

## Part 7. Fundamental Limits in Computation

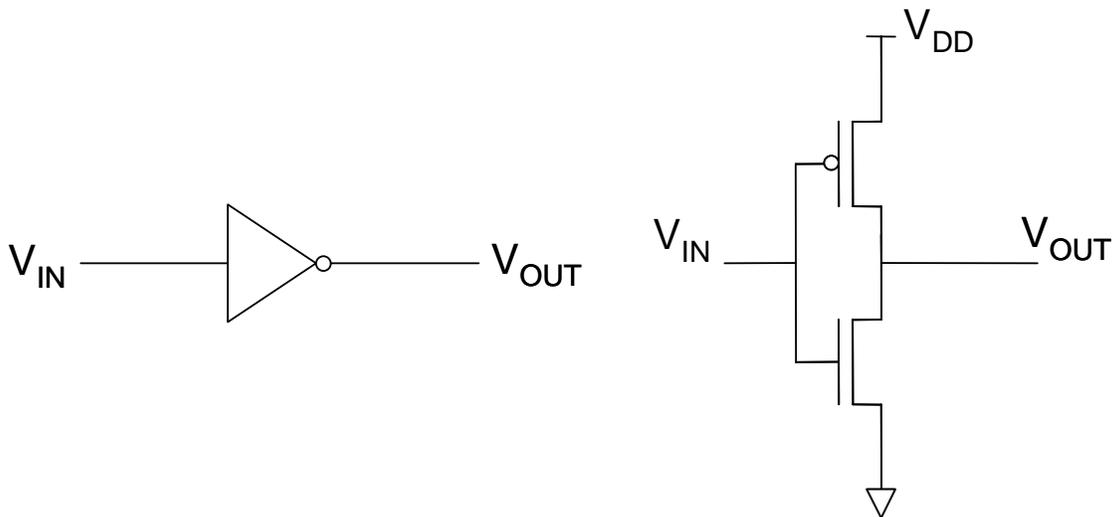
### Part 7. Fundamental Limits in Computation

This course has been concerned with the future of electronics, and especially digital electronics. At present digital electronics is dominated by a single architecture, Complementary Metal Oxide Semiconductor (CMOS), which is built on planar silicon field effect transistors. Steady improvements in the performance of CMOS circuits have been achieved by shrinking the feature sizes of the component transistors. This remarkable progress in electronics achieved over a period of  $> 30$  years has come to underpin much of our economic life.

In this section, we address both practical and thermodynamic limits to silicon CMOS electronics. It is likely that these limits will dominate the future of the electronics industry.

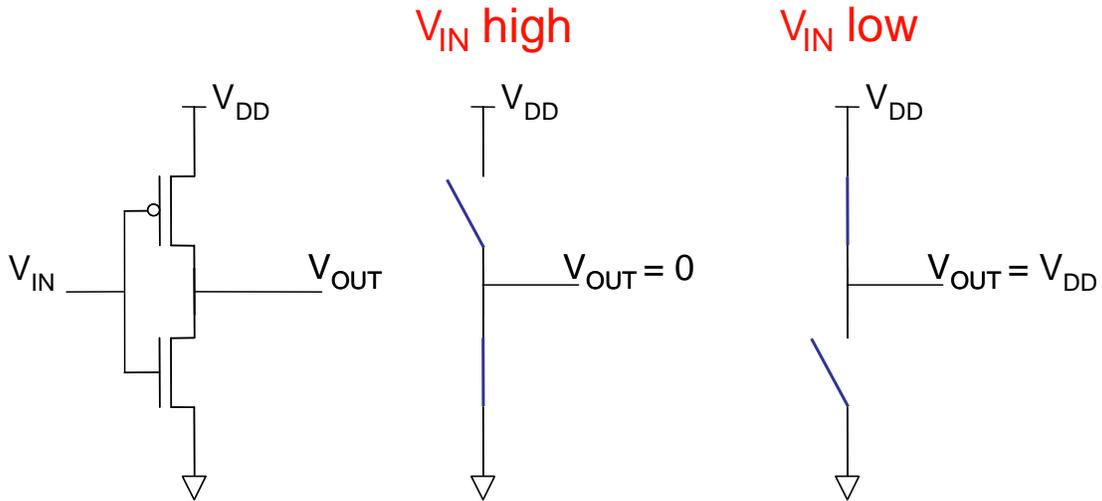
#### Speed and power in CMOS circuits

As you should remember from 6.002, the archetype CMOS circuit is shown in Fig. 7.1. It is composed of two complementary FETs: the upper MOSFET is off for a high voltage input, and the lower MOSFET is off given a low input. The circuit is an inverter.



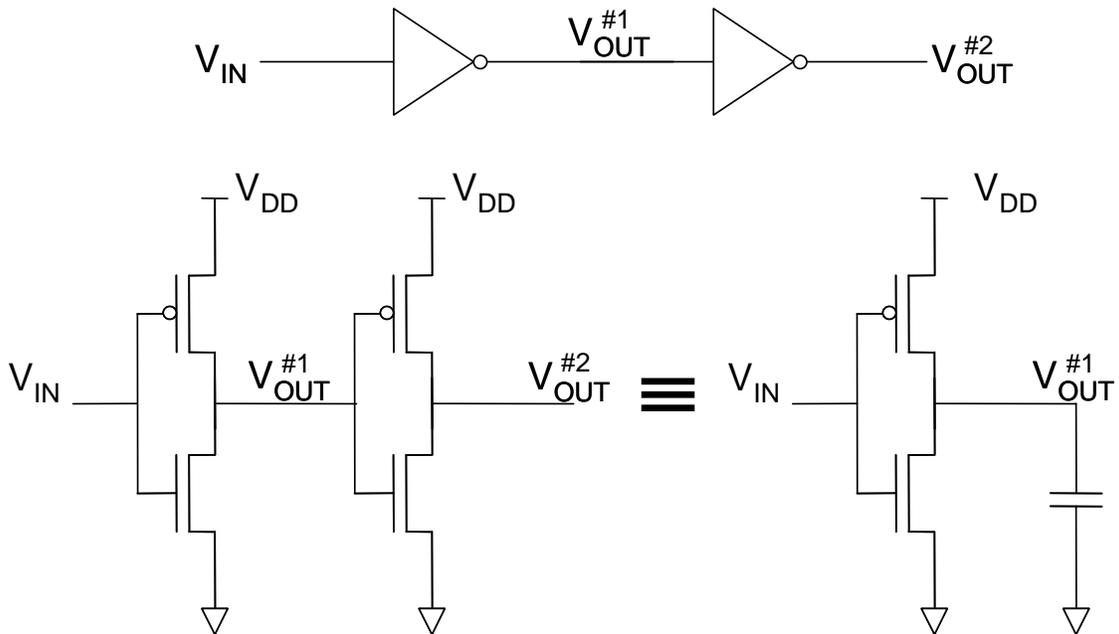
**Fig. 7.1.** A CMOS inverter consists of two complementary MOSFETs in series.

For constant voltage input, the circuit has two stable states, as shown in Fig. 7.2. Because one of the transistors is always off in steady state, the circuit ideally has no *static* power dissipation.



**Fig. 7.2.** The two steady state configurations of the inverter. No power is dissipated in either.

But when the input voltage switches the circuit briefly dissipates power. This is known as the *dynamic* power. We model the dynamics of a CMOS circuit as shown in Fig. 7.3. In this archetype CMOS circuit one inverter is used to drive more CMOS gates. To turn subsequent gates on an off the inverter must charge and discharge gate capacitors. Thus, we model the output load of the first inverter by a capacitor.



**Fig. 7.3.** Cascaded CMOS inverters. The first inverter drives the gate capacitors of the second inverter. To examine the switching dynamics of the first inverter, we model the second inverter by a capacitor.

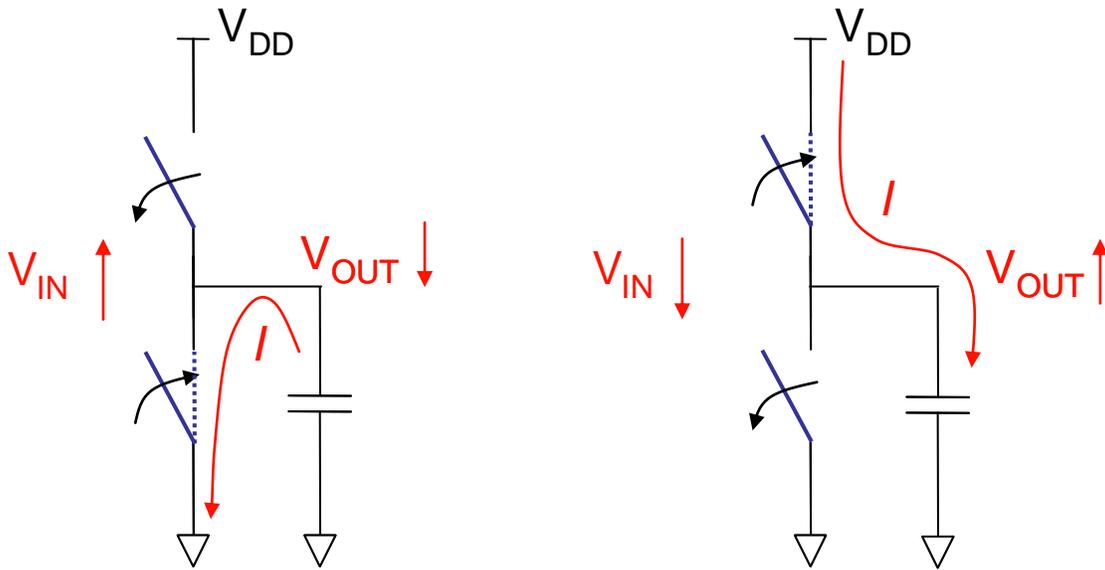
## Part 7. Fundamental Limits in Computation

We now consider the key performance characteristics of CMOS electronics.

### The Power-Delay Product (PDP)

The power-delay product measures the energy dissipated in a CMOS circuit per switching operation. Since the energy per switching event is fixed, the PDP describes a fundamental tradeoff between speed and power dissipation – if we operate at high speeds, we will dissipate a lot of power.

Imagine an input transition from high to low to the inverter of Fig. 7.1.



**Fig. 7.4.** Changes in the input voltage cause the output capacitor to charge or discharge dissipating power in the inverter.

If the output capacitor is initially uncharged, the energy dissipated in the PMOS FET is given by:

$$W = \int_0^{\tau/2} dt (V_{DD} - V_{OUT}) I \quad (7.1)$$

The current into the capacitor is given by:

$$I = C \frac{dV_{OUT}}{dt}, \quad (7.2)$$

Combining these expressions:

$$W = C \int_0^{\tau/2} dt (V_{DD} - V_{OUT}) \frac{dV_{OUT}}{dt} = C \int_0^{V_{DD}} dV_{OUT} (V_{DD} - V_{OUT}) = \frac{1}{2} C V_{DD}^2. \quad (7.3)$$

Similarly, in the second half of the cycle, when the capacitor is discharged through the NMOS FET, it is straightforward to show that again  $W = \frac{1}{2} C V_{DD}^2$ . Thus, the energy dissipated per cycle is:

$$PDP = C V_{DD}^2. \quad (7.4)$$

## Switching Speed

The dynamic model of Fig. 7.4 relates the switching speed to the charging and discharging time of the gate capacitor.

$$f_{max} = \frac{I}{CV_{DD}} \quad (7.5)$$

Thus, switching speed can be improved by

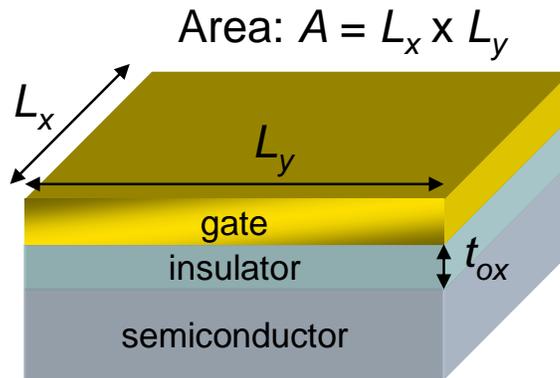
- (i) increasing the on current of the transistors
- (ii) decreasing the gate capacitance by scaling to smaller sizes
- (iii) decreasing the supply voltage (thereby decreasing the voltage swing during charge/discharge cycles)

## Scaling Limits in CMOS

Equation (7.4) demonstrates the importance of the gate capacitance. The capacitance is

$$C = \frac{\epsilon A}{t_{ox}} \quad (7.6)$$

where  $A$  is the cross sectional area of the capacitor,  $t_{ox}$  is the thickness of the gate insulator and  $\epsilon$  is its dielectric constant.



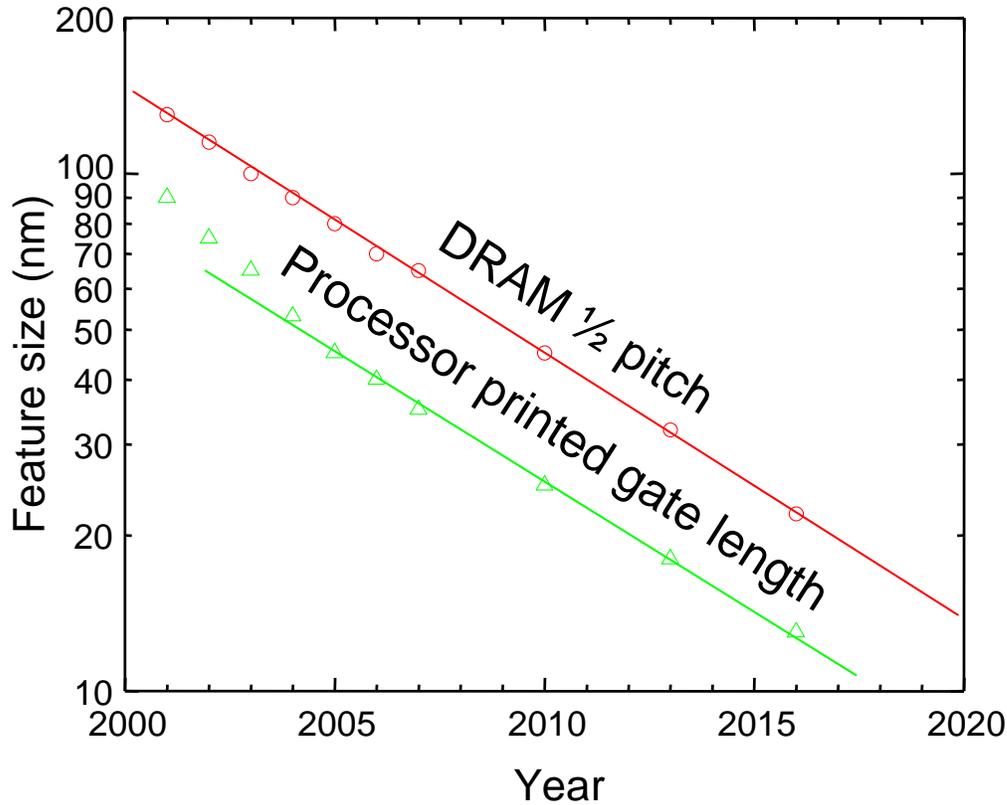
**Fig. 7.5.** The dimensions of a gate capacitor.

Now, if we scale all dimensions down by a factor  $s$  ( $s < 1$ ), the capacitance decreases:

$$C(s) = \frac{\epsilon s^2 A}{s t_{ox}} = s C_0 \quad (7.7)$$

From Eq. (7.4), reductions in  $C$  reduce the PDP, allowing circuits to run faster for a given power dissipation. Indeed, advances in the performance of electronics have come in large part through a continued effort of engineers to reduce the size of transistors, thereby reducing the capacitance and the PDP; see Fig. 7.6.

## Part 7. Fundamental Limits in Computation



**Fig. 7.6.** The semiconductor roadmap predicts that feature sizes will approach 10 nm within 10 years. Data is taken from the 2002 International Technology Roadmap for Semiconductors update.

At present, however, there are increasing concerns that we are approaching the end of our ability to scale electronic components. There are at least two looming problems in electronics:

(i) *Poor electrostatic control.*

We saw in part 5 that gate control over charge in the channel requires  $t_{ox} \ll L$ , where  $L$  is the channel length. Now as the channel length,  $L \rightarrow 10$  nm,  $t_{ox} \rightarrow 1$  nm, i.e. the gate insulator is only several atoms thick! But the electric field across the gate must remain high to induce charge in the channel. Thus, reductions in feature sizes will eventually place severe demands on the gate insulator.

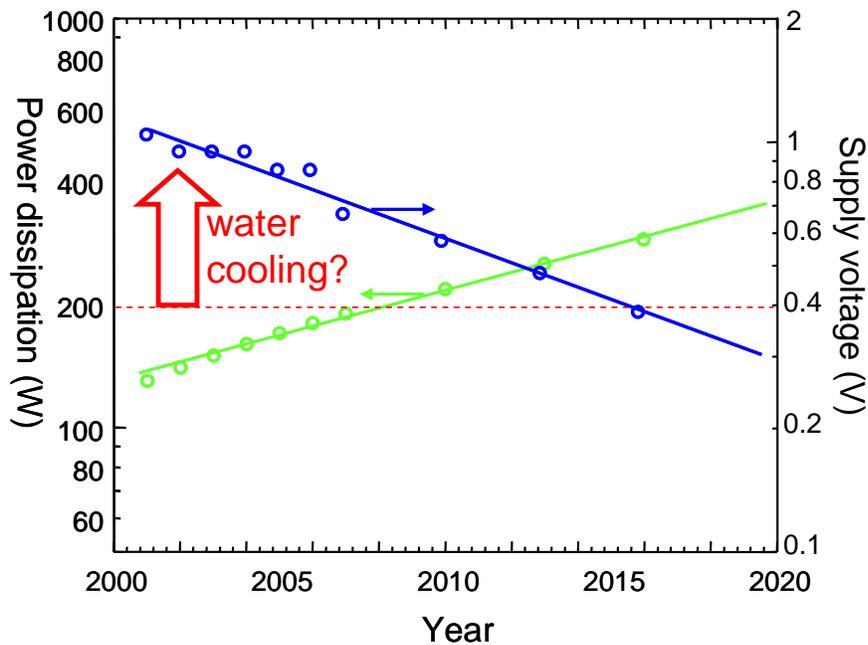
(ii) *Power density*

The electrostatic problem is fundamental, but it is possible that power concerns may obstruct the scaling of CMOS circuits prior to the onset of electrostatic issues. Power density is a particular concern since it does not benefit from continued reductions in component size. If the dimensions of a MOSFET are scaled down by a factor  $s$  ( $s < 1$ ),  $C \propto s$  (recall that capacitance is proportional to cross sectional area, and inversely

## Introduction to Nanoelectronics

proportional to the spacing between the charges). But even if the PDP scales as  $s$ , the power *density* may increase because the number of devices per unit area increases as  $1/s^2$ .

The power densities of typical integrated circuits are approaching those of a light bulb filament ( $\sim 100 \text{ W/cm}^2$ ). For comparison, the power density of the surface of the sun is  $\sim 6000 \text{ W/cm}^2$ . Removal of the heat generated by an integrated circuit has become perhaps *the* crucial constraint on the performance of modern electronics. Indeed, the fundamental limit to power density appears to be approximately  $1000 \text{ W/cm}^2$ . In practice, using water cooling of a uniformly heated Si substrate with embedded micro channels, a power density of  $790 \text{ W/cm}^2$  has been achieved with a substrate temperature near room temperature.



**Fig. 7.7.** The semiconductor roadmap predicts that supply voltages will drop to nearly 0.4V within 10 years. Power dissipation per chip is expected to increase to above 200W by 2008. It is expected that power dissipation in the shaded region will require significantly more expensive cooling systems. Data is taken from the 2002 International Technology Roadmap for Semi-conductors update.

As is evident from Eq. (7.4) above, the PDP also depends on the supply voltage  $V_{DD}$ . Ensuring that the total power dissipated per chip  $\ll 200 \text{ W}$  has driven  $V_{DD}$  from 5V in early CMOS circuits to nearly 1V today. If the industry conforms to roadmap predictions, the supply voltage will eventually reach 0.4V by 2016.

But what is the ultimate limit to the PDP?

## Part 7. Fundamental Limits in Computation

### Brief notes on information theory and the thermodynamics of computation

We now examine the thermodynamics of computation.

#### (i) Minimum energy dissipated per bit

Assume we have a system, perhaps a computer, with a number of possible states. The uncertainty, or *entropy* of the computer is a measure of the number of states. Recall from thermodynamics that the Boltzmann-Gibbs entropy of a physical system is defined as

$$S = -k_B \sum_{i=1}^N p_i \ln p_i, \quad (7.8)$$

where the system has  $N$  possible states, each with probability  $p_i$ , and  $k_B$  is the Boltzmann constant.

The opposite of entropy and uncertainty is information. When the uncertainty of the system decreases, it gains information.

Now, the second law of thermodynamics can be restated as “all physical processes increase the total entropy of the universe”. Let’s separate the universe into the computer, and everything else. The corresponding entropy of each system is given by

$$S_{universe} = S_{computer} + S_{everythingelse}. \quad (7.9)$$

Thus, thermodynamics requires

$$\Delta S_{universe} \geq 0. \quad (7.10)$$

It follows that

$$\Delta S_{everythingelse} \geq -\Delta S_{computer}, \quad (7.11)$$

*i.e.* if the information within a computer increases during a computation, then the entropy decreases. This change in entropy within the computer must be at least balanced by an increase in the entropy of the remainder of the universe. The increase in entropy in the remainder of the universe is obtained by dissipating heat,  $\Delta Q$ , from the computer.

According to thermodynamics the heat dissipated is

$$\Delta Q = T \Delta S_{everythingelse} \geq -T \Delta S_{computer} \quad (7.12)$$

Uncertainty and entropy can also be measured in bits. For example, how many bits are required to describe the computer with  $N$  states?

$$2^H = N. \quad (7.13)$$

Here,  $H$  is known as the **Shannon entropy**. If the states are equally probable, with probability  $p = 1/N$ , then the uncertainty reduces to:

$$H = \log_2 N = -\log_2 p. \quad (7.14)$$

Or more generally, if each state of the computer has probability  $p_i$ .

$$H = \langle -\log_2 p_i \rangle = -\sum_{i=1}^N p_i \log_2 p_i \quad (7.15)$$

Comparing Eq. (7.8) with Eq. (7.15) and noting that  $\ln p_i = (\ln 2) \log_2 p_i$  gives

## Introduction to Nanoelectronics

$$\Delta Q = -k_B T \ln(2) \Delta H_{\text{computer}} \quad (7.16)$$

The heat must ultimately come from the power supply. Thus, the minimum energy required per generation of one bit of information is:

$$E_{\min} = k_B T \ln(2). \quad (7.17)$$

This minimum is known as the **Shannon-von Neumann-Landauer (SNL) limit**.

### (ii) Energy required for signal transmission

Recall Shannon's theorem for the capacity,  $c$ , in bits per second, of a channel in the presence of noise.

$$c = b \log_2 \left( 1 + \frac{s}{n} \right), \quad (7.18)$$

where  $s$  and  $n$  are the signal and noise power, respectively, and  $b$  is the bandwidth of the channel. The noise in the channel is at least  $n = bk_B T$ .

The energy required per bit transmitted is:

$$E_{\min} = \lim_{s \rightarrow 0} \left\{ \frac{s}{c} \right\} = \lim_{s \rightarrow 0} \left\{ \frac{s}{b \log_2 (1 + s/n)} \right\}. \quad (7.19)$$

L'Hôpital's rule gives

$$E_{\min} = k_B T \ln(2). \quad (7.20)$$

consistent with the previous calculation of  $E_{\min}$ .

### (iii) Consequences of $E_{\min}$

It has been argued that since the uncertainty in energy,  $\Delta E$ , within an individual logic element can be no greater than  $E_{\min}$ , we can apply the Heisenberg uncertainty relations to a system operating at the SNL limit to determine the minimum switching time, *i.e.*<sup>†</sup>

$$\Delta E \Delta t \geq \hbar \quad (7.21)$$

Eq. (7.21) gives a minimum switching time of

$$\tau_{\min} = \frac{\hbar}{\Delta E} = \frac{\hbar}{k_B T \ln(2)} = 0.04 \text{ ps} \quad (7.22)$$

Assuming that the maximum power density that we can cool is  $P_{\max} \sim 100 \text{ W/cm}^2$ , the maximum integration density is

$$n_{\max} = \frac{P_{\max}}{E_{\min} / \tau_{\min}} = \frac{\hbar P_{\max}}{E_{\min}^2} \quad (7.23)$$

At room temperature, we get  $n_{\max} \sim < 10^{10} \text{ cm}^{-2}$ , equivalent to a switch size of 100 x 100 nm. This is very close to the roadmap value for 2016.

At lower temperatures, the power dissipation *on chip* is decreased, but the overall power dissipation actually increases due to the requirement for refrigeration.<sup>4</sup> Since the

---

<sup>†</sup> This argument, due to Zhirnov, *et al.* "Limits to Binary Logic Switch Scaling - A Gedanken Model", *Proceedings of the IEEE* **91**, 1934 (2003), has been used to argue that end of the roadmap Si CMOS is as good as charge based computing can get.

## Part 7. Fundamental Limits in Computation

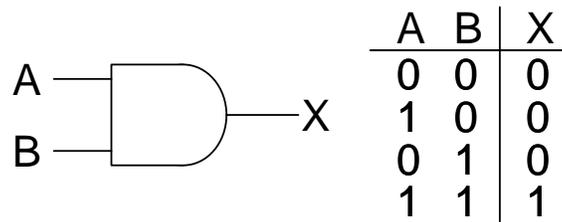
engineering constraint is likely to be on chip power dissipation – refrigeration may be one method for further increasing the density of electronic components.

### Reversible computers

In the previous section, we defined computation as a process that increases information and decreases uncertainty. But if uncertainty (i.e. entropy) decreases within the computer, entropy must increase outside the computer. This is an application of the second law of thermodynamics, which states that all physical systems can only increase entropy over time.

Of all physical laws, the second law of thermodynamics is famous for defining the ‘arrow of time’. The implication of the second law is that *computation is irreversible*, at least if the computation changes uncertainty.

For example, let’s consider a two input AND gate. If one of the inputs to the AND gate is a zero, then the information in the other input is thrown away. Thus, the total number of states decreases when the inputs propagate to the output of an AND gate. Consequently, entropy decreases, heat is dissipated and AND gates are not reversible.



**Fig. 7.8.** AND gates are not reversible. If the output is zero, the inputs cannot be reconstructed.

The heat dissipated in the AND gate is calculated as follows. There are four possible input states. Assuming each is equi-probable the Shannon entropy is

$$H_{in} = -\log_2 \frac{1}{4} = 2 \text{ bits} \quad (7.24)$$

There are two possible output states. The probability of the output  $X = 0$  is  $\frac{3}{4}$  and the probability of  $X = 1$  is  $\frac{1}{4}$ .

$$H_{out} = -\frac{3}{4} \log_2 \frac{3}{4} - \frac{1}{4} \log_2 \frac{1}{4} \approx 0.811 \text{ bits} \quad (7.25)$$

Thus,

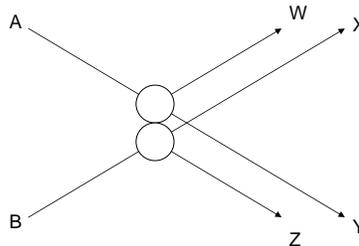
$$\Delta E = -k_B T \ln(2) \Delta H \approx 3.4 \times 10^{-21} \text{ J} \quad (7.26)$$

But what if we designed a gate that did not throw away states during the computation? Such a system would be reversible, and more importantly it would not need to dissipate energy.

In fact, several reversible logic elements have been proposed. Perhaps the best known irreversible computer is the billiard ball computer pioneered by Fredkin.

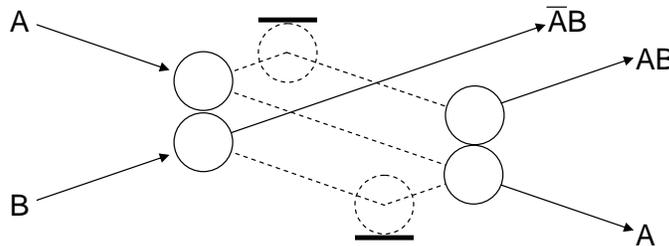
## Introduction to Nanoelectronics

An example of a billiard ball logic gate is shown in Fig. 7.9. Billiard balls are fired into the logic gate from positions A and B. If there is a collision, the balls are deflected to positions W and Z. If one ball is absent, however, an output at either X or Y is generated. We also need to assume that the balls obey the laws of classical mechanics; there is no friction and the collisions are perfectly elastic. Note that the number of states in a billiard ball logic elements does not change – the billiard balls are neither created nor destroyed.



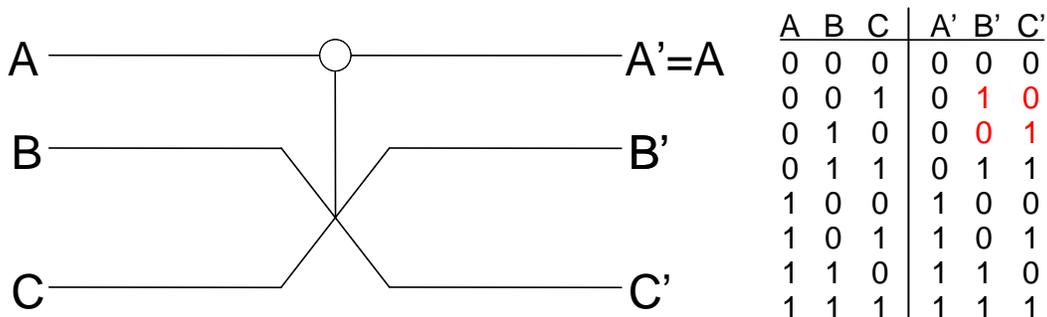
**Fig. 7.9.** A two ball collision gate. After Feynman, *Lectures on Computation*. Editors A.J.G. Hey and R.W. Allen, Addison-Wesley 1996.

More complex devices are possible by adding ‘redirection gates’ (walls). For example, Fig. 7.10 shows a switch made from collision and redirection gates.



**Fig. 7.10.** A billiard ball switch. After Feynman, *Lectures on Computation*. Editors A.J.G. Hey and R.W. Allen, Addison-Wesley 1996.

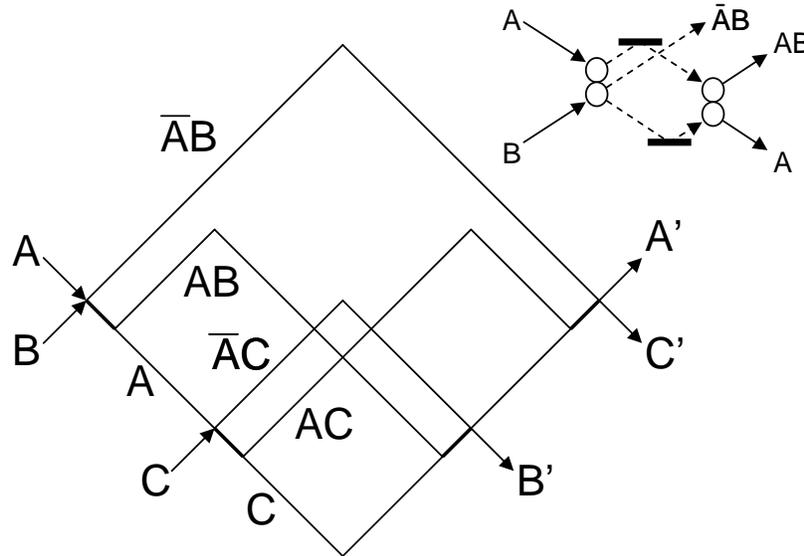
But given that many logic gates such as the AND gate are inherently non-reversible, the question arises: Can an arbitrary algorithm be implemented entirely from reversible elements? The answer is yes. Reversible computers can be constructed entirely of a fundamental reversible element known as the Fredkin gate, shown in Fig. 7.11.



**Fig. 7.11.** The symbol for the Fredkin gate. A is unchanged. If  $A = 0$  then B and C switch. If  $A = 1$  then B and C remain unchanged. All logic elements may be formulated from reversible Fredkin gates. After Feynman, *Lectures on Computation*. Editors A.J.G. Hey and R.W. Allen, Addison-Wesley 1996.

## Part 7. Fundamental Limits in Computation

An implementation of a Fredkin gate with billiard balls is shown in Fig. 7.12.



**Fig. 7.12.** A Fredkin gate constructed from four billiard ball switches. After Feynman, *Lectures on Computation*. Editors A.J.G. Hey and R.W. Allen, Addison-Wesley 1996.

### Reversible computers and noise

Reversible computers, however, remain extremely controversial in engineering circles. The catch is noise. Shannon's theorem, for example, requires  $E_{min} = k_B T \ln(2)$  for the transmission of one bit of information in a noisy channel. This applies even in a reversible system such as the billiard ball collision gate. In fact, billiard ball gates are extremely sensitive to errors. Given a slight error in the trajectory or timing of one ball and a billiard ball computer would accrue a large number of errors.

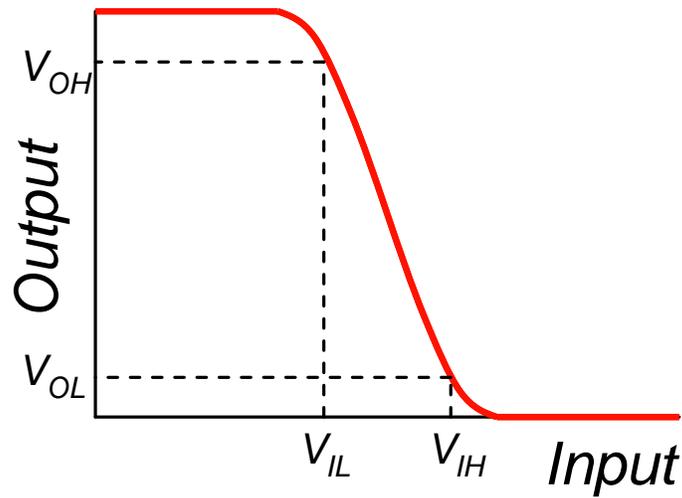
A billiard ball computer could be made more robust and noise resistant by including trenches to guide the balls. But the trench guides the balls by dissipating that component of the ball's momentum that would otherwise drive it off its designed trajectory. Thus, the trenches inevitably lead to energy dissipation.

In contrast, let's briefly look at noise in CMOS circuits. The transfer function of a CMOS inverter is shown in Fig. 7.13. We see that close to the switching voltage, the inverter has very large gain,  $A_V$ :

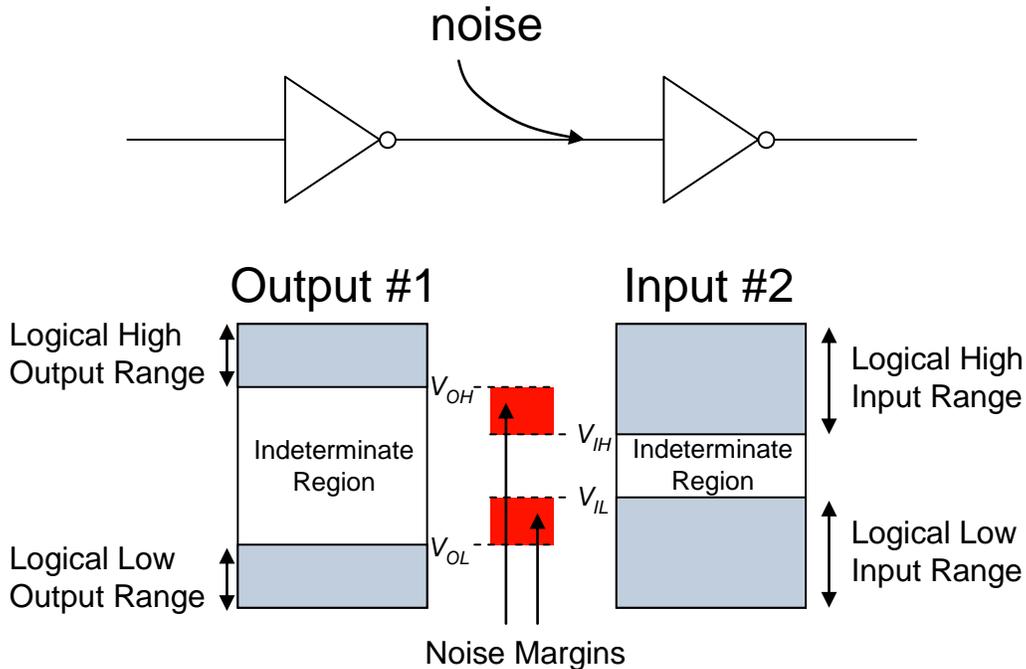
$$A_V = \frac{dV_{out}}{dV_{in}} \gg 1 \quad (7.27)$$

The gain protects the inverter against noise. For example, consider two cascaded inverters. Assume some noise is added to the output of the first inverter. The *noise margin* tells us the minimum amount of noise required to cause an error at the output of the second inverter; see Fig. 7.14.

Thus, many device engineers argue that without gain no computation system is practical. And since reversible computers do not dissipate power it is not clear how they can amplify a signal, rendering them always subject to the adverse effects of noise.



**Fig. 7.13.** Transfer characteristics of a CMOS inverter.  $V_{IL}$  and  $V_{IH}$  are defined as the threshold of low and high inputs, respectively. Note that the large gain means that  $V_{OL} < V_{IL}$  and  $V_{OH} > V_{IH}$ , helping protect signal integrity against the effects of noise.



**Fig. 7.14.** The noise margin in a digital circuit is the minimum input noise voltage required to cause an error at the output of the next gate. The greater the gain, the greater the noise margin.

## Part 7. Fundamental Limits in Computation

### The future of electronics?

The immediate path is clear: we have not yet reached the limits of scaling, or the fundamental limits of field effect transistors. The electronics industry will push to smaller length scales to minimize the power delay product. It will also seek to exploit ballistic conduction in low dimensional materials, thereby increasing switching speeds.

It is realistic to expect that a future MOSFET might possess:

- (i) ballistic transport and operation at the quantum limit of conductance
- (ii) switching on and off at the optimum FET subthreshold slope of  $kT/q$
- (iii) scaling of all dimensions with a gate insulator thickness of  $\sim 1$  nanometer

Traditionally, substantial materials development efforts have been devoted to improving the mobility of transistor channels. But because devices are already at the ballistic limit, the electrostatic design of nanotransistors will be a likely focus of materials development. We have seen that good electrostatic control of the channel can be achieved by maximizing the gate capacitance. For example, with a nanowire channel, the gate could be implemented as a concentric ring. Or a channel that consists of a single atomic layer (such as a grapheme sheet) might be preferable from the electrostatic viewpoint to a thicker layer of silicon, even though both will operate at the ballistic limit. Manufacturing such advanced structures may require a substantial amount of further development.

Beyond this, there appears to be only one major weakness of conventional FET technologies. There is a strong possibility that new technologies will demonstrate subthreshold slope far superior to  $kT/q$ . As we have seen, this will allow for dramatic reductions in operating voltage, and hence significantly lower power dissipation.

From a fundamental viewpoint, all transistors that operate in thermodynamic equilibrium, must exhibit an energy difference between their ON and OFF states. For example, the potential energy difference between the ON and OFF states of a FET is  $\Delta E = \frac{1}{2}CV^2$ , which can also be expressed as  $\Delta E = \frac{1}{2}QV$ , where  $Q$  is the total charge on the gate capacitor and  $V$  is the supply voltage. The fundamental limit in the OFF state current is the probability of thermal excitation from the OFF state to the ON state. That is:

$$I_{OFF} = I_{ON} \exp\left[-\frac{1}{2}QV/kT\right] \quad (7.28)$$

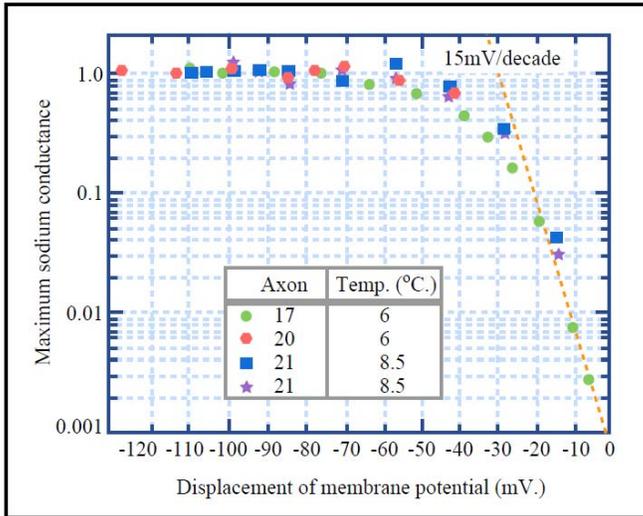
where  $I_{ON}$  is the maximum current associated with the ON state. But as we have seen, modern FETs do not operate at this limit because each electron in the channel is independent. In contrast to Eq. (7.28), the FET follows:

$$I_{OFF} = I_{ON} \exp[-qV/kT] \quad (7.29)$$

Except for a FET that operates with a single electron in the channel, the difference is substantial: a subthreshold slope of  $kT/Q$  versus  $kT/q$ . Indeed, at present transistor dimensions  $Q \gg 10^3 q$ .

**So how can we approach the subthreshold limit?**

It is thought that if all the charge in the channel behaves *collectively*, (*i.e.* all or none of the charge contributes to current) then it might be possible to switch closer to the limit. Perhaps the best examples of this principle are the voltage-dependent ion channels of biology, in which conformation changes may enable subthreshold slopes as sharp  $\approx 10$  mV/decade.<sup>†</sup>



**Fig. 7.15.** The switching characteristics of voltage gated sodium ion channels from the giant squid axon. Note the extremely sharp switching characteristics. Reproduced from Hodgkin and Huxley’s classic 1952 series of papers.

Figure by MIT OpenCourseWare.

Hodgkin and Huxley, *J. Physiol.* **116**, 449 (1952a)

Below, we show the structure and mechanism of the mechanical change in a voltage dependent K<sup>+</sup> ion channel, as determined by MacKinnon, *et al.*<sup>§</sup> The channels sit in a membrane; when open they allow the diffusion of ions from one side of the membrane to the other.

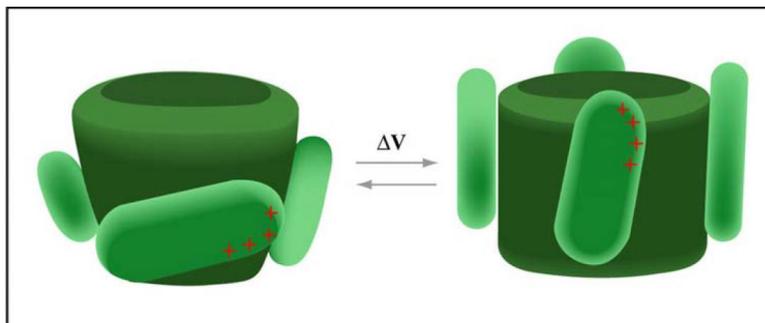


Figure by MIT OpenCourseWare.

**Fig. 7.16.** The voltage dependent K<sup>+</sup> ion channel has 4 charged paddles that rotate in an electric field, opening and closing a mechanical gate at the base of the channel. Reproduced from MacKinnon, *et al.*

<sup>†</sup> Hodgkin and Huxley, *J. Physiol.* **116**, 449 (1952a)

<sup>§</sup> Y. Jiang, A. Lee, J. Chen, V. Ruta, M. Cadene, B.T. Chait and R. MacKinnon. *Nature.* **423**. 33-41 (2003)

## Part 7. Fundamental Limits in Computation

Consider a membrane where there are  $N$  closed channels and  $N^*$  open channels. The ratio of open to closed channels is determined by the Boltzmann relation:

$$\frac{N^*}{N} = \exp\left[-\frac{U_{open} - U_{closed}}{kT}\right] \quad (7.30)$$

where  $U_{open}$  and  $U_{closed}$  are the energies of the open and closed conformations respectively. Under an electric field, we assume that  $Z$  charges move through a potential of  $\Delta V$ , *i.e.*:

$$U_{open} = U_{closed} - Zq\Delta V. \quad (7.31)$$

The current through the ion channel is proportional to the number of open channels,  $N^*$ .

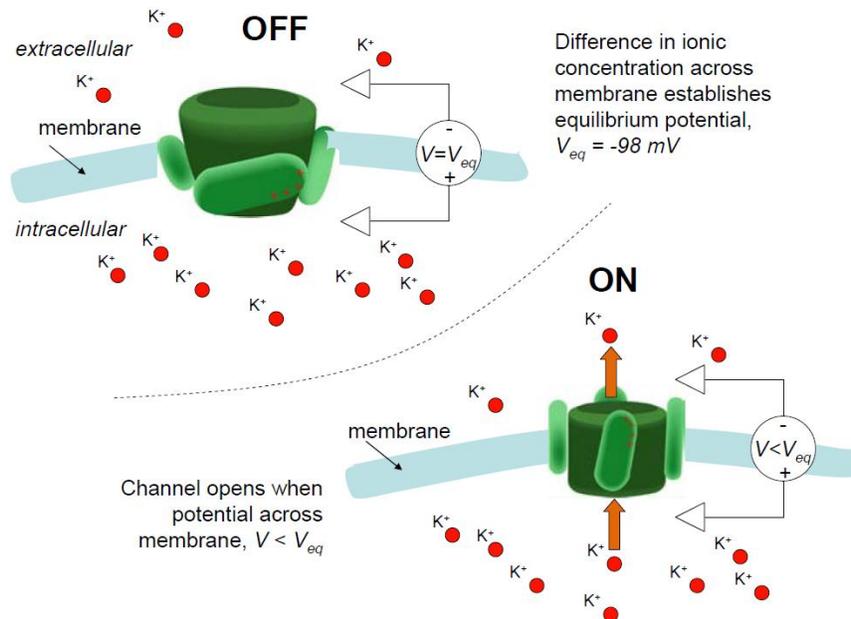
$$I \propto N^* \quad (7.32)$$

Since  $N + N^*$  is a constant

$$I \propto \frac{N^*}{N + N^*} \approx \frac{N^*}{N} = \exp\left[\frac{Zq\Delta V}{kT}\right] \quad (7.33)$$

That is, the subthreshold slope is sharpened by a factor,  $Z$ , the effective<sup>†</sup> number of charges on the movable paddles.

$$\frac{\Delta V}{\log_{10} I} = \frac{kT}{Ze} \frac{1}{\log_{10} e} \approx \frac{60}{Z} \text{ mV/decade}. \quad (7.34)$$



**Fig. 7.17.** Ion channels modulate the diffusion of ions through a membrane. The direction of ion current is determined by the concentration gradient. Typically, the ion channel preferentially passes ions of a particular size and charge. When it is open, the channel illustrated above selectively allows K<sup>+</sup> ions to diffuse.

<sup>†</sup> Note that the effective number of charges is usually less than the actual number of charges on the movable structures in the ion channel because the charges are not usually free to move through the full potential  $\Delta V$  across the membrane (the motion of the paddles is somewhat restricted).

## Introduction to Nanoelectronics

The conclusion is that transistors are possible with subthreshold characteristics superior to those of conventional FETs. The ion channel shows that mechanically-coupling the charges together is one path to achieving the *collective* behavior that we desire. But the reliability of mechanical devices is questionable. Instead, it is possible that another collective phenomenon, like the switching of a magnetic domain in a ferromagnet, may be exploited to improve switching.

### And beyond?

Researchers are currently pursuing a few ideas:

#### 1. *Reversible computing*

The absence of power dissipation makes this a big prize, but concerns remain as to its noise immunity and fundamental practicality.

#### 2. *New information tokens*

Transistors today use electrons to carry information. Instead, we might seek to use a different information token such as the spin of an electron or position in a mechanical switch. A change in information token could revolutionize electronics. But at present it is not clear what, for example, a spin-in spin-out transistor might look like, nor do we have a clear idea of the potential benefits of spin-based technology. For example, could it escape the Shannon-Von Neumann- Landauer limit?

#### 3. *Integration*

More transistors per chip have traditionally meant more computing power. If we can't make transistors any smaller, perhaps we could shift to three dimensional circuits? A transition from two to three dimensional circuits could massively increase integration densities. But apart from the difficulty of fabricating such structures, we must also figure out how to cool them.

#### 4. *Architecture*

The computing power of the brain clearly demonstrates the virtue of different approaches to certain problems such as pattern recognition. But it is not clear that our current model of electronics is suited to say, a shift to a neural network type architecture.

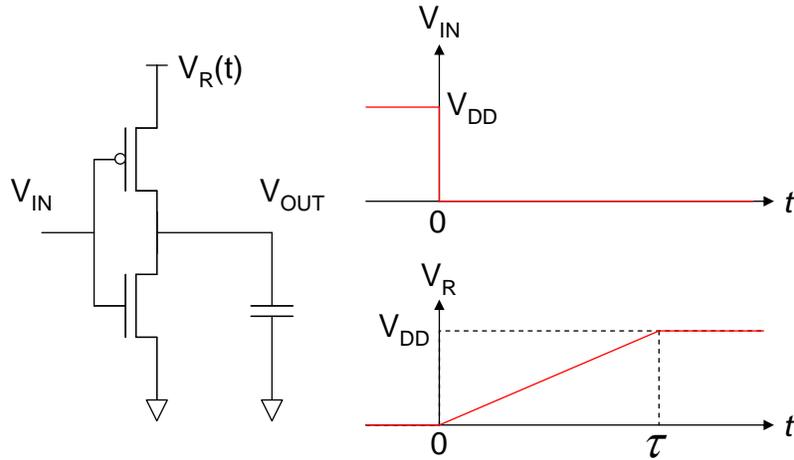
Whatever happens the stakes are high. As we approach the limits of CMOS, slow technological progress may reduce the need to update computers every few years. But the economic model of the electronics industry has come to rely on rapid technological change. Consequently, the rewards may be especially great for the next revolution in electronics technology.

## Part 7. Fundamental Limits in Computation

### Problems

#### Q1. Adiabatic Transistors

Consider the inverter shown below.



**Fig. 7.18.** An adiabatically-driven inverter.

Unlike in a conventional CMOS inverter, in this device, the supply voltage,  $V_R$ , adjusts during the switching operation. Initially the voltage on the output capacitor is zero, but at  $t = 0$  the input voltage drops to zero. Also at  $t = 0$ , the supply voltage ramps from zero to the logic high voltage,  $V_{DD}$ .

Assume that the PMOS FET is modeled by a resistor,  $R$ .

(a) Show that the energy dissipated during the switching operation is

$$E = \frac{RC}{\tau} CV_{DD}^2 \text{ for } \tau \gg RC.$$

This is known as an adiabatic switch, since switching occurs (in the limit) with no energy dissipation, i.e. we are adding charge to a capacitor using a vanishingly small excess voltage.

[Hint: You may assume  $V_{OUT}$  of the form  $V_{OUT} = a + b \exp[-t/RC] + ct$  where  $a$ ,  $b$ , and  $c$  are constants to be determined.]

(b) Show also that the energy dissipated reduces to the standard CMOS switching energy

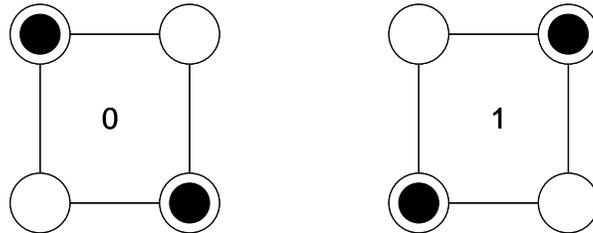
$$E = \frac{CV_{DD}^2}{2} \text{ for } \tau \ll RC.$$

(c) The above example shows adiabatic switching when the capacitor voltage changes from low to high. Can it be implemented generally? i.e. consider the case when the capacitor voltage changes from high to low. And what happens when the capacitor does not change voltage during a cycle?

## Q2. Cellular Automata

This question refers to a proposed architecture for molecular electronics: *Molecular Quantum-Dot Cellular Automata*. The figures are drawn from the reference.

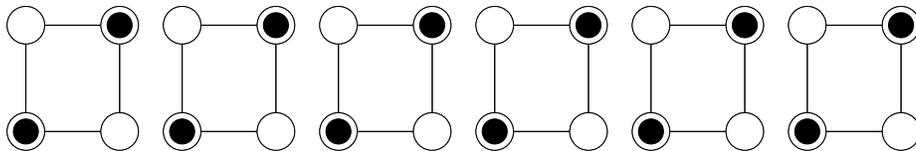
In this architecture information is stored in bistable cells. An example cell is shown below:



**Fig. 7.19.** A bistable cell for use in a cellular automata computer.

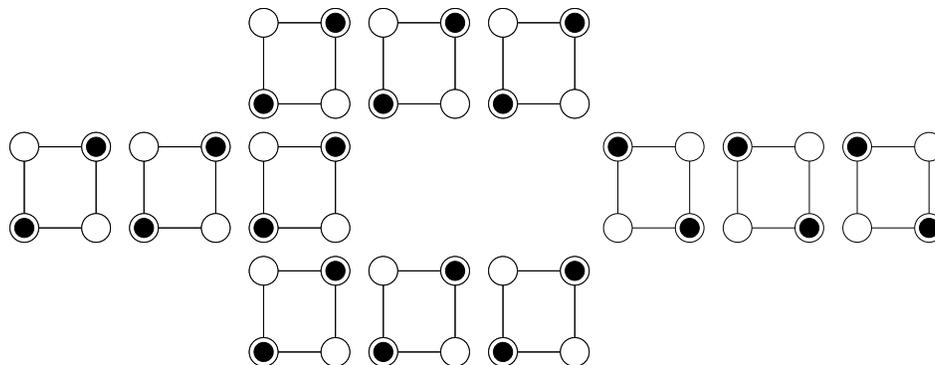
This cell consists of four electron traps positioned at the corners of a square. Only two of the traps are filled. From electrostatics, there are two stable states with the electrons at opposing corners of the square.

To transmit information, the cells are placed in a line. Information then propagates electrostatically, without current flow. It is argued that power dissipation is therefore eliminated and no interconnecting wires are required.



**Fig. 7.20.** A cellular automata wire.

By changing the topology, it is possible to make logic gates. For example, below we show an inverter.



**Fig. 7.21.** A cellular automata inverter.

*Question continued on next page....*

## Part 7. Fundamental Limits in Computation

(a) A proposed ‘majority gate’ is shown below. The output  $Z$  is the majority of the inputs,  $A$ ,  $B$  and  $C$ . *i.e.* if there are more 1 inputs than zero inputs then  $Z = 1$ , otherwise  $Z = 0$ . Use this gate to design a two input AND gate.

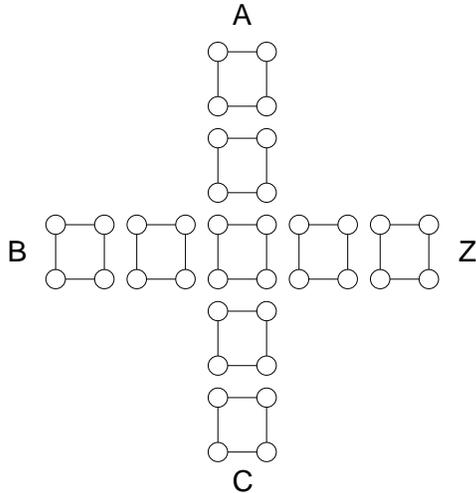


Fig. 7.22. A proposed majority gate.

(b) Is the majority gate truly dissipationless? Hint: calculate the entropy before and after a majority decision.

**Reference:** Lent, “Bypassing the transistor paradigm” Science **288** 1597 (2004)

### Q3. Power delay products at the nanoscale

The power delay product is the minimum energy dissipated per bit of information processed. For a CMOS inverter the PDP is:

$$PDP = CV^2$$

where  $V$  is the supply voltage and  $C$  is the load capacitance as seen by the inverter. In this question, we will assume that the supply voltage is fixed.

(a) Determine the load capacitance as a function of the gate and quantum capacitances. Assume we can neglect all other capacitances.

(b) Consider a 2d field effect transistor (where  $C_Q \rightarrow \infty$ ). If its dimensions are scaled by a factor  $s$ , how does the PDP scale?

(c) Now consider a quantum *wire* field effect transistor with  $C_Q \ll C_G$ . Its gate capacitance is given by

$$C_G = 2\pi\epsilon \frac{l}{\log(r/a_0)}$$

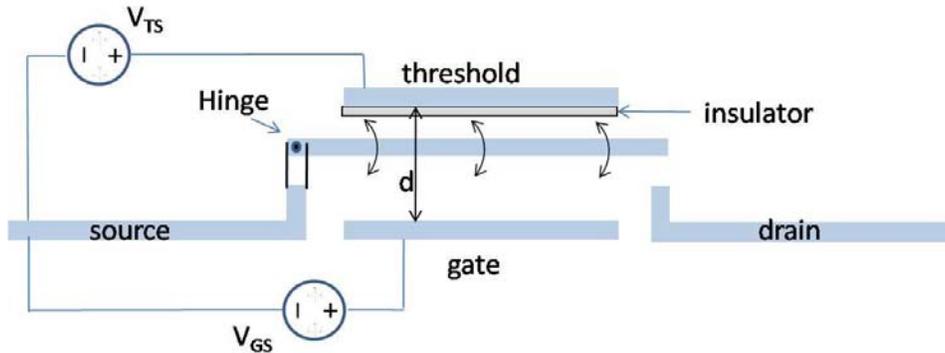
where  $\epsilon$  is the dielectric constant of the gate insulator,  $l$  is the gate length,  $r$  is the gate radius and  $a_0$  is the 1d wire radius.

Assume that  $l$  and  $r$  are scaled by a factor  $s$ , how does the gate capacitance for a quantum *wire* field effect transistor scale?

**(d)** Now consider the impact of the quantum capacitance on the PDP on the quantum *wire* field effect transistor. How does the overall PDP scale? Is the scaling faster or slower than the equivalent PDP using large quantum *well* field effect transistors?

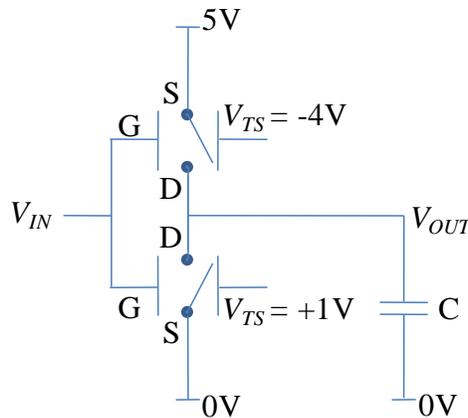
### Q4. Mechanical transistors

Consider a mechanical switch.



**Fig. 7.23.** A mechanical switch.

The conductor is pulled towards the gate electrode when  $|V_{GS}| > |V_{TS}|$ , switching the device on, and towards the threshold electrode when  $|V_{GS}| < |V_{TS}|$  switching the device off. Assume two switches are wired together in a complementary logic circuit that drives a capacitive load as shown below.



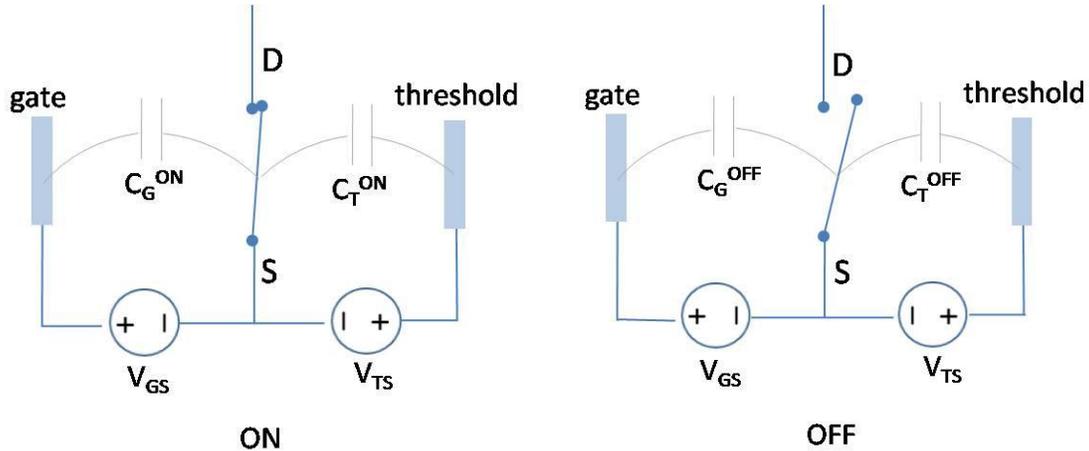
**Fig. 7.24.** A complementary logic circuit featuring mechanical switches.

**(i)** Plot steady state  $V_{OUT}$  versus  $V_{IN}$ , where  $V_{IN}$  ranges from 0 to 5V. Show that the circuit is complementary.

## Part 7. Fundamental Limits in Computation

(ii) Assume  $V_{IN}$  is switched from 0V to 5V and then back to 0V. How much energy is dissipated?

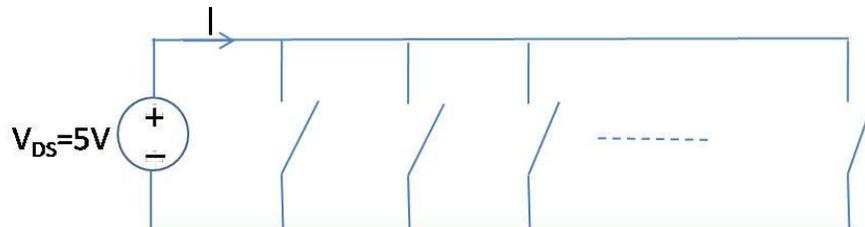
(iii) Consider one of the switches. Let  $C_T^{ON}$  and  $C_G^{ON}$  be the threshold-conductor capacitance, and the gate-conductor capacitance, respectively, in the ON state, and let  $C_T^{OFF}$  and  $C_G^{OFF}$  be the capacitances, respectively, in the OFF state. See the figure below.



**Fig. 7.25.** Capacitive models of the switch in the ON and OFF configuration.

What is the energy stored in these capacitors in the (a) ON and in the (b) OFF positions as a function of  $V_{GS}$  and  $V_{TS}$ ?

Now connect  $N$  switches all wired in parallel.



**Fig. 7.26.**  $N$  switches all wired in parallel.

Each switch has  $V_{TS} = +1V$  and resistance,  $R = 100\Omega$ . Assume all the gate electrodes are wired together at a potential  $V_{GS}$ . To simplify the analysis assume that  $C_G^{ON} \gg C_T^{ON}$  and also that  $C_T^{OFF} \gg C_G^{OFF}$ . Furthermore, take  $C_G^{ON} = C_T^{OFF} = C$ .

(iv) Considering Boltzmann statistics, and the potential energy difference between the OFF and ON states, out of the  $N$  switches, what is the probable number of switches that are ON as a function of  $C$ ,  $V_{GS}$  and  $V_{TS}$  when  $|V_{GS}| < |V_{TS}|$ ?

(v) Find  $I$  for the  $N$  switches as a function of  $V_{GS}$  and  $V_{TS}$  for  $0 < V_{GS} < 5V$  (for  $V_{TS} = 1V$ ).

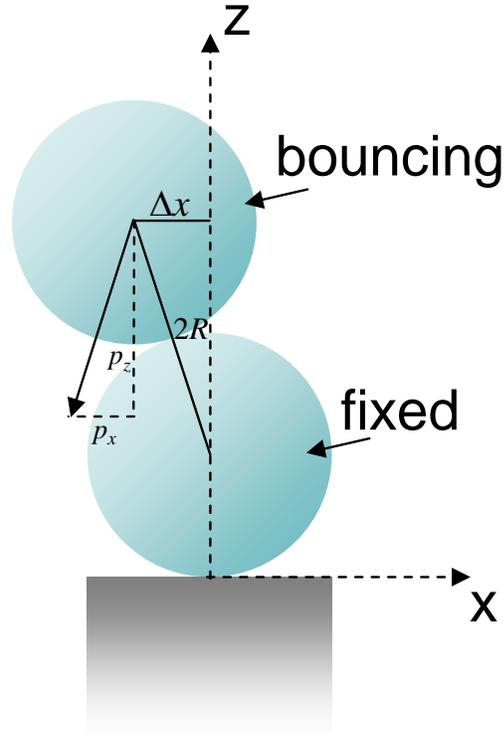
(vi) Does the mechanical switch exhibit any benefit over conventional CMOS?

**Q5. (a)** Consider two identical balls each 1cm in diameter and of mass  $m = 1g$ . One is kept fixed, and the second is dropped directly on it from a height of  $d = 10cm$ . From the uncertainty principle alone, what is the expected number of times the moving ball bounces on the stationary ball before it misses the latter ball altogether? Assume the ball is dropped from an optimal initial state.

**Hint:** some parts of this problem can be solved classically.

(b) Discuss the implications of (i) for billiard ball computers.

**Fig. 7.27.** An off-center collision between the fixed ball and the bouncing ball.



**Q6.** The following question refers to ion channel mechanical switches at  $T = 300K$ .

a) Assume that any given ion channel is either open with conductance  $G = G_0$ , or closed with conductance  $G = 0$ . Using Boltzmann statistics, write an expression for the conductance of a giant squid axon (with  $N$  ion channels in parallel) as a function of the applied membrane potential,  $V$ . Assume that the number of open channels at  $V = 0$  is  $N_0$ .

Hint: Given Boltzmann statistics, the relative populations  $N_1$  and  $N_2$  of two states separated by energy  $dU$  are  $N_1/N_2 = \exp(-dU/kT)$ .

b) Where possible given the data in Fig. 7.15, evaluate your parameters.

c) Sketch a representative IV of a single ion channel.

MIT OpenCourseWare  
<http://ocw.mit.edu>

6.701 / 6.719 Introduction to Nanoelectronics  
Spring 2010

For information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>.