I. INTRODUCTION

In the early part of this decade, Chatterjee et al. considered the implications of continued evolutionary downscaling. They foresaw fundamental problems that lie ahead which will eventually derail the impressive advances made by down scaling ICs.

They concluded that voltage levels must be lowered for devices with critical dimensions below 0.5 \( \mu \)m and cryogenic operation will be required below the 0.25-\( \mu \)m limit. Furthermore, for large chip sizes, delays involved with driving long interconnects will dominate the speed of circuit operation.

It has been proposed that a revolutionary technology based upon quantum effect devices are a possible vehicle to permit continued downscaling. The rationale included the fact that continued downscaling would lead to size scales where quantum mechanics would effect electron transport. Instead of attempting to design devices that operated in spite of quantum size effects, it is proposed that devices could be built whose operation was based upon these effects.

The suggestion that quantum transport effects could be seen in semiconductors dates back to 1957 when Schrieffer indicated that the narrow confinement potential of an inversion layer may lead to the observation of quantum transport behavior. This was shown in 1966 by Fowler and co-workers by low-temperature magnetotransport, which started an intense effort in exploring two-dimensional electron gases (2DEGs). The advent of ultrathin epitaxial film growth techniques ushered in the era of reduced dimensionality physics. These structures are inherently two dimensional; thus investigations concentrated on heterostructures, where quantum effects are due to confinement in the epilaxial direction; i.e., quantum wells (more properly, quantum planes) which are two-dimensional systems.

Now, however, we are rapidly approaching a level of solid-state structural fabrication in which the energy and length scales are such that macroscopic quantum effects are dominant. We have, for the first time, reached a stage where it is possible to observe and perhaps even control large scale quantum effects in a variety of materials and structures. Advances in microfabrication technology have allowed laboratories around the world to impose additional lateral dimensions of quantum confinement on two-dimensional systems with lateral length scales approaching the thickness of epitaxial layers used in quantum confinement experiments.

II. QUANTUM TRANSPORT

There are a number of ways to fabricate one-dimensional electron gases (1DEGs); the most common is to construct a 2DEG. Initial measurements on these structures demonstrated localization, weak quantum interference or universal conductance fluctuations (UCF), and trapping phenomena. However, a tremendous improvement in the understanding of these structures came about when the size of the samples could be made smaller than the "phase-breaking" or coherence and elastic scattering lengths.

In this regime, the structures act like electron waveguides, since the lateral "channel" dimensions are on the order of an electron wavelength and only a few laterally defined modes are occupied. The structures act like large, resonant systems, since the electron wave function penetrates across Classically unallowed parts, like voltage probes and adjacent unconnected structures.

This regime, the limit of electronic transport, is fascinating in many respects. Van Wees et al. (Philips) and Wharam et al. (Cambridge) recently discovered that the conductance of quantum-size 2DEG constrictions becomes quantized in units of \( 2e^2/h \) just before conduction ceases. These are called quantum point contacts (QPCs). This phenomena is due to depopulation of successive one-dimensional (ballistic) conducting subbands, and verifies in general the Landauer formula for the limit of electron conduction. This is a fascinating laboratory for the limits of electronic transport and coupling of single electron subbands.

There have been a number of attempts to utilize this, and
other, quantum conductance and interference phenomena for electron devices. The concept of quantum interference is illustrated in Fig. 1(a); consider a one-dimensional quantum wire, split into two branches, and then recombined in a ring type structure. The electron wave that propagates through this ring first splits when it encounters the ring, then subsequently recombines. How electron waves split and recombine determines the resistance of the structure. In the absence of any perturbation, ground state wave functions will have traveled the same distance and will recombine in phase and produce a low resistance. However, since the two waves propagate in different directions around the ring, their relative phase may be changed by a magnetic field creating an interference effect as the magnetic field is changed. This is the Aharonov–Bohm (AB) effect, and is very pronounced in electron waveguide rings. The resistance of these structures oscillate with the applied magnetic field as the phase is varied through $2\pi$ and the electron waves interfere constructively (low resistance) or destructively (high resistance). A geometrically distinct but operationally similar structure is shown in Fig. 1(b); an electron waveguide “stubtuner,” where the interference now comes from adjusting the “length” (electrostatically) of a stub on a waveguide, just as in microwave circuits a factor of a million larger. A transistor based on this structure has been proposed, and initial effects that may be due to this phenomena have been reported.

These structures demonstrate fascinating physics. There is a new realm of micro- and nanostructure induced phenomena that needs to be explored in order for progress to be made in quantum effect devices. However, are these structures actually useful electronic devices, and if not, what will it take to make a nanoelectronic structure a useful electron device?

There are a few fundamental requirements for digital electron devices, as we know them: gain, fanout, isolation of input from output, inversion, and well defined logic levels. In addition, though not a fundamental requirement, operation at noncryogenic temperatures is desirable. This can be summed up more precisely; equilibrium electronic transport, like that used in most 1DEGs, UCF, QPCs, AB and stubtuned structures, will require cryogenic temperatures, will not have good isolation and will not exhibit gain. This is a consequence of the embodiments of these quantum structures; though the effects are largest in semiconductor structures where the scattering lengths are long (with respect to metallic materials) the transport is still metallic and near equilibrium. Additionally, the nonlocality observed in these structures, when device length scales becomes smaller than the scattering length, becomes a limitation on scaling due to the loss of isolation. Sources and sinks of momentum such as heavily doped contact regions must be used for isolation. One must search for nonequilibrium transport for the necessary requirements.

III. HETEROJUNCTIONS

In order to consider how quantum effect devices could be fabricated which operate in nonequilibrium conditions and might satisfy criteria for useful digital electronic devices we should consider what we need in order to construct an electronic device.

Electron devices control the flow of charge with potential barriers. The two types of barriers in use are PN junctions and electrostatic depletion regions. In both of these cases the abruptness of the barrier which controls the flow of charge is limited by the width of the depletion regions. While the size of these depletion regions is reduced by increasing the doping level, tunneling current between valence and conduction band edges become a limiting factor, and there are practical limits to the amount of impurity atoms which may be introduced into semiconductor crystals.

A heterojunction between two semiconductors with different bandgaps, on the other hand, may be atomically abrupt. Unlike PN junctions the barrier height does not change with applied potential. In order to achieve ultimate limits in downscaling, it seems clear that new device technology must avoid depletion layers which will limit device size and should instead employ heterojunction technology. One

---

**Fig. 1.** (a) Schematic representation of an Aharonov–Bohm ring. (b) A proposed electron waveguide “stubtuner” device.
way to construct electronic devices based on quantum effects is to use heterojunctions to form tunneling barriers and quantum wells. The tunnel barrier is a thin slab of wide band gap material clad by narrow band gap material. When the layer is physically thin enough, (comparable to the electron wavelength) quantum mechanical tunneling through the barrier is permitted.

The quantum well is complementary to the tunnel barrier. It is a slab of sufficiently thin narrow band gap material surrounded by wide bandgap material. Electrons confined to this layer will have quantized energy levels where the ground state energy is higher than the conduction band edge found in a bulk crystal.

While the majority of work in nanoelectronics has utilized compound semiconductors, there is no fundamental requirement that quantum effect devices be limited to compound semiconductors. In fact some quantum effect structures have been demonstrated with the Si/Ge system.\textsuperscript{23} If a silicon heterojunction technology could be developed which included wider bandgaps as well as narrower bandgaps than silicon, then there would be strong incentives to investigate silicon quantum effect devices.

IV. RESONANT TUNNELING DEVICES

In any heterojunction material system, tunnel barriers and quantum wells may be combined to form a rudimentary quantum effect device known as a resonant tunneling diode (RTD)\textsuperscript{21,22} and is schematically represented in Fig. 2. Electron states in the quantum well of the RTD are now quasi-bound since there exists a chance that they will tunnel out of the well. At an appropriate voltage bias across the device the electron state in the well will energetically align with electrons in the emitter region producing a relatively large current flow as electrons tunnel through the structure via the quantum well state.

As the bias is increased to higher voltages the quantum well state drops below the conduction band edge significantly reducing the current in the device. The nonzero current at this bias is the result of inelastic processes that allow electrons to change energy and tunnel through the quantum state. There may also be a component of current from the tails of the Fermi distribution which have sufficient energy to go over the top of the barriers. At higher biases this becomes the dominant current component. The peak to valley current ratio is a frequently quoted figure of merit. Ratios in excess of 30:1 at room temperature and 70:1 at cryogenic temperatures have been demonstrated.\textsuperscript{25}

The potential for high speed resonant tunneling devices is real. Workers at Lincoln Laboratory have demonstrated RTD oscillations at 400 GHz.\textsuperscript{22} This two-terminal device may have some useful high frequency oscillator applications, but (because it is only two-terminal) it is doubtful that it would find use in conventional digital electronics.\textsuperscript{19}

A number of groups have introduced RTDs into some part of a conventional transistor to achieve a three-terminal device. While these integrations have produced some interesting devices, they have not done anything to circumvent the aspects of conventional transistors which limit the ability to continue downscaling. To attack this problem, devices like the originally proposed resonant tunneling transistor (Mead, 1960\textsuperscript{15}) are needed; i.e., ones that directly make contact to ultrasmall regions defined by heterobarsriers, and thereby control tunneling current through these regions.

Let us ignore for a moment the difficulty of making independent contact to a quantum well and consider how the potential of the quantum well is to be modulated. Potential is modulated of course by injecting or removing charge. The problem with the typical RTD structure is that electrons are not confined in the quantum well; they are free to tunnel out of the well. Any attempts to independently control the potential of this well are seriously compromised by this leakage path.

It is possible to create a resonant tunneling structure where there are confined charge carriers in the quantum well. This requires a structure with three different band gap materials as shown in Fig. 3. The wide band gap barriers surround the narrow band gap well (or base). Mid band gap contacts (emitter and collector) must be used. When prop-

![Fig. 2. Schematic representation of a resonant tunneling diode. Point A is below resonance, point B at resonance, and point C is past resonance.](image)

![Fig. 3. Band diagram of a resonant tunneling transistor where the potential of the quantum well may be controlled via carriers in the ground state of the valence band or conduction band in the well.](image)

erly designed the ground state in the quantum well is below the conduction band edge of the contacts, i.e., a "hidden" state. Electrons in this ground state are confined to the well since they may not tunnel out of the well and may be used to control the potential of the well. The tunneling current can pass through an excited state.

Note that the ground states of holes in the valence band of the well are similarly hidden. Therefore an electron tunneling current may be controlled by modulating of the well via holes in the valence band well ground state. This is the basis of a device called the bipolar quantum resonant tunneling transistor (BiQuaRTT). The BiQuaRTT structure includes a heavily doped $P$-type quantum well base. It is possible to bring excited quantum well states into resonance with the conduction band while keeping the ground hole state in the quantum well hidden. The output characteristics shown in Fig. 4(a) look very much like a typical bipolar transistor. Room temperature operation is observed with typical current gains of 50 and as high as 450. It should be stressed that the familiar peak and valley of a RTD does not appear in the collector output because the potential of the well is controlled independently and is not strongly affected by changing collector bias.

The signature of a resonant tunneling transistor is a transconductance which changes sign with applied bias to the base. As voltage increases the saturated collector current will rise to a peak, fall as one of the tunneling states goes out of resonance, and rise again as another state comes into resonance. This is easier to see if we plot $I_c$ at a fixed value of $V_{ce}$ as a function of $V_{be}$. This representation in Fig. 4(b) reveals the familiar peak and valley (we refer to this as a serpentine transconductance), which is the three-terminal analog of a two-terminal negative differential resistance. Both a superlattice version, which simulates a mid band gap material with a GaAs–AlGaAs superlattice, and a pseudomorphic version, which uses GaAs contacts/AlGaAs barriers/InGaAs narrow band gap quantum well, have been realized. However, because the BiQuaRTT does contain $PN$ junctions, it is not an ultimately scalable device.

A unipolar version of a RTT controls the well potential by electrons in the hidden ground state of the well and is what we refer to as the quantum excited state transistor or QuESTT. This device structure has been independently conceived of by Schulman et al. at Hughes and Haddad at University of Michigan. Without the benefit of $PN$ junctions for isolation it is more difficult to achieve contact to the quantum well while maintaining isolation from the emitter and collector regions. There are several approaches to building a QuESTT which would result in a unipolar resonant tunneling transistor but still might rely on electrostatic depletion for isolation. Figure 5 is a schematic of one approach to fabricating a QuESTT where base to emitter and base to collector isolation is accomplished with heterojunctions.

This embodiment of a QuESTT should be scalable down to very small sizes (currently only the epitaxy is ultra small). When lateral sizes approach deBroglie wavelengths, there will be additional quantum effects. To understand how those quantum effects will affect device performance, let us consider laterally quantized RTDs.

V. LATERAL CONFINEMENT—QUANTUM DOTS

Quantum wells produce confinement in one dimension only. These states in quantum wells are actually parabolic.
energy bands. When both lateral dimensions are reduced to dimensions on the order of electron wavelengths these bands are split into discrete electron states. These are defined as "quantum dots."

Quantum dot diodes have been fabricated by electron-beam lithography and reactive ion etching. Contact to a single diode is made by planarizing with polyimide and etching back to uncover the top contact. The lateral confining potential is not a heterojunction but the surface depletion from the Fermi level pinning at the GaAs surface.

Tunneling through these structures shows evidence of discrete states. Recent work in modeling these structures has improved our understanding of this spectra. It is important to note that this structure connects the quantum dot with a quantum wire which has its own set of subbands. The spectra of these structures is well described by tunneling from the subbands of the quantum wire emitter through the states confined in the quantum dot. The structures are an excellent nanolaboratory to investigate issues such as momentum nonconservation and the role of high angular momentum states.

Double dot diodes have also been fabricated where tunneling from discrete state to discrete state has allowed the discrete spectra to be observable up to higher temperatures than the single dot structures. Though their spectra is not well developed, it shows clear evidence of the overflap of single electron states in a nanofabricated electronic system.

Can a quantum dot transistor be built? A QuESTT scaled to its limits, would be such a device. If this is accomplished, will the ultimate barrier have been reached? In order to extend the increases in computational power well into the next century, it may be necessary to look beyond the transistor.

This is a difficult point of view to espouse when the entire electronic industry has been built around this basic switching device. Clearly circuits constructed from two terminal devices are not viable. Instead, it is suggested that multiple input devices be used in novel ways to build integrated circuits. In other words a revolutionary approach both in devices and circuit architecture appears necessary.

In order to understand the reasons why such a fundamental change may be in order, it is necessary to consider the limitations imposed by conventional von Neumann circuit architecture.

VI. NEW CIRCUIT ARCHITECTURES

The architecture employed by VLSI designers requires devices which are spatially removed by relatively large distances to be able to communicate with one another. With downscaling, devices are required to drive long interconnects which are not shrinking along with the devices. As current drive capabilities reduce with the device size, we will enter a regime where the principle limitation on circuit speed will not be the switching speed of the device but the charge-up time of the device load. Superconducting interconnects would eliminate one component of impedance but would not effect the input and output resistances of the devices and will not reduce the capacitance or inductance of long interconnects.

Another significant limitation of our present circuit architecture is relatively inefficient serial circuit operation. In other words only one or a small number of operations can be carried out in any one instant while most of the chip's devices are idle. Today's system designers are already attempting to circumvent this problem by employing multiple processors in their designs. It will be necessary to incorporate far more parallelism in the operation of next generation ICs.

The two architecture limitations we mentioned were long interconnects and serial processing. The solution would ideally be an architecture which had short or local interconnections and parallel circuit operation. One architecture which satisfies both criteria is a broad class of architectures known as cellular automata (CA).

LIFE is the best known example of a cellular automata. LIFE consists of a two-dimensional homogeneous array (all devices interact according to the same logical function) which is binary and synchronous. The interaction rules are: a device communicates with its 8 nearest neighbors, an off device will turn on if exactly three of its neighbors are on and an on device will stay on only if two or three of its neighbors are on. A rich variety of dynamic interaction is available from these simple interaction rules. However, it is not LIFE that we propose for doing computation.

The example of the type of CA architecture we are considering consists of a two dimensional array of devices. Each device is connected to only its four nearest neighbors. Information is input and retrieved from the periphery of the array. CA circuitry computes by "taking to" its four neighbors where each device functions according to its own "interaction rules." There are a very large number of possible interaction rules which include synchronous as well as asynchronous operation and may include multivalued logic.

We are also investigating CA arrays which are heterogeneous, in other words different devices in the array have different interaction rules. Function is designed or programmed by arranging devices with different interaction rules in the array to accomplish the desired computation. We have identified several different sets of synchronous interaction rules, any of which, would allow us to realize arbitrary boolean functions.

There is another line of reasoning which comprehends an interrelationship between devices and circuit architecture. Current architectures demand a high degree of isolation between neighboring devices. This sort of isolation is difficult to maintain as downscaling places devices ever closer together. This requirement is in contrast to CA where proper device operation may be based on interaction with neighboring devices.

CA cannot, in general, compete with current VLSI architectures for efficiency on a state/function basis. VLSI allows higher connectivity than a CA, which implies a more direct (in space and time) mapping an arbitrary function into the architecture. In cases where parallelism can speed up the execution of a function, CA will be able to compensate for lower interconnect dimensionality by distributing the computation over a smaller subset of interconnect dimensions. There are broad classes of these applications that quantum CAs may find early niches for functional acceleration. However, the ability of quantum devices to be scaled far beyond
VLSI suggests that another measure of efficiency, functional density, will tip the scales in favor of quantum CA for an even broader class of functions.

In order to present one concrete example of a set of interaction rules which can lead to useful computation, we will discuss a simple device which we call a Shannon cell. This device has three inputs and two outputs. The control input \( C \) selects whether the value at the \( X \) input or the \( Y \) input will be transferred to the two outputs as depicted in Fig. 6.

A homogeneous array of Shannon cells forms a functionally complete system where any combinatorial boolean function may be performed. Note that in this realization the data inputs are taken from not only the edges of the array but also from the \( C \) electrodes which are also used to program function. There are variations which allow data to be entered strictly from the periphery.

To see how we might be able to realize such a device with quantum effects, consider that the input \( C \) might control the potential of a quantum dot \( C. \) The \( C \) dot could then be used to select between tunneling from the \( X \) dot or the \( Y \) dot to the output. Remember that we have demonstrated the process of controlling tunneling current via control of the potential of a quantum state with the BiQuaRTT.

Note that there is more to the problem. We must level shift the energy of the outputs back up to input energy of \( X \) and \( Y \) for the next stage.

There is an impressive number of possible interaction rules for cellular automata. The Shannon cell example is a relative simple one where a quantum implementation seems possible. As we learn more about quantum transport and operation of resonant tunneling structures we will hopefully find a match between quantum mechanical coupling of quantum dots and a set of interaction rules which will allow us to build useful quantum effect integrated circuits.

**VII. SUMMARY**

Advances in microfabrication have permitted the investigation of quantum transport and the construction of quantum effect devices which open the possibility of building integrated circuits which will operate at a size scale below the predicted limits of conventional semiconductor devices. We believe that such devices must use heterojunction technology. A bipolar quantum resonant tunneling transistor has been demonstrated and unipolar versions have been proposed. Three dimensionally confined quantum dot structures have been fabricated and improvements in understanding these structures have been made. In order to realize quantum coupled ICs the interconnect problem must be overcome, which makes it necessary to revolutionize not only device operation but device type and circuit architecture as well. We have identified cellular automata as a promising candidate for a new architecture where there may be a quantum connection.

There are many challenges which must be met before quantum effect devices can become a viable IC technology. Quantum coupling between ultra-small structures must be harnessed in a manner which leads to useful and efficient computation. Successful application of CA circuitry would also require a revolution in software. Many other constraints must be met such as attention to the power dissipation of high density quantum devices, fault tolerance, and compatible interfaces with conventional ICs. There are also significant challenges to the community of microfabricators. Processing techniques with subteny micrometer resolution and unprecedented dimension control will certainly be required.

**ACKNOWLEDGMENTS**

This work has been sponsored in part by AFWAL, ARO, DARPA, and ONR. The authors wish to thank Bob Bate, Bill Frelsman, Chia-Hung Yang, Alan Seabaugh, and Jim Luscombe for useful discussions.

---


*R. T. Bate, Sci. Am. 258, 96 (1988).*


*M. Butiker, Phys. Rev. Lett. 57, 1761 (1986).*


An excellent example and access to the literature can be found in F. Capasso, S. Sen, A. Y. Cho, and D. L. Sivco, Appl. Phys. Lett. 53, 1056 (1988).


