NUMBER 14 FALL 1986

Los Alamos Science

LOS ALAMOS NATIONAL LABORATORY

LOS ALAMOS NATIONAL LABORATORY
3 9338 00350 7109

This issue is dedicated
to the memory of
John Wheatley

Inside This Issue

It is appropriate that this issue, devoted to frontier topics in science, technology, and technology transfer, be dedicated to the memory of John Wheatley, one of the great scientists and technologists of the twentieth century. His untimely death at age fifty-nine occurred before he and his Los Alamos collaborators, Greg Swift and Al Migliori, had completed the article on natural heat engines that appears in this issue. Natural engines are a departure from traditional heat engines in that their design capitalizes on irreversible processes instead of trying to minimize them. These are engines with no moving parts whose simplicity and reliability make them ideal for applications in space. John loved them because they could be understood on the basis of classical thermodynamics and their design took full advantage of nature's way of doing things. He lived to see the successful operation of the liquid sodium acoustic engine and would have been enormously pleased that the article on his pet project includes some good solid data on this particular natural engine.

Our experience with John bore out his reputation for meticulous care, sound pedagogy, and stubbornness in doing things the "right" way-which meant his way. Although enthusiastic about the initial draft written by our staff writer Roger Eckhardt (with help from Greg and Al), John, as predicted, chose to spend an entire week rewriting the manuscript so that it would contain the "right" approach to understanding the thermodynamic principles underlying the operation of natural engines. We are honored to be publishing this superb and easily accessible article.

John Wheatley was an important presence wherever he went, and his passing was a shock to all who knew him. To provide our readers with a glimpse of John and his numerous contributions to low-temperature physics, we organized a round table, shortly after his death, among those of his close collaborators who were then at Los Alamos. The title of the discussion, "Pushing The Limits," sums up John's approach to everything he did at work and at play. We hope the portrait that emerges will convey at least in part the impact of this inspired and inspiring man.

One of John's great contributions to technology was to push the low-temperature frontier down to a few thousandths of a degree above absolute zero. Another important technological frontier is defined by the shortest time scale on which events can be measured and controlled. An article by Bob Hammond, entitled "Photoconductivity and Picosecond Signals," describes the breakthrough being made with photoconductors in generating and measuring electronic signals on picosecond time scales and in tracking time-varying radiation with picosecond accuracy. The picosecond rise and fall times of photoconductivity were not known until photoconductors could be stimulated by ultrashort laser pulses. The rapid response of photoconductors is being exploited by the Electronics Division at Los Alamos in ultrafast power switches for generating accurately timed, short-rise-time, high-power electrical pulses and in ultrafast radiation detectors that can follow the rapidly varying signals from laser fusion experiments and nuclear weapons tests. Perhaps the most exciting new photoconductive device is one that can measure the performance of tiny segments of integrated circuits. Such a device enabled Bob Hammond and his colleagues to observe for the first time the picosecond phenomenon of carrier-velocity overshoot that had been predicted by theory. This phenomenon may provide a basis for even faster integrated circuits.

Frontiers defined by cultures provided almost as much of an adventure to John Wheatley as those defined by technology. He took his low-temperature work to the far reaches of Argentina and to the scientific community of Finland and made his technological innovations available worldwide through his founding of the SHE Corporation. In this issue we report on another type of adventure in technology transfer, this time in the Caribbean Basin and with economic rather than scientific goals in mind. Los Alamos is a center of research in geothermal energy and in energy resources planning and development. The Caribbean Basin project is designed to transfer that expertise to countries in Central America and the Caribbean Islands whose economic difficulties might be alleviated by the development of indigenous energy resources. Bob Hanold and Verne Loose, leaders of this unusual project, along with about thirty-five geologists, geochemists, geophysicists, and economists, have been traveling back and forth between the Laboratory and the Caribbean Basin. They are working with local professionals to evaluate economic and energy needs and to explore and develop the most promising natural resources. Accompanying Bob's and Verne's overview of this precedent-setting project are a series of short pieces that describe the technical work being accomplished by Los Alamos scientists in the jungles and cities of Central America.

This issue closes with a reminiscence by Herb Anderson in honor of Nick Metropolis on his sixtieth birthday. It concerns the early days of one of the most important frontiers of our time. the frontier of electonic computing. Herb's wonderful stories of Nick's innovative work on the MANIAC with his many illustrious collaborators is somehow a fitting end to an issue dedicated to the memory of John Wheatley. ■

Necia Grant Cooper

Los Alamos Science

NUMBER 14 FALL 1986

Editor
Necia Grant Cooper

Art Director
Gloria Sharp

Science Writers
Roger Eckhardt
Nancy Shera

Assistant Editor
Dixie McDonald

Contributing Editor
Hollie Sowerwine

Design and Illustration
Katherine Norskog

Illustration
Jim Cruz, Jeff Segler
Ward Zaelke

Photography
John Flower

Photo Laboratory Work
Ernesto Archuleta, Gary Desharnais
Debbie Fisher, Chris Lindberg
Ken Lujan, Bryan O'Hare
Mike O'Keefe, Cordy R. Roybal
Roger Volz, Gary Warren

Phototypesetting
Chris West
Dolly Lucero, Joyce Martinez
Kathy Valdez,

Circulation
Dixie McDonald

Printing
Jim E. Lovato
Guadalupe Archuleta

Los Alamos Science is published by Los Alamos National Laboratory, an Equal Opportunity Employer operated by the University of California for the United States Department of Energy under contract W-7405-ENG-36.

Address Mail to
Los Alamos Science
Mail Stop M708
Los Alamos National Laboratory
Los Alamos, New Mexico 87545

Cover photo by John Flower. **Stainless steel plates used in a liquid-to-liquid heat exchanger for the liquid propylene Stirling engine, one of several novel heat engines studied at Los Alamos.**

THE NATURAL HEAT ENGINE

by John C. Wheatley, Gregory W. Swift, and Albert Migliori

Heat engines are a compromise between the crisp ideals discussed in thermodynamic textbooks and the clanking, hissing realities of irreversible processes. This compromise produces wonderful machines, such as the automobile engine and the household refrigerator. In designing real devices, the goal is not to approach thermodynamic ideals by reducing irreversibilities but to balance cost, efficiency, size, power, reliability, simplicity, and other factors important to the needs of particular applications.

Simplicity is the most striking feature of a *natural* engine, a reciprocating heat engine with no moving parts. As we will see, the basic operating cycle of the natural engine is so straightforward it can be applied to a wide variety of systems with working media that range from air to paramagnetic disks.

Although the natural engine is new in concept, the underlying thermodynamic principles and processes are shared with conventional engines, such as the Stirling and Rankine engines. To set the stage for natural engines, we will first discuss a few conventional idealized thermodynamic cycles and the practical engines they suggest.

Conventional Heat Engines and Cycles

In principle, any idealized thermodynamic heat engine cycle is *functionally* reversible in the sense that it can be made to operate in either of two modes: prime mover or heat pump* (Fig. 1). In a prime mover, heat flows from high to low temperatures, and the engine converts a portion of that heat to work. In a heat pump, the flows of heat and work are reversed; that is, work done on the engine causes it to pump heat from low to high temperatures. Few practical engines are functionally reversible. The internal combustion engine is a prime mover only; the household refrigerator is a heat pump only; neither engine is ever operated in both modes.

Figure 1 shows how the first and second laws of thermodynamics place an upper limit on the *efficiency of a prime mover* (the fraction of the heat input converted to work). The efficiency of a *thermodynamically* reversible cycle—that is, one in which all parts of the system are always in thermodynamic equilibrium—is equal to that upper limit. (One statement of the second law of thermodynamics is that all reversible engines operating between the same two temperatures have the same efficiency.) Figure 1 also shows the upper limit for the *coefficient of performance (C.O.P.) of a heat pump* (the amount of heat rejected at the higher temperature per unit of work). Both theoretical limits depend only on the temperatures involved.

Carnot. The most fundamental engine cycle operating between two temperatures is the functionally and thermodynamically reversible cycle propounded by Sadi Carnot in 1824. The cycle consists of alternating adiabatic and isothermal steps (Fig. 2). During an adiabatic step, no heat flow occurs ($Q = 0$) and entropy ($\int dQ/T$) remains constant. Thus any flow of work causes a corresponding change in the temperature of the working medium. During an isothermal step, the temperature remains constant, and flows of entropy, work, and heat occur.

In the Carnot cycle, the entropy change of one isothermal step exactly balances the entropy change of the other isothermal step. Over a complete cycle, no entropy is generated. If an engine could be made to follow a Carnot cycle, its efficiency would equal the theoretical upper limit given in Fig. 1. Although the upper limit applies to any reversible engine, this efficiency is usually called the Carnot efficiency.

Building an engine that approximates a Carnot cycle requires that all processes in its cycle are carried out very near equilibrium. If not, the resulting irreversibilities due to temperature and pressure gradients generate entropy and cause a loss of efficiency. For example, the

A prime mover is often called an engine and a heat pump a refrigerator. Here we use the term engine to denote both *thermodynamic functions, and our use of the term heat pump includes the refrigerator. Strictly speaking, however, the purpose of a heat pump is to* reject *heat at the* higher *temperature, whereas the purpose of a refrigerator is to* extract *heat at the* lower *temperature.*

The release of acoustic energy by a simple natural heat engine, the Hofler tube, made evident by the white plume at the upper end. The device consists of a two-piece copper tube, closed at the bottom, and a short set of fiber glass plates that run parallel to the tube's axis in the region of the flanges. The acoustic energy results spontaneously when a temperature gradient is applied across the plates. In this case, the gradient was produced by holding one end of the tube while immersing the other end (frosted) in liquid nitrogen.

temperature differences across the heat exchangers that move heat in or out of the engine are frequently a source of irreversibility that greatly cuts efficiency. (See "The Fridge" for a quantitative accounting of this and other losses in a practical heat pump.)

Although one may approach near-equilibrium conditions by designing the engine so as to reduce these gradients, the end result is a very slow cycling of the engine and a very low power output. An important point (originally made by F. L. Curzon and B. Ahlborn and generalized by S. Berry, J. Ross, and their collaborators) is that Carnot-like cycles operating between two temperatures with imperfect heat exchangers have quite different efficiencies depending on whether work per cycle or power is being maximized. Real engines, especially high-speed reciprocating engines, cannot approximate Carnot's cycle closely.

Stirling. The Stirling *engine*, invented in 1816 by the Reverend Robert Stirling some eighteen years before Carnot's ideas were published and originally called the hot-air engine, is a reciprocating engine that is functionally reversible and, in principle, thermodynamically reversible. The ideal Stirling *cycle* has the Carnot efficiency. From a practical standpoint, implementing the Stirling cycle suffers from some of the problems of implementing the Carnot cycle. However, the introduction of a second thermodynamic medium provided the means by which high-speed Stirling engines of good efficiency could be built.

The Stirling *cycle* (the solid black curve in Fig. 3) differs from the Carnot cycle in that the adiabatic steps are replaced with steps that are reversible by virtue of being *locally* isothermal. This type of cycle is achieved by using *two* thermodynamic media. The first is the working fluid, which typically can be either a gas or a liquid. (There are Stirling cycles that use solids, but we do not discuss them here.)
continued on page 6

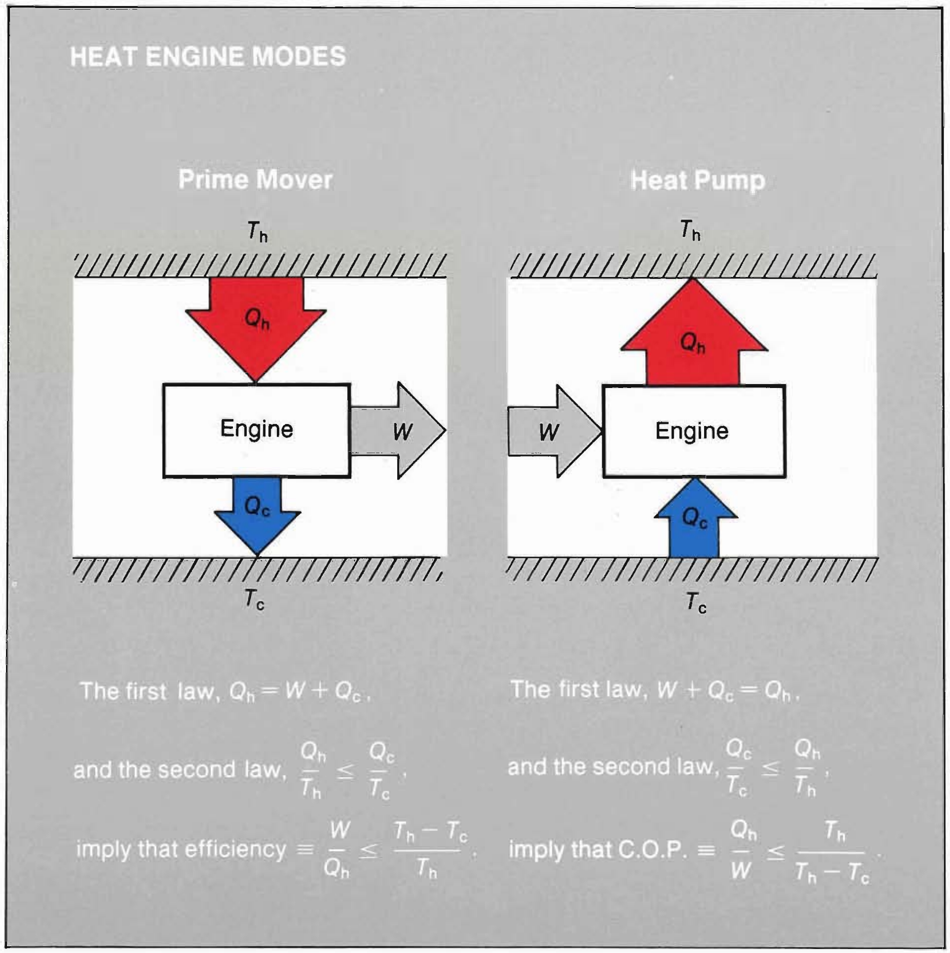
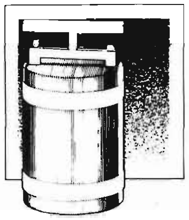


Fig. 1. (a) A heat engine operating as prime mover converts some of the heat that is flowing from a hot temperature T_h to a cold temperature T_c into work. The first law of thermodynamics tells us that Q_h, the heat that passes into the engine at the hot temperature, equals Q_c, the heat put back into the environment at the cold temperature, plus W, the work done by the engine. The second law tells us that the entropy per cycle generated by the system must be positive or, at best, zero. Since the engine is assumed to be in a steady state, the entropy change in the environment due to the heat flow out of the engine, Q_c/T_c, is greater than or equal to that due to the heat flow into the engine, Q_h/T_h. Together, these two laws give an upper limit for W/Q_h, the efficiency of the engine. Note that a prime mover can only approach its highest efficiency of unity when $T_c \ll T_h$. (b) In a heat engine operating as heat pump, all flows of heat and work are reversed. Thus work done on the engine causes it to draw heat out of the environment at the cold temperature and place it into the environment at the hot temperature. Consideration here of the first and second laws leads to an upper limit on the coefficient of performance (C.O.P.), Q_h/W, which is the reciprocal of the efficiency of a prime mover. (For a *refrigerator*, the C.O.P. is better defined as the ratio of the heat extracted at the *lower* temperature to the work done on the machine, that is, Q_c/W.)

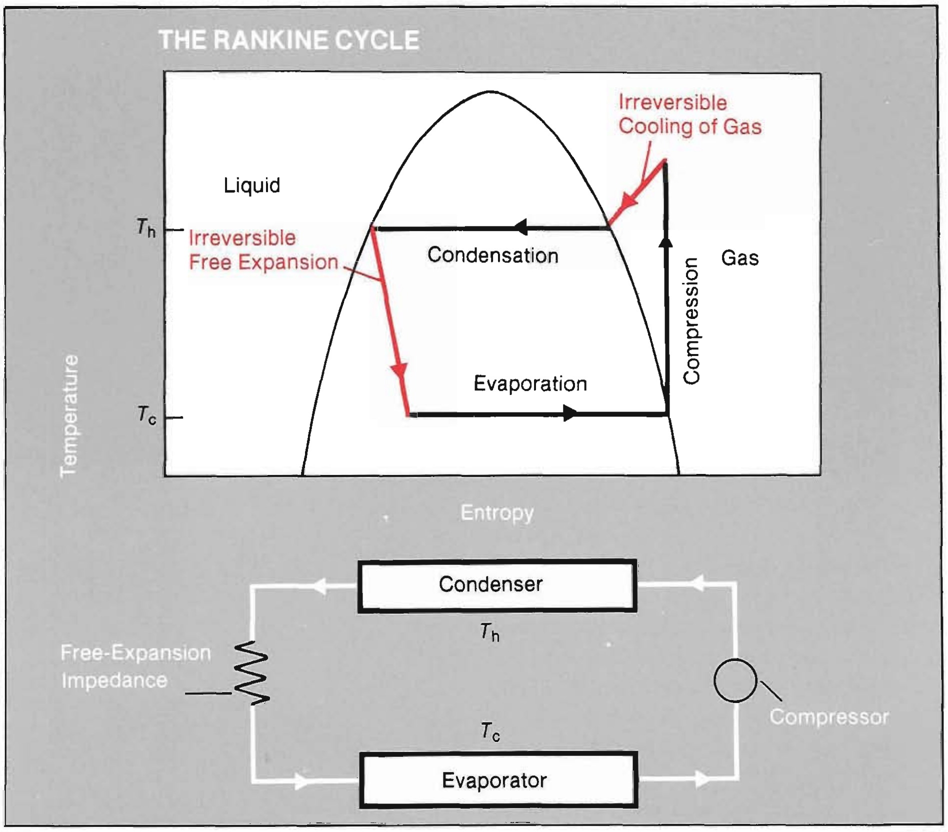


The Fridge

The basis for the household refrigerator is the Rankine cycle, which, as shown in the figure, duplicates a portion of the Carnot cycle in that it has one adiabatic step and two isothermal steps. A key feature of this cycle is a phase change in the working fluid, and the two isothermal steps correspond to condensation of the fluid at T_h and evaporation at T_c. Also, the engine operates with continuous flow rather than by reciprocating: the working fluid cycles through its various thermodynamic states by being forced around a closed loop.

This cycle has intrinsic irreversibilities associated with the free expansion of the liquid and the cooling of the gas to the temperature at which condensation occurs. Thus one expects the Rankine cycle to have less than ideal Carnot efficiency—even *before* accounting for such losses as those due to temperature differences at the heat exchangers. Nevertheless, Rankine engines remain the design of choice in many applications because they are simple and powerful. Many refrigerators will run thirty years with little or no maintenance, and overall cost is low.

The Rankine cycle can also be used in an air-to-air heat pump. Table 1 illustrates the effects of various irreversibilities on the coefficient of performance for such a pump—one designed to keep a house at 20°C when outside air is 5°C so that, ideally, $T_h - T_c$ is 15°C and the Carnot coefficient of performance is 19.5. The largest drop in the the estimated coefficient of performance occurs when ideal heat exchangers are replaced by practical heat exchangers—ones both small enough to get through the door of a house and cheap enough to cost less than the house. A small, cheap heat exchanger can only transfer large amounts of heat if a large temperature difference occurs across it. The net effect in our example is that the temperature difference, $T_h - T_c$, of the working fluid increases from 15°C to 50°C, causing the coefficient of performance for the Carnot cycle to drop from 19.5 to 6.4.



The Rankine cycle, used in the household refrigerator, is based on a liquid-gas phase change. The cycle is shown here superimposed on the phase diagram for the working fluid; a schematic of the heat pump is also shown. The Rankine cycle resembles the Carnot cycle in that there are two isothermal steps and, on the compression side, an adiabatic step. The two parts of the cycle (shown in red) that differ from the Carnot cycle—the cooling of the gas at constant pressure to the condensation temperature T_h and the free expansion of the liquid—are intrinsically irreversible. ▲

Table 1

Losses in the coefficient of performance (C.O.P.) due to irreversibilities for an air-to-air heat pump (adapted from *Heat Pumps* by R. D. Heap, 1983).

Cycle	Irreversibilities	T_c(°C)	T_h(°C)	C.O.P.
Carnot	none	5	20	19.5
Carnot	real heat exchangers	−5	45	6.4
Rankine	real heat exchangers, intrinsic irreversibilities	−5	45	5.1
Rankine	real heat exchangers, intrinsic irreversibilities, compressor losses	−5	45	4.0
Rankine	real heat exchangers, intrinsic irreversibilities, compressor losses, miscellaneous	−5	45	3.0

The C.O.P. drops to 5.1 when one takes into account the intrinsic irreversibilities of the Rankine cycle. Further decreases occur because of losses in the compressor (due to friction and the imperfect conversion of electrical power to shaft power) and miscellaneous losses (such as power to run the fans, the thermostat, and the controls). The final C.O.P. for a practical, operating Rankine heat pump is 3.0, more than a factor of 6 lower than the C.O.P. for an ideal engine. ■

continued from page 4

The working fluid is displaced at constant volume through a *regenerator* containing the second medium, which is typically a solid. The second medium can be metal plates or just the walls of the vessel, but its heat capacity should be large compared to that of the working fluid. A small temperature gradient exists along the length of the regenerator, the total temperature change being the temperature difference between the hot and cold heat exchangers at the ends of the regenerator. If we ensure good thermal contact between the two thermodynamic media (say by making the distance between any fluid element and its adjacent regenerator plate small enough), the fluid can temporarily store heat in the regenerator and recover it later under nearly reversible isothermal conditions. Of course, the steeper the gradient along the regenerator or the faster the displacement of the working fluid through the heat exchangers and the regenerator, the greater the irreversible losses.

During one part of the cycle, fluid enters the cold end of the regenerator, picks up heat from the second medium, and exits hot. During another part of the cycle, fluid enters the hot end of the regenerator, deposits heat in the second medium, and exits cold. The net heat stored in the second medium over a complete cycle is zero (provided, as is the case for an ideal gas, the specific heat of the fluid does not depend on pressure). The regenerator, therefore, enables us to change the temperature of the working fluid from the temperature of the hot reservoir T_h to the temperature of the cold reservoir T_c and back again without the adiabatic expansions and compressions of the Carnot cycle. In other words, *locally* isothermal reversible steps have replaced the adiabatic reversible steps for changing the temperature of the working fluid. As a result, the efficiency of the Stirling cycle is the same as that of the Carnot cycle.

But what about the Stirling *engine*? Typically, Stirling engines do *not* follow a Stirling cycle but rather follow an adiabatic pseudo-Stirling cycle (the dashed curves in Fig. 3). This confusing nomenclature is illustrative of the compromises made between the concept of a thermodynamic cycle and the construction of an operating engine. Unfortunately, because the same person's name can become attached to both the cycle and the engine, confusion abounds.

What changes the Stirling cycle to a pseudo-Stirling cycle is related to the temperature of the working fluid at the heat exchangers. An adiabatic compression warms the fluid prior to its displacement through the hot heat exchanger and into the regenerator, and, at the other end of the cycle, an adiabatic expansion cools the liquid prior to its displacement in the opposite direction. These adiabats partially replace the isotherms of the original cycle, necessitating extension of the constant-volume displacement steps.

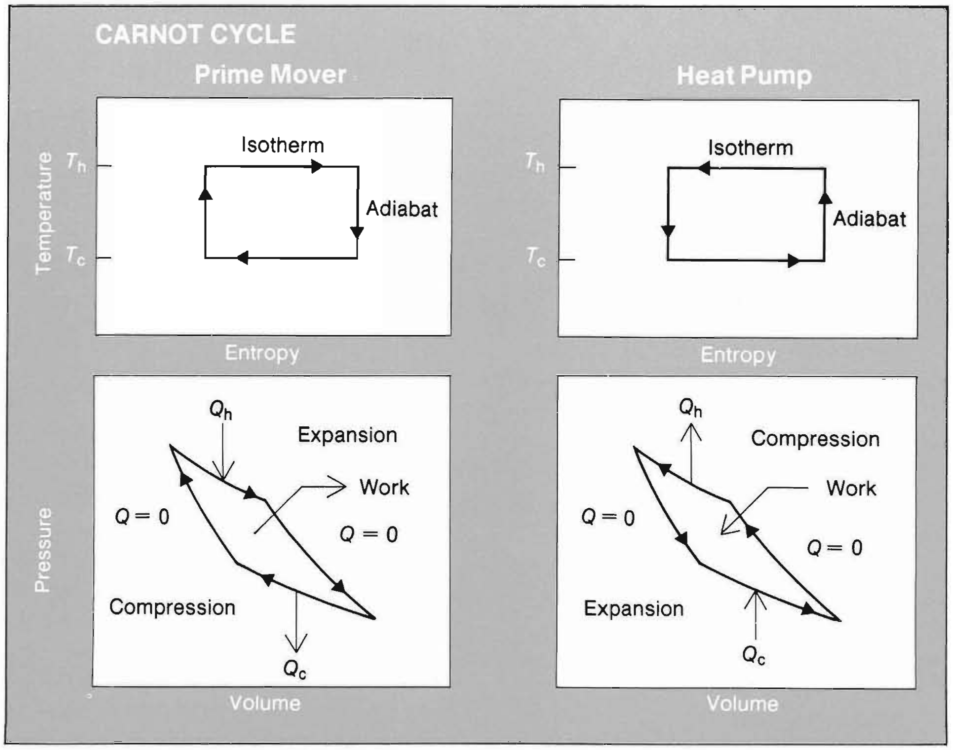


Fig. 2. Temperature-entropy and pressure-volume diagrams for the prime-mover and heat-pump modes of a Carnot cycle. When the engine is operating as a prime mover, the first part of the expansion stroke is the addition of heat to the engine at T_h. Because this process is isothermal, the heat energy is used to expand the working medium and do work on the surroundings. In the second step, further expansion occurs adiabatically, that is, with *no* addition of heat or change in entropy. Because work continues to be done by the fluid, the temperature of the medium must drop. The third step is isothermal compression in which heat is rejected from the engine to the lower temperature T_c and the entropy drops. Finally, an adiabatic compression raises the temperature of the medium. The Carnot cycle for a heat pump is just the reverse of that for a prime mover. The area enclosed by the pressure-volume diagrams equals the net work done by or on the engine in a full cycle.

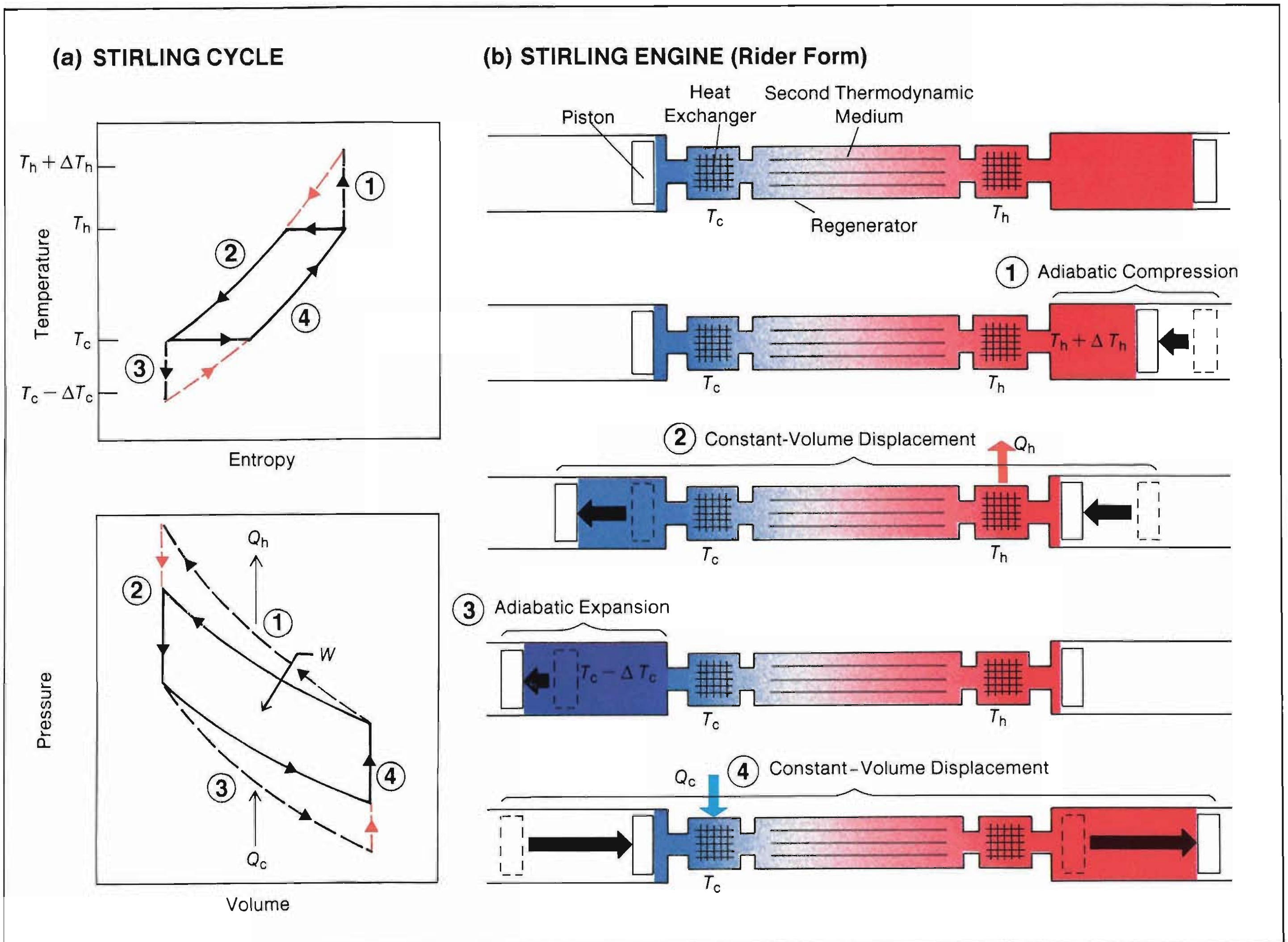


Fig. 3. (a) The ideal Stirling heat-pump cycle (black) consists of isotherms and constant-volume steps. The adiabatic pseudo-Stirling cycle replaces the isotherms with adiabats and extensions of the constant-volume displacement steps (dashed curves). It is the pseudo-Stirling cycle that frequently serves as the basis for practical Stirling engines. (b) One variation (the Rider form) of a Stirling engine following the adiabatic pseudo-Stirling cycle. All such engines are based on the ideal of local isothermal steps made possible through the use of a second thermodynamic medium in the regenerator. The first step of the cycle depicted here is adiabatic compression in the cylinder on the right that raises the temperature of the fluid above T_h. In the second step, both pistons move, displacing the fluid to the left. The heat Q_h generated by the compression is rejected in the heat exchanger on the right. Because of the small longitudinal temperature gradient and good lateral thermal contact along the regenerator, heat is transferred between the two media under essentially isothermal conditions, cooling the fluid from T_h to T_c. In the third step, adiabatic expansion cools the fluid in the left cylinder. Constant-volume displacement of the fluid to the right then causes heat Q_c to be drawn in at the left heat exchanger and the heat stored in the second medium during step 2 to be returned to the fluid. Irreversibility occurs at the beginning of both constant-volume displacements (dashed red in part (a)) when the fluid at one temperature contacts the heat exchanger at a different temperature.

Since the above alterations introduce intrinsic irreversibilities, the maximum efficiency possible for the pseudo-Stirling cycle is lower than that for the true Stirling cycle. In particular, fluid that has been warmed by adiabatic compression (and thus raised to temperature $T_h + \Delta T_h$) is pushed into the hot heat exchanger during the displacement step, where it makes thermal contact irreversibly with the exchanger at temperature T_h. The same type of irreversiblity occurs in the other heat exchanger after the adiabatic expansion step. Such effects are departures from the ideal of locally isothermal conditions.

Although a Stirling engine is not as simple conceptually as a Carnot engine, practical Stirling engines that operate at moderately high frequencies can indeed be built. As before, other irreversible losses occur because there must be significant temperature differences to drive heat through the heat exchangers. Also, if the working fluid is a liquid (see "The Liquid Propylene Engine"), an additional type of irreversibility arises: the specific heat of a liquid is pressure-dependent, making the recovery of heat in the regenerator imperfect. This irreversibility is not an intrinsic feature of the cycle but is a material property that cannot be avoided. As such, it is of a more fundamental nature than the limitation, say, of the heat exchangers.

Phasing of the various moving parts in a heat engine is another factor necessary to its operation. Although the engine depicted in Fig. 3 is a heat pump, if the phasing of the two pistons is altered so that expansion occurs on the hot-temperature side when most of the fluid is hot and compression occurs on the low-temperature side when most of the fluid is cold, heat flow will be reversed and the engine will become a prime mover. As we shall see, both phasing and the second thermodynamic medium are of key importance in natural heat engines also, although there are significant differences in the way in which the second medium is used.

Internal Combustion. One way to circumvent the loss of efficiency from irreversibilities at the heat exchangers is to generate the heating or cooling effects *inside* the engine rather than outside. In 1893 Rudolf Diesel envisioned such an engine and, in fact, intended it to follow a Carnot cycle of adiabats and isotherms. His idea was to provide the heat for the isothermal expansion by burning coal dust that was injected into the engine at just the proper rate to maintain isothermal conditions. Cooling for the isothermal compression was to be provided by spraying water into the chamber. So far, no one, including Diesel, has been able to implement this cycle, and we are once again confronted with confusing nomenclature: the modern Diesel engine does not follow the Diesel cycle.

The idea of internal combustion, of course, survived, and modern Diesel engines work very well indeed. But internal combustion introduces new practical irreversiblities. For example, the addition and burning of the fuel in a typical piston engine causes differences between the pressure and temperature in the cylinder at the end of the cycle and at the beginning. A considerable irreversible loss occurs as heat and pressure are vented in the exhaust. Thus, internal combustion engine cycles differ from the Carnot and Stirling heat engine cycles described earlier in that the working medium is not returned to its original state.

Nevertheless, the use of phased and controlled internal combustion eliminates the problem of bringing heat in through a firewall. The diesel and gasoline internal combustion engines are used today because they are simple, both in principle and in practice, their power density is very high, and their efficiency is relatively good, sometimes very good. Practical diesel engines approach a level of efficiency in which the useful work is nearly half the heating value of the fuel.

Otto and Brayton. Two common heat engine cycles that will help illuminate the characteristics of a natural engine are the Otto cycle (black curve in Fig. 4) and the Brayton cycle (red curve). Each of these cycles, both of which are typically implemented irreversibly, has two adiabatic steps and two nonadiabatic steps. In the Otto cycle, the nonadiabatic steps are the addition and removal of heat at constant volume; in the Brayton cycle, these steps are carried out at constant pressure.

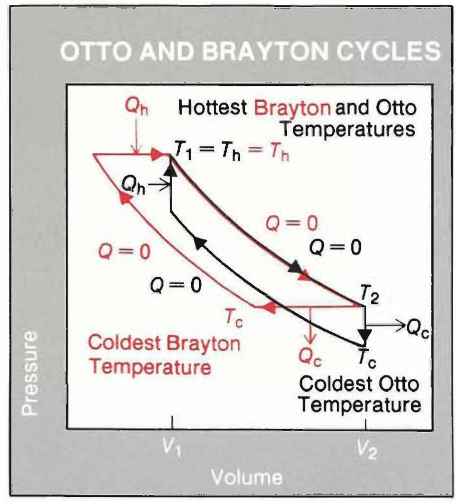


Fig. 4. The Otto (black) and Brayton (red) heat engine cycles, which consist of two adiabatic steps that alternate with two nonadiabtic steps—the latter steps being the addition or removal of heat at constant volume in the Otto cycle and at constant pressure in the Brayton cycle. Only the prime mover mode is shown. Note that for both cycles the highest temperature T_h equals the temperature T_1 at the upper extreme of the adiabatic expansion step but that the coldest temperature T_c is lower than the temperature T_2 at the lower extreme of the adiabatic expansion step.

If the working fluid is an ideal gas, both cycles have the same efficiency η given by

$$\eta = 1 - \left(\frac{V_1}{V_2} \right)^{\gamma - 1}, \qquad (1)$$

where γ is the ratio of the specific heat at constant pressure to that at constant volume. What is interesting about this formula is that efficiency for these cycles is determined by *geometry* (the ratio V_1/V_2 of the volumes at the extremes of the adiabatic expansion step) and by a *fluid parameter* γ but *not* by T_h and T_c, the temperatures of the hot and cold reservoirs.

Since for an ideal gas the quantity $TV^{\gamma-1}$ is constant along an adiabatic path, efficiency can also be expressed in terms of the temperatures, T_1 and T_2, at the extremes of the adiabatic expansion step:

$$\eta = 1 - \frac{T_2}{T_1}. \qquad (2)$$

The diagrams for the Otto and Brayton cycles show that in both cycles T_1 equals T_h but T_c is lower than T_2. This difference is due to further cooling, after the adiabatic expansion step, along a nonadiabatic step (removal of heat at constant volume in the Otto cycle and at constant pressure in the Brayton cycle). If we now examine the limiting case of zero heat transferred during the nonadiabatic steps, we see that T_2 approaches T_c and the efficiency approaches the Carnot efficiency. Of course, at the same time, the area enclosed by either cycle, and thus the work output, shrinks to zero.

We will find that all of these features of the Otto and Brayton cycles have counterparts in the natural engine.

The Natural Heat Engine

One guiding principle in the development of most heat engine cycles has been to minimize irreversibilities because they generate entropy and decrease efficiency. In the development of practical engines, however, irreversibilities are often deliberately introduced to increase power, decrease maintenance, or simplify design and manufacture, enabling one, for example, to build small engines, or high-speed reciprocating engines, or cheap engines.

On the other hand, irreversibilities play a more fundamental role in the natural heat engine. Rather than tolerating irreversibilities for the sake of expediency, the natural heat engine takes advantage of them. For example, heat conduction across a temperature gradient is central to the operation of a natural heat engine known as the acoustic heat engine. Without this irreversibility, the engine would not work. The result of such an approach is a significant leap in simplicity and, for certain applications, a leap in power and efficiency.

Thus, whereas engines that approximate, say, the Stirling cycle are intrinsically reversible (though possibly irreversible in practice), natural heat engines are intrinsically *ir*reversible—they cannot work if irreversibilites are eliminated. Nature abounds with useful irreversible processes, so, for the sake of a short, appropriate, and easily remembered name, we call intrinsically irreversible engines *natural* engines.

Acoustic Engines. Work in Los Alamos on natural engines began with an acoustic heat-pumping engine. Our work, however, was preceded by two conceptually related devices, which we will describe without, for the moment, explaining their somewhat surprising behavior.

W. E. Gifford and R. C. Longsworth invented what they called a pulse tube (Fig. 5a). Part of this closed tube was fitted with a set of Stirling-type regenerator plates intended to promote locally isothermal processes along their length, and part of the tube was left empty. Pulses were produced at the regenerator end of the tube by switching between high- and low-pressure gas reservoirs at a rapid rate (1 hertz). The extreme inner end of the regenerator plates got very cold, whereas a heat exchanger withdrew heat at the empty end of the tube. The pulse tube demonstrated the pumping of heat with acoustic energy in the presence of a second thermodynamic medium.

The other significant precursor to our work, and one of which we were initially unaware, was the half-wave resonator of P. Merkli and H. Thomann (Fig. 5b). In this apparatus, a piston drives pressure fluctuations in air at nearly half-wave resonance in a simple closed tube. Merkli and Thomann observed that the center of the tube *cooled*, whereas the ends of the tube *warmed*. At first, these results seem surprising. Naively, one might expect heating everywhere rather than cooling in one region. Further, the cooling occurred in the center, which, at a quarter of an acoustic wavelength, is coincident with a maximum, or antinode, in acoustic velocity and thus where one would surely expect a warming due to viscous scrubbing of the air on the walls.

The first acoustic heat pump built at Los Alamos used a speaker at one end of a closed tube to drive the acoustic resonance and has a stack of fiber glass plates positioned toward the opposite end (Fig. 5c). The plates constitute a second thermodynamic medium but not a Stirling-like regenerator because they are spaced so far apart that locally isothermal conditions do *not* prevail. With such an arrangement, it is easy to produce a 100-centigrade-degree temperature difference across a 10-centimeter-long stack of plates in only a minute or so.

Subsequently, Tom Hofler built a device (opening photograph and Fig. 6) to show his Ph.D. candidacy committee at the University of California, San Diego. The device, which we call the Hofler tube, consists of a quarter-wave acoustically resonant metal tube closed at one end and a stack of fiber glass plates that run parallel to the axis of the tube. Short copper strips glued at each end of each fiber glass plate provide heat exchange by making contact with two flanges encircling the tube.

If the closed end of the tube is heated, say by holding it in a warm hand, and its open end is cooled by dipping it in liquid nitrogen, the resulting steep temperature gradient causes the air in the tube to vibrate, and the person holding the tube will feel his or her whole arm begin to shake. When the tube is removed from the liquid nitrogen, the sound of the acoustic oscillations is very intense. Peak-to-peak pressure oscillations at the closed end have been found to be as high as 13 per cent of the atmospheric pressure! Thus, the tube operates as a prime mover, and heat is converted to acoustic work.

How do this and other acoustic engines work? The Hofler tube is the grandchild of the Sondhauss tube, famous in thermoacoustics and explained qualitatively by Lord Rayleigh over a hundred years ago. Theoretical understanding of these and related devices has been promoted by Nikolaus Rott in a series of papers published over the last fifteen years. The same conceptual foundation can be used to understand quantitatively not only the Hofler tube but the other acoustic devices mentioned above as well.

As mentioned before, an important factor in the operation of traditional engines is phasing: pistons and valves have to move with correct relative timing for the working medium to be transported through the desired thermodynamic cycle. The natural engine contains no obvious moving parts to perform these functions, yet the acoustic stimulation of heat flow and the generation of acoustic work point to some type of cycling, or timed phasing of thermodynamic processes.

The key to phasing in natural engines is the presence of two thermodynamic media. In the Hofler tube, gas was the first medium, the fiber glass plates were the second. Consider a parcel of gas that moves back and forth along the plates at the acoustic frequency. As it moves, the parcel of gas will experience changes in temperature. Part of the temperature changes come from adiabatic compression and expansion of the gas by the sound pressure and part as a consequence of the local temperature of the plate itself. The heat flow from gas to plate that occurs as a consequence of these temperature differences does not produce *instantaneous* changes. Rather a *thermal lag* in the heat flow between the two media creates the phasing between temperature and motion

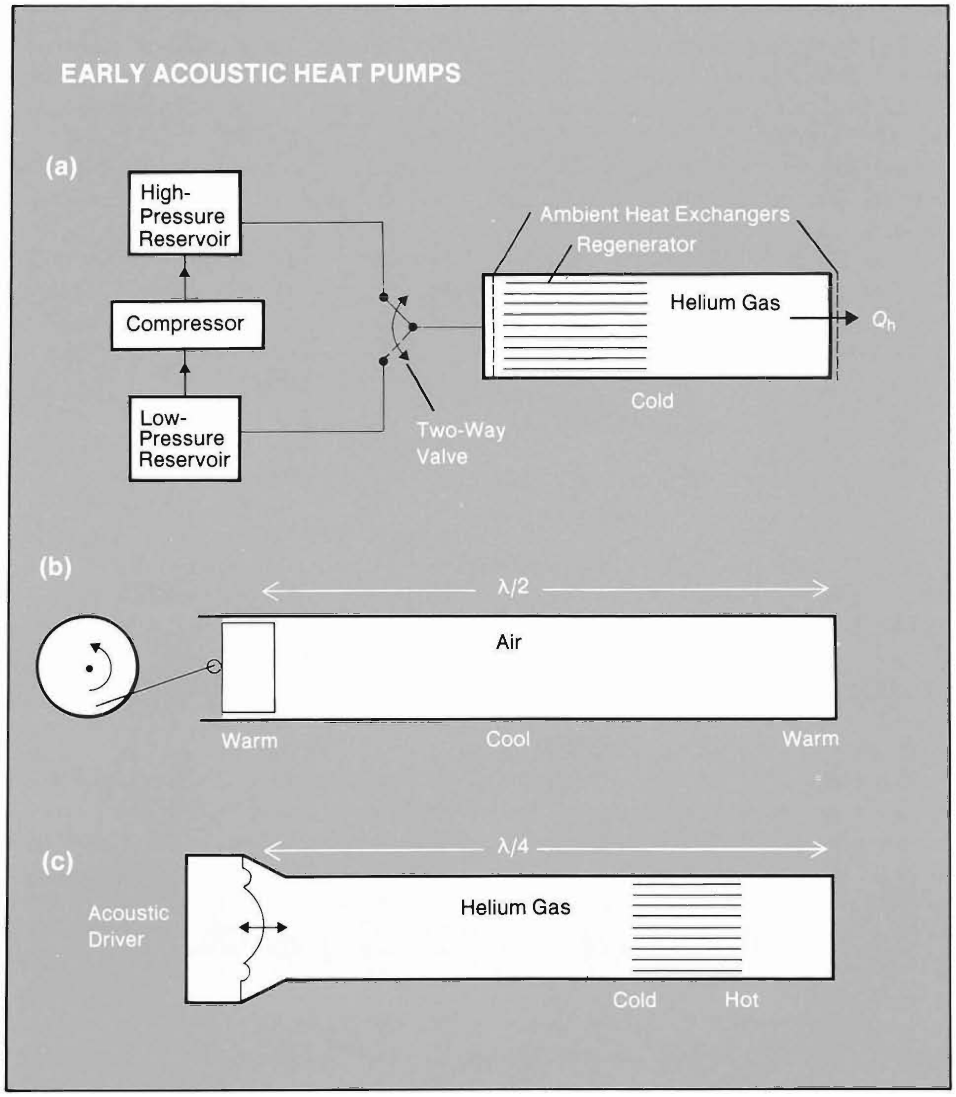


Fig. 5. (a) At the left end of the Gifford and Longsworth pulse tube, pressure pulses in a gas are generated at 1 hertz (Hz) by switching between high-pressure (5 bars) and low-pressure (1 bar) reservoirs. In conjunction with *two* second thermodynamic media (a Stirling-type regenerator and the walls of the open section of the tube), the pulses cause heat to be pumped from the middle of the tube to the far right. (b) The half-wave resonator heat pump of Merkli and Thomann is a simple closed tube whose acoustic resonance is driven on the left by a reciprocating piston. Contrary to one's intuition, the center of the tube, where the acoustic velocity is greatest, cools rather than warms. (c) The first acoustic heat pump built at Los Alamos contains a stack of fiber glass plates and helium gas as the working fluid. The quarter-wave acoustic resonance is driven on the left by a speaker. The stack acts as a second thermodynamic medium but is not a Stirling-like regenerator because the wide spacing of the plates does not promote locally isothermal conditions.

that is needed to drive the engine through a thermodynamic cycle. This is why a natural but irreversible process—heat flow across a temperature difference—is intrinsic to the operation of the engine.

An interesting contrast exists between the Stirling engine and natural engines. In the Stirling engine, good thermal contact between the working fluid and the second medium helps ensure reversible operation and high efficiency. In the acoustic heat engine, *poor* thermal contact is necessary to achieve the proper phasing between temperature and motion of the working fluid.

One additional condition is important to the operation of the acoustic heat engine: thermodynamic symmetry along the direction of relative motion must be broken. The concept of thermodynamic symmetry is fundamental, yet conceptually simple. In the natural engine, the two thermodynamic media are undergoing reciprocating relative motion along one direction and are interacting thermodynamically in a direction transverse, or laterally, to the motion. If the lateral interaction does not change as we move in the direction of relative motion, we say there is thermodynamic symmetry. But if the lateral interaction changes with the longitudinal coordinate, the symmetry is said to be broken. Where the symmetry is broken there is always some thermodynamic consequence, such as a change of temperature or a heat flow to an external reservoir.

Thermodynamic symmetry can be broken in a variety of ways. For example, in the heat pump depicted in Fig. 5c, it is broken *geometrically* at the longitudinal ends of the fiber glass plates. In our description of some variable stars as natural engines, it is broken by changes in opacity that alter the effective thermal contact between the stellar matter and the radiation field. It can also be broken *dynamically* by, for example, nonlinear localization of the acoustic energy in the primary medium.

The dramatic effects of breaking thermodynamic symmetry can be shown experimentally by fixing several thermocouples to the central plate of a simple acoustic heat pump (Fig. 7). When the acoustic driver or speaker is turned on, the temperature of thermocouples at the ends (where thermodynamic symmetry is broken) changes rapidly and by large amounts, whereas the temperature of other thermocouples further in along the plates changes only by small amounts. In an acoustic natural engine, the heat exchangers are, of course, located at positions where thermodynamic symmetry is broken.

Before explaining in more detail the operation of the acoustic heat engine, we summarize by pointing out that *all* natural heat engines possess the following elements:

- ☐ two or more thermodynamic media in reciprocating relative motion,
- ☐ an irreversible process that causes phasing of a thermodynamic effect with respect to the motion, and
- ☐ broken thermodynamic symmetry along the direction of relative motion.

The Cycle. Figure 8 displays the cycles of an acoustic engine serving as prime mover and as heat pump and also follows a typical parcel of gas as it oscillates alongside one of the fiber glass plates. In a real

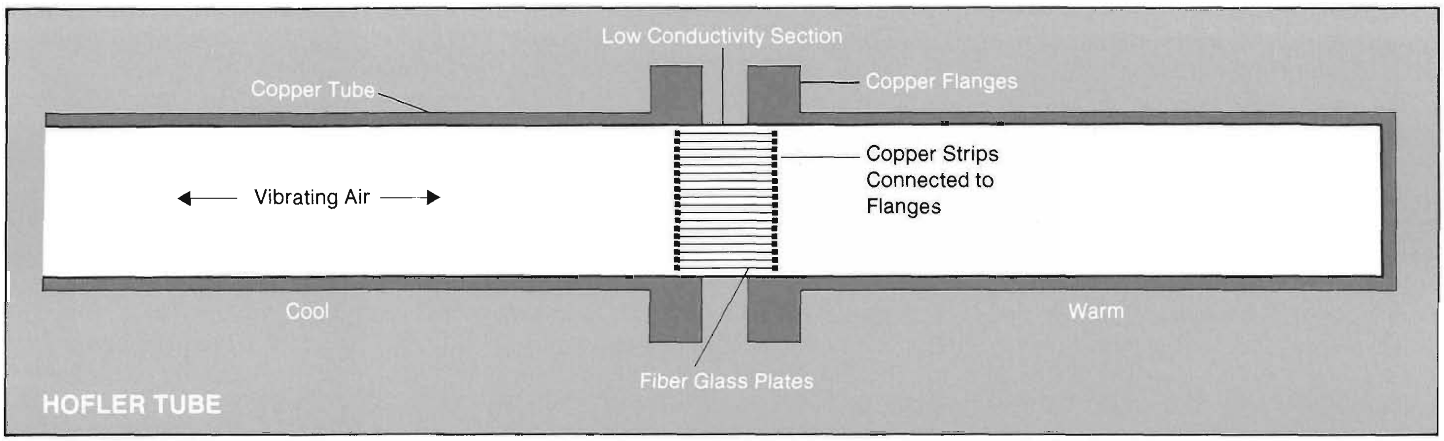


Fig. 6. The Hofler tube, a simple acoustic prime mover that consists of two thermodynamic media—air and fiber glass plates—inside a quarter-wavelength acoustically resonant tube closed at one end. If a steep temperature gradient is applied across the plates, the air in the tube vibrates strongly. The plates are 1.65 cm long, 0.38 mm thick, and spaced 1 mm apart. The stack of plates, here seen from the side, is placed about midway in the tube. Thermal contact between the plates and the tube at both ends of the stack is provided by thin copper strips that run along the longitudinal edges of each plate and into the thick encircling copper flanges.

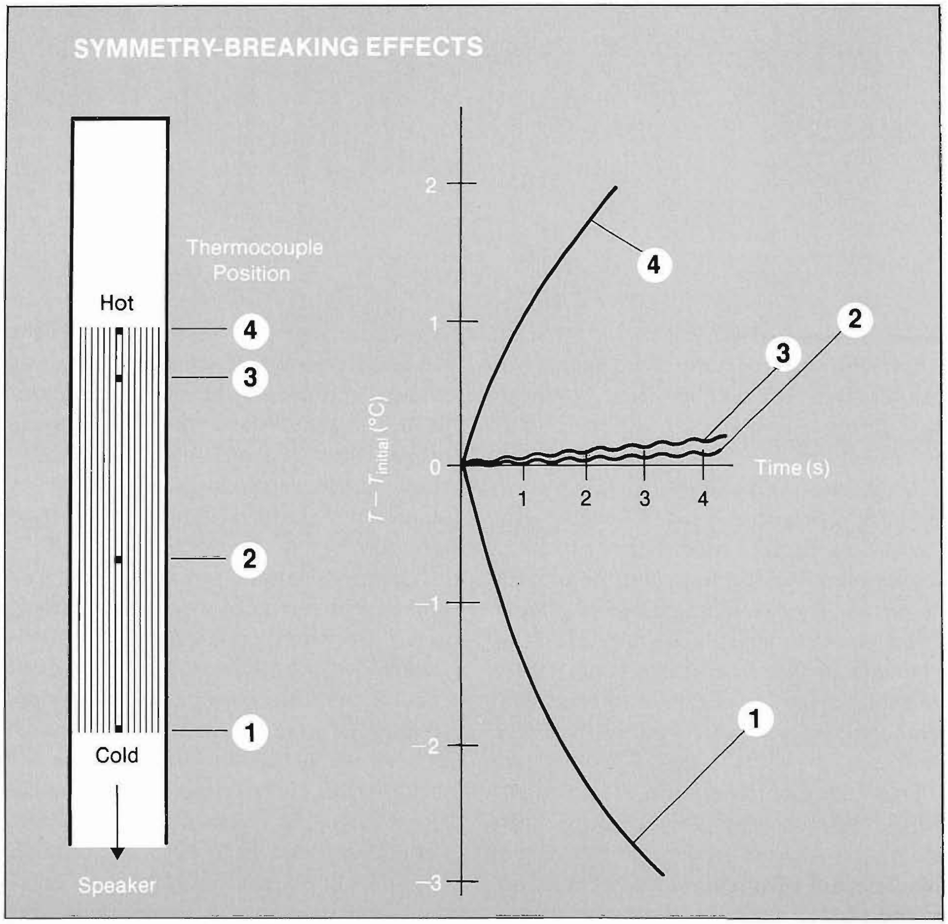


Fig. 7. The temperature change $T - T_{initial}$ of thermocouples placed in a stack of plates of a simple acoustic heat pump shows the effect of symmetry breaking. Application of acoustic power to the tube at time zero immediately produces large changes at the two ends (positions 1 and 4) where thermodynamic symmetry is broken geometrically. Much smaller changes occur at the middle of the stack (position 2) and relatively close to the end (position 3) that are a consequence of a weak dynamic symmetry breaking due to viscosity and the nonuniformity of the acoustic pressure and velocity fields.

acoustic engine, the oscillations are sinusoidal, producing elliptical cycles. For simplicity we consider square-wave, or articulated, motion so that the basic thermodynamic cycle can be pictured as consisting of two reversible adiabatic steps and two irreversible constant-pressure steps, as in the Brayton cycle.

Just as in the Stirling engine, relative phasing of motion (steps 1 and 3 in Fig. 8) and heat transfer (steps 2 and 4) determines whether the acoustic engine is a prime mover or a heat pump. In the Rider form of a Stirling engine, phasing is effected externally by altering the order in which pistons are moved. In an acoustic engine, however, phasing is a result of the natural time delay in the diffusion of heat between the two thermodynamic media. The sign of the relative phasing, and thus the mode of the natural heat engine, is determined by the magnitude of the temperature gradient along the fiber glass plates—a remarkable quality and a substantial gain in simplicity.

During the compressional part of the acoustic standing wave, the parcel of gas is both warmed and displaced along the plates. As a result, two temperatures are important to that parcel: the temperature of the gas after adiabatic compressional warming and the temperature of the part of the plate next to the gas parcel after compression (and displacement). If the temperature of the gas is *higher* than that of the plate, heat will flow from the gas to the plate. If the temperature of the gas is *lower*, heat flows in the opposite direction from plate to gas. Both heat and work flows can thus be reversed and the engine switched between functions by altering the size of the temperature gradient. A zero or low gradient is the condition for a heat pump; a high gradient is the condition for a prime mover. This engine is intrinsically irreversible but *functionally* reversible.

The gradient that separates the two modes is called the *critical* temperature gradient ∇T_{crit}. For this gradient, the temperature change along the plate just matches the temperature change due to adiabatic compression, and no heat flows between the gas and the plate. (Because of losses in a real engine, the maximum temperature gradient that can be produced by a heat pump is somewhat less than ∇T_{crit}, and the minimum gradient needed to drive a prime mover is somewhat greater than ∇T_{crit}.)

Thermoacoustic Couple. The thermoacoustic couple is a simple thermoacoustic device. A calculation of the properties of the thermoacoustic couple demonstrates a good deal of the physics of natural thermoacoustic engines and can be done quantitatively from first principles (see "The Short Stack"). When suitably calibrated, the device can also be used as a probe to measure both acoustically stimulated heat flow and acoustic power.

Typically, such a probe is a single short thin plate of the type used in an acoustic engine (or a small stack of such plates) that can be moved to various longitudinal positions in an acoustically resonant tube. A speaker at the open end of the tube drives the acoustic oscillations.

The material of the plate has a large thermal conductance so that no *substantial* temperature gradient can build up along its length, ensuring that the couple operates under a low temperature gradient as a heat pump. As the probe is moved to various locations in the standing acoustic wave, it measures a flow of heat generated by its presence by detecting a small temperature drop across its length.

Data taken with such a probe (Fig. 9) fit a simple sine curve whose period is *half* the wavelength of the acoustic standing wave. By noting how the sign of the temperature difference varies with respect to the plate's location in the sound wave, we see that heat always flows in the direction

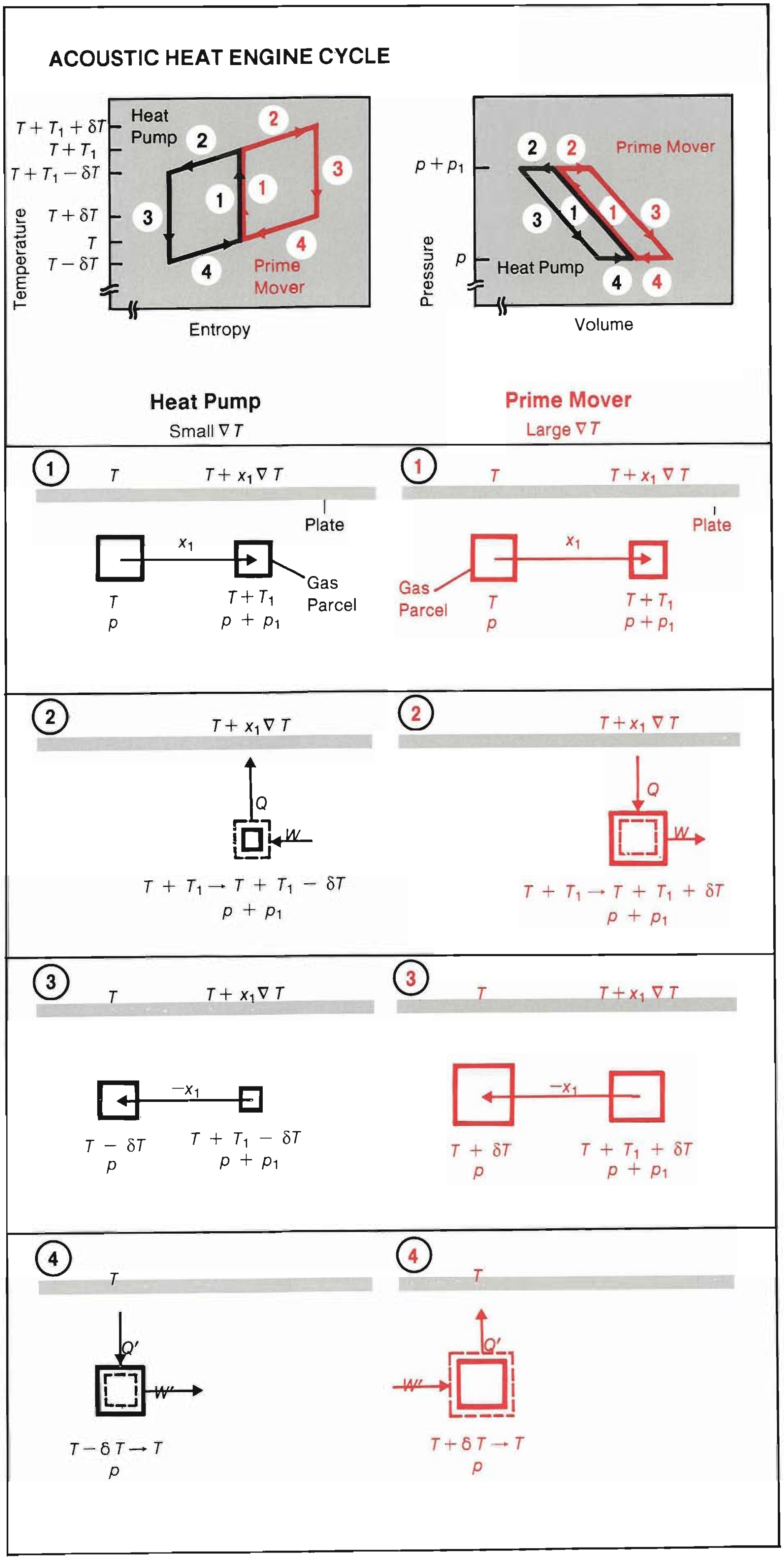
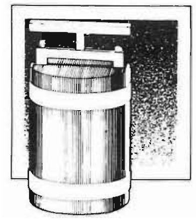


Fig. 8. The thermodynamic cycles (top) of the gas parcels in an acoustic heat engine consist of reversible adiabatic steps and irreversible constant-pressure steps (the acoustic mode is here simplified to articulated rather than sinusoidal motion). This cycle is identical to the Brayton cycle. If we follow a parcel of gas as it moves alongside a fiber glass plate, we see that the prime-mover mode (red) occurs when the temperature rise seen by a gas parcel on the adjacent plate due to displacement of the gas along the gradient ($x_1\nabla T$) is larger than the temperature rise of the gas due to adiabatic compression heating of the gas (T_1). The heat-pump mode (black) occurs under the opposite conditions, that is, when the gradient on the plate is zero or low. In the prime-mover mode, the pressure ($p + p_1$) during the heat-flow expansion step is larger than the pressure (p) during the heat-flow compression step, so net work is added to the acoustic vibration. All flows are reversed in the heat-pump mode, and work is absorbed from the acoustic vibration.◀

of the closest pressure antinode. This effect is expected from the description of the heat pump in Fig. 8 because a parcel of gas moving in the direction of a pressure antinode is compressionally warmed and will transfer heat to the low-gradient plate; a parcel moving toward a pressure *node* is cooled by expansion and will draw heat from the plate. This explains the surprising results of the half-wave resonator heat pump of Merkli and Thomann (Fig. 5b).

At both the pressure antinodes and the pressure nodes, heat flow in the couple drops to zero. This effect occurs because the pressure and the gas velocity in a resonant acoustic wave are spatially 90 degrees out of phase. Thus, a pressure antinode is also a velocity node, and heat flow drops to zero because there is no displacement of gas. On the other hand, a pressure node has zero heat flow because no compression

continued on page 16



THE SHORT STACK

To calculate thermodynamic efficiency for an acoustic heat engine, we need to know the hydrodynamic heat flow and the work flow. A heuristic derivation of these two quantities and the resulting efficiency for the particular case of a short stack follow. We then briefly discuss the effects of viscosity.

Heat Flow

Consider a stack of plates in a heat engine whose length is short compared to the acoustic wavelength and to the distance from the stack to the end of the tube. If that length is short enough, we can ignore the change in the longitudinal acoustic velocity magnitude u_1 and the change in the dynamic, or acoustic, pressure magnitude p_1 with respect to longitudinal distance x (measured from the end of the acoustically resonant tube). Further, if we ignore the effects of fluid viscosity, u_1 does not depend on *lateral* distance from the plates. Next, we can take the lateral distance between plates to be large compared to the thermal penetration depth δ_κ (the characteristic length for heat transfer in the fluid during a given cycle of the acoustic wave). Thus, any effects we estimate for a stack of plates will be the same as for a single plate having the same overall perimeter Π (measured transverse to the flow).

The adiabatic temperature change T_1 accompanying the pressure change p_1 can be derived from thermodynamics and is

$$T_1 = \frac{T_m\beta}{\rho_m c_p}\, p_1 \ , \tag{1}$$

where T_m is the mean absolute temperature, β is the isobaric expansion coefficient, ρ_m is the mean density, and c_p is the specific heat at constant pressure.

The change of entropy for a parcel oscillating in the manner depicted in Fig. 8 of the main text is just the lateral heat flow from the second medium divided by T_m or $\rho_m c_p \delta T/T_m$ per unit volume, where δT is the change in the fluid temperature due to that heat flow. The volume transport rate for that part of the fluid that is thermodynamically active is $\Pi\delta_\kappa u_1$. We thus can estimate the flow of hydrodynamically transported heat $\dot{Q}$ as the product of these two quantities times T_m; that is,

$$\dot{Q} \sim \Pi\delta_\kappa u_1 \rho_m c_p \delta T \ . \tag{2}$$

Now from Fig. 8 we also see that

$$\delta T = T_1 - x_1\,\nabla T = T_1\left(1 - \frac{\nabla T}{T_1/x_1}\right) \ , \tag{3}$$

where ∇T is the temperature gradient along the plate and x_1 is the fluid displacement. The value of ∇T that makes $\delta T = 0$ is the critical gradient, so

$$\nabla T_{\rm crit} = \frac{T_1}{x_1} \ . \tag{4}$$

Combining these equations and defining the temperature gradient ratio parameter as $\Gamma \equiv \nabla T/\nabla T_{\rm crit}$ gives an estimate for the hydrodynamic heat flow as

$$\dot{Q} \sim -\,\Pi\delta_\kappa (T_m\beta) p_1 u_1 (\Gamma - 1) \ . \tag{5}$$

The parameter $T_m\beta$ is what we call the *heat parameter* of the fluid. The presence of the $\Pi\delta_\kappa$ factor is obvious because it is the thermodynamically active area in a plane perpendicular to the longitudinal acoustic motion. The formula shows that when $\Gamma < 1$, heat flows up the temperature gradient, as for a heat pump; when $\Gamma = 1$, there is no heat flow; when $\Gamma > 1$, heat flows down the temperature gradient, as for a prime mover.

Work Flow

Now that we have estimated the heat flow, we need to calculate the work flow, which is given by the work per cycle (the area $p_1\delta V$ enclosed by the pressure-volume diagram in Fig. 8 of the main text) times the rate at which that work occurs (the angular acoustic frequency ω). The volumetric change δV that will contribute to the net work is just

$$\frac{\delta V}{V} = \beta\delta T \ , \tag{6}$$

where δT is the temperature change of Eq. 3. V, the total volume of gas that is thermodynamically active, is given by

$$V = \Pi\delta_\kappa \Delta x \ , \tag{7}$$

where Δx is plate length.

We can now simply put these pieces together and, using Eqs. 1, 3, 6, and 7, write down the work flow as

$$\dot{W} \sim p_1 \delta V \omega$$
$$\sim \Pi\delta_\kappa \frac{T_m\beta^2}{\rho_m c_p}\, p_1^2 \omega \Delta x (\Gamma - 1) \ . \tag{8}$$

From thermodynamics we know that

$$\gamma - 1 = \frac{T_m\beta^2 a^2}{c_p} \ , \tag{9}$$

where the quantity $\gamma - 1$ is what we call the *work parameter* of the fluid, and a is the speed of sound, so we can rewrite the expression for work flow as

$$\dot{W} \sim \Pi\delta_\kappa(\gamma - 1)\,\frac{p_1^2}{\rho_m a}\,(\Gamma - 1)\,\frac{\Delta x}{\lambda\!\!\!^{-}}\ , \qquad (10)$$

where $\lambda\!\!\!^{-} = a/\omega$ is the radian length of the acoustic wave.

The formulas for estimating $\dot{Q}$ and $\dot{W}$ (Eqs. 5 and 10) have a very similar structure, which is expected since they are closely related thermodynamically. The heat parameter $T_m\beta$ appears in the formula for $\dot{Q}$, and the work parameter $\gamma - 1$ appears in the formula for $\dot{W}$. Both $\dot{Q}$ and $\dot{W}$ are quadratic in the acoustic amplitude p_1 or u_1; both change sign as Γ passes through unity.

Efficiency

A quantitative evaluation of $\dot{W}$ and $\dot{Q}$ for this case of the short stack but for sinusoidal p_1 and u_1 would give the same results except each formula has a numerical coefficient of 1/4. Thus the efficiency η of a short stack with no viscous or longitudinal conduction losses is

$$\eta = \frac{\dot{W}}{\dot{Q}} = \frac{\gamma - 1}{T_m\beta}\,\frac{\omega\Delta x p_1}{\rho_m a^2 u_1}\ . \qquad (11)$$

For our standing acoustic wave, $u_1 = u_0 \sin x/\lambda\!\!\!^{-}$ and $p_1 = \rho_m a u_0 \cos x/\lambda\!\!\!^{-}$, where x is the distance of the stack from the end of the tube. Then the efficiency can be rewritten simply as

$$\eta = \frac{\gamma - 1}{T_m\beta}\,\frac{\Delta x}{\lambda\!\!\!^{-} \tan x/\lambda\!\!\!^{-}}\ . \qquad (12)$$

In the important limit of $x \ll \lambda\!\!\!^{-}$, the efficiency is simply

$$\eta = \frac{\gamma - 1}{T_m\beta}\,\frac{\Delta x}{x}\ . \qquad (13)$$

Thus, in either case, efficiency depends only on geometry and fluid parameters, just as for the Brayton and Otto cycles discussed in the text. The temperatures T_h and T_c do not enter.

As the *actual* temperature gradient approaches the critical temperature gradient ∇T_{crit}, the temperature difference δT approaches zero, so that even at the acoustic angular frequency ω the heat transfer rate and the power output approach zero, just what is needed to give the Carnot efficiency in the Brayton and Otto cycles. What happens in this engine? We use Eqs. 1, 4, and 9 and the fact that $u_1 = x_1\omega$ to rewrite the efficiency formula (Eq. 11) in general as

$$\eta = \frac{\Delta x \nabla T_{crit}}{T_m}\ . \qquad (14)$$

Because $\Delta T = \Delta x \nabla T_{crit}$ when $\nabla T = \nabla T_{crit}$, we have at the critical temperature gradient

$$\eta = \frac{\Delta T}{T_m}\ . \qquad (15)$$

The Carnot efficiency is $\eta_C = 1 - T_c/T_h$. But if $T_c = T_h - \Delta T$, and if $\Delta T/T_h$ is small so that T_h can be replaced by T_m, we get, with our approximations, the same formula for η_C as Eq. 15. So the acoustic engine approaches Carnot's efficiency as the power output and heat transfer rates approach zero, just like the Otto and Brayton cycles.

What About Viscosity?

So far we have assumed that the working fluid is inviscid. What if it is not? We know how to do the theory quantitatively for this more general case, but the resulting expressions for $\dot{Q}$ and $\dot{W}$ are terribly complicated and opaque. We can simplify them by assuming that the Prandtl number (the square of the ratio of the viscous penetration depth δ_v to the thermal penetration depth δ_κ) is small. In that case we obtain

$$\dot{Q} = \frac{1}{4}\,\Pi\delta_\kappa(T_m\beta)p_1u_1(\Gamma - 1) - \frac{1}{4}\,\Pi\delta_v(T_m\beta)p_1u_1\ , \qquad (16)$$

$$\dot{W} = \frac{1}{4}\,\Pi\delta_\kappa(\gamma - 1)\,\frac{p_1^2}{\rho_m a}\,\frac{\Delta x}{\lambda\!\!\!^{-}}\,(\Gamma - 1) - \frac{1}{4}\,\Pi\delta_v\rho_m a u_1^2\,\frac{\Delta x}{\lambda\!\!\!^{-}}\ . \qquad (17)$$

To lowest order, then, the effect of viscosity on heat flow is just to decrease $\dot{Q}$ by a term proportional to the viscous penetration depth. This simply means that viscosity prevents a layer of fluid of thickness δ_v adjacent to the plate from moving acoustically and contributing to the acoustically stimulated heat transport. Similarly, the work flow is decreased by a term proportional to δ_v; this term is simply the energy lost from the acoustic wave due to viscous drag on the plate.

For simplicity in Eqs. 16 and 17 we have kept our old definition of ∇T_{crit}, even though another effect of viscosity is to make the concept of a critical temperature gradient less well defined. In fact, with viscosity present there is a lower critical gradient below which the engine pumps heat and a higher critical gradient above which the engine is a prime mover. Between these two gradients the engine is in a useless state, using work to pump heat from *hot* to *cold*.

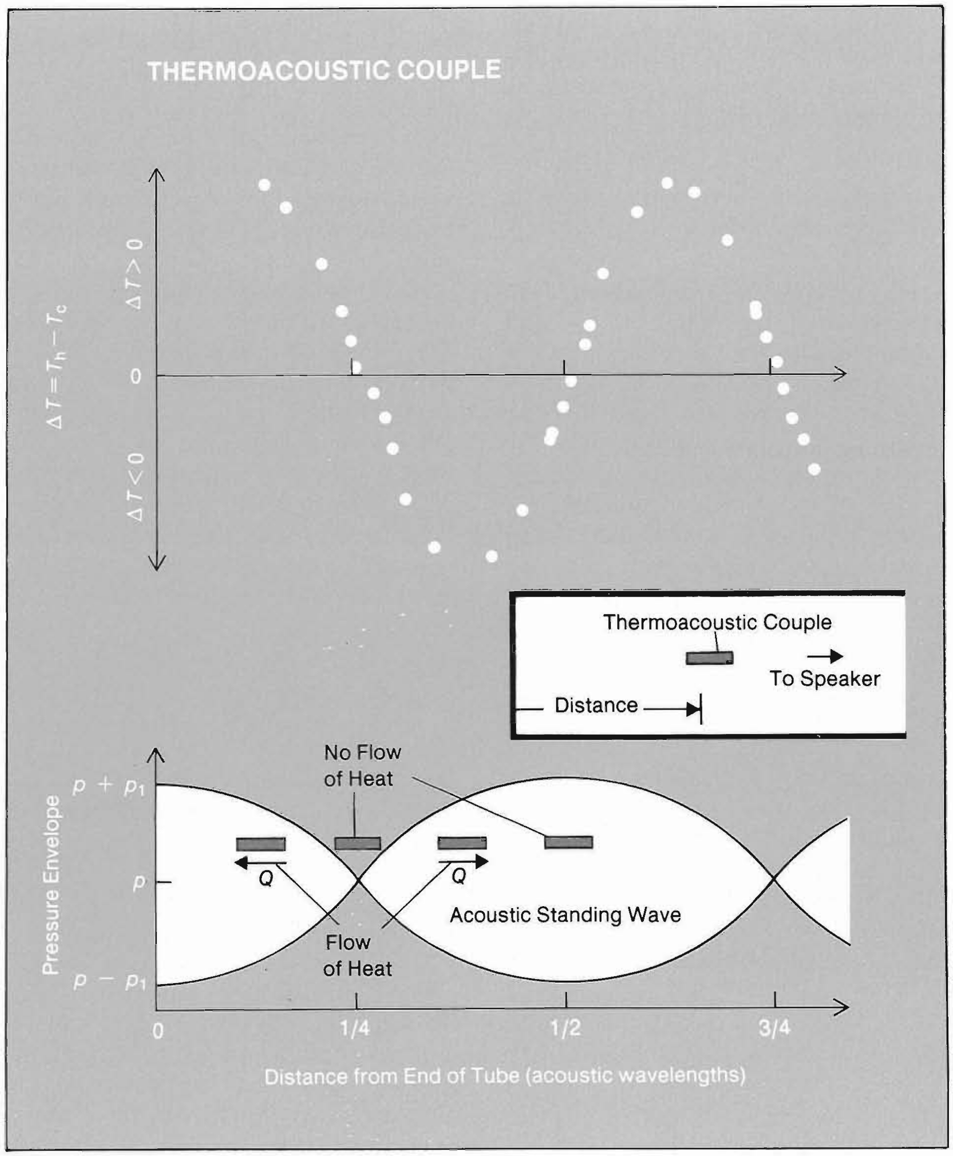
The Prandtl number for helium gas is about 0.67, so that viscous effects are very significant for our gas acoustic engines (and, in fact, Eqs. 16 and 17 are rather poor approximations). On the other hand, the Prandtl number for liquid sodium is about 0.004, so that viscous effects are much smaller. ■

continued from page 13

or expansion takes place. In other words, acoustic heat flow depends on both the acoustic pressure and the fluid velocity.

Figure 9 also illustrates the dramatic effect the positioning of a plate in the acoustic wave has on the operation of a natural heat engine. A plate or stack of plates placed completely within a quarter of a wavelength of the end of the tube operates in the manner depicted in Fig. 8. If that same stack is repositioned in the second quarter of a wavelength, the pictorial analysis of Fig. 8 still applies, but the directions of all heat flows and longitudinal temperature gradients are reversed. A stack that extends beyond an adjacent node-antinode pair, however, has heat flows that counter each other, canceling part of the overall transport of heat from one end of the plates to the other.

Also important is the stack's position *within* a given node-antinode pair separated by a quarter of an acoustic wavelength (Fig. 10). For an engine in the heat-pump mode, a stack close to a pressure antinode—say, the end of the tube—can develop steep temperature gradients. Why? In such a region the acoustic pressure change in a parcel of gas is large and thus the rise in temperature from compressional warming is large. This region is also near a velocity node, so displacement of the gas parcel is small. Large temperature changes over small displacements, of course, result in large temperature gradients. (Or one can say that ∇T_{crit}, which bounds the region between the heat-pump and prime-mover modes, is large close to a pressure antinode.)



As a plate or stack of plates is moved away from the pressure antinode, the temperature gradient developed becomes smaller. At a quarter of a wavelength, no gradient forms (∇T_{crit} equals zero). This positioning effect is important in the design of a refrigerator, because, together with the length of the plates, it places an upper limit on the maximum temperature drop possible across the stack.

Positioning also affects the losses that characterize an engine. For example, a stack close to a pressure antinode is close to a velocity node, and viscous losses will be small at that position. However, because temperature gradients are steep there, losses from ordinary diffusive thermal conduction in the plates and working fluid will be increased. The problem of ordinary conduction losses is especially critical for an engine acting as prime mover because such an engine needs a temperature gradient higher than ∇T_{crit} to work. Thus, the positioning of the set of plates is a tradeoff between viscous losses, losses from longitudinal conduction, the desired temperature span across the engine, and power output.

Heat and Work. We can now better understand the natural acoustic engine by examining what happens near a short plate positioned between a node and an antinode (Fig. 11). If the displacement of a given parcel of gas is small with respect to

Fig. 9. The temperature difference ΔT measured across a thermoacoustic couple as a function of the plate's position in the acoustic standing wave. Note that heat flows toward the closest pressure antinode, making that end of the couple hottest. However, at both the pressure antinodes and nodes there is no flow of heat ($\Delta T = 0$). These data were obtained for an acoustic wavelength of approximately 80 cm.◀

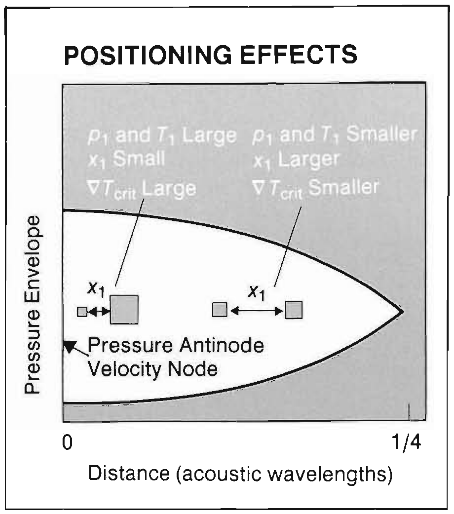


Fig. 10. Close to a pressure antinode a typical parcel of gas experiences large changes in pressure p_1 and thus large changes in temperature T_1 due to compressional heating. At the same time, displacement x_1 of the parcel is small, so that $\nabla T_{crit} = T_1/x_1$ is large; in fact, since x_1 is proportional to the distance x from the pressure antinode, $\nabla T_{crit} \propto x^{-1}$. In the heat pump mode, the maximum temperature gradient that can be developed is equal to ∇T_{crit} (since heat flow between gas parcel and plate stops when that gradient is reached), which means that close to a pressure antinode we can expect large temperature gradients. Further from the pressure antinode, pressure and temperature changes become smaller whereas displacements become larger, so the maximum temperature gradient that can be developed is smaller.▲

the length of the plate, there will be an entire train of adjacent gas parcels, each confined in its cyclic motion to a short region of length x_1 and each reaching the same extreme position as that occupied by an adjacent parcel half a cycle earlier. What is the net result of all these individual cycles on the flow of heat and work?

If the motion of the parcels is sinusoidal, only those about a thermal penetration depth* from the nearest plate are thermoacoustically effective. Parcels close to a plate transfer heat to and from the plate in a locally isothermal and reversible manner, just like the fluid in the regenerator of a Stirling engine. Parcels far away have no thermal contact and are simply compressed and expanded adiabatically and reversibly by the sound wave. However, parcels that are at about a thermal penetration depth from a plate have good enough thermal contact to exchange some heat with the plate but, at the same time, are in poor enough contact to produce a time lag between motion and heat transfer.

During the first part of the cycle for the heat-pump mode, the individual parcels will each move a distance x_1 toward the pressure antinode and deposit an amount of heat Q at that position on the plate. During the second half of the cycle, each parcel moves back to its starting position and picks up the same amount of heat Q from the plate. But this heat was deposited there a half cycle earlier by an adjacent parcel of gas. In effect, an amount of heat Q is merely passed along the plate from one parcel of gas to the next in the direction of the pressure antinode. Thus, as in the Stirling engine, the second thermodynamic medium is used for the temporary storage of heat.

At the ends of the plates, the thermodynamic symmetry is broken. Parcels of gas that move farther from the end of the plate than a thermal penetration depth idle through part of their cycle without accepting or rejecting heat. For example, if a parcel of gas at the end closest to the antinode is in equilibrium with the plate on one half of the cycle but then moves out of the range of thermal interaction, it has nowhere to deposit the heat resulting from its adiabatic warming. As this parcel completes its cycle, it cools adiabatically back to the temperature of the plate. The heat transferred to the plate from the next adjacent parcel down the line is uncompensated, so there is a net heat transfer to the plate on that end, and the temperature of the plate increases there. In similar fashion, heat drawn from the end closest to the node is not replaced, and that end cools. We can take advantage of the net effect—a flow of heat from one end to the other—by bringing the ends of the plates into contact with heat exchangers.

During each cycle an individual parcel of gas transports heat Q across only a small temperature interval along the plate that is comparable to the adiabatic temperature change T_1. However, because there are many parcels in series, the heat Q is shuttled down the stack, thereby traversing the temperature interval $T_h - T_c$, which can be much larger than T_1. Within the limits of a quarter of a wavelength, the flow of heat is not a strong function of plate length (in fact, for a stack much shorter than a quarter of a wavelength, heat flow does not depend on plate length at all).

If, on the other hand, we examine this train of gas parcels with respect to the flow of *work*, we realize that *each* parcel has a net effect. For example, a parcel of gas near the plates in an engine operating in the heat-pump mode absorbs net work because its expansion is at a lower pressure than the corresponding compression. But since the same is true for every parcel in the train, the total work done on the gas is roughly proportional to plate length (for a very short stack, work flow *is* proportional to plate length).

**The thermal penetration depth δ_κ is the characteristic length describing heat diffusion through the gas during one period of the acoustic cycle. Mathematically, $\delta_\kappa = \sqrt{\kappa/\pi f}$, where κ is the thermal diffusivity of the gas and f is the frequency of the sound.*

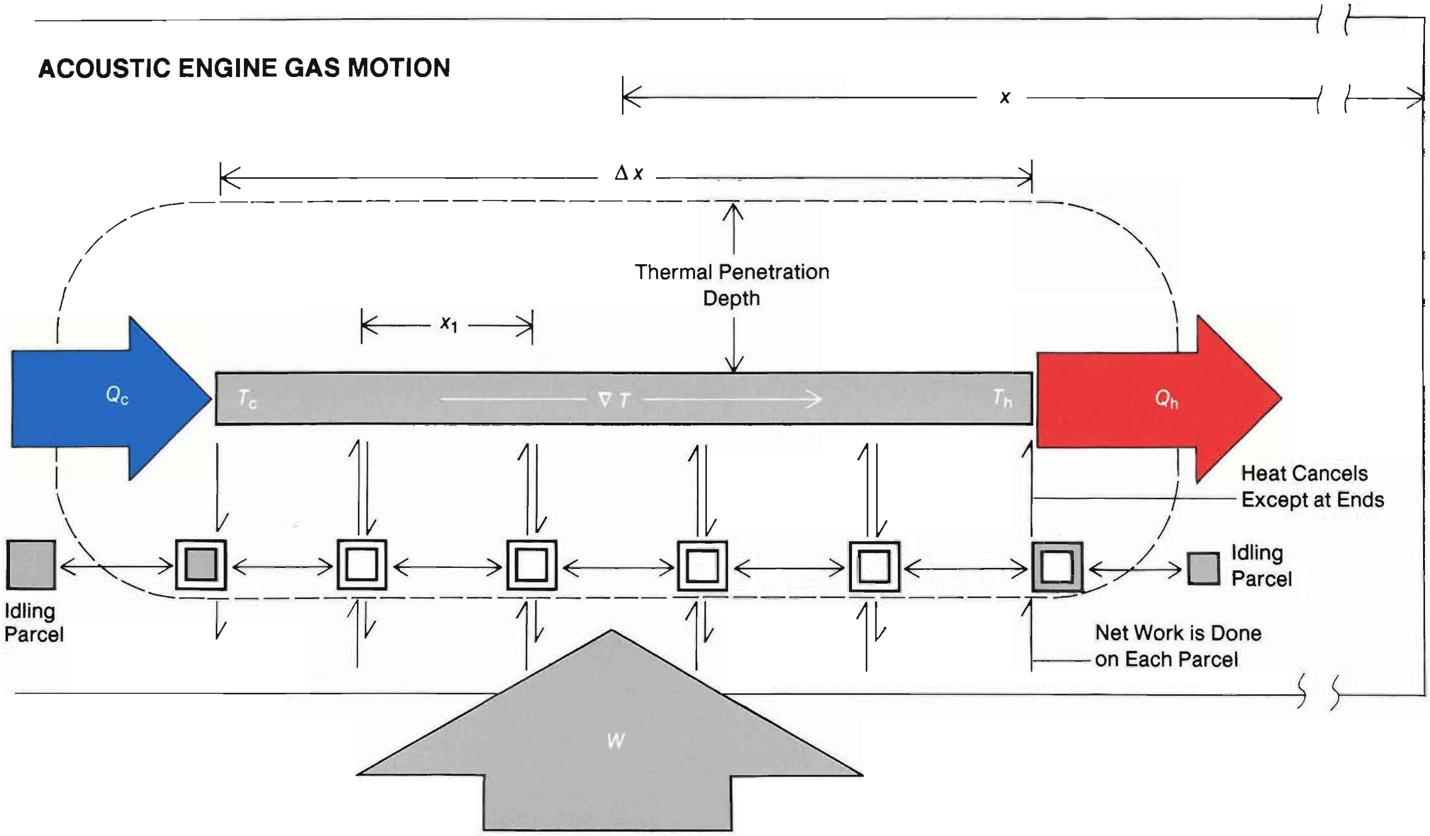
Efficiency. A calculation of heat and work flows for an acoustic heat engine with a short stack close to the end of the resonator tube and no viscous losses (see "The Short Stack") yields a limiting efficiency given by

$$\eta = \frac{\gamma - 1}{T_m \beta} \frac{\Delta x}{x} \qquad (3)$$

where T_m is the mean absolute temperature of the plates, β is the isobaric expansion coefficient, Δx is the plate length, and x is the distance of the plates from the pressure antinode (usually the end of the tube). This efficiency depends on properties of the working fluid (the *work parameter* $\gamma - 1$ and the *heat parameter* $T_m\beta$) and on the geometry of the engine ($\Delta x/x$), as in the Otto and Brayton cycles, rather than on the hot and cold temperatures, as in the Carnot and Stirling cycles.

It can also be shown that when the temperature gradient along the plates equals ∇T_{crit}, Eq. 3 reduces to the Carnot efficiency. This result is expected because, for such a gradient, rates of heat transfer approach zero, all processes approach reversibility, and no entropy is generated during the cycle. Once again, however, the approach to maximum efficiency means an approach to zero power.

A Natural Magnetic Engine. To help emphasize the generality of natural engines, we now discuss a hypothetical natural *magnetic* engine. The acoustic engine uses a fluid as its primary medium; our postulated natural magnetic engine uses a solid magnetic material. Further, the operation of the acoustic engine is based on the adiabatic change of temperature with pressure in a fluid, whereas the mag-



▲
Fig. 11. An acoustic heat engine can be thought to have a long train of adjacent gas parcels, all about a thermal penetration depth from the plate, that draw heat from the plate at one extreme of their oscillatory motion and deposit heat at the other extreme. However, idling parcels at both ends oscillate without removing or depositing heat. Adjacent heat flows cancel except at the ends; the net result is that an amount of heat *Q* is passed from one end of the plate to the other. Adjacent work contributions do not cancel, so that each parcel of gas contributes to the total work.

Tom Hofler attaching the resonator of the cryocooler to the housing of the cooler's driver and hot heat exchanger.

netic engine is based on the *adiabatic change of temperature with magnetization.*

The primary medium of our hypothetical apparatus (Fig. 12) consists of a stack of magnetic disks.* Each disk has a high internal thermal conductance, but each is also thermally insulated from the others so that a large temperature gradient can be sustained in the longitudinal direction.

The collection of disks is placed in a tube whose walls constitute the second medium. Like the first medium, the second has a high lateral thermal conductance, a large heat capacity, and a low longitudinal thermal conductance. Thus, it, too, can sustain a large temperature gradient parallel to the direction of relative motion between the two media.

The device is positioned between the poles of a permanent magnet in such a way that the disks of the primary medium are in the nonuniform fringing field at the side. The disks are linked mechanically to an external mechanism so that they can be moved in a reciprocating fashion. A gas fills the small annular space around the magnetic disks, providing lateral thermal contact with the second medium, but this contact is poor enough to create the necessary phasing for the engine. There is also some means for heat exchange with external reservoirs at each end of the second medium.

If we follow an element of the first thermodynamic medium in a magnetic engine through an articulated cycle (Fig. 13), we see that the various steps are analogous to those of an acoustic heat engine. For example, in the first step of a heat-pump cycle, the element is moved quickly and adiabatically to a region of higher magnetic field. As a result, its temperature rises. (Temperature changes of a few degrees per tesla are typical for ferromagnetic and strongly paramagnetic materials.) In the second step, the element thermally relaxes, its temperature adjusting to that of the adjacent region in the second medium, which, in the heat-pump mode, means that heat flows from the first medium to the second. As the element cools, its magnetization increases. The third step is motion back to a region of lower field; the fourth is another thermal relaxation. As in the case of acoustic engines, the phasing between motion and heat transfer is a result of the natural time delay caused by diffusion of heat between the two media.

**We simplify our discussion by assuming the magnetic material is an ideal Curie-law paramagnet, for which the magnetization m is given by $m = \lambda H/T$ and the entropy S is given by $S = S_0 - (\lambda/2)\, H^2/T^2$, where H is magnetic field, T is temperature, and λ and S_0 are constants. Thus, for adiabatic processes m is constant; for constant-field processes m decreases as T increases.*

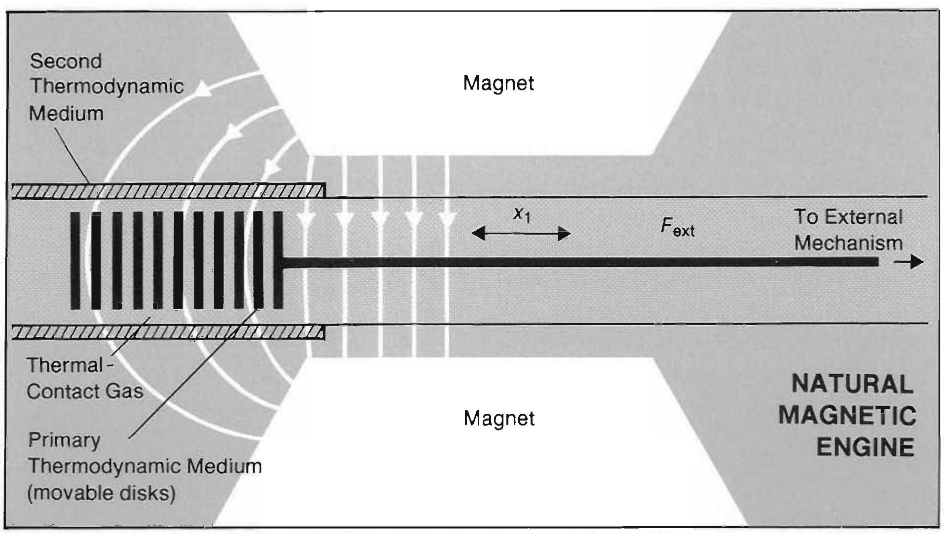


Fig. 12. A hypothetical natural magnetic engine in which the primary medium consists of magnetic disks placed in the fringing field at the side of a permanent magnet. This placement allows an external mechanism to displace the disks in a reciprocating fashion in the presence of a magnetic field gradient. A gas surrounding the disks makes thermal contact with the second medium (the walls of the tube), but conductivity of the gas is poor enough to create the necessary phasing for the engine. In both media, thermal conductance is poor longitudinally so that large temperature gradients can be supported. ▲

Some Applications of Natural Engines

What happens now if these ideals of natural engines are put into practice? What are the clanking, hissing realities of real natural engines?

Cryocooler. As part of his Ph.D. thesis, Tom Hofler designed and built a device called the cryocooler (Fig. 14) in which the numerical aspects of the design were based on the general thermoacoustic theory of Rott for ideal gases.

The cryocooler is an acoustic cooling device with a number of important features. Perhaps the most important is the fact that the acoustic resonance is driven from the *hot* end of the stack. All the early cooling engines were arranged with the stack near the closed end of the acoustic resonator tube and with the acoustic driver at the opposite end. Very large temperature differences (about 100 centigrade degrees) could be easily induced across the entire stack this way, but the cold end was seldom less than 20 degrees below ambient temperature. The problem was that the driver (at ambient temperature) and the cold end of the stack maintained moderately good thermal contact with one another by means of *acoustic streaming*. This phenomenon is a second-order, steady circulatory flow of the working gas that is superimposed on the oscillatory motion. The effect of the acoustic streaming was to use up a substantial amount of the refrigeration available at the cold end of the stack trying to cool down the driver.

Now while it is necessary for work to flow into the stack to pump heat, we realized that it is of no real importance whether that flow occurs at the cold or the hot end. Putting the driver at the hot, or closed end, means that none of the available refrigeration is used to cool the driver. Thus, with the "closed" end replaced by a movable piston acting at high dynamic pressure and low displacement, performance is improved.

The stack with its heat exchangers was placed rather close to the driver piston, that is, rather close to a pressure antinode. As noted earlier (see Fig. 10), such a region has the high critical temperature gradient needed for a heat pump. Of course, some separation between driver and stack is necessary because acoustically stimulated heat transfer is proportional to the displacements of the parcels of gas.

Such a configuration means the remainder of the resonator tube can be at the cold temperature, allowing it to be just a thermally insulated straight tube roughly half an acoustic wavelength long. However, losses due to the dynamical effects of viscosity and thermal conduction along the walls of the resonator reduce the externally available refrigeration. Roughly half this loss could be eliminated by using a quarter-wavelength resonator with one end open, but an open end eliminates the use, say, of several atmospheres of helium as the working fluid and revives the original heat load problem of acoustic streaming—here between the driver and the atmosphere. Moreover, an open end is downright noisy, radiating useful work out into the room.

The simple solution is to replace approximately half the half-wavelength resonator with a closed container of substantial volume. Dynamic pressure will be small in a region of large volume, making the losses correspondingly small. The res-

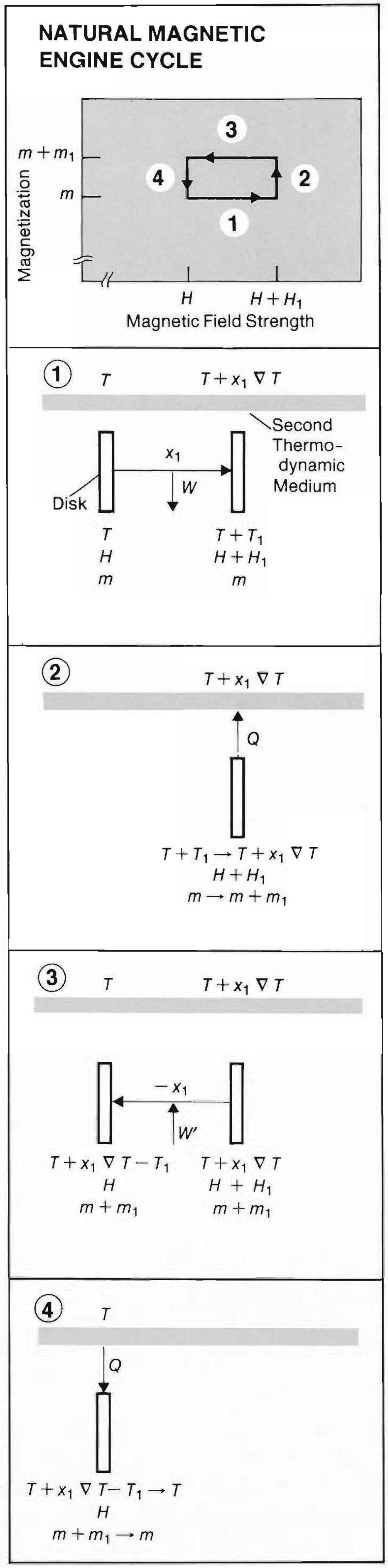
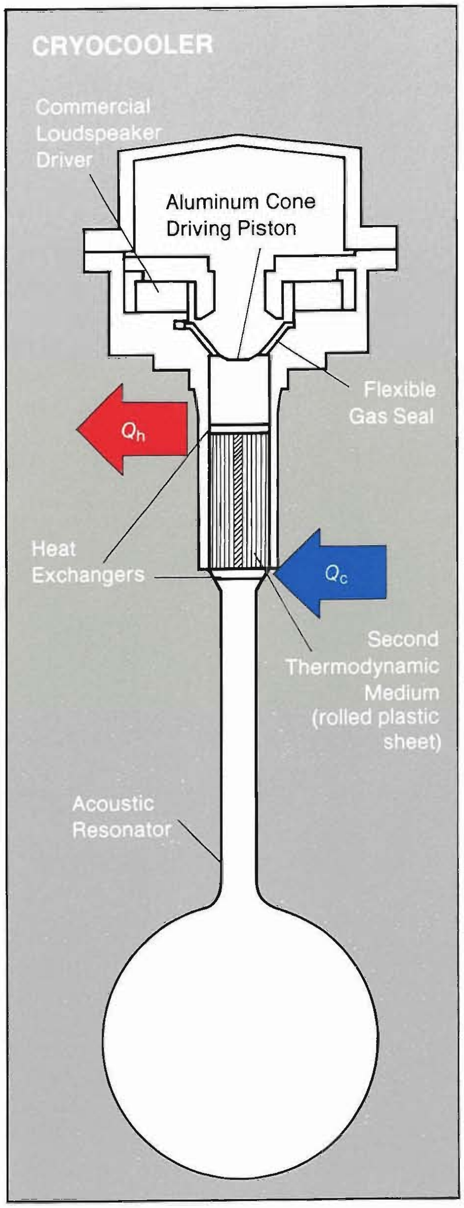
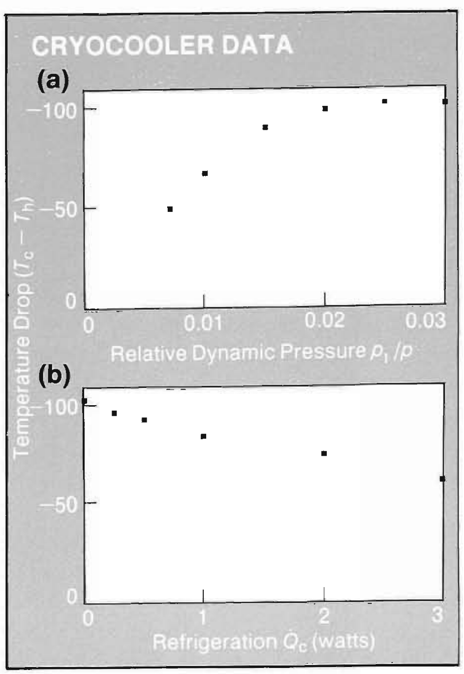


Fig. 13. The cycle shown here for the heat-pump mode of a hypothetical natural magnetic engine is analogous to the cycle for the heat-pump mode of the acoustic heat engine (Fig. 8) with magnetic field strength H taking the role of pressure and magnetization m taking the role of volume. Thus, the cycle consists of reversible adiabatic steps and irreversible constant-field steps. For an ideal Curie-law paramagnet, $m \propto H/T$. Thus, in the first step, when the disk moves adiabatically to a region of higher magnetic field, magnetization remains constant and temperature rises with increasing H. In the second step, heat flows to the lower temperature of the second medium, causing the disk to cool at constant H and the magnetization to increase. The net result of all four steps is the transport of heat up the gradient as a result of the work (which will equal m_1H_1) needed to move the disk through its cycle.





onator below the stack was modified further by *decreasing* the diameter of the confining tube and shortening its length. This last modification, at first sight, would appear to be of negative value as one would expect viscous losses to go up; but, for small decreases in neck diameter, dynamic thermal-conduction losses go down faster than viscous losses go up, and there is a net decrease in the overall losses. These surprising qualities explain the general shape and configuration of the cyrocooler.

Performance is rather good (Fig. 15). As the relative dynamic pressure amplitude increases, the temperature difference that can be pumped up for zero external heat load increases, eventually topping out at about −100°C when the acoustic pressure amplitude is about 2 or 3 per cent of the mean pressure. At that point, the cryocooler can handle a significant refrigeration load and still maintain a rather low temperature. This type of refrigeration capability is very suitable for cooling instruments and sensors.

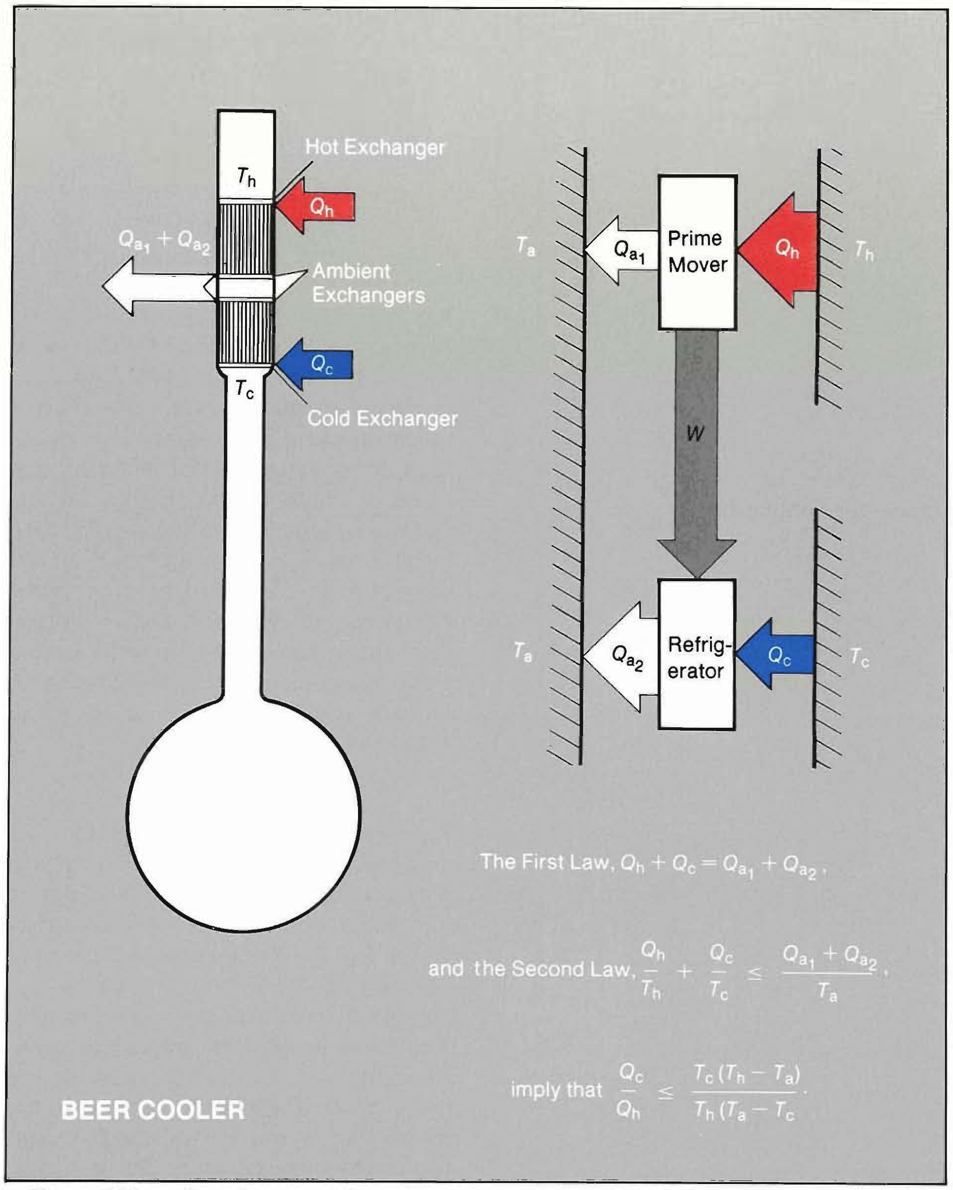
Fig. 14. The driver in the acoustic cyrocooler is an ultralight aluminum cone attached to the voice coil of a commercial loudspeaker. The second thermodynamic medium, rather than being a set of parallel plates, consists of a sheet of Kapton rolled about a vertical rod and spaced with 15-mil nylon fishing line aligned vertically. Copper heat exchangers are attached at both ends. The form of the bulb and neck, including the constriction, were chosen to reduce viscous and thermal losses by reducing surface area. The device is drawn to scale and is about 50 cm long. ◀

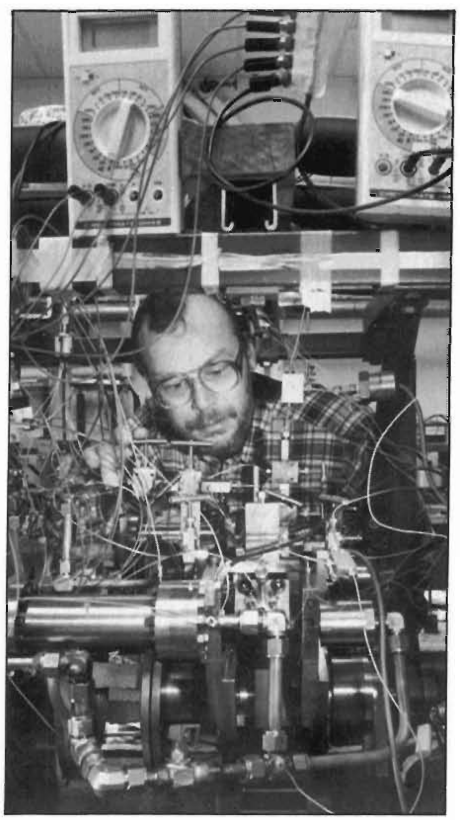
Fig. 15. Experimental data for the cryocooler of Fig. 14, obtained with a mean helium pressure of about 10 bars and acoustic frequencies in the range of 540 to 590 Hz. For thermal isolation the engine was placed in an evacuated vessel and surrounded by superinsulation. The frequency was adjusted electronically so the dynamic pressure and velocity were always in phase at the driver. Part (a) shows how the temperature difference between the hot heat exchanger at approximately 26°C and the cold heat exchanger increases with relative dynamic pressure amplitude (the ratio of the acoustic pressure amplitude p_1 at the pressure antinode to the mean pressure p). No heat load was applied to the cold heat exchanger. Part (b) shows how, for a relative dynamic pressure amplitude of 0.03, the temperature difference gradually drops with increasing refrigeration load at the cold heat exchanger. ◀

A Heat-Driven Acoustic Cooler. Natural acoustic engines are functionally reversible: they can be either prime movers that use heat to produce sound or heat pumps that use sound to refrigerate. Why not combine these two functions in one device and use heat to cool? Such an engine would have heat flow through the walls but no external flow of work.

A key problem in the design of a heat-driven acoustic cooler is where to position the two sets of plates—one set acting as prime mover, the other acting as refrigerator. Ideally, the refrigerator plates should be positioned as they are in the cryocooler, that is, close to the end of the tube where the velocity of the gas and the viscous losses are low but ∇T_{crit} is high. Although it would also be good to keep viscous losses low for the prime-mover plates, it is more important to have these plates near a velocity antinode where ∇T_{crit} is small enough for the stack to develop adequate power. These considerations imply that the refrigerator stack needs to be closer to the end of the tube than the prime-mover stack. However, such a configuration would put the hottest region (the hot end of the prime mover) next to the coldest region (the cold end of the refrigerator), creating a difficult thermal-design problem.

Fig. 16. The upper set of plates in this cooler is a prime mover that draws heat from a heater (at about T_h = 390°C) and rejects waste heat to cooling coils (at T_a = 23°C or room temperature), generating acoustic work. The lower engine uses that work to reject heat to the cooling coils (at T_a) and to draw heat from an even lower temperature (T_c = 0°C or the ice point). The acoustic tube is about half a meter in length, terminates in a 2-liter bulb, and contains helium at a pressure of 3 bars that resonates at a frequency of 585 Hz. Both sets of plates are made of 10-mil (0.025 cm) stainless steel, and the spacing between plates in both sets is 0.08 cm. The hot heat exchanger is made of nickel strips; the ambient and cold heat exchangers of copper. ▶





Bob Oziemski adjusting the flow of the working fluid in the liquid propylene Stirling engine. ▲

To avoid large heat inputs to the cooler, the positions of the two stacks can be reversed and the various temperatures arranged in decreasing order along the tube. What now becomes paramount is for conditions to be such that the prime mover is able to adequately drive the cooler. Amplification of acoustic fluctuations occurs only above a critical value of Γ ($\equiv \nabla T/\nabla T_{crit}$). One way to increase Γ is to shorten the prime-mover stack but keep the temperature difference across the stack constant, thus increasing ∇T. Unfortunately, this type of change increases the heat loss due to conduction down the stack.

Another way to increase Γ is to lower ∇T_{crit} by moving the prime-mover stack away from the pressure antinode at the end of the tube. Because of the intervening prime-mover stack, the refrigerator stack is *already* much farther from the pressure antinode than in the cryocooler, and additional movement of the prime-mover

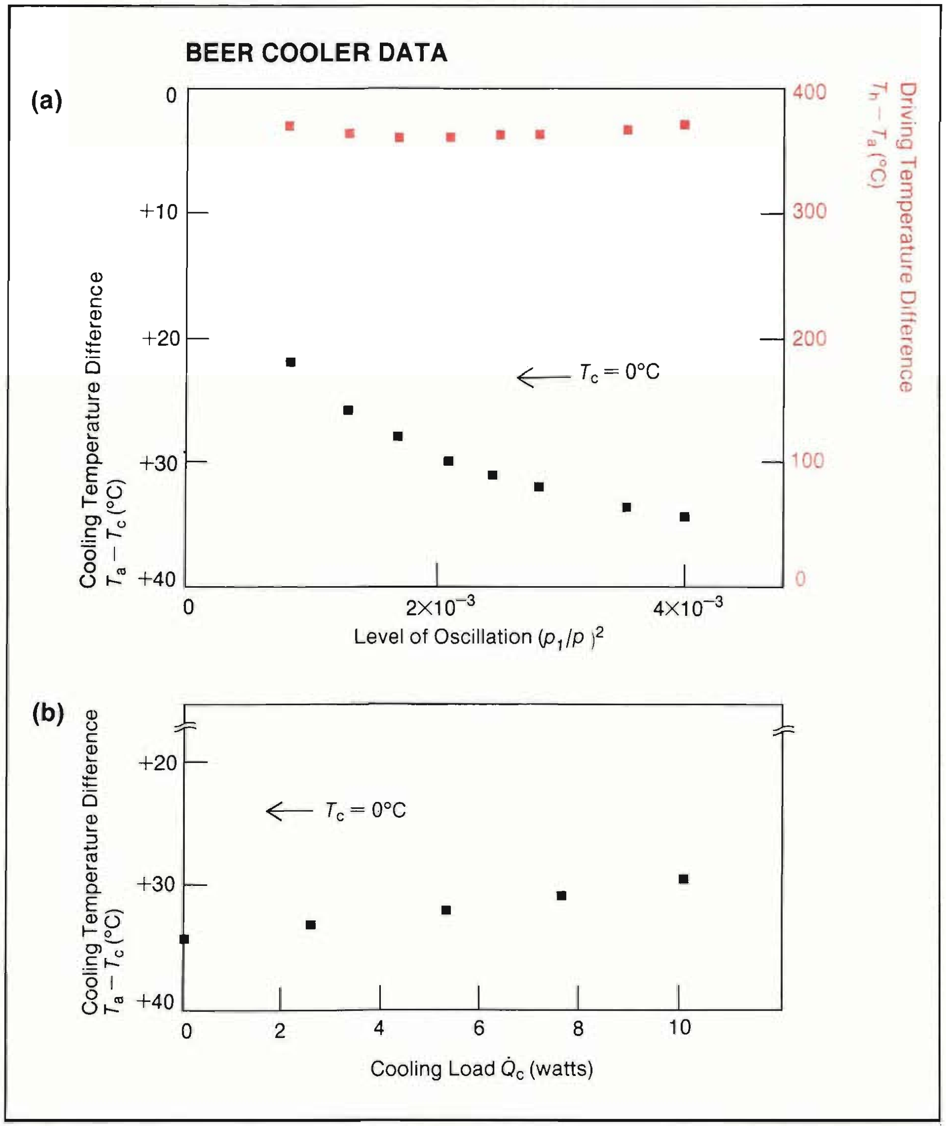


Fig. 17. For these measurements on the beer cooler, the refrigerator stack and the resonator were located in an evacuated space for thermal insulation. Part (a) shows how the cooling temperature difference (black) across the refrigerator stack and the driving temperature difference (red) across the prime-mover stack vary with the level of oscillation (here given by the square of the relative dynamic pressure amplitude) when no external refrigerator load is placed on the cooler. The level of oscillation is determined by the rate at which heat is supplied and removed across the prime-mover stack, but, as can be seen, this results in little change in the driving temperature difference. At the same time, the cold temperature drops gradually below the freezing point of water. (b) For a level of oscillation of about 4 × 10^{-3} (that is, p_1 is about 0.19 bar), we see that the beer cooler can handle a small heat load of 10 watts and still remain below freezing. ◄

stack pushes the refrigerator stack even more from its optimum position. Once again, changing an idea into a practical heat engine entails a set of compromises.

A schematic of an operational heat-driven cooler in which the prime-mover stack is between the end of the tube and the refrigerator stack is shown in Fig. 16. Because of the above considerations—especially those related to ∇T_{crit}—the refrigerator stack in this device cannot be expected to cool much below ambient temperature. Although unable to produce cryogenic temperatures, the cooler ought to be able to produce temperatures low enough to cool a can of beer. For this reason we have affectionately dubbed the engine the "beer cooler."

As in the case of the cryocooler, the rather complex design was carried out numerically, and many of the features important to the cryocooler apply to the beer cooler. For example, the resonator is similar to the resonator in the cryocooler, and the driver is on the "hot" side of the refrigerator stack, thus reducing viscous and acoustic-streaming losses.

In our model engine the plate material in the stacks (stainless steel) and the spacing of the plates were dictated by ease of fabrication as much as by anything else. The refrigerator assembly was placed in an evacuated container but not otherwise thermally insulated. The working fluid was helium, whose pressure was chosen experimentally to minimize the cold temperature. A major problem, yet to be satisfactorily solved, was efficient exchange of heat at the heat exchangers—especially the hot exchanger made of nickel.

In spite of the engine's compromises, it still sings along, performing rather well (Fig. 17). As the heat supplied to the hot end of the prime mover is increased, the level of oscillation increases—the largest peak-to-peak dynamic pressure amplitude measured at the ambient exchangers exceeding a tenth of the mean pressure. In agreement with our understanding, the temperature drop across the prime-mover stack does not change much as the dynamic pressure amplitude increases; the small changes seen in the data result from the diffusive flow of heat across the gaps of gas and through the heat exchangers from the heat source to the ambient heat exchanger. Figure 17b shows that the beer cooler can manage a 10-watt cooling load while keeping T_c 5 centigrade degrees below the freezing point of water—a rather encouraging result for the first laboratory model.

A number of issues concerning the practical use of this engine concept and of the cryocooler remain to be resolved. It is likely that the most important is the matter of heat exchange. This problem, as we've mentioned, has always been a key one in the development of heat engines—classical or otherwise.

The Liquid Sodium Acoustic Engine. As man moves from Earth into space, so does his need for reliable power. However, differences in the requirements and in the operating environment in space may prompt radical changes in the engines that provide such power. An idea stimulated by such differences is the liquid sodium acoustic engine, which not only is a natural, rather than a conventional, engine but uses a liquid instead of a gas as its working fluid.

The concept of using a liquid can be traced to a 1931 paper by J. F. J. Malone in which he pointed out that certain liquids have important thermodynamic qualities that make them suitable for use in heat engines. Although concerned about its chemical reactivity, Malone knew that liquid sodium was one of these "good liquids," but materials technology was then inadequate for him to consider its use.

Today's materials technology suggests revival of these ideas, and we had been working on the liquid propylene Stirling engine (see "The Liquid Propylene Engine") as a modern example of an innovative but more conventional engine that uses liquids. Thus, when we learned from then Associate Director Kaye Lathrop of the need in space for a reliable, moderately efficient electrical generator, it was not difficult for us to propose a natural acoustic engine based on liquid sodium. Especially ideal for this application is the high "cold" temperature (at least 400 kelvins) of the liquid sodium engine. This fact is important because the cold sink for any heat engine in space must ultimately be a black-body radiator whose size would be proportional to T_c^{-4}

Liquid sodium has many potential advantages as a working substance in a natural engine. The heat and work parameters are acceptably large. For example, at 700°C, which is roughly in the middle of the temperature range of a possible high-power engine, liquid sodium has a very high expansion coefficient and a large ratio of specific heats so that $T\beta = 0.28$ and $\gamma - 1 = 0.43$ (compared to a monatomic gas such as helium, for which $T\beta = 1$ and $\gamma - 1 = 2/3$).

For a given Mach number,* the power density in the stack is proportional to ρa^3, where ρ is the density of the working fluid and a is the speed of sound. The density of liquid sodium is about 500 times greater than that of helium at the pressures used in our gas acoustic engines, and the speed of sound is more than a factor of 2 greater. Thus, the power density for a liquid sodium acoustic engine should be more than 10^3 times greater than for a helium acoustic engine, a definite advantage.

This dramatic increase is not without its drawbacks, however. The heat capacity per unit area for the sodium within a thermal penetration depth of the second

**The Mach number is the ratio of the fluid speed to the speed of sound in the fluid.*

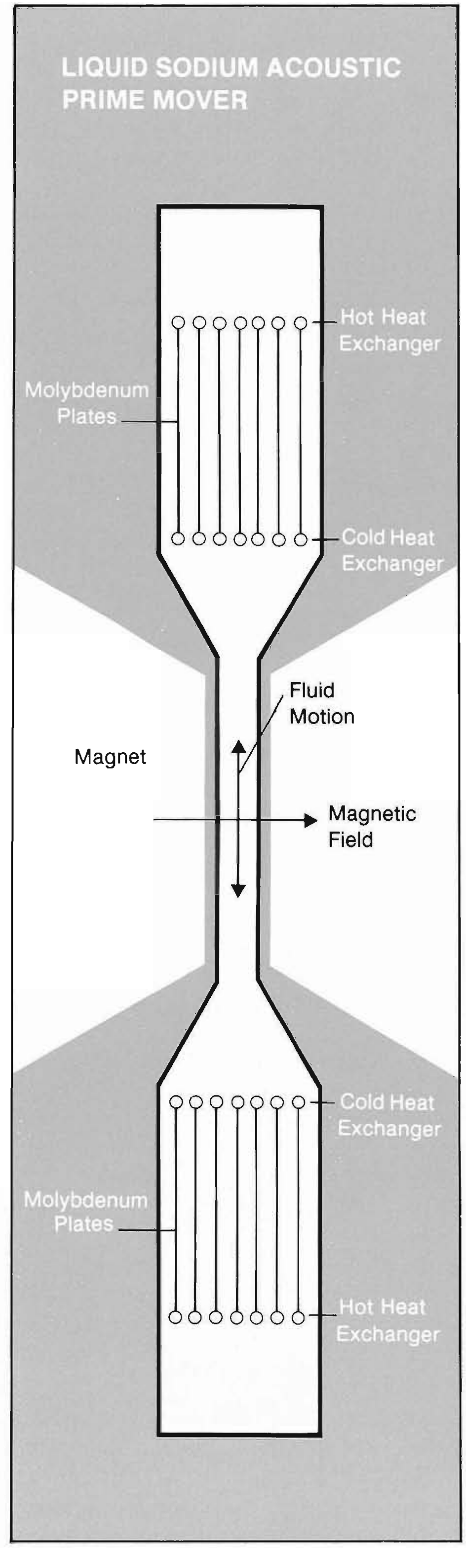


Fig. 18. The temperature drop applied across both stacks of molybdenum plates causes the liquid sodium in this proposed engine to oscillate back and forth between the poles of the magnet. A magnetohydrodynamic effect is used to convert acoustic to electric energy. ▶

Chris Espinoza welding heat exchange manifolds onto the resonator tube of the liquid sodium natural heat engine.

medium is so large that the usual assumption of infinite heat capacity of the second medium is not valid. As a consequence, the power density drops. Moreover, the acoustic impedance ρa of the sodium is relatively high—roughly equal to that of solids—which means that in a sodium engine motion of the stack and container can be expected to send heavy vibrations throughout the entire engine (unlike the beer cooler, for example, in which a peak-to-peak dynamic pressure oscillation of 10 per cent of the mean pressure produces only a pleasantly audible tone in the room). To counter this effect, stiff, high-density materials like molybdenum or tungsten need to be used in the stack, and the walls of the resonator need to be made of heavy stainless steel. Even with such strong walls, high Mach numbers cannot be achieved because the high acoustic pressures would burst the resonator.

Liquid sodium has other very desirable features. For example, its Prandtl number, which can be thought of as the square of the ratio of the viscous penetration depth to the thermal penetration depth, is extremely low (about 0.004 for sodium at 700°C compared to 0.667 for helium gas). The reason for such a low Prandtl number is that liquid sodium is a metal. As a result, its kinematic viscosity is rather normal for a liquid, but, owing to electronic contributions to the conduction of heat, its thermal diffusivity is high. The consequences are important. In helium, viscous shear extends into the gas from a boundary about as far as the temperature gradients that drive the flow of heat. This shear drains energy, decreasing efficiency and making it difficult for a gaseous heat engine to work. The low Prandtl number of liquid sodium means that heat can be transported between working fluid and the plates for a volume fifteen times larger than the volume being affected by viscosity, and viscous losses are correspondingly small. Again, however, a price must be paid: diffusive heat conduction in the sodium down the stack increases.

The fact that liquid sodium is a metal has yet another important consequence. Electrical current can be generated from the sound via magnetohydrodynamic coupling. Such coupling means electric power can be produced from heat without using moving parts (ignoring the non-negligible motion of the vessel containing the sodium!). This feature, of course, is one of the main reasons for the expected reliability of the engine. Figure 18 is a schematic of a possible liquid sodium prime mover that uses a half-wavelength resonant tube, two driving stacks (one on each side of the magnet), and magnetohydrodynamic power coupling.

To design a model liquid sodium engine, we constructed a thermoacoustic theory for liquids and then evaluated it numerically. The calculated characteristics of a reasonably designed engine are given in Table 1. Note that the dynamic pressure almost equals the mean pressure of the sodium and that efficiency is calculated to be about 18 per cent (31 per cent of the Carnot efficiency).

A complete engine has not yet been built, but work (supported by the Division of Advanced Energy Projects in DOE/BES) has been done separately on the magnetohydrodynamics and the thermoacoustics. In both cases preliminary results are encouraging, though technical problems remain.

First, a magnetohydrodynamic converter was built that consisted essentially of a liquid sodium acoustic resonator with a central rectangular channel for guiding the sodium in the transverse direction between the poles of a magnet. Electrodes for picking up the electric current were attached to the channel. The device was tested by exciting an acoustic standing

Table 1

Characteristics of a reasonably designed liquid sodium prime mover.

Frequency	1000 Hz
Hot temperature	1000 K
Cold temperature	400 K
Mean pressure	200 bars
Dynamic pressure	198 bars
Plate spacing	0.0373 cm
Plate thickness	0.0280 cm
Distance of hot end from tube end	8.65 cm
Length of stack	8.0 cm
Average $\dot{Q}_h$	300 W/cm^2
Average $\dot{W}$	55.1 W/cm^2
η	0.184
η/η_{Carnot}	0.307

wave (by temporarily putting electric power *into* the magnetohydrodynamic converter!) and then letting the energy stored in the acoustic resonance flow through the converter into a resistive load across the electrodes. The efficiency—defined as the ratio of the measured electric energy delivered to the load to the calculated stored acoustic energy—is already quite high in this first prototype (Fig. 19) and a number of improvements are possible. From a technological point of view, it is very significant that the maximum efficiency is still reasonable in a magnetic field of only 0.9 tesla, suggesting that a permanent magnet is appropriate with a consequent simplification and decrease in weight.

The thermoacoustic prime mover tested had a single stack of molybdenum plates (Fig. 20) inside a straight half-wavelength tube. For this test the cold heat exchanger was filled with pressurized water at 125°C and the hot heat exchanger with heated sodium at various temperatures ranging from 440°C to 645°C. Although the test was preliminary, it was successful. The application of various temperature drops across the stack resulted in the data of Fig. 21 and, above a 350°C drop, in an obvious acoustic vibration of the entire assembly.

We obtained the rate of heat supplied to the engine $\dot{Q}_h$ by monitoring the flow rate and the inlet and outlet temperatures of the sodium flowing through the hot heat exchanger. For a low temperature drop (ΔT) across the stack, the heat flow through the engine is due solely to the simple conduction of heat by the sodium, molybdenum, and stainless steel. However, for a ΔT of around 400°C, $\dot{Q}_h$ begins to increase dramatically above the value for simple conduction. This result agrees with the fact that acoustic oscillations at 906 hertz (Hz) were first detected at a ΔT of 350°C. By the time ΔT had reached 520°C, the resonator was oscillating at high enough amplitude that the sound in the room was unpleasantly loud and the apparatus was vibrating strongly.

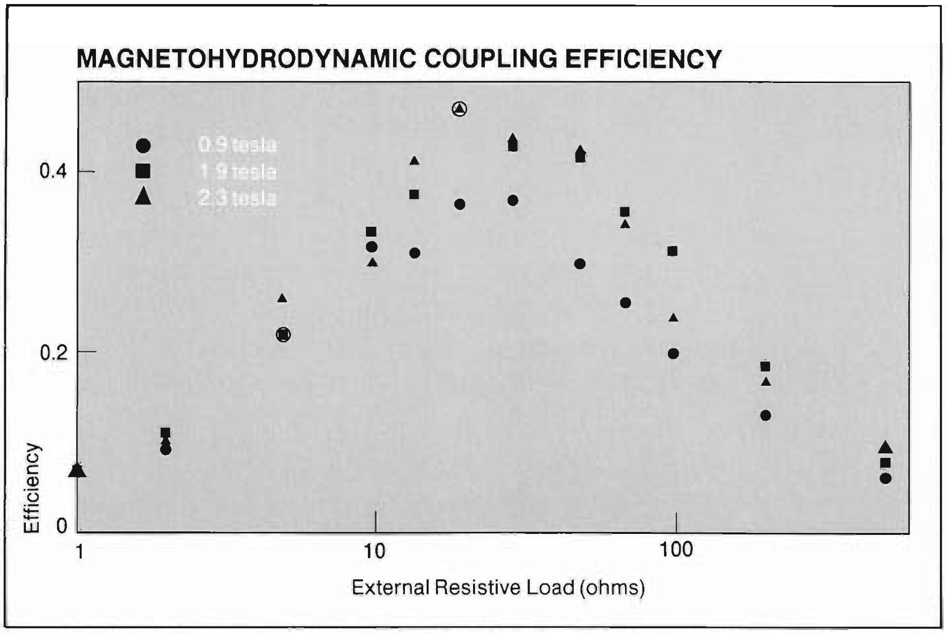
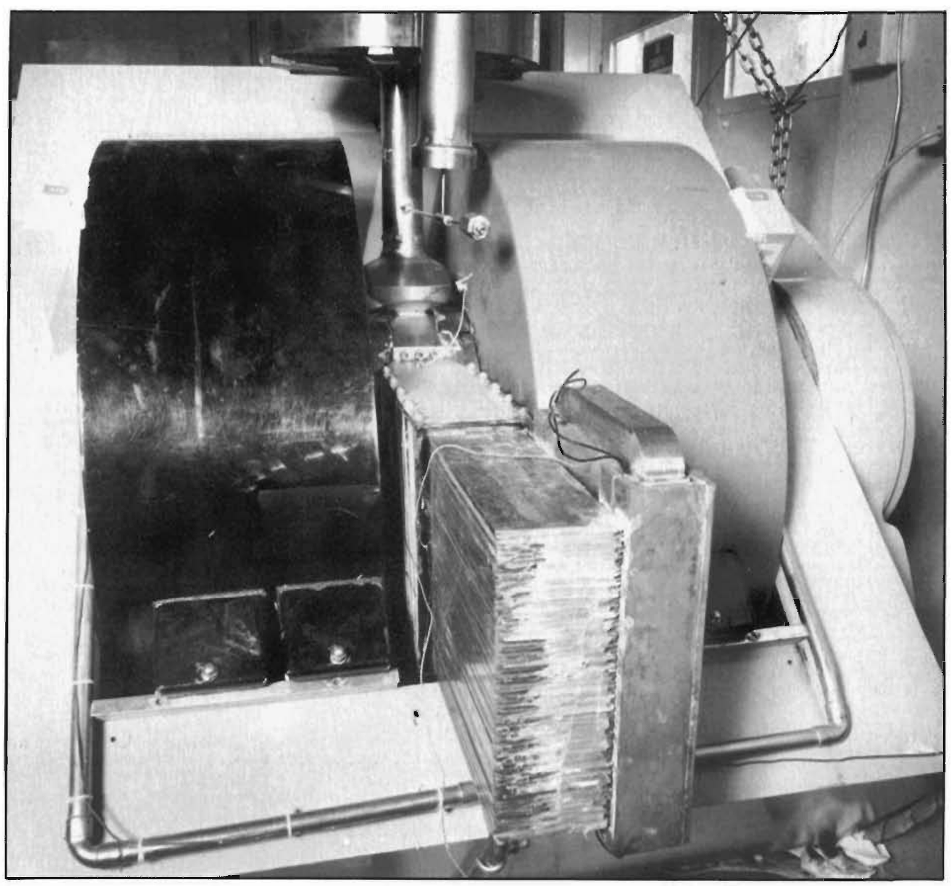


Fig. 19. These initial data demonstrate the efficiency with which acoustic energy in liquid sodium was converted to electric energy via magnetohydrodynamic coupling as a function of the resistance of an external load and for three different magnetic field strengths. In the apparatus used to obtain this data, the central rectangular channel holding the liquid sodium is 1.2 cm thick in the direction of the magnetic field, 7.6 cm thick in the direction of electric current flow, and 31 cm long; however, only 20 cm of that length is actually in contact with the electrodes. The central channel is part of a 1-m-long acoustic resonator filled with liquid sodium at a temperature of 130°C. ▲



The magnetohydrodynamic converter used to test power coupling for the liquid sodium heat engine.

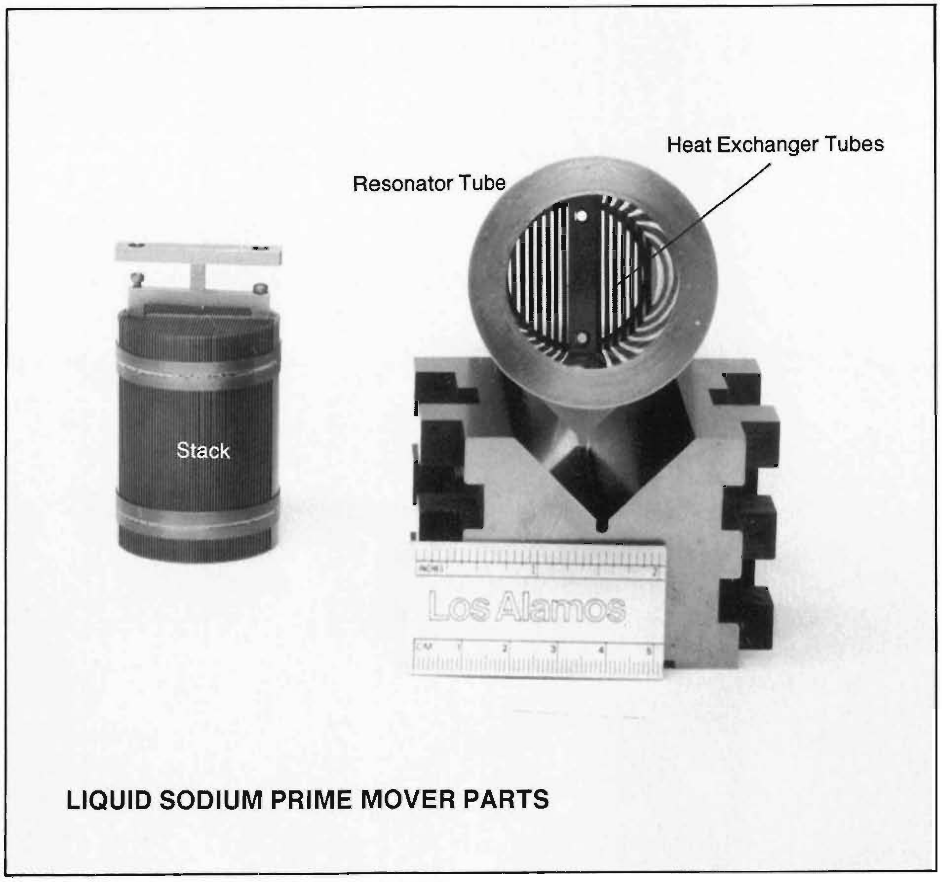


Fig. 20. Our first operating liquid sodium prime mover has a single stack of molybdenum plates (left) that were fabricated at Los Alamos from a solid rod using electric discharge machining. Plate thickness is 0.3 mm; spacing between plates is 0.38 mm; the length of the stack is only 5.2 cm so that the engine would oscillate at a reasonably low ΔT. The transverse tubes of the two heat exchangers (one set can be seen at the end of the cylindrical section of the resonator tube on the right) were made from stainless steel hypodermic needles. Hot liquid sodium at various temperatures (T_h) is circulated through one heat exchanger and pressurized hot water ($T_c = 125°C$) through the other. The stack just fits inside the half-wavelength resonator tube, which has a length of 106 cm. The plates are positioned in the tube at about $x = \lambda/14$ from the end. The acoustic resonant frequency is 906 Hz.

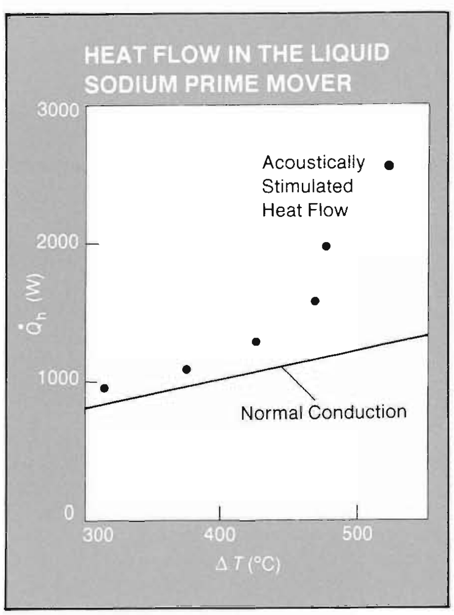


Fig. 21. The data shown here (dots) represent heat flow $\dot{Q}_h$ into the liquid sodium prime mover at the hot heat exchanger as a function of the temperature drop ΔT across the stack, whereas the solid curve represents the calculated flow of heat due to conduction with no thermoacoustic effects. At low ΔT heat flow is due to normal conduction across the stack. At high ΔT, however, the sharp rise in $\dot{Q}_h$ is indicative of acoustically stimulated heat flow.◀

There are some disagreements between the experimental results and our theoretical calculations. A calculation for the particular geometry and acoustic frequency of the device predicts that it should begin to oscillate at a ΔT of 260°C rather than 350°C. We also expected a maximum $\dot{Q}_h$ of 2000 watts at a ΔT of 520°C, whereas the measured value was 2600 watts. We do not yet understand these quantitative disagreements but are extremely encouraged by the initial success of the engine.

Molecular Natural Engines. Heat engines of any sort transform energy between the random thermal motion of atoms and the coherent motion needed for useful work. The concepts of heat and temperature—implicit to the understanding of heat engines—are statistical in nature. Hence, for these variables to be well defined, a system must have large numbers of atoms. But what is the smallest system that will still allow us to apply these concepts?

If we take the error in statistical quantities in thermodynamics to be approximately the reciprocal of the square root of the number of degrees of freedom and if we assume for a heat engine that errors of a few per cent are tolerable, then only several hundred to a few thousand atoms are sufficient. Rather nice systems of this mesoscale size, consisting of large organic molecules, are common. Furthermore, such systems behave much like a purely classical collection of masses and springs.

Within such systems, nonlinear intermolecular potentials give rise to phenomena directly related to the thermal expansion needed for heat engines. Intramolecular and intermolecular interactions provide connections between regions of vibrational energy that, if large enough, can be considered to be heat reservoirs. In other words, all the ingredients for an engine are present.

Do such engines then exist? And, if so, do they serve a useful function in nature, say perhaps as tiny engines in biochemical systems? Will the concepts of natural engines apply not only to these small sizes but also to the high frequencies associated with molecular vibrations?

Heat pumping might occur in mesoscale systems if coherent vibratory motion

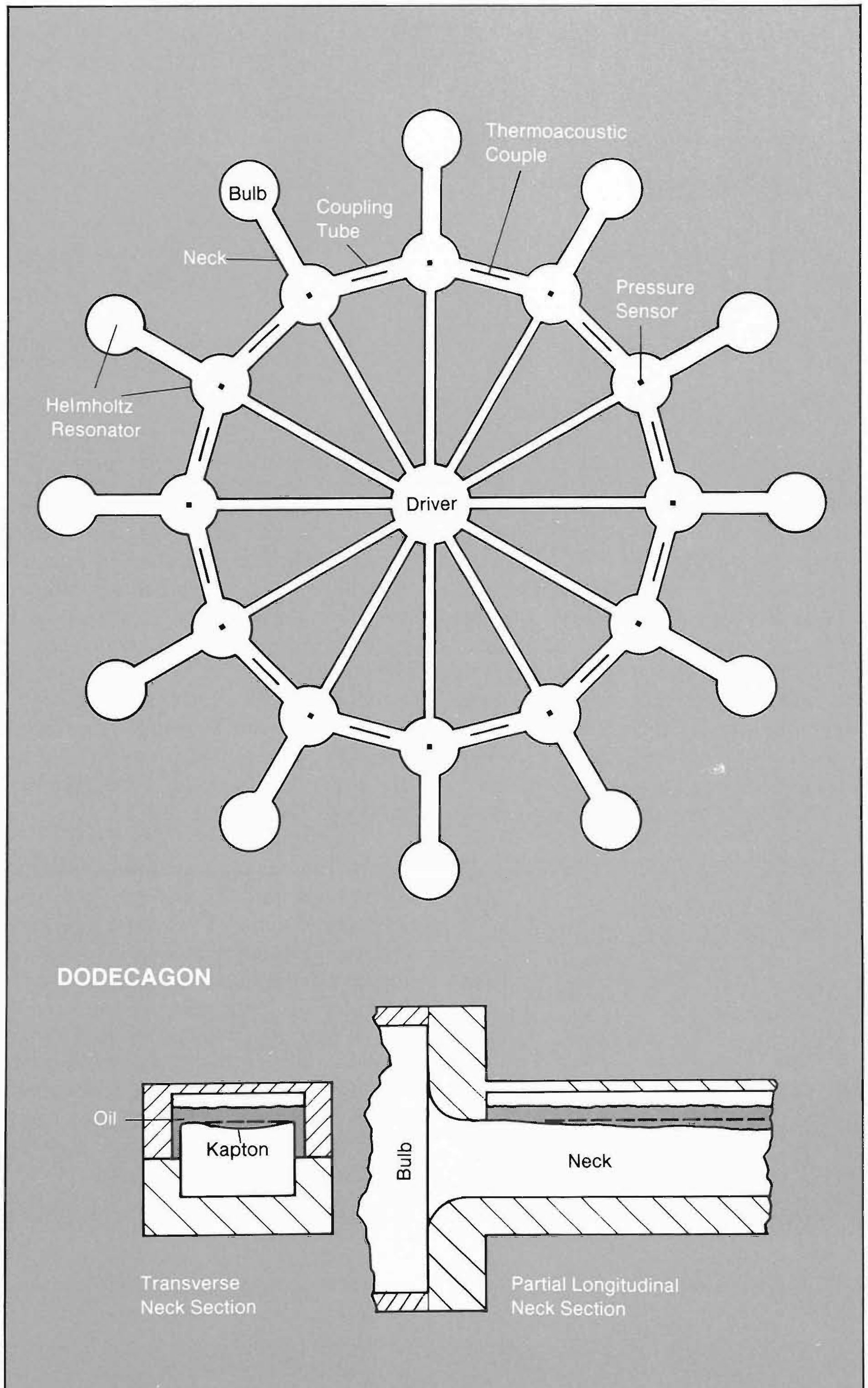


Fig. 22. Localization of acoustic energy was studied by coupling twelve nonlinear Helmholtz resonators together in a ring and measuring the direction of heat flow with the thermoacoustic couples positioned in the coupling tubes and measuring the level of vibration in individual resonators with the pressure sensors. The construction of the neck of each resonator (see details below the dodecagon) introduces a nonlinearity because vibrating gas that rushes through the neck causes the Kapton to flex, altering the resonant frequency of that resonator. The entire system is driven by a loudspeaker at the center. ▶

can first be established and then survive long enough to have a significant effect. Also, if the concepts of temperature and temperature gradients are to be useful, then the mean free paths of the heat-carrying excitation should be small compared to the classical thermal penetration depth and to the size of the mesoscale object. Using an angular frequency of 10^{11} Hz, we estimate the penetration depth to be about 14 angstroms. Hence, mesoscale objects perhaps 50 to 100 angstroms in size and vibrating at frequencies of order 10^{10} Hz might be large enough and slow enough to be natural engines, providing their level of coherent excitation is high enough.

Rather than building an object of such small size, the same effects may be realized in a natural way via a concept from nonlinear science—the solitary wave. An acoustic heat engine with its stack of plates centered at a pressure antinode will pump heat from both ends of the stack toward the middle. If we alter this idea by using a continuous stack that has, owing to dispersive and nonlinear effects, a *localized*, or *solitary*, vibrational disturbance in the longitudinal direction, then heat is pumped from the wings to the center of the disturbance. Because the stack is continuous, the thermodynamic symmetry is not broken geometrically; rather it is broken dynamically. We call such a device a *nonlinear natural engine*. In principle, such a localized disturbance could be a vibrational excitation of a mesoscale object.

Localized waves in lower-dimensional vibrational systems have received a great deal of theoretical attention because of their potential application to biological processes. However, macroscopic modeling experiments in a water wave trough at the University of California, Los Angeles, have been very valuable in developing insight about solitary waves. (An outstanding example is the Wu-ton, a nonpropagating soliton in water surface waves.) As a result, we decided to build an acoustical model that might give insight into how a coherently vibrating molecular system might behave. If such objects are indeed found to be real, we believe the field of potential applications will be much broader than just lower-dimensional systems.

Our apparatus, which we call the dodecagon, has been likened to a 12-element benzene ring. It consists of a circle of twelve coupled acoustic Helmholtz resonators with a nonlinear element included within each resonator (Fig. 22). We introduce the nonlinearity by building the resonator from two bulbs connected by a neck with a thin Kapton plastic film that flexes with changes in pressure. To prevent the neck from flapping at the acoustic frequency or its harmonics, we loaded the plastic film with oil.

According to one mathematical analysis, localization of energy can occur if the resonant frequency of any given resonator decreases as the amplitude increases. In our resonators, as the dynamic pressure of the acoustic wave increases, the velocity of fluid through the neck increases, which means, from Bernoulli's principle, that the average pressure there decreases. The Kapton neck then flexes inward, reducing the cross-sectional area and, thus, the resonant frequency (which is proportional to the square root of the area).

We installed a thermoacoustic couple in each tube linking resonators to measure the direction of heat flow and also put a dynamic pressure sensor in each resonator to measure its level of vibration. The whole system was driven symmetrically from the center by an acoustic driver. When we drove the system at a frequency less than the low-amplitude resonance frequency, localization of energy *did* occur above a certain threshold amplitude. Further, heat was pumped toward the region of high amplitude. But the localization was stronger and occurred at a much lower drive amplitude than expected. This localization was also attended by a low-frequency modulation—typically at 1/11 or 1/12 of the drive frequency but often with components a factor of 100 or more times lower than the drive.

What happened? Our resonators performed as expected so far as alteration of the resonant frequency was concerned. However, we had unwittingly introduced a second set of vibrational systems into the experiment: plate-like vibrations on the Kapton-oil system. We believe that under suitable conditions the driver resonantly excites the Kapton-oil system and induces the film to make a hysteretic transition to a different geometry that facilitates the localization.

Our acoustical model experiments have been helpful in inspiring thought on molecular-scale or mesoscale systems. Localization of energy and heat pumping did occur. More important, though, attending and preceding the localization, we observed behavior that changed on an entirely different time scale than the acoustic phenomenon.

We conjecture that in molecular and mesoscale systems it is important to have two or more interacting, or coupled, "fields." These coupled fields could be some of the normal optical vibrational modes of a molecular system. In particular, torsional or librational modes of motion are almost certainly coupled nonlinearly with the longitudinal modes of motion. We expect a time-dependent conformational change, say in the longitudinal field, to attend localization of vibrational energy. The fundamental molecular vibrational frequencies are of the order of 10^{12} Hz or greater, but a time-dependent conformational change in the longitudinal field could be at a much lower frequency—possibly low enough to create natural engine effects. These conjectures have motivated us to begin experimental work with a number of other collaborators on the general question of the localization of vibrational energy in materials. This is a case where we think we know what we are looking for, but we don't know what we will find.

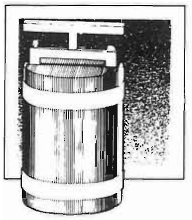
Thus, whether shaking loudly in the laboratory or, perhaps, vibrating soundlessly in a molecule, natural engines may be a widespread phenomenon of general importance. Not only are natural engines simple, they use a necessary thermodynamic evil—irreversibilities—as a positive feature of the engine. We hope an understanding of these concepts will serve mankind well in his quest for appropriate engines and will help us to comprehend better the behavior of molecular vibrational systems. ■

Further Reading

John Wheatley, T. Hofler, G. W. Swift, and A. Migliori. 1985. Understanding some simple phenomena in thermoacoustics with applications to acoustical heat engines. *American Journal of Physics* 53:147.

J. C. Wheatley, T. Hofler, G. W. Swift, and A. Migliori. 1983. An intrinsically irreversible thermoacoustic heat engine. *Journal of the Acoustic Society of America* 74:153.

G. W. Swift, A. Migliori, T. Hofler, and John Wheatley. 1985. Theory and calculations for an intrinsically irreversible acoustic prime mover using liquid sodium as primary working fluid. *Journal of the Acoustic Society of America* 78:767.



The Liquid Propylene Engine

An ideal use of geothermal energy is to warm buildings by extracting heat from ground water at temperatures of only about 10°C. This application involves the pumping of large amounts of heat across small temperature differences (of the order of 30°C). An efficient way to effect such heat transfer is from one liquid to another. As a result, a heat pump that appears well suited for this purpose is a conventional reciprocating heat engine using a *liquid* for a working substance.

We have been studying just such an engine—a Stirling engine that uses liquid propylene as its working fluid. Our discussion of this device will both contrast the simplicity of natural engines with the complexity of more traditional engines and, more important, will introduce the use of a liquid as a thermodynamic working substance. (The section in the main article called "The Liquid Sodium Acoustic Engine" discusses a *natural* heat engine that uses a liquid as its primary thermodynamic medium.)

It is a common misconception that liquids behave much like an idealized hydraulic fluid, with density independent of temperature and pressure. In fact, especially near the critical point (where the liquid and gaseous phases become indistinguishable), a typical real liquid is somewhat compressible, has a large thermal expansion coefficient (comparable to or larger than that of an ideal gas!), and has other attractive thermophysical properties. These facts were first appreciated by John Malone, who in the 1920s built several Stirling prime movers that used liquid water with pressures as high as 700 bars as the working substance. We chose liquid propylene (C_3H_6) for our work because its critical temperature is just above room temperature and its Prandtl number (which can be thought of as a measure of the material's viscous losses in relation to its thermal transport capacity) is lower than that of other fluids with similar critical temperatures.

A major advantage of a liquid working substance is that liquids have a very large heat capacity per unit volume compared to gases, making it possible to build efficient and compact heat exchangers and regenerators. This point is illustrated by the compact propylene-to-water heat exchanger we have developed for our engine (Fig. 1). The exchanger is made of hun-

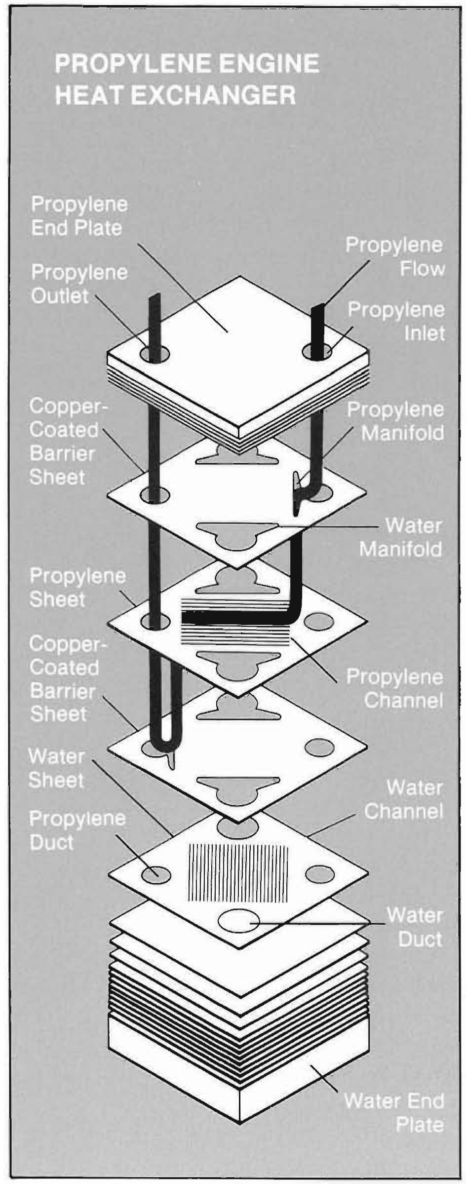


Fig. 1. In this propylene-to-water heat exchanger, made up of a stack of hundreds of stainless steel sheets copper-brazed together at Los Alamos, the propylene flows in at the top right of the stack and across through the propylene manifolds and channels, then moves up and out through the other propylene duct. The arrow in the figure traces the path through just *one* of the sets of channels and manifolds; similar flow occurs through the other, lower propylene channels and manifolds. At the same time, water flows in and up through one water duct and across the stack (but through alternate sets of plates and across the plates in a direction perpendicular to the corresponding propylene flow) until it returns, exiting through the other water duct. Because of the intimate thermal contact between fluid and stainless steel, heat can be transferred at a rate of 230 *W*/°C. ▶

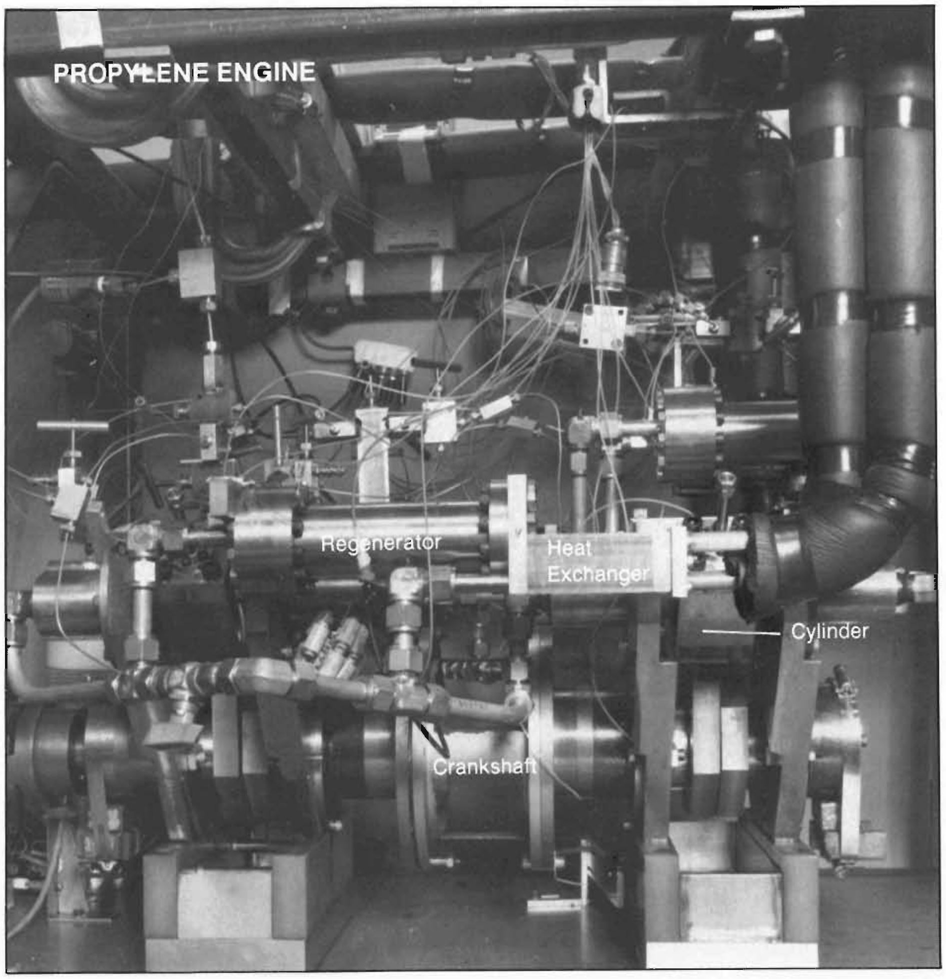


Fig. 2. The heat engine shown here consists of four Stirling engines of the Rider form operating from a common crankshaft but phased 90 degrees apart. The working medium is liquid propylene, and heat exchange between water and the propylene takes place in the stainless-steel exchangers depicted in Fig. 1. ◀

dreds of chemically milled stainless-steel sheets copper brazed together (several of the individual plates are shown on the cover). Although the exchanger (4 by 4 by 9 centimeters in size) entrains only a few cubic centimeters of propylene, it transfers heat between the two fluid streams at a rate of 230 watts per °C with only a few watts of power required to pump the fluids through the exchanger.

Another advantage of a liquid working substance is that liquids are typically much less compressible than gases. Thus the large pressure amplitudes needed to pump large amounts of heat can be achieved with only small displacements of a piston, even for a substantial volume of entrained liquid in the thermal elements. Because of this quality, it is possible to build a high-power engine that uses a short stroke, making the mechanical elements very efficient without compromising on the size and efficiency of the thermal elements.

Our laboratory-scale liquid-propylene Stirling engine (Fig. 2) uses the same configuration of parts shown in Fig. 3 of the main article (the Rider form of the Stirling engine), except that we have *four* such assemblies. These assemblies operate from a common crankshaft and are mechanically phased 90 degrees apart so that the shaft torque oscillations are minimized, eliminating the need for a big flywheel. Although much of the wiring in Fig. 2 is for diagnostic purposes, the photograph, when contrasted with photographs of natural engines (see the main article) is nevertheless a dramatic representation of the complexity of a more conventional reciprocating engine.

In its heat-pump mode, our engine uses work supplied by an electric motor to transfer heat from a source at or below room temperature to a heat sink consisting of flowing water at or above room temperature. For convenient measurement, the low-temperature source is an electric heater. Mean pressure, oscillating pressure amplitude, volumetric displacement, shaft rotation frequency f, and hot and cold temperatures are all independently controllable. We can measure both the rate at which heat is pumped away from the heat source $\dot{Q}$ and the shaft torque τ, the latter giving us shaft power $\dot{W} = 2\pi f \tau$.

In addition, our laboratory engine has valves that quickly change it from the ordinary heat-pump configuration to one in which there is no flow of propylene through the regenerators and heat exchangers, even though crankshaft and piston motion, pressure amplitudes, temperatures, and so forth remain the same. This feature allows us to accurately measure just the torque *difference* $\Delta\tau$ required to pump the heat, with the background torques due to bearing and seal friction, piston blowby, and the like eliminated.

Large amounts of heat can be pumped by the engine (Fig. 3a)—around 1300 watts at a crankshaft rotation frequency of 4.5 Hz—and the data points match very well curves predicted from theory for the particular geometry of the engine and for the use of propylene as the working fluid.

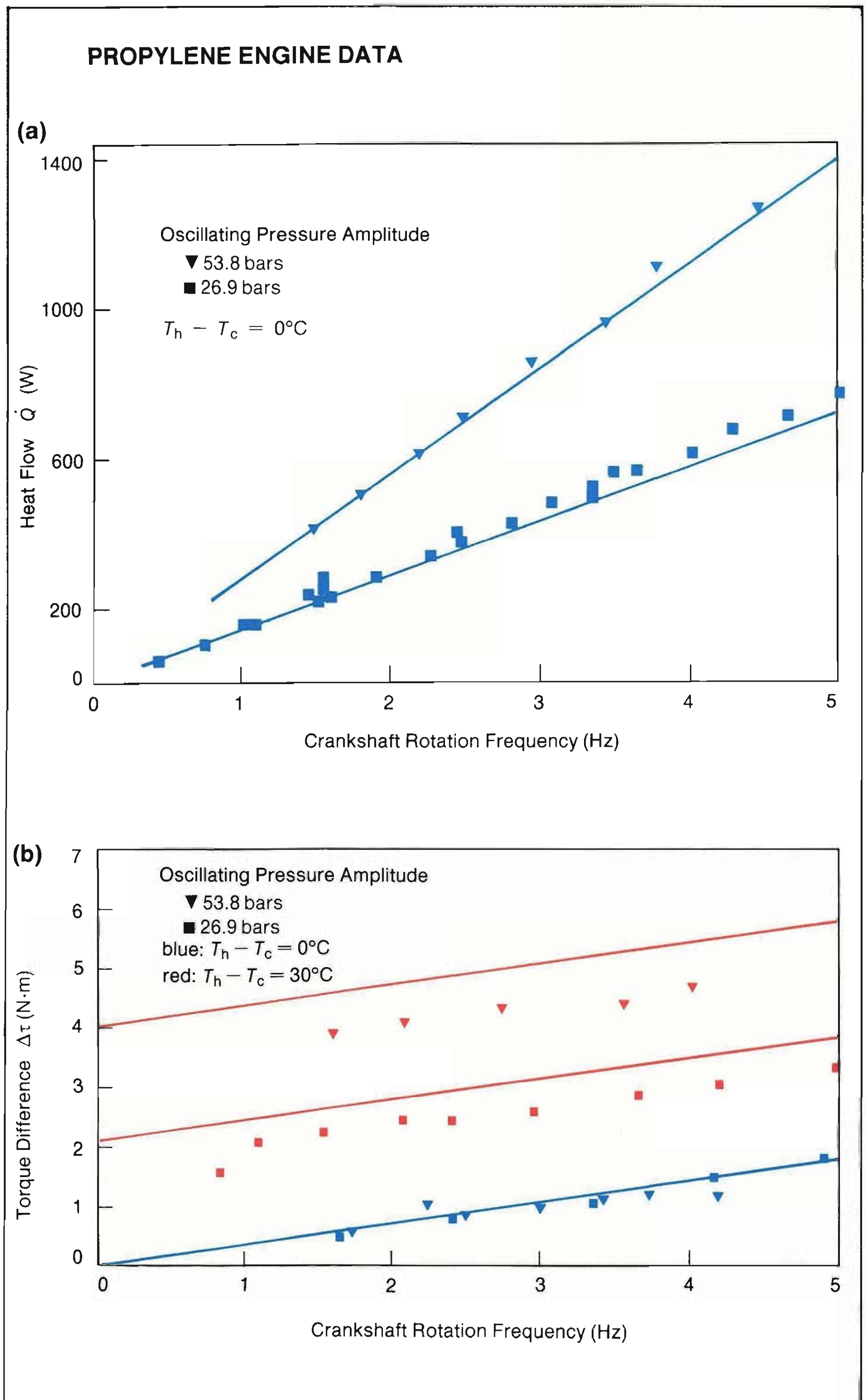


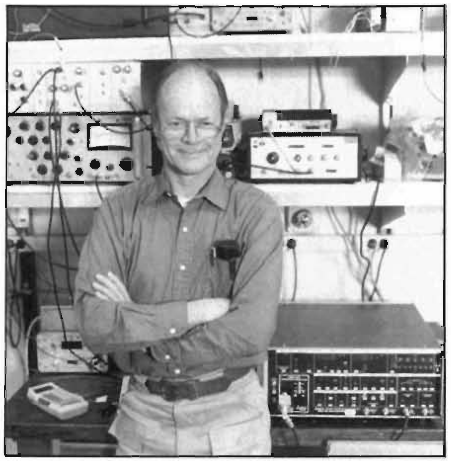
Fig. 3. (a) The rate at which the propylene engine pumps heat $\dot{Q}$ as a function of crankshaft rotation frequency f at two different oscillating pressure amplitudes agrees very well with theoretical curves predicted from the physical properties of propylene and the geometry of the engine. (b) The torque difference $\Delta\tau$, here also plotted as a function of f, is just that part of the torque needed to pump the heat. In both graphs the blue data points represent no temperature difference across the regenerators, whereas the red data points represent a 30°C difference. ◀

The lines drawn on Fig. 3b represent the torque required by an engine with the Carnot efficiency to pump the observed amount of heat added to the torque associated with just the viscous losses of pushing the fluid through the regenerators and heat exchangers. Our measured torque differences agree well with these theoretical curves.

Our laboratory engine is very far from a practical, economically useful device. Its scale and most of its design are appropriate for experimental measurements and for the understanding of principles, not for optimized efficiency or low manufacturing or operating costs in a specific application. But, as expected, we are learning that liquids *are* good heat engine working substances. Liquid engines may ultimately be of great technological importance.

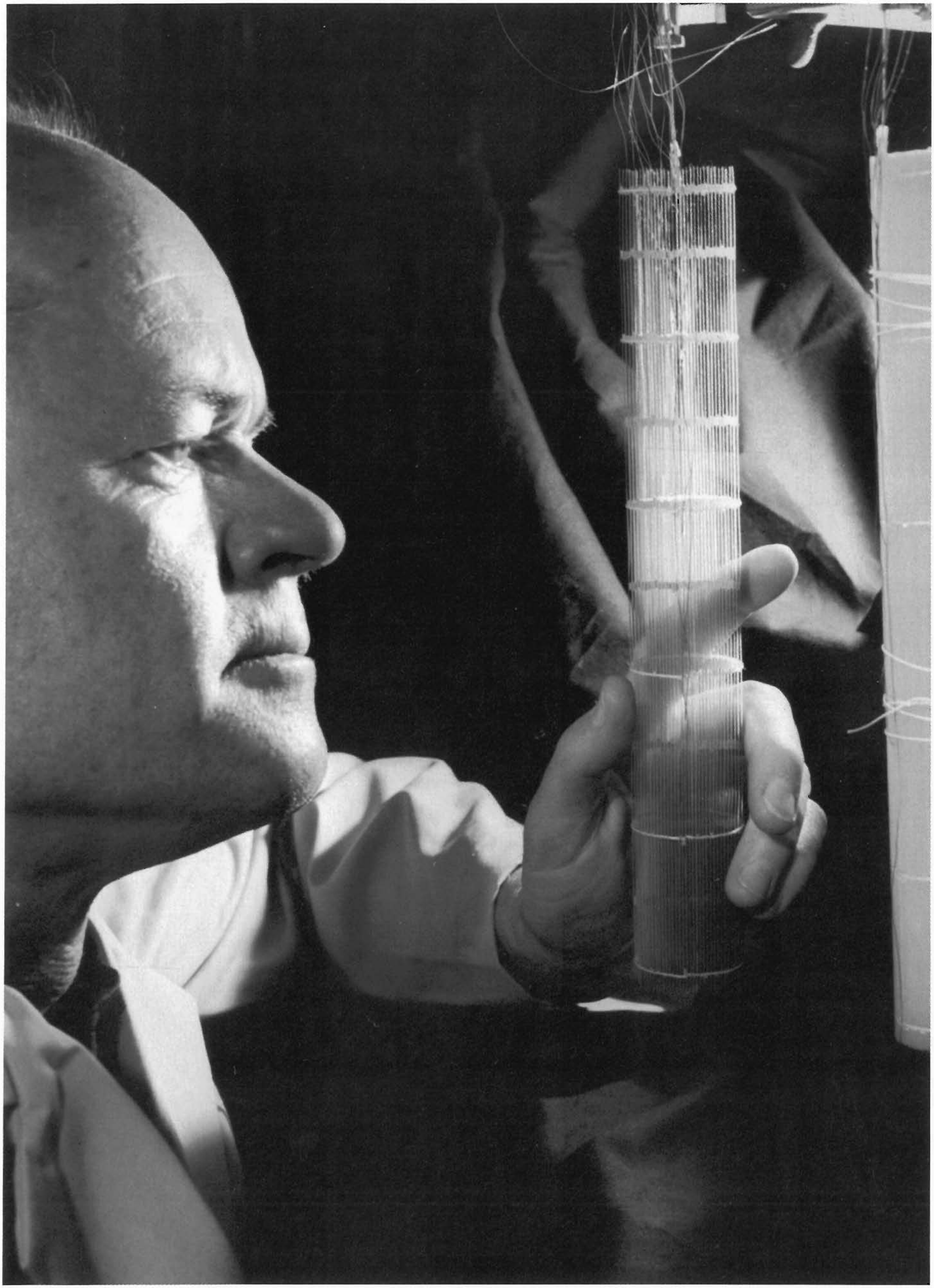
We are also learning much about the practical details of the use of liquids in engines. For example, we suspect that the next logical step in the development of practical liquid engines is to abandon the reciprocating Stirling engine entirely. Instead, we would use the liquid in, say, a Brayton engine with rotary compressors and expanders. Such a configuration would reduce losses from such things as bearing and seal friction that, until now, we have regarded as quite uninteresting. ■

John C. Wheatley (1927-1986) joined Los Alamos in 1981. During his tenure here, he performed experiments on novel heat engines and on the fundamentals of thermal and statistical physics. He received his B.S. in electrical engineering in 1947 from the University of Colorado and his Ph.D. in physics in 1952 from the University of Pittsburgh. He was elected a member of the National Academy of Sciences in 1975 and appointed to the Academy of Finland in 1980. His many honors include the two top awards given by the low-temperature physics community: the Simon Memorial Prize and the Fritz London Memorial Award. At the time of his death, he was the first joint Fellow of the University of California, Los Angeles, and Los Alamos National Laboratory.



Albert Migliori earned his B.S. in 1968 from Carnegie-Mellon University and his Ph.D. in physics in 1973 from the University of Illinois, where he studied superconducting thin films. He then joined Los Alamos as a postdoctoral fellow and studied high-field and self-field behavior of hard type II superconductors. In 1975 he was awarded a National Science Foundation Fellowship to study internal and surface magnetic fields in current-carrying superconductors with the Mössbauer effect. In 1976 he became a staff member of the Condensed Matter and Thermal Physics Group.

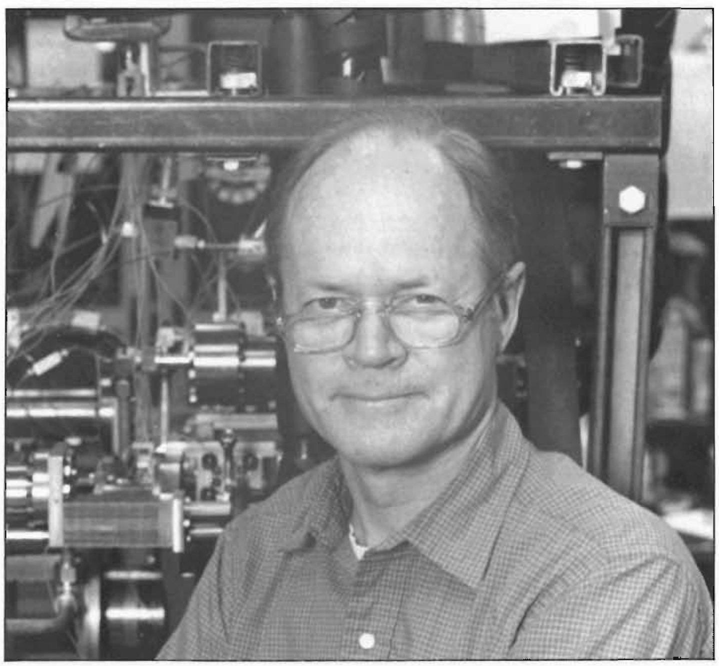
Gregory W. Swift is a staff member in the Condensed Matter and Thermal Physics Group, where he has been working on novel heat engines, acoustics, and superfluid helium-3 since 1981. He received his B.S. in physics and mathematics from the University of Nebraska and his Ph.D. in physics from the University of California, Berkeley. From 1983 to 1985 he held an Oppenheimer Fellowship at Los Alamos.



PUSHING *the* LIMITS

John Wheatley (1927-1986)

John Wheatley, one of the great low-temperature experimental physicists of the twentieth century, died suddenly this spring while bicycling to his lab at UCLA. His premature death left unfinished a large number of fascinating projects both in Los Angeles and in Los Alamos, among them the one on natural heat engines that he was writing about for this issue of *Los Alamos Science.*



As a tribute to John and his brilliant contributions to science and technology, a group of close associates shared with us their insights about the man and his achievements. What made him a great scientist? How did he succeed in carrying out high-precision experiments at such low temperatures? Why did other experimentalists frequently aim to prove him wrong? Why is research at a few thousandths of a degree above absolute zero so tricky? How did John Wheatley interact with theorists, graduate students, administrators? These are some of the topics addressed in the following round table. The participants included both theorists and experimentalists. Theorists David Pines (the current Bernd Matthias Visiting Scholar at the Los Alamos Center for Materials Science) and Gordon Baym from the University of Illinois had worked with John at Urbana in the sixties on liquid helium-3 and dilute solutions of helium-3 in helium-4; theorist Al Clogston, long-time member of Bell Laboratories and now a member of the Center for Materials Science, worked with John during the last three or four years on new ideas about nonlinear excitations in molecular systems. The theorists trusted John's physical intuition and knew they could count on his results.

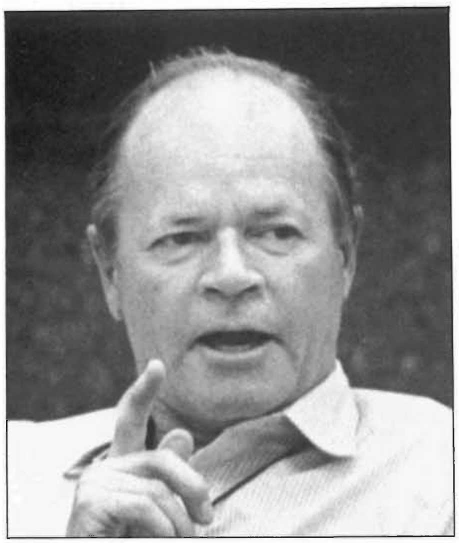
Three experimentalists at the round table got down to the nitty-gritty, giving a vivid picture of John's genius in the laboratory. Matti Krusius from Otaniemi, Finland, had worked with John during the seventies at the

University of California, San Diego, on the superfluid phases of helium-3 and at Los Alamos with John on spin-polarized hydrogen. Greg Swift worked with John at Los Alamos on superfluid helium-3 and natural engines, and Al Migliori worked with him here on natural engines and nonlinear excitations in molecular systems.

Sig Hecker wanted very much to be a participant but was unavoidably traveling. We interviewed him later, inserting his comments where appropriate. He gives a moving description of how his not-so-easy collaboration with John on establishing the Center for Materials Science grew into a strong friendship.

The result is a portrait of a man who inspired those around him by his extraordinary drive for excellence, his intense interest in science, and his joy at being in the lab doing the best experiment that could be done. Although his insistence on perfection, his intolerance of incompetence, and his confidence that he was right could often be a source of friction, he will primarily be remembered for his contagious enthusiasm and ingenious skill in pushing the limits of science.

Pines: John was the pre-eminent low-temperature physicist of his generation. He made absolutely major contributions both to low-temperature technology and to understanding the physics of the helium liquids. Between the late fifties and the mid seventies, he carried out most of the key experiments on liquid helium-3 and the dilute solutions of helium-3 in helium-4 that either provided a basis for a theoretical understanding or else confirmed theoretical predictions. He also



carried out a number of the key experiments on the superfluidity of helium-3 and missed, by really a shadow, identifying the superfluid phases.

John was not just the outstanding experimentalist in the low-temperature community; he was also its conscience. He paid attention to what other people were doing and was willing to take the time to sort out why their results were different from his own. In that respect he was unique. And he did it all with great style, verve, honesty, and a sense of humor.

John as the conscience of the low-temperature physics community is epitomized by the following anecdote. In 1964, at the Eighth International Low-Temperature Physics Meeting in Columbus, Ohio, one of the leading Soviet low-temperature experimentalists, V.P. Peshkov, presented the details of his previously announced "discovery" of the long sought-after superfluid phase of helium-3. Following Peshkov's presentation, John got up and, in the most careful, honest, objective way, pointed out what he thought were the fatal flaws in the experiment. John demolished Peshkov but not in any personal sense. He just demolished the way in which Peshkov had arrived at his temperature scale—one of the problems in low-temperature physics is knowing what temperature you're at—and then pointed out the effects that may have led Peshkov to erroneously conclude that he was dealing with a superfluid phase of helium-3. John had done experiments down to lower temperatures, he was sure of his temperature scale, and he knew he hadn't seen superfluidity.

Migliori: Later, when John in fact had superfluid helium-3 in his own lab, he didn't know it, although he was right about his temperature scale.

Krusius: Unfortunately, that one mistake probably cost him the Nobel Prize. He did most of the pioneering experiments on liquid helium-3, both before and after the discovery of superfluidity by Osheroff, Lee, and Richardson in 1972.

Pines: Another measure of John the scientist concerns the debate during the years 1980 to 1983 over the correct low-temperature specific heat of liquid helium-3. The results obtained in 1980 by a group working in Helsinki differed by some 40 per cent from the results that John and his collaborators had found in their classic work in the mid sixties. This discrepancy was of great concern because helium-3 is the benchmark liquid in all ultralow-temperature work. After the announcement of the Helsinki group's results, John wrote a long letter to the experimentalists in the field discussing all the possible things that could go wrong and all the consistency checks that were needed to do an accurate measurement. Within about a year and a half of that letter, Dennis Greywall at Bell Laboratories carried out what is likely to remain the definitive experiment. It led to results that differed from John's by about

10 per cent. No one understands to this day why the Helsinki experiments were so far off. But certainly John led the way in suggesting what could have caused an erroneous result.

Clogston: John was definitely a driving force in the low-temperature community, always stretching things to the limit. Take, for example, the Argentinean adventure.

Pines: Yes, John spent two years [1962 and 1963] founding the low-temperature group in Bariloche, a city off in the Argentinean countryside not far from the Chilean border.

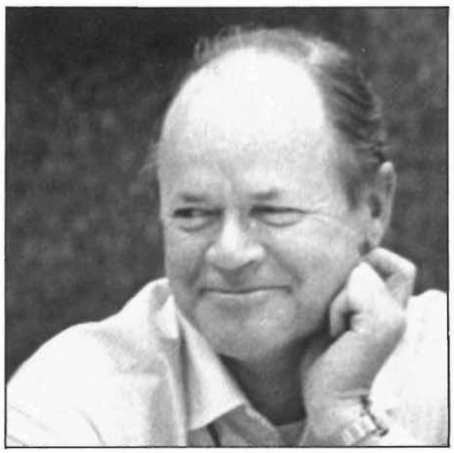
Krusius: At first the conditions at Bariloche were very primitive with hardly any electricity or water. Everything had to be started from scratch. They even had to build the liquefiers to make liquid helium and liquid hydrogen.

Pines: He went there because he liked the people, liked the adventure, and liked the opportunity afforded by the research atmosphere of Argentina at that time. Before he left, he set up a two-way radio on the roof of the Urbana physics building so he could talk regularly to his colleagues there and order much-needed equipment for the lab in Bariloche. On his return he used the set to stay in touch with the people in Bariloche. John was thoroughly successful in his Bariloche enterprise. A number of his former students continue to work there, and the lab is recognized as a center of excellence in the South American experimental physics scene.

Krusius: Even for a scientist of John's stature, a highly recognized international standing does not come automatically. John won international recognition and had many, many friends abroad because he cultivated and worked with his foreign colleagues. In 1975 he received the Fritz London Memorial Award, the highest recognition of the international low-temperature physics community, and an honorary degree of Doctor of Science from the University of Leiden. In 1980 he was appointed to the Academy of Finland. John really valued those recognitions. The Academy of Finland, which consists of only some thirty members, is small enough that if one member says he's interested in studying the use of hydrogen gas as fuel for the diesel cycle, he'll be whisked off in a helicopter to an icebreaker to see large-scale diesel engines in action.

Science: *When did John begin his work in low-temperature physics?*

Pines: He came to the University of Illinois in 1952 as a nuclear physicist, working on paramagnetic resonance and nuclear magnetic moments.



Krusius: John decided that the proper way to study the interactions of magnetic moments in condensed matter was to polarize the moments using low temperatures and high magnetic fields. So he had an immediate need to get acquainted with low-temperature techniques. In 1954 he spent a year in Leiden at the Kamerlingh Onnes Cryogenics Laboratory, the renowned birthplace of low-temperature physics. During his stay he undoubtedly became acquainted with the concoction of myths and cookbook recipes that made up low-temperature technology. When he returned to Illinois, he set out to correct this situation by methodically establishing the basic techniques of present-day low-temperature refrigeration and thermometry.

At first he used the adiabatic demagnetization of cerium magnesium nitrate as a cooling method for all his low-temperature work. When helium-3 became available in the late fifties, he added helium-3 evaporation as a precooling step and was able to extend the low-temperature limit by a factor of 20, down to a few millikelvins. At the same time he developed the whole technology needed to work at these low temperatures. For example, he compiled a list of materials according to their magnetic susceptibility at low temperatures so one could use materials for the apparatus that would not interfere with the measurement of very small magnetic moments.

Science: *Tell us about the atmosphere at Illinois in the fifties and sixties.*

Pines: John and I both came to the University of Illinois in 1952—as did Hans Frauenfelder and Francis Low. John Bardeen, Fred Seitz, and Charlie Slichter were already there. Later, in the sixties, Gordon Baym, Tony Leggett, and Chris Pethick came. It was a remarkable group; we very much enjoyed talking and working together. At first, during his nuclear physics phase, John Wheatley sat somewhat apart, but soon his work on the helium liquids made him a central figure in our discussions.

Science: *Why was helium so interesting?*

Pines: Helium-4 was always of great interest to both the experimental and theoretical low-temperature community because it remains a liquid down to the lowest temperatures—as long as you don't squeeze it—and because helium-4 becomes a superfluid, with all sorts of fascinating properties, below 2.19 kelvins. When helium-3 became available in large quantities in the late fifties, the attention of both theorists and experimentalists turned to the properties of this new quantum liquid. Because helium-4 is a spinless particle, a boson, it condenses to a superfluid in which all the atoms are in the same lowest energy state. Helium-3 with a spin of ½ is a fermion and can have only one particle at a time in a given state; hence it was expected that the quantum liquid properties exhibited at very low temperatures would be quite different from those of helium-4.

In the early days Los Alamos had more

helium-3 than anybody else because of the work done with tritium for the Super [the first H-bomb design], and the Lab also had a very active low-temperature group. It is this group that John eventually joined.

Baym: Of course it was no accident that helium-3 research began in Los Alamos in the late fifties. The half-life of tritium, the parent of helium-3, is 12.5 years, so it was natural that by the mid fifties substantial stocks of helium-3 would have begun to build up. Landau's paper on his Fermi-liquid theory of helium-3 was published in 1957, just one tritium half-life after the end of the Second World War.

Pines: The theoretical challenge was to understand why Landau's theory of Fermi liquids worked as well as it did for helium-3. The experimental challenge was to achieve temperatures low enough to cause the strongly interacting system of fermions, helium-3, to behave like a collection of weakly interacting elementary excitations. This is the not-so-obvious prediction of Landau's model. The elementary excitations in the model are helium-3 quasiparticles and quasiholes near the Fermi surface, analogous to the particle and hole excitations of electrons in a metal. For helium-3 at a low enough temperature that quasiparticle and quasihole excitations govern its properties, the model predicted that the specific heat would vary linearly with temperature, the spin susceptibility would be independent of temperature, and a collective mode, called zero sound, would arise. Zero sound is a density wave that propagates by means of the forces between the atoms, rather than by collisions maintaining local equilibrium, as in ordinary sound.

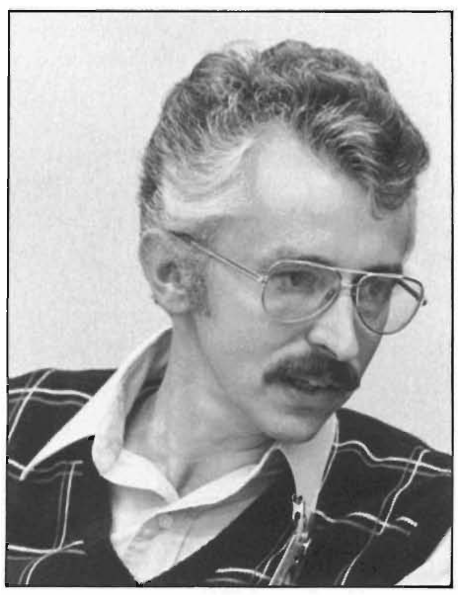
Science: *What were John Wheatley's contributions to the study of helium-3?*

Pines: First he developed the technology to cool to temperatures below 100 millikelvins, as Matti described. Once he had the technology, his work on helium-3 took off. John and his collaborators demonstrated that the transport properties—thermal conductivity, viscosity, ultrasonic attenuation, and spin diffusion—of liquid

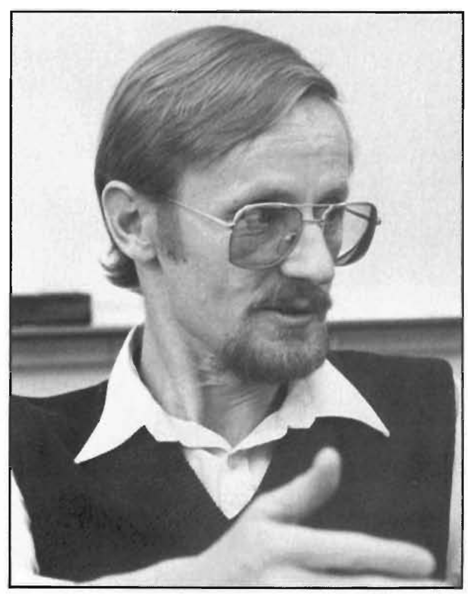
Round Table Participants

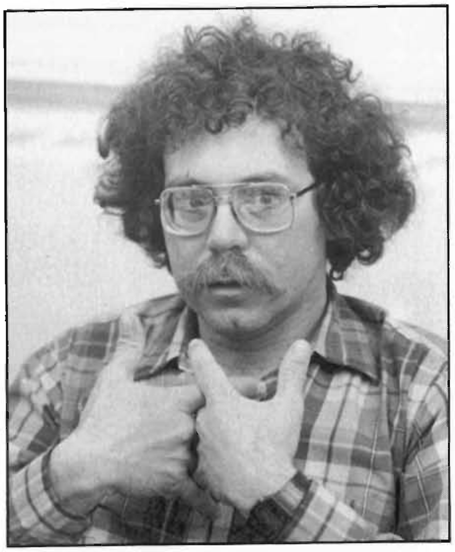
John's character was dominated by a systematic drive for excellence in all things.
—Greg Swift

John was always the first to perform a new type of experiment. He would quickly set the benchmark, and it would be a correct one.
—Matti Krusius



When any technical idea was brought up, John wouldn't let us continue until he understood every aspect of it. That's what made him such a pain and so beautiful too.
—Sig Hecker

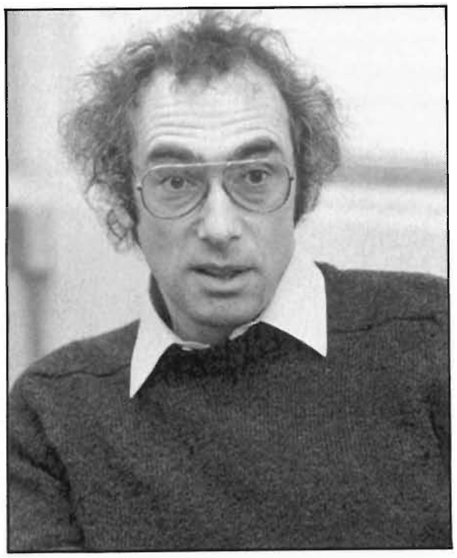


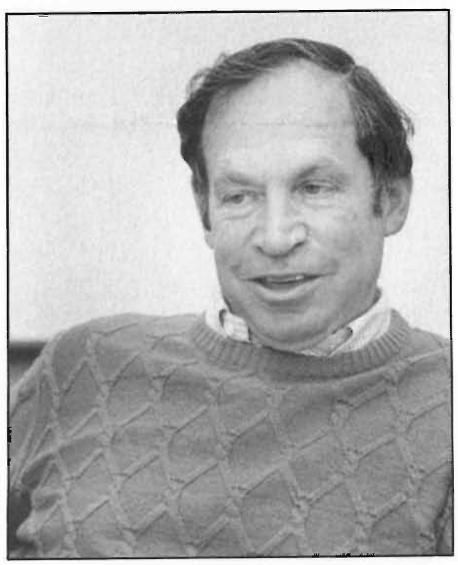


"... The irreversibility will be the thing that makes the engine work.*" He liked that idea because then the engine, the natural engine as he called it, wouldn't have a single extra thing wrong with the technology.*
—Al Migliori

John Wheatley developed his profound understanding of his experiments through mechanical models and simple theories that captured the essence of the physics involved.
—Al Clogston

By understanding the numbers, he was always way ahead of the theorists; he could tell quickly whether a theoretical guess was right or wrong.
—Gordon Baym





John was not just the outstanding experimentalist in the low-temperature community; he was also its conscience. ... In that respect he was unique.
—David Pines

helium-3 agreed with the predictions for a Fermi liquid. John was also the person who directly found zero sound in helium-3. Wilks had done an indirect measurement based on acoustic mismatch, but Wheatley, Abel, and Anderson made the direct observation in 1966.

Krusius: John was always the first to perform a new type of experiment. He would quickly set the benchmark, and it would be a correct one. This happened so consistently you often heard at conferences that the driving motivation for a later experiment had been a feeling that the early Wheatley experiment could not possibly have been right.

Swift: The basic principle here is that John was always right. Low-temperature experiments have many pitfalls, the most important being the determination of an accurate temperature scale. John was way ahead of everyone: he consistently worked sixty-hour weeks, had a fantastic memory for everything he did in the lab, and outlined his experiments in meticulous detail.

Krusius: An illustration is his work on the temperature scale in the millikelvin range. All through the years John had collected his own measurements on the absolute temperature scale, which were done with cerium magnesium nitrate. He then used them to finally summarize the helium-3 superfluid transition curve as a function of the externally applied pressure and absolute temperature. Many groups have since employed different techniques, but his initial tabulation has resisted challenges remarkably well.

Clogston: I think of John Wheatley as one of the world's premier experimentalists. He certainly wasn't a theorist; I'm not even sure he had a very high regard for theory. But he certainly interacted, David, with you, Gordon, and John Bardeen very deeply and intensively. What was the nature of that interaction?

Pines: Whenever John observed anything new and puzzling about helium-3, he'd climb up to the lair of quantum-liquid theorists two floors above his lab to ask questions. John always had a strong feeling for the essence of what was happening and a real interest in finding the simplest theoretical description. On the other hand he never got involved with theoretical technology. For instance, he understood the theoretical construct of a Fermi liquid but never learned the mathematical details at the heart of the theory.

Baym: The most intense interaction was on dilute solutions of helium-3 in superfluid helium-4; we were back and forth almost daily. John was always trying to force us as theorists to explain in his terms how the mixture was working. He insisted on very detailed explanations. By understanding the numbers, he was always way ahead of the theorists; he could tell quickly whether a theoretical guess was right or wrong.

Pines: The helium-3/helium-4 mixtures were a good system to work on because we understood helium-4 completely, and we could add helium-3 a little at a time, really checking theory against experiment as we went along. You could study a range of densities for helium-3 that's impossible to obtain in pure gaseous helium-3. Because of John Wheatley, Gordon, John Bardeen, and I attacked a fascinating theoretical puzzle, the nature of the effective interaction between helium-3 atoms in the helium-4 background at temperatures so low that the helium-4 behaves like a mechanical vacuum.

Science: *What do you mean by a mechanical vacuum?*

Clogston: Helium-3 binds more strongly to helium-4 than to itself, making it energetically more favorable under some circumstances to be in a mixture than to separate into different phases.

Migliori: As a result, the amount of helium-3 that will dissolve in helium-4 at atmospheric pressure can be as large as 6 per cent down to absolute zero temperature. At the same time, superfluid helium-4 has no viscosity, so helium-3 moves nearly as if nothing is there. It acts like particles in a vacuum, except, since you're in the liquid phase, you have a lot more atoms present per unit volume.

Baym: An example of John's influence on theory arose from his measurements of the thermal conductivity and spin diffusion of dilute solutions of helium-3 in helium-4. It appeared impossible to construct any theory of the effective interaction that would explain both of those experiments. From the discrepancy with the experiments—and John stood by his results—we discovered that the solutions to the Landau kinetic equation we were using were in fact not very accurate. John's measurements really inspired the subsequent work that eventually led to the exact solution. If John had not done the experiments and pushed on the theorists, that theoretical advance would likely not have been made for quite a while.

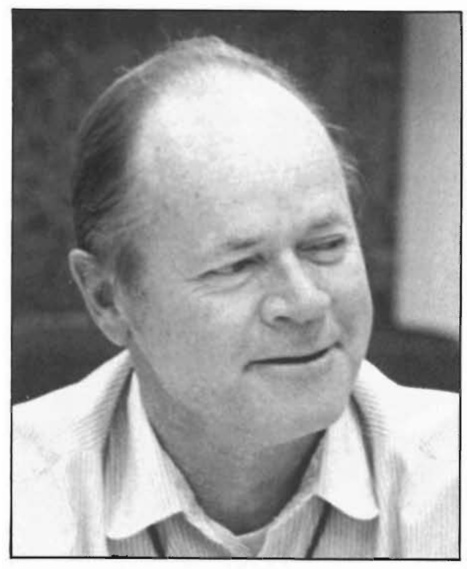
Migliori: John really understood what it meant for helium-4 to act as a mechanical vacuum for the helium-3 particles.

Baym: He realized that if you cooled by evaporating helium-3, the binding of helium-3 to helium-4 allowed you to achieve a higher vapor pressure at a given temperature than you could in the ordinary vacuum. If you just evaporated helium-3, you could cool down to about a third of a kelvin, but if you evaporated helium-3 into helium-4, you could go to much lower temperatures. That's the principle of the dilution refrigerator, which John developed to a practical device.

Swift: John was able to get a factor of a hundred lower in steady-state low temperature with the dilution refrigerator. Prior to this development work, the lowest continuously available temperature was about 0.3 kelvin, produced by evaporating helium-3 into a vacuum. London suggested the principle of cooling by diluting concentrated helium-3 in helium-4, but John engineered it into a practical reality, using heat exchangers, distillation units, and pumps to circulate the helium-3 continuously. This was an extremely important technological development. A factor of 100 reduction in temperature is as important in condensed-matter physics as a factor of 100 increase in energy is in particle physics.

Pines: In some ways John was more excited about his work on dilution refrigerators than he was about having sorted out the low-temperature experimental properties of helium-3. He worked very hard on the theoretical papers he wrote in connection with the dilution refrigerator. Right before he left Urbana in 1966, he had a refrigerator running, and he, of course, developed them further after he left. In 1970 during a symposium talk on experiments of the future, he spoke about ultralow temperatures of 30 millikelvins. Now, thanks to John's dilution refrigerators, one can reach those temperatures fairly easily.

Krusius: His dilution refrigerator chopped the continuous-cooling frontier down to 4.5 millikelvins. Dilution refrigerators



now go down to 2 millikelvins, but they require much larger pumps than John had in the early work.

Pines: When you combine the dilution refrigerator with demagnetization techniques, you can get down another factor of 10 to about 0.2 millikelvin.

Swift: But remember, there's a big difference between dilution refrigeration, which is continuous, and demagnetization, which is one-shot.

Baym: Dilution refrigerators are now being used at the high-energy laboratories, including LAMPF and Brookhaven, to cool targets to very low temperatures for polarized target experiments.

Getting back to John, when he left Urbana he had a very strong urge to get into a different field; he wanted to work on geophysics. However, superfluid helium-3 was discovered in 1972, and then John couldn't break away.

Science: *How did John happen to miss the discovery of the superfluid phase?*

Krusius: Ironically, John had made measurements below the superfluid transition temperature earlier but missed identifying it because he had his own inimitable and sometimes stubborn way of doing things. He was a very organized and meticulous worker, but he was sometimes reluctant to resort to the most modern type of equipment. Doug Osheroff discovered the transition early in 1972 during his compressional adiabatic cooling experiments at Cornell. Osheroff found a glitch in the pressure as a function of time—a very small glitch, only a few per cent of the total. This tiny glitch was the superfluid transition everyone had been looking for. About a year and a half before, John had also been developing adiabatic compressional cooling to obtain low temperatures and to look for the transition. During those experiments, rather than reading the pressure as a function of time from a strip-chart recorder attached to a pressure transducer, John had Rich Johnson, the graduate student doing his thesis work on this experiment, sit on a stool and shout out numbers from a pressure gauge with a needle. These discrete points did not show an obvious glitch. Afterwards, knowing where the transition occurred, they went back and plotted their data above and below the glitch and saw the transition.

Swift: After Osheroff's discovery John became very serious about making measurements on these new superfluid phases. He brought Matti from Helsinki to La Jolla as a postdoc, and, together with a couple of students, they built a cryostat that used a dilution refrigerator as the precooler and

JOHN WHEATLEY—CAREER HIGHLIGHTS

EDUCATION

1947 B.S. in electrical engineering, University of Colorado
1952 Ph.D in physics, University of Pittsburgh

THE URBANA YEARS—1952-1966
(University of Illinois)

Research

Paramagnetic resonance and nuclear magnetic moments, low-temperature refrigeration and thermometry, development of the helium-3 dilution refrigerator, transport and physical properties of liquid helium-3

Honors and Special Appointments

1954-55 Guggenheim Fellow and Fulbright Research Scholar, the Netherlands
1962-63 Fulbright Research Scholar, Argentina
1965-66 Member, University of Illinois Center for Advanced Study
1966 Simon Memorial Prize

THE LA JOLLA YEARS—1966-1981
(University of California, San Diego)

Research

Low-temperature phases of helium-3, dilution refrigeration, magnetic properties of dilute alloys, development of the point-contact SQUID, properties of superfluid helium-3, liquid working fluids in heat engines

Honors and Special Appointments

1968 William Pyle Philips Lecturer, Haverford College
1969 Loeb Lecturer, Harvard University
1975 Member, U.S. National Academy of Sciences
Doctor of Science, honoris causa, University of Leiden
Ninth Fritz London Memorial Award
1980 Academician, Academy of Finland

THE LOS ALAMOS YEARS—1981-1985
(Los Alamos National Laboratory)

Research

Conventional and natural heat engines, superfluid helium-3, Rayleigh-Bénard convection in mixtures of helium-3 and helium-4, spin-polarized hydrogen

Honors and Special Appointments

1983 Fellow, American Academy of Arts and Sciences
1984 Fellow, Acoustical Society of America
Distinguished Graduate Award, University of Pittsburgh

THE LOS ANGELES YEAR—1986
(University of California, Los Angeles)

Research

Nonlinear localization of vibrational energy

Honors and Special Appointments

1986 First Joint Fellow, UCLA and Los Alamos
1986 Chosen to be the first holder of the President's Chair at UCLA, a new position recognizing the most outstanding faculty member on campus

adiabatic demagnetization as the final cooling step. It was basically the Helsinki design, but it's remarkable that it worked just fine the first time they tried it. At 1 millikelvin a heat leak of only a billionth of a watt can be disastrous, so this is difficult work. Their first experimental run produced about ten publications, half of them in *Physical Review Letters*. This at a time when most experimentalists in the field were struggling just to get cold.

Krusius: John studied superfluid helium-3 systematically all through the seventies, producing one of his most important contributions to physics.

Baym: The superfluid phases of helium-3 are interesting in that they resemble a superconductor, except that the Cooper pairs have one unit of angular momentum instead of zero, which makes the description of the ordering much more complicated. There are two phases, called A and B. The A phase has an anisotropy axis, making it like a quantum liquid-crystal. A lot of John's work was on the dynamics of this anisotropy axis—orienting it with magnetic fields and such and using the anisotropy of zero-sound attenuation to see what was going on.

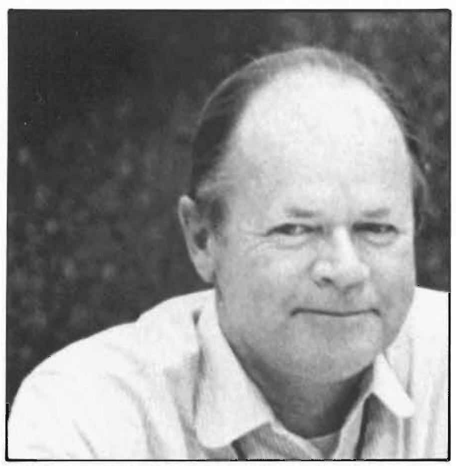
Pines: It seems that John began his study of superfluid helium-3 by, once again, improving the experimental technology.

Krusius: In particular, John was one of the key persons to develop the point-contact SQUID, a superconducting quantum-interference device that can measure very small magnetic fields and voltages.

Migliori: John used the SQUID to measure magnetic fields as well as temperature in his superfluid helium-3 experiments. A SQUID voltmeter was a good way of making a low-power, precise measurement of the magnetization curve of a dilute magnetic salt, and if you understand the physics of the salt reasonably well, its magnetic susceptibility defines the temperature scale. So the point-contact SQUID became a very important tool for determining temperature.

Krusius: John was probably the first person to use the SQUID for that. He also, I think, was the first person to use the SQUID to measure thermal noise in a resistor, letting that define a temperature scale also.

Migliori: The gold-iron SQUID thermometer is also a nice tool, sufficiently well developed by John's graduate students that it was written up in an instrument publication. That tool is used here at Los Alamos to study Rayleigh-Bénard convection in dilute helium-3/helium-4 solutions. These experiments yield signal-to-noise ratios equal to that of a compact disc, that is, 90 decibels. As a result, one can measure a long series of period doublings as the system undergoes a period-doubling transition to turbulence. These



are the finest measurements of the convecting state—in how it's initiated and how it evolves—and it's a real measurement, not a computer simulation.

Pines: The point-contact SQUID is now being used to measure electrical currents in the brain through the tiny magnetic fields that are generated. A number of people at Los Alamos and elsewhere are interested in trying to measure magnetic brain waves and to use a magnetic map of the brain for medical diagnosis.

When John went to San Diego after leaving Urbana, he not only devoted himself to research on superfluid helium-3 but founded the SHE [Superconductivity, Helium, and Electronics] Corporation. He did this, in part, to exploit some of his own inventions and also to make available excellent low-temperature equipment in a way he knew that no commercial company was going to do.

Swift: Two other low-temperature physicists were also founders of SHE, but it was John who guided it on a daily basis. Their main products have been dilution refrigerators and SQUIDs. John's technical expertise was indispensable in getting SHE started. I think the University faculty frowned on this involvement. It looked to them like John was developing these technologies at the University and at government expense and was trying to profit from them personally. Actually he never made a cent. A lot of his former students and postdocs work there now. That was one way he transferred his low-temperature technology from the lab to SHE.

Pines: At heart, John was a kind of missionary. Although he was a very private person, in another sense he was not at all secretive. He didn't develop these wonderful techniques to keep to himself and his immediate associates. He was always willing to share. He not only built the Argentinean low-temperature community but also played an important role in building a major low-temperature community in Finland.

Science: *What about John's work at Los Alamos these last five years?*

Krusius: Before he came to Los Alamos, he was interested in doing something new. He began thinking about thermal physics with heat engines, but he had a feeling such work could not be done at UCSD.

Migliori: When I first talked to John in San Diego about his decision to come to Los Alamos, he expressed this love of technology that has been a theme throughout his life. His interest in heat engines has a similar quality to his work on the dilution refrigerator. He wanted to do technology, to make things that did something. For that reason the acoustic engine work here at Los Alamos was one of the high points in his life.

Pines: His recent work on heat engines was an outgrowth of his profound working knowledge of thermodynamics and heat cycles and what you could do with them. Having solved essentially all those problems in low-temperature physics, he wanted a new major challenge. He went back to the work of the great thermodynamicists and engine creators of the nineteenth century, such as Carnot and Kelvin, and took a fresh look at that technology to see what you could do.

Clogston: I had the impression, working with him these last three or four years, that he always based his thinking on some kind of classical model. Perhaps he was one of the last great classical physicists.

Migliori: He loved these engines, especially the acoustic engines, more than anything else because the physics could be understood on the basis of classical thermodynamics. A normal person could understand it. Although there was nothing quantum mechanical about acoustic or natural engines, they were, nevertheless, a completely new development, and they were sufficiently rich and complex that they challenged everyone's understanding.

Science: *What was the new idea behind natural engines?*

Swift: In traditional heat engine designs, the idea is to minimize irreversible processes because they lead to inefficiencies.

Migliori: Anything you build is going to be irreversible, so from the traditional point of view, there'll always be some process that messes you up. John said to himself, "I'm going to make that irreversibility work for me. The irreversibility will be the thing that *makes the engine work*." He liked that idea because then the engine, the natural engine as he called it, wouldn't have a single extra thing wrong with the technology.

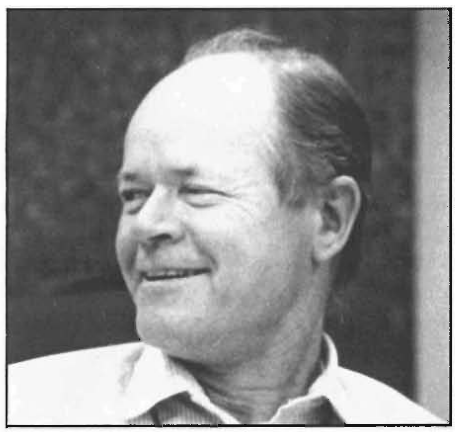
Swift: It's like taking a liability and turning it into an asset.

Migliori: The acoustic engine was the first natural engine John developed. Merkli and Thomann and then Nikolaus Rott had discovered the important acoustic engine principles, but John was trying out his new principle to see if it had absolute global importance.

Pines: In the summer of eighty-three, he organized a meeting to see whether people in various parts of physics would agree with him that this was a whole new approach to understanding engines.

Baym: I had pointed out to John the relation of his idea to instabilities in stars, and eventually he and Art Cox wrote a paper for *Physics Today* on the connection.

Pines: There are also a certain number of natural engines in your body, and that fascinated John. He wanted to see if he could make engines that operate at a molecular level.



Migliori: I think one of his chief motivations in all this work was his desire to make engines work well.

Swift: He liked the promise that something practical would come out of the research. Eventually he started using a liquid instead of a gas as the working substance in a heat engine—an idea that had lain dormant for fifty or sixty years. This idea led us to the liquid sodium acoustic engine, which we're working on now. John recognized all along that liquids were good things to work with and kept his eyes open for opportunities.

Migliori: The sodium acoustic engine is an example of taking an idea and implementing it with exactly the right working fluid. But it takes about fifty years to get an engine working properly in the economic sense—and you have to compete with existing technology. So even though the sodium engine has no moving parts, it's going to take time to yield a big payoff.

Clogston: The need for efficient engines in space is so great that the payoff may come sooner.

Pines: Well, SDI may push it up a little.

Science: *So the liquid sodium engine is truly a capstone on a prolific experimental career. What influence did John's experimental style—his way of doing things—have on other people?*

Swift: We need to talk about John's graduate students, because that's what his experimental career was all about.

Migliori: John had great skill getting students interested in topics he wanted to pursue. He displayed excellent taste, picking out topics that were important, and then, through force of personality or charm or just brute force, got people to work on them.

Swift: All the measurements in the fifties, sixties, and seventies on the properties of liquid helium-3 were made with John tightly controlling a handful of graduate students. That was the key to his great productivity. He was in the laboratory with them day and night, calling all the shots. They were reading the meters, and he was writing the numbers in the lab notebook. When he came to Los Alamos, he brought a number of students with him from the University of California. We still have a handful here. In the last few years, though, he started to let them take more initiative. He was mellowing.

Pines: You know, John had a killer instinct when he worked on something. He really wanted to get at it and get there fast. But it's a delicate point. When you want to get something done, the very best thing is to do it yourself. Sometimes, though, you need help, and you enlist a graduate student or a postdoc. If you want the answer in a hurry, you are on that person's neck every moment of the day. On the other hand it's not a very good way for a student or a postdoc to learn. The mellowing that Greg referred to was John's willingness to wait another day or two for the answer and let people make their own mistakes.

Krusius: I think John had a very positive interaction with his students. In low-temperature work you can lose an enormous amount of time if you do things wrong. John would have the student think about the measurement and come up with a proposal. Then they'd go through the plans together, and he'd press on the things he thought wouldn't work. The student would see that his proposal might not be a secure way of starting the experiment. That guidance was very useful. Otherwise, the student might waste a year or two.

Migliori: A simple example is using too much current through a thermometer so that it heats itself. That's a subtlety that might be missed by a new student.

Krusius: John's approach with his graduate students changed after his heart bypass operation three and one-half years ago, because time became immensely valuable to him. He really wanted to do physics on his own terms and not have too many people involved. John was the most organized person I know about doing work in the physics laboratory. When he came to the lab at seven or eight, he had a list in his mind of what he was going to do that day, and he really wanted to carry out all the things on that list. He didn't have much time for discussions. One graduate student in La Jolla solved this problem by bicycling home with John in the evening. During that bicycle ride he'd talk about his experiment and get advice for it.

Swift: In later years John would respond to a student's question by giving him as much time as he needed, but John would never seek out students to make sure they were doing the right thing that day.

Science: *What about his family?*

Swift: There was his wife, Martha, and two sons. His career would not have been possible without Martha because she devoted herself entirely to making his life easy. She was the foundation that gave him the freedom to do all the great things we remember him for.

Science: *Was his whole life devoted to work?*

Swift: No. Although he did spend some of his weekend time at work, a lot was spent with Martha riding bicycles on longer trips, or sometimes hiking or skiing. He had a passion for bicycling and, years ago, a passion for his motorcycle.

Clogston: That brings to mind the time in 1976 when John drove all the way from La Jolla to Urbana on his motorcycle and arrived just in time for a party honoring John Bardeen.

Pines: It was quite a dramatic event. We were holding a symposium on new directions in condensed-matter physics to honor Bardeen's retirement. John had allowed himself just enough time to make the trip by motorcycle. He appeared at the University about an hour and a half before the meeting was to begin, having driven through not one but two blizzards. But he loved it; he was so excited to have brought it off and to have arrived on time. That exhilaration of being out on the edge with the unknown is what attracts most of us to physics and keeps us there. John loved getting to lower temperatures than anyone else and to sort out tricky experimental aspects that might cause someone else to slip up. He loved living on the edge with his motorcycle, and he loved pushing himself on his bicycle.

Swift: The only thing that kept John from riding his bicycle every day was substantial snow or ice on the road. A couple of years ago, he took his bike with him to an Enrico Fermi Summer School in Italy and rode it from the Milan airport to Varenna—perhaps forty miles or so. During the two weeks of the school, he rode it around Varenna every day, and then he rode it back to the airport. That was pretty good for a fellow in a foreign country whose only fluent words in the native language were "more ice cream, please." At Los Alamos, John bicycled to work every morning, then home for lunch a little before noon, back in after lunch, and then back home again about six.

Photograph courtesy of the Wheatley family

John and Martha Wheatley congratulating their son Bill at his wedding.

Pines: John never liked to feel hemmed in. When he was at Urbana, he and his family lived in St. Joseph, a rural village about eight miles away. When they went to La Jolla, they didn't live near the beach; instead, they lived inland about ten miles.

Science: *Let's talk about John's impact at Los Alamos. I understand Jay Keyworth was responsible for bringing him here.*

Pines: Jay certainly played a major role. John called Bill Keller (then head of the experimental low-temperature physics group at Los Alamos) to say that he might be interested in moving to Los Alamos because of its strength in low-temperature work and its possibilities for his developing technological interests. Bill told Jay, who then launched a major campaign to secure the funds and space needed to attract John to Los Alamos. John would never have come without this kind of all-out effort.

Hecker: I remember Jay saying Wheatley must be gotten at any price because he did really high quality research.

Pines: John was pleased as punch with the space he had at Los Alamos and the possibilities of technical help. Every time I visited his lab, John would give me a guided tour and show me one more room and one more group of students doing another set of experiments. The move to Los Alamos was very liberating for him. He felt he could move out in a whole set of directions at once. This was simply not possible at La Jolla because of the lack of physical space as well as a lack of psychological space.

Hecker: John was incredibly protective of space. Part of his dowry in coming to Los Alamos was a good part of the old cryogenics building.

Swift: He had a dozen people—students, postdocs, staff members, technicians—working with him here on a whole set of problems that ranged from superfluid helium-3 to liquid Stirling-cycle heat engines. He spent most of his time managing, bringing his wisdom and good judgment to bear on the problems. It was incredible the way a little time with John could help point you in the right direction.

Clogston: John certainly brought an element of excellence and drive to Los Alamos that I think must have been unique. Also, John was one of the founding fathers of the Center for Materials Science. He worked closely with Sig and a few other people, and he maintained an enduring interest in the Center.

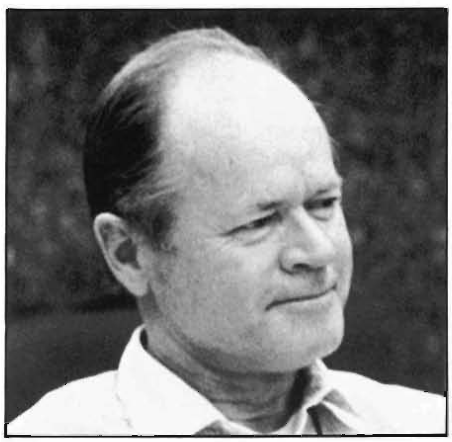
Swift: All John really wanted was to do physics. He'd do anything to get that to work, and, in the long run, he thought the Center would help him there.

Clogston: About two years ago he was made a member of the Center and that pleased him enormously, especially since, with his support coming from the Center, he could fund another postdoc. Again, that was exactly what he wanted.

Hecker: When we decided to start the Center for Materials Science, I wanted John on the internal advisory committee. His name would provide instant credibility with a large part of the solid-state, condensed-matter physics community. John somewhat reluctantly agreed, and then, for the next three months, I was sorry I'd asked. He just gave us hell. He made our meetings take twice as long as we anticipated because he was always a stickler for detail. When any technical idea was brought up, John wouldn't let us continue until he understood every aspect of it. That's what made him such a pain and so beautiful too.

Science: *Did he have definite ideas about the Center?*

Hecker: Absolutely! Early on, I gave what



I thought was an excellent seminar explaining my vision of the Center and emphasizing the sort of equipment we'd have, the type of building, and all kinds of other grand things. John thought it was a dreadful talk. The only thing that mattered to him was to get the best people; he assumed everything else would follow. In that sense he was quite an idealist. I, on the other hand, had to be a realist, because to build a center you have to know how to fit it into the existing structure. Eventually I realized just how valuable John was. I learned from him that you have to insist on excellence and insist on the quality of people. We had many disagreements, but gradually we learned from each other how to implement the Center, and John began to recognize that no one is going to just throw a million dollars at you.

Clogston: The Center was built adjacent to John's lab. Agreements were finally reached about how space was going to be shared, but I must say I detected no signs of mellowness whatsoever in John during those discussions.

Hecker: Even before that, in my eagerness to get the Center started, I tried to convince everyone we could start in a corner of the warehouse. That was the only space available at the time where we could build laboratories. Again, John thought that was a dreadful idea. He wanted the Center to be around him so he could interact closely with the people of quality that the Center was meant to attract. He knew he was being very selfish, and my first reaction was somewhat negative. I never expected to get space close to where he worked, but John helped make it happen. I remember his words well: the goal was "to build an intellectual community in materials at Los Alamos and have that community in a place where you get people rubbing elbows." The area close to his lab was clearly the right place since it's the site of a large share of the condensed-matter physics at the Laboratory.

Pines: I think John had no interest in exercising power for the sake of power. He had a very clear image of what the Laboratory could become—just very little opportunity to put those ideas to work. I'm not totally clear why John decided to return to an academic environment, but I've been told he felt he needed another army of graduate students.

Migliori: I think the key words are "army" and "graduate students."

Swift: Many people asked him why he was leaving, and I've collected seven or eight different answers. I think this is just part of his privacy. I don't think he wanted anybody to know why he wanted to leave.

Hecker: He *was* very unusual that way; he'd think things out totally beforehand. He came to me for advice only after he had made his decision to leave Los Alamos.

Migliori: Originally, John came to Los Alamos because he felt it was a good place to do technology. Eventually, though, he was disappointed that a few of the hoped-for services never materialized. When he found that he had to go outside for such things as electron-beam welding and plastic molding that were supposedly readily available at the Laboratory, some of his enthusiasm diminished.

Krusius: One reason he left was that, in the end, he realized he was a university teacher—an important part of his life that he missed.

Hecker: When we discussed his move to UCLA, he did say that maybe he was meant to be an academic person. Even here at Los Alamos, he ran a lot of his shop as if it were a university. He came, in part, because he felt he was a technologist and the Laboratory was a fantastic place to do technology. When he was ready to go back to the University, he said that maybe he was more physicist than technologist after all. He still felt Los Alamos was a fantastic place to do technology, but he wasn't quite clear what his role in that ought to be. Also, he missed the academic life and freedom. For instance, he always hated the fence around the site where his lab was located. He'd say, "Sig, tell me one thing. When are you going to get rid of this fence?" It turns out the fence went down about a year or so ago, not because of anything I did, but because of the process of fixing up the Center for Materials Science. When the fence went down, John was delighted. It's a pity that now the fence is gone, John is gone also.

Migliori: There were certainly many reasons for his leaving. Another was that he'd fought so long and hard over many issues that most of his blue chips at the Laboratory were gone. Of course, one of the fascinating things about John was that he argued a lot, but mostly he turned out to be right.

Swift: When John formed an opinion, it was very carefully thought out, and he knew he was right.

Hecker: After deciding to go to UCLA, John asked my advice on how to break the news and how to restructure his relationship with the Laboratory. I spent a lot of time with John carving out the idea that he proposed to Don Kerr, our Director, to become the first University of California-Los Alamos fellow. I wanted John to maintain his connection to Los Alamos. Not only was he doing very important work, but he provided a unique kind of leadership. There was just no substitute for having John Wheatley around.

Science: *Was the appointment successful?*

Hecker: The appointment is a very interesting one because it promotes a closer tie to the University of California and a better link to their students. The way John interacted with students was crucial to the way he did business. He was a natural teacher, and he recognized that the only way to get the best students is to be where the action is. He arranged to spend six months at UCLA and six months at Los Alamos. In fact his six months at UCLA were almost up when he died.

Clogston: We should list the experiments John was working on when he died.

Swift: Nucleation of the superfluid helium-3 B phase out of helium-3 A, measurements on sticking coefficients of spin-polarized hydrogen on superfluid helium, Rayleigh-Bénard convection in mixtures of helium-3 and helium-4, and a multitude of heat engines, including a liquid propylene Stirling engine, an acoustic cryocooler, a heat-driven acoustic cooler, and a liquid sodium acoustic prime mover. At UCLA he was doing nonlinear experiments on the localization of vibrational energy using the vibration of a thin cylindrical shell.

Migliori: In the last year and a half John, along with Scott Buchanan and me, became very interested in the localization of vibrational energy through nonlinear effects. At first the work was related to heat engine concepts, but it has since left those concepts far behind. Now, the idea is to establish whether such localized objects exist at the molecular level and whether they are as good an elementary excitation as anything else. To guide our thinking, John invented a classical model using the fact that a thin shell of stainless steel can exhibit some of the same effects that collections of tens or hundreds of molecules exhibit. Through this model one can make contact with the molecular system on an intuitive mechanical level by dealing with things you can hold in your hand. John, Scott, Seth Putterman at UCLA, and I planned to attack this problem in a major way. Of course, we're still going to do it.

Clogston: Within the last five years there's been a renaissance in classical physics as people developed tools to study nonlinear phenomena. This turn of events must have been very thrilling to John. As we discussed, classical physics is the area in which he really worked naturally, in which he could model things in his head. The surge of interest in nonlinear phenomena must have been very inspirational to him.

Migliori: But John felt, and I agree, that a lot of the principles are simple enough that one doesn't need the highest power theory to attack them. In fact, if you are going to do experiments, it's better to understand these things on a more fundamental and simpler level than merely to rely on buzz words and jargon.

Pines: Looking back on John's career, it seems that he'd concentrate on the technology for a period and suddenly there'd be a great outpouring of papers dealing with the physics made possible by that technology. His research at La Jolla was really based on the technology that he developed at Urbana with the dilution refrigerator. Toward the end of the time in La Jolla, he was beginning his work on heat engines. All the work in helium-3 was, as Matti said, based on solving the technical problems needed to do accurate experiments below a hundred millikelvins. He opened up new fields in science again and again either through his physics experiments or his interest in technology. One learned very soon that what John proposed, no matter how way out it might sound at first, had to be taken seriously.

Clogston: My experience was that John had enormous physical intuition, a really deep intuitive understanding of physics.

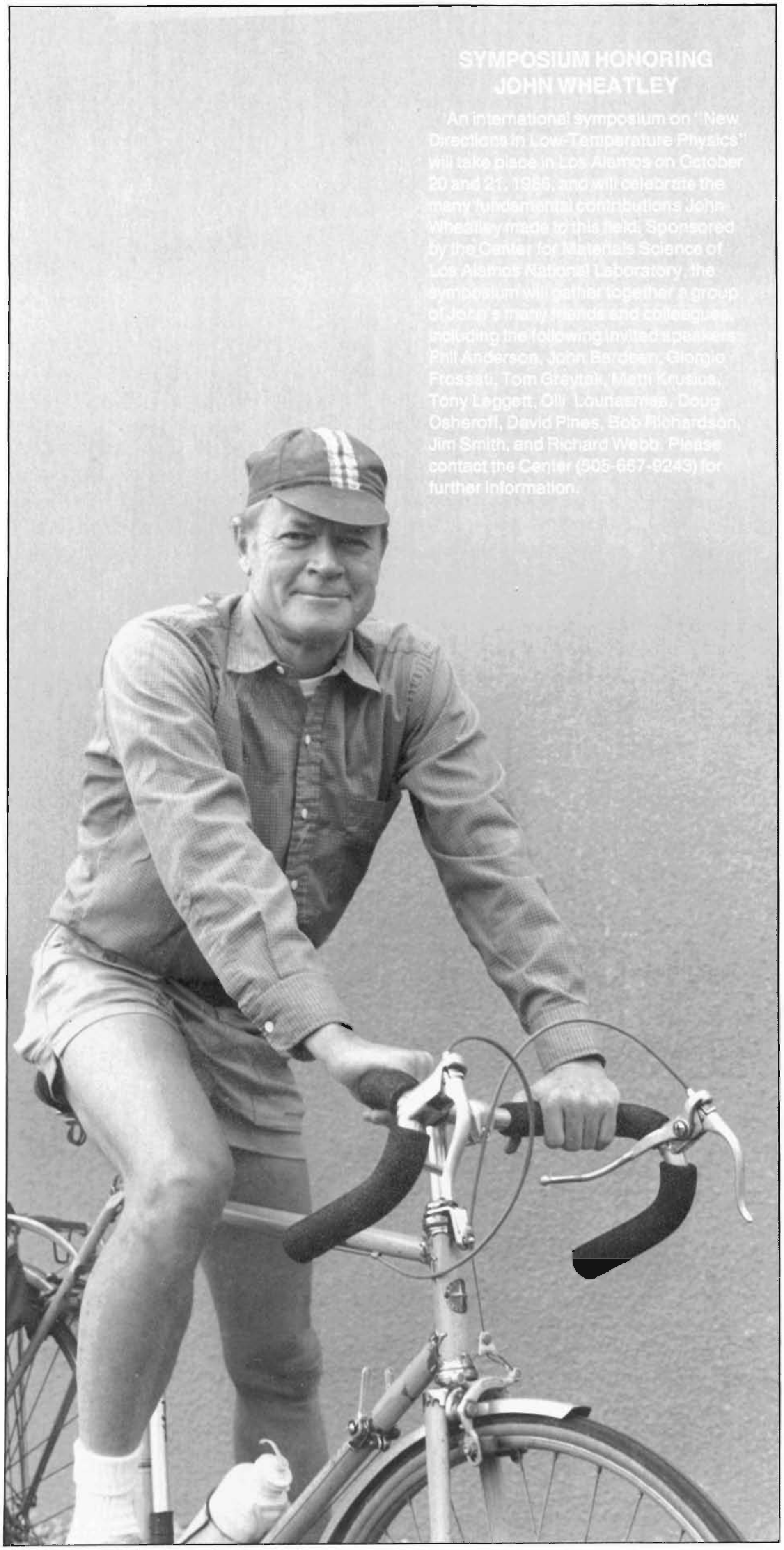
Hecker: I always respected John as a scientist, but I got to respect and love him as a human being. I learned later that he likewise had gained respect for me over a period of months and years. As a result, I was able to talk to him more frankly than to almost anybody else here at the Laboratory. He'd come to me for advice about his programs, his space, and his equipment because he knew I was a hardliner. There's no question he was difficult to deal with, but that was mostly because of his insistence on excellence. He wanted to be the best at everything he did, and he felt in order to do the best, he had to have the best. He insisted on it.

Pines: We don't miss just John the scientist; we miss John the person. He had an independent and special view about almost any topic. You couldn't anticipate it. He was never one to run with the crowd; he was just fun to talk to.

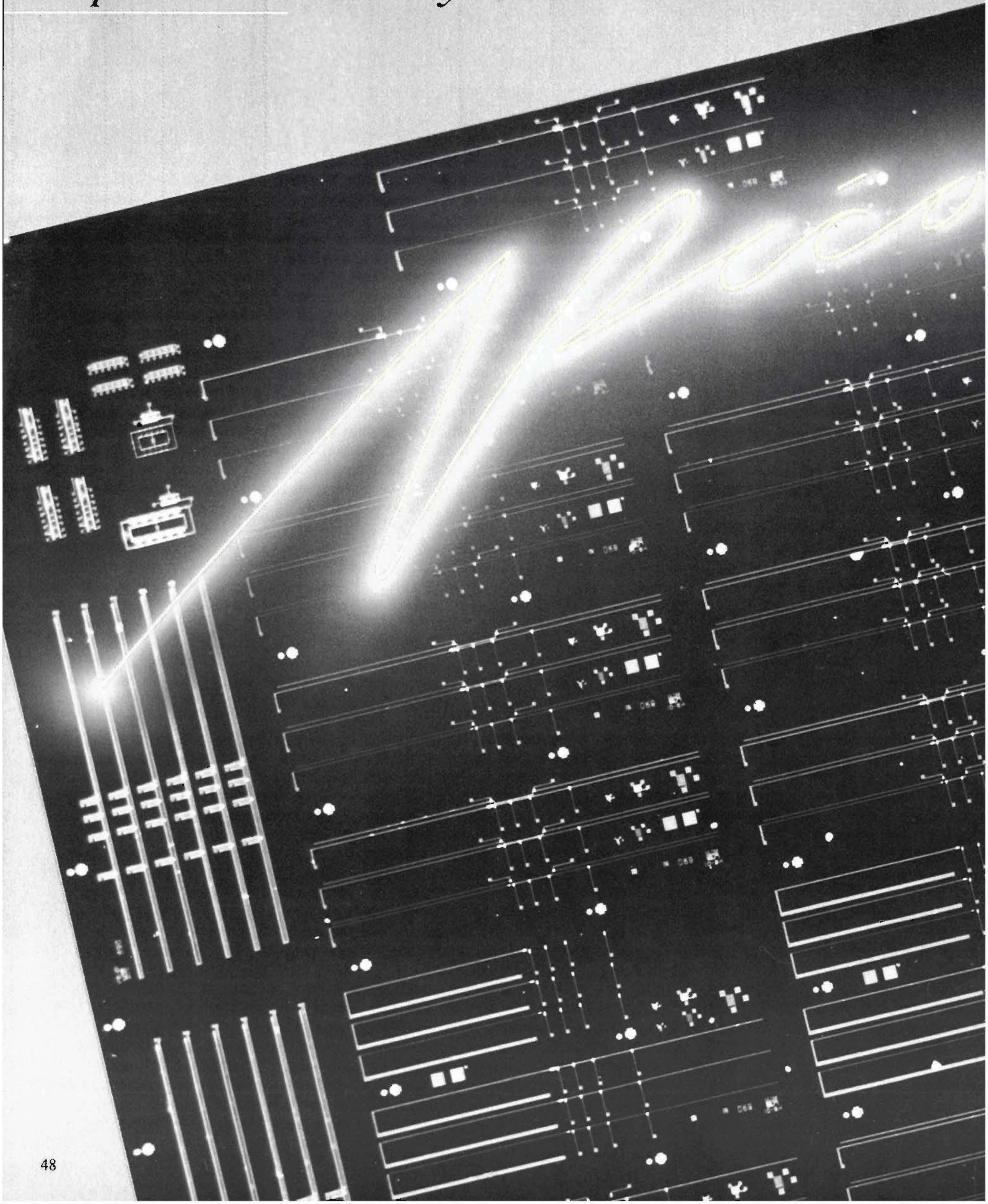
Clogston: Maybe that sums it up—he was fun to talk to. I'm going to miss him tremendously.

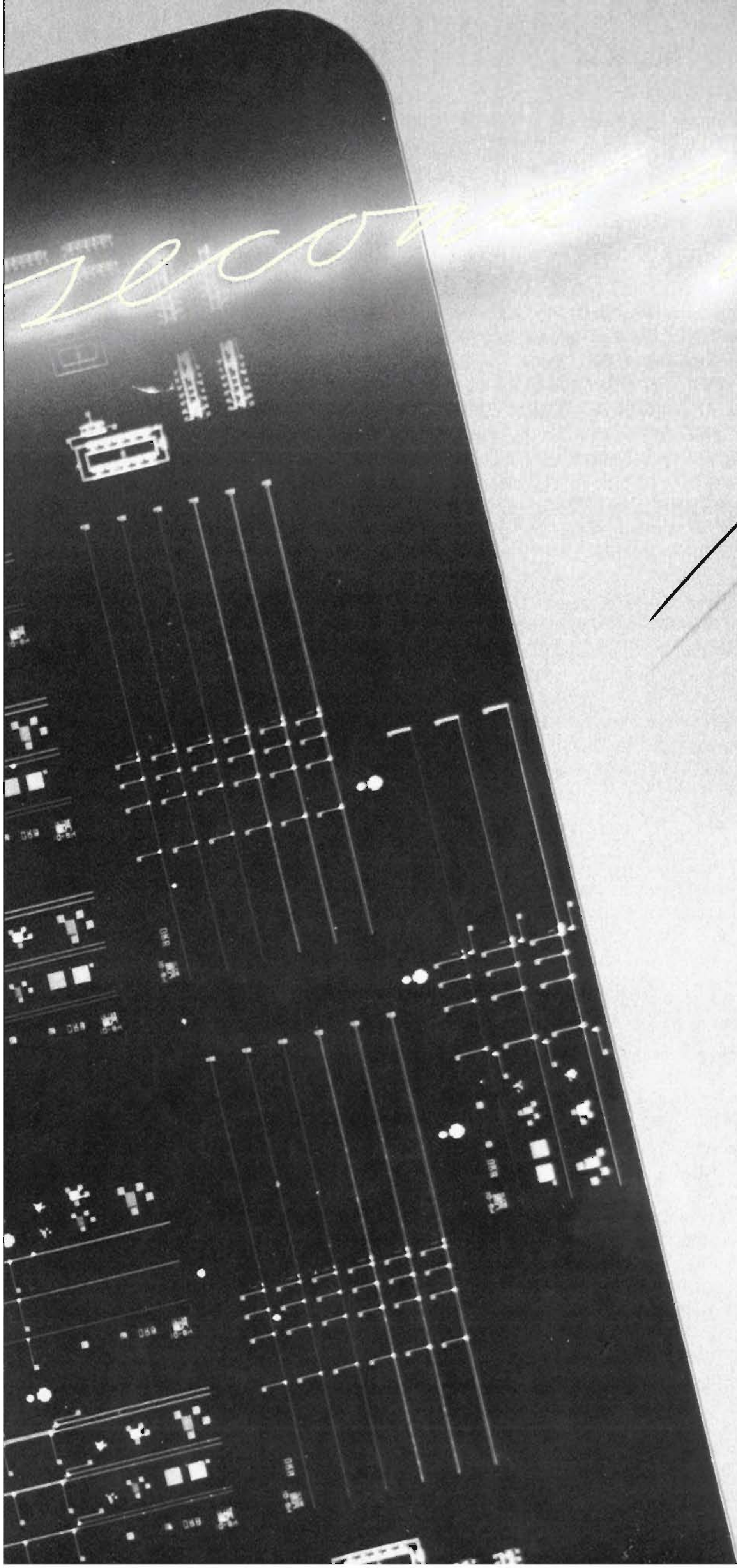
Krusius: Those of us who worked intimately with him over the years have lost a colleague, a mentor, and an example. He was a true experimentalist who found pleasure and inspiration in life from the search for new understanding. He was totally dedicated to this cause. I shall always keep in mind his disciplined and analytical thoughtfulness as he pursued a problem and his excitement and joy as he approached a solution. But beyond John's professional excellence, we have also lost a personal friend with whom we shared thoughts and countless ups and downs in both the laboratory and on bicycle rides. He was a friend who was always available for help and advice at difficult moments.

Swift: What John really liked most was to turn the screwdriver, to make the measurement, to do the whole scientific process himself. The vision of John that I'll always keep in mind is of him sitting on a hard wooden lab stool in front of a bunch of equipment, wearing a plaid shirt and khaki short pants—those great-looking legs of his on display—peering at instruments through his glasses, and writing numbers down in his lab book. ■



photoconductivity and





by Robert B. Hammond

Photoconductivity is by no means an unheard-of phenomenon, having been used for decades as a mechanism for converting light to electrical signals and as a tool for studying the physics of solids. But two recently discovered features of photoconductivity—an extraordinarily rapid onset and, in some materials, a similarly rapid decay—have made it the basis for a new class of electronic devices with exciting applications in diverse areas of research and technology.

The Electronics Division at Los Alamos is a leader in developing the new photoconductive devices. We have demonstrated switches for initiating near billion-watt pulses of electric power in less than a nanosecond (10^{-9} second), detectors for tracking the rapidly varying radiation from fission and fusion reactions, and instruments for studying the behavior of integrated-circuit components during intervals as short as a picosecond (10^{-12} second). Before describing these devices and their applications in detail, a brief review of the history and phenomenology of photoconductivity itself is in order.

A portion of a silicon integrated circuit used in measurements of the velocity, attenuation, and dispersion of high-frequency electrical signals as they propagate along various microstrip transmission lines. Photoconductive pulse generators and sampling gates fabricated within the integrated circuit permit such measurements to be made at frequencies previously inaccessible and with unprecedented accuracy.

As its name suggests, photoconductivity is an increase in the electrical conductivity of a material when it is illuminated by photons or other forms of radiation, such as electrons or alpha particles. The effect was first observed in 1873 in selenium and was soon incorporated in an 1884 patent for a precursor of a television camera. Initially thought to be a rarity, photoconductivity has now been observed in so many materials that it is assumed to occur in nearly all solids, whether nonmetallic or metallic, amorphous or crystalline. In some materials the conductivity increase is large and varies linearly (or nearly so) with light intensity. These materials find practical application in such familiar devices as photographic exposure meters, xerographic copiers, and vidicon television cameras.

With the formulation of the band theory of electronic states in crystals came an explanation for photoconductivity. As illustrated in Fig. 1, photons with appropriate energy excite electrons across the band gap (forbidden energy region) separating the valence band (the highest filled energy band) and the conduction band (the next higher energy band). Since the conduction band offers many unoccupied, very closely spaced energy levels, electrons in that band are free to acquire momentum and energy from an electric field and thus act as current carriers. (The same mechanism, triggered by thermal energy rather than incident radiation, is responsible for the modest conductivity of pure semiconductors and the very low but finite conductivity of pure insulators at temperatures above absolute zero.) Excitation of electrons to the conduction band creates holes (unoccupied energy levels) near the top of the valence band. These holes also contribute to the photocurrent, acting in effect as carriers of positive charge. As one might expect, the photoconductivity of semiconductors and insulators is quite pronounced, whereas that of metals is but a small fraction of their very high inherent conductivity.

The minimum energy required to create an electron-hole pair in a perfect crystal equals the width of the band gap, which is about 1 eV (electron volt) for the semiconductor silicon and about 6 eV for the insulator diamond. (These energies correspond respectively to near-infrared and ultraviolet photons.) However, the defects present in all real crystals, such as lattice vacancies or substitutional impurities, decrease the energy necessary for excitation by creating allowed energy levels within the band gap. Information about the energies of such levels can be deduced from measurements of photoconductivity as a function of photon energy. Photoconductivity has thus proved a valuable tool for research in solid-state physics.

Not until 1975, more than a century after its discovery, was the ultrafast onset of photoconductivity demonstrated, by D. H. Auston of Bell Laboratories. The long delay was not due to oversight. Rather, this feature was simply not observable before the development, in the late sixties and early seventies, of lasers capable of producing very short pulses of light. Auston measured the rise time of the voltage pulse generated across a high-purity, crystalline silicon sample by a 5-picosecond pulse of laser light. The results implied a rise time for the photoconductivity of less than 10 picoseconds. Five years later Auston and his coworkers determined a fall time of about 5 picoseconds for the photoconductivity in a sample of amorphous silicon.

A short rise time is an inherent characteristic of photoconductivity and was suspected long before being demonstrated experimentally. The creation of electron-hole pairs is virtually an instantaneous process and occurs throughout the volume of material illuminated within the very short time required for light to reach its optical absorption depth, or maximum depth of penetration. (For an optical absorption depth of 1 micrometer, a fairly typical value, the penetration time is about 10 femtoseconds, or 0.01 picosecond.) Thus almost no delay need be involved in establishing a conductive path across a photoconductor. In contrast, the introduction of carriers into a conventional semiconductor device (such as a transistor or diode) is localized at junctions between *p*- and *n*-type regions, and establishing a conductive path across the device involves a relatively slow process, the physical transport of carriers under the influence of an applied voltage (the bias voltage).

The decay of photoconductivity with cessation or decrease of illumination is not inherently rapid since the 'death' of carriers, unlike their creation, is far from instantaneous. (Still, however, no carrier-transport time is involved.) The number of carriers decreases with time by a process known as recombination, the mutual annihilation of an electron and a hole. The time required for the number of carriers to decrease by a factor of e is called the carrier lifetime. This parameter is usually determined by the number and nature of defects in the material, which determine the number and energies of levels within the band gap.

The lifetime of carriers in a crystalline material usually increases with crystal perfection. Carrier lifetimes in common high-purity, crystalline semiconductors are typically on the order of nanoseconds to milliseconds but can be decreased by introducing defects or impurities into the lattice, although often at the expense of other desirable properties.

Demonstration of the fast onset and the potential for fast decay of photoconductivity piqued the interest of many who saw the need for better methods of rapidly initiating electrical signals or of tracking rapidly varying radiation or electrical signals. New photoconductive devices are now helping to fill that need, which arises in activities as disparate as accelerating particles, testing nuclear weapons, and investigating nonequilibrium transport phenomena in semiconductors. The new devices described below are those developed at Los Alamos, some in collaboration with other institutions, and by no means exhaust the possibilities.

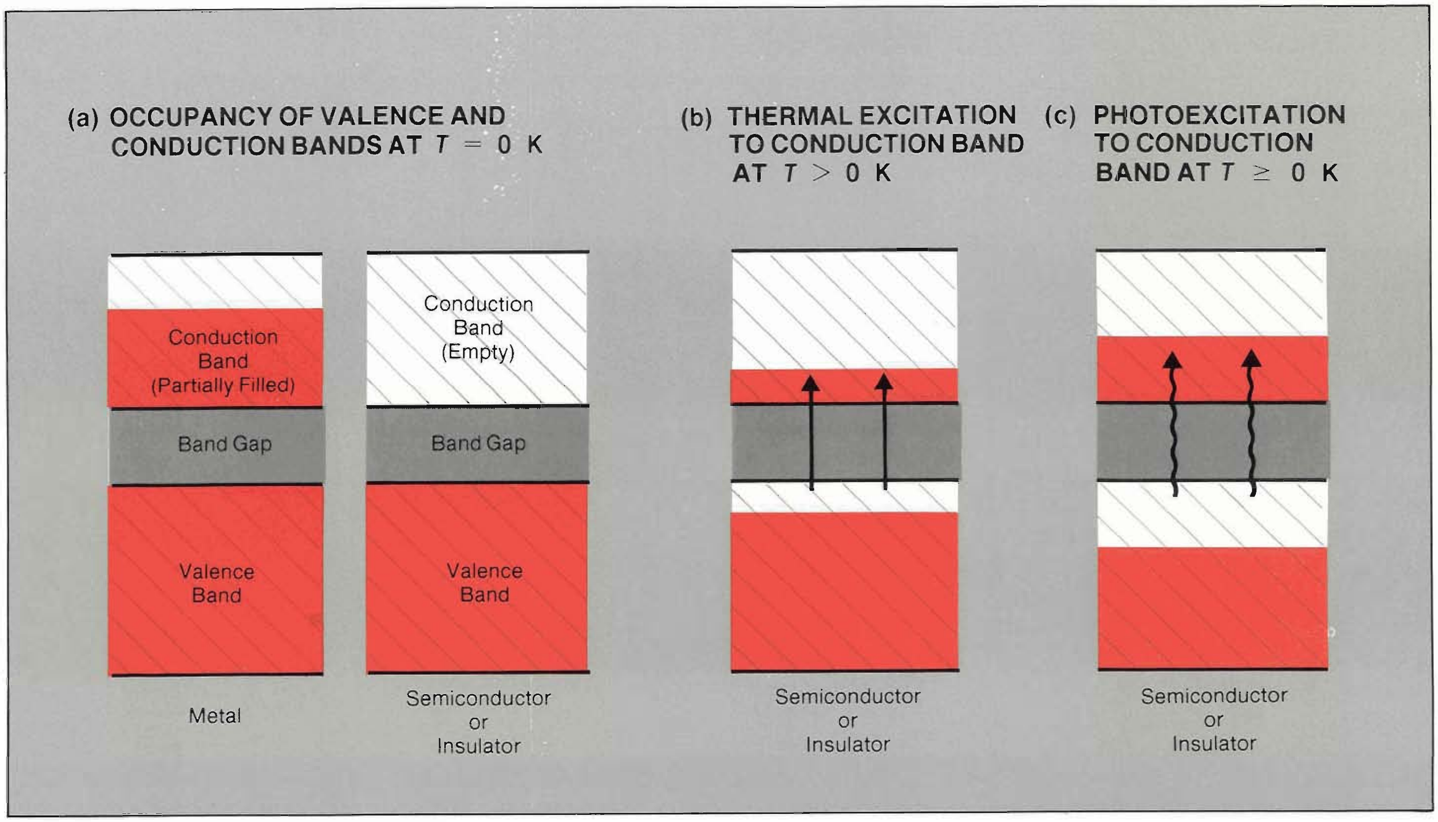


Fig. 1. The band theory of electronic energy levels in solids provides an explanation for electrical conductivity in general and for photoconductivity in particular. According to that theory the energy-level structure consists of bands of very closely spaced allowed energies (hatched regions) separated by gaps of forbidden energies (dark gray regions). The intrinsic electrical conductivity of a crystalline material depends on the occupancy of the uppermost bands, which in turn depends on the electronic configuration of the atoms constituting the material and on the structure of the crystal. Shown in (a) are the two possibilities for the occupancy of the uppermost bands in a crystalline material at absolute zero. In one case the highest nonempty band is only partially occupied, as indicated by red (occupied) and white (unoccupied) regions. Available to the electrons within this conduction band are many unoccupied levels of only slightly higher energy and momentum. A modest electric field can cause transitions to those levels, thereby creating a net momentum and a flow of current. Such a material is called a metal. In the other case the highest nonempty band is fully occupied, and the nearest levels available to the electrons within this valence band lie above the band gap in the conduction band. A modest electric field cannot supply the energy necessary to bridge the band gap, and no current flows. Such a material is called a semiconductor or an insulator depending on the width of the band gap. (b) The conductivity of semiconductors and insulators at temperatures greater than absolute zero is low but nonzero since some electrons possess sufficient thermal energy to bridge the band gap. This thermal excitation accounts for the strong temperature dependence of the conductivity of semiconductors and insulators. (c) Photons and other forms of radiation can also supply the energy necessary to bridge the band gap and cause the increase in conductivity known as photoconductivity. The effect is illustrated for semiconductors and insulators, which exhibit a much greater relative increase in conductivity than do metals. The photoconductivity of many materials is directly proportional to the intensity of the incident light. Note that unoccupied energy levels, or holes, are created near the top of the valence band when electrons are excited to the conduction band, whether by thermal energy or photons. Occupation of these holes makes a contribution to the current, a contribution identical to that expected of carriers with positive charge. The energy-level structures depicted are simplified in the sense that the band gap is completely devoid of allowed energy levels, a situation that obtains only in the ideal of a perfect crystal.

Ultrafast Photoconductor Power Switch

By the late seventies a number of researchers had used the combination of a photoconductor and a pulsed laser to generate electrical pulses with very short rise times. Currents up to 100 amperes at voltages up to 10 kilovolts were reported. Some back-of-the-envelope calculations on our part indicated that much higher currents at much higher voltages might be attained, without sacrificing short rise times, simply by increasing the dimensions of the photoconductive volume. This finding led naturally to the idea of a photoconductor power switch, a device for producing high-power, short-rise-time, accurately timed pulses of electricity. Other hopes for the switch included relatively long duration of the output pulses and economy of operation.

The photoconductor power switch is basically very simple in design (Fig. 2). It consists of a pulsed laser and a small volume of photoconductive material connected to a source of voltage (the operating voltage) and to the load. The electrical connections to the voltage source and the load should have a high injection efficiency; that is, they should permit efficient replacement of the electrons swept out of the photoconductor during its on state. To help realize a short rise time and also prevent energy loss due to electromagnetic radiation, the photoconductor should be incorporated in a transmission line (for example, a coaxial or microstrip transmission line).

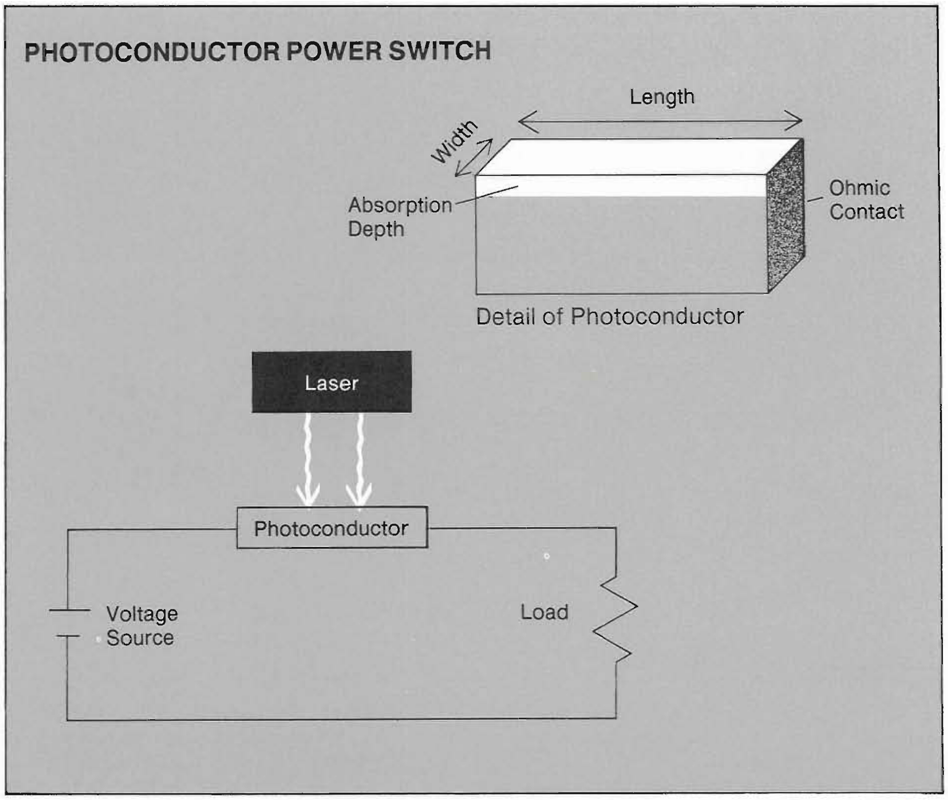


Fig. 2. A small volume of photoconductive material of relatively high intrinsic resistivity, a high-voltage source, and a pulsed laser are the basic components of a photoconductor power switch. When the photoconductive material is illuminated by a pulse of photons from the laser, its conductance rapidly increases and a short-rise-time current pulse flows through the load. In the absence of a laser pulse, the photoconductive material acts in effect as an open switch. The ohmic contacts to the photoconductive material permit efficient replacement of the carriers swept out of the active volume during its conducting state. In practice the photoconductive material is incorporated into a coaxial transmission line to help achieve a short rise time and to prevent energy loss due to electromagnetic radiation. As discussed in the text, the power capacity of the switch can be increased simply by increasing the length and width of the photoconductive material.

For our proof-of-principle experiments we chose the combination of a neodymium-glass laser and very pure, crystalline silicon. A Nd:glass laser is readily available and inexpensive and can be operated to produce high-intensity optical pulses. A semiconductor is preferable to an insulator as the photoconductive material because electron-hole pairs can be created in a semiconductor by the infrared photons from the Nd:glass laser, and yet the dark resistivity of a semiconductor is sufficiently high that the current through the switch in its off state would be very low. The choice between two strong candidates for the semiconductor, gallium arsenide and silicon, was based on consideration of the properties listed in the accompanying table. Gallium arsenide offered the advantages of higher dark resistivity and higher carrier mobility. Higher carrier mobility implies higher carrier velocity for a given operating voltage, which in turn implies higher current per carrier. We chose silicon rather than gallium arsenide, however, because of its greater absorption depth for the 1.06-micrometer photons from the Nd:glass laser and, most important, its longer carrier lifetime. A greater absorption depth permits excitation of a larger cross-sectional area and thus generation of higher currents. A longer carrier lifetime is an economic advantage, permitting produc-

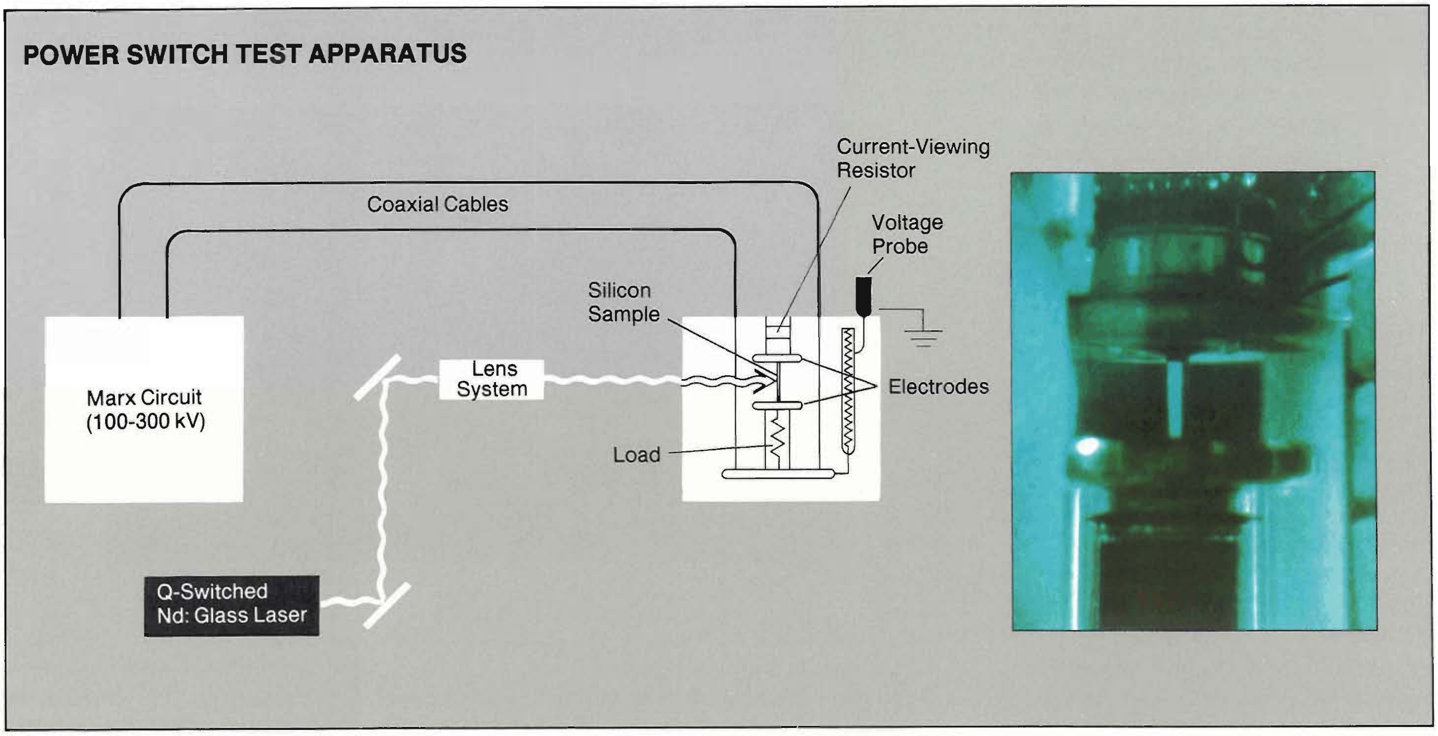


Fig. 3. The performance of high-resistivity, crystalline silicon in a photoconductor power switch was evaluated with an apparatus including a Marx circuit for supplying the operating voltage, a low-inductance current-viewing resistor for measuring the current through the switch, and a capacitance-compensated voltage divider for measuring the voltage across the load. The impedance of the load, a copper sulfate solution, was adjusted to match that of the coaxial cables. Good electrical contact between the silicon sample and two brass electrodes was ensured by a spring arrangement. The entire test assembly at the end of the coaxial cables was immersed in a dielectric fluid to prevent high-voltage arcs between circuit points. The photograph shows a portion of the apparatus, including the silicon sample, the brass electrodes, and the copper sulfate load solution.

Table

The choice of photoconductive material for a particular application is usually a compromise based on a number of properties. Listed below for silicon and gallium arsenide are properties of importance to a photoconductor power switch activated by 1.06-micrometer photons from a Nd:glass laser. An absorption depth for gallium arsenide is not listed because 1.06-micrometer photons are not sufficiently energetic to excite carriers to the conduction band in pure gallium arsenide. Thus the absorption depth is extrinsic, varying from sample to sample with the concentration of impurities or defects.

Property	Silicon	Gallium Arsenide
Carrier mobility	1925 $cm^2/V \cdot s$	9300 $cm^2/V \cdot s$
Carrier lifetime	1 ns to 1 ms	0.1 to a few ns
Intrinsic resistivity at 300 K	0.23 $M\Omega \cdot cm$	100 $M\Omega \cdot cm$
Absorption depth for 1.06-μm photons	1 mm	---

tion of a satisfactorily square current pulse for a longer time without repeated excitation by the laser.

What physical and material properties determine the power capacity (the operating voltage and current) of a silicon power switch? Although the voltage gradient across the silicon during its off state must not exceed its dielectric strength (about 100 kilovolts per centimeter), the operating voltage can be increased simply by increasing the length of the photoconductive volume (the distance between the contacts). Similarly, although the carrier density in the silicon must not exceed about 10^{18} per cubic centimeter (or, equivalently, the current density (at room tem-

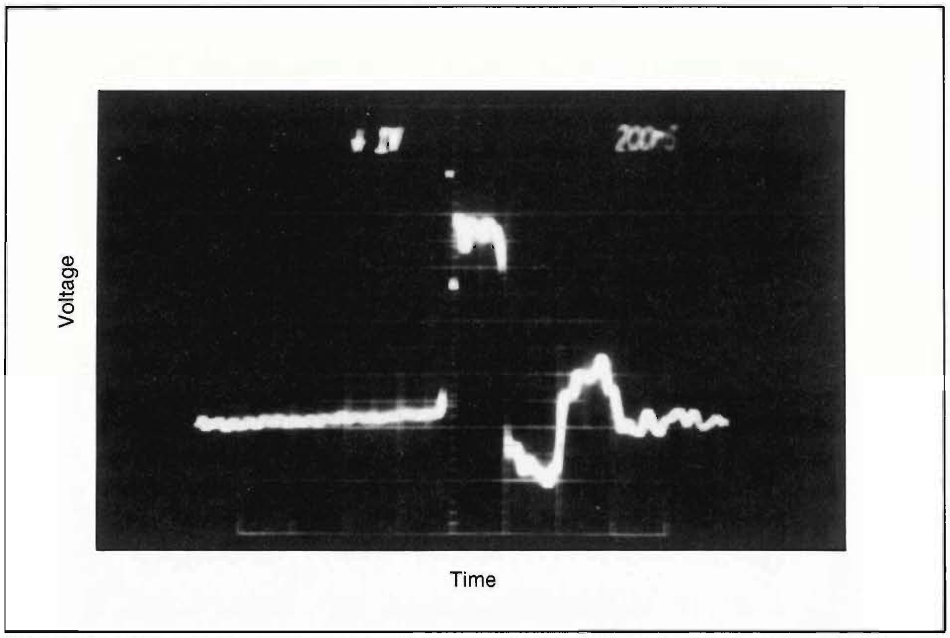


Fig. 4. Oscilloscope trace of a current pulse generated by a 0.5- by 0.5- by 2.5-centimeter bar of high-resistivity silicon in the apparatus of Fig. 3. The current pulse was induced by a 10-millijoule, 20-nanosecond laser pulse. Note the rapid rise of the current to its peak value of about 1.8 kiloamperes. This current, at the operating voltage of 170 kilovolts, corresponds to a switched power of about 150 megawatts. The duration of the pulse is about 200 nanoseconds.

perature in an electric field of 1 kilovolt per centimeter) must not exceed about 300 kiloamperes per square centimeter), the current through the silicon can be increased simply by increasing the cross-sectional area of the photoconductive volume. Since the depth of the cross-sectional area is fixed at the optical absorption depth of silicon, the width is the dimension varied to accommodate a desired operating current. This ease of scaling to arbitrary power capacities is due to the bulk nature of photoconductivity.

The limit on the carrier density in silicon, and ones of similar magnitude for other semiconductors, is necessary to minimize two phenomena that would adversely affect the performance of the switch. At carrier densities above 10^{18} per cubic centimeter, the rate of Auger recombination and the magnitude of free-carrier absorption increase rapidly. Auger recombination decreases the fall time of the photoconductivity and thus the duration of the current pulse; absorption of laser light by free carriers produces no additional carriers and thus wastes laser energy.

The rise time of the current pulse is determined by two factors: the inherent rise time of the photoconductivity and the inductance of the circuit. The low inductance necessary for a short rise time can usually be achieved by increasing the width of the photoconductive material. Therefore high currents and fast rise times are compatible goals.

Although some applications envisioned for the photoconductor power switch require only isolated current pulses, others require a periodic sequence of pulses. The rate at which the pulses can be repeated is limited to between about 10 and 1000 times per second. This limit is necessary to avoid thermal runaway, a mounting cycle of resistive heating, increased thermal generation of carriers, increased current, increased resistive heating, and so on, that culminates in complete conduction and melting of the switch. Were it not for the high specific heat and thermal conductivity of silicon, the limit on the repetition rate would be even lower.

We tested a number of switches, using the apparatus shown in Fig. 3, and the results confirmed our expectations for the potential of the device. For example, with one switch fashioned from a 2.5- by 0.5- by 0.5-centimeter bar of 1000-ohm-centimeter silicon and illuminated by a 20-nanosecond (full width at half maximum) laser pulse, we obtained a peak switched current of about 1.8 kiloamperes at an operating voltage of about 170 kilovolts (Fig. 4). These values for the current and voltage correspond to a peak switched power of about 150 megawatts. The rise time of the current pulse was approximately 5 nanoseconds; its duration was about 200 nanoseconds.

Our tests revealed nothing to prevent operation of a photoconductor power switch at voltages and currents approaching the maxima imposed by the limits on voltage gradient and carrier density mentioned above.

The major advantages of the photoconductor power switch over present high-power switching technologies are greater ease of scaling, output pulses with much shorter rise times, and independent optical control. Independent control eliminates interference from the switching circuit itself, and optical control permits greatly increased accuracy in the timing of a succession of pulses from one switch or of a temporal sequence of pulses from an array of switches. Other advantages include small size, simplicity, and efficiency at transforming optical energy to electrical energy. (The test switch mentioned above, which received an incident laser energy of 10 millijoules, transferred about 30 joules to the load.) The overall efficiency of the device is limited, however, by the inefficient production of optical energy by the lasers now commercially available. That

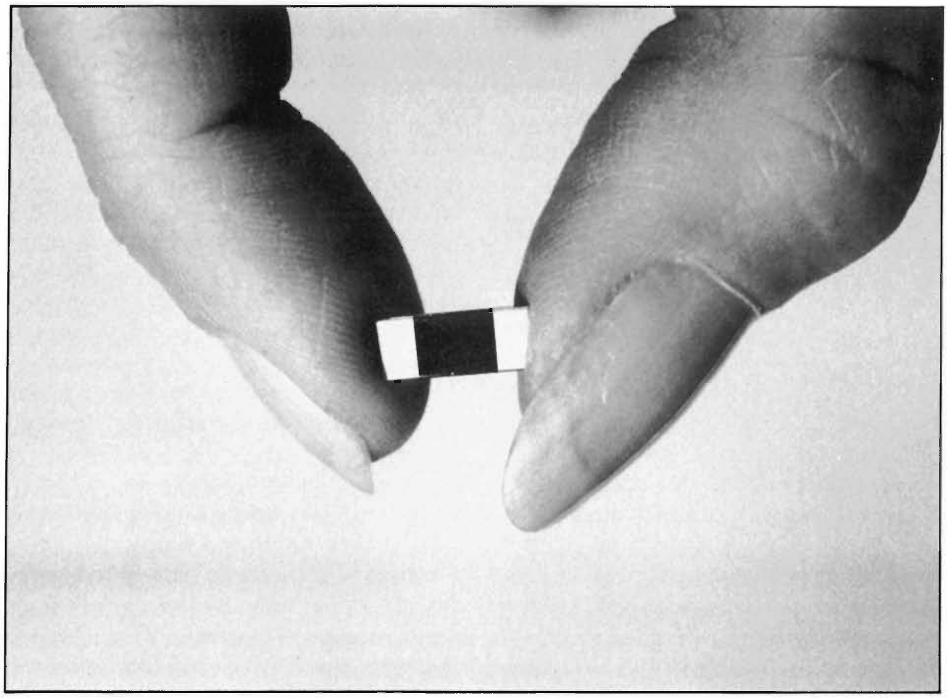


Fig. 5. One of the great advantages of a photoconductor radiation detector is the size of the active volume, which is many times smaller than that of other detectors with comparable sensitivity. The photograph shows the active volume of one of the larger of our InP:Fe detectors.

fault will soon be remedied since Nd:YAG (yttrium aluminum garnet) lasers with efficiencies up to 40 percent have recently been demonstrated.

Among the possible applications of photoconductor power switches are many systems that require high-power, short-rise-time, accurately timed pulses of relatively short duration (less than about 1 microsecond) at relatively low repetition rates (less than a few hundred per second). Such systems include particle accelerators, lasers for initiating thermonuclear fusion, devices for simulating nuclear-weapons effects, and directed-energy weapons.

Applications requiring longer pulses must await development of efficient methods for maintaining the carrier density at the desired level. One promising approach is to initiate the electrical pulse with a laser pulse and to maintain the carrier density with a beam of electrons. (We have shown that an electron beam can efficiently generate carriers in a photoconductor.) This approach may find applications in systems for producing ac power. Potential uses extend from very-low-voltage systems, systems in which the output voltage is less than the typical *p-n* junction drop (about 0.5 volt), to the very high-voltage systems that condition electrical power for long-distance transmission.

Ultrafast Photoconductor Radiation Detector

Assessing the results of, say, a laser-fusion experiment or a nuclear-weapon test involves diagnostic measurements on very unusual radiation events, events in which the intensity of the radiation varies extremely rapidly over an extremely wide range. Those measurements require equally unusual radiation detectors, detectors with ultrashort response times and high but constant sensitivities. We have developed a new class of detectors with just such properties, capitalizing not only on the rapid onset of photoconductivity but also on its potential for rapid decay.

(I point out immediately that our photoconductor radiation detectors do not count individual quanta of radiation nor, per se, provide information about the energies of the quanta. They are thus no substitutes for conventional semiconductor radiation detectors, such as lithium-doped germanium or silicon junction detectors and silicon surface-barrier detectors.)

The response time of a radiation detector is a measure of how well it can resolve rapid variations in intensity; it may be defined as the full width at half maximum of the current pulse induced in the detector by a very short radiation pulse. The response time of a conventional semiconductor radiation detector is shortened by decreasing the distance between the electrical contacts and thus the transit time for carriers. To prevent electrical breakdown, the decrease in contact spacing is accompanied by a corresponding decrease in applied voltage. But since the applied voltage and the magnitude of the current through the detector are directly related, a decrease in applied voltage produces a decrease in sensitivity (ratio of output signal to radiation intensity). Therefore short response time and high sensitivity are incompatible goals for a conventional semiconductor radiation detector. In contrast, the response time and sensitivity of a photoconductor radiation detector are much less closely coupled.

A photoconductor radiation detector is even simpler in design than a photoconductor power switch, since the radiation being studied replaces a laser as the carrier-excitation agent. In essence the detector consists of a small volume of photoconductive material (Fig. 5) with contacts to a voltage source and to a device, such as a fast os-

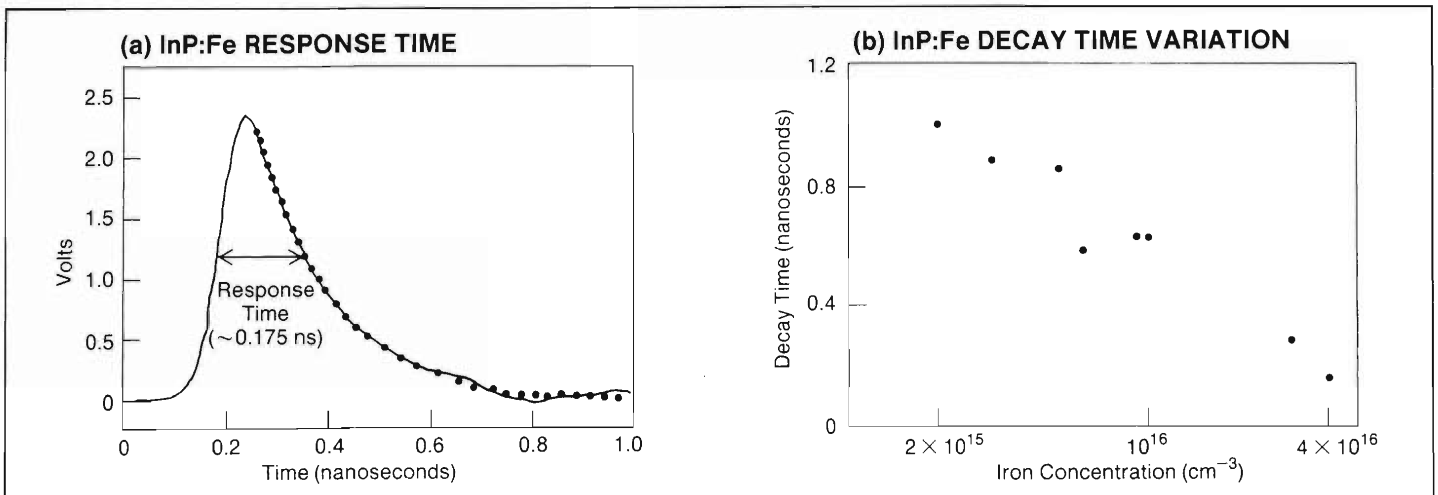


Fig. 6. (a) Results of a response-time measurement on one of a series of iron-doped indium phosphide detectors. The detector was activated by a 15-picosecond laser pulse; the resulting voltage pulse was measured with a sampling oscilloscope. Both the circuit-limited rise time and the decay time of the carriers contribute to the response time. (The decay time, a component of the carrier lifetime, is a measure of the rate at which electrons are removed from the conduction band by being trapped at iron sites. Recombination occurs when a trapped electron and a hole annihilate.) The decay of the pulse was fitted with an exponential (solid circles) corresponding to a decay time of 155 picoseconds. (b) The decay time, and hence the response time, decreases with iron concentration. However, decay times less than about 100 picoseconds cannot be attained because of the limited solubility of iron in indium phosphide.

cilloscope, for recording the photocurrent, which is a measure of the intensity of the radiation. The photoconductive material is integrated into a coaxial or microstrip transmission line, just as it is in the photoconductor power switch.

The key to a short response time for the detector is a short carrier lifetime. As mentioned above, the lifetime of carriers in a crystalline photoconductor can be reduced by introducing impurities or defects into the crystal lattice. The challenge is to achieve a suitably short lifetime without unduly decreasing the resistivity or carrier mobility, both of which should be as high as possible. Another desirable property of the photoconductive material is a linear variation of photocurrent with radiation intensity (that is, a constant sensitivity), since a nonlinear variation complicates analysis of the data.

We considered a number of photoconductive materials and found two with an attractive combination of lifetime, resistivity, and mobility: iron-doped indium phosphide (InP:Fe) and neutron-damaged gallium arsenide. Of these materials we have investigated InP:Fe the most extensively, both in laboratory experiments and in the field.

Using a variety of pulsed radiation sources, we measured the response times and sensitivities of a number of InP:Fe detectors and found them to be superior to those of other fast detectors, such as photodiodes or Compton-electron detectors. The detectors were fabricated from semi-insulating indium phosphide containing concentrations of iron ranging from 0.2×10^{16} to 4×10^{16} atoms per cubic centimeter. The sensitivity of the detectors to gamma rays is about 10^{-8} coulombs per rad for radiation events lasting less than about 10 nanoseconds. (For longer radiation events the photocurrent varies nonlinearly with radiation intensity.) The response time of the InP:Fe detectors (Fig. 6) decreases with increasing iron concentration, but unfortunately the solubility of iron in indium phosphide imposes a lower limit on the response time of about 100 picoseconds.

InP:Fe detectors can detect and image gamma rays, hard and soft x rays, and charged and neutral particles. (Neutrons, however, must be converted to protons before being detected.) The response of InP:Fe to soft x rays (those with energies less than a few keV) is surprising because radiation of that type has such a short absorption depth (about 10 nanometers) that it does not penetrate beyond an inactive layer present on the surface of most materials. This unusual feature of InP:Fe suggested the possibility of its application to diagnostic measurements on beams of synchrotron radiation, which are used for basic research in, for example, plasma physics and for high-resolution lithography. We tested this possibility by evaluating the performance of the detectors at the Stanford Synchrotron Radiation Laboratory. Our initial studies indicated that the sensitivity of the detectors to x rays was constant over

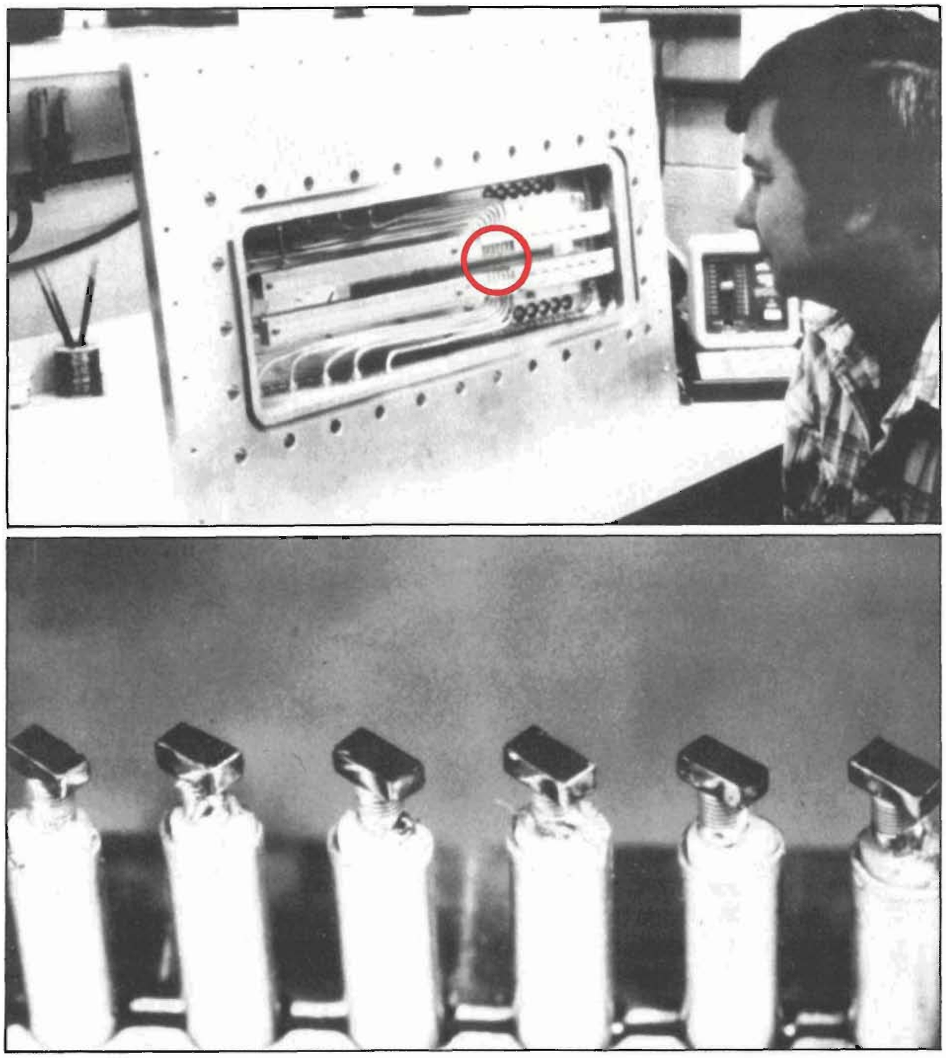


Fig. 7. A charged-particle spectrometer composed of an array of InP:Fe radiation detectors. The close-up shows the individual detectors. An electromagnet (not pictured) deflects particles with different energies to different elements of the array. Such spectrometers have provided time- and energy-resolved images of the radiation produced by nuclear weapons. ◀

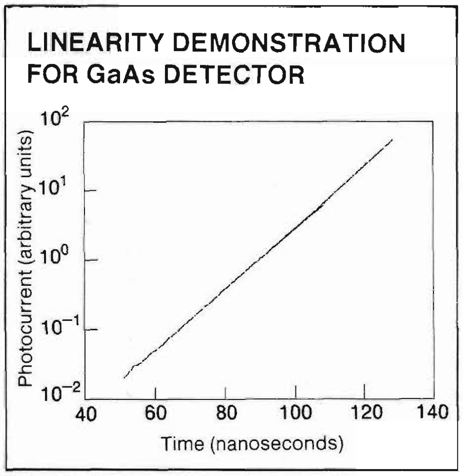


Fig. 8. This graph depicts the photocurrent generated by a gallium arsenide detector when exposed to a radiation source whose intensity varies exponentially with time. Note the nearly perfect linearity over an intensity range of 10^4. Such linearity is a very desirable feature of a radiation detector. ▲

the energy range 0.2 to 3 keV. More recently we have found that InP:Fe is uniformly sensitive to *all* x-ray photons. In contrast, the sensitivity of the vacuum photodiodes commonly used for beam diagnostics decreases with x-ray energy.

We have also developed charged-particle spectrometers in which particles with different energies are deflected by a magnetic field to different elements of an array of InP:Fe detectors (Fig. 7). Such spectrometers provide time- and energy-resolved images of radiation events.

Our InP:Fe detectors have proved their worth during tests of nuclear weapons at the Nevada Test Site by supplying previously inaccessible information. In addition, they hold promise of permitting a much wider variety of diagnostic experiments. The Naval Surface Weapons Center at White Oak, Maryland, also has used our InP:Fe detectors in the field for diagnostic measurements on the bremsstrahlung from pulsed electron beams that simulate the effects of nuclear weapons.

Our development efforts on photoconductor radiation detectors have concentrated recently on neutron-damaged gallium arsenide, a material that offers two major advantages over InP:Fe. First, its response time, which is not subject to a solubility limit, can be as short as 1 picosecond. Second, its sensitivity, which is similar in magnitude to that of InP:Fe, is constant over an exceptionally wide range of intensities (Fig. 8) and for radiation events lasting as long as about 100 nanoseconds.

Undoubtedly, other applications, and other materials with properties specific to those applications, will be found for this new class of detectors.

Ultrafast Electrical Measurement Device

The first integrated circuit, demonstrated in 1958, contained five components (a transistor, a capacitor, and three resistors) fashioned by modifying the electrical properties of discrete regions within a chip of semiconducting material no larger than a match. Today a chip of similar size contains millions of components. Such astounding densities—and the resulting signal speed—are the basis for many electronic marvels, including the single-chip computer microprocessor.

However, to take full advantage of the speed possible now, and even greater speed promised in the future, circuit designers need a new body of information: the response of components to signals within times as short as a few picoseconds. (The term 'response' here refers to the amplitude and phase changes effected in an electrical signal by a component.) This information is lacking because no measuring device with sufficient temporal resolution has been available. (The temporal resolution of a sampling oscilloscope, for example, is at best about 25 picoseconds, being limited by the on time of its sampling gate and by jitter in the timing of its pulse generator and sampling gate.)

Building on early ideas of D. H. Auston, we have now demonstrated that the missing information can be provided by a new device composed of two photoconductive circuit elements, one serving as a pulse generator and the other as a sampling gate, both fashioned on chip and activated by subpicosecond laser pulses (see "The World's Fastest Laser Oscillator"). The most significant feature of the new device is the generation and sampling of high-frequency signals *on chip*, that is, *within* the circuit containing the component being tested. This feature eliminates two formidable tasks: the transport of high-frequency signals into the circuit from an external generator and their transport out

continued on page 60

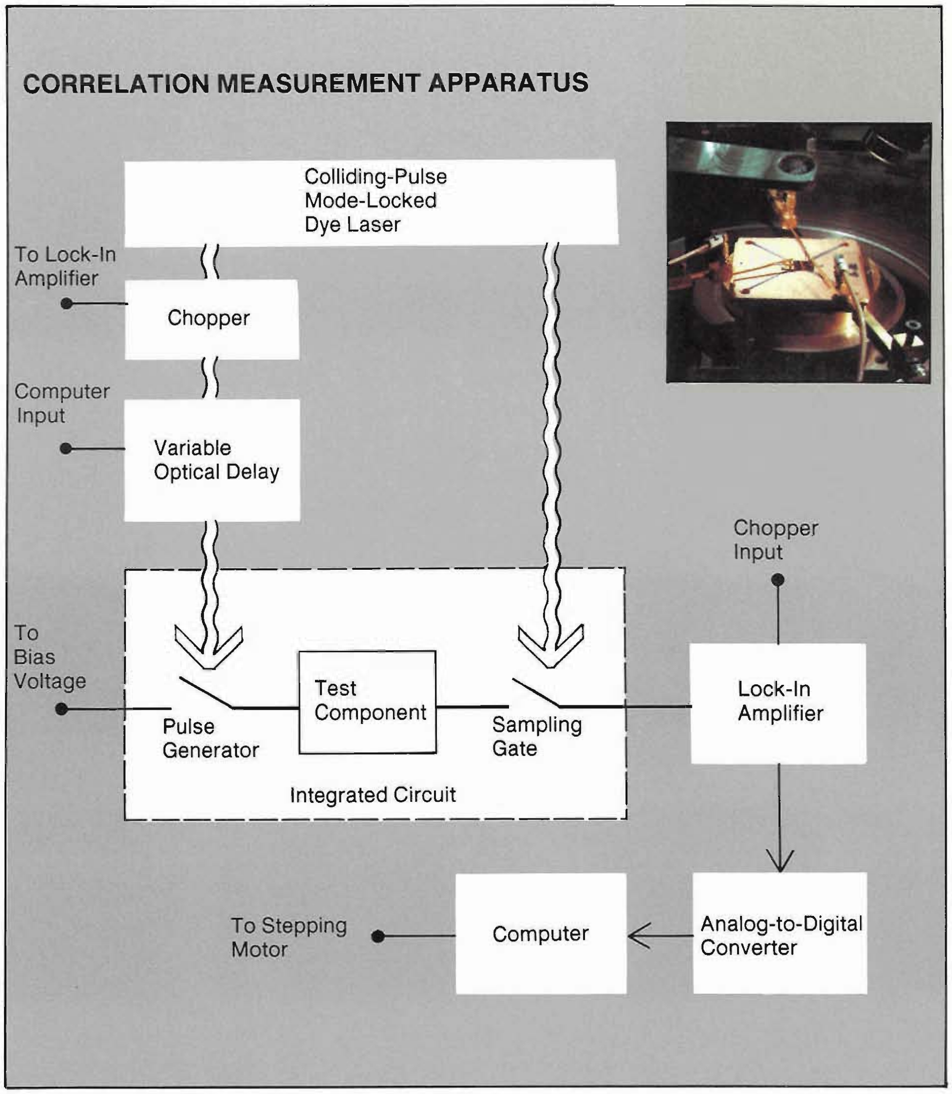


Fig. 9. Schematic diagram of an experimental setup for determining the response of an integrated-circuit component with photoconductive circuit elements (a pulse generator and a sampling gate) fabricated on chip (that is, as an integral part of the circuit). The laser produces two synchronous trains of subpicosecond pulses, one of which activates the biased pulse generator and the other the sampling gate. The pulse generator produces a current pulse, which propagates through and is modified by the component being investigated. The sampling gate feeds the modified current pulse to the lock-in amplifier during an interval (the sampling aperture) that is short compared to the duration of the pulse. The amplifier measures the average amplitude of the current pulse during that interval, which corresponds to a certain small portion of the pulse cycle. The relative timing of the pulse generator and the sampling gate is then varied (by varying the path length of one laser pulse train with a corner reflecting cube mounted on a stepping-motor driven stage), and the amplitude measurement is repeated. The result is an amplitude versus delay curve (a correlation function) from which the response of the component can be extracted. The mechanical chopper increases the signal-to-noise ratio of the amplitude measurements by imposing a frequency of 808 hertz on the laser pulse train that activates the pulse generator. The insert is a photograph of the platform built to facilitate the correlation function measurements. The integrated circuit containing the component being investigated is mounted in the center of the platform, the two laser pulse trains are focused onto the photoconductive circuit elements through two microscope objective lenses (one of which is visible at the top of the photograph), and three coaxial probes (extending from brass holders) couple the bias voltage to the pulse generator and the signals from the sampling gate to the lock-in amplifier.

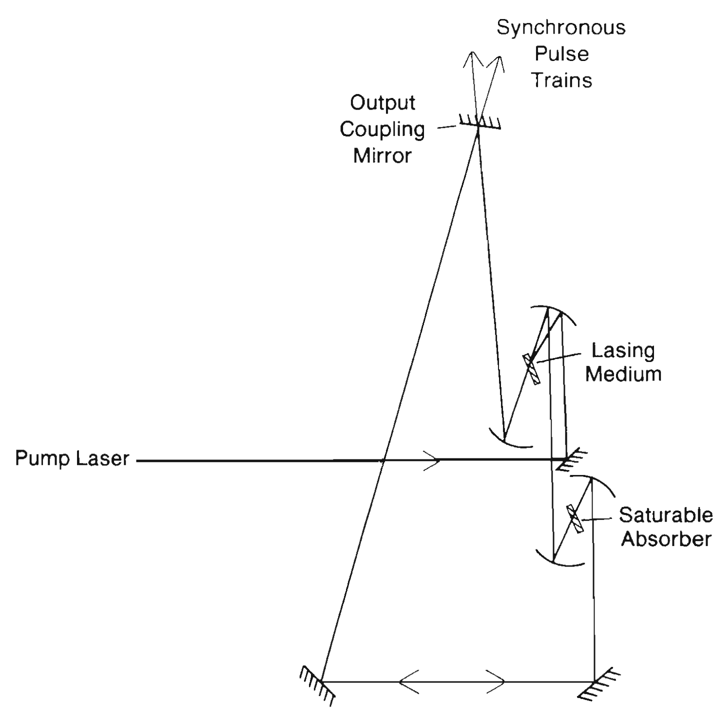
The World's Fastest Laser Oscillator

How short an electrical pulse can be generated by a photoconductor in response to an optical pulse? This question has fundamental significance as well as practical implications. Obtaining the answer requires that the photoconductor be excited by an optical pulse whose duration is short compared to the time scale of carrier decay. Optical pulses of the requisite brevity are produced by a laser first demonstrated in 1981 by R. L. Fork, B. I. Greene, and C. V. Shank of Bell Telephone Laboratories. This so-called CPM (for colliding-pulse mode-locked) laser is illustrated schematically in the accompanying figure, and the accompanying photograph shows the CPM laser built for our research on ultrafast photoconductive circuit elements.

The lasing medium (an organic dye, rhodamine 6-G) is pumped by a continuous-wave argon-ion laser. Pulses of light from the lasing dye travel in both directions through the laser cavity along a roughly triangular path that includes a saturable absorber (another organic dye, 3,3′-diethyloxadicarbocyanine iodide). Interaction of the counterpropagating pulses with the saturable absorber causes a locking in phase of many resonant cavity modes (mode locking). The result is a succession of relatively high-intensity pulses separated by the time required for light to traverse the cavity (about 10 nanoseconds). Two synchronous trains of pulses are extracted from the cavity through the output coupling mirror.

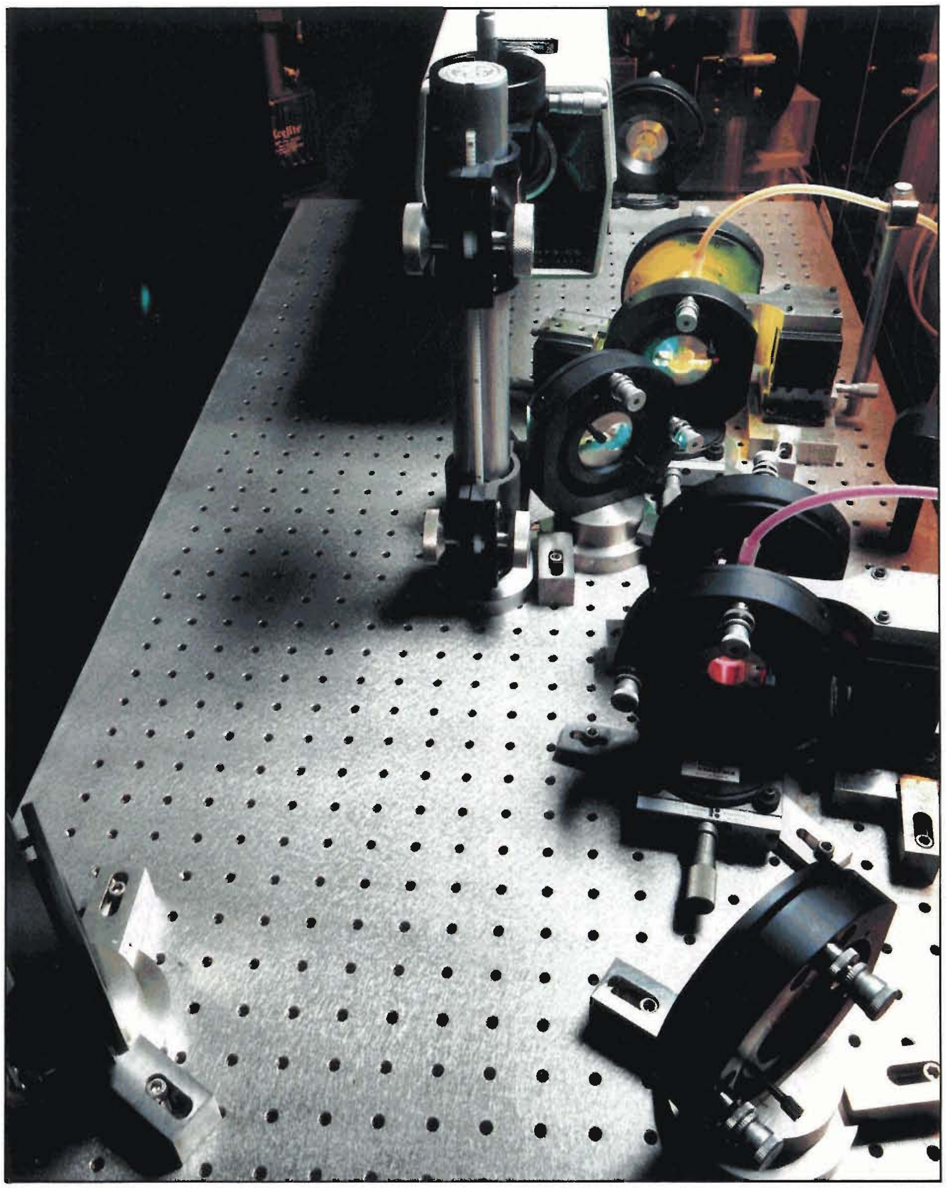
The first CPM laser produced 0.09-picosecond pulses; later versions with prisms in the cavity to compensate for dispersion from the laser mirrors have produced 0.027-picosecond pulses—the shortest available today.

CPM LASER SCHEMATIC



Schematic diagram of a CPM laser. The focusing mirrors for the lasing medium and the saturable absorber have radii of about 10 and 5 centimeters, respectively. The overall cavity length is about 300 centimeters.

A natural question to ask is how the duration of such short pulses can be determined. The standard technique is one known as autocorrelation by second-harmonic generation. A beam of pulses from the laser is split by a beam splitter, one of the resulting beams is fed through a variable optical delay, and both are then focused on a potassium dihydrogen phosphate (KDP) crystal. Nonlinear interaction of two out-of-phase but otherwise identical pulses produces a second harmonic whose maximum amplitude is a



Photograph of the CPM laser built by the Laboratory's Electronic Research and Exploratory Development Group. The lasing medium (yellow) and the saturable absorber (purple) circulate through the tubes at the right, the pump laser is visible at the top, and the output coupling mirror is in the far background.

function of the delay between the two original pulses. Measurement of that amplitude as the delay is varied yields a correlation function (an autocorrelation function, since the two pulses are identical) from which the duration of the pulses is derived. This technique is analogous to that by which the response of an integrated-circuit component is determined from measurements of signal amplitude versus the delay between activation of a photoconductive pulse generator and a photoconductive sampling gate (see main text). ■

continued from page 58

of the circuit to an external sampling instrument. These tasks have in fact proved nearly impossible for signals with frequencies above about 25 gigahertz. (Frequency-domain measurements with a 3-decibel bandwidth of about 25 gigahertz correspond to time-domain measurements with a resolution of about 15 picoseconds.)

Figure 9 shows schematically an experimental setup for investigating a circuit component with the device. Its operation is in many respects similar to that of a sampling oscilloscope. Briefly, a laser pulse induces the biased pulse generator to produce an electrical signal. This signal passes through the component, being modified in the process by its response. The sampling gate, activated by a second laser pulse, feeds the signal during a short interval (the sampling aperture*) to external circuitry that measures its average amplitude during that interval. By varying the relative timing of the two laser pulses, an amplitude versus delay curve known as a correlation function is obtained. Embedded within this correlation function is the response of the component, together with the responses of the pulse generator, the sampling gate, and the interconnections. Extracting the component response requires knowledge or reasonable estimates of the other responses.

The temporal resolution of the device is determined by the sampling aperture, which in turn is determined by the lifetime of carriers in the sampling gate. Thus short carrier lifetime is the key property for the sampling gate, just as it is for the photoconductor radiation detector. In addition, the material composing both photoconductive circuit elements should have high resistivity and high carrier mobility. Furthermore, for greatest utility

**The sampling aperture of a photoconductive sampling gate is defined as the full width at half maximum of the pulse induced in the gate by an ultrashort optical pulse. It is essentially identical to the property defined in detector parlance as response time.*

PHOTOCONDUCTIVE INTEGRATED-CIRCUIT ELEMENTS

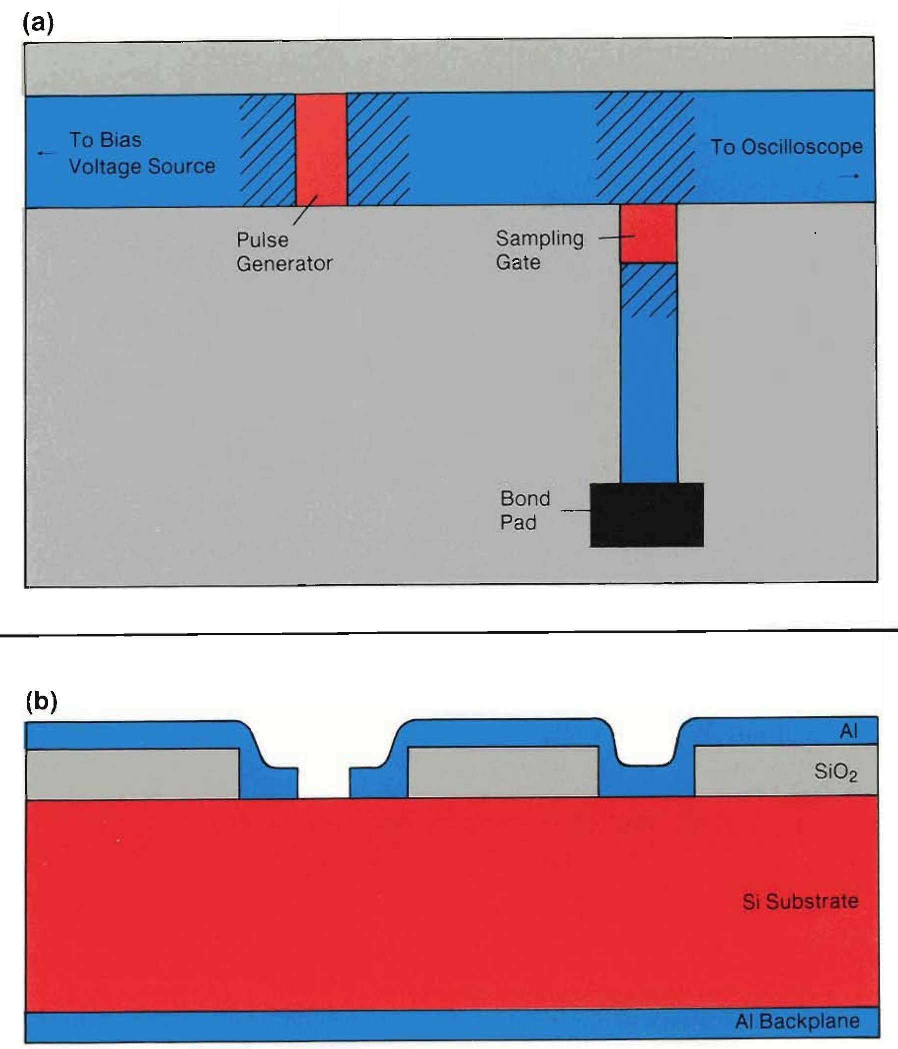


Fig. 10. (a) Top view (schematic) of a portion of the silicon integrated circuit used to determine the characteristics of photoconductive circuit elements (pulse generators and sampling gates) fabricated within the silicon substrate. The pulse generator and sampling gate consist of gaps (red) in the field oxide (gray) covering the silicon substrate. Aluminum microstrip transmission lines (blue) connect the pulse generator and sampling gate to other portions of the circuit. Areas where the aluminum lines contact the silicon substrate are hatched. Signals are extracted through the bond pad (black) for amplitude measurements. (b) A cross-sectional view of one of the photoconductive circuit elements. Photoconductive circuit elements can also be fabricated within a layer of polycrystalline silicon on a silicon substrate and within the substrate employed for gallium arsenide integrated circuits.

the device should be realizable within both silicon and gallium arsenide integrated circuits (that is, on substrates of both crystalline silicon and crystalline gallium arsenide). Therefore the necessary properties must be achieved in a material and by methods compatible with the technology of fabricating those types of integrated circuits. (Silicon integrated circuits dominate the industry, but the more advanced gallium arsenide circuits are in greater need of ultrafast measurements.) Finally, the material must be a very efficient photoconductor, capable of transforming the relatively low (about 50-picojoule) energy content of subpicosecond laser pulses to electrical pulses of relatively high amplitude.

Silicon Measurement Device. We first investigated the possibility of using crystalline silicon, the substrate for silicon integrated circuits, as the material for the photoconductive circuit elements. We fabricated a number of 'integrated circuits' on a typical silicon substrate, each circuit consisting simply of an aluminum interconnection line in contact, through holes in the field oxide (silicon dioxide) covering the substrate, to regions of the substrate that were to serve as the pulse generator and sampling gate (Fig. 10). The sampling-gate region (and sometimes also the pulse-generator region) were bombarded with various ions (deuterium, helium, neon, or oxygen) to decrease carrier lifetime.

Figure 11 shows the correlation function obtained for one of the test devices. Since the interconnection has very little effect on the signal issuing from the pulse generator, the correlation function provides a very close approximation to the response of the device itself (that is, the response of the pulse generator to the laser pulse and of the sampling gate to the resulting electrical pulse), which in turn provides the rise time of the pulse generator and the sampling aperture of the sampling gate. The shortest sampling aperture obtained with our early test de-

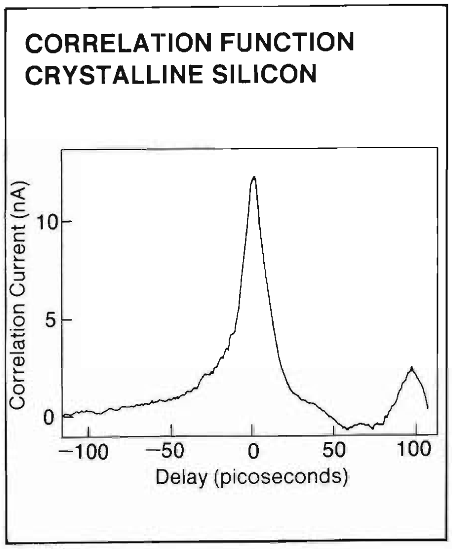


Fig. 11. This correlation function for a photoconductive pulse generator and sampling gate, both of crystalline silicon damaged with 30-MeV oxygen ions, was obtained with the experimental setup of Fig. 9. The left and right portions of the correlation function are dominated, respectively, by the rise time of the pulse generator and the lifetime of carriers in the sampling gate. Analysis of the correlation function yields a rise time (10 to 90 percent) for the pulse generator of about 5 picoseconds and a sampling aperture for the sampling gate of about 20 picoseconds. The small peak on the right is due to reflections from the bond pads at the ends of the microstrip trans mission line. ◀

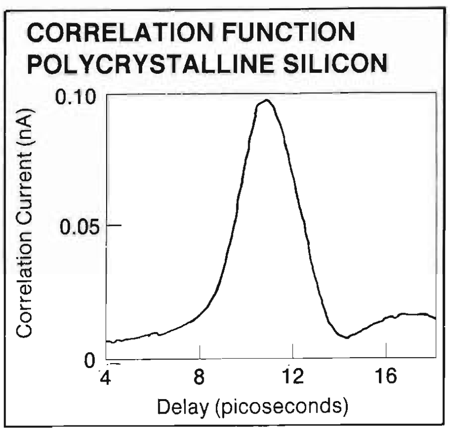


Fig. 12. Analysis of this correlation function for a photoconductive pulse generator and sampling gate, both of polycrystalline silicon damaged with 250-keV silicon ions, yields a sampling aperture of about 2 picoseconds, which is considerably shorter than the sampling aperture attainable with crystalline silicon. Such a short sampling aperture implies that the performance of the electrical measurement device is limited by the capacitances of the sampling gate and the pulse generator rather than by carrier lifetime. ◀

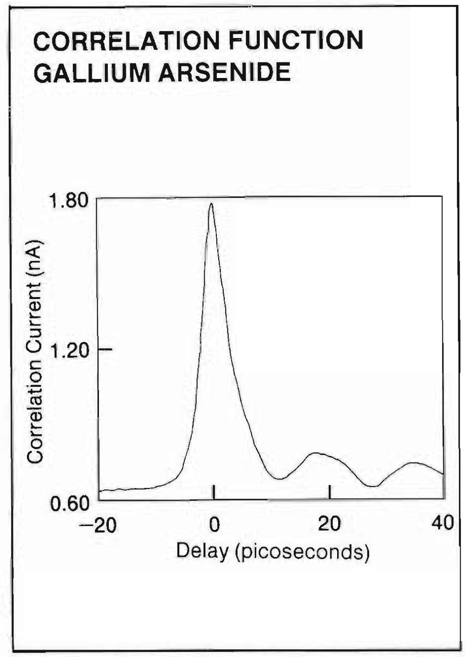


Fig. 13. Analysis of this correlation function for a photoconductive pulse generator and sampling gate, both of gallium arsenide damaged with 6-MeV alpha particles, yields a sampling aperture of about 1 picosecond, the shortest we have attained. A measuring device with such a short sampling aperture is required to determine the response of gallium arsenide components for advanced integrated circuits.▲

vices was about 20 picoseconds; by increasing the amount of radiation damage, we have since produced crystalline silicon sampling gates with sampling apertures of about 12 picoseconds. All the devices exhibited pulse rise times of about 5 picoseconds.

Our experiments with photoconductive circuit elements of crystalline silicon proved the principle of the device but also revealed a major disadvantage of that material. We found that the bombarding ions must penetrate much farther into the silicon than do the laser pulses. Otherwise, carriers created at depths below the extent of the radiation damage are long-lived, and these long-lived carriers, which are not electrically isolated from the radiation-damaged region, make the device worthless. Sufficiently deep radiation damage can be produced only by ions with energies much greater than the few hundred keV provided by the ion implanters available to the integrated-circuit industry.

To eliminate the need for deep radiation damage, we searched among the other materials found in silicon integrated circuits for one that could be electrically isolated from the silicon substrate. Polycrystalline silicon, a material used for interconnections and the gate electrodes of field-effect transistors, seemed a likely candidate. It can be isolated from the substrate by the field oxide, its intrinsic resistivity is reasonably high, and we found that the carrier mobility could be increased to an acceptable value by annealing at 1100°C. (Annealing increases the sizes of crystal grains in the material and thus reduces the grain-boundary scattering that limits carrier mobility in normal polycrystalline silicon.)

Figure 12 shows a correlation function for a test device containing photoconductive circuit elements of polycrystalline silicon. The sampling aperture of the device is so short (about 2 picoseconds) that it is limited not by carrier lifetime but by the capacitances of the photoconductive gaps. The temporal resolution of such a device exceeds that of sampling oscilloscopes by a factor of about 10.

We are pursuing several near-term applications of the polycrystalline silicon device, at present concentrating on its use to measure the velocity, attenuation, and dispersion of high-frequency signals as they propagate through various microstrip transmission lines on silicon substrates. (The opening figure shows an integrated circuit fabricated for this purpose.) The high accuracy of these measurements, made possible by the performance characteristics of the device coupled with essentially zero timing jitter, will permit us to verify the predictions of a model we had developed earlier for the transmission characteristics of such interconnections. (Another aspect of our research is the development of models for use in computer-aided circuit design that more adequately describe today's high-speed integrated circuits. For example, interconnections are commonly modeled as lumped circuit elements, but that approach is valid only when the wavelengths of the signals being transmitted are small compared with the dimensions of the interconnections.) Early results indicate that the model predicts propagation velocity with an accuracy of 1 percent and attenuation and dispersion with an accuracy of 10 percent.

Gallium Arsenide Measurement Device. The near monopoly of silicon as the substrate material for integrated circuits is now being broken by gallium arsenide, which offers two major advantages over silicon. First, lasers and other light-emit-

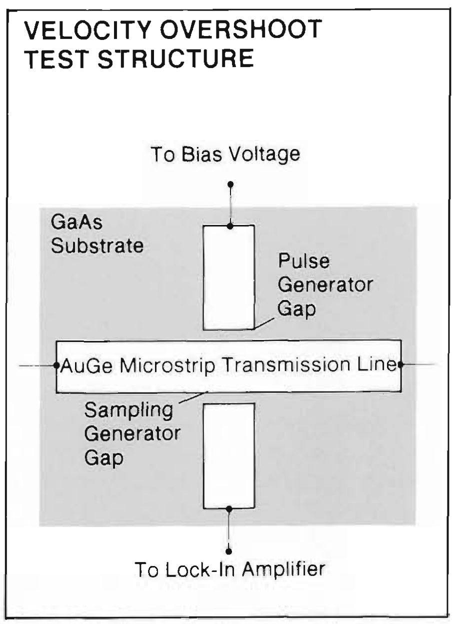
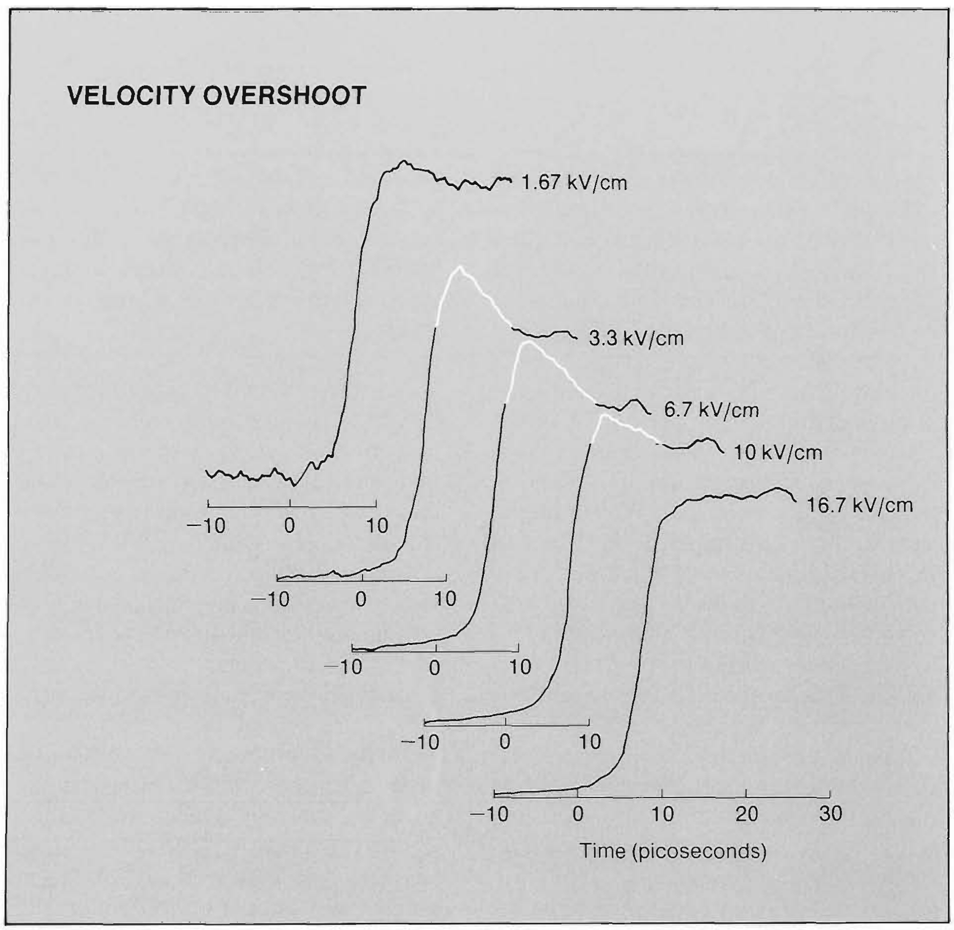


Fig. 14. Schematic diagram of a portion of the integrated circuit used to study velocity overshoot in gallium arsenide. The photoconductive pulse generator and sampling gate consist of gaps, patterned by lift-off photolithography, between the microstrip transmission lines. Gold metallization on the backside of the gallium arsenide substrate served as a backplane.

Fig. 15. These data for the current induced in a gallium arsenide pulse generator by 0.17-picosecond laser pulses represent the first direct observation of velocity overshoot. The observed dependence of the overshoot on electric field is qualitatively consistent with that predicted theoretically: at low fields no overshoot is expected; at high fields the overshoot, although large in magnitude, is so short in duration that it is obscured by the *RC* time constant of the photoconductive gaps; and at intermediate fields the overshoot, although smaller in magnitude, is sufficiently long to be resolved.▼



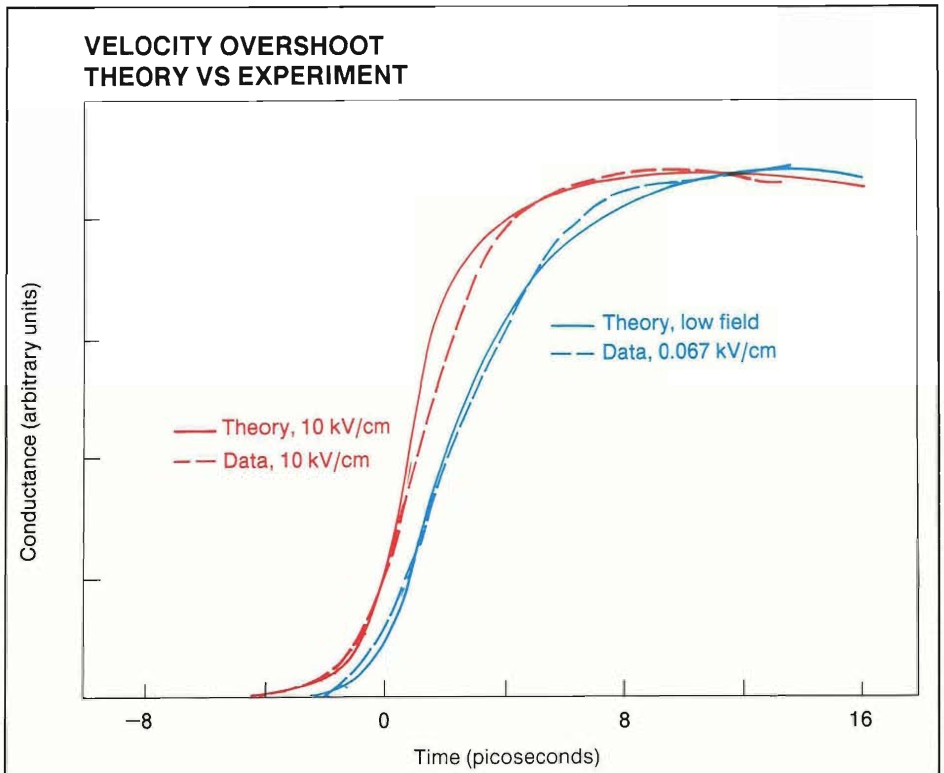


Fig 16. A convolution of the predicted velocity overshoot with the *RC* time constants of the photoconductive gaps yields very good quantitative agreement between the observed and calculated values for the conductance of a gallium arsenide pulse generator excited by 0.17-picosecond laser pulses. The much greater magnitude of the overshoot in high electric fields causes a steeper and earlier rise of the conductance.

ting devices can be fabricated within gallium arsenide, and second, gallium arsenide transistors are several times faster than silicon transistors. (Silicon's indirect band gap prevents it from lasing, and its lower carrier mobility decreases transistor speed.) These advantages imply that gallium arsenide integrated circuits will become increasingly common, and we have expended substantial effort on realizing ultrafast photoconductors within gallium arsenide substrates.

Correlation functions for early test devices fabricated on gallium arsenide substrates showed that the sampling aperture was limited by the *RC* time constants of the photoconductive gaps. Thinning the substrate, a usual method of reducing this limitation, proved extremely difficult, but replacing the microstrip transmission line with a coplanar waveguide transmission line yielded extraordinarily short sampling apertures—less than 1 picosecond (Fig. 13).

We have applied a gallium arsenide device to measuring the step response of a gallium arsenide transistor. The resolution of this measurement, 11 picoseconds, was, at the time, the highest ever achieved.

Velocity Overshoot. Our work on gallium arsenide circuit elements also provided the first direct observation of 'velocity overshoot.' This phenomenon, predicted theoretically in 1972 and eagerly sought ever since, is a transient acceleration of carriers, in a constant electric field, to velocities above the steady-state drift velocity. Velocity overshoot is of great technological interest because it could be the basis for very-high-speed transistors, provided the transistors were appropriately small.

Using the integrated circuit shown schematically in Fig. 14, we measured the current generated by a gallium arsenide photoconductor in response to 0.17-picosecond laser pulses. Early results (Fig. 15) were qualitatively consistent with the predicted dependence on electric field of the duration and magnitude of the velocity overshoot. By increasing the uniformity of the electric field imposed on the photoconductor (with improved contacts), we obtained further data that are in excellent *quantitative* agreement with theory (Fig. 16). These experiments not only proved the reality of velocity overshoot but also demonstrated that photoconductive circuit elements can be used to study the phenomenon, paving the way to its utilization in components for advanced integrated circuits.

Summary

This article has highlighted new devices made possible by the rapid onset and decay of photoconductivity. However, one goal of our efforts here at the Laboratory does not concern the devices themselves and their immediate practical benefits. Rather, development of the devices, in particular the photoconductive circuit elements, is a necessary first step toward new avenues of basic research on semiconductors. Our observation of velocity overshoot in gallium arsenide is an example of such research. The effect, one of a class known as nonequilibrium carrier-transport phenomena, is due to collisions of carriers with the vibrating crystal lattice. The mean time between collisions ranges from a few femtoseconds to a few picoseconds. It is that world of fleeting events we aim to explore. ■

Acknowledgments

Many others contributed to the success of the work described in this article. The author wishes in particular to acknowledge the collaboration of Jeffrey M. Bradley, Alan J. Gibbs, A. Evan Iverson, Donald R. Kania, Nicholas G. Paulter, Darryl L. Smith, and Ronald S. Wagner of Los Alamos, Douglas R. Bowman of the U.S. Military Academy at West Point (formerly of Stanford University), Robert W. Dutton of Stanford University, William R. Eisenstadt of the University of Florida (formerly of Stanford University), F. A. Lindholm of the University of Florida, and William C. Nunnally of the University of Texas, Arlington (formerly of Los Alamos).

Further Reading

General

D. H. Auston. 1975. Picosecond optoelectronic switching and gating in silicon. *Applied Physics Letters* 26:101-103.

D. H. Auston, A. M. Johnson, P. R. Smith, and J. C. Bean. 1980. Picosecond optoelectronic detection, sampling, and correlation measurements in amorphous semiconductors. *Applied Physics Letters* 37:371-373.

Gerard Mourou, editor. *Picosecond Optoelectronics.* Volume 439 of *Proceedings of SPIE.* Bellingham, Washington: The Society of Photo Optical Instrumentation Engineers, 1983.

D. H. Auston and K. B. Eisenthal, editors. *Ultrafast Phenomena IV: , Proceedings of the Fourth International Conference, Monterey, California, June 11-15, 1984.* Berlin: Springer-Verlag, 1984.

G. A. Mourou, D. M. Bloom, and C.-H. Lee, editors. *Picosecond Electronics and Optoelectronics: Proceedings of the Topical Meeting, Lake Tahoe, Nevada, March 13-15, 1985.* Springer Series in Electrophysics, edited by Walter Engl, volume 21. Berlin: Springer-Verlag, 1985.

Photoconductor Power Switch

W. C. Nunnally and R. B. Hammond. 1984. 80-MW photoconductor power switch. *Applied Physics Letters* 44:980-982.

W. C. Nunnally and R. B. Hammond. Optoelectronic switch for pulsed power. In *Picosecond Optoelectronic Devices*, edited by Chi H. Lee. Orlando, Florida: Academic Press, Inc., 1984.

R. B. Hammond. Photoconductor switching: Some physical mechanisms. In *Final Report of the Workshop on New Directions in Solid State Power Switches, Weber Research Institute, Polytechnic Institute of New York, Farmingdale, New York, August 28-30, 1985*, pp. 239-285. B. Senitzky, chairman. Washington, D.C.: Office of Naval Research, 1985.

Photoconductor Radiation Detector

Robert B. Hammond, Nicholas G. Paulter, Ronald S. Wagner, Thomas E. Springer, and Martin D. J. MacRoberts. 1983. InP:Fe photoconductors as photodetectors. *IEEE Transactions on Electron Devices* ED-30:412-415.

R. B. Hammond, N. G. Paulter, R. S. Wagner, and T. E. Springer. 1984. Excitation and Fe concentration dependences in the impulse photoconductance of InP:Fe. *Applied Physics Letters* 44:620-622.

D. R. Kania, R. J. Bartlett, R. S. Wagner, R. B. Hammond, and P. Pianetta. 1984. Pulsed soft x-ray response of InP:Fe photoconductors. *Applied Physics Letters* 44:1059-1061.

Ronald S. Wagner, Jeffrey M. Bradley, and Robert B. Hammond. 1986. Picosecond photoconductors as radiation detectors. *IEEE Transactions on Nuclear Science* NS-33: 250-253.

D. R. Kania, A. E. Iverson, D. L. Smith, R. S. Wagner, R. B. Hammond, and K. A. Stetler. Flat-response detectors for the vacuum ultraviolet and soft x-ray region. Accepted for publication in *Journal of Applied Physics.*

Electrical Measurement Device

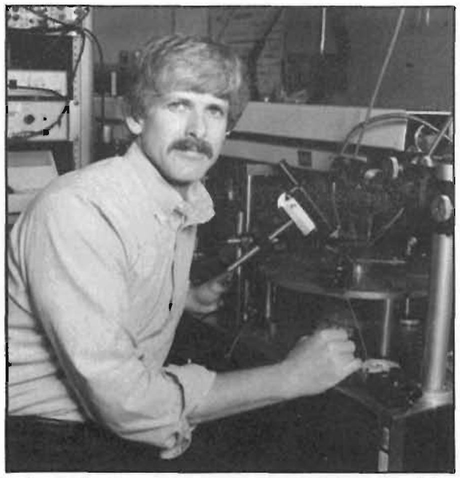
R. B. Hammond, N. G. Paulter, and R. S. Wagner. 1984. Observed circuit limits to time resolution in correlation measurements with Si-on-sapphire, GaAs, and InP picosecond photoconductors. *Applied Physics Letters* 45:289-291.

D. R. Bowman, R. B. Hammond, and R. W. Dutton. 1985. Polycrystalline-silicon integrated photoconductors for picosecond pulsing and gating. *IEEE Electron Device Letters* EDL-6:502-504.

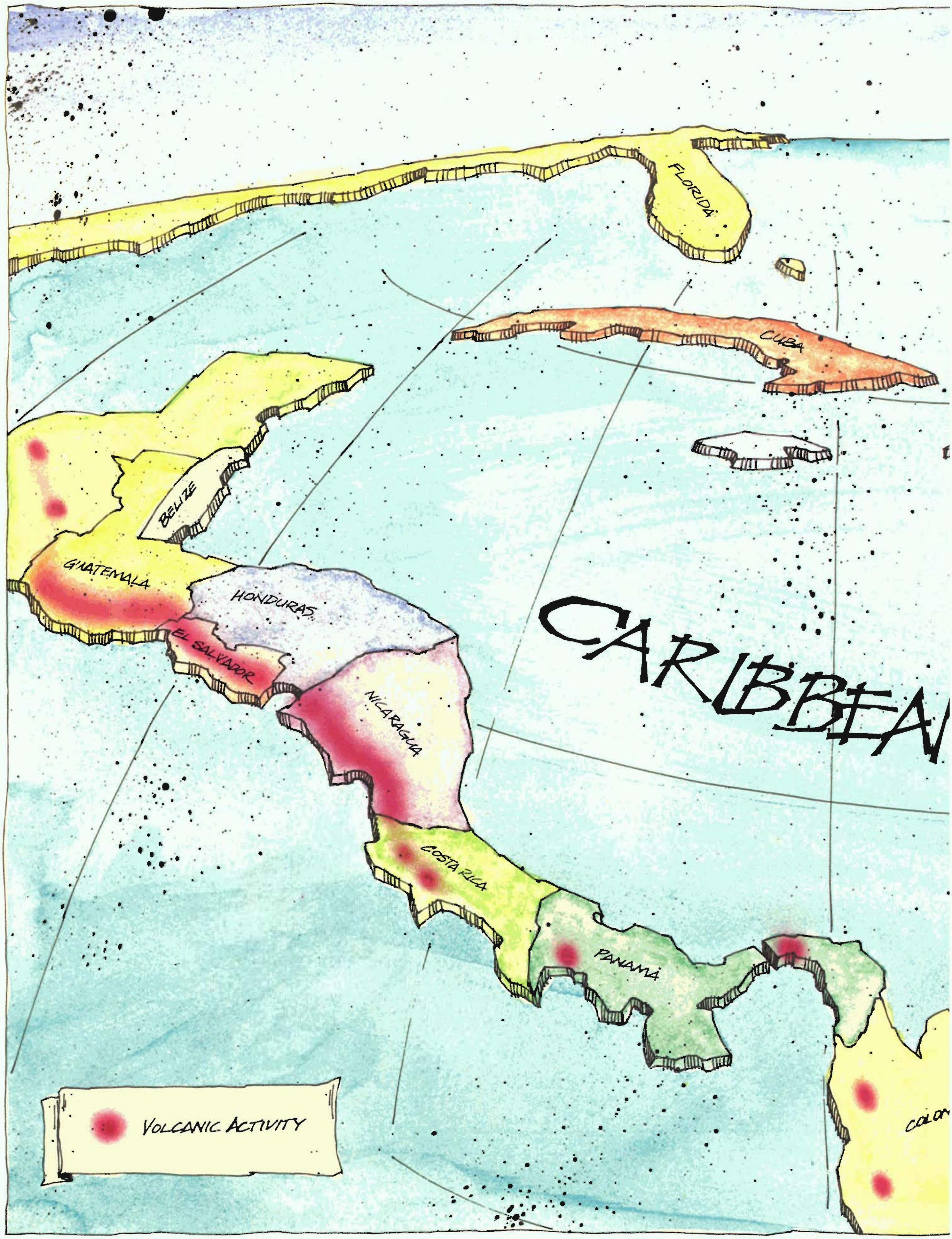
D. R. Bowman, R. W. Dutton, and R. B. Hammond. 1985. New integrated polysilicon photoconductor for ultrafast measurements on silicon. In *Technical Digest of the International Electron Devices Meeting, Washington, DC, December 1-4, 1985*, pp. 117-120. New York: Electron Devices Society of IEEE, 1985.

R. B. Hammond. 1985. Electron-velocity overshoot observed in an impulse-excited GaAs photoconductor. *Physica* 134B:475-479.

R. B. Hammond and N. M. Johnson. 1986. Impulse photoconductance of thin-film polycrystalline silicon. *Journal of Applied Physics* 59:3155-3159.

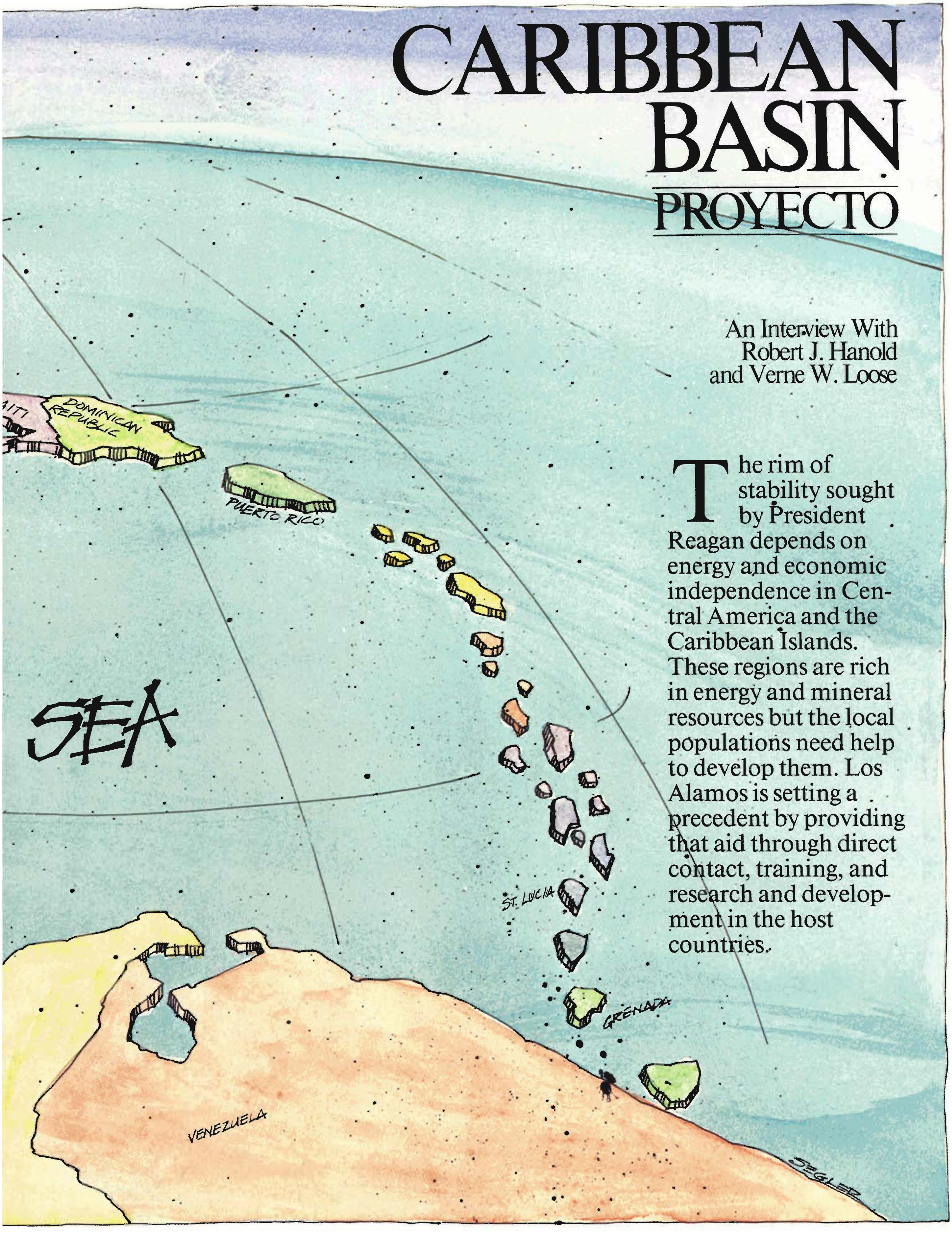


Robert B. Hammond received his scientific education at the California Institute of Technology, earning a B.S. in physics (with honor) in 1971, an M.S. in applied physics a year later, and a Ph.D. in applied physics in 1975. Since 1973 his research has focused on experimental and theoretical aspects of semiconductor physics. In 1975 he joined the Air Force Weapons Laboratory in Albuquerque, New Mexico, where he helped develop a chemically pumped iodine laser. In 1976 he joined the Electronics Division at Los Alamos. His work here has included development of a position-sensitive detector for infrared lasers and of a high-damage-threshold Smartt interferometer for pulsed infrared lasers, measurement of the phase diagram for the electron-hole liquid in silicon, experimental and theoretical studies of exciton kinetics in silicon, and passivation studies of III-V semiconductors. Since 1981 his attention has been devoted primarily to the topics reported on in this article. He has authored over fifty technical papers and is a member of the American Physical Society, the Institute of Electrical and Electronics Engineers, the Optical Society of America, Sigma Xi, and the Bohmische Physical Society.


FLORIDA
CUBA
BELIZE
GUATEMALA
HONDURAS
EL SALVADOR
NICARAGUA
CARIBBEAN
COSTA RICA
PANAMA
VOLCANIC ACTIVITY

CARIBBEAN BASIN PROYECTO

An Interview With
Robert J. Hanold
and Verne W. Loose



The rim of stability sought by President Reagan depends on energy and economic independence in Central America and the Caribbean Islands. These regions are rich in energy and mineral resources but the local populations need help to develop them. Los Alamos is setting a precedent by providing that aid through direct contact, training, and research and development in the host countries.

Science: *The Los Alamos travel office is busy sending scientists and equipment to the Caribbean Basin. What is going on there?*

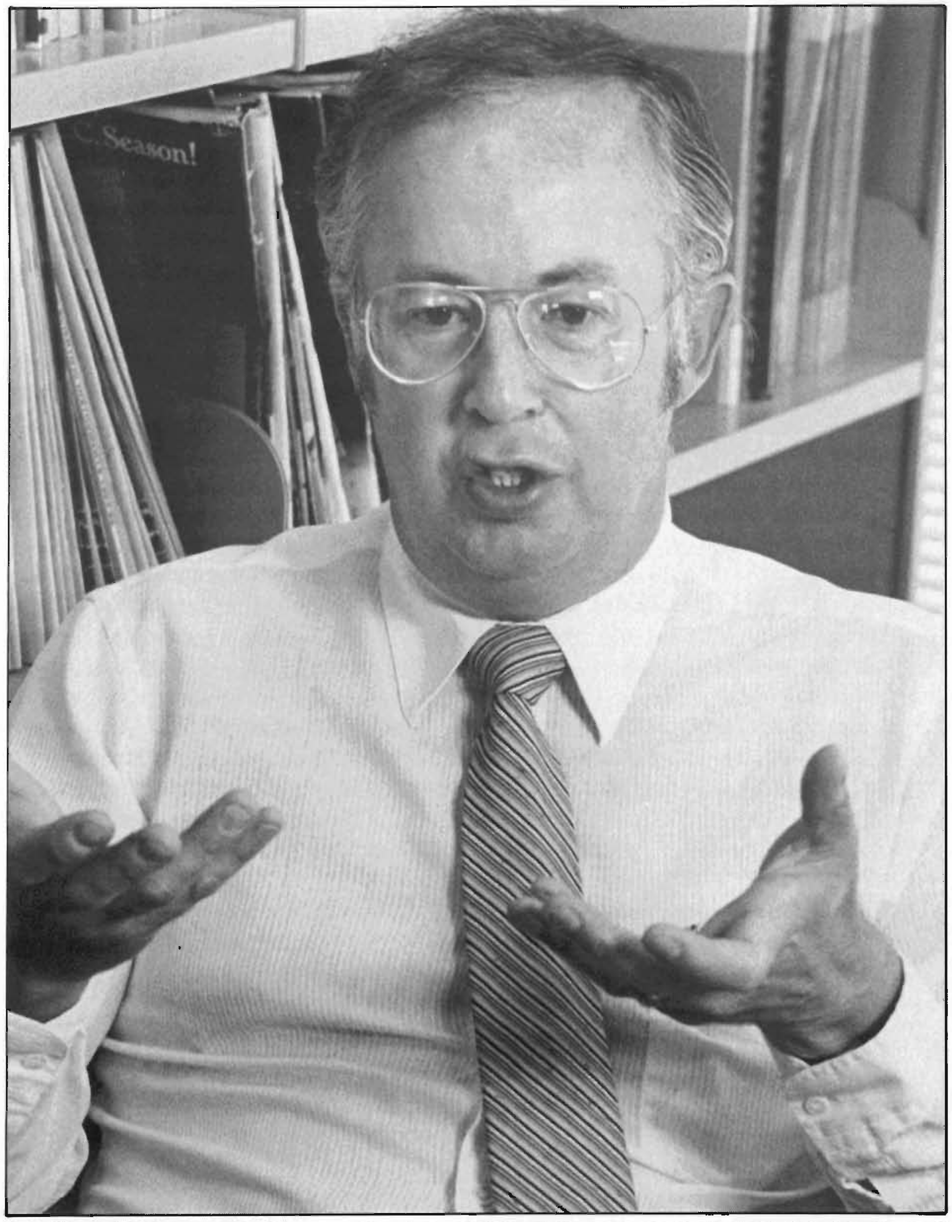
Loose: We are involved in a project designed to address the energy problems of certain countries there, the main problem being the need to import large quantities of petroleum and the consequent buildup of huge foreign debts.

Science: *Are the economies of these countries very weak?*

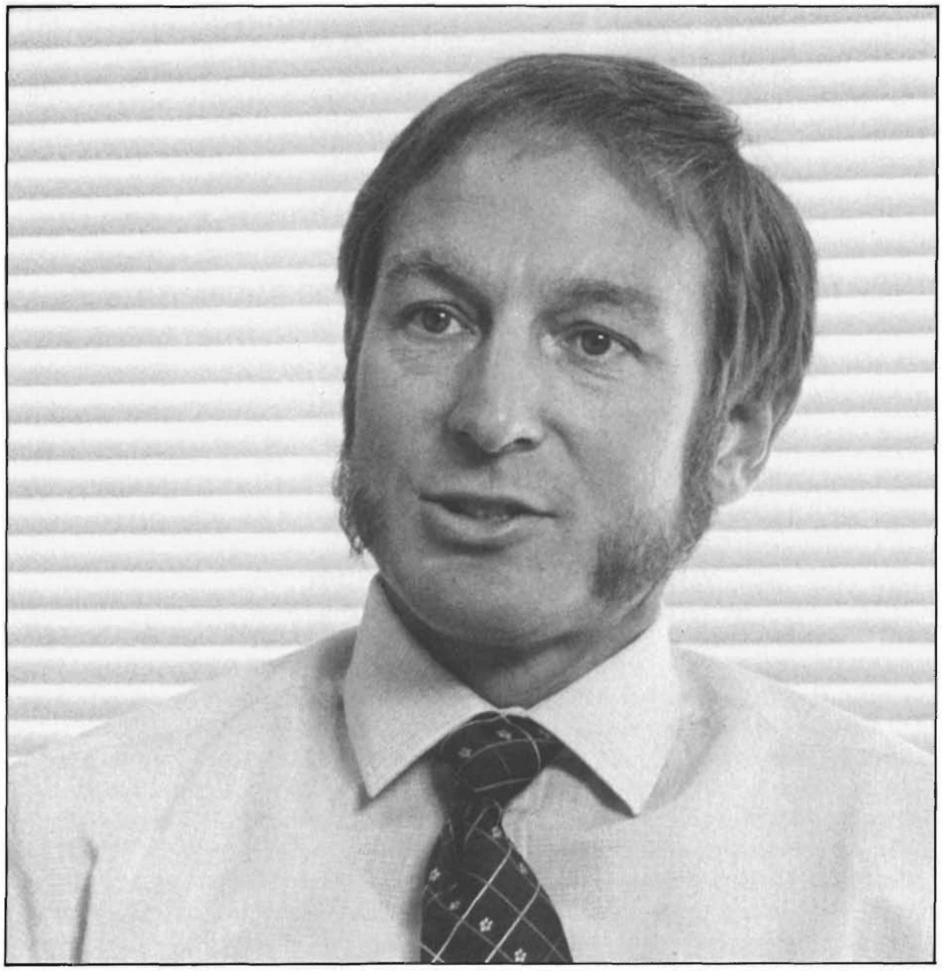
Loose: They are fragile. Between World War II and 1978 economic expansion was fairly rapid due to periodic booms in the price of their agricultural products, increases in manufacturing, and heavy foreign borrowing. Since 1978, however, commodity prices have fallen, and manufacturing has slowed down due to worldwide recession. Sustained insurgent activity in Nicaragua, El Salvador, and Guatemala has contributed an additional burden to the economies of those countries. Moreover, the cost of imported petroleum grew so rapidly after the Arab oil embargo in 1973 and the Iranian cutback of 1979 that by the early eighties 20 to 25 percent of every dollar earned from exports went to buy petroleum. It is this acute economic situation that we are trying to remedy.

Hanold: St. Lucia is a good example. This Caribbean island generates all of its electric power with diesel engines. When we made our first visit there in 1983, the import of diesel fuel was imposing a burden in excess of a million dollars a month. That's not much in our economy, but for a small Caribbean island with a total population of 120,000 and a relatively small industrial economy, the burden is intolerable. The prime minister of St. Lucia was aware of the efforts of the United States to develop alternative forms of energy. Since St. Lucia is entirely volcanic in nature, he knew it would probably have an excellent geothermal resource that, if developed properly, might displace the imported petroleum.

Science: *Who is the prime minister?*



Bob Hanold, Program Manager for International Energy Activities in the Earth and Space Sciences Division, is responsible for the technical management of the geothermal projects in Central America and the Caribbean. He accepted a postdoctoral fellowship at Los Alamos in 1966, immediately after earning a Ph.D. in engineering science at Case Institute of Technology, and became a staff member in 1968. He has extensive experience in research and development of all aspects of geothermal projects, including well stimulation by hydraulic fracture and chemical treatment, chemical scale control, pumping systems, fracture diagnostics, and high-temperature instrumentation. He has also been a leader in the development of cost-shared field experiments with industry. His warm personality combined with his vast technical knowledge makes him a very valuable asset to technology transfer initiatives.



Verne Loose is Program Manager for International Initiatives in the office of the Assistant Director for Industrial and International Initiatives and is also Program Manager of the Central American Energy Resources Project. He earned his Ph.D. in natural resources and energy economics from the University of British Columbia and then worked for the government of British Columbia as an energy economist. He has extensive experience in evaluation of natural resources development projects. Since joining the Laboratory in 1977, he has been doing research in energy economics, including analyses of utility investment and oil and gas substitution and mathematical studies of optimal oil and gas reservoir production. Before becoming Program Manager for International Initiatives, he was Leader of the Economics Group for three-and-a-half years. He has spearheaded the studies of energy economics for the Central American project during the past two-and-a-half years.

Loose: His name is John Compton. He is a very capable and charismatic leader and a very delightful gentleman. He was trained as an attorney in England and is also the owner of a banana plantation. His ancestors were slaves under the English. The control of St. Lucia, like that of many other Caribbean islands, has alternated back and forth during the last two or three hundred years among the English, the French, and the Spanish. Most recently St. Lucia was an English protectorate. English is the official language, although most of the population speaks a dialect of French.

Science: *Who is affected by the adverse balance of payments? The wealthy sector of the population?*

Hanold: Since imported petroleum is the source of all of the island's electricity, anyone who purchases electricity has to bear the brunt of its cost.

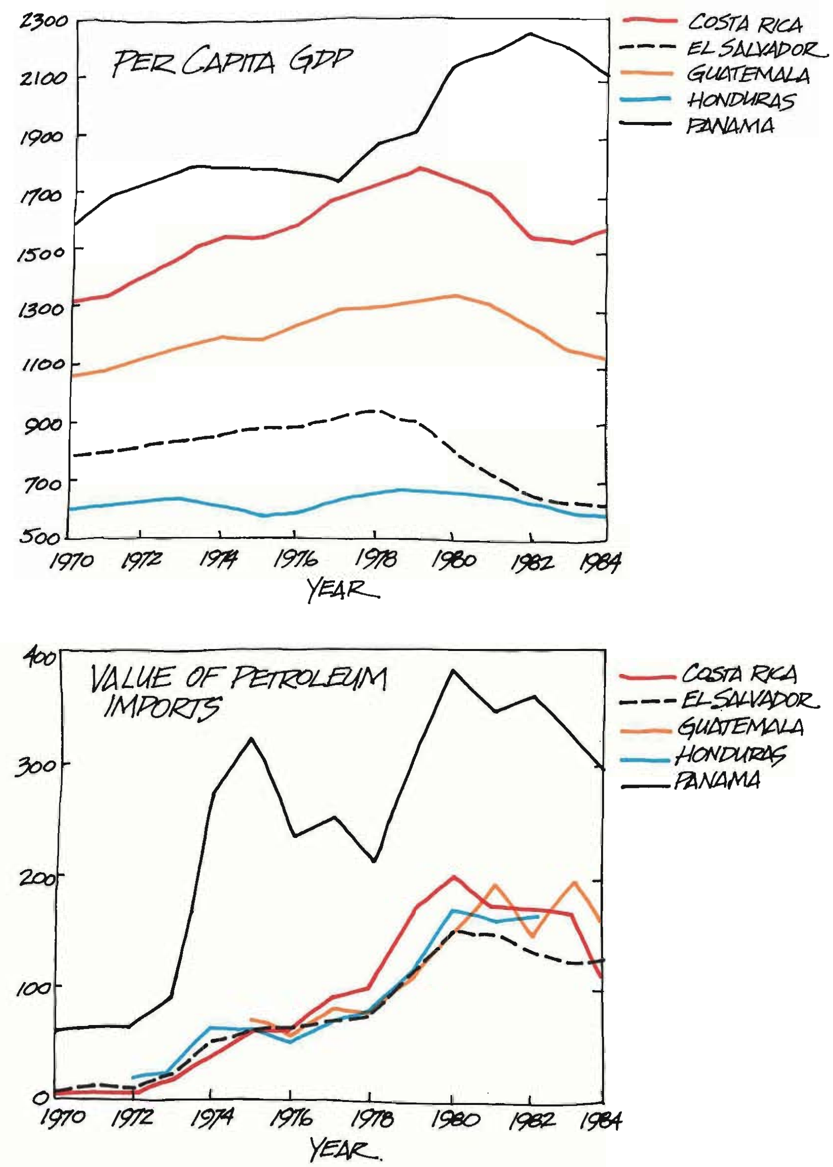
Science: *Did the Los Alamos project start with work in St. Lucia?*

Loose: Officially, yes. But Ron Lohrding, the Laboratory's Assistant Director for Industrial and International Initiatives, has been laying the groundwork for this project during the last five years or so. Initially he developed contacts with European and international energy organizations to see how the Laboratory's expertise in energy technology might be transferred to other countries. About three years ago his interest was channeled to the Caribbean islands and Central America, in part by Reagan's Caribbean Basin Initiative. This initiative was intended to develop a rim of stability in what the Monroe Doctrine defines as our nation's area of influence. Ron worked with people in Washington to find ways in which institutions like Los Alamos could help to support the president's policy. In addition, Ron and John Whetten, then leader of the Earth and Space Sciences Division, traveled extensively in the Caribbean. They visited various mission offices of the United States Agency for International Development [the AID] as well as government officials in the host countries. During that trip they met and briefed Mr. Compton. They emphasized the Laboratory's expertise in geothermal energy in part because of our hot dry rock geothermal project but mostly because they knew the prime minister was interested in geothermal development. This mutual interest led to requests in Washington on behalf of Mr. Compton for that type of technical assistance. Los Alamos secured funds from the State Department's Trade Development Program, one goal of which is to give projects a boost toward commercialization. We began to use the funds for field work at the Qualibou Caldera in St. Lucia in August 1983.

Hanold: A few months later the Kissinger Commission, the president's Bipartisan Commission on Central America, asked us to brief them on the technical assistance needs of Central America. They were particularly interested in identifying projects that would generate employment and promote economic development in the region.

Science: *Was Los Alamos the only national laboratory invited to give a briefing?*

Loose: Yes. I suppose we were chosen for several reasons—our familiarity with the Hispanic culture of the Southwest, our technical expertise in geothermal energy and economic analysis, our technical collaboration with the Petroleum Institute and the Nuclear Research Institute of Mexico, as well as our recent experience in St. Lucia. Prior to that meeting John



Whetten spent two weeks in Costa Rica with the Organization of American States to identify employment-generating projects in the energy and mineral sectors. Information gained from that mission, as well as an analysis of the region by Los Alamos, enabled Whetten to identify peat as a neglected resource in Costa Rica and possibly in other Central American countries. In addition to geothermal and peat development, we emphasized the potential for minerals development at the Kissinger Commission briefing. With the exception of Cuba, where Soviet influence is prevalent, mineral resources contribute very little to the gross national products of the Caribbean Basin countries. In contrast, minerals developed with Soviet investments account for a substantial fraction of Cuba's gross national product.

Hanold: Training the local population was another issue we discussed with the Commission. It is of paramount importance in any technical assistance project. The Soviets know this and are sponsoring graduate and undergraduate training of many Central American students.

Loose: Following the Kissinger Commission briefing we prepared a proposal to the AID for $10.2 million to provide technical assistance in energy development, mineral development, and energy and economic planning to five Central American countries—Honduras, Costa Rica, El Salvador, Panama, and Guatemala. The goals are to identify resources, provide technical training, and help find funding for development.

Science: *Does this project have any precedents?*

Loose: It is new for Los Alamos, and it is a little different from most American technical assistance projects. Usually the AID mission office in a country gets ideas from the local government personnel as to what types of technical assistance would be beneficial. In this case Los Alamos helped to identify the needs and to convince the AID mission offices of the benefits to be gained from meeting them.

Hanold: Also, the technological level of the project is higher than is typical.

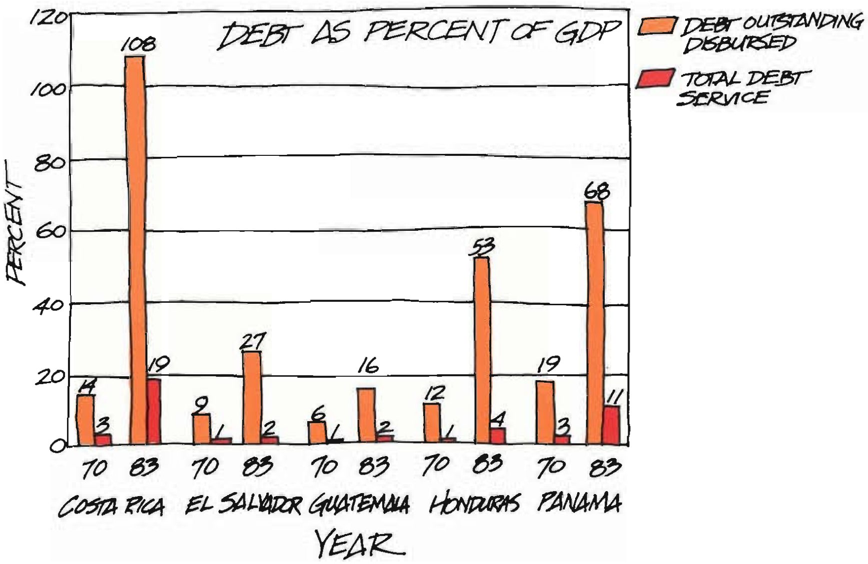
Loose: The AID usually tries to help the poorest of the poor. They focus on basic human needs, such as health, potable water, sanitation, and education. In contrast, this project, which addresses the middle ground of technology, is focused on the industrial sector of the country. It is being done in the spirit of the Alliance for Progress of the early sixties, which gave assistance to Latin America.

Science: *Did you have to convince the AID that the project was appropriate to their goals?*

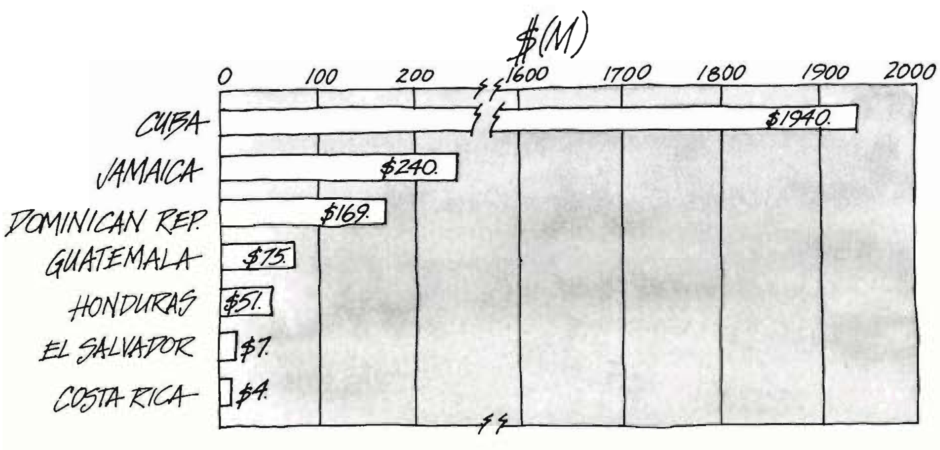
Loose: Yes, we had to create a demand for the project, and John Compton was very helpful in this regard by giving us a concrete example of what we could accomplish.

Hanold: Our job was made easier in St. Lucia because we had been preceded by technical delegations from other countries who had investigated the possibility of geothermal power. The Italians had done so in the early eighties. Before that the English had done a great deal of pioneering work on geothermal manifestations. The government of St. Lucia had seen a number of starts and stops and was anxious that the project be brought to fruition. With the basic geologic data already at hand, we were to choose the locations for drilling the geothermal wells and provide the government with enough confidence to solicit funds for drilling the wells and constructing the geothermal power plant.

Science: *Can you guarantee that the projects you start will reach fruition?*



Recent economic history of five Central American countries. The per capita gross domestic product (GDP) is the best indicator of the standard of living in these countries. The GDP is the value of all goods and services produced annually by citizens of a country (the gross national product) plus the net foreign income (the value of goods produced in a country by foreign companies minus the income earned by citizens living and working abroad). (The per capita GDP data were obtained from the Inter-American Development Bank and the World Bank.) Between 1970 and about 1980, the per capita GDP in these countries was on the rise as a result of industrial development and higher prices for agricultural commodities. The dramatic increase in oil prices after the Arab oil embargo of 1973 and the reliance on public borrowing to support economic growth led to the accumulation of huge foreign debts. (The foreign debt data were obtained from the World Bank.) Since 1980 the worldwide recession has led to a decrease in the per capita GDP and in the ability of these countries to import foreign oil.



Income generated from mineral resources. The larger income in Cuba compared with other Caribbean Basin countries is due to Soviet investment.

Hanold: There are no guarantees, but we will make every effort possible to finish what we start. We have already made dramatic strides in St. Lucia. Our work was accepted by the government of St. Lucia as a sufficient basis to begin drilling. Grants for about $5 million have now been obtained, from the United Nations and the AID, to drill three geothermal wells. Engineers from Los Alamos have been in St. Lucia putting together the plans for drilling, which will start sometime in mid 1986.

Of course drilling wells, whether for oil or geothermal water, is always a risky business. We're not home free until we hit a reservoir that has a good heat source and a good plumbing system through which the hot water can reach the well.

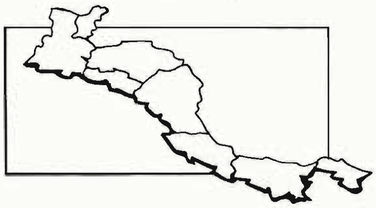
Science: *If the drilling is successful, will the project then be self-supporting?*

Loose: Self-supporting in the sense that money from the sale of electricity will pay for the power plant and its operation. The grants Bob mentioned will pay for the drilling. I would like to emphasize that the Los Alamos project is not focusing on paper studies that may end up collecting dust on somebody's shelf. We are focusing on real investments in energy production capacity. One of our aims is to establish a network of contacts in both the United States and the host countries to interest private investors in development. Private investors are interested in the St. Lucia development, and we would like to see that happen in Central America.

Hanold: Organizations from other countries have already approached us about investments in St. Lucia. For example, the Japanese, who are very strong in manufacturing equipment for electric power generation, requested information about the characteristics of the geothermal fluids and the kinds of equipment that would be suitable for generating the electricity.

Loose: Several American firms and one Canadian firm had discussed commercial development of geothermal energy with the St. Lucian government. We were not

continued on page 74



Energy Supply and Demand

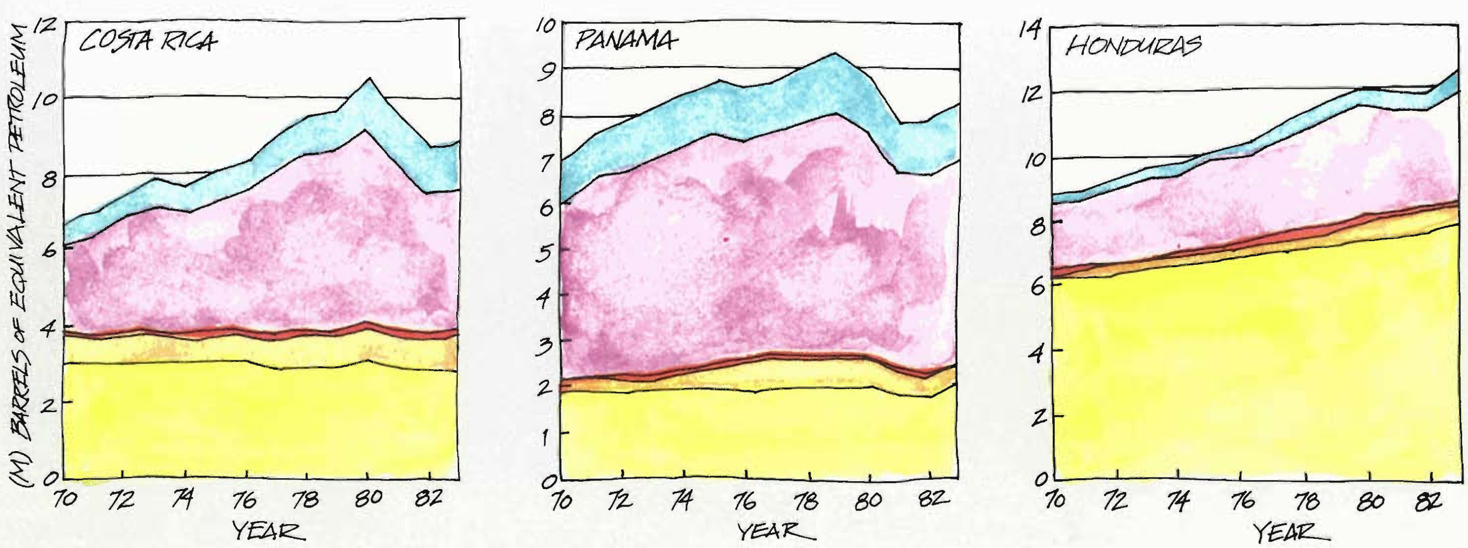
Like those of most developing countries, the Central American economies are dualistic in nature: the rural sector produces mainly traditional agricultural goods, while the rapidly growing urban sector is involved in more modern industrial and commercial pursuits. This dualistic nature is reflected in the pattern of energy consumption shown in the accompanying graphs. People in rural areas rely mainly on firewood to satisfy their energy needs, while those in the cities rely more on electricity and oil products. Consequently, countries with greater degrees of urbanization and higher per capita incomes use relatively more oil products and electricity and less fuelwood than the poorer countries. In 1983, for example, Panama and Costa Rica, the countries with the highest per capita incomes, relied on fuelwood for one-quarter to one-third of their energy needs, while the lower-income countries, Honduras, El Salvador, and Guatemala, relied on fuelwood for two-thirds to three-quarters of their energy needs.

(Data on fuelwood consumption in these and other developing countries are generally poor because much fuelwood does not enter commercial markets where its sale and use can be quantified. The apparent sharp increase in Guatemala's fuelwood consumption in 1979 (see graph) stems from a revision in the estimate of fuelwood consumption rather than a real increase in use.)

In all the Central American countries electricity composes a relatively small share of the total energy consumed, but it has shown the most rapid and variable growth in demand since 1970 (between 1970 and 1980 the demand for electricity grew at an average rate of 9 percent per year).

In the early 1980s Central American countries were hard hit by the worldwide recession. Total energy consumption decreased significantly in Costa Rica, El Salvador, and Panama, and the rate of growth in demand decreased in Guatemala and Honduras. These decreases were absorbed mainly in oil product and electricity demand. Unlike 1973, when oil price hikes could be balanced by debt-financed growth, the early 1980s were a time when large national debts and high interest rates made additional loans difficult and costly to obtain. To exacerbate the situation, the prices received for the main export commodities (bananas, coffee, and sugar) had dropped so low that foreign currency to pay for oil imports and to service foreign debts was in short supply. On a per capita basis Costa Rica and Panama are among the most indebted countries in the world.

With the exception of Guatemala, which produces 1.6 million barrels of poor-quality crude oil per year, the Central American countries have no proven oil reserves and must pay a burdensome price to import oil for transportation and, to a lesser extent, for industry. They all rely heavily on fuelwood, but this resource is threatened in some countries by growing deforestation due mainly to clearing of land for agricultural purposes. By the year 2000 they will undoubtedly face fuelwood



in Central America

by Linda K. Trocki and Steven R. Booth

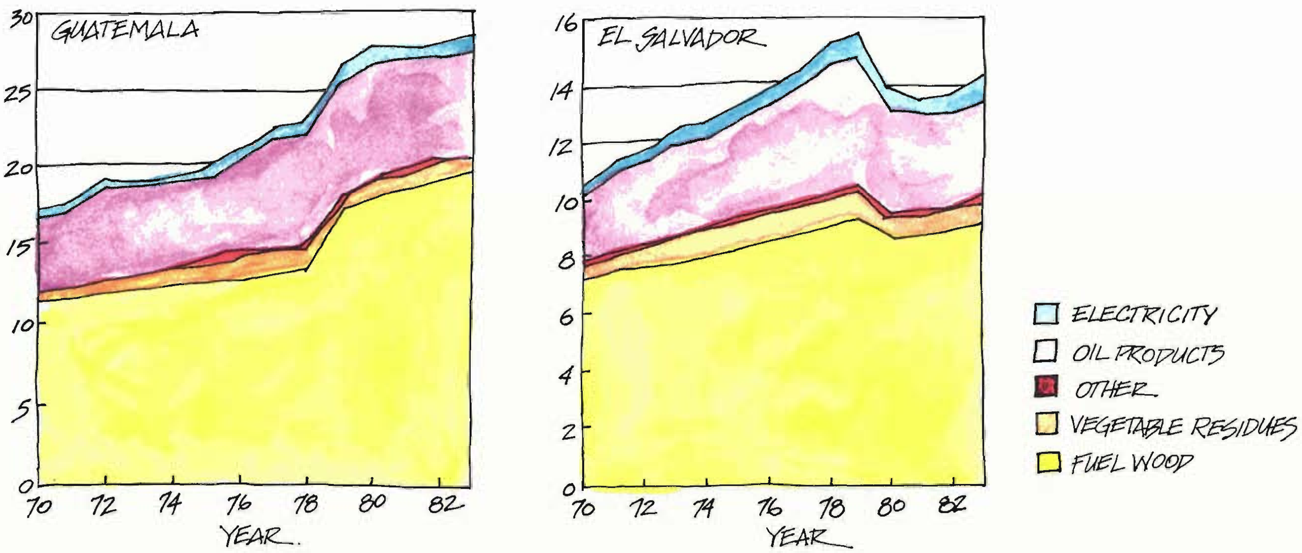
shortages unless substitution or conservation takes place or unless policies to increase fuelwood availability, such as tree farms, are implemented. Since fuelwood is generally gathered by individuals at zero or low cost, finding a similarly inexpensive substitute will represent a major challenge to many of the Central American countries.

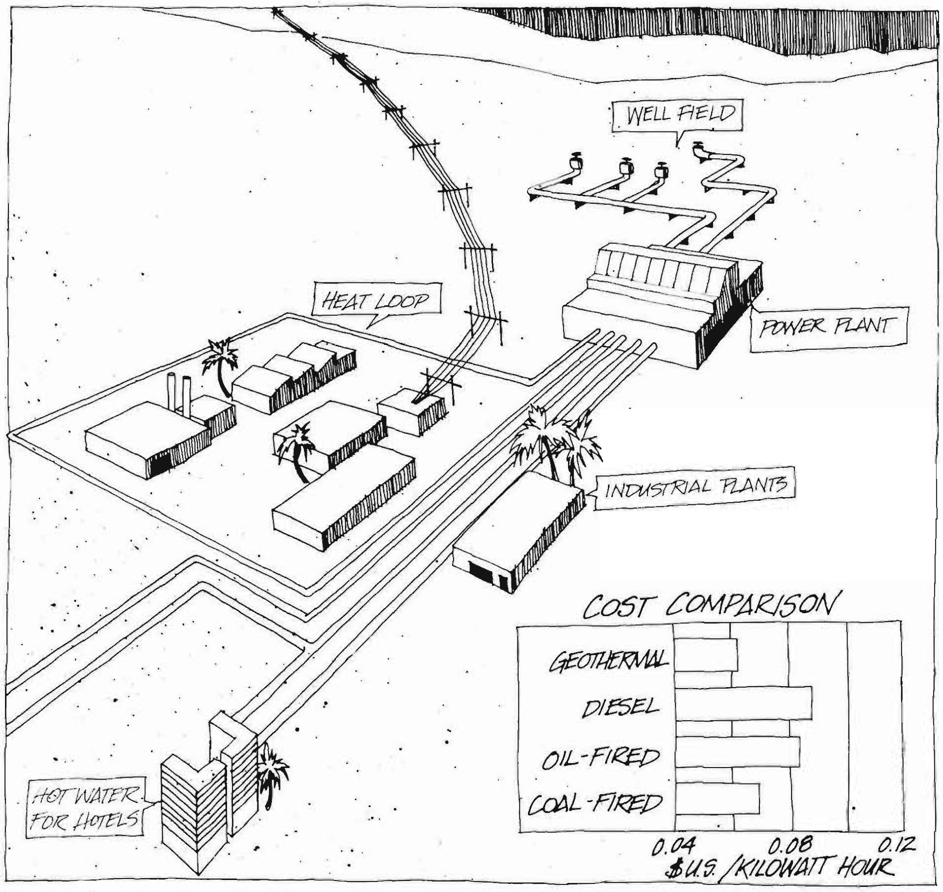
The Central Americans can reduce the demand for fuelwood and imported oil by further developing their large potential for hydroelectric energy and geothermal energy, as well as alternatives such as solar energy, crop residues, and peat. As is evident from the graphs, crop residues already play a signficant role in energy supply in most countries in the region. The residues are burned to provide process heat for the food-processing industries and in some cases to generate electricity. In addition, Costa Rica and El Salvador have begun to produce fuel alcohol from sugar cane.

The countries have also lessened their reliance on imported oil by exploiting their hydroelectric resources to generate electricity. The electric generating capacity in the Central American countries ranges from approximately 500 to almost 900 megawatts, and hydroelectric power constitutes more than 50 to 80 percent of this capacity in all the countries. Since 1979 all countries except El Salvador have greatly increased their electric generating capacity by constructing relatively large (250- to 330-megawatt) hydroelectric power plants. However, the Chixoy hydroelectric plant in Guatemala, commissioned in 1983, was down during much of 1984 and 1985 for repair of a tunnel associated with the dam. As a result, Guatemala incurred a large and unexpected requirement for oil-fired generation to meet its electricity needs. (Chixoy recently resumed operation but not at full capacity.) Construction of the large hydroelectric facilities, while reducing reliance on oil imports, has resulted in temporary overcapacity in Costa Rica and Honduras and significant debts to all the national utilities.

The development of indigenous geothermal energy resources represents an attractive alternative to meet the energy demand. Two countries in Central America already exploit geothermal energy for electricity generation—El Salvador and Nicaragua. (The latter is not included in the Los Alamos study.) By 1990 Costa Rica and Guatemala expect to begin generating electric power from geothermal sites now under development.

In summary, Central America, like most developing regions, relies heavily on two forms of energy—imported oil and fuelwood. Continued heavy reliance on these fuels could result in more serious economic repercussions in the future. For example, every dollar spent to pay the oil-import bill precludes the import of a dollar's worth of capital goods that could further production. And the strong market for fuelwood, which has already caused rapid price increases for that energy source, could lead to serious deforestation problems. Conservation and substitution of indigenous resources could ameliorate potential problems. ■





Costs of generating electricity by various methods in St. Lucia, as determined by Los Alamos economists. The data show that geothermal energy would be the cheapest source of electric power. Moreover, geothermal reservoirs could be used directly as a source of process heat for local industry and of hot water for tourist hotels.

continued from page 71

privy to the discussions that took place, but in the end no agreement was reached. So the United Nations and the AID are paying for the drilling. If geothermal fluid is found, as we expect it will, then the St. Lucia Electric Authority will have to find private money to build the power plants.

Science: *What will your role be during drilling and plant construction?*

Loose: The government of St. Lucia has asked us to serve as consultants as the project proceeds. They have confidence in our advice since we are independent and have no profit motive.

Hanold: Their dealings with private industry will involve areas in which they have very little experience, and they would like Los Alamos to stay and monitor those negotiations.

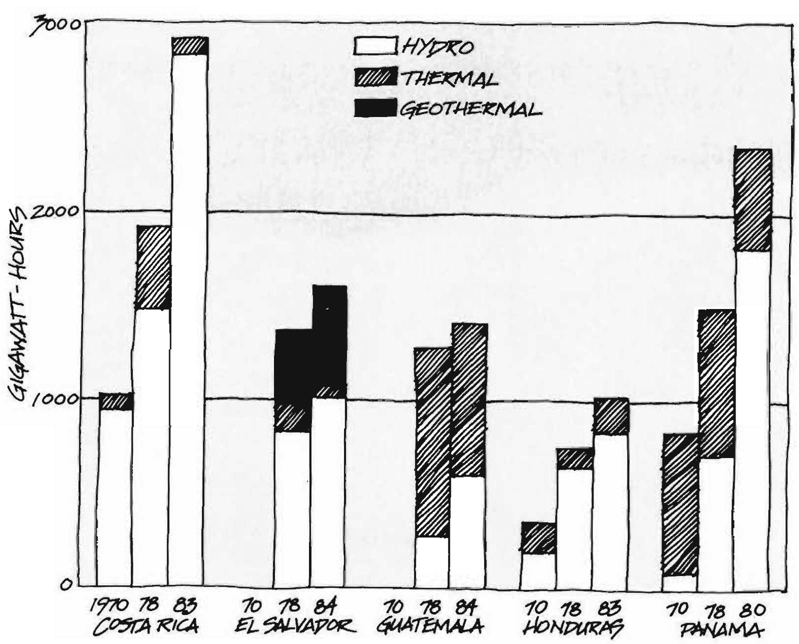
Science: *So you have become ambassadors?*

Hanold: Yes—of good will. The prime minister has shown extraordinary interest in the project. During the six or seven months of field work on the island, we briefed him frequently, and each time he was elated at the progress that was being made. He wants to get a geothermal power plant on line. Several times other organizations have raised the hopes of the St. Lucians for cheap geothermal power and then dashed those hopes by pulling out.

Loose: A geothermal power system would be a secure, low-cost source of electricity, and it could be expanded as the demand for electricity grew. This capacity would help to attract light to moderate industry and thereby alleviate the chronic unemployment on the island.

Science: *Does Central America have similar problems?*

Loose: In some ways yes. The Central American countries are also struggling with chronic unemployment, low per capita income, and energy-supply problems. [See "Energy Supply and Demand in Central America."] In addition they must deal with rapid population growth, class problems, demographic changes, unequal



Total amounts of electricity used in five Central American countries. Since 1970 the major increases have been supplied by recently constructed hydroelectric plants.

A native of St. Lucia examining sulfur deposits at the Sulphur Springs geothermal area, which lies within the Qualibou Caldera near Soufrière.

land distribution, and political instability. Their economies are more complex than those of the Caribbean islands. Their industries have been growing, but they lack the necessary industrial infrastructure and technical know-how for a secure industrial base. Moreover they have relied on loans from foreign countries and imported petroleum to support these industries. The region as a whole has amassed a foreign debt of about $12 billion. In the past our assistance has been directed to the rural population, which lives in very poor conditions. We have poured millions of dollars in this direction with little success in improving the quality of life. It has become apparent that our help should be directed toward strengthening their economies.

Science: *Is geothermal energy likely to be important in Central America as a cheap source of power and a means to reduce oil imports?*

Hanold: Yes. Right now hydroelectric power is the most common form of electric power in Central America. But it will be difficult to increase hydroelectric capacity at the rate at which we expect the demand to increase. Hydroelectric plants are very expensive to build. A large plant costs over half a billion dollars, and it is becoming increasingly difficult for Central American countries to borrow that kind of money. Also, two installations in Central America were plagued with technical difficulties that led to cost overruns and reductions in power output. For example, at a new plant in Guatemala, the tunnels that carry the water to the turbines are collapsing because the geologic formations the tunnels pass through are unstable. We have technical problems with dams in this country too. Some of the problems are beyond human control. For example, clearing a valley for a reservoir causes erosion, and the access roads to the plant make it easier for people to cut down trees in the watershed. The deforestation that may result causes more erosion and silting in the reservoir, which, in turn, shortens the useful life of the facility, sometimes dramatically.

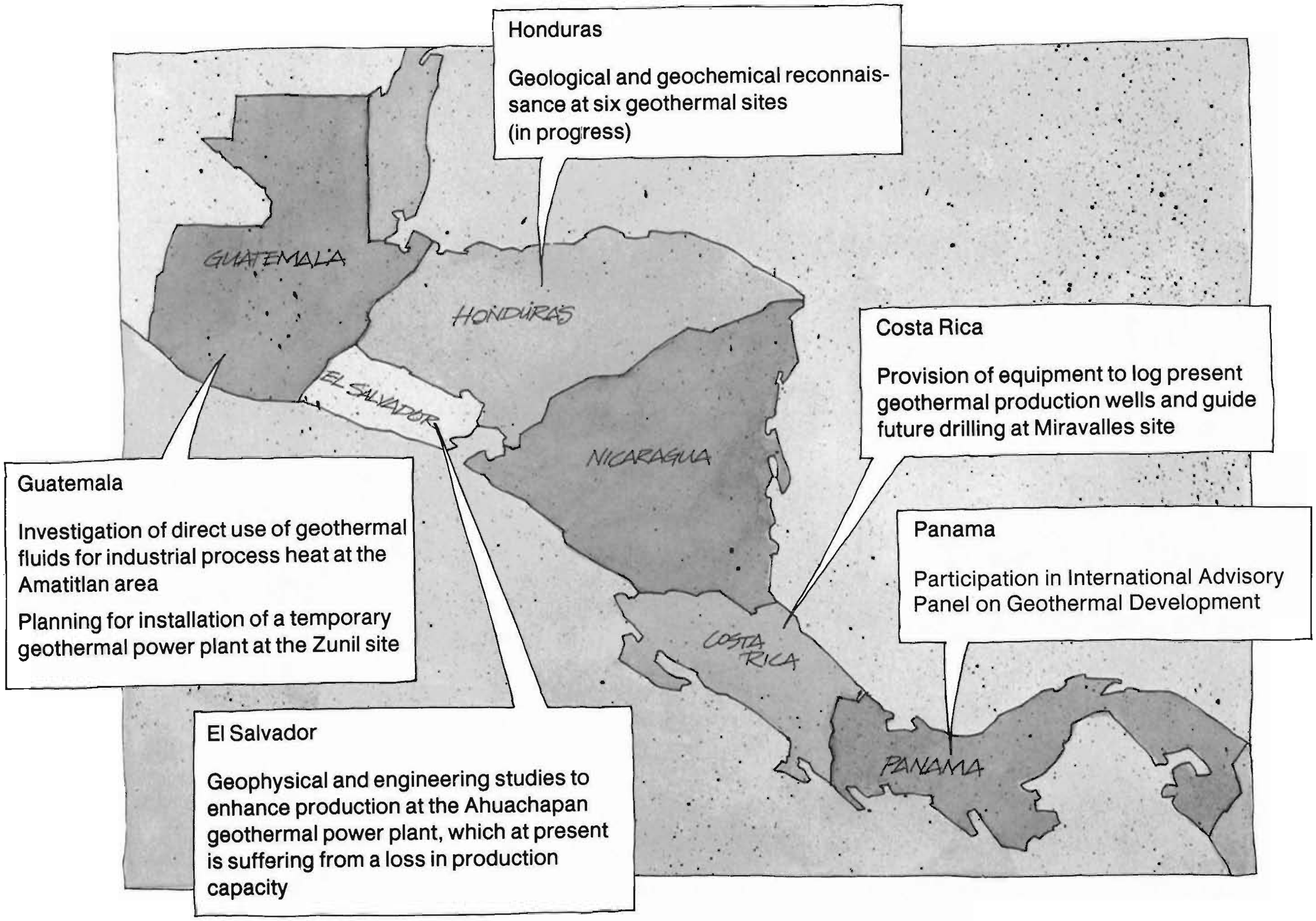
Loose: In our review of the available energy technologies in Central America, we found that Nicaragua and El Salvador had been generating substantial amounts of electricity from geothermal energy. El Salvador was generating over 40 percent of its power from a single geothermal plant called Ahuachapan. That plant has played an important role in the development of the country.

Science: *Where does imported petroleum fit into the picture?*

Hanold: Imported petroleum is needed for industrial and commercial uses and to fuel emergency power plants, the so-called thermal plants that come on line when demand is very high or when problems arise with the hydroelectric plants. One unavoidable problem is the dramatic decrease, by a factor of over a thousand, of the flow rates in the rivers from the wet to the dry season. Many of their plants are so-called run-of-river plants that can operate only when the water is flowing. During the dry season, when the flow ceases, they shut down.

Science: *So is geothermal power very appealing to the Central Americans?*

Hanold: It is ideally suited. The region has tremendous geothermal resources, as evidenced by the volcanic activity that extends from Mexico to the northern part of Panama. The experience of Nicaragua and El Salvador with the Momotombo and Ahuachapan geothermal plants and the recent drilling of successful geothermal wells in Guatemala and Costa Rica make it clear that almost all of the Central American countries can cash in on this indigenous resource. In addition, experience with the Ahuachapan plant makes it clear that the power can be very inexpensive. We are involved in geothermal development in all five countries mentioned earlier, but since Honduras lies



Roles played by Los Alamos scientists in the development of geothermal energy in Central America. Need and stage of development dictate the extent and nature of the support.

slightly outside the volcanic belt, defining the nature of its geothermal resources has provided us with the most challenging project in the area. We are now getting the field data back, and early indications are encouraging. The Honduran geothermal sites are not the typical volcanic variety found in El Salvador or Costa Rica but the Basin and Range type prevalent in the western United States. The Hondurans expect to reach a hydropower shortfall by 1993 and need to know by 1988 whether geothermal development can provide an alternative source of power. We still have a lot of work to do before we can assure them of the success of geothermal development. [See "Geology of Honduran Geothermal Sites" and "Geochemistry at Honduran Geothermal Sites."]

Science: *Tell us something about your personal reactions to working in Central America. For example, how do you feel about working in countries where the political situation is volatile?*

Hanold: It is not without strain, but we're getting used to it. When I first went to El Salvador, I was rather disconcerted by the presence of armed guards at the entrances to the offices of many government organizations. The directors of these organizations are usually political appointees, and one can't help but be aware that their personal safety may be at risk. Even our technical colleagues in these countries are not immune from political difficulties. All these things are upsetting at first. But the Central Americans are so pleasant to work with, so enthusiastic about our presence, and so appreciative of our help that I really look forward to the trips. The political atmosphere is just a fact of life; it's not ideal, but it doesn't prevent our doing an important job for these people.

Loose: And the Central Americans see our job as very important. Our activities are given a big play in the national presses. Even a series of lectures by Los Alamos scientists is accompanied by a great deal of fanfare and publicity. Sometimes the respect they give us is a little bit embarrassing because our technical opinions carry so much more weight with the local authorities than those of their own technical people. Fortunately, this does not seem to cause jealousy or ill feelings. On the contrary, our relationships with their technically trained people have been the most satisfying aspect of the entire project. They watch our people work in the field, they see them get their hands dirty, they see them work long hours with little supervision, and they admire what they see. Our ease of getting things done is surprising to them because they would require many levels of supervision and intricate advanced planning to accomplish similar tasks.

Science: *Are the Central Americans at all wary of your presence because of past experiences?*

Loose: Wariness is an appropriate word because they have been disappointed in the past. Some projects by foreign investors never get finished. We are working to build a strong relationship with these countries, not in the political sphere but rather through one-on-one relationships with the technical people of the country. By conveying a true image of American professionals, we are gaining trust.

Science: *In terms of the politics, does the State Department help you?*

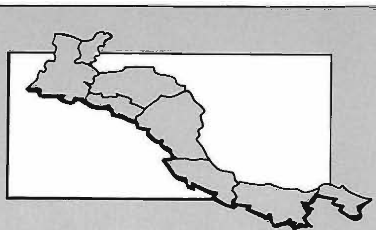
Hanold: Yes, through the AID. Normally our first stop in a country is at the AID mission office. We brief them on the purpose of our trip, and they give us an up-to-the-minute report on the country, advising us about any tense situations and any regions of the country that we should avoid. As we are leaving the country, we give the mission office a report about what was done during the trip and how successful it was.

Loose: The AID missions have been very helpful because they have a corporate memory of the country's history and an extensive network of contacts. Many of the AID people speak fluent Spanish. Most important, they have a finger on the pulse of the country. They are able to give us reliable information on how to get the job done and on which people and organizations are likely to be effective.

Hanold: They are also very helpful with more mundane things, such as providing transportation and meeting us at the airport on late-night flights and seeing that we get to our hotel safely. Many of our field operations take place in very, very remote areas, and the field crews may come through a city only every five or six days. The AID mission has our itinerary and knows the people involved. If a problem were to arise, we feel confident that they would take it upon themselves to go out after our people.

Loose: The AID and Los Alamos have one aim in common—to see that when something is started, it is finished. We have an effective working relationship with the AID, and it is improving daily.

Hanold: In technical areas we act as a filter for the AID. We hear requests from local experts for training or for equipment, and we separate the technically less important requests from those that have merit. We submit to the AID the proposals that we think will have the greatest economic impact on the country. From the beginning our intention has been to involve people from the host countries in all stages of the



The Costa Rican Peat Project

by Gary R. Thayer, K. D. Williamson, Jr., and Arthur D. Cohen

Scientists from the Laboratory and Refinadora Costarricense de Petróleo (RECOPE) are working together to assess the development potential of peat, an indigenous, unused resource in Costa Rica. This carbon-rich organic sediment, produced in swamps and marshes from partially decayed organic matter, could become a significant asset in a number of different ways. If made into briquettes and used as a fuel for heating and cooking, it could help reduce the heavy dependence on fuelwood, which now supplies 50 percent of Costa Rica's energy. If used to fuel electric power plants, it could help reduce oil imports. Since harvesting of peat is a labor-intensive operation, its development would provide jobs for the people in the areas where it is found. Further, its availability as a fuel might bring industries to those areas. Eventually peat might become the basis of "high-tech" industries converting this resource into liquid and gaseous fuels or valuable organic chemicals such as waxes, resins, and medicinals.

Despite all this promise and the extensive literature documenting the wide and growing use of peat in northern Europe, Ireland, and the Soviet Union, the development of peat in Costa Rica entails facing many unknowns regarding harvesting methods, appropriate and acceptable end uses, and overall economic impact.

So far we have surveyed the literature on harvesting and end uses and have made projects, including planning. Even for our original proposal to the AID, we solicited their opinions concerning their biggest energy- and mineral-related problems. And then we tried to address those problems squarely in the proposal.

Loose: We worked with geologists, engineers, economists, and fairly senior administrators of both utilities and government organizations such as ministries of energy and mines.

Science: *Had they had much contact with Americans before?*

Hanold: Yes. Since many Central American schools do not offer advanced degrees, many technical people have pursued such degrees in the United States. As a result most of our Central American counterparts speak English very well.

Loose: The AID regards education as a basic human need. In addition to their program of education at the primary and secondary levels, they help people with university potential to get an appropriate education at American universities.

Hanold: The local culture and pride are such that they don't want someone just coming down, doing a job, and walking away. They appreciate the assistance but want to participate in the doing. They want to be involved technically and physically. In all our interactions we stress working with the people in the region. We have the techniques to do certain tasks, and through our work they get exposed to the cutting edge of geology, volcanology, geochemistry, and geophysics. That experience will be left behind. In some cases we are actually leaving equipment behind so that they can continue on their own.

Loose: The people are very nationalistic and very proud of their countries. They have definite ideas on how to use the new knowledge we are giving them. A concrete example is the peat project in Costa Rica. We were very interested in the possibility of briquetting the peat for use as a cooking fuel in place of wood. It was a way to save the fast-disappearing forests. The national oil refinery, whose charter is to promote high-tech industry, was initially more interested in using the peat to produce petrochemicals, waxes, and resins and to generate electricity. Through discussion we have reached a mutual agreement to investigate both high- and low-tech uses of peat. That gives you an idea of how the input of local people is reflected in the projects. [See "The Costa Rican Peat Project."]

a preliminary field assessment of Costa Rica's peat resources. Two moderate-size peat bogs have been identified. Other sites exist but have yet to be explored. One identified site is in a sparsely populated region on the Nicaraguan border whose settlement would promote Costa Rica's national security. The other is a jungle site near the Caribbean coast. Both sites are large enough to provide fuel for a 10-megawatt electric power plant for 100 years or more. One of these locations may be chosen as the site for a demonstration peat project.

Peat, in its natural state, contains up to 95 percent water and must be dried to a water content of 50 percent or less before it can be burned. In Europe solar drying is used almost exclusively to reduce the water content. Milled peat is produced by draining the top few centimeters of a peat bog, scraping off the exposed layer, allowing the sawdust-like product to dry in the sun for a day to a week, and then collecting it with rakes or large vacuum cleaners. Sod peat is produced either by cutting out

A view of the flora characteristic of the peat site located in a tropical jungle region of Costa Rica near the Caribbean coast. The peat deposit here is extraordinarily thick, at least 12 meters. In the center of the photograph is a species of palm that is often associated with peat deposits.

Science: *What difficulties do you face in carrying out your work?*

Loose: Coordination, communication, and logistics are among our biggest difficulties. We have Washington looking over our shoulders as well as the local AID missions. In addition, we maintain contacts with government people in the five countries. The wide range of the projects, which include economic analysis, mineral exploration, and development of geothermal power, compounds our difficulties.

Hanold: Another difficulty is that sometimes the technology involved in a project is unfamiliar to the people in the host country, and they find it hard to understand how the project will get from A to Z. For instance, a geothermal energy project starts with geologists mapping the area and geochemists collecting water and gas samples from hot springs. These activities may seem mystifying. What do hot water samples have to do with electricity on line? To lessen this problem, we recently took two Honduran visitors to an operating geothermal power plant in the United States. These gentlemen were from the national organization that generates much of the electricity in Honduras, mostly from hydropower. They were at Los Alamos to help make long-range plans for a geothermal project in their country, and we spent a day at the Geysers plant in northern California as preparation. The Geysers, the largest geothermal power plant in the world, supplies much of the power for San Francisco. We took them up on drilling rigs, showed them how the wells are drilled, talked about the site's geology, geochemistry, and geophysics, and showed them how the geothermal steam is extracted, collected, and run through turbines at the power plant. The tour gave them a concrete understanding of the whole technology. I think many of our projects will require a similar educational effort.

Science: *Are you in contact with the rural populations of the Central American countries?*

Hanold: Very much so, particularly during our geothermal work in the remote areas of Honduras and our explorations for minerals in very remote areas of Costa Rica. Since these areas have no conventional hotels, arrangements are made with local people for sleeping accommodations and food. We usually hire someone local to buy and cook food for the field parties

---

blocks of peat by hand and stacking them to be dried or by cutting sections of peat by machine, grinding the peat, and extruding it in 2- to 10-centimeter-diameter cylinders. These "sods" are allowed to dry and then collected. Since both methods involve a solar drying step, they may be impractical at the two Costa Rican sites, which receive between 3 and 5 meters of rain per year. Instead, the peat may have to be dried artificially. Alternatively it can be collected in a slurry and heated to about 200°C to initiate exothermic oxidation reactions that produce free carbon, which can be collected and compressed into a coal-like substance, or, if the oxidation is carried to completion, heat for industrial use. Such wet harvesting methods are generally more expensive than the traditional methods described earlier, so their use in Costa Rica will make the economics of peat development different from that documented in the literature.

The environmental impact of peat harvesting and the economics of various end uses are also being examined. Mining of peat results either in changed drainage patterns or lake formation, both of which can be beneficial. Changed drainage patterns might permit reclamation of the mined areas as prime farm land, or the lakes could, according to a study done in Jamaica, be used for aquaculture.

Appropriate end uses for peat depend, first of all, on the quality and size of the resource. The higher ash content of presently known Costa Rican peat versus European peat may affect production technology and costs as well as quality of the end products. This possibility needs to be investigated. We are hopeful that the Costa Rican peat will be suitable as a fuel for electric power plants. Although these plants will be smaller in scale than typical European plants and the cost per unit output will therefore be higher, the additional cost might be offset by reduction in oil imports. We are most excited about the use of peat as a cooking fuel in place of wood, but the local population may not be equally enthusiastic, especially in areas where wood is plentiful and free for the gathering. The Costa Ricans are particularly interested in high-tech uses, such as the production of gaseous fuels and organic chemicals, but these may be too ambitious technologically for a first attempt at using the resource.

We and our Costa Rican colleagues are gathering information relevant to all these issues. We will also test the performance of peat in local cooking stoves and in larger scale combustion and gasification applications. The results, plus detailed information on specific peat sites, economics of harvesting, environmental effects, potential for reduction of oil imports, and development of remote areas, will be used to choose a demonstration peat project. We, as well as the Costa Ricans, look forward to finding a way to make this resource an economic success. ■

and to provide some structure, even an empty house, for sleeping. It is definitely not a downtown Marriott.

Science: *Are geothermal sites part of the culture of the people?*

Hanold: In the more primitive parts of Honduras, people use the hot water for simple needs such as sterilizing baby bottles, boiling eggs for lunch or cooking chickens for dinner. Information on the locations of many hot springs has come from the natives. They'll watch from a distance and then approach us and indicate that if we like that hot spring so well, they know of another one over there. We always make sure that at least one member of the field team speaks Spanish fluently so that communication is always possible.

Science: *The interactions must be quite different from those you have been accustomed to in this country.*

Hanold: Very definitely. Through the DOE we have been technically involved in many geothermal energy programs with American industries. Our experiences in Central America have a very different flavor. The people are exuberant and naturally inquisitive. After working in the field with our people, they ask question after question about what we did and why we did it. So we are training them as we go.

Science: *Which Laboratory divisions are involved in the project?*

Hanold: Geologists, volcanologists, and geochemists from the Laboratory's Earth and Space Sciences Division are doing the geothermal reconnaissance in Honduras and Costa Rica and the mineral exploration in Costa Rica. Some of our people have been in the field almost every month since the project started. These people have a great deal of experience with geothermal systems in the Rocky Mountain region of the United States and throughout the world—Europe, Asia, and St. Lucia. Soon engineers will be needed to log the geothermal wells, that is, to measure the temperature, pressure, and flow rate of the fluids in the wells and determine the nature of the rock formations in the reservoir. Central America has almost no equipment for logging hot wells. The measurements in Costa Rica, Guatemala, and El Salvador will be made with special equipment developed at Los Alamos to function in the high-temperature—up to 240 degrees Centigrade—environment of geothermal reservoirs. [See "High-Temperature Borehole Measurements at Miravalles, Costa Rica."]

Loose: Economists and energy technologists from the Laboratory's Systems Analysis and Assessment Division are also heavily involved in the project. All told, between sixty and seventy people from throughout the Laboratory contribute directly, but that number does not include the many people who provide support, such as working on publications, purchasing equipment, and arranging travel. These people have been very helpful about expediting or bending the Laboratory's procedures to get the job done.

Hanold: The project would rapidly grind to a halt if we followed all the conventional foreign travel rules. Thanks to the Travel Office staff, the approval time for foreign travel has been reduced from sixty days to two hours. They really deserve a pat on the back for accommodating our people.

Loose: So do the people who process our purchasing contracts in record time. We do other unorthodox things, such as hiring Central Americans as consultants and paying for local management offices. The Laboratory does not normally do business that way.

Hanold: We have many logistics problems. Our field people are only now getting access to places to store their gear in the host countries so they don't have to bring it back after each field trip. We had to arrange special assistance from the host countries to get equipment and samples through customs into the host countries and samples back to Los Alamos for analysis.

Science: *What types of equipment do you send down?*

Hanold: To sample the springs in Honduras, for example, we needed thermometers, pH meters, conductivity meters, and bottles for shipping samples back to the Laboratory for further analysis.When we start the geophysical studies and well logging, we will bring in very large pieces of equipment—high-voltage power supplies, diesel generators, a logging truck, and so on. The logging truck is equipped with a spool of special cable about 10,000 feet long, a data-acquisition system, a computer for data processing, and its own power supply. The cable is used to lower various tools into the well and to transmit signals from those tools back to the data-acquisition system in the truck. One can collect real-time data on what is taking place in the well with this self-contained truck. A similar setup is used at the Laboratory's Fenton Hill hot dry rock geothermal site near Los Alamos and at various locations in Nevada and California. We may have some trouble getting this equipment and the drilling rigs into some of the remote sites in Central America. So far we have needed only small four-wheel-drive vehicles and have had no trouble, but we have yet to experience a full rainy season.

Science: *How much money is dedicated to the project?*

Loose: During this first year and a half we have $10.2 million, but in charting our course for the next five years, we see the costs increasing to about $15 million annually by 1986. So the project represents a substantial effort for the Laboratory and for Central America. A fair fraction of the funds will be spent on salaries of the core group of between thirty and forty Los Alamos people who work on the project.

Hanold: The remainder will be spent on goods and services. We are fabricating and buying equipment that will be left behind in Central America for use throughout the region. Since we will also be renting equipment from Central American countries to drill shallow wells, a fair portion of the money will be spent in the host countries.

Loose: A significant portion of the total will go to the U.S. Geological Survey, who

Wilfred Gutierrez standing next to a boiling spring at Platanares, Honduras. Note the white deposits of silica sinter, an indication of subsurface reservoir temperatures equal to or greater than 150°C. (Photo by Fraser Goff.)

will participate in the mineral and geothermal work. Consultants, universities, and other organizations will also participate and receive a portion of the funds.

Hanold: One of the exciting parts of the project is the opportunity to sponsor graduate students. Central Americans who are capable of doing excellent research but have been unable to get funding will be supported to do research in their countries. I think this aspect of the project will prove to be one of its biggest successes.

Science: *Let's talk in a little more detail about the technical work. What do the economists do?*

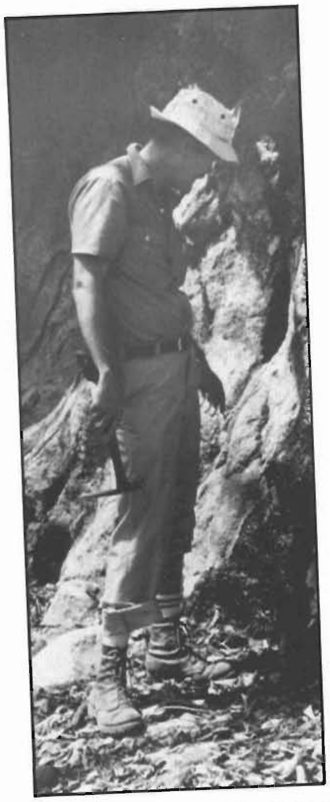
Loose: We have had teams in all of the countries except El Salvador. They interact with the economic ministries and energy-planning ministries to collect data on energy consumption, energy production, and energy resources. This information will be used as a database for analyses—with small computers we will provide—to determine directions for energy development. The economists here at the Laboratory have a lot of experience with such analyses for our own country, which is asking the same question. How can our petroleum imports be reduced? In the case of the geothermal projects, our economists work with the local economists and energy planners to help characterize the available energy options. In St. Lucia, for example, we compared the costs of producing electricity by various methods and the impacts of these methods on the overall economy and on petroleum imports. The St. Lucians asked for these cost comparisons, and our analysis was based on the information provided by Los Alamos geologists and engineers. We also looked at the macroeconomy of the country to forecast the growth in energy demand and to determine the energy capacity that should be installed to meet that demand.

Science: *Bob, would you outline the geothermal work?*

Hanold: A geothermal power system requires a hot fluid, usually called brine, that consists mainly of mineralized water and steam. The temperature of the brine should be at least 180 degrees Centrigrade to generate electricity efficiently. The rock containing the fluid must be well fractured or porous, since fluid tightly trapped in impermeable rock cannot be accessed. In Honduras we began the search for suitable geothermal reservoirs by studying in greater detail the geology of six areas where high-temperature springs were known to exist. Then we sent a geochemistry team to collect and analyze samples of the surface waters and predict the temperatures of the underground reservoirs. Over the last ten or fifteen years geochemists have developed empirical correlations between the properties of surface fluids and the temperatures of underground reservoirs. This area of geochemistry is called geothermometry. Our geochemists have a great deal of experience with geothermometric techniques, which worked very well as predictive tools in St. Lucia.

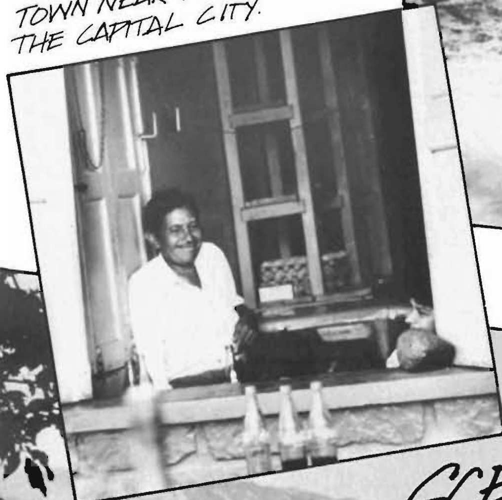
One of the best known geothermometric techniques is use of the sodium-potassium-calcium geothermometer, which provides an estimate of the temperature of a reservoir from the relative abundances of those elements in the surface fluids. According to this geothermometer, the reservoir temperatures at two Honduran sites, Platanares and San Ignacio, are well above the 180 degrees required.

SCENES FROM HONDURAS



GRANT HEIKEN EXAMINING PADRE MIGUEL IGNIMBRITES AT THE PLATANARES GEOTHERMAL SITE.

HONDURAN IN A RESTAURANT IN SUYAPA, A SMALL TOWN NEAR TEGUCIGALPA, THE CAPITAL CITY.



LA MASICA HOT SPRINGS, SAMBO CREEK GEOTHERMAL SITE.

REGION NEAR THE PAVANA GEOTHERMAL SITE.

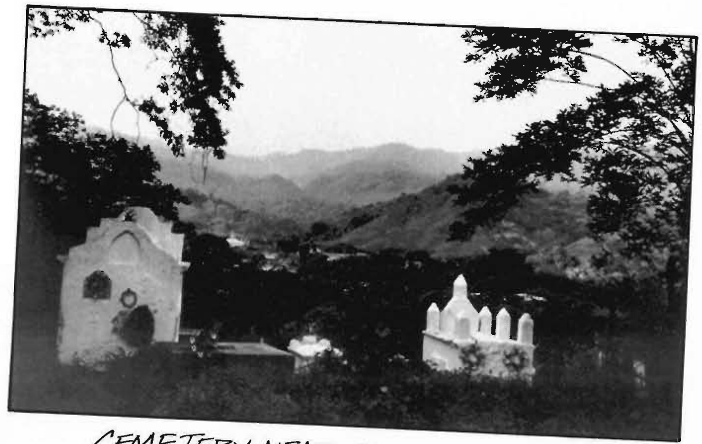
LITTLE GIRL IN AZACUALPA.

FRANK PERRY AND SCOTT BALDRIDGE WORKING ON THE AZACUALPA TECHNICAL REPORT AT LAGO DE YOJOA.

FARM ANIMALS NEAR THE SAMBO CREEK GEOTHERMAL SITE. HONDURAS HAS A BURGEONING BEEF INDUSTRY, AND STEAK IS OFTEN THE CHEAPEST MEAL ON THE MENU.

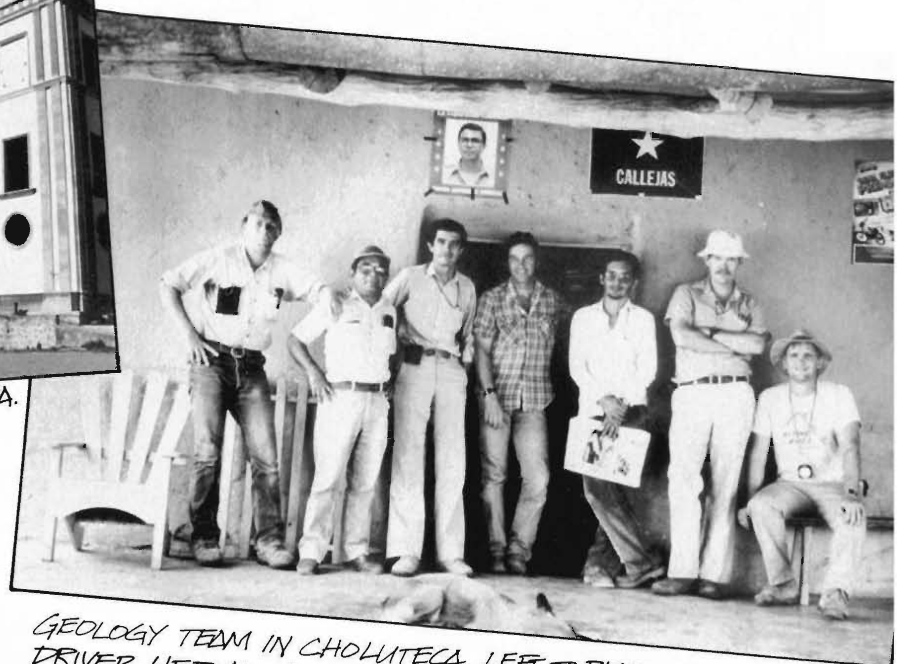
HONDURANS IN A LOCAL STORE, SAN FRANCISCO DE OLIVERA.

GIRLS CARRYING WATER FROM A WELL NORTH OF EJOC.

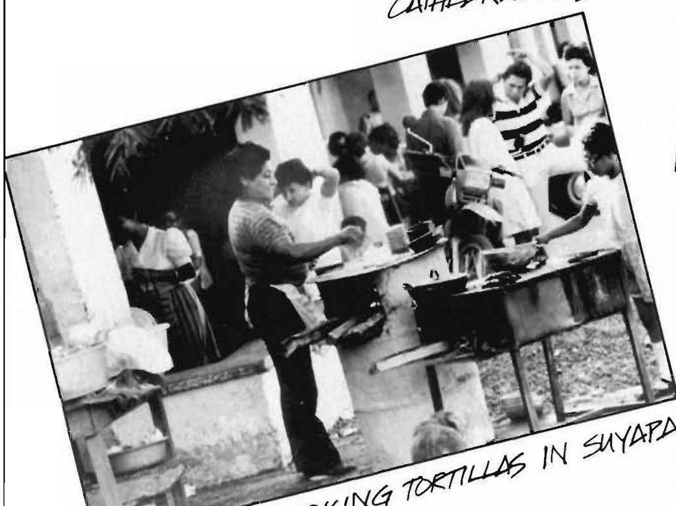


CEMETERY NEAR SAN PEDRO ZALAPA.

CATHEDRAL AT SUYAPA.



GEOLOGY TEAM IN CHOLUTECA. LEFT TO RIGHT, DEAN EPPLER, DRIVER HERMAN, WILMER FLORES, WENDELL DUFFIELD, RODRIGO PAREDES, GRANT HEIKEN, KEN WOHLETZ.



HONDURAN COOKING TORTILLAS IN SUYAPA.

DEAN EPPLER PASSING OUT POLAROID PRINTS TO CHILDREN IN AZACUALPA.

CAMPESINOS NEAR SAN FRANCISCO DE OJUERA.

Some participants in the Caribbean Basin Proyecto. Front row, left to right: Ken Wohletz, Tony Montoya, Flavio Gurule, Ron Lohrding. Middle row, left to right: Steve Bolivar, Gary Thayer, John Altseimer, Gloria Bennett, Anne Tellier, Bob Hanold, Annette Turpin, Dean Eppler, Pat Aragon, Bill Laughlin. Back row, left to right: Joe Frank, Duane Marr, Fred Edeskuty, Grant Heiken, Ken Williamson.

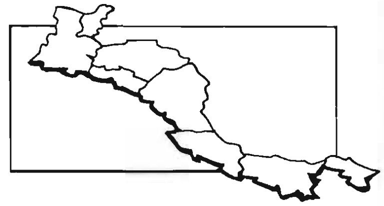
Other data corroborate these measurements, so our confidence in the geothermal potential of Honduras is growing. [See "Geochemistry at Honduran Geothermal Sites."]

We also determine a gas geothermometer temperature from the relative abundances of various gases in the surface fluids. These gases are either dissolved in the fluid or are contained in the steam that reaches the surface through a fumarole, a steam vent. Geologic data, such as rock composition and age and the geologic history of the site, are also valuable in geothermal exploration. For instance, magma, molten lava, is usually closer to the surface in younger geologic systems. So geothermal reservoirs in younger systems are easier to develop because higher temperature fluids occur at shallower depths. We have come to understand that, in Honduras, the high temperature gradients near the surface are due to thinning of the earth's crust by tectonic processes. [See "Geology of Honduran Geothermal Sites."]

We also send our geophysicists to some sites to perform electrical measurements. With a high-voltage generator they establish an electric field in the ground, and then with probes they measure the electric field at various distances from the source. If the ground had a constant resistivity, the field would vary with distance in a certain way. Any perturbations on that variation reflect changes in resistivity, which in turn indicate the presence, for example, of fluid-filled faults and fractures. From the data the geophysicists can determine the locations, both horizontally and vertically, at which such features occur. This information tells us about the plumbing of the reservoir. However, since the electrical measurements are time-consuming and expensive, we limit them to only those sites with the most promising geologic and geochemical properties.

Science: *Over how big an area are the electrical resistivity measurements made?*

Hanold: In St. Lucia we ran a 5-kilometer-long resistivity survey cable through steep,



wet, dense jungle. The entire trail had to be cut with machetes. We also measure temperature gradients by drilling holes with fairly small diameters to a depth of 300 to 500 meters. The higher the temperature gradient is, the more promising the site is. After all this information is collected and analyzed, we pick locations to drill the production wells. In St. Lucia we picked three locations along the 5-kilometer resistivity line.

Science: *Does the fluid have to be pumped out of the reservoir?*

Hanold: Fortunately, pumping is unnecessary for most high-temperature geothermal reservoirs. The fluid boils as it rises in the well, and the emerging steam lifts the fluid by lowering its density.

Science: *Does a geothermal reservoir cool off as it is used?*

Hanold: Not by very much. It can lose its pressure but only gradually and after long production. Some reservoirs are very large, extending vertically for 2000 to 3000 feet and horizontally for miles. The rock is essentially saturated with brine. Much of the fluid that reaches the surface through the well is injected back into the earth and is reheated as it percolates through the hot rock.

Science: *Will the economist summarize the point of the project?*

Loose: If oil were still $1.50 a barrel, we wouldn't be carrying out this project. Economics provides the rationale for our helping these countries identify and develop their indigenous energy resources. If oil were still cheap, the best thing we could do would be to continue to burn oil.

Hanold: I'd like to say a little more about the people and the cultures we've encountered. The Central Americans are very adaptable and jump at the chance to learn new techniques. They are very eager and aggressive. If they take a short technical course, they want a diploma to verify their training because it may have an impact on getting a promotion or a raise. So we plan to grant diplomas for a course, now being developed, on conducting field work for geothermal reconnaissance.

Loose: In the Latin way of doing business, the bosses have a great deal of power, more than we are used to, and command the respect of the people who work for them. Respect for authority is very much a part of their culture. As a consequence, you have to deal with organizations at the appropriate level of authority. If you go too low, the person can't help you. If you go too high, you have made a faux pas. You have to understand and work within the hierarchy.

Science: *Does the style of communication differ from ours? When you want to get something done, can you be direct about it?*

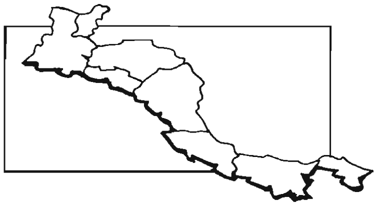
Hanold: We have been very direct. Meetings follow a pattern similar to ours, and it has certainly been my experience that we can deal around a conference table with the Central Americans much as we would with our fellow citizens.

Science: *What is the outlook for this project?*

Loose: We have been working in Central America for seven months and are now beginning to see more clearly both the obstacles and the promise. The initial euphoria of getting funding and the novelty of working in exotic places is being replaced by the real difficulties of implementing this multicultural, multipurpose project. We are learning that good intentions are not enough. Communication among the various arms of the project across thousands of miles and between cultures requires management skills beyond those normally required. We need to do more detailed planning and closer monitoring to avoid misunderstandings. On the other hand, we are more certain than ever that the project can have a significant impact. The peat work is particularly exciting because it may lead to a brand new technology in Central America. The geothermal work is very important because it may result in reduced oil imports and foreign debts. It will take tremendous dedication and perseverance to realize the goals of the project, but we and our Central American friends are up to the challenge. ■

Geothermal Projects

- **Geology in Honduras**
- **Geochemistry in Honduras**
- **Borehole Measurements in Costa Rica**



Geology of Honduran Geothermal Sites

by Dean B. Eppler

Since March 1985 a team of Laboratory geologists has been working with counterparts from the Empresa Nacional de Energía Eléctrica (ENEE) of Honduras and from four American institutions on a project to locate, evaluate, and develop geothermal resources in Honduras. The team, headed by Grant Heiken and funded by the U.S. Agency for International Development, has so far completed three trips to Central America to study in detail the geology of six geothermal spring sites.

Basic Geology of Honduras

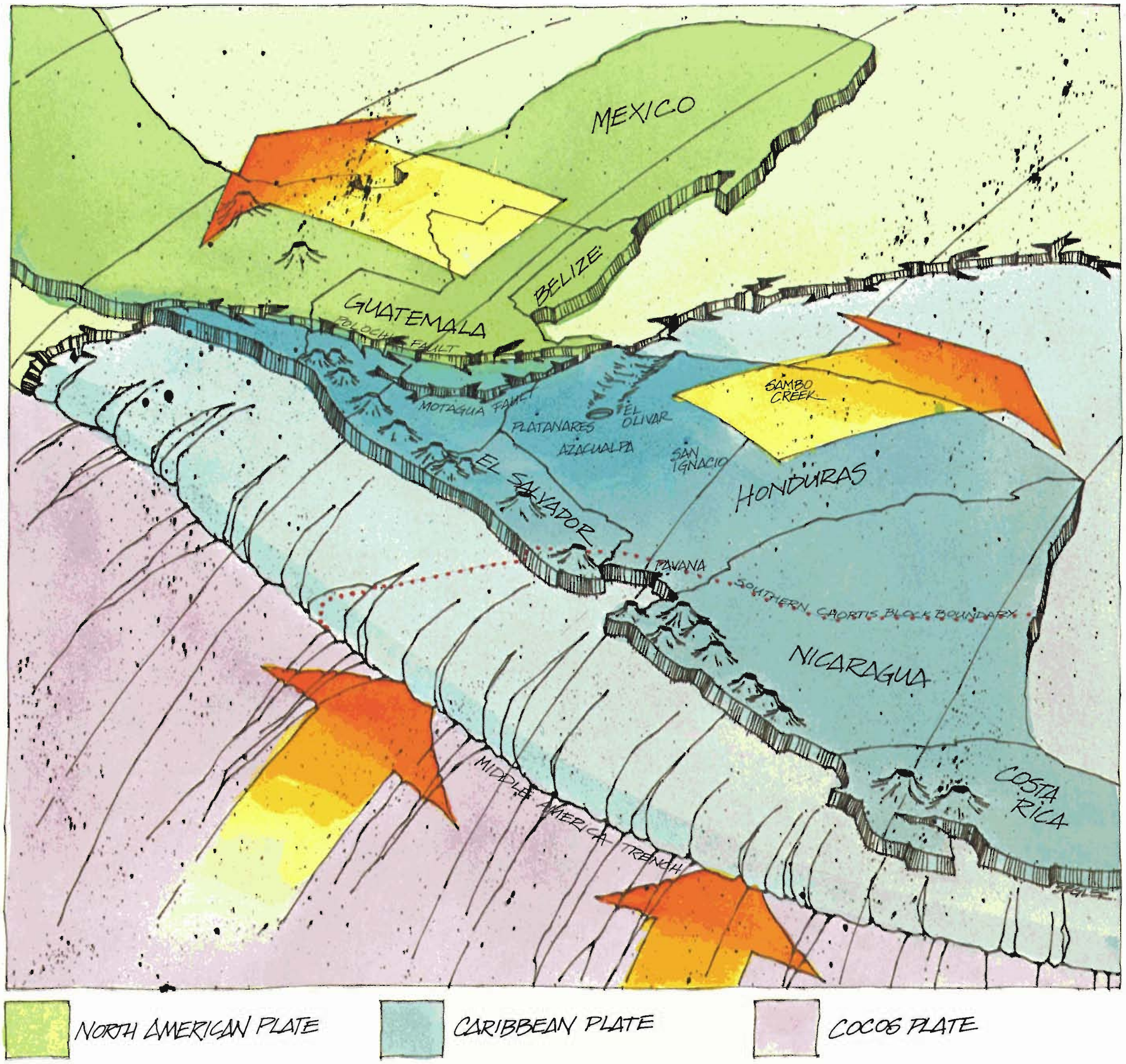
Honduras, the largest and most rugged country in Central America, is perhaps the least known geologically. Its steep terrain, dense vegetation, and paucity of roads hampered basic geologic studies until the late 1960s. Since then studies sponsored by American universities, including Ph.D. dissertations by project collaborators Bob Fakundiny and Rick Finch, have meshed with a greater level of in-country expertise to produce a basic understanding of the geology of the country. Such an understanding is an essential first step in any geothermal exploration. It has been particularly useful in Honduras as we set out to determine the nature of the geothermal heat source and the "plumbing system" through which the geothermal waters reach the surface.

The geology of Central America is extremely complex. The meeting of three tectonic plates in western Guatemala and southern Mexico has resulted in an unusual juxtaposition of structures and rock types whose geologic history has yet to be unraveled. Textbook reconstructions of tectonic-plate motions very often sidestep the problem of how Central America developed through geologic time by never showing its existence until the present time.

As shown on the accompanying map, Honduras lies on a portion of the Caribbean tectonic plate called the Chortis Block. This block, composed of rocks deposited in a continental environment, is bounded on the north by large strike-slip faults in sourthern Guatemala (the Matagua and Polochic faults) that form the boundary between the Caribbean plate and the North America plate.

The continental rocks of the Chortis Block are bounded on the south by younger rocks in Nicaragua that were deposited in an oceanic environment. The western boundary of the Chortis Block lies along the Central American volcanic chain and the Middle America Trench, a subduction zone where the Cocos plate is being thrust under the Caribbean plate. The complex geology of Honduras is the result of its proximity to the intersection of the three tectonic plates. In some areas of the country, major faults lie less than 10 meters apart. Most of these are normal faults, developed as a result of stress that is literally pulling the country apart along an east-west axis. Although Honduras has been spared the devastating earthquakes that have rocked much of Central America, we suspect that deformation is taking place continually; in some areas faults cut stream gravels that are only several thousand years old. The result of this faulting, as shown in the accompanying photo, is rugged topography dominated by north-south oriented fault basins and adjacent fault-block mountains very similar to those found in the Basin and Range physiographic province of the western United States.

The rocks of Honduras were deposited in rapidly changing environments, and the resulting stratigraphy is as complex as the structures that modify it. Precise dating is difficult because of the absence of identifiable fossils and the rapid changes in rock types over short geographic distances. However, three distinct age groups are apparent: a basement complex of Paleozoic low-grade metamorphic rocks about 245 million years of age (Horne, Clarke, and Pushkar 1976); an overlying section of Mesozoic limestones and redbeds that is estimated to be between 100 million and 200 million years of age (Mills et al. 1967); and a thick upper sequence of volcanic rocks from two distinct episodes of volcanism. The Matagalpa Formation, a



Honduras is positioned on the Chortis Block near the junction of three tectonic plates: the North American, the Cocos, and the Caribbean. The large arrows indicate the direction of motion of the plates. The Cocos plate is being thrust under the Caribbean plate along the Middle America Trench. The Motagua and Polochic faults are large strike-slip faults separating the North American plate from the Caribbean plate. The plate-tectonic and geologic histories of the area are not known well enough to explain how and when Central America was formed. For example, the southern boundary of the Chortis Block, where continental rocks end and oceanic rocks begin, is indicated by a dashed line because its exact location in the jungles of Nicaragua has not been determined. We do know that plate movements are continuing to create faulting throughout Honduras and pulling the country apart along an east-west axis. Rainwater circulating through the fault regions has created numerous geothermal systems. The map also shows the locations of the six geothermal sites now being evaluated as indigenous sources of energy.

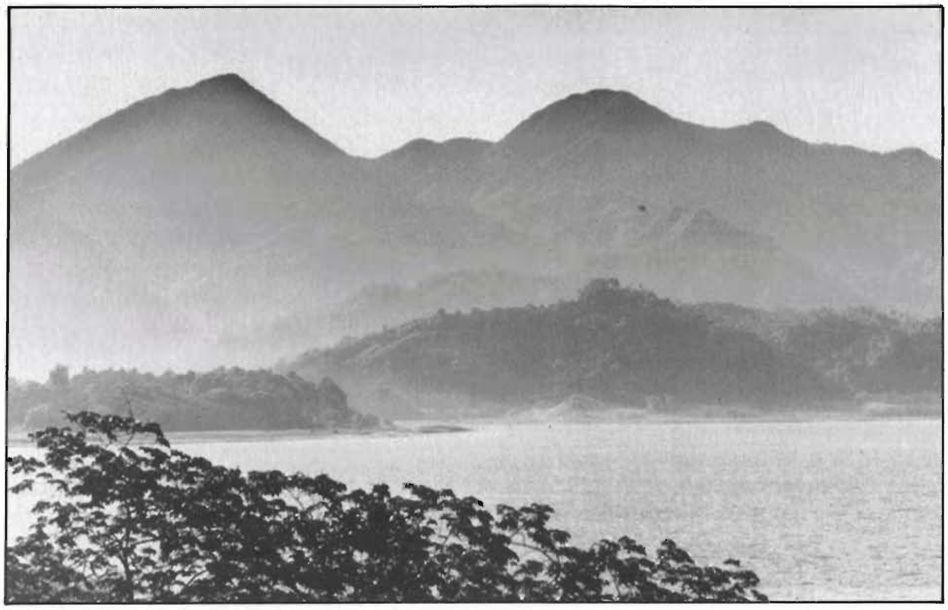
series of early Tertiary interbedded lava flows, pyroclastic flows, debris flows, and interbedded water-laid sediments, is between 40(?) million and 60(?) million years of age (McBirney and Williams 1965). The Padre Miguel Group, the result of the second episode of volcanism, is a thick sequence of ignimbrite similar to the Bandelier Tuff and is found throughout the southern half of Honduras; it is between 15 million and 20 million years of age (Williams and McBirney 1969).

This bare outline of the geology of Honduras will have to be filled in by studies of individual drill holes before we can infer with any confidence the nature of the plumbing system at each geothermal site.

Studies of Geothermal Sites

In the late 1970s several American firms began preliminary geothermal explorations in Honduras but were unable to complete them because of economic difficulties. These reconnaissance efforts allowed selection of six promising geothermal sites. However, the origin of the geothermal resource was misunderstood and incorrectly attributed to recent volcanism rather than, as our studies now indicate, to tectonic processes. Identification of the nature of the geothermal resource is a major contribution to the project. The amazing abundance of hot springs in Honduras suggests a large geothermal resource. Consequently, the project has two goals: selection of two geothermal sites for further development on the basis of detailed studies, by Los Alamos and ENEE geologists, of the six previously identified sites; and identification of other promising geothermal sites on the basis of a country-wide inventory of hot springs by ENEE with technical support, as necessary, from Los Alamos.

Detailed geologic studies have so far been carried out at three sites: Platanares, San Ignacio, and Azacualpa. Concurrently a team of geochemists from the Laboratory, the U.S. Geological Survey, and ENEE has sampled and analyzed the thermal waters to determine their chemistry and estimate the temperatures of the geothermal resevoirs (see "Geochemistry at Honduran Geothermal Sites").



Fault block mountains on the east side of Lago de Yojoa, Honduras.

Platanares. This site, located in the western portion of Honduras, is similar to many being developed in the Basin and Range province of Nevada. That is, water is heated deep underground and rises to the surface along faults. The numerous hot springs at Platanares are found in lavas, tuffs, and tuffaceous sediments of the Padre Miguel Group. The faults appear to be extensional, and the presence of wedges of gravel perched above the present water level in the Quebrada del Agua Caliente (Gorge of Hot Water) suggests relatively recent movement on these faults. The hottest springs are associated with faults that trend mostly northwest and north. Thermal energy is being released from boiling springs and numerous fumaroles. Since the stream that flows through the gorge is 10 to 15°C hotter in the area of the hot springs than it is upstream, additional energy is probably being released from submerged springs. Estimates of the thermal power of this area are given in "Geochemistry at Honduran Geothermal Sites."

San Ignacio. This site, located on the north side of the fault-bounded Siria Valley, also appears to be a geothermal system of the Basin and Range type. Hot springs are located at the intersection of a young northwest-trending fault scarp with older north-trending faults. These faults are also extensional, and, again, recently cut deposits of stream gravel suggest recent movement. The rocks within the area are primarily Paleozoic metamorphic schists containing some remnant patches of Padre Miguel Group tuffs. More than one hundred springs were mapped, many of which surface in terraces formed in deposits of silica-cemented gravel.

Azacualpa. This site, located in highly faulted sedimentary rocks that bound a major fault basin (the Santa Barbara graben), also appears to be a geothermal system of the Basin and Range type. The hot springs and fumaroles are surfacing along segments of the Zacapa fault, which cuts limestones and redbeds of Cretaceous age.

A view of the Platanares geothermal site in Copán, Honduras. Water from the hot springs enters the stream in the foreground, which flows through the Quebrada del Agua Caliente. Note the clouds of steam rising from individual fumaroles.

Summary

Our studies so far suggest that the geothermal manifestations in Honduras originate in a Basin and Range type of geothermal system, in which meteoric water (rainwater) flows downward along extensional faults, is heated, and rises back to the surface along other faults. In the Basin and Range geothermal systems in the United States, the heat is a by-product of the elevated geothermal gradient that develops when the earth's crust has been thinned by tectonic processes. We suspect the same heat source is responsible for the geothermal systems in Honduras, since the Padre Miguel Group of volcanic rocks is too old for residual heat to be the source of thermal energy.

Geophysical surveys are being planned for the spring of 1986 to answer questions about the size, depth, and location of the geothermal reservoirs, the regional heat flow, and the thickness of the crust. Plans are also under way to begin drilling shallow (about 500-meter) boreholes to measure the geothermal gradient. By combining all this information, we should be able to estimate the size and quality of the geothermal resources and to make recommendations to ENEE for future exploitation. ■

Participants in the geologic studies (and their institutional affiliations, if other than Los Alamos National Laboratory) are Jim Aldrich, Scott Baldridge, Wendell Duffield (U.S. Geological Survey), Dean Eppler, Bob Fakundiny (New York State Geological Survey), Richard Finch (Tennessee Technological University), Wilmer Flores (ENEE), Grant Heiken, Rodrigo Paredes (ENEE), Frank Perry, Napolen Ramos (ENEE), Alexander Ritchie (College of Charleston), and Ken Wohletz.

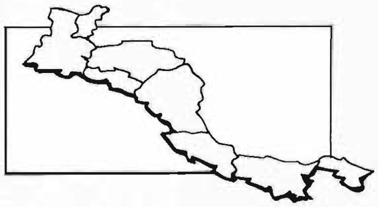
References

Horne, Gregory S., Clarke, George S., and Pushkar, Paul. 1976. Pre-Cretaceous rocks of northwestern Honduras: Basement terrane in Sierra de Omoa. *The American Association of Petroleum Geologists Bulletin* 60:566-583.

McBirney, Alexander R. and Williams, Howel. 1965. *Volcanic History of Honduras.* University of California Publications in Geological Sciences, vol. 55. Berkeley:University of California Press.

Mills, R. A., Hugh, K. E., Feray, D. E., and Swolfs, H. C. 1967. Mesozoic stratigraphy of Honduras. *The American Assoication of Petroleum Geologists Bulletin* 51:1711-1786.

Williams, Howel and McBirney, A. R. 1969. *Volcanic History of Honduras.* University of California Publications in Geological Sciences, vol. 85. Berkeley:University of California Press.



Geochemistry at Honduran Geothermal Sites

by Fraser Goff, C. O. Grigsby, Lisa A. Shevenell, and J. Wilfred Gutierrez

In May 1985 a team from Los Alamos National Laboratory, the United States Geological Survey, and the Empresa Nacional de Energía Eléctrica (ENEE) carried out hydrogeochemical studies at six major hot-spring systems in the western half of Honduras. The locations of these systems are noted on the map in "Geology of Honduran Geothermal Sites." The team analyzed water samples for concentrations of major and trace elements, stable isotopes, and tritium, gas samples for concentrations of carbon dioxide, hydrogen sulfide, methane, and other gases, and rock samples for concentrations of carbon and oxygen isotopes. The results of the analyses were used to assess the suitability of the sites for geothermal development. The team also studied many cold springs throughout Honduras to obtain background information about the concentrations of deuterium, tritium, and oxygen-18 in Central American waters.

View east of the silica sinter terrace at San Ignacio, Honduras. Boiling springs, which are used for cooking, discharge all around the perimeter of the 100- by 150-meter terrace. (Photo by Fraser Goff.)

View north of the La Cueva area, Azacualpa, Honduras. Note the steam from boiling springs at the mouth of the cave, which is formed of old carbonate travertine undercut by the creek in the foreground. (Photo by Fraser Goff.)

Table 1

Concentrations of silica, arsenic, lithium, boron, bromine, and ammonium in surface waters at six Honduran hot-spring sites and the temperatures of those surface waters. High surface concentrations of these species may indicate high temperatures in the underground reservoirs. Also listed, for comparison, are the concentrations found in a fluid sample from the Valles Caldera geothermal site in New Mexico. This sample was collected at a depth of 1500 meters (at the entry to Baca well #13); the temperature of the fluid there, after being corrected for steam flash, is 278°C.

	Hot-Spring Temperature	Concentration (mg/l)					
Site	**(°C)**	**SiO_2**	**As**	**Li**	**B**	**Br**	**NH_4**
Azacualpa	115.4	211	0.07	0.94	1.59	<0.1	1.09
El Olivar	75.9	120	<0.05	1.38	8.02	0.3	10.00
Pavana	101.8	128	0.11	0.27	1.43	<0.1	0.17
Platanares	99.5	288	1.26	4.04	16.70	<0.1	10.40
Sambo Creek	99.5	133	<0.05	0.17	0.09	<0.1	0.12
San Ignacio	99.0	214	<0.05	1.44	3.81	<0.5	2.78
Valles Caldera		488	1.16	17.20	14.90	5.3	1.52

The six sites studied were Azacualpa, El Olivar, Pavana, Platanares, Sambo Creek, and San Ignacio. Geologic evidence indicates that the hot-spring systems in Honduras are not associated with recent silicic volcanism, as is the case, for example, at Yellowstone National Park in Wyoming and the Valles Caldera in New Mexico. Rather, the setting in Honduras resembles that of Nevada: water circulates deep into the earth, is heated conductively, and rises convectively along faults and fractures. In agreement with the geologic evidence, the surface waters were found not to be acid-sulfate in character, which is indicative of an origin in near-surface steam reservoirs, and to be instead neutral to alkaline-chloride in character, which is indicative of an origin in subsurface reservoirs. Boiling and/or superheated hot springs are present at all sites except El Olivar; the temperature of the springs there is less than 76°C. Several of the spring systems have deposited silica (SiO_2) as terraces or as gravel cements, a feature that usually indicates subsurface reservoir temperatures greater than 150°C.

The concentrations of certain chemical constituents in the surface waters at a geothermal site depend primarily on the subsurface reservoir temperature and the rock type and to a lesser extent on the amount of circulating water and the flow rate. Significant concentrations of silica, arsenic, lithium, boron, bromine, and ammonium, for example, usually indicate a high equilibrium temperature in the reservoir. Table 1 lists the concentrations of these constituents in typical water samples from the six Honduran sites and, for comparison, in a sample from a reservoir in the Valles Caldera of New Mexico, which is known to contain high-temperature fluid. We use the Valles Caldera for comparison because it is a classic geothermal system, well known among geologists, and its rock types are very similar to those found at the Honduran sites (primarily welded tuffs and ancient sedimentary rocks such as limestones, sandstones, and shales). Nevertheless, since the Valles

Caldera fluid is generated in a volcanic environment and, in addition, the sample for that reservoir came from a very-high-temperature well, only qualitative conclusions can be drawn from such a comparison. The data suggest that the Platanares site is the hottest of the six Honduran sites but is not as hot as the Valles Caldera reservoir.

A better way to assess equilibrium reservoir temperatures is to use chemical geothermometers. Table 2 lists, for the six Honduran sites and for the Valles Caldera, the subsurface reservoir temperatures estimated with two widely used geothermometers, quartz and sodium-potassium-calcium. The quartz geothermometer relates quartz (SiO_2) concentration to temperature through the laboratory-measured solubility curve of this mineral. The solubility of quartz rises steeply between 100 and 300°C. Since precipitation is quite sluggish with falling temperature, the silica concentrations found in the surface waters are good indicators of the subsurface reservoir temperature. The sodium-potassium-calcium geothermometer, an empirical relation between relative concentrations of these elements in surface water and reservoir temperatures, is based on data gathered from many high-temperature geothermal systems around the world. The Platanares site again comes out ahead, but the temperatures of two other sites, San Ignacio and Azacualpa, are greater than 180°C, the minimum temperature required for economical generation of electric power.

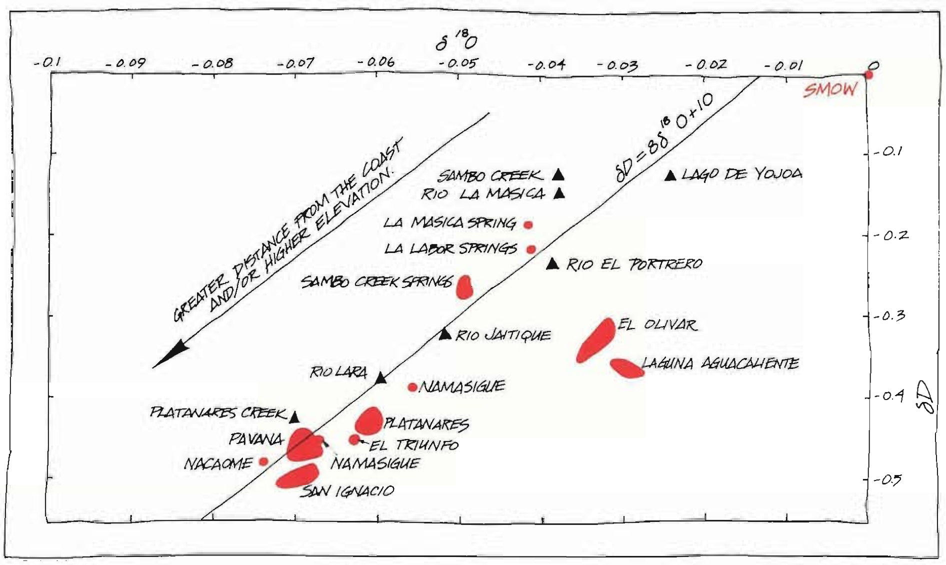
Our results from chemical geothermometry generally agree with those from gas geothermometry. One gas geothermometer uses the relative concentrations of carbon dioxide, hydrogen sulfide, methane, and hydrogen as an indicator of temperature. The relationship is empirical but has been supported by theoretical studies of equilibration among these gases at high temperature and by comparison with the gas chemistry of many explored geothermal fields.

Table 3 lists the minimum electric

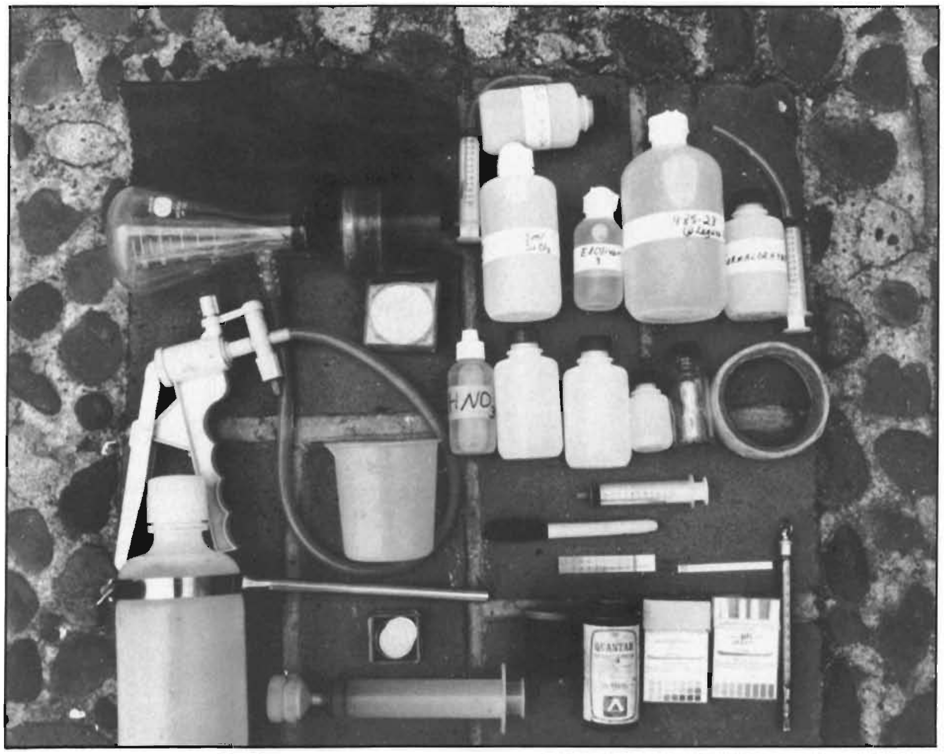
Table 2

Estimated temperatures of the underground reservoirs at six Honduran hot-spring sites. These estimates were obtained by using two chemical geothermometers, the quartz and the sodium-potassium-calcium geothermometers (see text). Also listed are similar estimates for the temperature of the underground reservoir at the Valles Caldera geothermal site in New Mexico. The observed temperature of the Valles Caldera fluid, measured at a depth of 1500 meters (at the entry to Baca well #13) and corrected for steam flash, is 278°C.

	Estimated Reservoir Temperature (°C)	
Site	**Quartz Geothermometer**	**Na-K-Ca Geothermometer**
Azacualpa	184	181
El Olivar	148	101
Pavana	151	138
Platanares	207	225
Sambo Creek	147	148
San Ignacio	185	208
Valles Caldera	249	282



Results of isotopic analyses of water samples from hot springs (red) and lakes, rivers, and cold streams (black) in Honduras. Shown is a plot of $\delta D = (^2H/^1H)_{sample} - (^2H/^1H)_{SMOW}$ versus $\delta^{18}O = (^{18}O/^{16}O)_{sample} - (^{18}O/^{16}O)_{SMOW}$, where ^2H, ^1H, ^{18}O, and ^{16}O are isotopic concentrations and SMOW stands for standard mean ocean water. Data points for all surface waters worldwide are found to fall near the line $\delta D = 8\delta^{18}O + 10$. The distance along that line between the data point for SMOW and the data point for a sample is indicative of the distance from the ocean and the elevation of the sample's origin. These isotopic analyses indicate that the Honduran geothermal reservoirs contain recycled rainwater from the local area around each site.



Some of the lightweight, compact equipment used to collect samples of water from hot springs. Six samples are collected from a spring, one each for anion, cation, tritium, deuterium and oxygen-18 (as water), carbon-13, and oxygen-18 (as sulfate) analysis. The samples for anion and cation analysis are filtered through a 0.45-micrometer filter; the sample for cation analysis is acidified with HNO_3 to a pH less than 2; the sample for carbon-13 analysis is treated with saturated $SrCl_2$ and concentrated NH_4OH to precipitate $SrCO_3$; and the sample for analysis of oxygen-18 as sulfate is treated with formaldehyde to preserve the sulfate. Conductivity, temperature, pH, and chloride are measured at the site. Gas sampling requires a different array of equipment.

Table 3

Estimates of thermal power and equivalent electrical power from surface discharges at six Honduran hot-spring sites. Also listed are values for the parameters involved in the estimates: the estimated surface discharge rates, the best available estimates for the temperatures of the underground reservoirs, and the ambient temperatures. The thermal power was approximated as the product of surface discharge rate and the difference between the heat content of the fluid at the temperature of the reservoir and at ambient temperature. An efficiency of 20 percent was assumed for the conversion of thermal power to electrical power, which is practical only if the reservoir temperature exceeds 180°C. The power potential of these sites may be much greater than these estimates of power from surface discharges.

Site	Estimated Surface Discharge (l/min)	Estimated Reservoir Temperature (°C)	Ambient Temperature (°C)	Thermal Power (MW)	Electrical Power (MW)
Platanares	3150	225	27	44.9	9
San Ignacio	1200	190	28	13.8	2.8
Azacualpa	1200	185	28	13.4	2.5
Pavana	1000	145	30	8.1	---
Sambo Creek	2000	150	30	16.9	---
El Olivar	200	120	30	1.3	---

power potential of these sites based on the estimated reservoir temperatures and the estimated surface discharge rates. Here too, Platanares looks very promising.

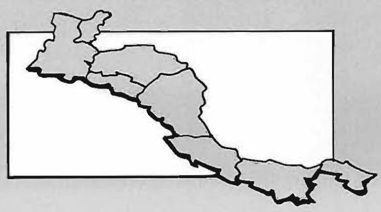
Ratios of deuterium to hydrogen-1 concentrations and oxygen-18 to oxygen-16 concentrations in a water sample provide information about the source of the water. Measured values of these isotopic ratios in samples of surface water from the Honduran geothermal sites indicate that recycled rainwater is feeding the reservoirs (see accompanying figure).

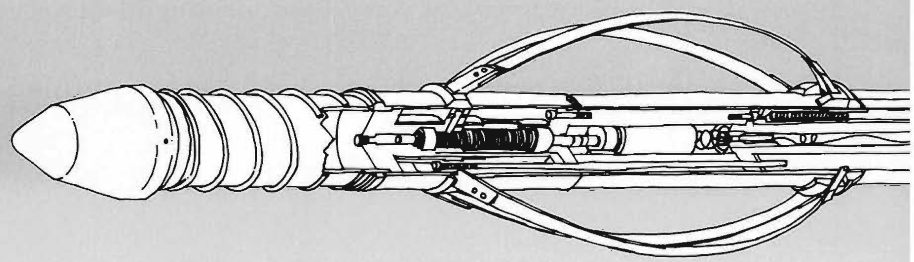
The tritium content of a geothermal fluid can be related to its age through equations describing the circulation of fluids in the geothermal system. The equations include the known input of tritium from the atmosphere as a function of time and location of the system. Analytical solution of the equations indicates that Honduran geothermal waters are between 34 and 7500 years old and are most likely several thousand years old. The better sites should therefore provide a stable, long-lasting source of geothermal power.

The ratio of carbon-13 to carbon-12 concentrations in a sample of surface water is an indicator of rock types through which the water flows. Measured values of this ratio in bicarbonate (HCO_3) from the Honduran hot springs indicate that the springs are flowing through sedimentary rocks and/or rocks containing hydrocarbons and other organic compounds.

By combining the information obtained from geochemical and geologic studies, the temperature and flow dynamics of a site can be evaluated before the more expensive step of drilling begins. ■

The geochemical work reported here was done by Dale Counce, Fraser Goff, Chuck Grigsby, Wilfred Gutierrez, Lisa Shevenell, and Pat Trujillo of Los Alamos National Laboratory, Alfred Truesdell and Cathy Janik of the U.S. Geological Survey (Menlo Park), and Rodrigo Paredes of ENEE.





High-Temperature Borehole Measurements at Miravalles, Costa Rica

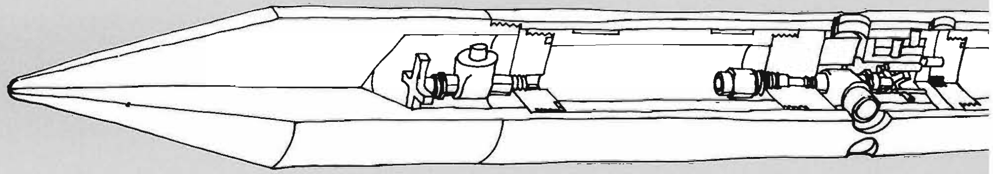
by Bert R. Dennis and Robert J. Hanold

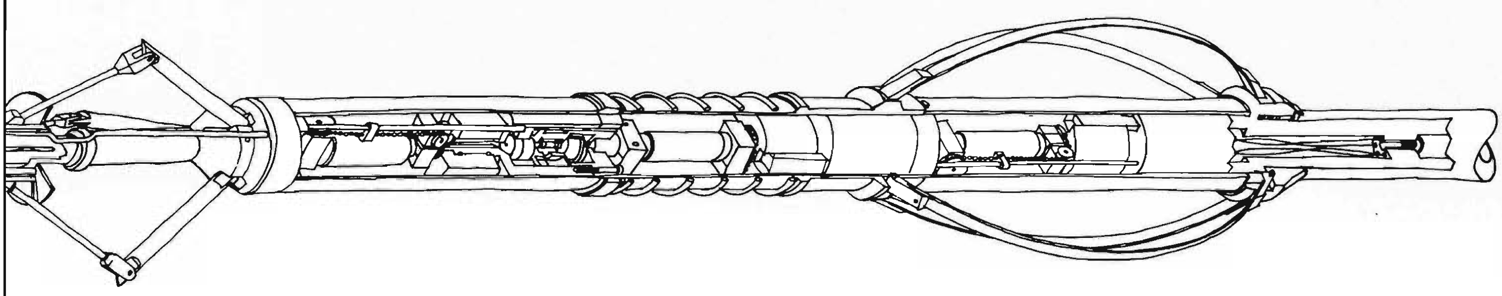
Costa Rica is developing its first geothermal power plant on the southern flank of the Miravalles Volcano in the Guanacaste Volcanic Range. If successful, this development will complement the vast hydroelectric resources of the country and help eliminate the need for fossil-fueled power plants. At present the import of petroleum contributes significantly to the trade imbalance of the country.

Well-logging equipment on site at the Miravalles geothermal field.

The development at Miravalles began about ten years ago with reconnaissance efforts, sponsored by the United Nations, that identified the slopes of Miravalles and Rincon de la Vieja volcanoes as potential sites for development of geothermal resources. The Power Planning Division of the Costa Rican Institute of Electricity (ICE) then began drilling deep production wells at Miravalles. The results were encouraging; the production wells have penetrated a 240°C-reservoir of geothermal brine at a depth of less than 2 kilometers, and the flow rates in the wells are very high (39 to 76 kilograms per second). So far so good. But when the initial wells were flow-tested, ICE engineers detected the presence of calcite ($CaCO_3$) deposits in the well bores. Since they lacked instruments to make measurements in the high-temperature environment downhole, they had no way to assess the scope of the problem.

Then, while attending a 1984 Los Alamos workshop for Central Americans on geothermal energy development, ICE engineers learned of the specialized instruments developed by the Laboratory's Earth Science Instrumentation Group to satisfy the diagnostic needs of the Los Alamos hot dry rock geothermal energy





▲Fig. 1. The three-arm caliper and contour tool developed at Los Alamos yields precision measurements of borehole dimensions and in situ casings. An electric motor extends or retracts the caliper arms on command from the surface logging rig.

program. These unique logging tools are capable of operating at temperatures up to 300°C and pressures up to 15,000 psi for durations between 8 and 30 hours. The Costa Ricans explained their problem, and with support from the U. S. Agency for International Development, these instruments were made available to ICE for downhole diagnostic measurements at the Miravalles wells.

Los Alamos engineers and technicians overhauled a surplus well-logging rig and equipped it with 3 kilometers of special cable and a cablehead assembly, designed at Los Alamos, for interfacing with the downhole tools. A computer-driven data-acquisition system was installed in the logging cab. After being tested in a local geothermal well, the unit was shipped to Costa Rica together with logging tools for measuring temperature and pressure as a function of depth, flow rate throughout the production layer, and the contour and average diameter of the well casing and for collecting samples of brine from the reservoir without loss of dissolved gases.

Two wells at the Miravalles field were logged. A single borehole instrument measured temperature, pressure, and flow rate. This new tool significantly increases the efficiency of measurements in a hot, high-pressure well because only a single entry and removal through the pressure lock is required. Figure 1 shows the tool used to measure the contour and average diameter of the well casing. Figure 2 shows the tool used to collect pressurized fluid samples at various locations in the well.

We are currently analyzing the logging data, and the brine samples are enroute to Los Alamos for chemical analysis. Bottom-hole temperatures in both wells approach 240°C, an excellent temperature for efficient generation of electricity.

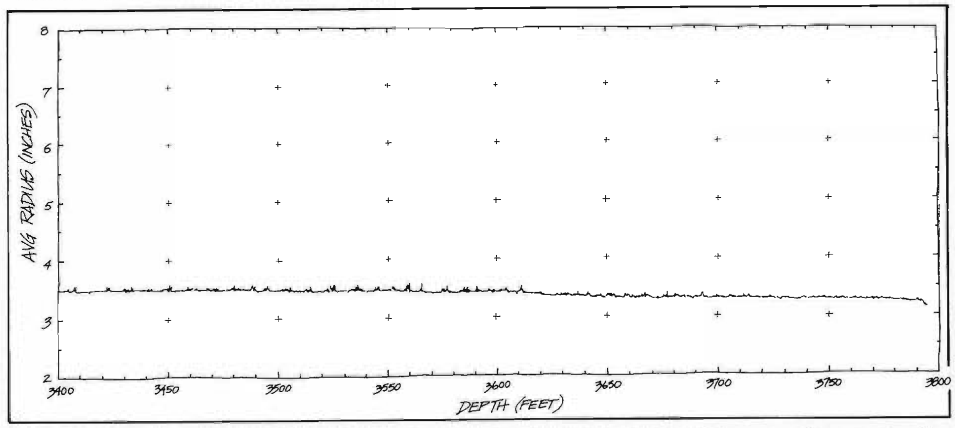


Fig. 3. Results of a caliper survey of a well at the Miravalles geothermal site. Accumulated scale deposits have caused a decrease in the inside radius of the slotted production liner from 3.5 inches at a depth of 3400 feet to 3.2 inches at 3800 feet. The slots in the production liner, which provide the passageway for the reservoir fluids into the production well, are evident in the caliper data at shallower depths but essentially disappear at depths below 3600 feet. These slots are apparently being plugged with calcite deposits.▲

The caliper surveys confirmed the suspicions of the ICE engineers by indicating considerable buildup of calcite in the lower sections of one of the wells (Fig. 3). Such deposits will ultimately reduce the flow rate from the well.

The initial logging experiences at Miravalles indicated the need for some modifications in the logging tools to improve their durability in the high-flow-rate wells. When these modifications are completed, the equipment will be returned to Costa Rica and used to log additional production wells. Data from these logging surveys could lead to an improved drilling strategy for the rest of the production wells. ■

Participants in the logging efforts included David Anderson, Gloria Bennett, Lynn Brewer, Pete Chavez, Benny Garcia, Ray Jermance, Jerome Kolar, Richard Maestas, and Evon Stephani of Los Alamos National Laboratory and Rodrigo Corrales and Manuel Corrales of ICE.

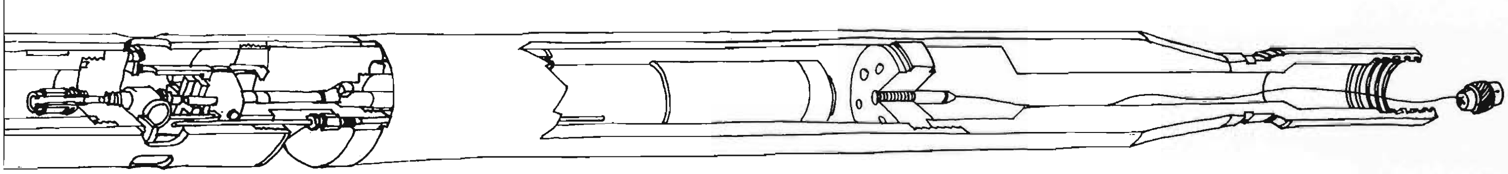
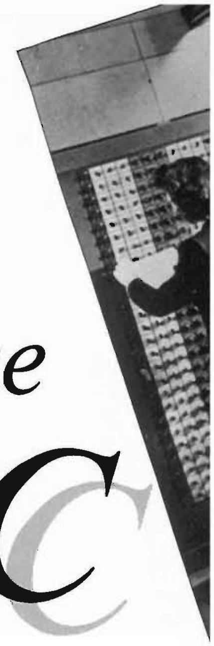


Fig. 2. The fluid sampler tool has two chambers for collecting samples. The motor that opens and closes the sample chambers is activated on command from the surface logging rig.▲



Metropolis, Monte Carlo, and the MANIAC

by Herbert L. Anderson

In 1942 Nick Metropolis was working with Edward Teller on the reactor project at the University of Chicago when J. Robert Oppenheimer invited the young physicist to continue his collaboration with Teller, but at Los Alamos. There Metropolis joined the Manhattan Project as a member of the Theoretical Division, having been encouraged by Teller to move from experimental to theoretical physics. His first assignment was to develop equations of state for materials at high temperatures, pressures, and densities.

Over the years Metropolis turned increasingly to mathematics and computer design, and by 1948 he was leader of a Los Alamos team that designed and built the MANIAC, one of the first electronic digital computers. Many of the country's foremost scientists were eager to try their experiments on the wonderful new machine and came to the Laboratory to work with Metropolis. A few years later, together with Teller, John von Neumann, Stanislaw Ulam, and Robert Richtmyer, Metropolis developed techniques and algorithms for using the Monte Carlo method (so named by Metropolis) on the new computers.

The Monte Carlo method is an application of the laws of probability and statistics to the natural sciences. The essence of the method is to use various distributions of random numbers, each distribution reflecting a particular process in a sequence of processes such as the diffusion of neutrons in various materials, to calculate samples that approximate the real diffusion history. Statistical sampling had been known for some time, but without computers the process of making the calculations was so laborious that the method was seldom used unless the need was compelling. The computer made the approach extremely useful for many physics problems.

Metropolis was also involved in the development of an importance-sampling scheme, called the Metropolis algorithm, that improves the effectiveness of the Monte Carlo method. In the past twenty years his work has included nonlinear problems and combinatorial theory as well as Monte Carlo calculations. He was named a Senior Fellow of the Laboratory in 1980.

In September 1985 more than one hundred researchers from around the world met in Los Alamos for a four-day conference, in honor of Nick Metropolis, on the frontiers of quantum Monte Carlo. One of the speakers was Herb Anderson, from the Laboratory's Physics Division. Anderson's presentation was a fascinating reminiscence about the first modern calculating machines and the scientists who used them, about the intellectual ferment in the physics community that began early in this century and still continues, and about Nick Metropolis and the MANIAC. This article is adapted from Anderson's presentation.


Conversion of input to hexadecimals
3. Punching input

This story is about the MANIAC and about Nick Metropolis, who conceived the MANIAC, built it, and saw how to use it for a wide variety of problems. In the period just after World War II, other computers were being built, many of them, like the MANIAC, modeled on the von Neumann principle, the principle of the stored program. Von Neumann organized a group at the Institute for Advanced Study and started building a computer based on that principle. Other institutions got into the act, too, because they all realized the importance of building computers. There was one at Argonne called the AVIDAC and one at Oak Ridge called the ORACLE. Then there were the SEAC at the National Bureau of Standards and the ILLIAC at the University of Illinois. But the MANIAC at Los Alamos was special.

You see, the circumstances in postwar Los Alamos were special. The war had brought together there an exceptional group of scientists, with whom Nick had established a close relationship. John von Neumann, Enrico Fermi, Hans Bethe, Edward Teller, Stan Ulam, Dick Feynman, George Gamow, Tony Turkevich, and Robert Richtmyer, among others, were drawn to Nick because they enjoyed working with him. When you went to Nick with a problem, he took a deep interest in it and worked hard on it with you, invariably making some essential contribution. It was a treat to work with Nick, and, if you understand that, you will understand a lot about how this story evolved.

Those who returned to Los Alamos after the war were drawn irresistibly to Nick and his MANIAC, to what this wonderful electronic computer could do for them. They went away enriched by their new experience and rewarded by new scientific results in their fields of interest. This is a happy tale of how one of the first of the modern computers got its start and what it was able to do in its early years.

Fermi's Brunsviga

Let's go back about forty-five years to the beginning of World War II. Nick was twenty-five years old. In those days we had no computers as we now know them. We used slide rules and adding machines—hand-operated machines. Machines with electric motors were the exception rather than the rule.

I remember particularly the hand-operated machine that Fermi used with remarkable deftness. It was made by Brunsviga, a German firm in the town of the same name, famous as the place where Gauss was born. This machine had a crank that you rotated by hand. To multiply, for example, you set the machine and then rotated the crank the number of times called for by each digit of the multiplier, shifting the carriage for each successive digit and turning the crank again. Fermi had one of those machines when he was working in Rome and brought it along with him when he came to Columbia University in 1939. He was using it when I started working with him on the chain reaction, soon after his arrival. Whenever I used my slide rule to make a calculation, he started cranking his machine. By the time I announced my result, he was waiting—and grinning. He could beat me every time. But that situation changed when I got myself a Marchant. When it became clear that he couldn't even keep up with me, let alone beat me, he gave up the Brunsviga and got a Marchant of his own. Fermi could never resist the opportunity to calculate faster.

It seemed to me that Fermi was always calculating something. It was Fermi's view that Nature revealed itself through the experiments you devised to test it. You can construct a theory to explain what is going on, but unless the numbers come out right, you can't be sure the theory is right. So you have to do a lot of calculations.

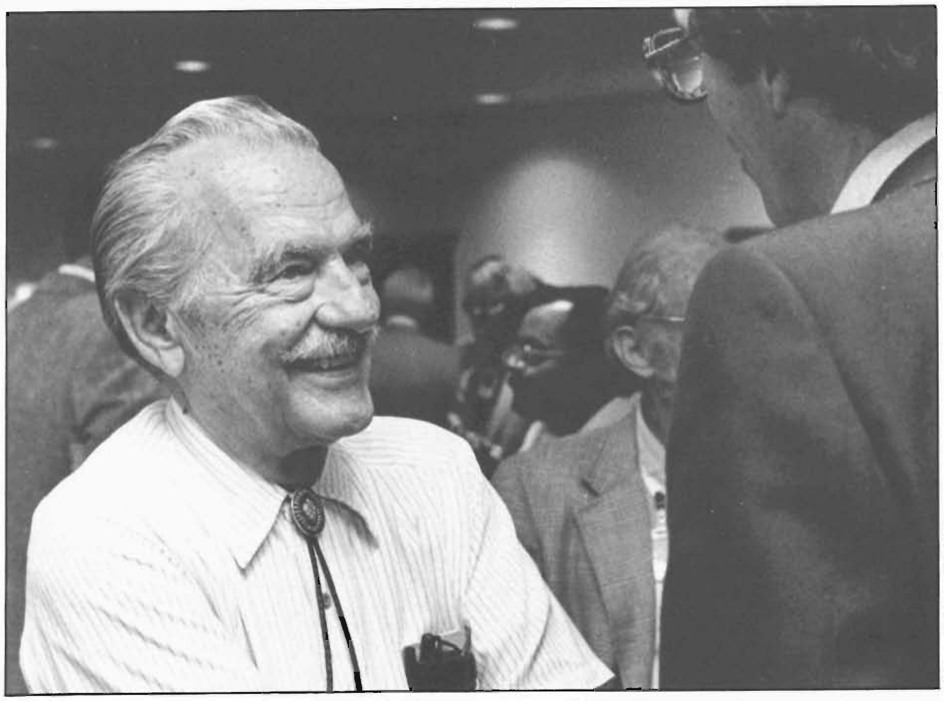
Fermi Monte Carlo

You might be interested to know that Fermi was one of the first to use the Monte Carlo method—in a rather simple form and hand-calculated—long before it had a name. I don't know whether he was the very first, but the story comes from Emilio Segrè, who told me that Fermi used that statistical sampling technique as early as 1934, when he was working on neutron diffusion in Rome.

In 1933 Frédéric Joliot and Iréne Curie had discovered the radioactivity induced in light elements by bombardment with alpha particles. The neutron had been discovered just one year before. These two facts gave Fermi the idea that neutrons, having no charge at all, would be much more effective than alpha particles in producing nuclear transformations. They would not be repelled by the Coulomb barrier and could therefore penetrate the nuclei of all atoms, whatever their charge, whereas the alpha particles could only get into the nuclei of light elements.

It was an exciting idea. He got a radon-beryllium neutron source and began a series of experiments with some of his young and enthusiastic collaborators: Amaldi, Segrè, Pontecorvo, d'Augostino, and Rasetti. Early in the course of their work, they found they were getting some very peculiar effects. The radioactivity they obtained depended a whole lot on where the irradiation was carried out. In particular, the activity induced in silver was much greater when they did the irradiation on a wooden table than when they did it on a table with a marble top. That was a great puzzle. They couldn't explain it. Then Fermi began to tell them that they didn't know how to experiment very well and that they didn't really do things properly. Of course this didn't make them very happy. To clear up the puzzle, Fermi decided to try filtering the neutrons through various substances. His first idea was to use lead, but then at the last minute, for no apparent reason, he substituted paraffin instead. The increase in the activity of the silver was phenomenal. Everyone went home mystified.

Now, one of Fermi's characteristics was that he liked to come into the lab early in the morning and surprise his colleagues



Nick Metropolis enjoying a break in the quantum Monte Carlo conference, September 1985.

with the answer to whatever problem they had been worrying about the night before. Unlike—or like—the early bird who catches the worm, Fermi had an affliction that helped him do this. He had insomnia, and he always got up at four in the morning. Now what do you do if you are wide awake at four in the morning? Well, in Fermi's case he either did theory or he did calculations. For making quick calculations he had a whole bag of tricks, and the hand-calculated Monte Carlo method was one of them.

On the morning following this great puzzlement in the lab he got up at four as usual, and in thinking about the problem he decided he knew what might be happening. Maybe the neutrons were being slowed down as they went through various substances, and if they were, then hydrogen nuclei would be especially effective. And with a greater slowing, you could expect a higher level of induced activity in the silver. Well, he came into the lab and made this pronouncement, and everyone was struck by the simplicity of it all and the theory turned out to be quite plausible.

Then Fermi, in working out the detailed theory, used Monte Carlo calculations to give him a physical insight and to help him choose a suitable functional form, Gaussian, exponential, or other, for representing the slowing down process.

The slowing down turned out to be a major discovery. Slow neutrons have very large cross sections for nuclear reactions, and with them Fermi produced a large number of new radioactive isotopes. This work won him the Nobel Prize in 1938. It led directly to the chain reaction in Chicago in 1942 and to the establishment of Los Alamos in 1943.

Fermi never wrote up his use of the Monte Carlo method, but he told the story to Emilio Segrè many years later when computers had made statistical sampling practical and the technique was coming into wide use. Segrè mentions it in his introduction to the neutron papers in *The Collected Works of Enrico Fermi.*

The Marchant Repairman

Now let's get back to Nick Metropolis. It is 1944, and the scene is Los Alamos. What is Nick doing? He is busily repairing Marchant calculators. And how did he get into that business? Well it happened in the following way. When Los Alamos was set up in 1943 there were no calculators. There was, however, an obvious and urgent need to carry out a lot of calculations, and the Lab went out and bought a whole lot of calculators—Marchants and Fridens, which were the best mechanical calculators at the time—and set up a hand-computing facility. The machines were heavily used and soon began to show signs of wear and tear. Too many were out for repair, and it took too long to send them to the manufacturer to be fixed. So Nick, together with Dick Feynman, set up a little repair shop. They took the machines apart and traced the mechanical linkages to find the sources of jams and slippage, and of course they learned how the machines worked. Pretty soon they could identify the difficulty rapidly, and the machines sent to their shop were quickly repaired and returned to service.

When the administrators came across this curious extracurricular activity, they regarded it as a problem. They issued some sharp criticism and stopped the repair service. But not for long. The demand for working machines was so great that the administrators decided they had better not interfere, and the service was soon reinstated.

In the fall of 1943 it became apparent that large computational problems were straining the capacity of the hand-calculators. That's where Dana Mitchell comes into the picture. I mention his name with great fondness because he was the man who got me into physics. He had come to Los Alamos to help with procurement, and he was familiar with the IBM punched-card machine. Soon a whole set of those machines arrived at the Lab. Metropolis and Feynman immediately decided to see whether these punched-card machines were really faster than their team of Marchant hand-calculators. They set up a test in which the two groups would calculate the same problem. For the first two days the two teams were neck and neck—the hand-calculators were very good. But it turned out that they tired and couldn't keep up their fast pace. The punched-card machines didn't tire, and in the next day or two they forged ahead. Finally everyone had to concede that the new system was an improvement.

As the atomic-bomb project entered its final phases in late 1944, the pressure for computation increased sharply. Nick became more and more involved in punched-card operations with Dick Feynman, who was put in charge. Their work continued at a frantic pace during 1945 until the Japanese surrendered on August 15, after which they began to relax a bit.

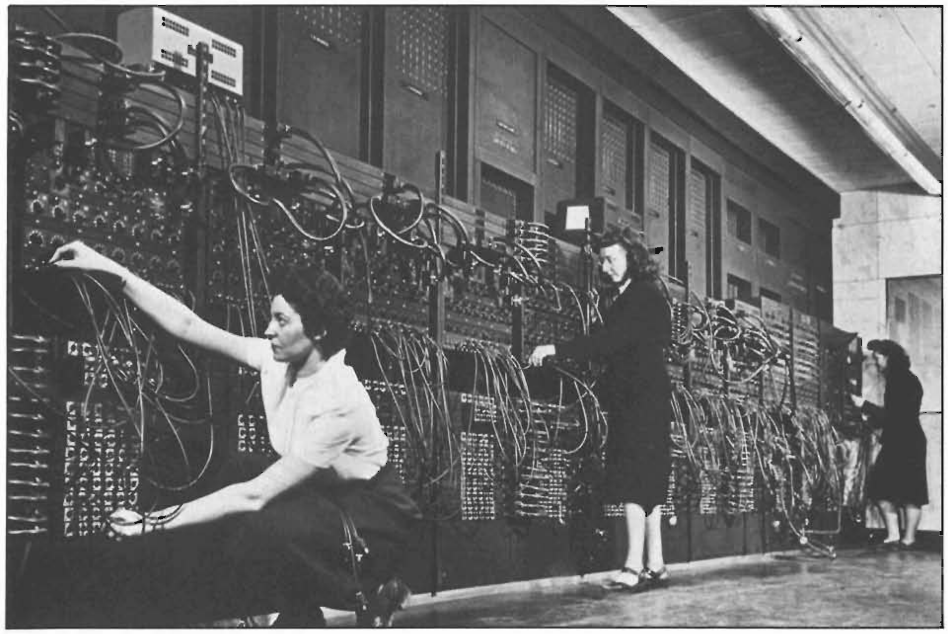


Fig. 1. Programming panels and cables of the ENIAC.

The ENIAC

The great step forward in computing was the introduction of electronics, and now the ENIAC entered the scene. The ENIAC was the first electronic, digital, general-purpose, scientific computer. It was designed and built for the Aberdeen Proving Grounds by a group of engineers under the direction of Pres Eckert and John Mauchly. It had 18,000 vacuum tubes and computed 1000 times as fast as its closest electromechanical competitor. The machine was built during wartime on the promise that it would calculate ballistic trajectories at least ten times faster than the mechanical differential analyzers then in use. As things frequently turn out, the machine was now working as promised—but the war was over, and suddenly no one cared that much about calculating ballistic trajectories.

The connecting link between the ENIAC and Los Alamos was Johnny von Neumann, who was a consultant both at the Aberdeen Proving Grounds and at the Lab. He was tremendously excited about the ENIAC. He took a deep interest in its design and thought a good deal about what could be done with it. When he came to Los Alamos early in 1945, he told Nick, Edward Teller, and Stan Frankel about what the Eckert and Mauchly team was doing. They were enchanted. You know how things sometimes work out if you are in the right place at the right time—and prepared for the opportunity. Well, von Neumann suggested that perhaps Los Alamos had a problem that could be worked out on the ENIAC and invited Nick and Stan to try the new machine. It was a great opportunity, and they seized it eagerly. And so it turned out that the first serious problem the ENIAC solved was the one Metropolis and Frankel put to it regarding the "super," the thermonuclear bomb.

By this time the idea of the stored program had already been conceived, principally by John von Neumann and his collaborators on the ENIAC. They had already begun to design the EDVAC, a computer that would have a stored program. The ENIAC was programmed by connecting cables and wires and setting switches on a huge plugboard that was distributed over the entire machine. Figure 1 is a photograph showing young women programming the ENIAC by interconnecting the electron tube registers with cables inserted in plugboards. It was an awkward and tedious way to tell the machine what to calculate and what to do with the results. When von Neumann was in Los Alamos, in about 1947, he described a suggestion made by Richard Clippinger of the Ballistic Research Laboratory on how the ENIAC might be converted to a limited stored-program mode. The idea was to rearrange the so-called function tables, normally used to store 300 twelve-decimal-digit numbers set by manual switches, to store up to 1800 two-decimal-digit numbers, each pair of numbers corresponding to an instruction. A particular problem would correspond to a sequence of such instructions. This sequence would be set on the function tables. A background control would interrogate these instructions sequentially, including so-called loops of instructions. Changing from one problem to another would be achieved by resetting the switches of the function tables to correspond to the new sequence of instructions. Figure 2 shows the function tables of the ENIAC.

This suggestion made a deep impression on Nick, but there was a missing element—the background control. On a visit to the ENIAC in early 1948, Nick learned that a new panel had been constructed to augment one of the logical operations. It was a one-input, one-hundred-output matrix, and it occurred to Nick that this matrix could be used instead to interpret the instruction pairs in the control mode proposed by Clippinger. Such a panel would greatly simplify the implementation of a background control. He told von Neumann about it and was encouraged to go ahead and try—and so he did. The scheme was implemented on the ENIAC forthwith, and Nick's set of problems—the first computerized Monte Carlo calculations—were run in the new mode.

The MANIAC

After the war Nick joined the faculty of the University of Chicago, to help set up a major computing facility. When that didn't materialize as quickly as he had hoped, he began to think of other possibilities, and about that time he got a call from Carson Mark, head of the Theoretical Division at Los Alamos, suggesting that he set up a computing facility here. Nick was ready, willing, and able. The moral of my story is that fortune favors the prepared mind. Nick was right there, well prepared to do just what he was asked to do, and that's how the MANIAC was born.

The Mathematical and Numerical Integrator and Computer—the MANIAC—was designed according to von Neumann's principles, which had been set forth in a remarkable publication by Arthur Burks, Herman Goldstine, and Johnny von Neumann. The MANIAC borrowed heavily from the IAS, the computer being built at the Institute for Advanced Study under von Neumann's direction. But because the MANIAC came later, Nick was able to avoid many pitfalls that delayed the IAS.

As I have mentioned, many computers were built in this outburst of activity after the end of the war. The ENIAC had started a revolution that continues to this day, with no end yet in sight. But the unusual success of the MANIAC was due primarily to the personality and motivation of Nick Metropolis and to the group of highly capable engineers and programmers he assembled at Los Alamos to help him build the machine and make it run. The original engineers were Dick Merwin, Howard Parsons, Jim Richardson, Bud Demuth, Walter Orvedahl, and Ed Klein. The importance of programming aids was recognized early on. About 1953 John Jackson led a study of assembly languages, and an assembler was produced. Mark Wells and others launched the development of the MADCAP, a high-level programming language and compiler. This was a critical development because it provided a convenient way to communicate with the MANIAC.

Fermi and Metropolis

Enrico Fermi had been at Los Alamos during the war and liked it so much that he claimed he would not have left if only the Lab were a university. Since it wasn't a university, he made the best possible compromise by accepting a position at the University of Chicago and spending his summers at Los Alamos.

When he came to Los Alamos in the summer of 1952, the MANIAC was up and running, and it would have been very hard to keep Enrico from that machine. He though the MANIAC was just wonderful. He could hardly wait to get his hands on it. I've told you how he loved to calculate, the faster the better, and here was his good friend Nick Metropolis with the fastest machine in the world, offering to introduce him to its mysteries and let him run it himself.

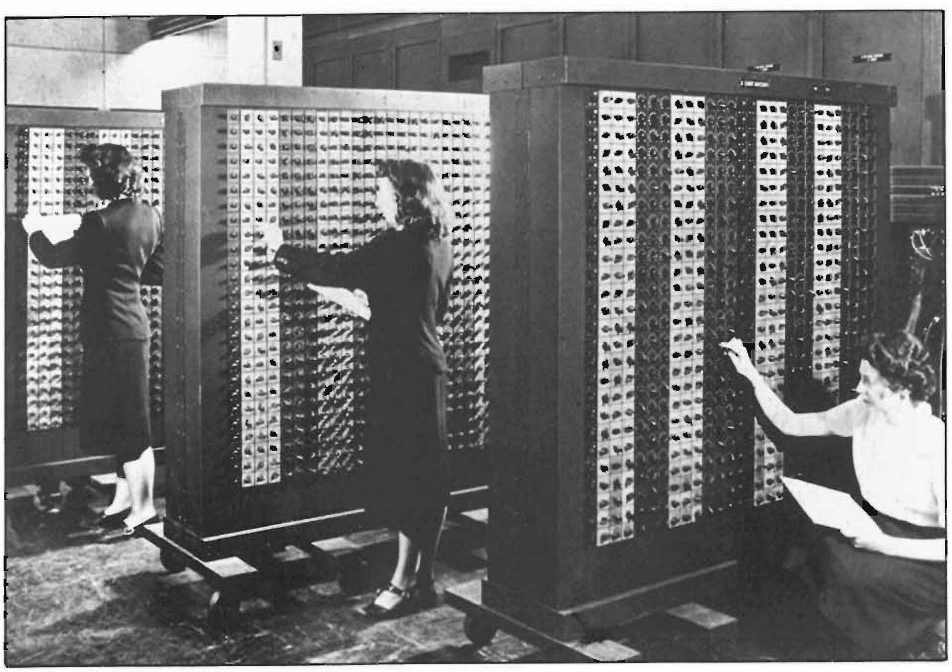


Fig. 2. Function tables of the ENIAC.

Nick must have been pleased by the praise and interest shown by so many of the illustrious scientists he had worked with in wartime Los Alamos, but the supreme accolade came from Enrico Fermi. When you build something, what could be more satisfying than to have one of the world's greatest physicists tell you not only that it's a great machine but that he wants to use it. Moreover, Fermi had a problem that was ideally suited to the machine. He wanted to analyze the pion-proton scattering experiments he had been carrying on in Chicago with his collaborators at the new 450-MeV synchrocyclotron.

My connection with this story is analogous to Nick's except that in this case I had built the cyclotron. I also helped build the apparatus and carry out the measurements. The other collaborators were Darragh Nagle and Earl Long, on the faculty, and my graduate students, Ronald Martin, Gaurang Yodh, and Maurice Glicksman.

The results of our pion-proton scattering experiments had been as striking as they were surprising. The large scattering cross section at a specific energy indicated the presence of a resonance.It was a major discovery, an excited state of the proton. We had uncovered a new particle now known as the delta, and it attracted the attention of the entire high-energy physics community. In order to extract the appropriate quantum numbers of the delta, Fermi wanted to do what he called phase-shift analysis, which tells you which quantum states have the biggest scattering amplitudes and which have the smallest.

I should mention that Fermi had a knack of coming up with problems whose computation matched the means available. Some years before, when the punched-card machines were the principal means for computing, Fermi posed the problem of calculating a table of atomic masses using a semiempirical mass formula he had devised on the von Weizsäcker model. Nick organized the calculation and the preparation of the tables.

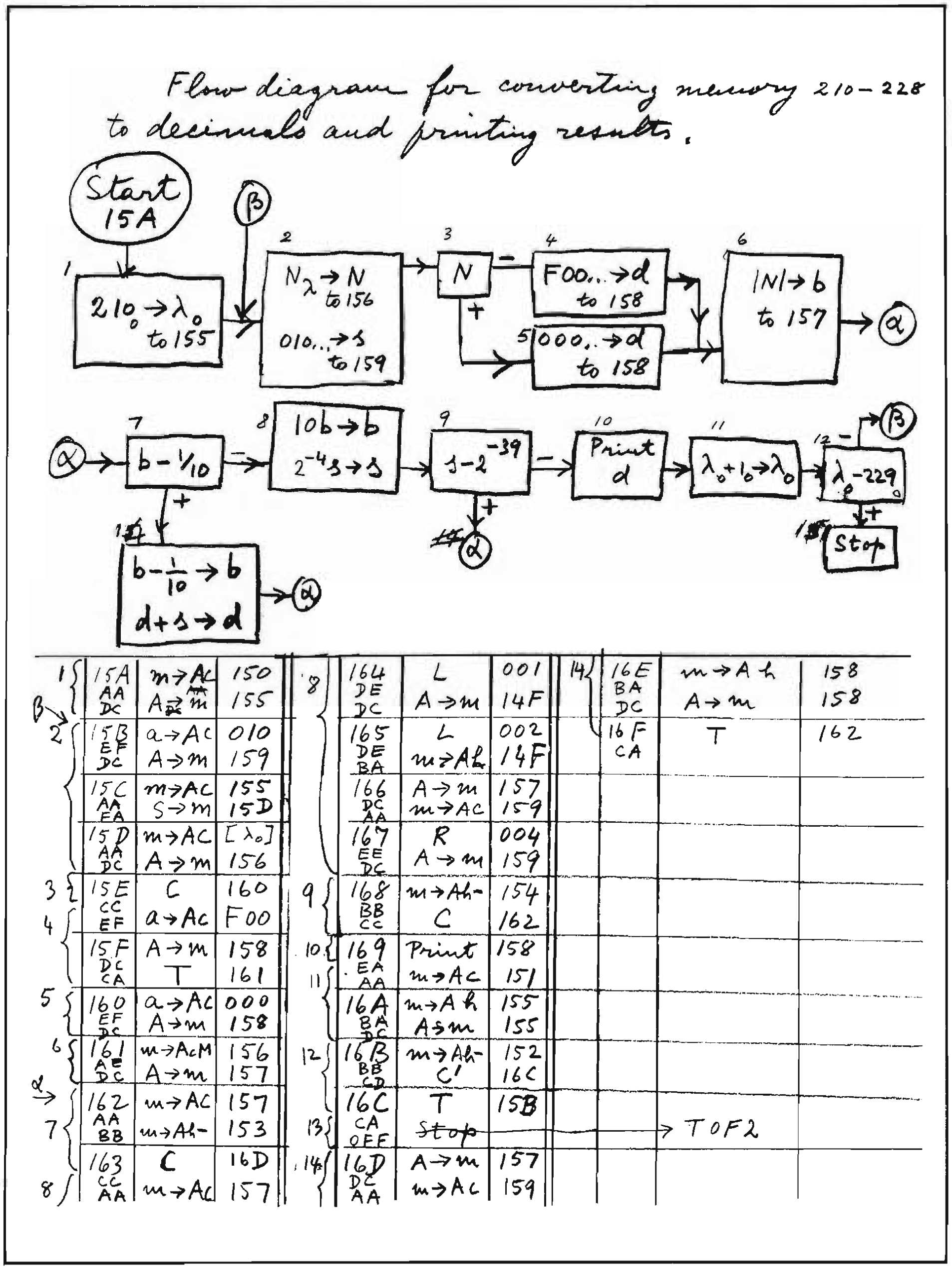


Fig. 3. A subprogram written by Fermi for converting data in memory from hexadecimal to decimal form and printing the results.

The tables turned out to be very useful and were widely used. I still have my copy.

So anyway, Fermi came to Nick with his phase-shift problem. As always, Nick was extremely helpful, and they carried out the work. I learned all about it when Fermi returned to Chicago in the fall of 1952, so steamed up about computers and the MANIAC that he announced he would give a series of lectures on digital computing. We were treated to a magnificent course—Fermi at his best. We learned for the first time about binary and and hexadecimal arithmetic, Boolean algebra, and linear programming. With this kind of introduction, we were easy converts to the cause of computers in science, and we even began to go out to Argonne, where by that time the AVIDAC was running, to learn how to program and run that machine. The gospel according to Nick Metropolis was taking effect.

There's an amusing story about Edward Teller that fits in here. Remember that Nick had been a member of Teller's group when they were working on the "super." Now, Teller was not one to let Fermi leave him behind. Anything Fermi could do, he could do too. So Teller also became a student of Nick's and learned how to program the MANIAC. When he came back to Chicago—he was on the faculty then—not to be outdone by Fermi he announced that he would give a colloquium on the subject of computers. But when the colloquium notice appeared, it didn't convey exactly the impression he had intended. It read,

Edward Teller

The MANIAC

To show how closely Fermi interacted with the MANIAC, I want you to see some of his programming efforts, done in his own hand. Remember, these were the days before FORTRAN. Programming was done at the lowest level, in machine language. Figure 3 is a subprogram Fermi wrote to convert the data in memory into decimals and to print the results. Figure 4 is a block diagram of the program for calculating the phase shifts by finding a minimum chi-squared in a fit to the data. And Figure 5 is a printout of the program from the MANIAC. Note the use of hexadecimal numbers. The comments are written in Fermi's hand.

Phase-Shift Analysis

In this period, 1953 and 1954, phase-shift analysis was such a hot subject that it occupied center stage in the elementary particle physics community. At the Rochester Conferences held in those and subsequent years, you could talk about alpha three three and alpha three one, and everyone understood that these were the phase shifts of the pion-proton scattering. The physics was important—the delta was a new particle.

In working with the phase-shift analysis program, we encountered, for the first time, solutions in hyperspace, many-functional space. You had to get used to the fact that this kind of space has its own problems of minimization, that you could easily fall into the wrong minimum and end up with wrong solutions. The virtuosity of the computer almost made us lose sight of the discovery of the proton

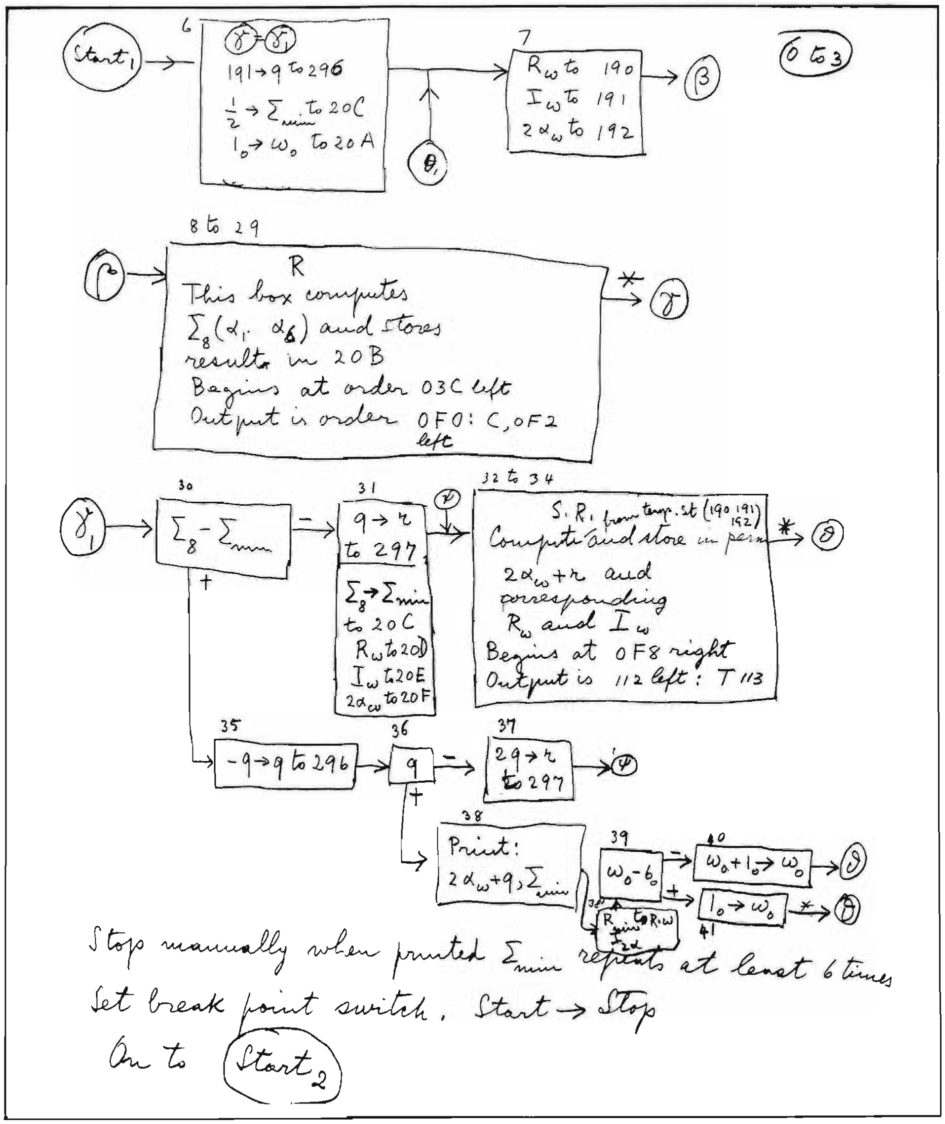
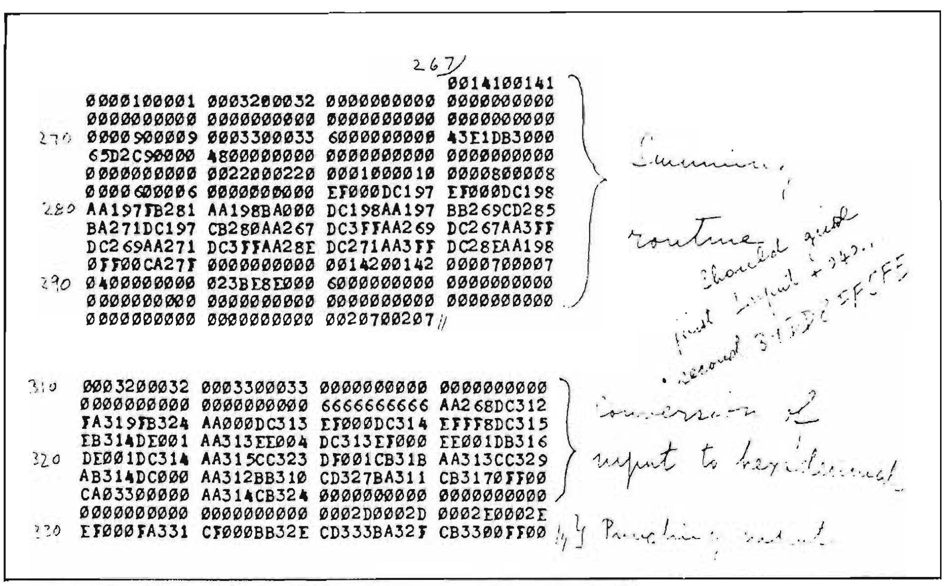


Fig. 4. A subprogram written by Fermi for calculating phase shifts by finding a minimum chi-squared in a fit to the data.



```
                                           0014100141
0000100001  0003200032  0000000000  0000000000
0000000000  0000000000  0000000000  0000000000
0000900009  0003300033  6000000000  43E1DB3000
65D2C90000  4800000000  0000000000  0000000000
0000000000  0022000220  0001000010  0000800008
0000600006  0000000000  EF000DC197  EF000DC198
AA197FB281  AA198BA000  DC198AA197  BB269CD285
BA271DC197  CB280AA267  DC3FFAA269  DC267AA3FF
DC269AA271  DC3FFAA28E  DC271AA3FF  DC28EAA198
0FF00CA27F  0000000000  0014200142  0000700007
0400000000  023BE8E000  6000000000  0000000000
0000000000  0000000000  0000000000  0000000000
0000000000  0000000000  0020700207

0003200032  0003300033  0000000000  0000000000
0000000000  0000000000  6666666666  AA268DC312
FA319FB324  AA000DC313  EF000DC314  EFFF8DC315
EB314DE001  AA313EE004  DC313EF000  EE001DB316
DE001DC314  AA315CC323  DF001CB31B  AA313CC329
AB314DC000  AA312BB310  CD327BA311  CB3170FF00
CA03300000  AA314CB324  0000000000  0000000000
0000000000  0000000000  0002D0002D  0002E0002E
EF000FA331  CF000BB32E  CD333BA32F  CB3300FF00
```

Fig. 5. A portion of the printout of the program containing the subprograms described in Figs. 3 and 4. The program is written in machine language in hexadecimal numbers.

resonance because the computer would find solutions that we didn't expect were there. We would put a program into the machine fully expecting the quantum state of the resonance to emerge. But no. A computer doesn't pay any attention to what Nature would like the solution to be; it has its own way of finding solutions. And all of a sudden it began to find solutions, many of them—six of them—that had nothing to do with the resonance. It found many sets of phase shifts that gave good fits to the data, leaving open the question whether the resonant solution was the correct one. The resonant solution was appealing in that it accounted for all of the unusual features found in the experiments, but the nonresonant solutions were not easy to rule out.

I won't say that this confused Fermi—that would be a little too harsh—but the result of all this was that he couldn't claim with certainty that we had discovered a resonance in our experiments. As long as the computer was turning out solutions, good fits of the data, good chi-squareds, that were nonresonant solutions, he was always forced to conclude that the result was ambiguous.

Now Hans Bethe, who had been head of the Theoretical Division at Los Alamos during the war, decided to get into the act. He made himself a great expert in phase-shift analysis. He wasn't satisfied with the way Fermi was handling the problem, and he went in with the idea that he wasn't going to be that naive and that he could do better by including additional physics arguments. So he enlisted Nick's aid and, with Fred deHoffmann to help him, mounted a second program in phase-shift analysis. So here we had Nick on both sides of the competition, an odd situation that only someone like Nick could handle. In matters of science, Nick had no favorites. In the end, I think, the problem was handled best by two of my graduate students, Ronald Martin and Maurice Glicksman, working separately. They didn't have a fancy computer and therefore had to use much simpler approaches to the problem. By using graphical methods with simplified but plausible assumptions, they came up with the resonant solution that turned out to be the correct one. It was a lesson in the use of computers that should be a caution to us all.

Scientific Triumphs

The nice thing about having the first computing machine is that almost anything you do on it is new and important. In Fig. 6 I have listed ten of the many scientific uses of the MANIAC. I chose these to emphasize the distinction of the men who came to work with Nick and the variety and importance of what they were able to accomplish. These projects are also examples of how the computer opened new possibilities for scientific investigation, sometimes with surprising results.

Nonlinear Oscillators. I have already discussed the first two calculations listed in Fig. 6. The third, the Fermi, Pasta, and Ulam study, turned out to be extraordinarily important. In the summer of 1953 Fermi raised the question of the nature of the approach to equilibrium of a vibrating, nonlinear string, initially in a single oscillatory mode. He thought it would be fun to use the MANIAC for this experiment, so he and Stan Ulam and John Pasta set up a test problem and began to run it. As they expected, the computations showed that the initial vibrational energy gradually transferred into neighboring modes and eventually achieved equilibrium, the time taken being the so-called relaxation time. Everybody was really quite happy with the result.

But the completely unexpected happened one day when they were computing a typical problem. While the machine was grinding away at this problem, they became engrossed in a heated discussion and let the computer go beyond its usual turn-off point. When they finally got around to looking at it, they found that the vibrational energy had returned to within a few percent of its initial state. Well, that was

Pion-proton phase-shift analysis (Fermi, Metropolis; 1952)

Phase-shift analysis (Bethe, deHoffman, Metropolis; 1954)

Nonlinear coupled oscillators (Fermi, Pasta, Ulam; 1953)

Genetic code (Gamow, Metropolis; 1954)

Equation of state: Importance sampling (Metropolis, Teller; 1953)

Two-dimensional hydrodynamics (Metropolis, von Neumann; 1954)

Universalities of iterative functions (Metropolis, Stein, Stein; 1973)

Nuclear cascades using Monte Carlo (Metropolis, Turkevich; 1954)

Anti-clerical chess (Wells; 1956)

The lucky numbers (Metropolis, Ulam; 1956)

Fig. 6. Scientific triumphs achieved with the MANIAC. Nick Metropolis was a coauthor of the publications resulting from these studies except for the ones on nonlinear coupled oscillators and anti-clerical chess.

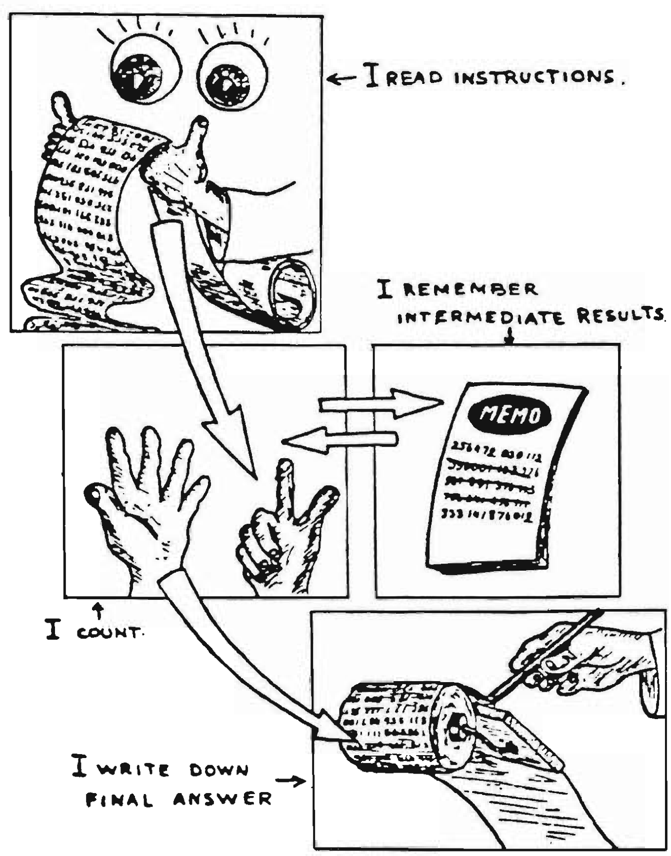


Fig. 7.

(For simplicity numbers in Fig. 7 are given in the decimal and not in the binary system.)

Fig. 7. Mr. Tomplins lears how the MANIAC works in an illustration from *Mr. Tompkins Learns the Facts of Life* by George Gamow. (Copyright 1953 by Cambridge University Press; reprinted with permission.)

Fig. 8. Paul Stein (left), Nick Metropolis, the MANIAC, and the 6-by-6 "anti-clerical" chess board.

such a tremendous surprise that at first they thought the machine had gone awry. They ran the problem again, and lo and behold, given enough time, the amplitudes all went back to the initial state, and then more new and surprising things happened. The rest is history—nonlinear systems were shown to have many fascinating aspects. The ideas of soliton theory emerged, and the subsequent outpouring of papers became a minor industry. Today this classic work is known as the FPU (Fermi-Pasta-Ulam) problem.

The Genetic Code and Mr. Tompkins. The Metropolis circle of famous scientists included George Gamow. At the time Gamow was introduced to the MANIAC, his interest had turned to biology, and so he got Nick to help him work on some problems of the genetic code. He was one of the first to have the idea that the genetic code was vested in four nucleic acids, which somehow were able to code for the twenty amino acids from which proteins are made. So some of these early studies of the genetic code were carried out on the MANIAC.

On the lighter side Gamow was writing a book called *Mr. Tompkins Learns the Facts of Life.* Gamow was so fascinated by the MANIAC that he decided to include the machine in his book. So when Mr. Tompkins is to learn how the brain works, Gamow introduces him to MANIAC and has the computer explain how his electronic brain works. Figure 7 is an inimitable Gamow cartoon from this delightful book; notice how appealingly the essentials are presented.

Importance Sampling. A significant advance in the use of the Monte Carlo method came out of Nick's collaboration with Edward Teller. Teller, obviously delighted at gaining access to such a marvelous toy, proposed that the MANIAC, and the Monte Carlo method, be used to carry out calculations on the equation of state in two dimensions for hard spheres. These calculations introduced the idea of what is now known as importance sampling, also referred to as the Metropolis algorithm. The scheme, which reduces the statistical error and thereby greatly improves the effectiveness of the Monte Carlo method, is widely used today, as participants in this conference have made evident.

Two-Dimensional Hydrodynamics. I won't say too much about two-dimensional hydrodynamics because it was always a major problem at the Lab—and still is. But Johnny von Neumann was one of the great experts in the field, and he showed how to attack the two-dimensional flow of two incompressible fluids under gravitational and hydrodynamic forces.

Iterative Functions. A striking thing about these early computer experiments is that in some cases the importance of the work wasn't recognized at the time the work was done. It was immediately obvious that the pion-proton phase-shift analysis was going to be important. But when Paul Stein, Myron Stein, and Nick Metropolis began to look into the problem of iterative functions, it was mainly to satisfy their curiosity. Quite some time later, this problem unexpectedly turned out to be very important.

One of the surprising consequences of these early studies of iterative transformations was the discovery of their universal properties. This work was an inspiration to Mitchell Feigenbaum, who took it up a few years ago and used it to show how such functions lead to a theory of the onset of turbulence and chaotic behavior. This subject has turned out to be as important as it is exciting. It has fired the imaginations of many who are intrigued by the curious aspects of nonlinear behavior. It is currently being widely developed, a good example of computer-driven mathematics.

A Few Others. Another noteworthy study was the classic study of the nuclear cascades induced by bombarding heavy nuclei with high-energy particles. Nick did this study with Tony Turkevich, using Monte Carlo techniques.

Then Mark Wells and others prepared the first program to develop a strategy for "anti-clerical" chess. This game was played on a 6-by-6 board with bishops removed. It was a highly amusing experience and had many implications for subsequent games of strategy. Figure 8 shows MANIAC I, Paul Stein, Nick Metropolis, and the chessboard.

Finally, I must mention some interesting work in number theory in which Stan Ulam and others introduced the notion of "lucky numbers," a generalization of the ordinary prime numbers with many similar properties. That was an attractive piece of work.

This list of ten is just a small sample of the many scientific uses of the MANIAC, and it includes only the most prominent names associated with Nick on those projects. These I think were the most important works—or at any rate the most interesting.

Death of MANIAC

Before closing, let me make clear how the MANIAC evolved over the years. MANIAC II succeeded MANIAC I in Los Alamos in 1956. The second MANIAC was more powerful than the first and, because it included floating-point arithmetic, was easier to use. MANIAC III, with the latest in solid-state circuitry, was developed at the University of Chicago when Nick returned there to head the newly formed Institute for Computer Research.

In 1965 Nick returned to Los Alamos from Chicago, but by this time the computing needs of the Laboratory had increased dramatically and were being supplied by commercial machines. Unfortunately, in this climate the decision was eventually made to abandon the kind of original research that was special to the computer project, and in 1977 MANIAC II was turned off.

To conclude this story on a happier note, I remind you that it's Nick's birthday. I've always had the idea that when there is a birthday there ought to be a poem. Here is my own modest offering.

Monte Carlo Metropolis

Nick, you'll remember those halcyon days;
The Monte Carlo method was in its earliest phase.
You went to the ENIAC and got it to load
Problems you wanted in a stored-program mode.
The Monte Carlos you gave it opened a crack
That helped you decide to build MANIAC.
MANIAC came out as a marvelous toy,
A machine you could work with and really enjoy.

You called on your friends to join in the fun,
And it wasn't too long before they'd begun.
There was Teller, Gamow, Turkevich too.
Of others not mentioned there were quite a few.
But these in particular knew that their goal
Was through Monte Carlo in the MANIAC's soul.
Those were seminal papers, seeds had been sown;
That's how Monte Carlo came into its owı

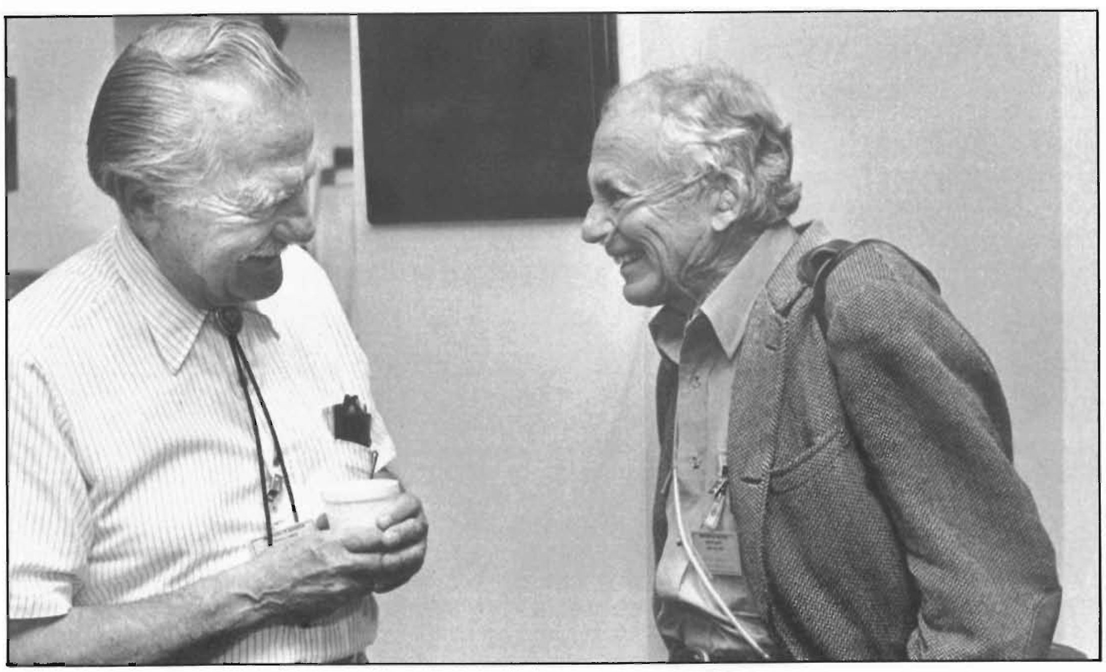


Photo by Fred Rick

Further Reading

Metropolis, N., Howlett, J., and Rota, Gian-Carlo, eds. 1980. *History of Computing in the Twentieth Century*. New York:Academic Press.

Burks, Arthur W., Goldstine, Herman H., and von Neumann, John. 1946. *Preliminary Discussion of the Logical Design of an Electronic Computing Instrument*. Princeton:Institute for Advanced Study.

Segrè, E., ed. 1965. *The Collected Works of Enrico Fermi*. Chicago:The University of Chicago Press.

Metropolis, Nicholas, Rosenbluth, Arianna W., Rosenbluth, Marshall N., Teller, Augusta H., and Teller, Edward. 1953. "Equation of State Calculations by Fast Computing Machines." *Journal of Chemical Physics* 21:1087.

Metropolis, Nicholas, and Ulam, S. 1949. "The Monte Carlo Method." *Journal of the American Statistical Association* 44:335.

Historical illustrations for this article were provided by N. Metropolis.

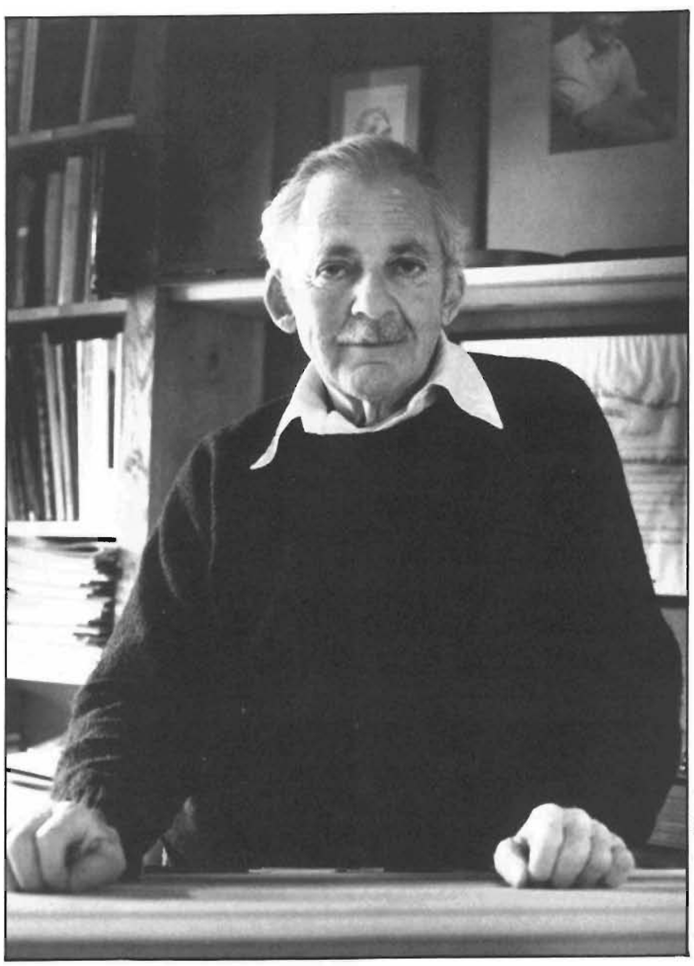


Photo by Fred Rick

Herbert L. Anderson received his Ph.D. in physics from Columbia University in 1940. As a graduate student he assisted in the construction of a 37-inch cyclotron at Columbia. He also began a close association with Enrico Fermi that lasted until Fermi's death in 1954. Anderson was the first in the United States to demonstrate the energy release in the fission of uranium in an ionization chamber/linear amplifier combination. He participated in the original experiments on neutron reproduction in uranium, which led directly to the development of the nuclear chain reaction. These experiments included the early studies on the fission of uranium, the emission of neutrons by uranium, the slowing down of neutrons in carbon, and neutron reproduction in a lattice of uranium and graphite. During 1942-44 he had a major role in the design and construction of the first chain reacting pile, in Chicago, the second (CP-2) pile, at Argonne, and the Hanford piles. At Los Alamos his experiments at the Omega reactor helped establish the critical size for nuclear explosions in uranium-235. His measurements at Alamogordo, New Mexico, in which he used fission-product analysis, established the yield of the first nuclear explosion in 1945. The method, adapted to air sampling, became a principal means for detecting and analyzing nuclear testing carried out by foreign countries. He returned to Chicago in 1945 as a member of the newly formed Institute for Nuclear Studies. He served as Director of the Enrico Fermi Institute from 1958-62. He was appointed Guggenheim Fellow in 1955-57 and Fulbright Lecturer in Italy in 1956-57. He was elected to the National Academy of Sciences in 1960 and to the American Academy of Arts and Sciences in 1978. In 1982 he received the Enrico Fermi Award. During the years in Chicago, he was a consultant and Visiting Fellow at Los Alamos, and in 1978 he joined the Laboratory as a staff member. He was appointed a Los Alamos National Laboratory Senior Fellow in 1981 and Distinguished Service Professor Emeritus at the University of Chicago in 1982. In the Physics Division at Los Alamos, he is currently working with an instrument of his own design to analyze the proteins made by living cells. The proteins are separated by two-dimensional electrophoresis, and the protein analyzer measures them by direct beta-ray counting. In a collaboration with biologist Theodore T. Puck, he is trying to identify the proteins specific to three human chromosomes.