2 Physics of Computation

2.1 Physical laws and information processing

2.1.1 Hardware representation

Information processing is often considered a purely abstract process, which can be discussed in purely mathematical terms. However, information is always represented in some physical entity, and processing and analyzing it requires physical devices [17, 18]. As a consequence, any information processing is inherently limited by the physical nature of the devices used in the actual implementation. While it is evident that an electronic chip with a high clock speed is more powerful as an information processor (in most respects) than a hand-operated mechanical computer, it is perhaps less obvious that the nature of the physical device does not just determine the clock rate, it can determine qualitatively the class of algorithms that can be computed efficiently.

This principle is often overlooked, but its consequences have often been discovered. Church and Turing asserted [19, 20] that most computers are equivalent with respect to computability (not with respect to speed), allowing one to disregard the details of the information processing device for determining if a given problem can be solved on a computer. This equivalence principle may well be considered the foundation of the computer science: it allows one to discuss information processing without reference to a specific hardware basis. However, the strong form of the Church–Turing hypothesis, which states that any problem that can be solved *efficiently* on one computer can be solved efficiently on any other computer, appears to be wrong: some problems have been established to be solvable efficiently if the computer operates according to quantum mechanics, but not on classical computers.

The physical laws governing the hardware that stores and processes the information determine, e.g., the amount of information that can be stored or the types of operations that can be applied to them and therefore the operations that can be included in an algorithm. They differ from mathematical limitations (e.g., complexity classes), which determine the number of logical operations needed to complete an algorithm, but not the speed at which it can be executed.

2.1.2 Physical laws and ultimate limits

Physical laws often allow us to determine the ultimate performance limits, even if the currently existing devices are very far from these limits. The best known examples are probably the speed of light, the conservation of energy or the thermodynamical limits on the energy efficiency of thermal engines, such as the Carnot cycle. These limits can not only indicate future roadblocks in the development of computer hardware, they also can be used as guidelines for the design of efficient devices. These limitations arise on all levels and relate to the performance of all computational steps, such as the storage of information, execution of logical operations, or the transfer of information between different parts of the computer. While they are also relevant for natural information processing devices (such as the human brain), we will consider here only artifacts, since their operation is still better understood and easier to quantify.

For this section, we will concentrate on physical laws that do not refer to a specific hardware basis chosen for implementing the information processing devices. We will refer to these issues as fundamental, in contrast to issues that depend on a specific technology, such as the speed at which a CMOS gate can be operated (which is, of course, also limited by physical laws). While most of our present information processing systems are still limited by technical rather than by fundamental physical limits, some systems are approaching these limits (e.g. the channel capacities of experimental fiber optics systems are close to the limit found by Shannon [21]) and other components will be approaching real or perceived limitations within the next few decades, provided that the current trends can be extrapolated. In the past, several apparent limitations could be overcome by conceptual changes.

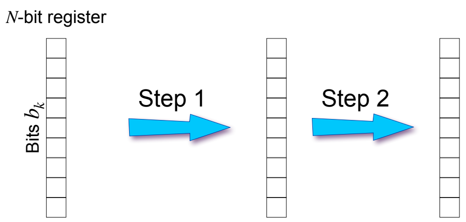

Figure 2.1: Model of computation: the information is stored in a register consisting of N (classical or quantum) bits. Computation is performed in discrete steps acting on this register. The subsequent registers represent the same register at different times.

Figure 2.1 shows schematically the model that we use to analyze the computational process: information is stored in N bits combined into a register. The computation is split into discrete steps executed in sequence. Each step uses information from the register to transform the register into a new state. For each step j, the state of bit $b_k(j+1)$ after the operation is determined by the state of all bits before this step,

$$b_k(j+1) = f_k^j(b_1(j), b_2(j), ...b_N(j)), \quad (2.1)$$

where the functions f_k^j together represent the logical operations acting on the register.

2.1.3 Quantum vs. classical information processing

Quantum and classical computers share a number of properties that are subject to the same physical limitations. As an example, the limits on processing speed that we discuss in the following section apply to both approaches. Similarly, the amount of information that can be stored in a system is limited by the number of distinguishable states of the system.

One of the major differences between classical and quantum computers is the existence of superpositions in the quantum computer, which implies that the amount of information processed by a single computational step is a single number of N bits in the classical computer, while the quantum computer processes typically 2^N numbers simultaneously.

Another, but less fundamental difference is that ideal quantum computers operate reversibly: logical operations are implemented by unitary transformations, which do not change the energy of the quantum state on which they operate and therefore (in the ideal case) do not dissipate any energy. As we discuss in more detail below, the operation of today's classical computers is irreversible. This is partly due to the logic foundations (Boolean logic uses non-invertible operations), and partly due to the hardware design. The progress in microelectronics is quickly reducing the dissipation per logical operation and considerations of the ultimate limits to the requirements on energy and power to drive logical operations are becoming relevant.

The quantum mechanical measurement process imposes some limitations on quantum computers that are not relevant for classical computers: the readout process will always change the information stored in a quantum computer, while its effect on a classical computer can be made arbitrarily small.

2.2 Limitations on computer performance

While some of the limits that physical laws set for the operation of computers are quite obvious (such as the speed of light as a limit for information transfer), others have only recently been established, while others have been shown not to be fundamental limits, if some of the concepts are adjusted.

2.2.1 Switching energy

One limitation that was held to be fundamental was that the operation of a logical gate working at a temperature T should dissipate at least the energy k_BT [22, 23]. At the time that these minimum energy requirements were discussed, actual devices required switching energies that were some ten orders of magnitude larger, so this perceived limit appeared quite irrelevant for any conceivable device.

As we discussed in the introduction, the situation has changed dramatically since then: the energy dissipated per logical step has decreased exponentially, at a rate of approximately a factor of ten every 4 years. This increase in energy efficiency is a requirement for the increase in speed and computational power and will need to continue if these other trends continue. Consider, e.g., a typical microprocessor with some 10^8 transistors being clocked at 1 GHz: if it were to dissipate 10 mJ per logical operation, as was typical in 1940, it would consume some 10^{15} W for a short time, probably disintegrating explosively within a few clock cycles.

It appears therefore quite likely that this trend must continue as long as the increase in speed and integration continues. As figure 1.4 shows, the extrapolation of this trend implies that the energy per logical step will reach the thermal energy k_BT ($T \approx 300$K) within about 10 years. This limit (k_BT) is relevant in at least two respects:

- $\frac{1}{2}k_BT$ is the average thermal energy per degree of freedom. Any environment that is at the temperature T will therefore inject this energy into switches that are not perfectly shielded from the environment, thus causing them to switch spontaneously.
- $k_BT\ ln2$ is the minimum energy that is dissipated by non-reversible gate operations, such as an AND operation.

We are therefore led to conclude that conventional electronic circuits will encounter problems when they reach this limit. However, as we discuss below, it is now established that information can be processed with techniques that dissipate less energy than k_BT per logical step. There is no lower limit for the energy required for a logical operation, as long as the switching time is not critical.

2.2.2 Entropy generation and Maxwell's demon

The flow of information in any computer corresponds to a transfer of entropy. Information processing is therefore closely tied to thermodynamics. As an introduction to these issues consider the Maxwell demon: As Maxwell discussed, in his "Theory of heat" in 1871,

> "If we conceive a being whose faculties are so sharpened that he can follow every molecule in its course, such a being, whose attributes are still essentially finite as our own, would be able to do what is at present impossible to us. For we have seen that the molecules in a vessel full of air at a uniform temperature are moving with velocities by no means uniform... Now let us suppose that such a vessel is divided into two portions, A and B, by a division in which there is a small hole, and that a being, who can see the individual molecules, opens and closes this hole, so as to allow only the swifter molecules to pass from A to B, and only the slower

> ones to pass from B to A. He will thus, without expenditure of work, raise the temperature of B and lower that of A, in contradiction with the second law of thermodynamics."

Clearly such a device is not in contradiction with the first law of thermodynamics, but with the second. A number of people discussed this issue, adding even simpler versions of this paradox. A good example is that the demon does not have to measure the speed of the molecules; it is sufficient if he measured its direction: He only opens the door if a molecule comes towards the door from the left (e.g.), but not if it comes from the right. As a result, pressure will increase in the right-hand part of the container. This will not create a temperature difference, but rather a pressure difference, which could also be used as a source of energy. Still, this device does not violate conservation of energy, since the energy of the molecules is not changed.

The first hint at a resolution of this paradox came in 1929 from Leo Szilard [24], who realized that the measurement, which must be taken on the molecules, does not come for free: the information required for the decision, whether or not to open the gate, compensates the entropy decrease in the gas. It is thus exactly the information processing, which prevents the violation of the second law.

While Szilard's analysis of the situation was correct, he only assumed that this had to be the case, he did not give a proof for this assumption. It was Rolf Landauer of IBM [22] who made a more careful analysis, explicitly discussing the generation of entropy in various computational processes. Other researchers, including Charles Bennett, Edward Fredkin, and Tommaso Toffoli showed that it is actually the process of erasing the information gained during the measurement (which is required as a step for initialising the system for the next measurement) which creates the entropy, while the measurement itself could be made without entropy creation. Erasing information is closely related to dissipation: a reversible system does not destroy information, as expressed by the second law of thermodynamics. Obviously most current computers dissipate information. As an example, consider the calculation $3+5=8$. It is not possible to reverse this computation, since different inputs produce this output. The process is quite analogous to the removal of a wall between two containers, which are filled with different pressures of the same gas.

The creation of entropy during erasure of information is always associated with dissipation of energy. Typically, the erasure of 1 bit of information must dissipate at least an energy of $k_B T \ln 2$. This can be illustrated in a simple picture. We assume that the information is stored in a quantum mechanical two-level system, the two states being labeled $|0\rangle$ and $|1\rangle$. Erasing the information contained in this bit can be achieved by placing it in state $|0\rangle$, e.g., independent of its previous state. This cannot be achieved by a unitary operation, i.e., by (energy-conserving) evolution under a Hamiltonian: such an evolution is reversible and the previous state could always be recovered by inverting the operation. The information would therefore not really be erased.

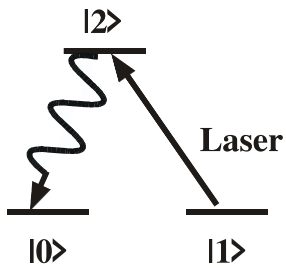

Figure 2.2: Erasing one bit of information, i.e., setting it unconditionally to the value $|0\rangle$ can be achieved by driving the transition from state $|1\rangle$ to an auxiliary state $|2\rangle$ with a laser.

Figure 2.2 shows a simple system that allows for initialization of a qubit by spontaneous emission. A laser drives the transition from state $|1\rangle$ to an auxiliary optically excited state $|2\rangle$. If this state has a non-vanishing probability to decay to state $|0\rangle$, it will eventually end up in this state, since this does not interact with the laser beam. It represents therefore a (re-)initialization of the qubit into state $|0\rangle$. For this scheme to work, the third state $|2\rangle$, must have an energy higher than that of state $|1\rangle$. If the system is initially in state $|1\rangle$, the pulse puts it in state $|2\rangle$. If it is initially in state $|0\rangle$, the pulse does nothing. From state $|2\rangle$, the system will undergo spontaneous emission; in a suitable system, the probability for this spontaneous emission to bring the atom to state $|0\rangle$ approaches unity. In systems where the probability is not high enough, the procedure must be repeated.

The minimum energy expenditure for this procedure is defined by the photon energy for bringing the system into the excited state. This energy must be larger than k_BT, since the system could otherwise spontaneously undergo this transition, driven by the thermal energy. Similar requirements hold in classical systems, where dissipation is typically due to friction.

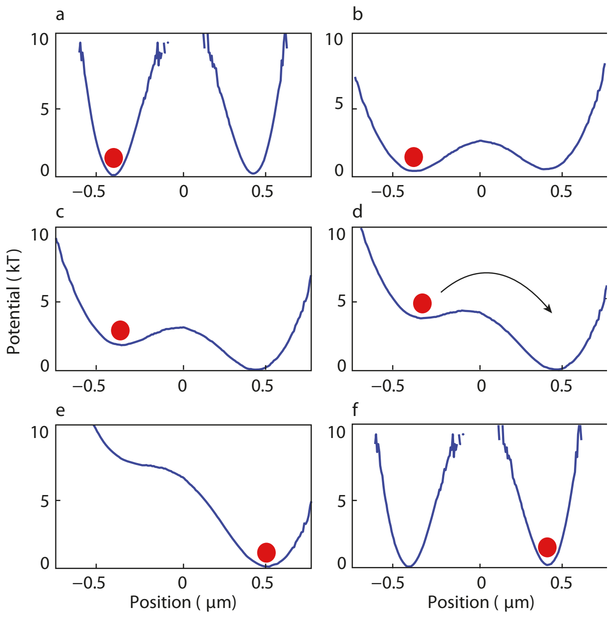

Figure 2.3: Erasing one bit of information in a time-dependent potential. [25]

This theoretical prediction was verified experimentally in 2012 by measurements on a colloidal particle trapped in a time-dependent potential [25]. The authors used a bistable potential, where the two minima represented the logical states of the system. They erased the information by first lowering the central barrier and then applying a tilting force. The figures represent the transition from the initial state, 0 (left-hand well), to the final state, 1 (right-hand well). With this procedure, the final state of the particle is always 1, irrespective of the initial state. The experiment used glass beads (2 µm diameter) manipulated by optical tweezers.

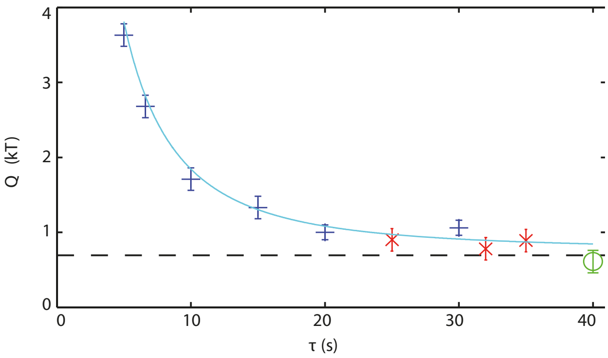

Figure 2.4: Heat generated by erasing one bit of information as a function of the switching time. [25]

As shown in Fig. 2.4, the heat generated by the erasing process approaches $k_BT \ln 2$ if the switching process is slow.

2.2.3 Reversible logic

As discussed before, conventional computers use Boolean logic, which includes the operations AND and OR. Both these operations, which have two input bits and one output bit, discard information, i.e., they reduce the phase space. When the system has fewer accessible states, its entropy is lower. Since the second law of thermodynamics does not allow a decrease in the entropy of a closed system, this decrease has to be compensated by the generation of entropy at some other place. The entropy generated by erasing a bit of information is $\Delta S = k_BT \ln 2$. Computers based on Boolean logic are therefore inherently dissipative devices, with the dissipation per logical step of at least $k_BT \ln 2$. This generation of heat during the computational process represents an obvious limitation on the possible speed of a computer, since no physical device can withstand arbitrary amounts of heat generation.

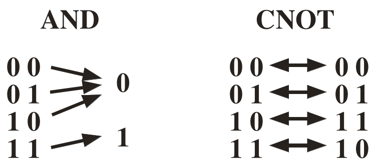

Figure 2.5: Examples of an irreversible (AND) and reversible (CNOT) gate.

It turns out, however, that computers do not have to rely on Boolean logic. They can use reversible logic instead, which preserves the information, generating no entropy during the processing [26]. Figure 2.5 shows an example of a reversible logic gate, the so-called controlled NOT or CNOT gate, which can be used to implement arbitrary algorithms. This particular gate is its own inverse, i.e., $\text{CNOT} \cdot \text{CNOT} = 1$.

Quantum information processors use unitary operations to perform computations. Since unitary operations are always reversible, they therefore require algorithms that use only reversible logical gates. For the example of a quantum computer, it is easy to prove that the energy dissipation during the computation vanishes. For this we calculate the energy of the quantum register at time t

$$\begin{aligned}\langle E\rangle(t) &= \mathrm{Tr}(\mathcal{H}\boldsymbol{\rho}(t)) = \\ &= \mathrm{Tr}(\mathcal{H}e^{-i\mathcal{H}t}\rho(0)e^{i\mathcal{H}t}) \\ &= \mathrm{Tr}(e^{i\mathcal{H}t}\mathcal{H}e^{-i\mathcal{H}t}\rho(0)) = \langle E\rangle(0),\end{aligned}$$

where we have used that $[e^{i\mathcal{H}t}, \mathcal{H}] = 0$. (The density operator $\boldsymbol{\rho}$ describes the state of the system, Tr denotes the trace, see Chapter IV.)

Figure 2.6 shows schematically how a reversible operation could be implemented by a time-modulated potential and a coupling between source and target. The double well potential represents the information: the bead in the left hand well corresponds to the logical value 0, the bead in the right hand well to the value 1. Each potential therefore stores one bit of information, with

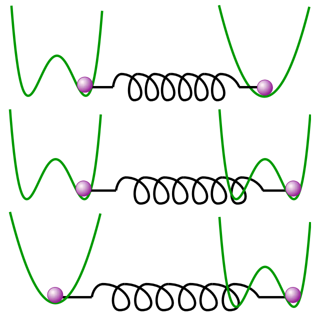

Figure 2.6: Reversible copy operation in a time-modulated potential.

the single minimum well representing a neutral state. The copy operation is achieved by modulating the potential between a monostable and a bistable state in such a way that no energy is expended. The modulation must be sufficiently slow that the system can follow it adiabatically. The spring, which is a passive device, assures that the bead in the second well falls into the left or right sub-well, depending on the position of the other bit.

Reversible computer

A general reversible computer can be represented as a system of states, corresponding to the information stored in the computer, and a sequence of logical operations, transforming one such state into the next. Since no information is discarded, it is possible to reverse the complete computation and bring the system back to its initial state, simply be reversing each of the logical operations. No minimum amount of energy is required to perform reversible logical operations. However, not discarding any information also implies that no error correction or re-calibration is done, since these processes also discard (unwanted) information. Reversible computation (which includes quantum computation) therefore requires very reliable gate operations.

2.2.4 Reversible gates for universal computers

The first proof that reversible logic gates can form the basis of a universal computer is due to Fredkin and Toffoli [27]. They proposed a three-bit gate that is now known as the Fredkin gate, which can be operated in a reversible way (details will be discussed in Section III). It can be described as a controlled-SWAP operation: the first bit acts as the control; if it is =1, the 2^{nd} and 3^{rd} bit are swapped. The control bit is always transmitted unchanged. Since the SWAP operation is it's own inverse, SWAP^2=1, the Toffoli gate is also its own inverse.

The Fredkin gate can be used to implement a reversible AND gate by identifying the inputs of the AND gate with two lines of the Fredkin gate and setting the third input to the fixed value 0. The corresponding (third) output line then contains the output of the AND gate, while the two other lines contain bits of information which are not used by the Boolean logic, but would be required to reverse the computation. The following table shows the outputs for all possible inputs of the first two bits (with the third set to 0):

Inputs 1,2	00	01	10	11
Output 1	0	0	1	1
Output 2	0	1	0	0
Output 3	0	0	0	1

Other reversible gates can be derived from the Fredkin gate in a similar way: the irreversible Boolean gate is embedded in the larger Fredkin gate.

When irreversible gates are embedded in larger reversible ones, some of the output lines are not used in the rest of the computation. They can be erased at the corresponding dissipation expense, or they can be used to reverse the computation after the result has been read out, thus providing a truly reversible operation of the machine at the expense of some additional bits whose number grows linearly with the length of the computation [26].

Another reversible computational architecture is the reversible Turing machine. As discussed in more detail in section III, a Turing machine consists of an infinitely long tape storing bits of information, a read/write head that can be in a number of different states, and a set of rules stating what the machine is to do depending on the value of the bit at the current position of the head and the state of the head. A reversible set of rules would be the set of operations represented in Table 2.1.

head state	bit read	change bit to	change state to	move to
A	1	0	A	left
A	0	1	B	right
B	1	1	A	left
B	0	0	B	right

Table 2.1: Reversible Turing machine

The information processing corresponds to a motion of the head. The motion is driven by thermal fluctuations and a small force defining the direction. The amount of energy dissipated in this computer decreases without limit as this external force is reduced, but at the same time the processing speed decreases. Overall the best picture to describe the operation of a reversible computer is that it is driven along a computational path. The same path may be retraced backward by changing some external parameter, thereby completely reversing the effect of the computation.

2.2.5 Processing speed

One limit for the processing speed can be derived from the uncertainty principle. It can be shown [28] that it takes at least a time

$$\Delta t = \frac{\pi \hbar}{2E} \tag{2.2}$$

for a quantum mechanical state to evolve into an orthogonal state, if E is the energy of the system. This condition is a requirement for two states to be distinguishable, which is one condition to qualify as a computational step. This limit therefore defines a minimal duration for a computational step given the available energy E. It does not imply, however, that this energy must be dissipated during this step. In an ideal system, the energy will remain available for the continuation of the computation.

Quantum computers work close to this limit if the energy is equated with the energy range of the eigenstates of the relevant Hamiltonian. This implies that only the energy in the system degrees of freedom is included in the calculation, not the (usually much larger) energy stored in bath degrees of freedom, in particular, not the rest mass of the system. In an NMR quantum computer, e.g., where the relevant degrees of freedom are the nuclear spins, the energy available to the computation is the Zeeman energy of the spins.

This system also permits a verification of the condition stated above. Setting the energy of the ground state $|\uparrow\rangle$ to zero, the excited state $|\downarrow\rangle$ has an energy $\hbar\omega_L$(where ω_L is the Larmor frequency of the spin, which is proportional to the magnetic field). An initial state

$$|\Psi(0)\rangle = \frac{1}{\sqrt{2}}(|\uparrow\rangle + |\downarrow\rangle)$$

then evolves into

$$|\Psi(t)\rangle = \frac{1}{\sqrt{2}}(|\uparrow\rangle + e^{-i\omega_L t}|\downarrow\rangle)$$

The final state is distinguishable from the initial state if the two are orthogonal, i.e.

$$\begin{aligned}\langle\Psi(0)|\Psi(t)\rangle &= \frac{1}{2}\langle(\uparrow + \downarrow)|(\uparrow + e^{-i\omega_L t}\downarrow)\rangle \\ &= \frac{1}{2}(1 + e^{-i\omega_L t}\downarrow) = 0.\end{aligned}$$

This can be simplified to $e^{-i\omega_L t} = -1$ or $\omega_L t = \pi$. Apparently the two states are orthogonal after $t = \pi/\omega_L$.

Since the energy of the ground state is 0, the energy of the excited state is $\hbar\omega_L$, and the population of both states is $1/2$, the (constant) energy of the superposition state is $E = (0 + \hbar\omega_L)/2 = \hbar\omega_L/2$. Solving for $\omega_L = 2E/\hbar$ and inserting it into $\omega_L t = \pi$, we recover the condition $\Delta t = \frac{\pi\hbar}{2E}$ given above.

An interesting aspect of this limit is that it does not depend on the architecture of the computer. While we generally expect computers containing many processors working in parallel to be faster than purely serial computers, this is no longer the case for a computer working at the limit just discussed: if the number of processors increases, the available energy per processor decreases and correspondingly its speed. The total number of logical operations per unit time remains constant.

2.2.6 Information content and speed

A limit on the amount of data stored in the computer can be derived from thermodynamics. According to statistical mechanics, the entropy of a system is

$$S = k_B \ln W, \tag{2.3}$$

where W is the number of accessible states. To store N bits of information, we need N two-level systems, which have 2^N states.

Assuming that all sates are roughly equally probable, that is, every state occurs with probability $p_i = 2^{-N}$, we calculate the entropy of the register using statistical thermodynamics as

$$\begin{aligned}S &= -k_B \sum_i p_i \ln p_i = \sum_i 2^{-N} \ln(2^{-N}) \\ &= N k_B \ln 2.\end{aligned}$$

This entropy is formally that of an ensemble at a given energy, while the actual system doing the computation is in a well-defined (pure) state, thus having zero entropy.

The information content I as given by the number of qubits is related to the entropy by

$$I = N = \frac{S}{k_B \ln 2}$$

and the maximum number of operations per qubit per second is

$$\frac{1}{I}\frac{2E}{\pi\hbar} = \frac{4k_B \ln 2}{\pi\hbar}\frac{E}{S}.$$

For many macroscopic systems (see below for an example) $\frac{E}{S}$ is roughly proportional to the temperature T, and thus

maximum number of operations per bit per second $\approx \frac{k_B}{\hbar}T$.

2.2.7 Additional details

Reversible gates (quantum or classical) do exist and are discussed elsewhere in this course. Let us briefly sketch the overall design of a general reversible computation. The computer memory is divided into three parts fulfilling different tasks: program storage, data storage, and work space. As illustrated in Figure 2.7, most of the garbage generated by running the program is taken care of by running the program backwards. Only in the final step entropy is generated (and energy dissipated) by erasing program and input data in order to prepare everything for the next computation.

2.3 The ultimate laptop

2.3.1 Processing speed

Some limits to the performance of computers have been summarised by Seth Lloyd [29] in a very popular style: he discusses the "ultimate laptop" , i.e., the maximum performance that a computer of 1 kg mass and a volume of 1 l may ultimately achieve. "Ultimately" means again that this approach does not consider any specific implementation; in fact, the conditions considered are such that it is highly unlikely that any device will ever be built that even remotely approaches the conditions that are derived here. Nevertheless, the considerations are instructive in showing that limitations will eventually become important, no matter what advances materials science will make in the future. Specific assumptions are that the mass of his laptop is one kilogramand the volume is one liter. He asked what kind of computation could be achieved with this system, given the values of the fundamental constants $\hbar$, c, k_B, and G (the Newtonian gravitational constant).

The limit on the processing speed discussed in Section 2.2.5 would be reached if all the mass of the computer were available as energy for driving the computation; it can be obtained from the condition (2.2) on the processing speed. An energy of

$$E = mc^2 = 9 \cdot 10^{16}\mathrm{J}$$

results in a maximum speed of

$$n = \frac{2E}{\pi\hbar} = \frac{2mc^2}{\pi\hbar} = \frac{1.8 \times 10^{17}}{3.2 \times 10^{-34}} = 5 \cdot 10^{50}$$

operations per second.

An additional limit derives from the necessity to include error correction. Detecting an error can in principle be achieved without energy dissipation. However, correcting it implies eliminating information (about the environment), thus generating dissipation. The dissipated energy will heat the computer and must be removed to the environment. We will assume here that energy dissipation is limited by blackbody radiation. An ideal blackbody radiates, according to the Stefan-Boltzmann law, a power of

$$P = \sigma A T^4,$$

where $\sigma = 5.67 \times 10^{-8}\mathrm{Wm}^{-2}\mathrm{K}^{-4}$ is the Stefan-Boltzmann constant, A the surface area and T the temperature. If this computer operates, e.g., at a temperature of $T = 6 \times 10^8\mathrm{K}$, with a surface area of $0.01\mathrm{m}^2$, the power of the blackbody radiation amounts to $P = 4 \times 10^{26}\mathrm{W}$. This energy throughput (which is required for error correction, not for operation) corresponds to a mass

program storage	data storage	work space
program → load from outside	input → load from outside	blank
⇓	⇓	⇓
run the computation		
⇓	⇓	⇓
program	output ← copy to archive	some stuff
⇓	⇓	⇓
run the computation backwards		
⇓	⇓	⇓
program erase	input erase	blank reuse

Figure 2.7: General scheme of reversible computation.

throughput of

$$\frac{dm}{dt} = P/c^2 = 1\frac{\text{kg}}{\text{ns}},$$

which must be fully converted to energy. If this is possible, the number of error bits that can be rejected per second is 7×10^{42} bits per second. With a total of 10^{50} logical operations per second, this implies that its error rate must be less than about 10^{-8} to achieve reliable operation.

2.3.2 Maximum storage density

A limit that may be easier to approach can be obtained from the following model: We assume that every atom of the system can store at most 1 bit of information. This is in principle fulfilled in NMR and ion trap quantum computers. For a mass of 1 kg, the number of atoms would be of the order of 10^{25}. At this density, it would thus be possible to store 10^{25} qubits of information in a computer. If optical transitions of these atoms are used for logical operations, gate times of the order of 10^{-15}s would be feasible, allowing a total of 10^{40} logical operations per second for the whole computer. The difference between this number and the 10^{50} quoted above is the energy content: here, we assume an electronic energy difference between the states of ≈ 1 eV, while the rest mass of the atom corresponds to an energy of $mc^2 \approx 10^{-23} \cdot 10^{17} \approx 10^{-6}\text{J} \approx 10^{13}\text{eV}$.

At such data rates, the different parts of the computer would not be able to communicate with each other at the same rate as the individual logical operations. The computer would therefore need a highly parallel architecture. If serial operation is preferred (which may be dictated by the algorithm), the computer needs to be compressed. Fully serial operation becomes possible only when the dimensions become equal to the Schwarzschild radius ($= 1.5 \cdot 10^{-27}$m for $m = 1$ kg), i.e., when the computer forms a miniature black hole.

While all these limits appear very remote, it would only take of the order of 100–200 years of progress at the current rate (as summarized by Moore's law) to actually reach them. It is therefore very likely that a deviation from Moore's law will be observed within this time frame, irrespective of the technology being used for building computers.

2.3.3 Monoatomic gas

The simplest example of a macroscopic system is the classical monatomic ideal gas with N particles of mass m, whose entropy is given by the

Sackur-Tetrode formula

$$S = \frac{5}{2} N k_B + N k_B \ln \left[\frac{V}{N} \left(\frac{4\pi m}{3h^2} \frac{E}{N} \right)^{\frac{3}{2}} \right].$$

Noting that $\frac{E}{N} = \frac{3}{2} k_B T$ and measuring T in appropriate dimensionless units we obtain

$$S = N k_B (a + b \ln T) \approx \text{const},$$

which means that the temperature dependence of S (logarithmic) is utterly negligible compared with that of E (linear), and thus, in this approximation,

$$\frac{E}{S} = \frac{3}{2} N k_B \frac{T}{\text{const}} \sim T,$$

as claimed above.

2.3.4 Massless particles

Since the classical monatomic ideal gas is not the only macroscopic system we have to find out what kind of system (within the given limits of one liter and one kilogram) provides most entropy at a given temperature T. Systems of massive particles are improbable since the density of particles of mass m in equilibrium is

$$\rho(m) \sim \exp\left(-\frac{mc^2}{k_B T} \right)$$

which is exceedingly small for $mc^2 \gg k_B T$. We thus have to focus on systems dominated by massless particles. A standard example in statistical mechanics is the Stefan-Boltzmann law for the energy of a gas of massless particles (for example, photons or acoustic phonons):

$$\frac{E}{V} = r_e \frac{\pi^2}{30} \frac{(k_B T)^4}{(\hbar c)^3} \sim r_e T^4,$$

where r_e is a particle-specific factor of the order of unity: the number of photon polarizations ($r_e = 2$) or phonon branches ($r_e = 3$). The free energy for massless particles is $F = -\frac{1}{3} E$ and thus we obtain the entropy:

$$S = -\frac{\partial F}{\partial T} = \frac{4}{3} \frac{E}{T} \sim E \left(\frac{r_e}{4} \right)^{\frac{1}{4}} = r_e^{\frac{1}{4}} E^{\frac{3}{4}}.$$

We see that the dependence on the number of available particle species r_e is very weak. The exact formula is

$$S = I k_B \ln 2 = \frac{4}{3} k_B \left(\frac{\pi^2 r_e V}{30 \hbar^3 c^3} \right)^{\frac{1}{4}} E^{\frac{3}{4}}.$$

At low temperature, only photons are present, and thus $r_e = 2$. With $V = 1$l and $E = 1\text{kg}c^2$and the operating temperature $T = 5.87 \cdot 10^8$K (see above), we then obtain

$$\begin{aligned} S &= 2.04 \cdot 10^8 \frac{\text{J}}{\text{K}} \\ \Rightarrow \quad I &= \frac{S}{k_B \ln 2} = 2.13 \cdot 10^{31} \text{bits}. \end{aligned}$$

(One might ask at this point wether $6 \cdot 10^8$ K really is a low temperature. This is answered by comparing the corresponding energy of roughly $6 \cdot 10^4$ eV to the rest energy of an electron, which is $5.11 \cdot 10^5$ eV.)

Correspondingly, the number of operations per bit per second is

$$\frac{1}{I} \frac{2E}{\pi \hbar} \approx 10^{19},$$

which is way beyond the clock speed of any processor available in 2013 and probably for some years to come. Of course none of the fundamental limits discussed here will be easily reached within the foreseeable future. However, the many orders of magnitude between today's laptops and the ultimate laptop serve to illustrate nicely that there is a lot of technical progress which can be made before hitting the hard walls set up by fundamental physical laws.

2.3.5 Parallel / serial operation

This gigantic memory and the enormous speed of the ultimate laptop can only be used efficiently in highly parallel computing. To see this we note that the speed of communication between different parts of the computer is limited by the speed of light, and thus the typical communication time is

$$t_{com} = \frac{2R}{c}$$

where $R \approx 0.1$m is the size of the laptop. This has to be compared to the time for a bit flip (state change)

$$t_{flip} = \frac{\pi\hbar}{2k_B \ln 2} \frac{S}{E}.$$

The ratio between these times then is

$$\frac{t_{com}}{t_{flip}} = \frac{k_B 4 \ln 2}{\pi\hbar c} \frac{RE}{S} \sim \frac{k_B T}{\hbar c} R \sim \frac{R}{\lambda_T},$$

where λ_T is the thermal de Broglie wavelength determined by the equality between the thermal energy and that of a massless particle of wavelength λ_T:

$$k_B T = \hbar c \frac{2\pi}{\lambda}.$$

For $R = 0.1$ m that means

$$\frac{t_{com}}{t_{flip}} = 10^{10}.$$

Since communication is extremely slow compared to the bit flip rate the computer must operate in an extremely parallel way: every single bit must go through 10^{10} operations on its own before it gets another chance to communicate with other bits.

In order to reduce the necessary degree of parallelization (to make the computer work more serially), communication time must be reduced by making the computer smaller (than the original 1 liter size). The ultimate limit of compression is reached in a black hole, with Schwarzschild radius R_S. The Schwarzschild radius is given by

$$\frac{Gm^2}{R_S} = \frac{mc^2}{2},$$

meaning that the classical Newtonian gravitational energy of two particles at distance R_S becomes comparable to the rest energy. The entropy (and information content) of a black hole is proportional to its surface area, and for $m = 1$ kg one obtains

$$\begin{aligned} R_S &= 1.5 \cdot 10^{-27}\text{m} \\ I &= 4 \cdot 10^{16}\text{bits} \\ t_{flp} &= I \frac{\pi\hbar}{2E} = \frac{4 \cdot 10^{16}}{2 \cdot 10^{-51}} \approx 7 \cdot 10^{-35} \\ t_{(com)} &= \frac{2R_S}{c} = \frac{3 \cdot 10^{-27}}{3 \cdot 10^{8}}\text{s} = 10^{-35}\text{s} \end{aligned}$$

leading to

$$\frac{t_{com}}{t_{flip}} \sim 1.$$

Thus the ultimate serial computer is a black hole.

Returning to the original one liter extremely parallel device we finally point out that parallelism is also related to the admissible error rate. In order to correct an error which has occurred, that error must first be communicated to another region of the computer, or to the outside world. Thus t_{com} is the smallest tolerable time interval between two successive errors of a given bit, and consequently the maximum admissible error rate is

$$\frac{t_{flip}}{t_{com}} \sim 10^{-10},$$

which is challenging, to say the least.

Further reading

A brief, nontechnical introduction into the thermodynamic aspects of computation is given in two articles in Scientific American [30, 31].

Problem

A metal sphere of radius 50 nm is used as a capacitor to store charge, or information. What is the capacitance? What is the voltage change for each additional electron stored? What is the energy of the capacitor if a single electron is stored? If the capacitor is charged to a voltage of 1 V? Compare these energies to $k_B T$ at room temperature. How large is the electric field at the surface in both cases?