GHz and 1 K when operated reversibly; its entropy coefficient thus comes out to be 10^{-5} bits/GHz. This is the lowest entropy coefficient that we have encountered so far.

Table 8.3 summarizes the above figures. Armed with them, we are now in a position to calculate the scale at which the various reversible technologies will beat the various irreversible technologies that we mentioned in §8.2. We will measure this scale first in terms of the number of devices required per unit area, then, in technologies for which we know the device volume, this can be used to find the necessary diameter or thickness of the machine.

Based on the analysis of section 6.2.2.1 (p. 128), we can express the number of reversible devices N_A per unit area required to achieve a given rate $\mathcal{R}_\mathrm{A}$ of operations per unit area as

$$N_\mathrm{A} = \mathcal{R}_\mathrm{A}^2 k_\mathrm{S}/F_\mathrm{S},$$

where as usual k_S is the entropy coefficient and F_S is the entropy flux per unit area.

Reversible device technology	Entropy coefficient k_S in bits/GHz
SCRL in HP14	4.3×10^7
SCRL in best available CMOS	6×10^6
Merkle quantum FET	1.2×10^0
Drexler rod logic	7.0×10^{-2}
Likharev parametric quantron	2.5×10^{-2}
Helical logic	1.0×10^{-5}

Table 8.3: Entropy coefficients k_S for some existing and proposed asymptotically reversible logic device technologies

To achieve the same rate of operation achievable by an irreversible machine that produces S bits of entropy per operation and uses the same cooling system, the number becomes

$$N_A = F_S k_S / S^2.$$

Table 8.4 shows the number of reversible devices in various technologies needed to beat the maximum per-area processing rate for the 3 combinations of irreversible technologies and cooling technologies that fall along the diagonal of table 8.2. The parenthesized numbers indicate cases in which the number of devices required may be determined by the maximum rate of operation of the devices, rather than by the entropy limits. The number given is the number of devices that would be required if the individual devices could run with as high a frequency as needed. The actual number required will most likely be higher.

To make sense of the non-parenthesized numbers in table 8.4, we estimate the volumes of the logic devices in various technologies. SCRL logic gates we will take to be about 10 μm $\times$ 10 μm $\times$ 1 μm = 100 μm^3. Merkle's quantum FET we estimate at about (10 nm)^3, a rod logic interlock as 40 nm^3 ([51], §12.4.2, p. 357), and a helical logic switch as 10^7 nm^3 ([126], §5.2, p. 330). Given these values we produce the results in table 8.5.

The parenthesized numbers in table 8.5 need explanation. The entries that say "any" indicate that even if the given reversible devices are arranged over a surface in only a single layer, they will still be faster than any machine built with the given irreversible technology within that surface. As for the 0.1 mm figure we calculated for 10^{12} helical logic devices per square centimeter beating the best possible irreversible technology given a 10^{26} bit/cm^2 entropy flux, it is probably inaccurate because the individual helical logic devices probably couldn't be made to run at the implied rate of 100 THz.

	Irreversible device and cooling technology combination		
	best CMOS/passive	RSFQ/convective	best/atomic
Entropy S, bits/op	10^6	2.1×10^4	1
Flux F, bits/s cm^2	3.5×10^{22}	3.8×10^{24}	10^{26}
Rate R, ops/s cm^2	3.5×10^{16}	1.8×10^{20}	10^{26}
Reversible Technology	Devices required per square cm to beat the above rate		
SCRL/best CMOS	2.1×10^8	5.2×10^{13}	6×10^{23}
Quantum FET	(42)	(107)	1.2×10^{17}
Rod logic	(2.4)	6×10^5	7×10^{15}
Helical logic	(3.5×10^{-4})	(86)	(10^{12})

Table 8.4: Numbers N_{A} of reversible machines per unit area required to beat different irreversible device technologies with different cooling strategies. Parenthesized numbers indicate lower bounds, where the real bounds depend on the maximum rate of operation of the devices.

	Irreversible device and cooling technology combination		
Reversible Technology	best CMOS/passive	RSFQ/convective	best/atomic
SCRL/best CMOS	0.21 mm	52 m	(4 au)
Quantum FET	(any)	(any)	1.2 mm
Rod logic	(any)	(any)	2.8 μm
Helical logic	(any)	(any)	(0.1 mm)

Table 8.5: Thicknesses d of reversible machines that can beat different irreversible technologies in terms of operations per unit area. “Any” indicates that even a single layer of the given reversible devices will suffice to beat the given irreversible technology.

The entry in the upper right corner of the table indicates that a machine built with current CMOS reversible technology, such as SCRL, would have to be the size of the inner solar system (!!) before it would be faster per unit area than the most efficient possible irreversible technology. Needless to say, a machine this large, composed mostly of solid silicon, would collapse under its own gravity.

In any case, the table indicates overall that most of the listed reversible technologies outperform most irreversible technologies, in terms of raw numbers of operations per second per unit area, for a wide range of cooling capabilities and for machines at a reasonable scale. Current CMOS reversible technology does not perform so well against the most efficient conceivable irreversible technologies, but it can still beat machines based on contemporary irreversible CMOS technology at reasonable scales.

One caveat to the above results is that in general a reversible device operation is not quite as computationally useful as an irreversible operation, due to the algorithmic issues we discussed in §3.3. However, for problems that have efficient reversible algorithms, like physical simulations (see §9.5.6), a small constant number of reversible device operations should suffice to do as much useful computational work as a single irreversible operation. The diameters in table 8.5 should therefore be increased by a factor of the same small constant.

8.4 Future device technologies—Conclusion

In this chapter, we listed a number of existing and proposed device technologies for both irreversible and reversible logic, and a variety of existing and hypothetical cooling technologies. Many of the technologies described cannot currently be built, but it is plausible that someday they might, and in any case all the technologies described serve as interesting points for comparison.

For each device technology, we gave the explicit numerical parameters determining its entropy generation, and from this, we determined limiting rates of operation per unit area for the irreversible technologies. Then, based on the superior scaling laws we have derived for time-proportionately reversible machines, we estimated the thickness of the reversible machines that would beat the irreversible machines' performance per unit area.

The upshot is that although present reversible technology is not so great, many of the proposed future reversible technologies would outperform *any* irreversible technology in terms of rate per unit area, even when considering only very thin layers—on the order of microns to millimeters thick—of packed logic devices in the given technology. This result holds firm unless a way is found to remove entropy from a system at a flux much higher than our rather arbitrarily-chosen maximum rate of 10^{26} bits per square centimeter per second. (A rate corresponding to 1 bit per cubic Ångstrom, moving at an arbitrary 1 m/s.)

These figures argue that in the long term, as computing technology moves down into the nanometer realm, and (eventually) away from conventional bulk-semiconductor techniques, reversibility will become a clear win in any macroscopic-scale computers built from such nano-scale devices.

This long-term trend makes it interesting and important to study reversible computer architectures and algorithms even today, because no matter the precise details of the future nano-scale device technologies that might become dominant, we can expect that using them in an asymptotically reversible way will confer substantially more computational power, in many applications for all but the smallest-scale machines. We will need reversible architectures and algorithms eventually; we can get a head start by designing them today. In the next chapter, we describe what we have learned along that direction so far.