Information, computing technology, and quantum computing

2006 J. Phys.: Condens. Matter 18 S703

(http://iopscience.iop.org/0953-8984/18/21/S01)

Institute of Physics Publishing | Journal of Physics: Condensed Matter
J. Phys.: Condens. Matter **18** (2006) S703–S719 doi:10.1088/0953-8984/18/21/S01

TOPICAL REVIEW

Information, computing technology, and quantum computing

Robert W Keyes

IBM Research Division, Yorktown, NY 10598, USA

Received 12 December 2005, in final form 13 January 2006
Published 12 May 2006
Online at stacks.iop.org/JPhysCM/18/S703

Abstract
Information has long been described by physical structures. The spectacularly successful modern computers use silicon transistors to hold and process information. A number of attempts to repeat the success with other kinds of solid-state devices have failed. The reasons for the unique success of silicon transistors are found in the requirements of computing, the properties of transistors, and the variability in devices manufactured in the large quantities needed to build large computing systems. New challenges will be met in building quantum computers to meet the same requirements.

Contents

1. Introduction S704
2. Information S704
3. Computers S705
4. Variability in devices S706
5. Computing with real devices S708
   5.1. Logic gates S708
   5.2. Noise, standards and gain in devices S709
   5.3. Transistor circuits S709
6. Alternative logic devices S711
   6.1. Tunnel diodes S711
   6.2. Negative resistance devices S711
   6.3. Josephson tunnelling S712
   6.4. RSFQ computer S713
   6.5. Threshold logic S713
   6.6. Optical logic S714
7. Quantum computing S715
8. Summary S717
References S717

0953-8984/06/210703+17$30.00

1. Introduction

Forty years of continuous progress in the performance and availability of electronic information processing with silicon transistors has suggested to some investigators that there are other, better paths to automated information processing with solid state devices and a number of massive development efforts have tried to realize computers that used other devices. Here the reasons for the uniform lack of success of alternatives are examined and compared with the success of the transistor. Quantum computing has yet to face the challenges of building systems of a great many devices.

2. Information

The late Rolf Landauer (1991, 1996) taught that 'information is physical', and mankind for millennia has relied on physical means to record and transmit information. Stones in a bowl, notches in sticks, and characters in a finite alphabet have been used to represent information, as have the length of a shadow, the position of mercury in a barometer and the orientation of a magnetic needle or a stick placed on the ground. Simple calculations could be done with simple representations of information but the larger problems that emerged with the advancement of civilization and knowledge called for methods more adaptable to larger numbers. The quest for better methods to deal with larger and larger volumes of information has long occupied technologists. Many methods have been suggested and tried but few have survived to find a useful place among the tools available to commerce, science, and engineering.

The modern era of automated information processing began when the 1890 Census of the United States stored the collected information in the form of holes in punch cards (Aspray 1990). The information was in binary form: there were two possible outcomes from reading a position on the card; there either was or was not a hole. At first punch card machines were used to sort data and data could be transferred from place to place in machine readable form as a stack of cards. Ways to add function were devised and punched cards and the machines that used them had a long run as a way to record and process information.

Reading cards electrically by making a contact through the hole was an obvious step and meant that information was embodied in electrical signals. Electrification allowed the function of a punch card processor to be controlled by changing connections. Relays were a way to open and close connections, and the functioning of a system of relays could be changed by other relays. Relays can hold binary information, but the step from having information in electrical form to performing Boolean logic operations such as the functions in figure 1 with relays was slow in coming. Eventually, however, large relay computers, for example, the Harvard Mark II with 13 000 relays, were built in the 1940s (Aspray 1990).

Slide rules, devices that represented numbers with a continuum of possible values, were widely used by scientists and engineers for many calculations before the invention of the large general purpose electronic computer. The accuracy obtained with a slide rule was limited by the precision and the finite size of the markings on it. Accuracy could be increased by lengthening the slide rule, and ingenious ways of doing so have been devised. Nevertheless, the improvement in accuracy is only proportional to the length and the cost increases rapidly. The limitations of precision manufacturing methods and the awkwardness of handling large objects cannot be avoided for very long. More accurate work used information in larger blocks and needed more physical resources that were available as motor-driven mechanical calculators and tables of logarithms. The difference between the slide rule and the mechanical calculators and numerical tables is the difference between analogue and digital representation of numbers. The accuracy of digital systems, wherein characters are drawn from a limited set, can be readily

A	B	A AND B	A NOR B	A XOR B
0	0	0	1	0
0	1	0	0	1
1	0	0	0	1
1	1	1	0	0

Figure 1. Typical binary logic functions AND, NOR, and XOR.

increased by adding digits, while increasing analogue precision requires expensive efforts. Adding another decimal to slide rule numbers by making the rule ten times larger does not appear as a practical alternative. The use of a value from a continuum to represent information is limited by the resolution of the method of reading the information and by imperfect knowledge of the characteristics of the physical artifacts that create and hold the physical quantity that contains it.

Among the early examples compass needles and the position of the shadow on a sundial are analogue information while the stones in a bowl and the notches in a stick are digital. The long-enduring example of digital is the abacus, which was specifically designed for accurate calculation with large numbers. Although used for decimal arithmetic, the abacus is a binary device, a bead has just two positions, fixed by mechanical barriers. No adjustment is needed, a flick of a finger suffices to move a bead between positions. In contemporary language each bead contains one ‘bit’ of information and the five bits in one vertical column represented a decimal digit. High numerical accuracy is obtained by use of many rows of beads. Binary representation has lasted into the electronic age and has been a key to successful automated information handling.

Representation of information as a series of choices between two alternatives has long been favoured because it is not difficult to create physical devices with two states. Binary devices are familiar to all for that reason: the electrical on–off switch is an example of a bistable device in daily use and a peg in hole as used in games stores binary information. A door may be either locked or unlocked. Tossing a fair coin provides a random binary choice.

3. Computers

Accepting that information is physical implies that physical entities must be used to store and transmit and use it. Computers have been built from a number of kinds of real physical objects. Analogue computers have served well and were a mainstay of the control of ordnance during World War II (Aspray 1990). Information is stored in capacitors and resistors and the position of wheels and sliders in analogue computers. Analogue computers are well suited to control problems and other small problems where appreciable effort can be focused on maintaining high accuracy in the manufacture of the limited number of components needed.

Electrically operated binary switches, relays, were used to construct the earliest large digital computers, e.g., the Mark II mentioned above. Relays were slow and unreliable, suffering from inertia and friction and dirt on the contacting surfaces. The quest for better devices that would permit larger, faster machines soon led to construction of computers from vacuum tubes. The vacuum tube is generally relegated to the dustbin of history now, but was the only source of electrical amplification from its invention in 1906 until its replacement by transistors after 1950. By eliminating mechanical motion vacuum tube computers were able to operate far faster than relay computers. The first successful general purpose computer was the ENIAC, built with 18 000 vacuum tubes and first operated in late 1945 (Aspray 1990). Vacuum tube computers were successful enough to form the basis of a small industry. However, the tubes' high power usage and their frequent failures limited the size of machines.

After World War II the reliance of the Bell System on relays and tubes was recognized as a problem for the rapidly growing telephone industry. AT&T Bell Laboratories inaugurated a search for a solid state switch that shortly led to the discovery of transistor action (Pearson and Brattain 1955, Brinkman *et al* 1997, Riordan and Hoddeson 1997). The invention soon spawned a major revolution in electronic technology. It was quickly realized that solid state devices, in addition to small size, offered higher reliability and lower power consumption than vacuum tubes and relays. The development of transistor circuits and applications flourished because experimental transistors could be made in laboratories and were widely available for experimentation. The invention of the transistor greatly increased attention to the physics of the solid state and spawned interest in other solid state devices.

The high reliability and low power consumption of transistors made collecting a much larger number of devices into a compact computing system feasible. Methods for the mass fabrication of transistors were developed and large steps in the economics of manufacturing solid state components were made possible by the invention of the integrated circuit in 1960. The mass production of integrated circuits precluded attention to individual devices and necessitated the acceptance of a range of component characteristics.

The first transistors were bipolar transistors but another type, the field-effect transistor or FET, was soon developed (Sze 1981, Taur and Ning 1998) and became the device of choice for computer electronics. The FET has made the rapid progress in the capabilities of computers through more than three decades possible. Both transistors resemble the vacuum tube in their dependence on the attraction between positive and negative charges. A fixed charge of one kind attracts a mobile charge of the opposite sign. This fundamental fact of physics has been the sole basis for electronic amplification through the century since DeForest's invention of the vacuum triode inaugurated the age of electronics.

Low cost mass fabrication methods initiated the rapid increase in the size and power of machines that continues to the present day. The steady growth of the power of computers led to interest in larger and larger problems that drove a continuing quest for even bigger and faster machines. The thrust of transistor development during the intervening decades has been miniaturization, making transistors smaller (Moore 1965, 1975, Dennard *et al* 1974, Frank *et al* 2001, Keyes 2005b). Devices today are made on highly integrated chips of semiconductor silicon. Small size means more devices per unit area of silicon and reduces the cost per device. The history of the number of transistors on a single chip of silicon is shown in figure 2; today there may be over 10^8 transistors on a single microprocessor chip. Small size also allows devices to operate faster. The mass fabrication of smaller and smaller transistors in larger and larger quantities has made the great power and widespread use of computers possible.

4. Variability in devices

A different approach to manufacturing was necessary to make the large systems with many components that opened the way to new applications and attacks on large, complex problems widely available. The processes for mass producing large numbers of solid state devices for computers pose a problem for the physical representations of information. Producing large numbers of devices at low cost involves a compromise with high precision and reproducibility of the product. There are several sources of the variability between devices, each of which is rather formidable.

Solids contain crystalline defects and impurities (Istratov *et al* 2000, Nevin *et al* 2001, Voronov and Falster 2002). Dislocations and precipitates of interstitial atoms and vacancies

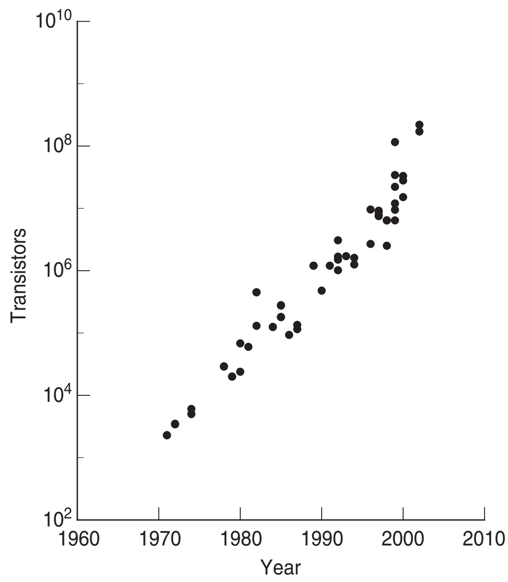

Figure 2. The growth of the number of transistors on microprocessor chips.

and of oxides can affect the properties of a device that contains them and can also affect the fabrication process. Many semiconducting substances consist of several isotopes that have somewhat different properties, specifically, different nuclear spins. As the wavefunctions of the electrons and holes in a semiconductor contact the nuclei and the properties of the electrons and holes may be affected by isotopic composition.

Semiconductor devices are fabricated on silicon wafers up to 300 mm in diameter. Perfect control of all of the parameters that characterize the tools and the substances used to make the devices cannot be achieved. Any variability or nonuniformity in something like the temperature of a reaction or an annealing step, in the mixing of a reagent, in cleanliness, or in the intensity of energy driving a process appears as variability in the parameters of devices. Perfect homogeneity and reproducibility in temperatures, in mixing of chemicals, in patterns of convection, in almost any physical quantity within a fabrication vessel is unattainable. Processes do not control the exact locations of dopants or impurities that are intentionally added to control device properties. It is hoped that their effects will be determined by a uniform average, but in miniaturized devices significant regions involve few dopant atoms and the number and location of dopants becomes a significant source of differences (Hoeneisen and Mead 1972, Keyes 1975, 1994, Mizuno *et al* 1994).

The patterns that define the structures to be created are imposed on the semiconductor by photolithography. Any lack of uniformity in the photosensitive layers may appear in the final product. The properties of light and lenses limit the precision with which a pattern can be transferred. Lenses can provide only a limited depth of focus of an image, sensitizing the patterns produced to any lack of perfect flatness or warping of a substrate. Twenty or more masking and exposure steps may be used, and imperfect alignment from step to step is reflected in the resulting devices.

Adherent contacts between materials with different coefficients of thermal expansion are found in devices. The materials are stressed by that mismatch during the heating and cooling cycles used in fabrication and some of the resulting strain is retained in the finished product (De Wolf 1996, Shen *et al* 1996, Noyan *et al* 1999) to alter the properties of devices.

Also, devices can and do change with age and use. The irreversible phenomena known by such names as creep, diffusion, bleaching, corrosion, electromigration, and thermomigration cannot be entirely avoided, and change properties of device structures during their use. Insulators often contain states that can trap charge that will be felt in a device. The trapped charge may depend on the operational history of the device. A system must tolerate a certain number of such changes in its devices throughout its life.

Cryogenic operation, by hindering atomic motion, diminishes differences arising from usage. However, temperature excursions between cryogenic and room temperatures exacerbate the changes that arise from thermal gradients and allow trapped charges to escape and move.

In addition to variability in the physical structure of devices, the fact that they are operated in a variety of environments must be taken into account. Temperature is probably the most important environmental variable. System specifications demand operation throughout a temperature range, and similar ranges are actually encountered in practice. Practically all physical properties of materials depend on temperature, and such things as Fermi levels, dielectric constants, and energy gaps are very directly reflected in device characteristics. The production of heat by working devices can lead to temperature gradients that cause differences between devices on a single substrate and patterns of usage may cause these differences to vary from moment to moment. Energetic radiation, especially the omnipresent cosmic radiation, can produce charge carriers along its tracks in silicon that are collected by devices and may change its state. This is the source of the ‘soft errors’ in memories that plagued the computer industry briefly around 1990 (Ziegler and Srinivasan 1996). Corrective actions controlled the problem then, but energetic radiation remains part of the environment that must be tolerated.

Semiconductor memory supplies an example of the influence of device variability in technology. One type of semiconductor memory stores binary digital information as charge on a capacitor: there either is or is not charge on the capacitor. Reverting to an analogue view, technologists have long asked themselves ‘Why not look at how much charge and store more than one bit on the capacitor?’ Distinguishing four levels of charge would store two bits, recognizing eight levels, three bits, etc. This has never been accomplished in DRAM, where the charge on the capacitor varies by leaking off and must be refreshed periodically. Charge does not leak in so-called ‘flash memory’ but it has proved possible to store no more than two bits, four charge levels (Atwood *et al* 1997, Fazio and Bauer 1997).

5. Computing with real devices

5.1. Logic gates

The two states in a binary physical representation of information may be designated as ZERO and ONE. Logic gates are simple circuits that perform logic operations of the type illustrated in figure 1 on the ONEs and ZEROs.

The functions in figure 1 illustrate a necessary property of digital logic gates: they receive and operate on inputs from more than one source. The connection of several inputs to a logic gate is known as fan-in. Similarly, the output of a logic gate must be able to provide fan-out, the output signal serves as the input signal to several other gates. The outputs are isolated from the inputs so that the calculation proceeds in a predetermined direction. For example, no information as to whether the output of the NOR gate is a ONE or a ZERO is reflected back to the inputs.

A complete set of logic functions must include inversion, that is, the conversion of a ONE to a ZERO and vice versa. It is apparent that a logic gate that implements the NOR function, figure 1, can perform inversion (if one of the inputs is fixed as 0 the output is the inverse of the other input). All Boolean operations can be implemented with combinations of NOR gates.

Devices for computing must handle lengthy calculations that often have great depth, meaning that the result of an operation is used in a succeeding operation, the result is used again, and so on, through tens, hundreds or thousands of steps. For example, in simulating the evolution of a system through time the outputs of an operation are reused as inputs many times. Any small error in an operation is passed on to become input to the next step and errors accumulate to produce a result with little meaning after a long series of steps. The accuracy of an analogue device, for example the slide rule discussed above, is evidently not adequate for problems that require many iterations.

Errors in operation may occur in several ways. The variability in the devices used for computing can cause the processing of a signal to differ from device to device. A large information processing system also needs a means for transmitting information from one component to another. This communication requirement is supplied by wires in electrical computers. Devices in a computing system are packed close together to minimize the time taken for signals to propagate between them and to take maximum advantage of the economies of mass fabrication. The dense packing of devices requires that the wires that carry signals from device to device are also closely spaced. The currents in the wires produce changing electromagnetic fields that induce voltages in nearby wires, causing crosstalk. Wires have resistance and signals are attenuated during transit. High frequency signals are also altered by dispersion. The signal received by a device that is to be used in a logic operation is likely to be a distorted version of the intended signal.

5.2. Noise, standards and gain in devices

Success in preventing signal degradation and information loss in the noisy variable environment of a large computing system is achieved by establishing reference signal levels that are separated by an amount larger than the noise from all sources throughout the system and resetting the signal level to its correct level at each step. The ability to refer to a standard, as in the abacus, prevents the deterioration of information during a long series of operations. Even if the representation of a digit is not perfect, if it can be recognized as, say, a one or a zero it can be restored to its standard value (Keyes 1985a, 1985b, 1989, 2001).

The advantage of binary representation in providing a standard is apparent: only two voltages, typically a power supply and ground, are needed to provide the standard signal levels throughout a system. The exact value of a received signal is not needed; it must only be interpreted as representing a zero or a one. The restoration of a signal level to its proper value at each step is effected with a response of the type shown by the curved line in figure 3 that can be produced with devices with high gain. The tolerable regions of inputs that are brought near to the desired standard values are called noise margins. The dashed curves in figure 3 that roughly parallel the solid curve illustrate the immunity to substantial variability in the position of the threshold between output signals. High gain provides rapid transition between the range of inputs that produce a one and the inputs that produce a zero.

5.3. Transistor circuits

A bistable electrical circuit makes an ideal vehicle for containing binary information. The CMOS circuit made from FETs and illustrated in figure 4 is the bistable element of the modern computer. The high gain of the transistors insures that the output terminals are connected to either the power supply V_b or ground. A positive input voltage turns the p transistors (in which current is carried by holes) off and the n transistors on, connecting the output to ground. Two negative inputs have the opposite effect and connect the outputs to V_b. The circuit executes a NOR operation: if neither input is positive then the output is positive. The output is

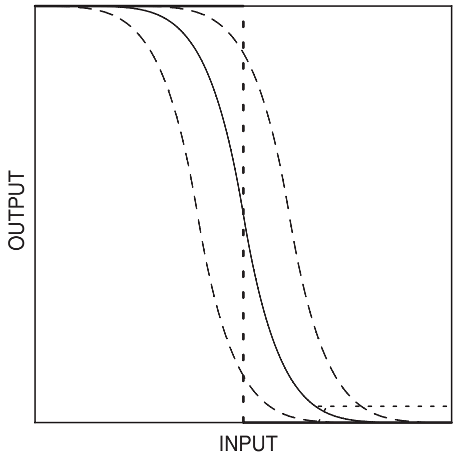

Figure 3. The solid curved line and the dashed curves show responses of an inverter that standardizes signals. Standardization does not depend on precision in the response.

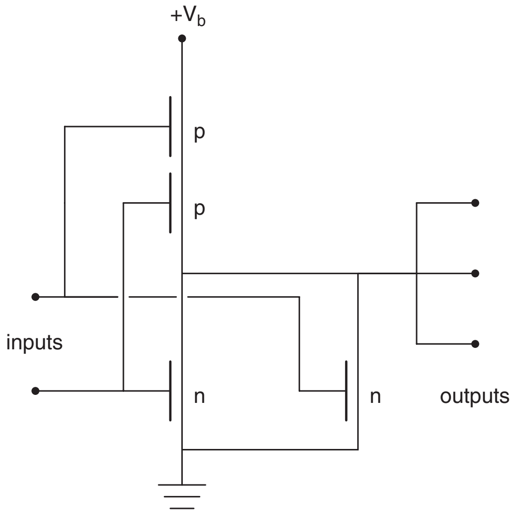

Figure 4. The FET NOR circuit.

standardized as shown in figure 4; levels are set by V_b and ground, not by the device. Even if the input signal has been degraded by 10 or even 20 percent the output is restored to its intended value.

An important feature of transistor circuits as exemplified by figure 4 is that the transistor is a three-terminal device. The gate is not sensitive to the condition of the output; there is good isolation of input from output. Both charge and voltage can be amplified. Current can flow through the transistors for a long enough time to charge or discharge a large number of following stages; excellent fan-out capability is provided. The circuit as drawn illustrates fan-in, the acceptance of a number of input signals as required for logic operations. Voltage can be amplified to restore a degraded input signal to the standard value. The circuit can switch in either direction in comparable amounts of time; no separate resetting operation is needed. Moreover, the inversion that is necessary for a complete logic system is available.

6. Alternative logic devices

Attempts to build computers from devices judged to be in some way superior to silicon transistors continue. The varied proposals use phenomena and device principles quite different from those involved in transistor action. Most have been driven by a quest for speed, based on a hope that a faster device will allow a faster computer to be realized. However, reliability of operation and noise immunity and power dissipation, among other things, are also major factors in determining the viability of a system. If, for example, devices are large or dissipate high power, cooling requirements may dictate that the distances and hence the transit times of signals between them will be large and system speed degraded.

The quest for a better device began shortly after the introduction of the transistor and has launched major development projects that have eventually been abandoned. These unsuccessful efforts have used large resources to no end. It behoves us, therefore, to try to identify what has been missing from proposed alternatives and compare them with the remarkable successes of silicon transistor technology.

6.1. Tunnel diodes

The tunnel or Esaki diode was invented a few years after the discovery of transistor action had attracted attention to semiconductors (Esaki 1958). Tunnel diodes carry current at high densities and were soon adopted as candidates for fast digital logic devices. The reasoning was that the high current density could charge the capacitance of a very small, thus very low capacitance, device in a short time.

A way of performing logic with tunnel diodes is shown in figure 5 (Gentile 1962). The power supply (shown as a battery) drives current in series through the tunnel diode and a resistor. The state of the circuit is determined by the intersection of the resistive load line with the tunnel diode characteristic, points marked A and B in the figure. State C is unstable. A three-input AND is realized as follows: current inputs are fed from preceding stages through the input at X and add to the current supplied through series resistor R. If the sum of all of the currents is greater than the peak current of the diode, then the intersection of type A no longer exists and the circuit is forced to switch to state B and a pulse of charge is transmitted from the output terminal to the next stage. The input currents are adjusted so that two will not cause the current to exceed the peak value while three will.

The difficulty met with in this kind of logic is that it depends on tight control of the signal amplitudes and the peak current. Tunnelling currents, in this case the peak current, are very sensitive to the details of the tunnelling barrier. A late study of tunnel diodes for switching circuits found that reproducibility of the peak current of plus or minus 25% was difficult to achieve (Holmes and Baynton 1967).

Additionally, the input is not well isolated from the output; signals can propagate back through resistors P. Many circuits using additional components to improve isolation were invented. The gain is not large; transformers were sometimes introduced to increase it. The lack of a way to switch the circuit in the opposite direction, from B to A, is another defect; a separate resetting procedure, perhaps turning the power off and restarting, must be provided. Nevertheless, interest in the application of the tunnel diode in computer logic continued well into the 1960s.

6.2. Negative resistance devices

The current–voltage characteristic of the tunnel diode TD in figure 5 is an example of a negative resistance, in this case of a heavily doped p–n junction. A negative electrical resistance, a

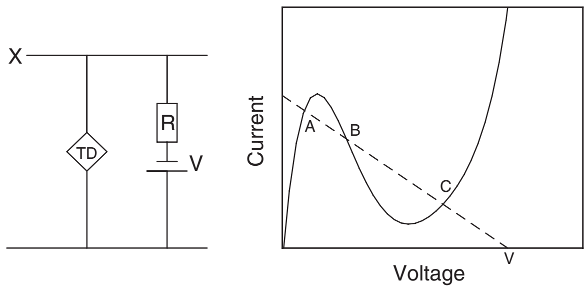

Figure 5. A negative resistance circuit and its use in digital logic. The dashed line is a load line fixed by the voltage source V and the resistance R. The solid line is the current–voltage characteristic of the tunnel diode TD. Intersections A, B, and C are points at which the current through the tunnel diode and the load R are the same and the voltages sum to the applied voltage.

region in which current decreases with increasing voltage, can also be produced with other devices. One that has received some recent attention is the resonant tunnelling diode or RTD (Goldhaber-Gordon *et al* 1997). The RTD is a semiconductor structure that contains electrons in several layers that are so thin that the electronic energy levels in them are quantized (particle in a box quantization). Tunnelling through the insulating layers between the thin regions is facilitated when their Fermi levels are aligned with one another by application of a voltage across the layer stack. As the voltage is varied the Fermi levels in the electron-containing layers are swept into and out of alignment with one another, causing the current to increase and decrease. A circuit like that of figure 5 can then be constructed and switched in the same way. However, such a circuit suffers from the problems already enumerated: inability to tolerate variability of the device parameters, lack of input–output isolation, and difficult resetting.

6.3. Josephson tunnelling

Interest in the development of logic based on the properties of Josephson junctions in superconducting circuits developed in the 1970s (Zappe 1983, Gheewala 1978, Hasuo and Imamura 1989). Josephson junctions can carry currents less than some threshold current with zero voltage drop across the junction. The zero-voltage threshold can depend on the magnetic flux through a circuit containing the junction, making it possible to change the threshold with another current by inductive coupling. Circuits containing Josephson junctions can be in two distinct states depending on whether or not the junctions in them are in the zero-voltage state and logic operations can be implemented by moving a circuit across the boundary between the states with another current as shown, for example, by figure 3 of Gheewala (1978).

The intent of the circuit is that a small change in control current should move the operating point across the boundary separating the zero-voltage from the voltage state. However, there can be substantial uncertainty in the location of the boundary between the two regions because it depends on the characteristics of the Josephson junction. In common with other devices that depend on tunnelling, the characteristics of Josephson junctions are difficult to control and reproduce from device to device (Vernik *et al* 1999, Mooij 2005). Large signals, comparable to the currents being controlled, would be needed to insure switching, leaving little room for gain to accommodate fan-out and dominate signal attenuation and distortion.

Josephson junctions are also subject to another limitation of cryogenic computers: they must be able to endure the stresses from the thermal expansion and contraction experienced in being cycled between cryogenic and room temperature for purposes of testing, repair, and modification. It has been estimated that a computer operated at cryogenic temperature will be subject to 300 such cycles during its useful life (Gomory 1983). Large systems of Josephson junctions intended to operate according to the prescriptions of Zappe (1983) and others were not successful.

6.4. RSFQ computer

The Josephson junction computer just discussed should not be confused with another superconducting computer design that has been more successful, called by its inventors the rapid single-flux-quantum computer (Likharev and Semenov 1991). The RSFQ circuit achieves a measure of signal standardization from the tendency of circuits containing a Josephson junction to find a state in which an integral number of flux quanta pass through the circuit. Information is transmitted from device to device with 'SFQ pulses' that have the property

$$\int_0^T V(t)\,\mathrm{d}t = \Phi_0.$$

Here $V(t)$ is a voltage pulse of length T and Φ_0 is the flux quantum. Small working systems have been demonstrated (Rylyakov *et al* 1999, Rylyakov and Polonsky 1998). Pursuit of RSFQ computers has been limited by the economics of providing a cryogenic environment for a computer and the difficulty of preparing large numbers of similar Josephson junctions (Vernik *et al* 1999).

6.5. Threshold logic

Novel technologies based on two-terminal devices, such as the examples above, resort to a method known as 'threshold logic' (Gentile 1962). Simply stated, the idea, as explained in the case of the tunnel diode, is that a small integer number M of ONE inputs to a device acting as a logic stage are summed, and if the sum is equal to or greater than some threshold K the stage switches. A two-input AND, in which an output is produced only if both inputs are ONE, is an example. Although working electrical circuits have been built in this way and used in early computers, the method places great demands on the accuracy of components and signal levels and is not suitable for large collections of solid state devices.

To illustrate the point, consider that the AND, figure 1, is to be implemented by adding two inputs with a nominal signal value S. Two signals must add to exceed a threshold T. Now assume for simplicity that both S and T may vary by a fraction f from their desired values. Then the following inequalities must be satisfied in all cases, including the worst. One input never exceeds threshold:

$$S(1+f) < T(1-f). \tag{6.1a}$$

Two inputs always exceed threshold:

$$2S(1-f) > T(1+f). \tag{6.1b}$$

For both of these relations to be true f must satisfy

$$(1+f)^2/(1-f)^2 < 2 \tag{6.2}$$

$$f < (\sqrt{2}-1)/(\sqrt{2}+1) = 0.172. \tag{6.3}$$

This may seem like a feasible goal, and it can easily be met by selecting devices in a laboratory, but to be used in a working computer it must be true for every one of thousands to millions of circuits that have been produced at low cost and placed in a large inhomogeneous noisy system. Since the signals and the threshold are of similar magnitude a fixed bias voltage might be added to lower the threshold, but such a bias rapidly decreases the latitude permitted in the signal size (Keyes 1989).

6.6. Optical logic

Starting in about 1970 optical physicists experimented with the performance of logic operations with beams of light. Unfortunately for optical logic the interaction between light and light is very weak compared to the strong interactions between electrical charges, leaving optics fundamentally unsuited to the performance of highly nonlinear logic operations. Interest in optical logic for digital computing centred on achieving optical bistability (Bowden *et al* 1981, 1984, Garmire *et al* 1985). Optically bistable elements could be made because of the existence of nonlinear effects that change the index of refraction of some materials at high light intensity. The use of high intensity light beams to obtain nonlinear effects implied high power that posed its own problems (Keyes and Armstrong 1969).

Optical bistability means that the light transmitted through a body of material may have two different values for the same incident intensity. Because of the weakness of optical nonlinear effects, bistable devices were made as optical etalons that trapped light between two reflective surfaces so that optical energy could be contained and accumulated to an intensity high enough to exhibit nonlinear effects. The changes in indices of refraction with light intensity altered the number of wavelengths in the light paths and thereby the efficiency with which the light was retained in the interferometer. The number of optical wavelengths in an interferometer determine whether it is in a highly transparent or relatively opaque state (Gibbs *et al* 1979), a combination that permitted bistable interferometers to be constructed in rough analogy to the bistable negative resistance circuits of the preceding section. Logical operations would be performed by making the light coming from one interferometer impinge on another one, thereby changing the light intensity and hence the transparency of the latter.

It is apparent, again, that this kind of logic does not possess the desirable properties of digital computing elements (Keyes 1985a, 1985b). The dependence of the interferometer on the relation of the wavelength of the light to the length of the cavity implies that close control of the length and of the index of refraction is required. Moreover, the large optical nonlinearities are most often associated with some kind of critical point, the GaAs exciton absorption in one example (Gibbs *et al* 1976, 1979). The wavelengths at which such points are found vary with temperature and the large effects would not be useful for digital device purposes, even though they appear attractive at first sight. As with electrical bistability, the output is not standardized to any reference but depends on the inputs and on the device. Means of achieving fan-out to drive many other devices are not apparent; it has been suggested that light amplifiers be added to make fan-out possible. The etalon is a two-terminal device; the controlled cavity is optically coupled to the cavity providing the input, allowing strong interaction between input and output. The proposals for performing elementary logic functions such as AND are implemented in threshold logic with its implications for precise control of all system parameters. It is not surprising that no attempts to construct even something as simple as a ring oscillator from all-optical components have been made.

Another thrust, also called ‘optical computing’, used light beams to carry information from place to place but performed the nonlinear operations that are essential to logic electrically with electro-optic devices (Jenkins *et al* 1984, Main *et al* 1994, Prise *et al* 1991). None of these demonstrated any convincing advantage in comparison to purely electrical computing.

Still another kind of optical computing is based on the filtering and transformation of images with masks and lenses or holograms. Since a large number of bits are necessary to describe an image, the methods can be said to process many bits in parallel. They differ from digital logic in that each information stream follows a predetermined path from light source to final transducer with no branching or referring to memory or intervening logic operations.

7. Quantum computing

Quantum computing has attracted considerable interest since the early work of Deutsch (1985) and Shor (1994), reviewed in DiVincenzo (1995) and Preskill (1998). A quantum computer is expected to be able to perform certain tasks that are far beyond the power of today's machines. The device of quantum computing is the qubit, a physical entity with at least two quantum states. Two states of the qubit are used to represent binary information that may be called ONE and ZERO, and information for computing is held in a register of qubits.

A quantum computer that applied the methods of nuclear magnetic resonance to spins of nuclei in molecules as qubits has been demonstrated (Chuang *et al* 1998, Vandersypen *et al* 2001). The two qubit states were the two orientations of spin 1/2 nuclei in a magnetic field. NMR computing, however, seems limited to less than ten-qubit systems (Chuang *et al* 1998) while it appears that a quantum computer that could attack worthwhile problems would need to have at least 10^6 qubits (Berggren 2004). The suggestions for development of a device technology that could be meet the need for larger systems have focused on a variety of solid state concepts, for example Averin (1997), Kane (1998), Loss and DiVincenzo (1998), Mooij *et al* (1999), Vrijen *et al* (2000), Stoneham *et al* (2003), and Petta *et al* (2005). Superconducting devices dominate demonstrations of working qubits and many solid state qubits based on the phenomena of superconductivity have been demonstrated. Short surveys of these have been presented by Mooij (2005) and You and Nori (2005). No large assemblages of qubits have been realized.

The interest in quantum computing stems not from a possibility of working faster than conventional computers, as did the cases discussed in the preceding section, but from its potential for solving problems that have proved intractable with present-day methods (Shor 1994). The key to this remarkable possibility is the representation of large amounts of information in superpositions of wavefunctions. An eight-qubit register might hold a byte of information as

$$|01010001\rangle.$$

However, quantum mechanics also allows information to be stored in a linear combination of the wavefunctions of the individual qubits, a superposition of many bytes of information:

$$a|01010001\rangle + b|00111101\rangle + c|01001110\rangle + d|00000110\rangle + \cdots. \tag{7.1}$$

The idea of quantum computing is that a series of operations on the physical qubits in (7.1) would affect all of the superposed bytes, thereby simultaneously processing a very large body of information.

The coefficients $a, b, c, \ldots$ that describe a superposition typified by (7.1) are analogue information: they may be selected from the entire complex number domain (subject only to a normalization condition) rather than from some finite set of values, such as zero and one. Superpositions can in principle hold vast amounts of information, being chosen from an infinitely dense continuous range of numbers. Steane and Rieffel (2000) propose that a 64-qubit register could hold 2^{64} bits (2×10^{19} bits). Still farther out, Brassard (2005) suggests that a 1024-bit register could hold 2^{1024} (10^{301} bits) but suggests disbelief by adding that this is more than the number of elementary particles in the universe. It has even been asserted that superpositions can hold an infinite amount of information (Barenco 1996). However, it is hard to reconcile such faith with the physical nature of information.

Solid state devices for quantum computing share the sources of variability that are found in devices used for other purposes. There is frequently another environmental variable to consider: a magnetic field that creates two energetic states of spin 1/2 particles that can function as qubits. Lack of homogeneity in the magnetic field is an additional source of variability between devices. A magnetic field homogeneous within one part in 10^9 within their sample was needed and was difficult to provide in the NMR computer of Chuang *et al* (1998).

The lack of precise knowledge of the physical properties of solid state qubits and, therefore, of their wavefunctions is a limitation of quantum processing. Consider the CNOT operation that involves the interaction between two qubits. When the first bit is a one the CNOT operation turns on a perturbation to the wavefunction of the second bit that inverts the second bit. (CNOT can be constructed from the XOR of figure 1.) The problem for hardware is that the perturbation must be turned off just at the time that the second bit is inverted. However, variability in physical devices implies variability and uncertainty in the wavefunctions of the qubit, in the effect of the perturbation, and in the turn-off time. The architects of NMR quantum computing found and emphasized the necessity and the difficulty of precise timing of computing operations (DiVincenzo 1995, Chuang *et al* 1998). Instead of

$$\mathrm{CNOT}(|10\rangle) \rightarrow |11\rangle \tag{7.2}$$

the defective knowledge may lead the CNOT to yield

$$\mathrm{CNOT}(|10\rangle) \rightarrow \alpha|11\rangle + \beta|10\rangle. \tag{7.3}$$

Synchronization of operations throughout a large system needs careful attention even in conventional computing systems.

Reading the second bit of the result, equation (7.3), yields a one with probability $|\alpha|^2$ and a zero with probability $|\beta|^2$. An erroneous value, such as might be found with equation (7.3), would be passed on to add to error in subsequent operations and cause a steady deterioration of information in the course of a computation.

The qubits of a quantum computer must interact with one another for purposes of communication, but the interaction of a qubit with an external entity, perhaps a phonon or an electromagnetic field, would destroy the quantum coherence of a computation. The developers of quantum computing are well aware (DiVincenzo 1995, Haroche and Raimond 1996, DiVincenzo and Terhal 1998, Preskill 1999) of the possibility of error caused by decoherence and have designed error correction procedures to combat it (Shor 1995, Steane 1996). The error correction methods are designed to detect and correct the change of a qubit from 0 to 1 or from 1 to 0. It remains to be seen whether the same procedures can overcome the limitations of the analogue aspect of quantum computing as expressed in a range of device characteristics and faulty timing of operations.

Quantum error correction is difficult because the information in a superposition is not like classical information. One of the counter-intuitive aspects of quantum mechanics is that the result of reading a qubit with contents described by a superposition of numbers sets the qubit to the value read. If the second bit of the CNOT in equation (7.3) is read as zero, then its value becomes zero and the memory of the probabilities α^2 and β^2 would be lost. Because error correction must avoid reading information, prescriptions for error correction involve a substantial increase in the number of devices in a computer and in the number of operations that must be performed in the course of a computation. The first error correcting codes were those of Shor (1995), which represented each bit (0 or 1) by one or the other of two superpositions of eight nine-bit words, and of Steane (1996), which was suggested by Hamming codes and used superpositions of eight seven-bit words for each bit. Bennett *et al* (1996) and Laflamme *et al* (1996) have presented a compact error correction coding in which each digit is represented by a superposition of 16 five-qubit words. It has been suggested that error correction will need to be iterative, that is, the operations of error correction must themselves be corrected in the same way. Apparently coding bits in these prolix ways will require a large increase in the number of devices and a very large computational overhead. This is part of the reason that a quantum computer needs a seemingly large number of qubits. Over 10^6 operations will be needed to execute a logical quantum gate on a qubit with concatenated error correction according to one estimate (Knill *et al* 1996).

The probability of decoherence is greatly reduced by minimizing thermal excitations, keeping qubits in the tens of millidegrees temperature range. However, the computing hardware is then exposed to the temperature cycling and the accompanying stresses that plagued the Josephson computer.

8. Summary

The high reliability and low power consumption of transistor circuits made possible the building of very large systems of active devices that have led to the wonders of modern electronics. Methods for producing large numbers of components at low cost made the power of computers widely available at the sacrifice of high accuracy in the reproducibility of devices. Computers use the high gain of transistors to restore signals to their intended values, thereby allowing reliable operation in the noisy variable imprecise environments that characterize large systems.

The many fruitless attempts, often carried out by very competent scientists and engineers, to find other device technologies that would advantageously replace the transistor in logic circuits cause one to wonder what so consistently went wrong. Why were the sources of failure not obvious without the expenditure of a large amount of development work? A devotion to device R&D easily entices one to focus on the device at the expense of the large system. The realities of computing that lie beyond the domain of science can be lost behind the allure of clever technology. Further, generous financial support for novel technologies is frequently generated by uncritical acclaim in the popular and the technical presses.

The laboratory environment is deceptive; when one or two devices are made to work in a laboratory by carefully adjusting everything to obtain a desired action it is tempting to extrapolate the success to a system of a great many devices. However, the system presents many obstacles to contend with: devices must work together in spite of differences in their physical characteristics and the demands for connectivity, fan-in, fan-out, and noise immunity made upon them. Each must be able to make reliable decisions concerning the meaning of signals received that may have been attenuated and distorted by resistance and noise. Silicon technology found the solution in circuits like that of figure 4 that correct signals at each step. The inability of bistable circuits to duplicate the standardization has led to their failure to find a place in solid state technology.

The future of quantum computing is obscure. Although physics is based on experimental knowledge of the world, quantum computing is almost entirely a construct of the mind, an exclusively theoretical subject. Only the first tentative steps have been taken by realizations of one or two solid state qubits. Quantum computing must await the knowledge to be gained from attempts to construct useful systems of millions or even systems of thousands of qubits that must all work simultaneously without the careful tuning possible in a laboratory and that include dense device to device communication channels.

References

Aspray W 1990 *Computing Before Computers* (Ames, IA: Iowa State University)
Atwood G *et al* 1997 *Intel. Technol. J.* **1** 1–8
Averin D V 1997 *Solid State. Commun.* **105** 659–64
Barenco A 1996 *Contemp. Phys.* **37** 375–89
Bennett C H *et al* 1996 *Phys. Rev.* A **54** 3824–50
Berggren K K 2004 *Proc. IEEE* **92** 1630–8
Bowden C M, Ciftan M and Robl H R (ed) 1981 *Optical Bistability* (New York: Plenum)

Bowden C M, Gibbs H M and McCall S L (ed) 1984 *Optical Bistability 2* (New York: Plenum)
Brassard G 2005 *Nat. Phys.* **1** 2–3
Brinkman W F, Haggan D E and Troutman W W 1997 *IEEE J. Solid State Circuits* **32** 1858–65
Chuang I L *et al* 1998 *Nature* **393** 143–5
Dennard R H *et al* 1974 *IEEE J. Solid State Circuits* **9** 256–68
Deutsch D 1985 *Proc. R. Soc.* A **400** 97–117
De Wolf I 1996 *Semicond. Sci. Technol.* **11** 139–54
DiVincenzo D P 1995 *Science* **270** 255–61
DiVincenzo D P and Terhal B 1998 *Phys. World* **11** (3) 53–7
Esaki L 1958 *Phys. Rev.* **109** 603–4
Fazio A and Bauer M 1997 *Intel. Tech. J.* **1** 19–22
Frank D J *et al* 2001 *Proc. IEEE* **89** 259–87
Garmire E *et al* (ed) 1985 *IEEE J. Quantum Electron.* **21** 1339–549
Gentile S P 1962 *Theory and Application of Tunnel Diodes* (Princeton, NJ: van-Nostrand Reinhold) chapter III.8
Gheewala T 1978 *Appl. Phys. Lett.* **33** 781–3
Gibbs H M *et al* 1979 *Appl. Phys. Lett.* **35** 451–3
Gibbs H M, McCall S L and Venkatesan T N C 1976 *Phys. Rev. Lett.* **36** 1135–8
Goldhaber-Gordon *et al* 1997 *Proc. IEEE* **85** 521–40
Gomory R E 1983 *Science* **220** 576–80
Haroche S and Raimond J-M 1996 *Phys. Today* **49** (8) 51–2
Hasuo S and Imamura T 1989 *Proc. IEEE* **77** 1177–93
Hoeneisen B and Mead C A 1972 *Solid-State Electron.* **15** 819–29
Holmes D L and Baynton P L 1967 *Gallium Arsenide, Conf. Series* No 3 (London: Inst. Phys. and Phys. Soc.) pp 236–40
Istratov A A, Hieslmair H and Weber E R 2000 *Mater. Res. Soc. Bull.* **2000** 33–8
Jenkins B J *et al* 1984 *Appl. Opt.* **23** 3455–64
Kane B E 1998 *Nature* **393** 133–7
Keyes R W 1975 *Appl. Phys.* A **8** 251–9
Keyes R W 1985a *Opt. Acta* **32** 525–35
Keyes R W 1985b *Science* **230** 138–44
Keyes R W 1989 *Rev. Mod. Phys.* **61** 279–87
Keyes R W 1994 *IEEE Circuits Devices Mag.* **10** 28–31
Keyes R W 2001 *Phil. Mag.* B **81** 1315–30
Keyes R W 2005a *Computer* **38** 65–9
Keyes R W 2005b *Rep. Prog. Phys.* **68** 2701–46
Keyes R W and Armstrong J A 1969 *Appl. Opt.* **8** 2549–62
Knill E, Laflamme R and Zurek W 1996 *Preprint* quant-ph/9610011 v3
Laflamme R *et al* 1996 *Phys. Rev. Lett.* **77** 198–201
Landauer R 1991 *Phys. Today* **44** (5) 23–9
Landauer R 1996 *Phys. Lett.* A **217** 188–93
Likharev K K and Semenov V K 1991 *IEEE Trans. Appl. Supercond.* **1** 3–28
Loss D and DiVincenzo D P 1998 *Phys. Rev.* A **57** 120–6
Main T *et al* 1994 *Appl. Opt.* **33** 1619–28
Mizuno T, Okamura J and Toriumi A 1994 *IEEE Trans. Electron. Devices* **41** 2216–21
Mooij J E 2005 *Science* **307** 1210–1
Moore G E 1965 *Electronics* **19** 114–7
Moore G E 1975 *Int. Electron Devices Meeting* **21** 11–13
Mooij J E *et al* 1999 *Science* **285** 1036–9
Nevin W A *et al* 2001 *J. Electrochem. Soc.* **148** G649–54
Noyan I C *et al* 1999 *Appl. Phys. Lett.* **74** 2352–5
Pearson G and Brattain W H 1955 *Proc. IRE* **43** 1794–806
Petta J R *et al* 2005 *Science* **309** 2180–4
Preskill J 1998 *Proc. R. Soc.* A **454** 469–86
Preskill J 1999 *Phys. Today* **53** (6) 24–30
Prise M E *et al* 1991 *Appl. Opt.* **30** 2287
Riordan M and Hoddeson L 1997 *Crystal Fire: The Birth of the Information Age* (New York: W W Norton)
Rylyakov A and Polonsky S 1998 *IEEE Trans. Appl. Supercond.* **8** 14–9
Rylyakov A, Schneider D and Polyakov Y 1999 *IEEE Trans. Appl. Supercond.* **9** 3623–7

Shen Y-L *et al* 1996 *J. Appl. Phys.* **80** 1388–97
Shor P W 1994 *Proc 35th Annual Symp. Foundations Comp. Sci.* (Los Alamitos, CA: IEEE Computer Soc.)
Shor P W 1995 *Phys. Rev.* A **52** 2493–6
Steane A M 1996 *Phys. Rev. Lett.* 793–7
Steane A M and Rieffel E G 2000 *Computer* **33** 38–45
Stoneham A M, Fisher A J and Greenland P T 2003 *J. Phys.: Condens. Matter* L447–51
Sze S M 1981 *Physics of Semiconductor Devices* (New York: Wiley)
Taur Y and Ning T H 1998 *Fundamentals of Modern VLSI Devices* (Cambridge: Cambridge University Press)
Vandersypen L M K *et al* 2001 *Nature* **414** 883–7
Vernik I V *et al* 1999 *IEEE Trans. Appl. Supercond.* **9** 4341–4
Voronov V V and Falster R 2002 *J. Electrochem. Soc.* **149** G167–74
Vrijen R *et al* 2000 *Phys. Rev.* A **62** 012306
You J Q and Nori F 2005 *Phys. Today* **58** (11) 42–7
Zappe H H 1983 *Advances in Superconductivity* ed B Deaver and J Ruvalds (New York: Plenum) pp 51–127
Ziegler J F and Srinivasan G R (ed) 1996 *IBM J. Res. Dev.* **40** 1–128