J. Matisoo

Overview of Josephson Technology Logic and Memory

This paper serves as an introduction to the other logic and memory papers in this issue. Basic concepts of superconductivity and electron tunneling underlying the operation of Josephson devices are outlined and an overview of the literature on the subject is presented, with emphasis on work performed at the IBM research laboratories since the beginnings of the Josephson computer technology program in 1965.

Introduction

Josephson logic and memory circuits make use of essentially conventional passive components and a rather esoteric active device, all made in an integrated manner as described by Greiner *et al.* [1] in this issue. The passive elements, resistors, capacitors, inductors, and transmission lines, while conventional, approximate the ideal components seen in textbooks because of the essentially lossless nature of superconducting metals.

The two phenomena underlying the operation of the *active device* are superconductivity and electron tunneling. The first of these, superconductivity, discovered by Kammerlingh Onnes, was explained in detail by Bardeen, Cooper, and Schrieffer. All four received Nobel prizes in physics for their work; Kammerlingh Onnes in 1913, and BCS in 1972 for the microscopic theory. Electron tunneling received its first practical application with the invention of the semiconductor tunnel diode by Esaki in 1957. The second important tunneling discovery was that of superconductive tunneling by Giaever. Lastly, Josephson predicted that magnetic field-sensitive supercurrents should flow through a tunnel junction with a frequency proportional to the voltage across the junction. The tunneling discoveries of Esaki and Giaever and the predictions of Josephson led to a "tunneling" Nobel prize in physics, awarded in 1973.

That the superconductive tunnel junction, combined with a means of controlling the magnitude of the zero-voltage current, forms a fast, low-power logic and memory device was recognized in 1967 by Matisoo [2]. Matisoo designed and fabricated such devices and established their static [3] and dynamic [4] properties in simple circuits [5]. The switching speeds and power levels, when combined with superconducting transmission lines, make extremely attractive devices for computer applications. Studies by Anacker [6] led to the undertaking of a coordinated research program which has evolved and progressed to the point where the design and construction of a small prototype computer system is currently underway. The computer, a Josephson Signal Processor (JSP), is described by Tsui [7] in this issue.

Basic concepts

This section will introduce the reader to some of the basic concepts of superconductivity and electron tunneling. The literature on both of these subjects is vast, with many books and review articles; the reader is referred to a partial bibliography at the end of this paper for more complete descriptions.

- *Superconductivity*

Most metals, particularly those which are poor conductors at room temperature, exhibit the superconducting state provided that certain conditions are met. These conditions, characterized by critical parameters, are that a) the temperature be below the transition temperature T_c; b) the applied magnetic field be below the critical field strength H_c; c) the current through the superconductor be below a critical value; and d) the fields (or currents) have time rates of change below some critical value ω_c. These critical parameters are all temperature-dependent, having their largest numerical values at 0 K and vanishing at T_c. The transition temperature and the zero temper-

ature values of the other critical parameters are material constants. The transition temperature of known materials covers a range from essentially 0 K to approximately 24 K. The materials of technological interest, lead and niobium, have T_cs of 7.2 K and 9.4 K respectively. Zero-temperature critical field strengths similarly cover a range from the vanishingly small to tens of kilo-amperes per meter, being in excess of 40 kA/m for lead and niobium. Critical currents (in terms of current density) are typically on the order of 10^7 A/cm^2 for materials of interest. Critical frequencies for lead and niobium are in excess of 1000 GHz. When the conditions for superconductivity are met, the superconductive state is characterized by two macroscopic properties, perfect conductivity and perfect diamagnetism. Transport and shielding currents are confined to within a distance λ of the surface of a superconductor; the interior is field-free. This distance λ, called the penetration depth, is again temperature-dependent and has a zero-temperature value which is characteristic of the material; typical zero-temperature values are approximately 100 nm. The temperature dependence of all of these parameters is weak once the temperature has been reduced below approximately one-half of T_c.

It is important to note that perfect conductivity and perfect diamagnetism are independent properties of the superconducting state. This means that a superconductor cooled through its transition temperature in the presence of a small magnetic field does not trap that field as might be expected of a perfect conductor, but rather expels it. This is the so-called Meissner effect. In principle, all field is expelled; however, in practice some flux trapping always occurs.

According to the BCS theory, the superconducting state is characterized by a condensation of the electrons near the Fermi level into a state of energy lower than the normal state. In this state the electrons form bound pairs of equal and opposite momentum. The binding energy is designated as $2eV_g$. (*Editor's note*: The symbol V_g is the same as $2\Delta/e$ used in certain other papers of this issue, where 2Δ is the bandgap energy in electron-volts.) The electrons forming a bound pair are on the average separated by a distance ξ, the coherence length, which has a typical value of approximately 50 nm. At zero temperature, all electrons are paired and form a highly coherent state in which the individual pair wave functions are superposed with zero phase difference to form a grand wave function describing the superconducting state. Roughly speaking, perfect conductivity follows from this picture. An impressed current simply results in the center of mass momentum of the pairs assuming a nonzero value. No scattering events occur until the kinetic energy of the pairs becomes sufficient to result in pair breaking, thereby resulting in single electrons which scatter from the lattice and lead to electrical resistance. When this occurs, the critical current of the superconductor has been reached.

It should be noted that in the superconducting state there exists an energy gap between the ground state formed of electron pairs and the excited states consisting essentially of the familiar single electrons. This gap value is temperature-dependent, having its largest value at $T = 0$ and vanishing at $T = T_c$. The zero-temperature value of the gap is directly proportional to the numerical value of the transition temperature of the material and is typically a few millivolts. The energy gap (bandgap) voltage, V_g, is $\approx$2.8 mV for lead and its alloys at the usual operating temperature of approximately 4.2 K. In many ways a superconductor is describable by the familiar bandgap model of a semiconductor, with some crucial differences, as discussed by Adkins [8].

The ground state of a superconductor is described in terms of a wave function which has a magnitude and a phase. In a multiply connected superconductor, such as a ring, gauge invariance, coupled with the requirement of single-valuedness for the wave function, demands that the flux enclosed (or trapped) by such a ring be quantized in units of the flux quantum $\Phi_0 = h/2e = 2 \times 10^{-15}$ Wb. This multiply connected geometry is of particular interest because it forms the basis of the memory technology.

Perfect conductivity and diamagnetism mean that the interior of a superconductor is free of both electric and magnetic fields. Josephson investigated the situation in which superconductivity was "weakened" by some means so that the response of the superconductor to electric and magnetic fields could be studied. His original calculations were specifically made with respect to the Giaever tunnel junction, as a prototype of the weakly superconducting region, in which the effects should be observable. Subsequently, he generalized the theory to any weakly superconducting region to make it amply clear that the predictions follow directly from the general properties of the superconductive state. The significant relations are the following:

1. The current flow between any two points of a weak superconductor is a periodic function of the phase difference of the wave functions at those points; *i.e.*, $j = j_1 \sin \phi$, where j is the current density between the two points, j_1 is the maximum current density which can occur, and ϕ is the phase difference.
2. The current that flows is a function of the magnetic field. This is expressed parametrically through ϕ as

$$\nabla^{(2)}\phi = \frac{2ed}{\hbar}\,(\mathbf{H} \times \mathbf{n}),$$

where e and $\hbar$ are the fundamental constants, electron charge and Planck's constant divided by 2π, respectively, d is the distance over which the field penetrates, $\mathbf{H}$ is the magnetic field, and $\mathbf{n}$ is a unit vector.

The equation is a close relative of flux quantization, arising from the same requirements of gauge invariance and single-valuedness for the superconductive wave function. Through the first relation, it makes the current flow between two points of the weak superconductor a periodic function of the magnetic field also.

The nature of the current between these two points also depends on the electric field through the equation

$$\frac{d\phi}{dt} = \frac{2e}{\hbar}\,V,$$

where V is the voltage between the points and t is time. Again, through the first relation, the current between these two points is a periodic function, but now of time, whenever the voltage is nonzero. This is a most interesting nonlinearity and has been of considerable scientific and practical interest. For example, the legal United States volt is now expressed in terms of this equation.

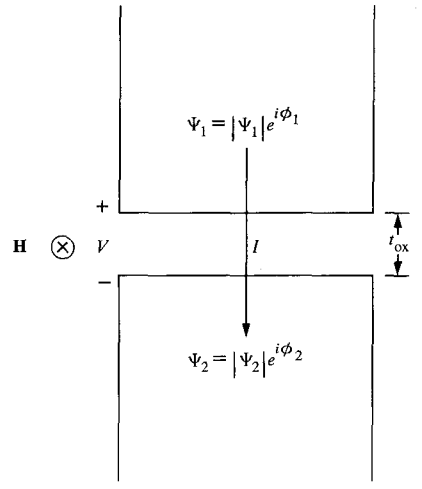
These three relations are the so-called Josephson equations and are the natural response functions of the superconductor to impressed electric and magnetic fields; their practical manifestations are most readily observed in the Giaever tunnel junction.

- *Tunneling*

A metal tunnel junction is a simple structure consisting of a sandwich of two metal electrodes separated from one another by an insulator thin enough to permit electrons to tunnel with a reasonable probability. Most frequently, and this is the case in this technology, the insulator is an oxide grown on the lower, or base electrode. At room temperature the current I can be written as

$$I = G_{nn}V + C\,\frac{dV}{dt},$$

where G_{nn} is the conductance and C is the capacitance of the oxide structure, corresponding to that flowing in a leaky capacitor, which is what the structure most nearly resembles. Indeed, if the oxide is sufficiently thick there is very little conduction and we have just a parallel plate capacitor; at the other extreme, a metallic contact having some resistance. What distinguishes this structure is that, if electron transport from one electrode to the other is indeed by electron tunneling, the conductance G_{nn} depends strongly on the magnitude of the potential barrier formed by the oxide and its thickness. The thickness dependence is of particular interest, since it is through the control of this parameter that the conductance is in turn controlled. The dependence is exponential with the conductance decreasing exponentially as the thickness of the oxide increases [9]. Greiner *et al.* [1] discuss the technological problems involved in controlling this quantity.



$$I = I_J + I_{qp} + I_C$$

$$I_C = C\,\frac{dV}{dt}$$

$$I_{qp} = G_{nn}/e\int N(E - eV)N(E)[f(E - eV) - f(E)]dE$$

$$I_J = Aj_1 \sin\phi \qquad \text{where } \phi = \phi_1 - \phi_2$$

$$\frac{\partial\phi}{\partial t} = \frac{2e}{\hbar}\,V$$

$$\nabla^{(2)}\phi = \frac{2ed}{\hbar}\,(\mathbf{H} \times \mathbf{n})$$

$$d = \lambda_1 + \lambda_2 + t_{ox}$$

$$j_1 = \frac{\pi\Delta G_{nn}}{2eA}\tanh\left(\frac{\Delta}{2kT}\right)$$

Figure 1 Currents in a tunnel junction and their dependence on electric and magnetic fields.

As this sandwich is cooled to a temperature well below T_c of the electrodes, the *I-V* characteristic of the junction becomes highly nonlinear and therefore useful. As shown in Fig. 1, a third current contribution, the Josephson current I_J, has also appeared; the ohmic current I_{qp} has become nonlinear in voltage. This nonlinearity is simply the result of the metals having become superconducting and having developed an energy gap $2eV_g$. In the gap there are no available electron states; consequently, no current

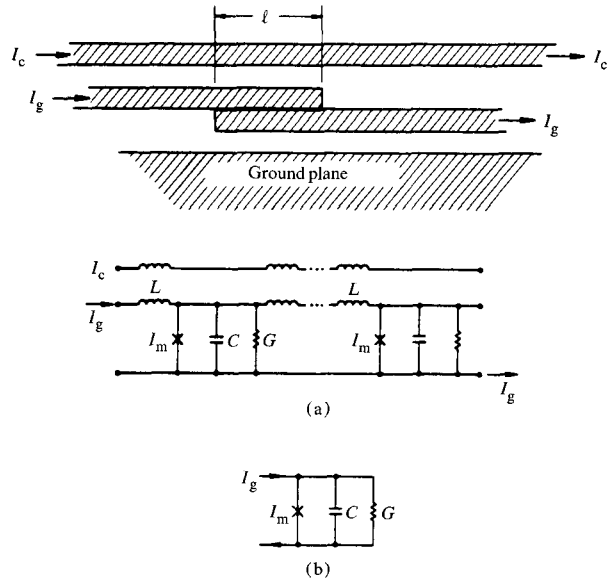


Figure 2 Equivalent circuit of a tunnel junction and nonlinear partial differential equations which describe its behavior: (a) for $\ell \gtrsim 3\lambda_J$ and

$$C\frac{\partial^2\phi}{\partial t^2} + G\frac{\phi}{\partial t} + \frac{2eI_m}{\hbar}\sin\phi = \frac{1}{L}\frac{\partial^2\phi}{\partial x^2};$$

(b) for $\ell \lesssim 3\lambda_J$ and

$$C\frac{\partial^2\phi}{\partial t^2} + G\frac{\partial\phi}{\partial t} + \frac{2eI_m}{\hbar}\sin\phi = I; \text{ where} \lambda_J^2 = \frac{\Phi_0}{2\pi\mu_0 dj_1}.$$

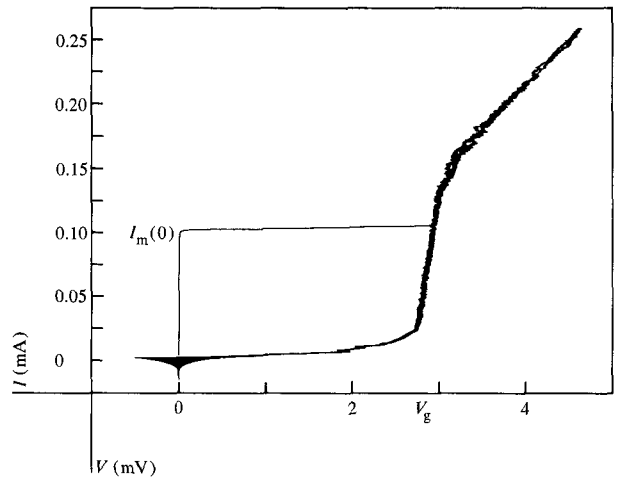


Figure 3 *I-V* characteristic calculated on the basis of model illustrated in Fig. 2, with typical parameters.

flows until the voltage across the junction is equal to V_g, at which time the electrons from one electrode can tunnel to states above the gap in the other electrode.

The Josephson current term appears because the interactions responsible for superconductivity extend, albeit attenuated, across the oxide. The attenuation, however, is sufficient in most cases to render the oxide region "weakly" superconducting and yet magnetic fields and electric potentials can be applied to the oxide—precisely the conditions postulated by Josephson for observing the response of the superconducting state.

Note, in Fig. 1, that apart from the displacement current I_C, the entire current scale of the tunnel junction is determined by G_{nn}. The voltage scale is fixed by the superconducting energy gap voltage V_g.

This picture of currents in a tunnel junction can be translated to familiar engineering terminology, resulting in an equivalent circuit shown in Fig. 2. The partial differential equation describes the time and spatial variation of voltages and currents in a tunnel junction. Since it is nonlinear, the solution of this equation, particularly in the presence of boundary conditions, can be performed only numerically. However, for a point junction (when spatial variations can be neglected) the linear portion of the equation is familiar and simple, the time response being determined by an *RC* time constant in the familiar way. Typical values of R and C are such that time constants on the order of a few tens of picoseconds arise.

The differential equation displayed in the caption of Fig. 2 is approximate, with certain assumptions and simplifications having been made in its derivation, and with certain known current terms not included. Nevertheless, the equation has been tested in many and diverse experimental situations and has been found to represent the real world with remarkable accuracy, to time scales on the order of a few picoseconds.

Two simple examples of tunnel junction properties calculated on the basis of the model are shown in Figs. 3 and 4. Figure 3 shows a computed *I-V* characteristic of the junction as it might be measured with an oscilloscope of very large bandwidth. Apparent are the key features of the *I-V* characteristic: the zero-voltage current of maximum value, $I_m(0)$, the nonlinear *I-V* characteristic with the gap voltage V_g, and the superposed oscillating currents whose frequency decreases as the voltage across the junction is reduced. A time-averaged version of this *I-V* characteristic is commonly observed in practice.

Figure 4 compares experiment and calculation for the magnetic field dependence of the maximum zero-voltage current I_m in a specific junction configuration with the electrical length of the junction in terms of the Josephson penetration depth λ_J as a parameter. This penetration

depth is analogous to λ, the superconducting penetration depth, and λ_J is the measure of distance over which the fields and currents penetrate the tunnel junction. Here I_g is defined as the current through the junction and I_c as the current through the control line.

• *Devices*

There are a number of ways in which practical switching devices can be configured from the basic tunnel junction. In each case, however, the operating principles are the same. All devices exhibit an *I-V* characteristic similar to that of Fig. 3. The two states of the device are the zero-voltage state and the resistive state. In operation, the device is current biased in the zero-voltage state with $I_g < I_m(0)$ and is caused to switch under the influence of an input (or control) current I_c, which either adds to I_g or reduces $I_m(0)$ so that the threshold is exceeded, and the device switches according to the external load to the resistive state. Switching from the resistive state to the zero-voltage state can occur in one of two ways. Either the device switches to the zero-voltage state upon the removal of the input (operation referred to as nonlatching), or by reducing the bias current I_g such that the voltage across the device becomes less than a characteristic voltage V_{min}. Either device operation can be obtained by appropriate choice of device and circuit parameters. A thorough discussion of the relevant considerations for this latching and nonlatching operation is given by Zappe [10]. For present-day technology, nonlatching operation can be obtained only for very small load impedances. Consequently, the majority of logic circuits are latching. This mode of operation places special requirements on the power supply. These will be discussed subsequently.

A very significant design feature of the device is the threshold characteristic, the locus of points in the I_g and magnetic field plane which forms the boundary between the $V = 0$ and $V \neq 0$ states of the device. It is essentially this feature and the operating current levels which form the distinction between various classes of device.

There are three types of devices. The oldest of these is the in-line gate with single or multiple controls [2]. This device was in use as the basic logic and memory device from its inception in 1965 until superseded by multijunction devices in 1974. The device structure consists of a tunnel junction with an overlaid in-line control, or set of controls, which couples magnetic field to the junction, thus providing control of the zero-voltage current. To obtain current gain (loosely defined here as the current supplied to the load, divided by the input current), the device current levels have to be sufficiently large that $\lambda_J \lesssim \ell/3$. This requirement restricts the devices for optimally driving relatively low-output impedances.

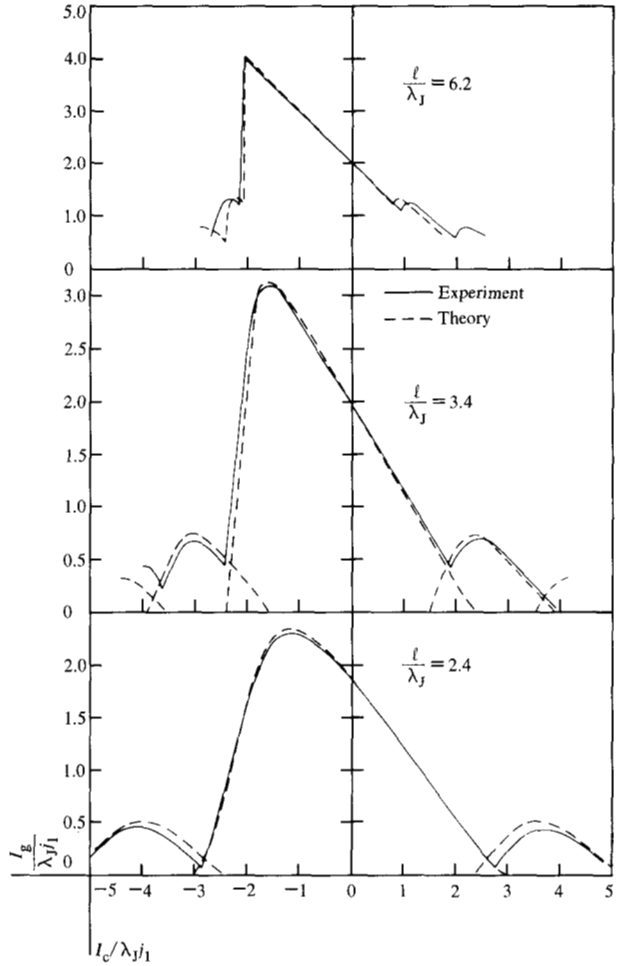


Figure 4 Threshold curves for in-line devices with ℓ/λ_J as a parameter.

As the technology has evolved from 25-μm minimum linewidths to the current 2.5-μm linewidth, the output impedances Z_0 have increased, leading to incompatibility with the in-line device properties and the introduction of multijunction devices. This incompatibility arises because the natural current scale is roughly V_g/Z_0 and since V_g is a fixed material constant, as Z_0 increases, the current levels decrease. In-line devices still find use, however, in applications in which large current levels are necessary, for example, as drive and logic devices in the DRO main memory chip described by Guéret, Moser, and Wolf in this issue [11].

The second class of device consists of the modern multijunction devices, which came into being in response to the need for device structures whose threshold currents were sensitive to smaller current levels. It was known through the work of Jaklevic *et al.* [12] that this could be achieved by interconnecting two point junctions with a

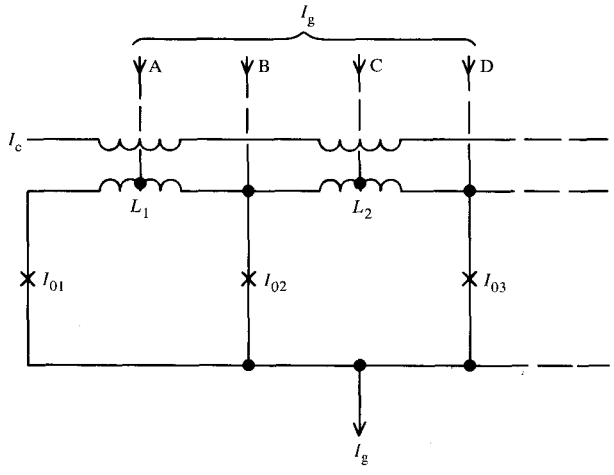


Figure 5 Generalized equivalent circuit for multijunction devices. Each cross represents a tunnel junction with its own equivalent circuit, as for example in Fig. 2. Note: the I_0s are the zero-field critical currents of the individual junctions.

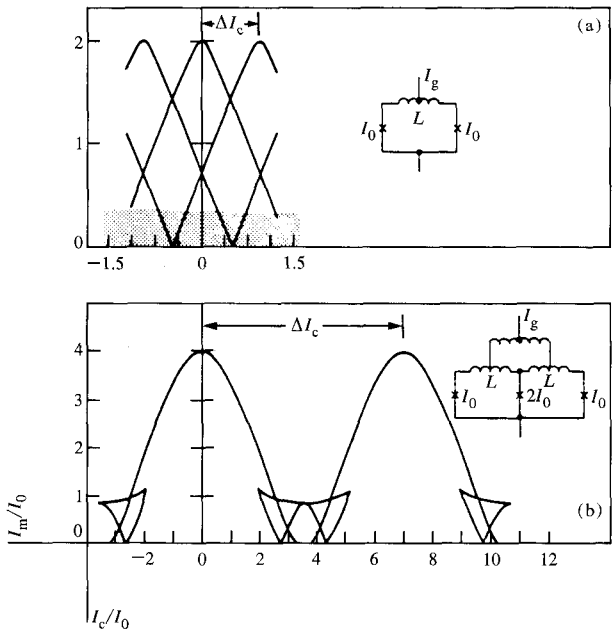


Figure 6 Threshold characteristics for multijunction devices: (a) two-junction device, $\Phi_0/LI_0 = 0.94$; (b) three-junction device, $\Phi_0/LI_0 = 7$.

superconducting line and maximizing the area of this loop or ring through which the field could penetrate. They had shown that the total supercurrent through this parallel structure is reduced to a small value whenever half a flux quantum is enclosed in this area. Since the flux is proportional to the area, it is clear that magnetic field sensitivity is increased by increasing the enclosed area. Such two-junction devices carry the acronym "dc SQUID" (dc superconducting quantum interference device) and have been used extensively as sensing elements of uniquely high-sensitivity instruments, magnetometers, gradiometers, and voltmeters. To make practical switching devices the magnetic field is coupled via an overlying control line as in the in-line gate.

The idea of two-junction devices can be generalized to three-, four-, five-junction and so on. A generalized equivalent circuit for such structures is shown in Fig. 5, in which each cross represents a point junction which in turn has the equivalent circuit shown in Fig. 2 [13]. The interconnecting inductances represent the interconnecting superconducting lines. The control line which carries current I_c is transformer-coupled to the loops. The bias current I_g can be fed to the device through a number of points. The two device structures of primary interest are the center-fed two-junction device and the split-feed three-junction structure, in which the point junctions have a zero-voltage current ratio of 1:2:1. The threshold characteristics of these devices are shown in Figs. 6(a) and 6(b), respectively. The threshold curves are determined by the L and I values and the injection points A, B, C, D. The lobe-to-lobe separation ΔI_c is given solely by Φ_0/L_m, where L_m is the inductance which couples to I_c. The fact that the lobe-to-lobe separation is *solely* and uniquely related to L_m permits the measurement of unknown inductances by incorporating them as the interconnecting inductance in a two-junction SQUID. This technique, pioneered by Henkels [14], was used by Jones and Herrell [15] in a beautiful set of measurements to determine the inductance matrix associated with the chip-carrier interconnections.

The threshold characteristic of the two-junction SQUID of Fig. 6(a) shows overlapping modes. Under the central lobe, no circulating current is present in the loop formed by the two point junctions and the interconnecting inductances, whereas in the two adjacent lobes a circulating clockwise or counterclockwise current, corresponding to a flux quantum, is present in addition to the externally applied currents. The fact that the modes overlap about the origin means that $+\Phi_0$ or $-\Phi_0$ can be stored in the two-junction SQUID with no bias. This device is utilized as the main memory cell of the DRO (destructive read out) memory being designed in the IBM Zurich research laboratory and described by Guéret, Moser, and Wolf [11].

For logic applications the mode overlap limits the usefulness of the device. For this purpose the three-junction SQUID with the threshold curve shown in Fig. 6(b) was introduced by Zappe in 1975 [16].

The use of this split-feed three-junction device with current ratios of 1:2:1 in the individual junctions, coupled with the proper choice of the LI_0 and feed points, leads to a large open area between lobes which is desirable for large operating margins of the device when used in logic circuits.

The third type of device is a two-junction SQUID in which the input currents are directly injected, rather than the control signals being transformer coupled through an insulated overlying control line. These are the nonlinear current injection devices. The device is basically an adder of inputs A and B, in such a way that if either input is present singly it requires a relatively large signal level for the device to switch, but when the two signals are present simultaneously a low threshold to switching exists. This approximates well the ideal AND function. These devices are described by Gheewala in this issue [17].

A major distinguishing feature between the current injection device and the previous two types is that those have excellent isolation between input and output, whereas in current injection the isolation between input and output is poor. Consequently, the injection devices are always used in combination with the electromagnetically coupled multijunction devices to provide the required isolation between input and output.

All of the multijunction device structures can be realized in at least two ways, the "bridge" structures and the "planar" structures. In the bridge devices, the interconnecting inductance forms a "bridge" between the two junctions; *i.e.*, the counter electrode which is common to both point junctions is lifted away from the ground plane by interspersing a layer of relatively thick insulation. The junctions can be defined entirely by the insulation window or only on two sides. The major advantage of the bridge structure is that it occupies a relatively small area. However, the interconnecting inductance between the junctions is formed about equally by both the base and counter electrodes, which makes the structure difficult to damp should resonances pose a problem.

The parallel *RLC* circuits, which the multijunction devices clearly are, can resonate at discrete frequencies. A two-junction device forms a single mesh circuit and thus has a single resonant frequency, whereas the three-junction logic device has two resonant frequencies. The significance of these resonances is that, through the first and third Josephson equations, these voltages are frequency modulated with resulting *zero-frequency* sidebands. These manifest themselves as steps in the device *I-V* characteristic. If their amplitude is sufficiently large, an output load line may intersect the steps and result in incomplete current transfer. The solution to this problem is to damp the *RLC* circuit appropriately. Zappe and Landman have analyzed this problem in general for a simple but practically interesting case of the symmetric two-junction device [18]. They find that the resonance amplitude depends strongly on circuit Q. The amplitude increases as Q increases and reaches a maximum of 0.6 $I_m(0)$ and subsequently decreases. In the high-Q region, although the resonance amplitude is statically small, it is found to be a function of the device current rise time having the same 0.6 $I_m(0)$ maximum for sufficiently rapidly rising device currents. In the low-Q region, however, the resonance contribution to the current is broad in voltage and low in amplitude. Zappe and Landman derived the optimum damping conditions for the low-Q case and verified the analysis experimentally [19].

The necessary damping conditions can be achieved in practice by a resistor in parallel with the interconnecting inductance. To achieve this in practical device structures, the planar interferometer structure was devised in which the interconnecting inductance is formed almost entirely by a strip line of the base electrode over the ground plane [19]. Properly damped three-junction devices for logic have been designed by Geppert *et al.* in a 5-μm technology [20]. That paper contains a detailed device model, enumerates the device parameters, and makes a detailed comparison between the calculated threshold characteristic and the actually measured one. Excellent agreement is obtained, implying highly accurate modeling. An advanced 2.5-μm minimum linewidth design, in which ground plane holes are utilized to increase the base electrode inductance per unit length, results in a very small device. It is described by Gheewala [17, 21]. Gheewala's paper also contains a description of the injection gates and shows that they are also well modeled.

The important device parameters are, of course, the threshold characteristic (determined by the LI_0 and the injection points), the device *I-V* characteristic from which $I_m(0)$, R_J, and V_g can readily be determined, and the device capacitance, including parasitics to ground. All parameters, with the exception of the inductances and the capacitances, are readily measurable. The device capacitance is most readily determined as a fitting parameter to the analysis of resonance structures.

The major parameters to be calculated are the self- and mutual inductances of various superconducting films disposed over a continuous or discontinuous superconducting ground plane. For large aspect ratios (films close to the ground plane as compared with their width), the inductance per unit length is given as $\mu_0 D/w$, where D is the oxide thickness plus the sum of the penetration

depths of the ground plane and the superconducting film, w is the width of the film, and μ_0 is the permeability of free space. For small aspect ratios, analytical formulae and numerical procedures have been developed by Chang [22] and by Alsop *et al.* [23]. The numerical analysis programs for self- and mutual inductances of such structures can also be applied to the modeling of package components, as demonstrated with considerable success by Jones and Herrell [15].

Overview of logic circuits

- *Introduction*

The earliest example of a logic circuit was the flip-flop circuit of Matisoo [5]. This circuit consisted of two two-input in-line gates interconnected in parallel with a superconducting line. This simple circuit can serve a variety of functions and can be operated in a number of different ways. For example, with a dc supply current, loop current can be flipped back and forth between the two parallel branches on the application of inputs to one or the other device. Under these conditions it forms a common storage element, the flip-flop.

With two inputs present at each device, the circuit can perform the two-input OR and two-input AND functions. This simple circuit can be used either in the "dynamic" or in the "clamped" mode. In the "dynamic" mode, current transfer to the inductive load is initiated by the first device crossing its threshold and current transfer proceeding according to the *RLC* dynamics, either in an underdamped, overdamped, or critically damped mode. The usual design choice is for the critically damped operation.

In the "clamped" mode of operation such a large input signal to the driving device is provided that it is essentially clamped into the resistive state and held there until current transfer is complete. To achieve this latter mode of operation, an additional device design constraint is imposed; namely, that the zero-voltage Josephson current can be suppressed to zero with reasonable input currents.

In the only present use of this circuit it operates in the "clamped" mode. The address driver and decoder circuits of the DRO memory utilize these circuits with specially shaped in-line devices to achieve the zero-voltage current suppression [11].

By switching the supply current on and off as needed, this same circuit can be made to trap a circulating current of a polarity determined by the polarity of the supply current, and thus store information as the prototype of a memory cell. The early memory cells designed by Anacker [6] and experimentally investigated by Zappe [24] were of this type. In each of these circuits a fan-out or a read-out device must be present in order to usefully obtain the functions described.

This class of circuit has not found use in a random logic environment because circuit dynamics in the "dynamic" mode must be customized for each circuit configuration; conversely, if these circuits were operated in a "clamped" mode, very large current gains would be required. These are unavailable. Furthermore, when compared with circuits about to be described, it is slow in operation (current transfer times are on the order of 100 or more ps, depending on the size of the circuit). These circuits are, however, very low-power circuits, which is the motivation for using them in the DRO memory.

To overcome the shortcomings of this type of circuit, Anacker and Matisoo [25] invented a class of circuits in which the devices drive a superconducting transmission line terminated in its characteristic impedance.

- *Terminated transmission line logic circuits*

The principles of terminated transmission line logic are simple. A device (or devices) is (are) loaded by a superconducting transmission line of characteristic impedance Z_0. The transmission line is terminated resistively with a resistor $R = Z_0$. Fan-out devices are distributed along the superconducting transmission line, which passes over the fan-out devices and acts as control or input to these devices. The major advantage of this circuit arrangement is that it represents the fastest possible signal transmission between devices, because the propagating current wave reaches its final value immediately. The fan-out is essentially unlimited.

The superconducting transmission lines, which are formed by the ground plane and the overlying superconducting line separated from one another by one or two dielectrics, Nb_2O_5 and SiO, are the ideal structures for transmission of signals. These lines have zero dc resistance and are nearly lossless and dispersionless until the frequencies approach the gap frequency. This frequency for lead is approximately 10^{12} Hz, which means that for rise times longer than a picosecond or so the transmission line can be viewed as ideal. Another significant feature of a superconducting line over a superconducting ground plane is that crosstalk between adjacent lines is minimal.

The characteristic impedance of the line is primarily determined by the dimensions and the dielectric properties, although the superconducting penetration depth plays a role. Similarly, the delay per unit length τ is determined primarily by the dielectric constant of the insulation, with the penetration depth of the superconductors entering

weakly. Complete expressions for Z_0 and τ are given by Gheewala [17].

The transmission line is terminated with a normal metal resistor, $AuIn_2$. For all rise times of interest the ac and dc resistances are the same, so that the terminating resistor can be chosen with no need to account for skin depth effects.

In-line multicontrol circuits With this circuit approach, several ways of performing logic functions have been devised. The earliest of these utilized the multicontrol in-line gate with three inputs. The principle is illustrated in Fig. 7. Inputs *A*, *B*, and *C* are all designed to be equivalent. Fig. 7(c) shows that by proper choice of input polarity with respect to the bias current I_g, a single three-input device can perform the OR function and the three-input AND and INVERT functions. The inverter is formed with the data input applied in the antiparallel sense as compared with the bias current followed by a timed "1" level in the parallel sense. Thus a single device driving the output constitutes the entire logic family needed to perform all logic functions.

This type of circuit was first investigated by Henkels [26], who carefully designed the output lines to have a constant characteristic impedance, even when they cross the fan-out devices. Detailed comparisons were made of the current waveforms as simulated and as measured with excellent results. The output impedance of the circuit was very low, approximately 0.5 Ω. The circuit rise time was measured to be 165 ps, very short for a circuit made with a minimum linewidth of 50 μm.

These circuits were utilized to construct relatively complex logic functions by Herrell; first, a one-bit adder circuit [27], and second, a serial four-bit multiplier [28]. The adder, fabricated with a 25-μm minimum linewidth technology, consisted of seven interconnected circuits with logic delays per stage of about 125 ps in this environment. The four-bit multiplier consisted of approximately 50 interconnected circuits, fabricated with a 25-μm minimum linewidth with loaded logic delays ranging from 236 to 275 ps per gate for fan-outs of one and four respectively. The average power dissipation per gate was 35 μW including the power supply. The multiplication was performed serially, with a four-bit adder with ripple carry, and a four-phase eight-bit accumulator shift register. The circuit functioned properly under all data conditions and operated with a minimum measured add-shift cycle of 6.6 ns, giving a four-bit multiplication time of 27 ns. These results were limited by the external test equipment. Simulations suggested that the add-shift cycle could be as short as 3.0 ns with a corresponding multiplication time of 12 ns.

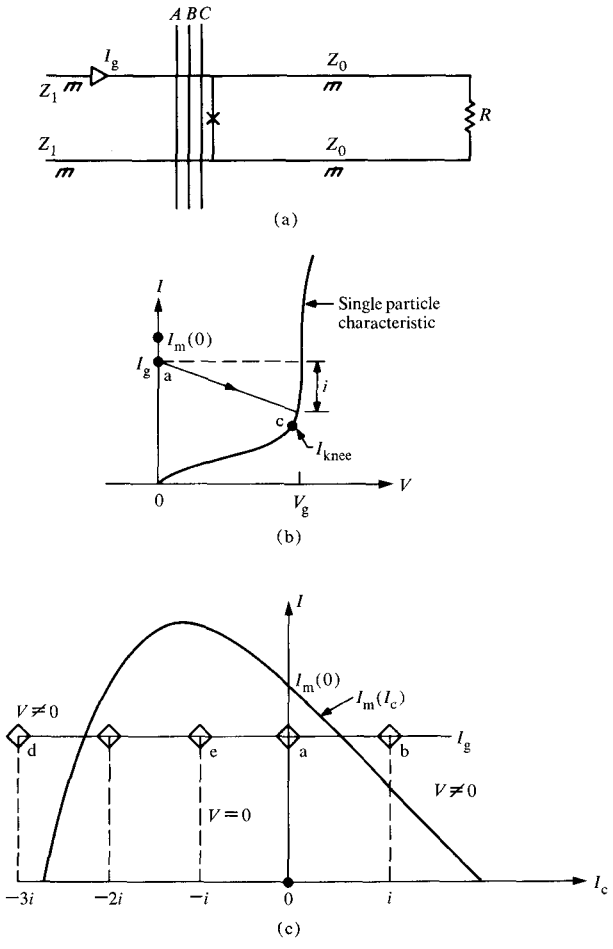


Figure 7 Principle of multicontrol circuits after Herrell [27]: (a) basic logic gate; (b) [and (c)] typical I_g *versus* V_g [and I_g *versus* I_c] characteristics observed for nonlinear in-line Josephson tunneling gates.

In both the adder and multiplier circuitry, bias currents were supplied externally to a group of serially interconnected circuits. This somewhat unusual power supply arrangement was subsequently found to be unsatisfactory for high-performance machines [29], and has been replaced by a transformer-coupled, locally regulated voltage supply. This will be described in somewhat more detail in a subsequent section.

The shift from 25-μm to 5-μm minimum linewidth technology, which occurred in 1977–1978, led to a number of significant changes in devices, circuits, and power supply, and to the definition of storage register elements and clocking philosophy. The latching nature of these logic circuits permits novel approaches to the solution of classical high-performance computer design problems; namely those associated with power supply regulation and disturbance and clock skew.

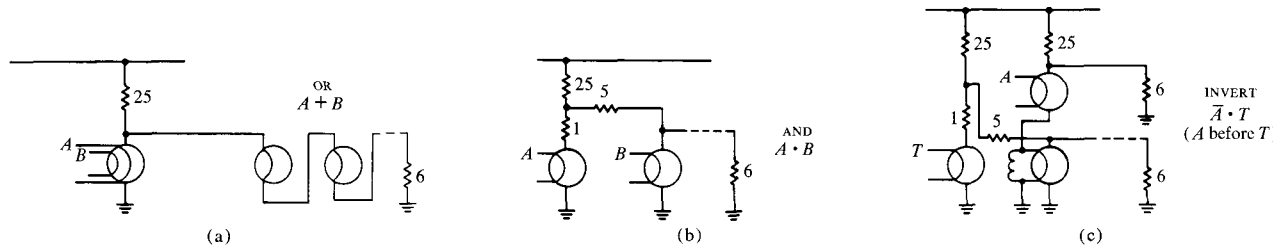


Figure 8 The 5-μm logic family as described by Klein and Herrell: (a) OR circuit; (b) AND circuit; (c) INVERT circuit.

Multidevice circuits As already discussed in the context of devices, the in-line multicontrol devices do not scale well. In a 5-μm linewidth technology they are too large, have too large a capacitance (device as well as parasitic), and are consequently too slow to be desirable. This led to a shift to multijunction devices or SQUIDs. With a 5-μm linewidth technology, technological difficulties in linewidth control are much more apparent and become a significant design issue. The multicontrol AND circuit, when scaled to 5-μm technology and using SQUID devices, was found to be marginal with respect to current density tolerances. As a consequence, a new multidevice AND circuit which had significant improvement in margins was devised by Zappe [13].

The complete set of the simplest members of this 5-μm technology logic family is shown in Fig. 8. Detailed functional, delay, and dynamic power supply analyses and experiments of this family have been performed by Klein and Herrell [30]. The basic building blocks of this family are a two-input OR, a two-input AND, and a timed inverter. These circuits utilize a basic three-junction 1:2:1 device with LI_0 of approximately $\Phi_0/4$, where L is the loop inductance and I_0 is the critical current of the smaller junction in the loop. These circuits operate with the supply current initially shorted to ground through the devices. When the devices switch, the "1" level current is transferred to the load which consists of the superconducting terminated transmission line of $Z_0 \approx 7\ \Omega$ controlling the fan-out gates. For these circuits the "1" level is approximately 0.2 mA with $I_m(0) = 0.4$ mA and a nominal supply current level I of 0.32 mA. Logically, the circuit of Fig. 8(a) is a two-input OR, for the device will switch with inputs to either A or B.

The AND gate in Fig. 8(b) has two devices in parallel, both of which must switch to the resistive state in order for an output to be developed at the load. With two control lines available at each gate, a logic function (A or B) and (C or D) can be generated. Fig. 8(c) shows the timed inverter circuit. It is an OR and an AND combined in such a way that the signal to be inverted is supplied to input A, whereas the timing pulse is supplied to input T. If a "1" is present at A, the AND is disabled and the subsequent timing pulse generates no output, performing the inversion, and conversely.

Klein and Herrell designed a series of experiments in which the delays and delay components of these circuits could be experimentally evaluated and compared with simulation results. A number of OR and AND gate chains were measured. The results give an average OR delay of 43 ps and an average AND delay (with the B input before the A) of 105 ps. The AND gate delay depends upon whether the signal is applied first to input A or B. Only the slower mode of B before A was measured. The delays of the faster mode of A before B were obtained by simulation after the experimental cases were found to agree with simulations. The simulated fast mode of A before B has a delay of 73 ps per stage. In these measurements the logic was performed correctly. The fan-out was 1. Fan-out delay was measured with the OR gate loaded with 11 OR gates having each pair of control lines connected in series, giving an effective fan-out of 22. The crossing inductance was separately determined. The delay per fan-out was measured to be 14.3 ps as compared with the predicted delay (for this particular case) of 14.4 ps. In these circuits the inductive discontinuity was tuned out by means of matching capacitors C_m to ground at the input of each fan-out gate. Provided that the signal rise time is $\gtrsim 3L_c/Z_0$, where L_c is the crossing inductance, Z_0 is the characteristic impedance of the line, and $C_m = L_c/Z_0^2$, the inductive discontinuity is tuned out, resulting in the entire output of the circuit being a uniform matched transmission line in which no reflections occur.

The delay measurements were compared with detailed simulations and yielded essentially perfect agreement. Careful analysis of the simulation and experimental results showed that the logic circuit delay consists of essentially four components. These four components are: a) the rise time, given by some fraction of the CZ_0 time constant, where C, the device capacitance including the parasitics to ground, was approximately 3 pF for these de-

vices; b) the propagation delay, which is approximately 0.013 ps/μm; c) the fan-out delay, given by L_c/Z_0; and d) the turn-on delay, which is the time lapse between the input signal crossing the threshold and the time at which the output signal begins its exponential rise. This turn-on delay has been analyzed in detail by Harris [31], who showed that it depends strongly on overdrive, becoming small with large overdrives. The turn-on delay is proportional to $(C/I_0)^{1/2}$, where C is the device capacitance.

In these experiments the circuits were powered with a trapezoidal bipolar ac power supply generated externally. Because these circuits once switched to the resistive state remain latched until the supply current is reduced, combinatorial logic based on these gates makes a transition from the "0" state to a "1" state only, followed by a reset of the complete combinatorial network.

It is extremely important to note that single logic levels are not clocked; rather, synchronization occurs at latches. Data are read from latches at the beginning of a cycle, propagate asynchronously through the logic, and are again synchronized at the output latches. The next few paragraphs describe the power supply, latches, and clocking concepts and experiments carried out thus far.

Power supply and network The power supply must provide a well-defined, well-regulated current level to the logic circuits, minimize disturbs, provide for reset and power-up at all portions of the machine as simultaneously as possible, maximize the duty cycle, keep current and voltage levels reasonable throughout the system, and add minimally to the total power dissipation of the system. These requirements are met with a power supply system originally conceived by Fang and Herrell [32]. Design details have been provided by Herrell, Arnett, and Klein [33] and by Arnett and Herrell [34]. This power distribution scheme provides switched, regulated power supply of alternating polarity to the logic circuits. No rectification is necessary since the devices are symmetric with respect to the supply current (and control current) polarity interchange. The master power supply at room temperature provides sine wave power of one-half the machine clock frequency (inverse of machine cycle time) through a tree of thin-film transformers with single primaries and multiple secondaries. Use of such a tree maintains low current levels throughout, and results in a small amplitude and phase skew for the power at the chip level. The sine wave power signals are clipped on-chip to the desired voltage level (typically $\approx$12 mV) by voltage regulators consisting of several large-area Josephson junctions in series. From the regulators, voltage is distributed across the chip by voltage busses. The supply resistors between the regulated voltage bus and logic circuits define the operating current bias.

On-chip, excellent voltage regulation and minimal disturbs are a result of three principles used in the on-chip power system design. First, power dissipation in the supply resistors is approximately an order of magnitude larger than that in the circuit "1" state. This means that the overall change in the power drawn from the power system as the circuit switches between "1" and "0" states is minimal, leading to a small disturb. This is reduced further by local regulation in which the inductance of the power bus between the regulator and the circuits is kept small. The resulting overall disturb is negligible.

The power bus impedance and length are carefully chosen to give good frequency response with minimum ringing, so that the power supply voltage can be established and removed with no overshoot or ringing.

The obtainable duty cycle is determined essentially by the amount of additional power one wishes to dissipate in the regulators. For example, in the circuits described by Klein and Herrell the power supply voltage was 8.2 mV with a supply resistor of 25 Ω. Thus, the logic circuit dissipation was $\approx$2.7 μW. When operated at 100 MHz frequency (for a 5-ns machine cycle time), the average dissipation per circuit including on-chip regulation is $\approx$4.5 μW with a duty cycle of 0.75.

The transformers in the power tree are formed by two overlaid superconducting lines crossing a hole in the superconducting ground plane. Transformers designed in the experiments described by Arnett and Herrell are one-to-one transformers with a mutual inductance of 325 pH and a coupling coefficient of 0.93. The transformer length was $\approx$700 μm with a width of 80 μm. It should be emphasized that a key aspect of the power system design is to minimize phase and amplitude skews of the power supply waveforms between different chips. Simulations of a JSP-sized power system indicate that amplitude and phase skews are indeed negligible. However, this remains to be confirmed experimentally.

Given that essentially simultaneous power-up and down of the logic circuits will occur throughout the machine, it is natural to tie the clocking to the power supply. This has indeed been done. At the beginning of a machine cycle, the machine state is contained in data registers and control latches. As the power comes up, the data are automatically gated from the latches and registers to the combinatorial network and thence to the output registers and latches, which are updated if control signals are present. This automatic gating is achieved by a

circuit called the self-gating AND (SGA), which locks to the data state of the latch at the beginning of the power cycle and holds that state regardless of subsequent changes in the state of the latch, permitting the latch to be updated as required. Thus, the logic is hazard- and race-free. (An exception to this is the use of the INVERT circuit. It is a timed inverter, in the sense that the data must be known valid before inversion, otherwise an error results. The timing pulse is appropriately delayed with respect to the data.) Such an SGA, in conjunction with a dc-powered flip-flop and input circuits to form latches, has been experimentally investigated by Davidson [35].

A condition to be satisfied by these latching logic circuits when operated with a bipolar power supply is that they reset with certainty to the zero-voltage state at the end of each cycle. It is well known [36, 37] that if the rate of change of the supply current exceeds a value I_{min}/τ, a nonzero probability exists that the circuit will not reset to the zero-voltage state upon transition of the supply current through zero to the negative portion of the cycle. If the circuits indeed behave as point junctions, I_{min} is just given by the Stewart-McCumber formula [10] and τ is the effective RC time constant consisting of the total junction capacitance and the load resistance. The measured limiting rate for the circuits used by Klein and Herrell is greater than 2-3 mA/ns, whereas the rate expected for a 5-ns machine cycle operated with 80% duty cycle and a 330-μA current level is only 0.5 mA/ns.

In summary, the 5-μm logic circuits of Klein and Herrell, when placed on a logic chip with an appropriate number of wiring channels, would result in approximately 300 circuits per 6.35 × 6.35-mm chip. The power dissipation per chip would be $\approx$1.5 mW including power supply and power supply regulation. The resulting power density is well within the capabilities of direct heat removal to helium, which would permit the dense three-dimensional packaging described by Brown in this issue [38]. It is estimated that machines with a 5-ns cycle time could be built with these circuits in which power is supplied at 100 MHz frequency and the operating duty cycle is 80%.

The package The package concept described by Anacker [39] and Brown [38] mounts chips on cards, plugs cards into a board, and interconnects cards via wiring modules. Mechanically, all major parts are made of silicon. Electrically, throughout the package superconductive transmission lines are used except at the chip-to-card interconnection, at the space expander interconnection, and at the card-to-board-to-wiring module interconnection. The properties of the superconducting transmission lines are the same as those on-chip and are made with the chip technology. The three classes of interconnection, however, constitute electrical discontinuities, especially to the extremely rapid rise times of the signals propagating from chip to chip on different cards. It is essential that the electrical properties of these interconnections be carefully modeled and experimentally verified. Only then can the necessary noise tolerances of the intercommunication circuits be specified.

The electrical characterization of the chip-to-carrier interconnections, *i.e.*, the so-called "controlled collapse chip connector" (C4) connections, for a single peripheral row of 84 C4s per chip with a spacing of 200 μm between centers, has been carried out by Jones, Herrell, and Yao [40] and is described in detail by Jones and Herrell [15]. Physically, this connector is superconducting solder of approximately cylindrical shape with a height of 30 μm and a diameter of $\approx$120 μm. These rather gross physical structures have a self-inductance, as well as mutual inductances to all of the other C4s on the chip. The inductance matrix can be evaluated as a function of the placement of the ground connectors by programs written by W. H. Chang [41]. Groups of four, six, and eight signal connectors between two grounds were considered. The analysis showed that grounds effectively isolate one group from another. Measurements of self- and mutual inductance were made on groupings corresponding to those calculated, with essentially perfect agreement between calculation and measurement.

The self- and mutual inductances are a function of the pin position between grounds. For example, for the grouping of six, the self-inductance of the center pin has the largest value at approximately 23 pH, whereas the end connectors closest to ground have a self-inductance of approximately 13 pH. The mutuals range downward from about 20 pH. Having established the inductance matrix, it then becomes possible to calculate the delay through the connector as well as the crosstalk which should be experienced under any set of experimental conditions. Jones and Herrell describe in detail the delay and crosstalk measured for specifically chosen signal I/Os. In this experiment signals were generated on one chip, propagated through three others, and returned to the same chip. The measured delays and crosstalk were compared with simulations utilizing the inductance matrices measured for these same configurations. Essentially perfect agreement was found here as well with a delay of about 10 ps per connector for the central one in the array of eight I/Os per ground. A worst-case crosstalk of 12% was measured in the array of eight I/Os per ground when five signals arrived essentially simultaneously at the connector site. Although specific to this particular experiment, these results form the basis for the eventual physical design of the connector structure and for the determi-

nation of the ground configuration necessary to meet a set of electrical objectives. It appears that delay through connectors is not a significant factor, and one would have to focus on the crosstalk as the major C4 issue.

The remaining package discontinuities, namely those at the space expander and the card-socket/wiring module, remain to be characterized. However, based on the geometry of those structures as compared with the C4s, it is expected that they will form a more significant discontinuity to the electrical signals than the C4s.

Current injection logic circuits Recently, Gheewala has devised a set of logic circuits based on the nonlinear current injection device previously described [42]. Because of the nonlinear addition of inputs, both the OR and the AND gates have excellent operating margins. Additionally, these logic circuits have high gain and overdrive capabilities. High gain results from switching of two units of gate current to the output (for a two-input circuit). This leads to shorter delays and the possibility of parallel fan-out. The overdrive capability, limited in the circuits described by Klein and Herrell, is large because in the CIL circuits the control current is applied only after the supply current is established in the isolation devices. Thus the amplitude of the control current has no upper bound. The overdrive capability results in short turn-on delays.

Detailed designs of these circuits have been made in a 2.5-μm technology and the circuits have been experimentally realized. Gheewala describes the design and experiment in this issue [17]. The logic circuits are very fast, having an average nominal logic delay of 36 ps per gate for an average fan-in of 4.5 and a fan-out of 3. The corresponding average power dissipation is 3.4 μW per gate. Gheewala has compared experimental measurements and simulations over a broad range of supply currents for both two-input and four-input OR and AND gates of varying fan-out, with excellent agreement found throughout. These experiments also contain a measurement of a 13-ps-per-gate delay, in a two-input OR circuit when operated at high bias. This measurement is of considerable significance, because the fan-out delay and propagation delay in this circuit are 7 ps, which means that the delay through the two-input OR circuit itself is only about 6 ps, the first known instance in which circuit delays of under 10 ps have been experimentally measured in any technology. The fact that the models still predict measured delays correctly suggests that logic circuit performance can be extrapolated to yet a faster speed range as the linewidth is reduced from the present 2.5-μm level. Indeed, Zappe [43] has projected the performance of logic circuits as linewidths are scaled to the submicron level. He finds that by retaining the present-day vertical structure, gate delays can be decreased to a few picoseconds with power levels of less than 1 μW per gate.

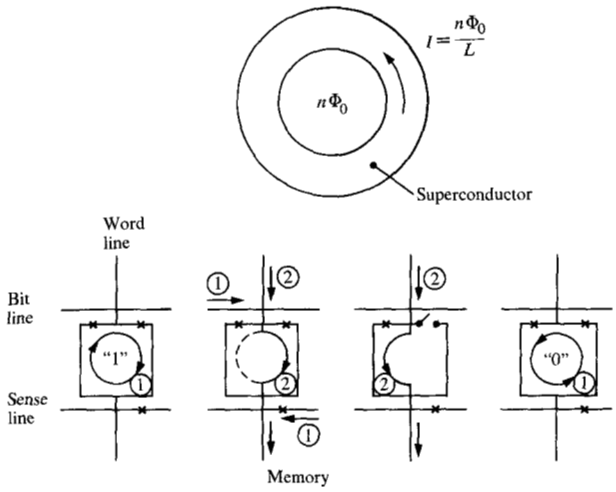


Figure 9 The basic NDRO memory cell idea.

To take advantage of the very high logic circuit performance already demonstrated by Gheewala, compatible packaging structures must be devised, essentially by extending the present 5-μm packaging scheme so that self- and mutual inductances of discontinuities are correspondingly smaller. Equally important, the memory technology must keep pace.

In what follows, we present an overview of the NDRO (nondestructive read out) memory work aimed at high-speed cache and the DRO (destructive read out) dense memory chip aimed at a main store.

Overview of NDRO cache memory

As already indicated, the fundamental idea is to store information by means of circulating, quantized, persistent currents in superconducting rings. It remains only to provide a means of selectively writing and reading the information. The basic principle for a bit-organized NDRO memory array is shown in Fig. 9. Here the memory cells are arranged in a rectangular array of rows and columns. The column line constitutes the word line, the upper row line the bit line, and the lower row line the sense line. Information is stored with a clockwise circulating current representing a "1" and a counterclockwise circulating current representing a "0." Information is carried according to the polarity of the word current. The bit-line current is unipolar, as is the sense-line current. Writing is performed by coincidence of the bit- and word-line currents, and reading is performed by coincidence of word-line and sense-line currents. Reading is nondestructive in

that the sense-line device merely senses the presence or absence of the control current immediately above it. In this case the cell contains three devices, in reference to a relatively old in-line gate design. Modern NDRO cells contain two devices and have a somewhat more complex accessing arrangement to improve write and sense margins.

The early NDRO cell experiments, such as those of Zappe [24], emphasized important questions concerning the individual cells such as cell dynamics, the ability to repeatedly read nondestructively, and so on. The necessity for critically damping the cell was demonstrated, along with the capability of repeatedly reading with no diminution in the quantized circulating persistent currents. In some experiments attempts were made to emphasize cell density [44, 45]. However, these early cells utilized the in-line device for both writing and sensing. Consequently, with miniaturization the current densities were driven to extremely large values, in some cases in excess of 20 kA/cm^2. Nevertheless, the basic cell was found to behave properly, store information as required, read nondestructively, and have very short (approximately 100-ps) current transfer times. Thus the cell itself was clearly shown to be a building block for a very high-speed memory.

NDRO memory development using in-line devices culminated in the design and testing of an experimental, fully decoded 64-bit random access memory chip, described in detail by Henkels and Zappe [46]. In this memory the cell is a ring containing a single diamond-shaped in-line write gate. The shaped devices suppress resonances which are observed to cause erratic current transfer [24]. Information is still stored as clockwise and counterclockwise circulating currents. The stored flux is $\gtrsim 100\Phi_0$, so quantization effects are small. With the elimination of one write gate, the bit current also becomes bipolar. The cell area is large, with no effort made to minimize it. Selection is with coincident current along the word and bit lines.

The two decoders, a word and a bit/sense decoder, are tree decoders. A decoder section consists of two branches, each containing four series-connected decoder gates. The gates are controlled by two lines, a dc-bias line shared by all the gates throughout both decoders, and a bipolar address line unique to each decoder stage. The idea in each stage is to always have one branch superconducting, with the application of the address current, which adds to the dc-bias line current, maximizing the control signal, and the other branch superconductive, where the address current subtracts from the dc bias, eliminating any control. The four gates per branch are intended to increase the resistance in the resistive branch, leading to a smaller L/R time constant for each decoder stage. Even so, the decoder turns out to be the weak link in these experiments. The overall access time of this array fabricated into 25-μm technology on the usual 6.35 × 6.35-mm^2 chip was measured to be approximately 4 ns (a cycle time of 5 ns) with full operating margins. With somewhat reduced sense margins the chip was operated at an access time of 2.3 ns and a cycle time of 3.5 ns. Simulation of the operation of the decoder and array lines with very simple models gave a good description of the decoder dynamics. Furthermore, it was found that the diamond-shaped in-line devices utilized as write and sense gates successfully eliminated the junction resonance problems previously observed [24]. The experiments also revealed that tree decoders suffer from leakage problems, which results in currents in nonselected paths.

This memory chip was for some time the most complex chip fabricated in that it contained a total of 322 individual devices.

These experiments did not emphasize cell density, memory access time, or cycle time, but rather demonstrated overall principles. As the technology developed to the 5-μm and 2.5-μm level, design emphasis shifted to the design of memory chips potentially usable in actual machines; consequently, margins, density, access time, and cycle time all became important design parameters with the usual resulting engineering compromises necessary to obtain a satisfactory design.

The major emphasis in the new designs was to improve the performance of the peripheral circuits. In the 64-bit memory array work, the decoder delay was more than an order of magnitude longer than the cell current transfer time.

General considerations are sufficient to indicate that small peripheral delays require low current levels. If terminated transmission lines are utilized to perform peripheral logic, the arguments are precisely the same as those in logic circuits, suggesting that low current levels be utilized. If, on the other hand, the peripheral logic employs simple current steering, the delays are to first order given by LI/V_g, where L is the inductance of the superconducting loop, I is the current transferred into the loop, and V_g is the gap voltage. To correctly utilize small current levels, interferometer devices must be used; consequently, the memory cells utilize interferometers as write and sense gates.

Detailed design considerations, and their experimental verification of small NDRO cells utilizing low current lev-

els, have been described by Henkels [47]. Henkels' cell contains a 1:2:1 bridge device as the write gate, and a two-junction device as the sense gate. The cell stores $8\Phi_0$ in the clockwise and counterclockwise mode. Cell dynamics is controlled by an external damping resistor which results in a slightly underdamped cell. No initialization cycle is required.

Writing is accomplished by a triple coincidence of I_Y (bipolar), I_X, and I_D, where Y, X, and D refer to the Y line, the X line, and the Diagonal line, respectively. Utilization of triple coincidence increases the discrimination between the selected and unselected cells, because the selected cell sees three currents, whereas the unselected cell sees only one. Further, the dual controls halve the total current, which implies the possibility of faster peripheral circuits.

Reading is accomplished in the usual manner of coincidence of I_Y and I_S. Discrimination is maximized by an asymmetric cell in which the write gate is contained in the smaller branch having roughly one-third of the total ring inductance.

The experiments confirmed the design in detail. Specifically, the margin evaluation revealed that the analysis based on quasistatic threshold characteristics is indeed valid, that the design approach to minimization of I_{min} by slight underdamping is indeed correct, and that provided $\Phi_0/LI_0 \geq 12$ is satisfied for the write gate, then the resonant voltage $V_r > V_{min}$. Under these conditions, the write gate can switch into the resonance without erratic current transfer occurring. This was a criterion established on theoretical grounds and was in fact supported by experiment.

In addition to verifying that the cell operated functionally with no set-up required, a nondestructive read test was performed in which a "1" was read 2×10^{12} times consecutively. The stored flux was $8\Phi_0$ before and after the test.

An unexpected and interesting phenomenon was discovered when performing a series of measurements of I_{min} as a function of effective damping, when the write gate control current level exceeded the level at which the first threshold lobe intersects the minor lobe; then there was no dependence of I_{min} on damping, and its value was uniformly zero. The fact that $I_{min} = 0$ regardless of cell damping is potentially a most useful phenomenon for increasing the cell tolerances. The explanation for this effect is that when the operating point of the write gate crosses the threshold curve from the central lobe, the interferometer either absorbs or emits a flux quantum. During this transition a voltage is developed, destroying the circulating current and leading to $I_{min} = 0$.

In conjunction with the memory cell design, a compatible decoder with a self-contained address register was also designed and experimentally investigated. Details of this work have been reported by Faris [48]. The decoder, called a "loop decoder," has the following features. The address register is formed from a series of superconducting loops into which the address is first established. In this design true and complement address inputs are required. For the decoder realized in the 5-μm technology, the address is established in 180 ps. The decoding, which is subsequently initiated by a start-decode pulse, requires ≈30 ps per stage. The experimentally realized decoder incorporated address loops sufficient for six-bit decoders, but only three bits were actually utilized. The experiments clearly establish the compatibility of this decoder with the memory cells so that an access time estimate of 1.8 ns for a 2K-bit array has been made [49].

As with logic, the memory technology has moved rapidly from a 5-μm minimum linewidth to one of 2.5 μm. As a consequence, present activity centers about the design of a 4 × 1K-bit cache chip in the 2.5-μm technology. An experiment which Henkels executed [47] was to operate the 5-μm memory cell in the 1,0 mode. In this mode a "1" is a circulating current and a "0" is no circulating current [50]. A desirable feature of such a cell is that all currents are monopolar; consequently, no polarity inverter circuitry is required. Also, the diagonal line can be replaced by a second Y line running vertically, which simplifies decoding and saves space. The undesirable feature of this mode is that the sense discrimination is inherently smaller than it is for the 1,−1 mode. However, Henkels' discovery, in which $I_{min} = 0$ independent of the damping resistor value (provided that the conditions for this are met), partially overcomes this inherent disadvantage and has led to a reconsideration of this mode of operation and its adoption as the cell in the 2.5-μm design.

The basic design of the cache memory chip is described by Faris, Henkels, Valsamakis, and Zappe in this issue [51]. The design objective is a 4K-bit cache chip with a nominal access time on the chip level of ≈500 ps. The basic design unit is a 1K-bit array. It is intended to place four of these on one 6.35 × 6.35-mm chip. The cell is the 1,0 cell. The decoders are the loop decoders modified for 2.5-μm design so that the requirement for both true and complement addresses is eliminated. A sense bus, which collects information from any selected sense line, is described. The sense bus detects the decay of the sense-line current upon reading a "1" and transmits the signal to the memory-to-logic interface driver. The current levels in all

components are nearly those used in logic with correspondingly short delays and the elimination of the need for amplifiers in the logic-to-memory interface. The flux stored in the cell has been further reduced to a value of $2\Phi_0$. The cell current transfer time is expected to be $\approx$35 ps, the addressing delay $\approx$100 ps, with the decoding delays of 20 ps per stage. The current transfer time from the sense bus is expected to be 95 ps, with the resultant overall access time under 500 ps. The paper by Faris *et al.* [51] describes the design of the interrelated components required for a full memory chip. This design has yet to be experimentally verified.

In the memory package, as described by Brown in this issue, the memory chips are to be bonded to cards with the same techniques as those used for bonding the logic chips. The electrical requirements imposed by memory on the package are appreciably less severe than those demanded by logic. Therefore it is expected that the package will be more than satisfactory for memory modules.

Overview of the DRO main memory

In a paper in this issue Guéret, Moser, and Wolf [11] have provided an overview of the main memory work, as well as previously unpublished work on array components. Hence, the discussion here will be brief.

The major design objectives of the main memory are essentially twofold; first, to pack as many cells on a chip as possible; second, to minimize the power consumption per chip. Speed can be sacrificed to some degree to meet these design objectives. The main memory chip utilizes two-junction bridge-type devices for cells and shaped inline devices as drivers and logic devices, for decoding and control. The logic principle is that of current steering to minimize power.

These ideas have been tested by a main memory cross-section chip consisting of 2048 single-flux-quantum (SFQ) cells, partial drivers, and a portion of the decoders [52]. However, no control or timing logic was incorporated. This cross section was successfully fabricated and in fact proved the basic principles. On the basis of this cross section, it is expected that a 16K-bit chip with an access time of $\approx$15 ns, dissipating $\approx$300 μW, can be built in a 5-μm/2.5-μm technology. This access time is adequate as main memory for a machine cycle time of a few nanoseconds when coupled with a cache with an access time of also a few nanoseconds.

Summary

This paper has reviewed the basic principles on which Josephson computer technology is based, the devices, the development of the logic, cache memory, and main memory. The technology is currently in transition between a 5-μm and 2.5-μm minimum linewidth capability. The designs for devices, circuits, and memory which were made for 5-μm technology and experimentally verified, when combined with the package as described in this issue, should yield a machine with a cycle time of <5 ns, even for fairly large processors.

With the evolution of the technology to 2.5 μm, an improved package is required. This package is, however, simply an extension of the 5-μm package with reduced inductive discontinuities. This technology should yield machine cycle times on the order of a few nanoseconds, even for relatively large systems.

References

1. J. H. Greiner *et al.*, *IBM J. Res. Develop.* **24** (1980, this issue).
2. J. Matisoo, *Proc. IEEE* **55,** 172 (1967).
3. J. Matisoo, *J. Appl. Phys.* **40,** 2091 (1969).
4. J. Matisoo, *Appl. Phys. Lett.* **9,** 167 (1966).
5. J. Matisoo, *Proc. IEEE* **55,** 2052 (1967).
6. W. Anacker, *IEEE Trans. Magnetics* **MAG-5,** 968 (1969).
7. F. Tsui, *IBM J. Res. Develop.* **24** (1980, this issue).
8. C. J. Adkins, *Phil. Mag.* **8,** 1051 (1963).
9. S. Basavaiah, J. M. Eldridge, and J. Matisoo, *J. Appl. Phys.* **45,** 457 (1974).
10. H. H. Zappe, *J. Appl. Phys.* **44,** 1371 (1973).
11. P. Guéret, A. Moser, and P. Wolf, *IBM J. Res. Develop.* **24** (1980, this issue).
12. R. C. Jaklevic, J. Lambe, J. E. Mercereau, and A. H. Silver, *Phys. Rev.* **140,** A1628 (1965).
13. H. H. Zappe, *IEEE Trans. Magnetics* **MAG-13,** 41 (1977).
14. W. H. Henkels, *Appl. Phys. Lett.* **32,** 829 (1978).
15. H. C. Jones and D. J. Herrell, *IBM J. Res. Develop.* **24** (1980, this issue).
16. H. H. Zappe, *Appl. Phys. Lett.* **27,** 432 (1975).
17. T. R. Gheewala, *IBM J. Res. Develop.* **24** (1980, this issue).
18. H. H. Zappe and B. S. Landman, *J. Appl. Phys.* **49,** 344 (1978).
19. H. H. Zappe and B. S. Landman, *J. Appl. Phys.* **49,** 4149 (1978).
20. L. M. Geppert, J. H. Greiner, D. J. Herrell, and S. Klepner, *IEEE Trans. Magnetics* **MAG-15,** 412 (1979).
21. L. M. Geppert, IBM Thomas J. Watson Research Center, Yorktown Heights, NY, private communication.
22. W. H. Chang, *J. Appl. Phys.* **50,** 7030 (1979).
23. L. E. Alsop, A. S. Goodman, F. G. Gustavson, and W. L. Miranker, *J. Comput. Phys.* **31,** 216 (1979).
24. H. H. Zappe, *IEEE J. Solid-State Circuits* **SC-10,** 12 (1975).
25. W. Anacker and J. Matisoo, U.S. Patent 3,758,795, 1973.
26. W. H. Henkels, *IEEE Trans. Magnetics* **MAG-10,** 860 (1974).
27. D. J. Herrell, *IEEE Trans. Magnetics* **MAG-10,** 864 (1974).
28. D. J. Herrell, *IEEE J. Solid-State Circuits* **SC-10,** 360 (1975).
29. D. J. Herrell, IBM Thomas J. Watson Research Center, Yorktown Heights, NY, private communication.
30. M. Klein and D. J. Herrell, *IEEE J. Solid-State Circuits* **SC-13,** 577 (1978).
31. E. P. Harris, *IEEE Trans. Magnetics* **MAG-15,** 562 (1979).
32. F. F. Fang and D. J. Herrell, U.S. Patent 4,092,553, 1978.
33. D. J. Herrell, P. C. Arnett, and M. Klein, *AIP Conf. Proc.* **44,** 470 (1978).
34. P. C. Arnett and D. J. Herrell, *IEEE Trans. Magnetics* **MAG-15,** 554 (1979).
35. A. Davidson, *IEEE J. Solid-State Circuits* **SC-13,** 583 (1978).
36. T. A. Fulton and R. C. Dynes, *Solid-State Commun.* **9,** 1069 (1971).

37. H. H. Zappe, IBM Thomas J. Watson Research Center, Yorktown Heights, NY, private communication. The corrected formula given by Fulton and Dynes in [40].
38. Alan V. Brown, *IBM J. Res. Develop.* **24** (1980, this issue).
39. W. Anacker, *IEEE Spectrum* **16,** 26 (1979).
40. H. C. Jones, D. J. Herrell, and Y. L. Yao, *IEEE Trans. Magnetics* **MAG-15,** 432 (1979).
41. W. H. Chang, IBM Thomas J. Watson Research Center, Yorktown Heights, NY, unpublished results.
42. T. R. Gheewala, *Appl. Phys. Lett.* **33,** 781 (1978).
43. H. H. Zappe, *Proceedings of National Science Foundation Workshop on Opportunities for Microstructure Science, Engineering and Technology*, Airlie, VA, Nov. 19-22, 1978, p. 209.
44. R. F. Broom, W. Jutzi, and Th. O. Mohr, *IEEE Trans. Magnetics* **MAG-11,** 755 (1975).
45. W. Jutzi, *Cryogenics* **16,** 81 (1976).
46. W. H. Henkels and H. H. Zappe, *IEEE J. Solid-State Circuits* **SC-13,** 591 (1978).
47. W. H. Henkels, *J. Appl. Phys.* **50,** 8143 (1979).
48. S. M. Faris, *IEEE J. Solid-State Circuits* **SC-14,** 699 (1979).
49. H. H. Zappe, IBM Thomas J. Watson Research Center, Yorktown Heights, NY, private communication.
50. P. Wolf, *IBM Tech. Disclosure Bull.* **16,** 214 (1973).
51. S. M. Faris, W. H. Henkels, E. A. Valsamakis, and H. H. Zappe, *IBM J. Res. Develop.* **24** (1980, this issue).
52. R. F. Broom, P. Guéret, W. Kotyczka, Th. O. Mohr, A. Moser, A. Oosenbrug, and P. Wolf, *IEEE J. Solid-State Circuits* **SC-14,** 690 (1979).

Received May 18, 1979; revised August 27, 1979

The author is located at the IBM Thomas J. Watson Research Center, Yorktown Heights, New York 10598.